Progress in Molecular and Subcellular Biology

Đurđica Ugarković
Editor

Long Non-Coding RNAs

Springer

Progress in Molecular and Subcellular Biology

51

Volumes Published in the Series

Progress in Molecular and Subcellular Biology

Subseries: Marine Molecular Biotechnology

Volume 35
RNA Trafficking and Nuclear Structure Dynamics
Ph. Jeanteur (Ed.)

Volume 36
Viruses and Apoptosis
C. Alonso (Ed.)

Volume 37
Sponges (Porifera)
W.E.G. Müller (Ed.)

Volume 38
Epigenetics and Chromatin
Ph. Jeanteur (Ed.)

Volume 39
Echinodermata
V. Matranga (Ed.)

Volume 40
Developmental Biology of Neoplastic Growth
A. Macieira-Coelho (Ed.)

Volume 41
Molecular Basis of Symbiosis
J. Overmann (Ed.)

Volume 42
Antifouling Compounds
N. Fusetani and A.S. Clare (Eds.)

Volume 43
Molluscs
G. Cimino and M. Gavagnin (Eds.)

Volume 44
Alternative Splicing and Disease
Ph. Jeanlevr (Ed.)

Volume 45
Asymmetric Cell Division
A. Macieira Coelho (Ed.)

Volume 46
Marine Toxins as Research Tools
N. Fusetani and W. Kem (Eds.)

Volume 47
Biosilica in Evolution, Morphogenesis, and Nanobiotechnology
W.E.G. Müller and M.A. Grachev (Eds.)

Volume 48
Centromere
Đurđica Ugarković (Ed.)

Volume 49
Aestivation
C.A. Navas and J.E. Carvalho (Eds.)

Volume 50
miRNA Regulation of the Translational Machinery
R.E. Rhoads (Ed.)

Volume 51
Long Non-Coding RNAs
Đurđica Ugarković (Ed.)

Đurđica Ugarković
Editor

Long Non-Coding RNAs

Springer

Editor
Đurđica Ugarković
Ruđer Bošković Institute
Bijenička 54
10001 Zagreb
P.O. Box 1016
Croatia

ISSN 0079-6484
ISBN 978-3-642-16501-6 e-ISBN 978-3-642-16502-3
DOI 10.1007/978-3-642-16502-3
Springer Heidelberg Dordrecht London New York

Cover design: SPi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

A major portion of the eukaryotic genome is occupied by DNA sequences whose transcripts do not code for proteins. This part of eukaryotic genome is transcribed in a developmentally regulated manner or as a response to external stimuli to produce large numbers of long noncoding RNAs (lncRNAs). Genome-wide studies indicate existence of more than 3,300 lncRNAs. Long ncRNAs are tentatively defied as molecules of ncRNA more than two hundred nucleotides long. Due to the complexity and diversity of their sequences and their mechanisms of action, progress in the field of lncRNAs has been very slow. Nonetheless, lncRNAs have emerged as key molecules involved in the control of transcriptional and posttranscriptional gene regulatory pathways. Although limited numbers of functional lncRNAs have been identified so far, the immense regulatory potential of lncRNAs is already evident, emphasizing that a genome-wide characterization of functional lncRNAs is needed. Here, we review this rapidly advancing field of long ncRNAs, describing their structures, organization, and function in diverse eukaryotic systems.

Although the evidence for diverse biological functions of lncRNAs exists across the wide evolutionary spectrum, the underlying molecular mechanisms are far from clear. In Chap. 1 of this book, Radha Raman Pandey and Chandrasekhar Kanduri discuss the epigenetic and nonepigenetic mechanisms by which lncRNAs regulate various biological functions in model systems, from yeast to mammals. Long ncRNA molecules take part in gene regulation from the single gene level to an entire chromosome via recruitment of chromatin-modifying complexes *in cis* or *trans*. At the posttranscriptional level, lncRNAs regulate the splicing, localization, stability, and translation of the target mRNAs by base-pairing with their target RNAs. Transcriptional repression is mainly done by long noncoding RNAs in contrast to translational repression executed mostly by short noncoding RNA. In Chap. 2, Riki Kurokawa overviews the recent publications regarding the transcription regulation by long ncRNAs. In addition, the relation between a random transcriptional activity of RNA polymerase II and the origin of long ncRNAs is discussed.

In mammalian female somatic cells, one of the two X chromosomes is inactivated, and in the last few decades, several *cis*- and *trans*-acting factors involved in the regulation of the X chromosome inactivation process have been identified. The two main regulatory factors are *Xist* and *Tsix* that both encode functional lncRNAs. In Chap. 3, Joost Gribnau and collaborators describe the current knowledge about the structure and function of *Xist* and discuss the important *cis*- and *trans*-regulatory elements and proteins in the X chromosome inactivation. The authors also highlight new findings with other ncRNAs involved in gene repression and discuss these findings in relation to *Xist*-mediated gene silencing.

Telomeres protect the ends of linear eukaryotic chromosomes from being recognized as DNA double-stranded breaks, thereby maintaining the genome stability. The highly heterochromatic nature of telomeres had, for a long time, reinforced the idea that telomeres were transcriptionally silent. In 2007, the longstanding dogma that telomeres are transcriptionally silent was overturned by the discovery that noncoding RNA molecules, named *TE*lomeric *R*epeat-containing *RNA* (TERRA), were found to emanate from and associate with telomeres. In Chap. 4, Claus M. Azzalin and collaborators provide an overview of telomere structure, function, and biology and extensively review the current knowledge about TERRA biogenesis, regulation, and potential functions.

In eukaryotic cells, correct segregation and inheritance of genetic information relies on the activity of specialized chromosomal regions called centromeres. Centromeric and pericentric regions have long been regarded as transcriptionally inert; however, a number of studies in the past 10 years provided convincing evidence that centromeric and pericentric sequences are transcriptionally active. In Chap. 5, Claire Vourc'h and Giuseppe Biamonti review the expression of these sequences in mouse and human cells and discuss the possible functional implications of centromeric and pericentric sequences activation and/or of the resulting noncoding RNAs. An overview of the molecular mechanisms underlying the activation of centromeric and pericentromeric sequences is provided.

Alu elements are the most abundant repetitive elements in the human genome and, recently, it has become evident that they play crucial and diverse roles in regulating gene expression. Audrey Berger and Katharina Strub in Chap. 6 review role of Alu and *Alu*-related RNAs in regulation of transcription and translation. Transcription from these elements occurs at low levels under normal conditions but increases transiently after stress, indicating a function of *Alu* RNA in cellular stress response. Alu elements provide a source for the biogenesis of miRNAs and, when embedded into mRNAs, can be targeted by miRNAs. Certain *Alu* elements evolved into unique transcription units with specific expression profiles producing RNAs with highly specific cellular functions.

The large noncoding *roX* RNAs have a central role in sex chromosome dosage compensation in flies, where they fulfill a role with similarities to that of *Xist* during mammalian dosage compensation. In Chap. 7, S. Kiran Koya and Victoria H. Meller summarize the current knowledge of the function of the noncoding *roX* genes in the process of dosage compensation in *Drosophila*. The unexpected discovery of a role for *roX* in the expression of heterochromatic genes is discussed.

Satellite DNAs are major heterochromatin constituents in many insect species found to be transcribed during all developmental stages. Transcripts play a role in heterochromatin establishment and regulation, although the detailed molecular mechanism and proteins involved are not elucidated yet. The satellite DNA transcription is associated with development and differentiation and is actively regulated by environmental factors such as temperature. In Chap. 8, Đurđica Ugarković and collaborators review the transcription of satellite DNAs in different insects. They also discuss the role of satellite DNA transcripts in regulation of heterochromatic genes as well as genes located in the vicinity of satellite DNA elements within euchromatin.

In contrast to small RNAs, much less is known about the large and diverse population of long noncoding RNAs in plants, and only few have been implicated in diverse functions such as abiotic stress responses, nodulation and flower development, and sex chromosome-specific expression. Moreover, many long noncoding RNAs act as antisense transcripts or are substrates of the small RNA pathways interfering with a variety of RNA-related metabolisms. As plants show a remarkable developmental plasticity to adapt their growth to changing environmental conditions, understanding how ncRNAs work may reveal novel mechanisms involved in growth control and differentiation. In Chap. 9, Virginie Jouannet and Martin Crespi discuss a major class of long noncoding RNAs and antisense transcripts in plants. They also introduce long noncoding RNAs interacting with specific RNA-binding proteins to modulate their action or localization.

Zagreb, Croatia Đurđica Ugarković

Contents

Contributors

Rajika Arora Institute of Biochemistry, ETHZ-Eidgenössische Technische Hochschule Zürich, CH-8093, Zürich, Switzerland

Claus M. Azzalin Institute of Biochemistry, ETHZ-Eidgenössische Technische Hochschule Zürich, CH-8093, Zürich, Switzerland, claus.azzalin@bc.biol.ethz.ch

Audrey Berger Department of Cell Biology; University of Geneva, 30 quai Ernest Ansermet, 1211 Geneva 4, Switzerland

Giuseppe Biamonti Istituto di Genetica Molecolare CNR, Via Abbiategrasso 207, 27100 Pavia, Italy, biamonti@igm.cnr.it

Josip Brajković Ruđer Bošković Institute, Bijenička 54, 10001 Zagreb, P.O. Box 1016, Croatia

Catherine M.C. Brun Institute of Biochemistry, ETHZ-Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland

Martin Crespi Centre National de la Recherche Scientifique; Institut des Sciences du Végétal, 91198 Gif-sur-Yvette Cedex, France, martin.crespi@isv.cnrs-gif.fr

Isidoro Feliciello Ruđer Bošković Institute, Bijenička 54, 10001 Zagreb, P.O. Box 1016, Croatia, Dipartimento di Medicina Clinica e Sperimentale; Università degli Studi di Napoli Federico II, via Pansini 5, 80131 Napoli, Italy, martin.crespi@isv.cnrs-gif.fr

Cristina Gontan Department of Reproduction and Development; Erasmus MC, University Medical Center, Room Ee 09-71, PO Box 2040 3000 CA Rotterdam, The Netherlands

Joost Gribnau Department of Reproduction and Development; Erasmus MC, University Medical Center, Room Ee 09-71, PO Box 2040, 3000 CA Rotterdam, The Netherlands, j.gribnau@erasmusmc.nl

Iris Jonkers Department of Reproduction and Development; Erasmus MC, University Medical Center, Room Ee 09-71, PO Box 2040, 3000 CA Rotterdam, The Netherlands, Department of Molecular Biology and Genetics; Cornell University, Ithaca, NY, USA

Virginie Jouannet Department of Stem Cell Biology; University of Heidelberg, INF230, 69120 Heidelberg, Germany, Centre National de la Recherche Scientifique; Institut des Sciences du Végétal, 91198 Gif-sur-Yvette Cedex, France

Chandrasekhar Kanduri Department of Genetics and Pathology, Rudbeck Laboratory; Uppsala University, Dag Hammarskjölds Väg 20, 75185 Uppsala, Sweden, Kanduri.Chandrasekhar@genpat.uu.se

S. Kiran Koya Department of Biological Sciences; Wayne State University, 5047 Gullen Mall, Detroit, MI 48202, USA, sk_koya@wayne.edu

Riki Kurokawa Division of Gene Structure and Function, Research Center for Genomic Medicine; Saitama Medical University, 1397-1 Yamane, Hidaka-shi, Saitama-Ken 350-1241, Japan, rkurokaw@saitama-med.ac.jp

Victoria H. Meller Department of Biological Sciences; Wayne State University, 5047 Gullen Mall, Detroit, MI 48202, USA, meller@biology.biosci.wayne.edu

Radha Raman Pandey Department of Genetics and Pathology, Rudbeck Laboratory; Uppsala University, Dag Hammarskjölds Väg 20, 75185 Uppsala, Sweden

Željka Pezer Ruđer Bošković Institute, Bijenička 54, 10001 Zagreb, P.O. Box 1016, Croatia

Katharina Strub Department of Cell Biology; University of Geneva, 30 quai Ernest Ansermet, 1211 Geneva 4, Switzerland, Katharina.Strub@unige.ch

Đurđica Ugarković Ruđer Bošković Institute, Bijenička 54, 10001 Zagreb, P.O. Box 1016, Croatia

Claire Vourc'h Université Joseph Fourier-Grenoble; INSERM U823; Institut Albert Bonniot, La Tronche BP170, 38042 Grenoble cedex 9, France, claire.vourch@ujf-grenoble.fr

Chapter 1
Transcriptional and Posttranscriptional Programming by Long Noncoding RNAs

Radha Raman Pandey and Chandrasekhar Kanduri

Abstract Recently, several lines of evidence have suggested that noncoding RNAs, which include both small and long noncoding RNAs (ncRNAs), contribute to a significant portion of the transcriptome in eukaryotic organisms. However, the functional significance of this wide-spread occurrence of ncRNAs, and in particular, the long ncRNAs (lncRNAs), for organismal development and differentiation is unclear. The available evidence from a subset of lncRNAs suggests that certain lncRNAs, and/or the act of their transcription, are involved in important biological functions at the transcriptional and posttranscriptional level. This chapter discusses the epigenetic and nonepigenetic mechanisms by which lncRNAs and/or their transcription are involved in the programming of various biological functions in model systems, from yeast to mammals.

1.1 Introduction

A major portion of the eukaryotic genome is occupied by DNA sequences, whose transcripts do not code for proteins. It has been proposed that the size of the noncoding portion of the genome is linked to the development of complex organisms (Mattick 2004; Taft et al. 2007), as the protein-coding portion of the genome, by and large, has remained constant while the noncoding portion has grown significantly during the evolution of more complex organisms from simpler life-forms (Mattick 2004). This hypothesis indicates that these sequences are not "junk" but perhaps play a major role in the generation of organismal complexity. In the initial attempt to define the mouse transcriptome by sequencing of mouse full-length cDNA clones, it was found that the majority of the nonprotein-coding DNA region is transcribed but produces RNA with little or no protein-coding potential

R.R. Pandey and C. Kanduri (✉)
Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Dag Hammarskjölds Väg 20, 75185 Uppsala, Sweden
e-mail: Kanduri.Chandrasekhar@genpat.uu.se

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3_1,

(Okazaki et al. 2002; Carninci et al. 2005). Moreover, the development of new highly sensitive and ultra high-throughput techniques such as second generation sequencing in combination with preexisting classical molecular biology techniques such as CAGE (Cap Analysis of Gene Expression) (Shiraki et al. 2003), 5′ and 3′ SAGE (Serial Analysis of Gene Expression) (Velculescu et al. 1995), ASSAGE (Asymmetric Strand-specific Analysis of Gene Expression) (He et al. 2008), and GRO (Global Run On Analysis) (Core et al. 2008) have provided us with a detailed overview of the extent of transcription in eukaryotes (Nagalakshmi et al. 2008). The results were surprising in that most of the eukaryotic genome is transcribed and produces a plethora of noncoding RNA (ncRNA) species during various stages of cellular differentiation (Kapranov et al. 2007a, b and references therein; Birney et al. 2007).

A ncRNA is defined as an RNA species with an open reading frame (ORF) of less than 100 amino acids, whereas protein-coding mRNAs have ORFs greater than 100 amino acids in length. Some of the ncRNAs are constitutively expressed in all cells, for example, ribosomal RNA, transfer RNA, and small nuclear and nucleolar RNA (snRNA, SnoRNA), and are hence known as housekeeping ncRNAs. The functions and mechanisms of action of the housekeeping ncRNAs have been investigated in greater detail in recent years. The ncRNAs, other than housekeeping ncRNAs, are broadly categorized into small ncRNAs (less than 100 nucleotides in length) and lncRNAs, which are longer than 200 nucleotides in length. The small ncRNAs are further divided into subgroups (miRNA, siRNA, piRNA, etc.) depending on their size, biogenesis, mode of action, and the proteins with which they are associated. Small ncRNAs regulate gene expression at the transcriptional level by guiding the repressive chromatin complexes known as RNA-induced transcriptional silencing and RNA-dependent RNA polymerase complexes (RITS-RDRCs) to cognate genes, and at the posttranscriptional level by guiding the effector complexes known as RNA-induced silencing complexes (RISCs) either to cleave the target mRNA or to bring about translational inhibition (Bartel 2004, 2009; Grewal and Jia 2007; Malone and Hannon 2009; Ghildiyal and Zamore 2009).

The lncRNAs are the least characterized of all the ncRNAs whose biological functions are, in any case, poorly investigated. The majority of the lncRNAs are transcribed by RNA polymerase II (RNA pol II) and possess a 5′ methyl cap and polyA tail. Depending on their location, with respect to the mRNA gene, they can be classified as (1) Sense, transcribed from the same strand as the mRNA; (2) Antisense, transcribed from the strand opposite the mRNA; (3) Intronic, the transcription unit of the lncRNA lies within an intron of another gene; and (4) Intergenic, transcribed from a region lying outside mRNA genes. Several thousand lncRNAs are predicted to be present in the eukaryotic genome; however, at present, the most difficult issue is the identification of functional lncRNAs from the vast pool of pervasively transcribed noncoding transcripts.

There is a possibility that a significant number of lncRNAs could arise from experimental artifacts. For example, genome tiling array experiments in different organisms reported thousands of *cis* natural antisense transcripts (*cis* NATs)

(Yamada et al. 2003; Bertone et al. 2004; Carninci et al. 2005; David et al. 2006; Samanta et al. 2006). However, a more recent study could only find less than half of the *cis* NATs in yeast when actinomycin D was included in the cDNA synthesis reaction to prevent false second strand synthesis (Perocchi et al. 2007), indicating that experimental artifacts could have contributed to the number of noncoding transcripts. In addition, many of the intronic lncRNAs could be fragments derived from the splicing of pre-mRNAs. Similarly, a large proportion of the intergenic transcripts could arise from the ripple effect of nearby transcription, which induces changes in nucleosome organization, thus providing an opportunity for the transcription machinery to produce transcripts of no significance from cryptic promoters (Ebisuya et al. 2008).

LncRNAs show a very low level of sequence conservation compared to protein-coding mRNAs. Nevertheless, the base substitution rate or constraint (ratio of the nucleotide substitution rate between functional sequences and neutral sequences) for ncRNAs is 90–95%, which is fairly high when compared with protein-coding sequences but still shows positive selection over the neutral sequences in the genome (Ponjavic et al. 2007), indicating that ncRNAs do possess important biological functions. The observations that lncRNAs display subcellular localization (Mercer et al. 2008), tissue- and cell type-dependent expression, specific expression in response to certain environmental cues (Cawley et al. 2004), and transcriptional regulation by key transcription factors such as p53, c-MYC, SP1 (Cawley et al. 2004), and CREB (Euskirchen et al. 2004) further emphasize that lncRNAs could play critical roles in cellular proliferation, differentiation, and the development of complex organisms.

Recently, several different approaches have been used to identify functional lncRNAs. In one approach, several hundred long intervening ncRNAs (lincRNAs) were identified using active chromatin signatures associated with RNA pol II transcription, i.e., the histone H3 lysine 4 trimethylation and histone H3 lysine 36 trimethylation domains (K4-K36 domains) (Guttman et al. 2009; Khalil et al. 2009). The studies identified 1,586 and 3,289 lincRNAs in different mouse and human cell types, respectively, and predicted that the total number of lincRNAs could be around 4,500. The lincRNAs show significant evolutionary conservation when compared to neutral sequences in the genome and many of them show changes in their expression patterns in response to different environmental stimuli, suggesting that lincRNAs could play critical roles in various biological functions (Guttman et al. 2009; Khalil et al. 2009). In another recent study, around 215 functional lncRNAs were identified based on their chromatin interaction properties (Mondal et al. 2010). The chromatin associated RNAs (CARs) also show significant evolutionary conservation and transcribed from both intronic and intergenic regions. Functional characterization of one of the CARs revealed that they regulate gene expression by regulating chromatin structure. Collectively, the above observations suggest that lncRNAs are an integral component of mammalian genetic programming.

Although the functional roles of lncRNAs are very much in evidence in diverse biological functions across the evolutionary spectrum (Bernstein and Allis 2005;

Mattick and Makunin 2006; Prasanth and Spector 2007; Amaral et al. 2008; Amaral and Mattick 2008; Sunwoo et al. 2009, and references therein), the underlying molecular mechanisms are far from clear. In this chapter, we discuss the epigenetic and nonepigenetic mechanisms by which lncRNAs regulate various biological functions in model systems, from yeast to mammals.

1.2 Pervasive ncRNA Transcription at Gene Regulatory Regions and the Link to Transcription

Several high-throughput approaches have uncovered widespread pervasive transcription across the promoter and terminator regions of annotated genes in yeast, mice, humans, and plants, which produce a complex repertoire of noncoding transcripts. These transcripts include small RNAs [miRNA, piRNA, and siRNA] as well as lncRNAs. A recent study (Kapranov et al. 2007a), aimed at profiling human and mouse transcriptomes from cell lines, used polyA+ RNA, longer than 200 nucleotides (nt), from nuclear and cytoplasmic fractions separately, and total cellular RNA of less than 200 nt in length to hybridize to tiling arrays at 5-nucleotide resolution. The study found three different RNA species: promoter-associated small RNAs (PASRs), terminator-associated small RNAs (TASRs), and promoter-associated long RNAs (PALRs). The PASRs and TASRs ranged in size from between 20 and 200 nt; however, a significant number of PASRs were between 26 and 50 nt long. PASRs were centered around the transcription start site of protein-coding genes in both directions, whereas TASRs were mostly oriented in the antisense direction at the 3′ termini of the host genes. This study further demonstrated that PASRs and TASRs are also present in mouse at the 5′ and 3′ ends of genes, respectively, indicating that these RNAs are highly conserved across the evolutionary spectrum and could have a potential role in gene regulation. PALRs are 100 nt to 1.0 kb long and map to 5′ regulatory regions, like PASRs, which suggests that many PASRs could be derived from PALRs. However, in the majority of cases, the expression of PASRs and TASRs is strongly correlated with the associated gene expression. The genes that were found to be highly enriched for PASRs and TASRs were also highly expressed and vice versa Kapranov et al. 2007a. PASRs are not produced by the Dicer-dependent cleavage mechanism as the PASR profile in mouse ES cells lacking Dicer remained unchanged (Kapranov et al. 2007a).

In addition to PASRs and TASRs, another category of highly unstable small and long ncRNAs, located close to promoters in yeast and human cells, have been described. In the budding yeast, *Saccharomyces cerevisiae*, these transcripts were upregulated in a mutant, which lacked components of the exosome machinery, and were therefore christened cryptic unstable transcripts (CUTs) (Xu et al. 2009, Wyers et al. 2005). The exosome is known to act as a surveillance pathway for the removal of unwanted RNA molecules from cells. The 3′ SAGE sequencing of CUTs peaked at

50 nt downstream and 550 nt upstream of known open reading frame (ORF) transcription start sites (TSSs). Since the average size of CUTs is around 250–300 nt, it can be concluded that they mostly originate from intergenic regions (Neil et al. 2009). CUTs are transcribed in both divergent and convergent configurations, but the former contributes to the most abundant class. To date, the functional significance of CUTs in various biological functions is still unclear.

Similar to CUTs in yeast, a subclass of promoter upstream transcripts (PROMPTs) were stabilized when HeLa cells were treated with an siRNA to knockdown hRrp40, a crucial component of the human 3′–5′ exoribonucleolytic exosome (Brower et al. 2001). PROMPTs can originate more than 2.0-kb upstream of the TSS with a peak around −1.0 kb. PROMPTs are transcribed in both the sense and antisense directions with respect to the TSS of the associated gene (Preker et al. 2008). The function of PROMPTs is largely unknown, but they may play a regulatory role since certain ncRNAs, known to exert regulatory functions, are located within PROMPT regions. Interestingly, one of the ncRNAs, *Khps1*, which is transcribed in the antisense direction from the TSS of sphingosine-kinase 1 (*SPHK1*), is stabilized in hRrp40-knockdown cells. The *Khps1* transcript has been linked to the demethylation of the *SPHK1* differentially methylated region (DMR) (Imamura et al. 2004); however, the mechanism by which *Khps1* mediates demethylation is not known. Taking the data from yeast, mouse, and human together, it is clear that the divergent transcription of ncRNAs surrounding the promoter regions of annotated genes is a common and conserved feature of eukaryotic RNA pol II transcription. This is demonstrated further by the broad distribution of RNA pol II near TSSs and by the bimodal distribution of active chromatin markers such as histone H3 lysine 4 trimethylation.

Several models have been proposed for the biogenesis of pervasive transcripts at gene regulatory regions. The TSSs for most of the promoter- and terminator-associated ncRNAs fall within the nucleosome-free region (NFR) of the related genes, suggesting that perhaps they originate from the spurious activity of RNA pol II on naked DNA in the promoter, as well as the terminator regions. Nucleosome positioning is known to suppress cryptic transcription by preventing the random access of RNA polymerase to the DNA. This is clearly demonstrated in yeast containing mutations in the *spt6* gene, where the ability to reassemble nucleosomes is lost in the RNA Pol II-elongated portions of coding regions, resulting in cryptic transcription from the NFRs (Cheung et al. 2008). Moreover, insertion of an enhancer with several LexA or Gal4 binding sites induced an NFR around the site of insertion, irrespective of the genomic location, leading to cryptic transcription from the 3′ ends of the LexA/Gal4 binding sites (Dobi and Winston 2007). Likewise, a very recent study using chromatin signatures specific to enhancer and promoter found that most of the extragenic RNA Pol II peaks overlapped the enhancer regions, indicating that long noncoding transcription is prevalent in the enhancer regions (De Santa et al. 2010). These examples clearly suggest that nucleosome positioning is critical for preventing aberrant transcription across the genome. Moreover, the majority of promoter- and terminator-associated RNAs are less abundant than protein-coding mRNA and rapidly degraded by nuclear quality control pathways in both yeast and human (Preker et al. 2008; Wyers et al. 2005),

indicating that they might possibly represent the by-products of RNA pol II spurious activity in NFR regions. However, the presence of an independent TSS for PASRs, TASRs, and PALRs, and the fact that they are conserved across the evolutionary spectrum, suggests that they are not by-products of RNA pol II spurious activity in NFR regions. Additionally, in yeast, a mutation in the TATA box of the *TPI1* gene affected expression of the mRNA but not of the sense CUT, further supporting the notion that CUTs originate from the assembly of an independent preinitiation complex (PIC) and substantiating their functional role in gene regulation (Neil et al. 2009).

The key question here is, "what is the role of pervasive transcription?" Since promoter- and terminator-associated transcripts are rapidly degraded, the transcript per se may not be directly involved in the gene regulatory process. Interestingly, the expression of promoter-associated RNAs in human cells (PASRs and PROMPTs), as well as in yeast (CUTs), correlates positively with the expression of sense mRNAs. However, when several synthetic sense and antisense PASRs, surrounding the *c-MYC* and connective tissue growth factor (*CTGF*) promoters, were transfected into HeLa cells (Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project 2009), the mRNA levels of both the *c-MYC* and *CTGF* genes were downregulated, in contrast to data suggesting a genome-wide positive correlation of PASRs with mRNA gene expression. This may explain why PASRs in human and CUTs in yeast are rapidly degraded by the exosome machinery.

Interestingly, a couple of recent investigations have further implicated PASRs in the negative regulation of cognate genes. For example, intergenic spacer regions in ribosomal gene clusters encode lncRNAs, whose promoters lie about 2.0-kb upstream of the rRNA promoters. In addition to the 2.0-kb lncRNAs, the spacer regions also contain 150–200 nt RNAs (pRNAs), which span the *rRNA* promoters, indicating that the pRNAs could be derived from the spacer lncRNAs. The pRNAs have been shown to interact with and recruit the nucleolar remodeling complex (NoRC) to *rRNA* gene promoters, and this, in turn, leads to the recruitment of components of the heterochromatin machinery, including HP1 (Mayer et al. 2006).

Like *rRNA* gene promoters, the *p21* promoter also contains promoter-associated RNAs in both sense and antisense directions. Interestingly, the generation of antisense promoter-associated RNAs, which correlates with the silencing and heterochromatinization of the *p21* sense promoter, is dependent on transcription from the *p21* antisense promoter in an Ago-1-dependent manner. This indicates that antisense pRNAs could be derived from the *p21* antisense RNA and play a critical role in the transcriptional silencing of the *p21* sense promoter (Kim et al. 2006; Morris et al. 2008). Alternatively, the transcription of PASRs and CUTs may be involved in establishing an open chromatin configuration, which would be required for high-level mRNA gene expression, or they could act as rheostats involved in maintaining a specific level of mRNA expression by competing for the same pool of transcription factors. This has been shown at least in the case of one antisense CUT promoter, where a mutation in the promoter of the *TPI1* mRNA gene resulted in several fold higher expression of the antisense CUT (Neil et al. 2009). Although there is no genome-wide study yet available to describe the function of the 5′ and 3′ associated small and long ncRNAs in the regulation of mRNA genes, several

studies covering individual CUTs/PARs highlight that different mechanisms are being used to control mRNA gene expression at various levels.

1.3 Transcriptional Silencing by Noncoding Transcription via Transcriptional Interference

Transcriptional interference (TI) refers to the suppressive effect of one transcriptional event on a second transcriptional event *in cis*. TI occurs when two promoters are convergent or in tandem. The elongating complex from one promoter can affect the transcriptional initiation (by interfering with preinitiation complex assembly), elongation, or termination step of the second promoter, depending on its physical relationship with the first promoter. For example, the first promoter only affects PIC assembly when the second promoter is in tandem, but can affect PIC assembly, transcriptional elongation or termination when the second promoter is transcribed convergently. Although few eukaryotic genes have been shown to be regulated by a transcriptional interference mechanism involving lncRNA transcription, the observation that most protein-coding genes in higher eukaryotes have overlapping transcription from promoters in the upstream intergenic region or from downstream intragenic sense and antisense promoters, suggesting that transcriptional regulation by TI could be a common mechanism for regulating protein-coding genes. Here, we provide the biological contexts in which noncoding transcription regulates protein-coding genes via TI.

In the yeast *S. cerevisiae*, a gene involved in the serine biosynthesis pathway, *SER3*, is transcribed in nutrient-poor media; however, in nutrient-rich media, the *SER3* gene is silenced due to the activation of a noncoding RNA gene promoter *SRG1*, located upstream of the *SER3* gene. In the presence of serine in the nutrient-rich media, a serine-dependent activator, Cha4, along with chromatin remodeling complexes such as SAGA and SWI/SNF, binds to the *SRG1* promoter to activate its transcription (Martens et al. 2004; 2005) across the *SER3* promoter, leading to repression of the *SER3* gene. Promoter competition for basal transcription factors is not involved in *SER3* transcriptional repression, as the incorporation of a transcription termination signal for the *SRG1* transcript, upstream of the *SER3* promoter, resulted in derepression of *SER3*. This indicated that it is not the *SRG1* ncRNA but its transcription across the *SER-3* promoter that is required for its transcriptional repression. More importantly, it has been shown that *SRG1* transcription across the *SER-3* promoter interferes with the binding of transcription factors (Fig. 1.1aI), resulting in *SER3* gene silencing (Martens et al. 2004).

The inhibition of transcriptional initiation and elongation as means of cell type-specific gene regulation by overlapping antisense ncRNA transcription is beautifully illustrated in the diploid and haploid cells of the budding yeast, *S. cerevisiae*. In nutrient-rich media, *S. cerevisiae* cells divide mitotically to produce more diploid cells, whereas during starvation, the yeast undergoes meiotic division to produce

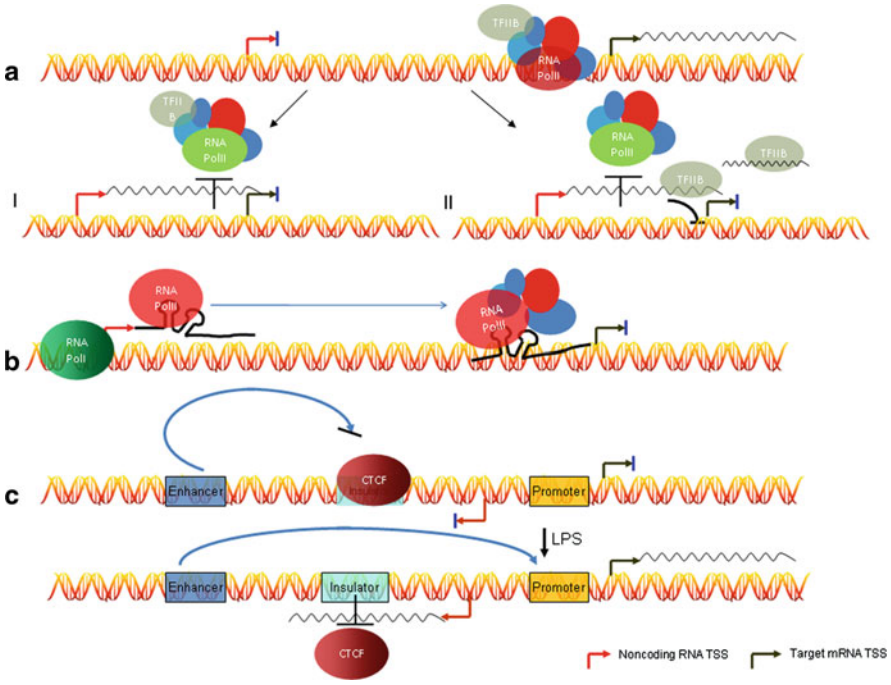

Fig. 1.1 Transcriptional silencing by lncRNAs via transcriptional interference. (**aI**), The transcription of a ncRNA through the promoter region of a target gene causes the occlusion of basal transcription machinery, thus repressing the transcription of the target gene. (**aII**) The lncRNA from DHFR minor promoter binds to TFIIB and titrates away the components of the preinitiation complex (PIC) from the DHFR major promoter. (**b**) The *Alu* and *B2* ncRNAs possess a modular structure, which includes two domains: an RNA pol II binding domain and a transcriptional inhibitory domain, which inhibits the transcription initiation step. The *Alu* and *B2* RNA inhibitory domains do not interfere with the binding of transcription factors to the ncRNAs but inhibit formation of the proper contact between RNA pol II and the DNA promoter elements required for the initiation of transcription. (**c**) In the chicken Lysozyme gene, CTCF target sites maintain silencing of the Lysozyme gene by preventing the communication of the upstream enhancer elements with the downstream Lysozyme promoter. In response to proinflammatory signals such as lipopolysaccharide (LPS), the lncRNA, LINoCR, transcription is activated across the CTCF target sites, resulting in the eviction of CTCF from its target site and activation of downstream Lysozyme promoter

haploid cells. This event is controlled by several genes, including *IME4* (initiator of meiosis). In diploid cells, only *IME4* sense mRNA was detected, whereas in haploid cells, an antisense ncRNA to the *IME4* gene was discovered, indicating that both sense and antisense *IME4* RNAs can affect each other's transcription (Hongay et al. 2006). Moreover, the separation of otherwise overlapping sense and antisense *IME4* transcription units resulted in the loss of the reciprocal effect on transcription, indicating that TI could be the mechanism in common between the sense and antisense transcriptional silencing effects *in cis*.

NcRNA transcription is not always involved in the repression of overlapping genes; sometimes it is engaged in activation of the associated gene by interfering with the binding of repressor complexes such as the chromatin insulator protein CTCF, which is known to function as a transcriptional repressor or an enhancer blocker (Kanduri et al. 2002; Phillips and Corces 2009, and references therein). The lysozyme gene in chicken has three enhancers at 2.7, 3.9, and 6.1 kb upstream of the TSS and is induced in response to proinflammatory signals such as lipopolysaccharide (LPS) in a chicken macrophage cell line. The silencing of the lysozyme gene is maintained by CTCF, whose target site maps to the region between the enhancers and the lysozyme promoter (Fig. 1.1c). The LPS induction of macrophages results in transcription of an ncRNA, LINoCR (LPS induced noncoding RNA). The transcription of LINoCR through CTCF target sites results in expulsion of the CTCF protein due to the positioning of a nucleosome over the CTCF target site (Lefevre et al. 2008). The expulsion of CTCF, and chromatin remodeling by LINoCR transcription, which further inhibits the binding of CTCF to its target site, facilitates enhancer/promoter communication, leading to lysozyme gene activation in response to the LPS proinflammatory signal (Fig. 1.1c).

Intriguingly, the interplay between the transcriptional processes of two intergenic noncoding transcription units in *S. cerevisiae* determines the transcriptional activity of the neighboring *FLO11* protein-coding gene (Bumgarner et al. 2009). *FLO11*, which encodes a cell-wall glycoprotein controlling cell–cell adhesion, has a variegated expression pattern; in some cells the gene is highly expressed, while in the other cells, it is completely repressed. This variegated or binary expression is the result of functional interplay between two *cis*-interfering lncRNAs, upstream of the *FLO11* gene. The 5′ regulatory region of *FLO11* is fairly long (3.4 kb) and harbors binding sites for several transcription factors, such as Sfl1 and Flo8, which overlap the two lncRNAs transcribed from opposite strands (Bumgarner et al. 2009). One of the ncRNAs, *ICR1* (Interfering Crick RNA), is transcribed from the same strand as *FLO11* and runs across the *FLO11* promoter, causing repression of the *FLO11* gene by the promoter occlusion mechanism. The second ncRNA, *PWR1* (Promoting Watson RNA), is transcribed from the complementary strand of *ICR1* and passes through its promoter, causing repression of the *ICR1* ncRNA and indirectly activating *FLO11* transcription. The transcription of *PWR1* is highly regulated. The Flo8 transcription factor specifically activates *PWR1*, resulting in the silencing of *ICR1* and, as a consequence, derepression of the *FLO11* gene. On the other hand, the transcriptional inhibitor, Sfl1, represses the *PWR1* promoter, causing repression of the *FLO11* gene via derepression of the ICR1 promoter, presumably by interfering with the binding of the transcriptional initiation machinery (Bumgarner et al. 2009). This is a very interesting example of how the interplay between two functional intergenic ncRNAs determines the activity of flanking protein-coding mRNA, and it highlights the fact that ncRNA-mediated transcriptional regulatory mechanisms are multilayered and highly complex.

Recent evidence suggests that gene regulation via TI constitutes one of the significant gene regulatory mechanisms in mammals. The functional role of TI in

transcriptional regulation is well characterized in the *DHFR* (di-hydro folate reductase) gene in quiescent cells. *DHFR* has two promoters, one major and one minor. In rapidly growing human cells, *DHFR* mRNA is transcribed from the major promoter to fulfill the high demand for DNA synthesis. In quiescent cells, a high level of *DHFR* gene transcription is not required; therefore, the DHFR gene needs to be silenced. Interestingly, transcriptional silencing of the major promoter is achieved by ncRNA transcription from the 5′ upstream minor promoter. The ncRNA produced from the minor promoter forms a triplex structure at the major promoter and interferes with the formation of the preinitiation complex. Furthermore, the ncRNA from the minor promoter also interacts with TFIIB, thus titrating away the components of the preinitiation complex (Fig. 1.1aII). These results indicate that both the ncRNA and the act of its transcription play a crucial role in the transcriptional repression of the *DHFR* major promoter via dissociation of the preinitiation complex (Blume et al. 2003; Martianov et al. 2007).

1.4 Heritable Epigenetic Gene Inactivation via Noncoding Transcription

Epigenetic gene silencing refers to the heritable mechanisms that mediate gene silencing without any changes in the primary DNA sequence. For example, post-translational histone modifications, such as di- and trimethylation of the histone H3 lysine 9 residue (H3K9me2 and H3K9me3) and trimethylation of the histone H3 lysine 27 residue (H3K27me3), and DNA methylation are often enriched at transcriptionally silenced genes (Kouzarides 2007 and references therein). Recent evidence suggests that transcriptional read-through of a neighboring gene by sense or antisense transcription results in heritable epigenetic gene inactivation, which has been shown to occur mostly in disease conditions. For example, the mismatch repair gene, *MSH2*, is often methylated or deleted in Lynch syndrome patients who are susceptible to colorectal and endometrial cancers. A recent study demonstrated that a deletion at the 3′ end of the *TACSTD1* gene resulted in extension of its transcription into the downstream *MSH2* gene, causing specific methylation and transcriptional inactivation of its promoter (Ligtenberg et al. 2009). However, it is not clear how transcriptional read-through across the *MSH2* promoter leads to its methylation.

A similar mode of action was detected as part of a disease mechanism in patients with an inherited form of alpha-Thalassemia, where transcriptional silencing of the *HBA2* gene was detected due to aberrant antisense transcription across its promoter (Tufarelli et al. 2003). In these patients, deletion of a region between the *HBA2* gene (α2 globin) and the *LUC7L* gene places the truncated *LUC7L* gene very close to the *HBA2* gene, resulting in transcriptional read-through from the *LUC7L* promoter into the normally expressed *HBA2* gene promoter. This transcriptional read-through causes DNA methylation and silencing of the *HBA2* gene.

Furthermore, a transgenic mouse model was used to show that antisense transcription through the *HBA2* promoter CpG island is necessary and sufficient to cause *HBA2* promoter DNA methylation and silencing (Tufarelli et al. 2003). In both instances, transcriptional silencing of protein-coding genes occurred due to aberrant transcriptional read-through, indicating that common mechanisms are used in aberrant and programmed silencing, and the only difference is the direction of transcription: in the former, it is sense, and in the latter, it is antisense.

Transcriptional silencing by aberrant natural antisense transcription across promoters appears to be a common feature in various diseases as it has also been documented in tumor suppressor genes such as *p15* and *p21*. The *p15* gene is a key tumor suppressor gene, and the loss of *p15* expression either by deletion, point mutation, or promoter hypermethylation is associated with a variety of tumors (Nobori et al. 1994). Recently, an ncRNA transcribed antisense to the *p15* gene (*p15AS*) was identified. This antisense RNA was shown to be expressed in leukemia cells at higher levels than in normal cells (Yu et al. 2008). Interestingly, *p15* antisense RNA transcription leads to enrichment of the repressive chromatin mark (H3K9me3) over the *p15* promoter and exon 1. The expression of *p15AS* is also correlated with *p15* promoter DNA hypermethylation. The epigenetic silencing of the *p15* promoter by *p15AS* is Dicer-independent, indicating that it is not mediated by RNA interference. Like the p15 gene, the *p21* gene is also often methylated and silenced in several cancers. Recent investigation has shown that bidirectional transcription of the *p21* gene is critical for its balanced expression. Suppression of steady state levels of the *p21* antisense RNA (*p21AS*) results in activation of the *p21* sense RNA. The repression of *p21* sense RNA by *p21AS* is mediated in an Ago-1-dependent manner via formation of heterochromatin over the *p15* sense promoter (Kim et al. 2006; Morris et al. 2008).

In the above four examples, the sense genes are silenced epigenetically via heterochromatin formation at the promoter due to aberrant transcription in the sense or antisense directions (Fig. 1.2a). Though heterochromatin formation over the silenced promoters is common in all the cases, it is not clear whether common mechanism(s) are involved. It is also not apparent, from the available data, whether the act of transcription, or the RNA itself, mediates transcriptional silencing. Although, in the case of *p15*, the data point towards a functional role for the RNA, it needs to be thoroughly investigated before the act of transcription is ruled out as the mechanism involved in transcriptional silencing.

1.5 LncRNAs Mediate Long-Range Gene Silencing Through the Recruitment of Polycomb Repressor Complexes

In mammals, subsets of genes are expressed from one of the parental alleles, while the other allele is often silenced by repressive epigenetic modifications. This allele-specific silencing is most prevalent in imprinted gene clusters and on the inactive

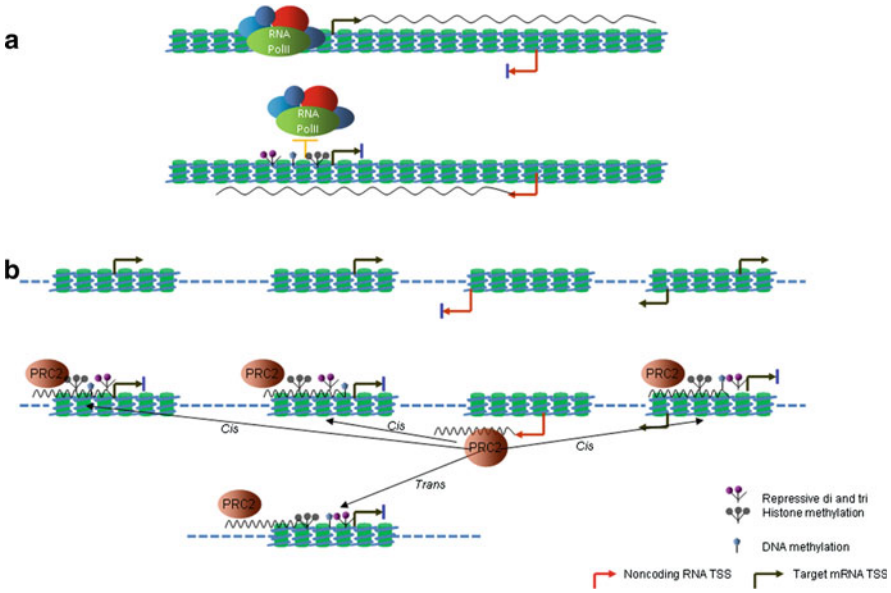

Fig. 1.2 Epigenetic reprograming of individual as well as domain-wide gene regulation by lncRNAs or its transcription. (**a**) An antisense ncRNA transcription across the promoter of the overlapping sense gene causes the formation of repressive chromatin environment via the enrichment of repressive modifications such as H3K27me3, H3K9me2, H3K9me3, and DNA methylation, thus repressing the overlapping sense gene. (**b**) The lncRNA-mediated regulation of gene expression in chromosomal domains via targeting of PRC2 complexes *in cis* or *in trans*

X chromosome in female mammals. In imprinted domains, allele-specific gene silencing occurs in a parent of origin-specific manner. In the case of the X chromosome in female mammals, allele-specific gene silencing also occurs in a parent of origin-specific manner (X-linked genes are silenced only on the paternal chromosome) in preimplantation embryos, whereas it occurs at random later in embryonic development. Interestingly, lncRNAs have been shown to play an important role in the establishment and maintenance of allele-specific gene silencing.

Cells in female mammals have two X chromosomes, whereas males have only one X. In order to equalize the dosage of X-linked gene products between males and females, one of the X chromosomes becomes inactivated during early embryonic development in female mammals (Payer and Lee 2008 and references therein, Chap. 3). The X chromosome inactivation center (XIC), a 500-kb region on the X chromosome, is implicated in X chromosome inactivation (XCI). The XIC harbors several genes for lncRNAs, for example, *Xist* (X inactivation specific transcript), *Tsix* (an antisense transcript to *Xist*), *Xite*, *DXPas34*, and *RepA* among others. *Xist* plays an important role in XCI by directing the heterochromatin machinery along the inactive X chromosome, and the other lncRNAs are involved in the regulation of *Xist* expression, and thus control the counting and choice processes of XCI (Payer and Lee 2008).

Mouse embryonic stem cells (ES cells) have been widely used to study XCI as they faithfully recapitulate the molecular events that serve to establish random XCI in the inner cell mass (ICM) of blastocysts. In ES cells, both X chromosomes are active and *Xist* RNA expression is maintained at very low levels on both chromosomes by pluripotency factors such as Nanog, Oct3/4, and Sox2 (Navarro et al. 2008). Upon differentiation, *Xist* RNA is upregulated on the future inactive X chromosome and spreads along the X chromosome *in cis*, accompanied by accumulation of repressive histone marks (H3k27me3 and H3K9me3), CpG DNA methylation, and deposition of the histone variant macroH2A1, thus establishing a repressive chromatin environment devoid of RNA Pol II. The A region, rich in repeats, at the 5′ end of *Xist* was shown to be critical for the establishment of XCI (Wutz et al. 2002). Deletion of this region compromised the accumulation of repressive histone modifications and silencing of X-linked genes *in cis*, suggesting that this repeat-rich region recruits the repressive histone modification machinery to the X chromosome *in cis*.

Recently, a new lncRNA (*RepA*) of 1.6 kb in length was discovered at the 5′ end of the *Xist* gene, covering the A repeat-rich region of the *Xist* gene (Zhao et al. 2008). *RepA* associates with the PRC2 complex members, EZH2 and SUZ12, before and during XCI. Interestingly, the PRC2 complexes are targeted to chromatin only at the onset of XCI. In light of the identification of a new member in the long list of lncRNAs involved in XCI, it would be interesting to investigate whether the *Xist* and *RepA* RNAs function synergistically in the XCI process or whether they have altogether different functions. However, an earlier study investigating the dynamics of XCI found that the *Xist* RNA forms a repressive compartment in the early phases of ES cell differentiation. The repressive compartment excludes the RNA polymerase II machinery from the genes to be silenced (Chaumeil et al. 2006), and this step is not dependent on the A repeat-rich region of *Xist* as ES cells in which the A region has been deleted still form the repressive compartment. However, the formation of the repressive compartment followed by accumulation of the H3K27me3 marks, and the translocation of X-linked genes into the core of the repressive compartment is dependent on the A repeat-rich region, indicating that the A-repeat plays a critical role in the transcriptional silencing of X-linked genes (Chaumeil et al. 2006). Together, these observations suggest that, at the onset of XCI, *Xist* organizes a repressive chromatin compartment, which includes all the genes to be silenced on the future inactive X chromosome. This is followed by *RepA*-dependent recruitment of the PRC2 complex members to stabilize the repressive compartment by repressive chromatin modifications (Zhao et al. 2008).

Similar to *Xist*/*RepA*-mediated XCI, a subclass of lncRNAs, including *Kcnq1ot1* and *Airn*, mediate transcriptional gene silencing in imprinted chromosomal domains in mouse. The molecular mechanism by which these two lncRNAs mediate gene silencing shows many similarities to the *Xist* RNA-mediated XCI. Both *Kcnq1ot1* and *Airn* are ~100 kb long RNA pol II-encoded ncRNAs, transcribed from the paternal allele of mouse chromosomes 7 and 17, respectively. They are responsible for the silencing of multiple genes spread over several hundred kilobases of the genome (Fitzpatrick et al. 2002; Sleutels et al. 2002;

Thakur et al. 2004; Kanduri et al. 2006). Both the lncRNAs have been shown to coat the chromatin of their target genes (Murakami et al. 2007; Nagano et al. 2008; Mohammad et al. 2008). *Kcnq1ot1* target genes show significant enrichment of the repressive chromatin marks, H3K27me3 and H3K9me3, but not the active chromatin marks H3K9ac and H3K4me3 (Pandey et al. 2008). Similarly, *Airn* ncRNA target genes show enrichment of H3K9me3 (Nagano et al. 2008). The presence of repressive chromatin marks over target genes is correlated with the association of *Kcnq1ot1* with the PRC2 members (EZH2 and SUZ12) and G9a (H3K9 histone methyltransferase) and of *Airn* with G9a (Nagano et al. 2008; Pandey et al. 2008). Collectively, these observations suggest that these lncRNAs interact with heterochromatin proteins and recruit them to the target genes, thus modifying the chromatin structure surrounding the promoters (Fig. 1.2b). Interestingly, both *Kcnq1ot1* and *Airn* have been shown to silence genes by organizing repressive chromatin compartments similar to that seen in case of *Xist* (Redrup et al. 2009). Another striking similarity between *Kcnq1ot1* and *Xist* is that, like *Xist*, *Kcnq1ot1* harbors a 0.9 kb silencing domain (SD) at the 5′ end of the RNA, which is crucial for the epigenetic silencing of its target genes (Wutz et al. 2002; Mohammad et al. 2008). Once the silencing of the target genes is established, it is equally important to maintain silencing through subsequent cell divisions, and it is possible that this is achieved by targeting the silenced gene to the heterochromatin nuclear compartments. Like *Xist*, *Kcnq1ot1* has been shown to maintain transcriptional silencing by recruiting genes to the perinucleolar space, which is enriched with heterochromatin factors such as Ezh2 (Mohammad et al. 2008; Zhang et al. 2007).

Intriguingly, lncRNAs have also been implicated in gene silencing *in trans*. In an elegant study using human primary fibroblast cells, it was shown that transcription of the *HOTAIR* lncRNA from the *HOXC* cluster correlates with the appearance of H3K27me3 marks over the *HOXD* cluster, which resides on another chromosome (Rinn et al. 2007). Depletion of *HOTAIR* using siRNA technology resulted in the loss of H3K27me3 marks over the *HOXD* cluster, indicating a link between *HOTAIR* expression from the *HOXC* locus and the enrichment of H3K27me3 marks over the *HOXD* cluster. Moreover, *HOTAIR* was shown to interact with the PRC2 members, EZH2 and SUZ12, in both in vitro and in vivo experiments. On the basis of the above observations, the authors speculated that *HOTAIR* interacts and guides the PRC2 complex to the *HOXD* cluster to silence the genes by H3K27me3 chromatin modification (Fig. 1.2b) (Rinn et al. 2007). Furthermore, a recent study demonstrated that the overexpression of *HOTAIR* in epithelial cancer cells resulted in genome-wide changes in the PRC2 complex occupancy and enhanced cancer invasiveness and metastasis (Gupta et al. 2010). This link between lncRNA-mediated epigenome reprogramming and cancer is most interesting.

Taken together, a consensus seems to be emerging by which lncRNAs are involved in epigenetic gene silencing. Upon transcription, these lncRNAs form ribonucleoprotein (RNP) complexes with repressive histone modification machinery. This could be achieved either by the interaction of proteins with a linear RNA sequence or by formation of an RNA secondary structure. The latter possibility is perhaps more likely as, even though there are no sequence similarities between the

above-mentioned lncRNAs, they still form RNP complexes with the same proteins. Supporting this idea, a 2-D structure of the *Xist* A region in mouse and human has been shown to be important for binding of the PRC2 complex to *Xist* (Maenner et al. 2010). The RNPs are then directed to the target genes, either *in cis* or *in trans*, by an unknown mechanism, thus resulting in higher order repressive chromatin formation and silencing of the associated genes. This silenced state can be further stabilized and maintained through subsequent cell divisions by targeting the silenced genes to the nucleolar or perinuclear region (Zhang et al. 2007; Mohammad et al. 2008).

Although our knowledge of lncRNA-mediated epigenetic gene silencing has significantly improved in the past few years, several key questions remain to be answered. First, how do lncRNAs maintain their high levels of expression in a repressive chromatin environment? Do they need a repressive chromatin environment for high expression levels, or do they have a different mechanism to combat this problem? For example, the presence of boundary elements flanking the lncRNA promoter and coding sequences, which prevent the spread of heterochromatin formation into the lncRNA gene, or the presence of strong promoter elements, which can overcome the heterochromatinization by recruiting p300/pCAF, or similar transcriptional activators (Pandey et al. 2004), or both. Second, how are RNP complexes targeted to specific genes, whereas other genes residing in between the target genes escape silencing? Since no sequence homology between lncRNAs and their target genes has been reported so far, it is unlikely that targeting is based on sequence similarity.

1.6 LncRNA-Mediated Targeting of Activator Complexes in Epigenetic Gene Activation

Some lncRNAs have been shown to activate genes through targeting activator complexes to gene regulatory regions. This is best exemplified in the case of the *roX* RNA-mediated hyperactivation of the X chromosome in *Drosophila melanogaster* (see Chap. 7). In contrast to mammals, where dosage of X-linked gene products between males and females is achieved via inactivation of one of the two X chromosomes in females, equal dosage of X-linked gene products between male flies with one X chromosome and female flies with two X chromosomes is achieved by hypertranscription of the lone X chromosome in males. The upregulation of X-linked genes is achieved by *roX* RNA-dependent targeting of the dosage compensation complex (DCC) at several loci along the X chromosome. The DCC consists of five proteins, MSL1 (male specific lethal), MSL2, MSL3, MLE (Maleless), MOF (Males absent on the first), and two lncRNAs: *roX1* and *roX2* (RNA on the X). MSL1 and MSL2 are necessary for DCC binding to DNA; MOF is an enzyme that catalyzes the acetylation of lysine 16 on histone H4 (H4K16ac), a modification crucial for the transcriptional upregulation of genes on the X chromosome (Gelbart et al. 2009); MLE is an ATP-dependent RNA/DNA helicase,

required for the incorporation of *roX* RNA into the DCC. The *roX1* and *roX2* ncRNAs are transcribed from the X chromosome and either of them is sufficient for correct localization of the DCC along the X chromosome. Deletion or mutation of both *roX* RNAs resulted in mislocalization of the DCC complex to the chromocenter and the heterochromatin regions of autosomes (Meller and Rattner 2002; Chap. 7).

In flies, hundreds of small GA-rich DNA elements, known as chromatin entry sites (CESs) or high affinity sites (HASs), are present across the X chromosome. The DCC can recognize and bind to CESs in the absence of *roX* lncRNAs; however, gene activation cannot be achieved (Alekseyenko et al. 2006, 2008; Straub et al. 2008), indicating that *roX* lncRNAs are an integral part of the DCC complex. Intriguingly, CESs are enriched only twofold on the X chromosome when compared to autosomes, suggesting that CESs alone are not sufficient for X chromosome recognition by the DCC. Moreover, autosomal transgene copies of *roX* can rescue male embryos carrying deletions of the *roX1/2* RNA genes. In these embryos, the DCC was localized to the X chromosome and also to limited autosomal loci, further suggesting that the mere presence of CESs on autosomes is not sufficient for correct targeting of the DCC to autosomes. The CES provides an entry point for the DCC; however, transcriptional upregulation of genes requires spreading of the DCC from the CES and the H4K16ac modification of chromatin (Gelbart and Kuroda 2009 and references therein). MSL3, another member of the DCC, contains a chromodomain, which has been shown to bind to nucleosomes with the H3K36me3 modification in vitro. The chromodomain of MSL3, along with MLE and MOF, is required for the spreading of the DCC complex (Sural et al. 2008). Although the exact role of the *roX* lncRNAs is not yet clear, it has been suggested that they are vital for the cotranscriptional assembly of the DCC, increasing the affinity of the DCC for the CES and in enhancing the enzymatic activity of MOF in the DCC complex (Gelbart et al. 2009).

LncRNA-mediated transcriptional activation through the recruitment of activator complexes has also been reported at the single gene level. For example, ncRNAs, encoded by polycomb/trithorax elements in the *Bxd* region in *Drosophila*, recruit a member of the trithorax complex, ASH1, to the downstream *Ubx* gene by forming base pair interactions with DNA. ASH1 is a histone methyltransferase containing a SET domain and its ncRNA-dependent recruitment to the *Ubx* gene promoter results in active chromatin formation and transcriptional activation of the *Ubx* gene (Fig. 1.3) (Sanchez-ELsner et al. 2006).

Epigenetic gene activation by lncRNAs is also implicated in the regulation of *Hox* genes during the primitive streak phase of embryoid body (EB) differentiation in mice (Dinger et al. 2008). *Evx1as* and *Hoxb5/6as* lncRNAs show concordant expression with the *Evx1* and *Hox5/6* genes, respectively. The *Evx1as* and *Hoxb5/6as* lncRNAs are enriched in the active chromatin compartment (H3K4me3) and also interact with MLL1 (a histone methyltransferase responsible for H3K4me3 methylation), which suggests that these lncRNAs activate flanking genes through the establishment of active chromatin structures (Fig. 1.3) (Dinger et al. 2008).

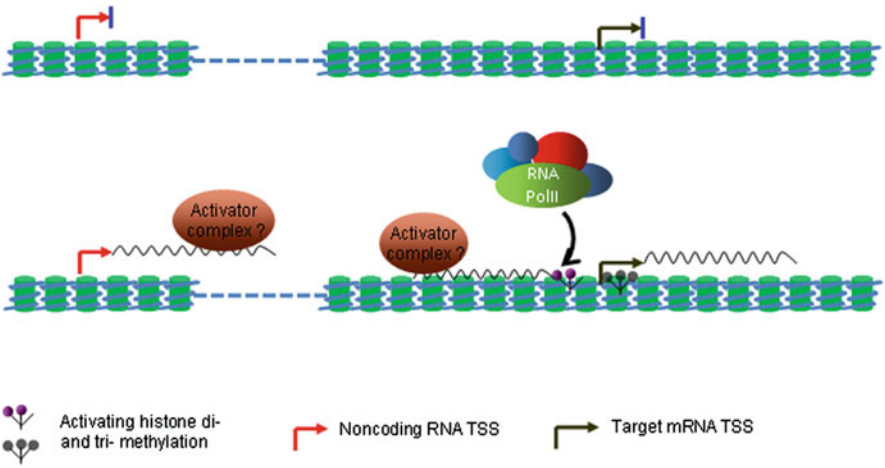

Fig. 1.3 Epigenetic gene activation through the targeting of activator complexes to the gene regulatory regions. Intergenic lncRNAs have been shown to associate with H3K4me3 histone methylatransferases such as ASH1 in Drosophila, and MLL1 in mammals, and target them to the promoters of nearby genes to activate their transcription through establishing active chromatin marks

However, the absolute requirement of *Evx1as* and *Hoxb5/6as* lncRNAs in the gene activation process has not been investigated.

Interestingly, in a recent investigation, a long intergenic ncRNA, *Intergenic 10*, was implicated in the activation of the flanking genes, *FANK1* and *ADAM12*, via the formation of active chromatin structures (Mondal et al. 2010). Downregulation of *Intergenic 10* in human fibroblasts resulted in significant loss of expression and active chromatin marks, such as H3K4me3, from the flanking genes, indicating that this lncRNA specifically activates its flanking genes. Except for *roX* lncRNAs, which act at the RNA level, it is not clear whether the process of transcription, or the ncRNA itself, takes part in the biological events involving lncRNAs described above.

1.7 Transcriptional Regulation of Heat Shock Response by lncRNAs

LncRNAs have been implicated in the global transcriptional upregulation of heat shock responsive genes and in the downregulation of housekeeping genes during the heat shock response. Transcriptional upregulation upon heat shock in mammals is mediated by heat shock factor 1 (HSF1). Under normal growth conditions, HSF1 is associated with hsp90 and other chaperones in an inactive complex, which cannot bind to heat shock elements (HSEs) found in the promoters of heat shock responsive genes. Upon heat shock treatment of cells, HSF1 is released from the inactive complex and forms an HSF1 trimer with the help of eEF1A (eukaryotic elongation factor 1A) and a lncRNA, *HSR1* (Shamovsky and Nudler 2008). The trimeric HSF1

then binds to HSEs to activate heat shock responsive genes. The lncRNA, *HSR1*, is ubiquitously expressed in cells growing under normal conditions. Heat shock causes a conformational change in the *HSR1* structure, which, together with eEF1A, facilitates *HSF1* trimerization and its DNA binding, leading to transcriptional activation of heat shock responsive genes.

Conversely, two other lncRNAs, *Alu* and *B2*, transcribed from *Alu* repeats in human and *SINE B2* (short interspersed elements B2) repeats in mouse, respectively, are known to inhibit transcription from housekeeping genes during the heat shock response (Mariner et al. 2008; Yakovchuk et al. 2009). The *Alu* and *B2* ncRNAs possess a modular structure, which includes two domains: an RNA pol II binding domain and a transcriptional inhibitory domain, both of which are essential for the transcriptional repression of target genes. It has been demonstrated that *Alu* and *B2* RNAs bind to RNA pol II before formation of the preinitiation complex and that the binding of the ncRNA with RNA pol II does not inhibit the association of RNA poll II with general transcription factors (Chap. 6). The *Alu* and *B2* RNA inhibitory domains inhibit formation of the contact between RNA pol II and the DNA promoter elements required for the initiation of transcription, perhaps by changing the structure of the transcription complex (Fig. 1.1b). Intriguingly, *Alu* and *B2* RNAs share no sequence similarity, yet they function via a similar mechanism (Yakovchuk et al. 2009), probably due to the similarity of their secondary structures, indicating that the lack of conservation at the primary sequence level does not necessarily mean lack of function and that secondary structures could harbor critical functional information.

1.8 LncRNAs Regulate Transcription by Modulating Protein Activity

Many transcription factors are localized in the cytoplasm of resting cells. In response to external stimuli, they are transported from the cytoplasm to the nucleus to activate the transcription of an array of genes. This cytoplasmic to nuclear transport is mediated by various different mechanisms generally thought to involve proteins. A genome-wide screen to identify lncRNAs that inhibit the NFAT (Nuclear Factor of Activated T cells) activity in a human cell line identified a noncoding repressor of NFAT (NRON) (Willingham et al. 2005). The NRON inhibits NFAT nuclear import by associating with members of the importin-beta superfamily, which are involved in the nucleocytoplasmic transport of protein cargos (Willingham et al. 2005). Although the exact mechanism of this inhibition in not clear, it suggests the importance of lncRNAs in such processes.

It is intriguing to note that lncRNAs can also modulate gene activation programs globally by regulating the functions of key transcription factors or signaling molecules. One such case is the regulation of the transcriptional activation of several genes by the glucocorticoid receptor (GR) in response to glucocorticoids.

The GR is a cytoplasmic protein, which upon ligand binding, moves into the nucleus and binds to glucocorticoid response elements (GRE) via its DNA binding domain. This results in the recruitment of transcriptional activators and coactivators to the regulatory regions of GR-responsive genes, and ultimately, in the activation of GR-responsive genes.

A noncoding transcript known as growth arrest-specific 5 (*Gas5*) accumulates in growth-arrested cells. Overexpression of *Gas5* inhibits GR binding to GRE elements in a dose-dependent manner, suggesting a direct role for the *Gas5* ncRNA in GR-mediated transcriptional reprogramming. Deletion studies, to pin down the *Gas5* ncRNA region responsible for the inhibition of GR binding to the GRE, revealed a short region forming a hairpin structure with a GRE-like sequence. Mutation in this GRE-like sequence, or in the DNA binding domain of GR, abolished GR binding to *Gas5*. Taken together, these results suggest that the GRE-like structure in the *Gas5* ncRNA titrates out the ligand-bound GR, thus inhibiting the activation of GR-responsive genes (Kino et al. 2010).

Furthermore, ncRNAs can also alter chromatin-bound protein activity by allosterically modifying protein structure. This has been elegantly demonstrated in the case of the cyclin D1 (*CCND1*) gene in response to DNA damage. The transcription of the *CCND1* gene is dependent on histone acetylation of its promoter, mediated by the histone acetyl transferase (HAT) activity of CREB binding protein (CBP). The *CCND1* gene is silenced when cells are exposed to agents that damage DNA, such as ionizing radiation. Upon exposure to ionizing radiation, an RNA binding protein, TLS (translocated in liposarcoma), is recruited to the *CCND1* gene promoter by ncRNAs transcribed from the *CCND1* 5′ regulatory region (Wang et al. 2008). These ncRNAs are not only responsible for TLS recruitment but also allosterically modify the TLS protein such that it inhibits the HAT activity of CBP (Wang et al. 2008). The examples described above further emphasize the complexity of gene regulation in higher organisms and the power of lncRNAs to regulate each and every step of transcriptional regulatory mechanisms.

1.9 LncRNAs Regulates mRNA Splicing, Stability, and Translation

Posttranscriptional control of gene expression is critical for the quick response of cells to changes in external stimuli. Posttranscriptional regulation involves the regulation of mRNA splicing, mRNA localization, and mRNA stability and translation, and evidence from recent investigations suggests that these steps are also regulated by lncRNAs. The epithelial to mesenchymal transition (EMT) is a crucial step in organismal development and involves the downregulation of the E-cadherin gene in mesenchymal cells. E-cadherin is downregulated by ZEB2, a transcriptional repressor (Guaita et al. 2002). Interestingly, the *Zeb2* gene is transcribed in both epithelial and mesenchymal cells, but in epithelial cells, its translation is prevented

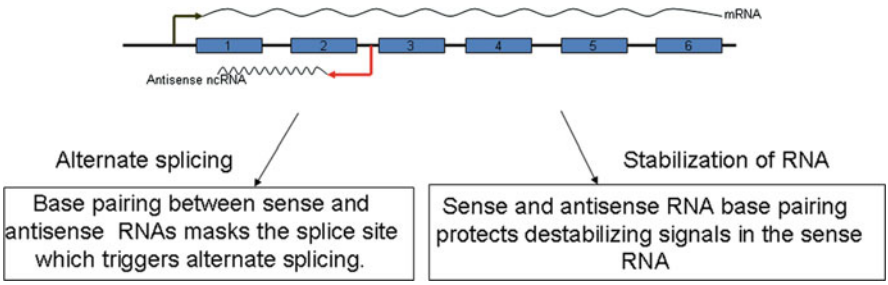

Fig. 1.4 Posttranscriptional gene regulation by an antisense ncRNA. Sense/antisense hybrid formation masks splice junctions or mRNA destabilization signals, leading to alternative splicing or stabilization of the sense transcript

by a splicing event, which removes the IRES (Internal Ribosome Entry Site) containing the 5′ UTR. In the mesenchymal cells, on the other hand, an antisense RNA overlapping the 5′ UTR splice site forms a sense–antisense RNA hybrid, which prevents splicing of the 5′ UTR and the IRES, thereby allowing translation of the *Zeb2* mRNA (Fig. 1.4) (Beltran et al. 2008).

A recent investigation has implicated long antisense ncRNAs in the pathogenesis of Alzheimer's disease. It has been shown that an antisense ncRNA, (*BACE1-AS*), against *ß-Secretase*, also known as *BACE1*, is upregulated in Alzheimer's patients, and that *BACE1-AS* upregulation is linked to the stabilization of the *BACE1* mRNA, and thus an increase in its protein level (Fig. 1.4) (Faghihi et al. 2008). This increase in BACE1 results in cell stress through the production of the amyloid β 1–42 peptide, which in turn increases the production of *BACE1-AS* in a feed-forward mechanism (Faghihi et al. 2008). It is not yet clear how *BACE1-AS* increases the stability of *BACE1*. Conversely, an antisense RNA (*aHIF*) originating from the 3′ UTR of the hypoxia-inducible factor 1 alpha (*HIF-1α*) has been proposed to reduce the stability of the *HIF-1α* mRNA (Rossignol et al. 2004). The *HIF-1α* mRNA 3′ UTR has AU-rich elements that are known to act as signals for RNA degradation. In cells expressing low levels of *aHIF*, the AU-rich elements of *HIF-1α* mRNA are not exposed due to complex secondary structure formation; however, when *aHIF* is present at higher levels, it is proposed to form an RNA–RNA hybrid with *HIF-1α* mRNA, thus exposing the AU-rich elements of *HIF-1α* and promoting its degradation (Rossignol et al. 2002).

1.10 Conclusions and Future Perspectives

The last few years have seen an increase in publications describing pervasive transcription in multicellular organisms, which results in the production of a large number of ncRNAs. Among these ncRNAs, the lncRNAs perhaps represent the most complex category of regulatory molecules in the multicellular organisms. So far, no sequence or structural similarity has been reported between those

lncRNAs shown to have a common mode of action. Due to the complexity and diversity of their sequences and their mechanisms of action, progress in the field of lncRNAs has been very slow. Nonetheless, lncRNAs have emerged as key regulators of developmental programs through their control of transcriptional and posttranscriptional gene regulatory pathways. The information from different biological contexts indicates that the functional roles of noncoding transcription and/or the ncRNAs are interpreted in different ways. While the noncoding transcriptional process often interferes with neighboring genes at the transcriptional level via TI mechanisms, on the other hand, ncRNA molecules take part in gene regulation from the single gene level to an entire chromosome via recruitment of chromatin modifying complexes *in cis* or *trans*. Transcription of an ncRNA through the regulatory region of its target gene can inhibit the assembly of transcription factors, or alternatively, the lncRNA can bind directly to key basal transcription factors, thus inhibiting PIC complex formation and leading to gene silencing. Similarly, the act of noncoding transcription or the ncRNA itself can negatively regulate the assembly of repressor complexes at the gene regulatory regions of target genes, thereby leading to transcriptional activation. LncRNAs also affect the target gene transcriptional output by targeting the repressor or activator complexes to the regulatory region of genes, a mechanism that is fairly well established in dosage compensation in mammals and *Drosophila*. At the posttranscriptional level, lncRNAs regulate the splicing, localization, stability, and translation of the target mRNAs by base-pairing with their target RNAs.

Although limited numbers of functional lncRNAs have been identified so far, the immense regulatory potential of lncRNAs in various developmental programs in multicellular organisms is already evident, emphasizing that a genome-wide characterization of functional lncRNAs is needed. Once the catalog of lncRNAs has been refined using biochemical and bioinformatic tools, genome-wide RNA interference (RNAi) screens, combined with powerful imaging techniques, such as those used in the identification of cell cycle regulatory proteins (Neumann et al. 2010; Walter et al. 2010), can be applied to characterize the roles of lncRNAs in different biological processes.

Acknowledgments This work was supported by the grants from the Swedish Cancer Research foundation (Cancerfonden), Swedish Medical Research Council (VR-M), and Swedish Childhood Cancer Society (Barncancerfonden) to Chandrasekhar Kanduri. Chandrasekhar Kanduri is a Senior Research Fellow supported by VR-M.

References

Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project (2009) Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. Nature 457:1028–1032

Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI (2006) High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome. Genes Dev 20:848–857

Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ, Kuroda MI (2008) A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. Cell 134:599–609

Amaral PP, Mattick JS (2008) Noncoding RNA in development. Mamm Genome 19:454–492

Amaral PP, Dinger ME, Mercer TR, Mattick JS (2008) The eukaryotic genome as an RNA machine. Science 319:1787–1789

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215–233

Beltran M, Puig I, Peña C, García JM, Alvarez AB, Peña R, Bonilla F, de Herreros AG (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev 22:756–769

Bernstein E, Allis CD (2005) RNA meets chromatin. Genes Dev 19:1635–1655

Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M (2004) Global identification of human transcribed sequences with genome tiling arrays. Science 306:2242–2246

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET et al (2007) Identification and analysis of functional elements in 1% ofthe human genome by the ENCODE pilot project. Nature 447:799–816

Blume SW, Meng Z, Shrestha K, Snyder RC, Emanuel PD (2003) The 5′-untranslated RNA of the human dhfr minor transcript alters transcription pre-initiation complex assembly at the major (core) promoter. J Cell Biochem 88:165–180

Brouwer R, Allmang C, Raijmakers R, van Aarssen Y, Egberts WV, Petfalski E, van Venrooij WJ, Tollervey D, Pruijn GJ (2001) Three novel components of the human exosome. J Biol Chem. 2;276(9):6177–84

Bumgarner SL, Dowell RD, Grisafi P, Gifford DK, Fink GR (2009) Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. Proc Natl Acad Sci U S A 106:18321–18326

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K et al (2005) The transcriptional landscape of the mammalian genome. Science 309:1559–1563

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, Wheeler R, Wong B, Drenkow J, Yamanaka M, Patel S, Brubaker S, Tammana H, Helt G, Struhl K, Gingeras TR (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116:499–509

Chaumeil J, Le Baccon P, Wutz A, Heard E (2006) A novel role for *Xist* RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev 20:2223–2237

Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, Hughes TR, Winston F (2008) Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. PLoS Biol 6:e277

Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322:1845–1848

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM (2006) A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci U S A 103:5320–5325

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G (2010) A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol 8:e1000384

Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Soldà G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 18:1433–1445

Dobi KC, Winston F (2007) Analysis of transcriptional activation at a distance in *Saccharomyces cerevisiae*. Mol Cell Biol 27:5575–5586

Ebisuya M, Yamamoto T, Nakajima M, Nishida E (2008) Ripples from neighbouring transcription. Nat Cell Biol 10:1106–1113

Euskirchen G, Royce TE, Bertone P, Martone R, Rinn JL, Nelson FK, Sayward F, Luscombe NM, Miller P, Gerstein M, Weissman S, Snyder M (2004) CREB binds to multiple loci on human chromosome 22. Mol Cell Biol 24:3804–3814

Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G 3rd, Kenny PJ, Wahlestedt C (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat Med 14:723–730

Fitzpatrick GV, Soloway PD, Higgins MJ (2002) Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet 32:426–431

Gelbart ME, Kuroda MI (2009) *Drosophila* dosage compensation: a complex voyage to the X chromosome. Development 136:1399–1410

Gelbart ME, Larschan E, Peng S, Park PJ, Kuroda MI (2009) *Drosophila* MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. Nat Struct Mol Biol 16:825–832

Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. Nat Rev Genet 10:94–108

Grewal SI, Jia S (2007) Heterochromatin revisited. Nat Rev Genet 8:35–46

Guaita S, Puig I, Franci C, Garrido M, Dominguez D, Batlle E, Sancho E, Dedhar S, De Herreros AG, Baulida J (2002) Snail induction of epithelial to mesenchymal transition in tumor cells is accompanied by MUC1 repression and ZEB1 expression. J Biol Chem 277:39209–39216

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464:071–076

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458:223–227

He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW (2008) The antisense transcriptomes of human cells. Science 322:1855–1857

Hongay CF, Grisafi PL, Galitski T, Fink GR (2006) Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. Cell 127:735–745

Imamura T, Yamamoto S, Ohgane J, Hattori N, Tanaka S, Shiota K (2004) Non-coding RNA directed DNA demethylation of Sphk1 CpG island. Biochem Biophys Res Commun 322:593–600

Kanduri M, Kanduri C, Mariano P, Vostrov AA, Quitschke W, Lobanenkov V, Ohlsson R (2002) Multiple nucleosome positioning sites regulate the CTCF-mediated insulator function of the H19 imprinting control region. Mol Cell Biol 22:3339–3344

Kanduri C, Thakur N, Pandey RR (2006) The length of the transcript encoded from the Kcnq1ot1 antisense promoter determines the degree of silencing. EMBO J 25:2096–2106

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR (2007a) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316:1484–1488

Kapranov P, Willingham AT, Gingeras TR (2007b) Genome-wide transcription and the implications for genomic organization. Nat Rev Genet 8:413–423

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 106:11667–11672

Kim DH, Villeneuve LM, Morris KV, Rossi JJ (2006) Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat Struct Mol Biol 13:793–797

Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal 3:ra8

Kouzarides T (2007) Chromatin modifications and their function. Cell 128:693–705

Lefevre P, Witham J, Lacroix CE, Cockerill PN, Bonifer C (2008) The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. Mol Cell 32(1):129–139

Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TY, Bodmer D, Hoenselaar E, Hendriks-Cornelissen SJ, Tsui WY, Kong CK, Brunner HG, van Kessel AG, Yuen ST, van Krieken JH, Leung SY, Hoogerbrugge N (2009) Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3′ exons of TACSTD1. Nat Genet 41:112–117

Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, Dubois A, Sanglier-Cianférani S, Van Dorsselaer A, Clerc P, Avner P, Visvikis A, Branlant C (2010) 2-D structure of the A region of Xist RNA and its implication for PRC2 association. PLoS Biol 8:e1000276

Malone CD, Hannon GJ (2009) Small RNAs as guardians of the genome. Cell 136:656–668

Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol Cell 29:499–509

Martens JA, Laprade L, Winston F (2004) Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. Nature 429:571–574

Martens JA, Py Wu, Winston F (2005) Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. Genes Dev 19:2695–2704

Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature 445:666–670

Mattick JS (2004) RNA regulation: a new genetics? Nat Rev Genet 5:316–323

Mattick JS, Makunin IV (2006) Non-coding RNA. Hum Mol Genet 15:R17

Mayer C, Schmitz KM, Li J, Grummt I, Santoro R (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. Mol Cell 22:351–361

Meller VH, Rattner BP (2002) The *roX* genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J 21:1084–1091

Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS (2008) Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A 105:716–721

Mohammad F, Pandey RR, Nagano T, Chakalova L, Mondal T, Fraser P, Kanduri C (2008) Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. Mol Cell Biol 28:3713–3728

Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C (2010) Characterization of the RNA content of chromatin. Genome Res 20(7):899–907

Morris KV, Santoso S, Turner AM, Pastori C, Hawkins PG (2008) Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet 4: e1000258

Murakami K, Oshimura M, Kugoh H (2007) Suggestive evidence for chromosomal localization of non-coding RNA from imprinted LIT1. J Hum Genet 52:926–933

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349

Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P (2008) The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 322:1717–1720

Navarro P, Chambers I, Karwacki-Neisius V, Chureau C, Morey C, Rougeulle C, Avner P (2008) Molecular coupling of *Xist* regulation and pluripotency. Science 321:693–695

Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457:1038–1042

Neumann B, Walter T, Hériché JK, Bulkescher J, Erfle H, Conrad C, Rogers P, Poser I, Held M, Liebel U, Cetin C, Sieckmann F, Pau G, Kabbe R, Wünsche A, Satagopam V, Schmitz MH, Chapuis C, Gerlich DW, Schneider R, Eils R, Huber W, Peters JM, Hyman AA, Durbin R, Pepperkok R, Ellenberg J (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 464:721–727

Nobori T, Miura K, Wu DJ, Lois A, Takabayashi K, Carson DA (1994) Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. Nature 368:753–756

Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, Yamanaka I, Kiyosawa H, Yagi K, Tomaru Y, Hasegawa Y, Nogami A, Schönbach C, Gojobori T, Baldarelli R, Hill DP, Bult C, Hume DA, Quackenbush J, Schriml LM, Kanapin A, Matsuda H, Batalov S, Beisel KW, Blake JA, Bradt D, Brusic V, Chothia C, Corbani LE, Cousins S, Dalla E, Dragani TA, Fletcher CF, Forrest A, Frazer KS, Gaasterland T, Gariboldi M, Gissi C, Godzik A, Gough J, Grimmond S, Gustincich S, Hirokawa N, Jackson IJ, Jarvis ED, Kanai A, Kawaji H, Kawasawa Y, Kedzierski RM, King BL, Konagaya A, Kurochkin IV, Lee Y, Lenhard B, Lyons PA, Maglott DR, Maltais L, Marchionni L, McKenzie L, Miki H, Nagashima T, Numata K, Okido T, Pavan WJ, Pertea G, Pesole G, Petrovsky N, Pillai R, Pontius JU, Qi D, Ramachandran S, Ravasi T, Reed JC, Reed DJ, Reid J, Ring BZ, Ringwald M, Sandelin A, Schneider C, Semple CA, Setou M, Shimada K, Sultana R, Takenaka Y, Taylor MS, Teasdale RD, Tomita M, Verardo R, Wagner L, Wahlestedt C, Wang Y, Watanabe Y, Wells C, Wilming LG, Wynshaw-Boris A, Yanagisawa M, Yang I, Yang L, Yuan Z, Zavolan M, Zhu Y, Zimmer A, Carninci P, Hayatsu N, Hirozane-Kishikawa T, Konno H, Nakamura M, Sakazume N, Sato K, Shiraki T, Waki K, Kawai J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Imotani K, Ishii Y, Itoh M, Kagawa I, Miyazaki A, Sakai K, Sasaki D, Shibata K, Shinagawa A, Yasunishi A, Yoshino M, Waterston R, Lander ES, Rogers J, Birney E, Hayashizaki Y, FANTOM Consortium, RIKEN Genome Exploration Research Group Phase I & II Team (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420:563–573

Pandey RR, Ceribelli M, Singh PB, Ericsson J, Mantovani R, Kanduri C (2004) NF-Y regulates the antisense promoter, bidirectional silencing, and differential epigenetic marks of he Kcnq1 imprinting control region. J Biol Chem 279:52685–52693

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32:232–246

Payer B, Lee JT (2008) X chromosome dosage compensation: how mammals keep the balance. Annu Rev Genet 42:733–772

Perocchi F, Xu Z, Clauder-Münster S, Steinmetz LM (2007) Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Nucleic Acids Res 35:e128

Phillips JE, Corces VG (2009) CTCF: master weaver of the genome. Cell 137:1194–1211

Ponjavic J, Ponting CP, Lunter G (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res 17:556–565

Prasanth KV, Spector DL (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev 21:11–42

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH (2008) RNA exosome depletion reveals transcription upstream of active human promoters. Science 322:1851–1854

Redrup L, Branco MR, Perdeaux ER, Krueger C, Lewis A, Santos F, Nagano T, Cobb BS, Fraser P, Reik W (2009) The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. Development 136:25–30

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129(7):1311–1323

Rossignol F, Vache C, Clottes E (2002) Natural antisense transcripts of hypoxia-inducible factor 1alpha are detected in different normal and tumour human tissues. Gene 299:135–140

Rossignol F, de Laplanche E, Mounier R, Bonnefont J, Cayre A, Godinot C, Simonnet H, Clottes E (2004) Natural antisense transcripts of HIF-1alpha are conserved in rodents. Gene 339:121–130

Samanta MP, Tongprasit W, Sethi H, Chin CS, Stolc V (2006) Global identification of noncoding RNAs in *Saccharomyces cerevisiae* by modulating an essential RNA processing pathway. Proc Natl Acad Sci U S A 103:4192–4197

Sanchez-Elsner T, Gou D, Kremmer E, Sauer F (2006) Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. Science 311:1118–1123

Shamovsky I, Nudler E (2008) New insights into the mechanism of heat shock response activation. Cell Mol Life Sci 65:855–861

Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajska A, Harbers M, Kawai J, Carninci P, Hayashizaki Y (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A 100:15776–15781

Sleutels F, Zwart R, Barlow DP (2002) The non-coding *Air* RNA is required for silencing autosomal imprinted genes. Nature 415:810–813

Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB (2008) The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. PLoS Genet 4: e1000302

Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL (2009) MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res 19:347–359

Sural TH, Peng S, Li B, Workman JL, Park PJ, Kuroda MI (2008) The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. Nat Struct Mol Biol 15:1318–1325

Taft RJ, Pheasant M, Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 29:288–299

Thakur N, Tiwari VK, Thomassin H, Pandey RR, Kanduri M, Göndör A, Grange T, Ohlsson R, Kanduri C (2004) An antisense RNA regulates the bidirectional silencing property of the Kcnq1 imprinting control region. Mol Cell Biol 24:7855–7862

Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet 34:157–165

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484–487

Walter T, Shattuck DW, Baldock R, Bastin ME, Carpenter AE, Duce S, Ellenberg J, Fraser A, Hamilton N, Pieper S, Ragan MA, Schneider JE, Tomancak P, Hériché JK (2010) Visualization of image data from cells to organisms. Nat Methods 7:S26–S41

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature 454:126–130

Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science 309:1570–1573

Woo CJ, Kingston RE (2007) *HOTAIR* lifts noncoding RNAs to new levels. Cell 129:1257–1259

Wutz A, Rasmussen TP, Jaenisch R (2002) Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. Nat Genet 30:167–174

Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, Libri D, Jacquier A (2005) Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121:725–737

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. Nature 457(7232):1033–1037

Yakovchuk P, Goodrich JA, Kugel JF (2009) B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. Proc Natl Acad Sci U S A 106:5569–5574

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302:842–846

Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature 451:202–206

Zhang LF, Huynh KD, Lee JT (2007) Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell 129:693–706

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322:750–756

Chapter 2
Long Noncoding RNA as a Regulator for Transcription

Riki Kurokawa

Abstract Investigation of noncoding RNAs is in rapid progress, especially regarding translational repression by small (short) noncoding RNAs like microRNAs with 20–25 nucleotide-lengths, while long noncoding RNAs with nucleotide length of more than two hundred are also emerging. Indeed, our analysis has revealed that a long noncoding RNA transcribed from cyclin D1 promoter of 200 and 300 nucleotides exerts transcriptional repression through its binding protein TLS instead of translational repression. Translational repression is executed by short noncoding RNAs, while transcriptional repression is mainly done by long noncoding RNAs. These long noncoding RNAs are heterogeneous molecules and employ divergent molecular mechanisms to exert transcriptional repression. In this review, I overview recent publications regarding the transcription regulation by long noncoding RNAs and explore their biological significance. In addition, the relation between a random transcriptional activity of RNA polymerase II and the origin of long noncoding RNAs is discussed.

2.1 Introduction

It has been strikingly reported that more than ninety percent of the human genome is potentially transcribed (Carninci et al. 2005; Kapranov et al. 2007; Willingham and Gingeras 2006). However, a whole fraction of human HeLa cell RNA at a denatured RNA agarose gel displays mostly the 18S and 28S bands of ribosomal RNA and just smear bands that include mRNA, tRNA, and noncoding (nc) RNA (Fig. 2.1). This observation implies that the human genome generates vast number of ncRNAs, but most of them are as low copy number RNA molecules. The number of ncRNA species is huge, although each copy number is very low, suggesting that

R. Kurokawa
Division of Gene Structure and Function, Research Center for Genomic Medicine, Saitama Medical University, 1397-1 Yamane, Hidaka-shi, Saitama-Ken 350-1241, Japan
e-mail: rkurokaw@saitama-med.ac.jp

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3_2,

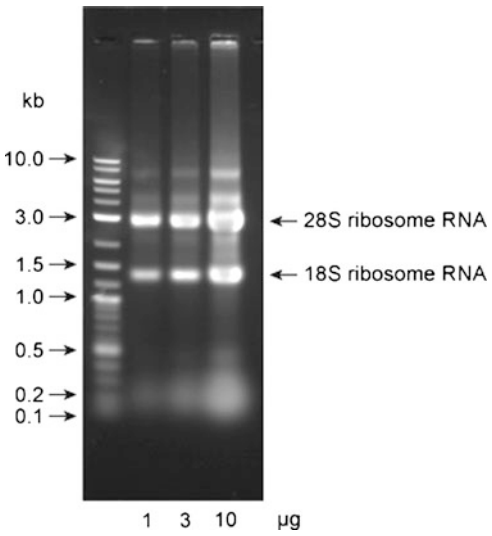

Fig. 2.1 Denatured electrophoresis of total RNA fractions of human fetal kidney cell line 293. Detection of RNA was performed with ethidium bromide

significant fractions of the ncRNAs might be involved in the regulation of various cellular functions instead of cellular structure. Actually, micro (mi) RNA, one of the most well-studied ncRNA functions as a translational repressor (Ambros 2001; Fire et al. 1998). Recently, transcription regulatory functions have been found in certain kinds of ncRNAs. Most of such kinds of ncRNAs have reported as "long" ncRNA of which length is more than 200 nucleotides (Kurokawa et al. 2009; Ponting et al. 2009). Mechanisms of the transcriptional regulations are divergent for various kinds of ncRNAs. In this review, I overview recent papers regarding the transcriptional regulation through the long ncRNAs and discuss heterogeneity of mechanisms of these transcriptional regulations.

2.2 Long ncRNAs

Long ncRNAs that regulate transcription are divergent molecules. Classification of long ncRNA is attempted in this section.

2.2.1 Length of Long ncRNAs

Long ncRNAs are tentatively defined as molecules of ncRNA more than 200 nucleotides long. Actually, their lengths are ranging from 200 bp to 2.2 kb of HOTAIR and 17 kb of Xist. Therefore, naming as "long ncRNA" is merely based on its nucleotide length.

2.2.2 Single- or Double-Stranded Long ncRNAs

There have been reported both single-stranded and double-stranded long ncRNAs. Sense and antisense strands of Alu repeats are transcribed and form a double-stranded RNA (Wang et al. 2008a). The functional consequence of the formation of a double-stranded ncRNA remains unclear. A possible explanation for double-strandedness of ncRNAs is that the double-stranded ncRNA might not bind a target molecule, and formation of double-strand of the ncRNA presents repression of the ncRNA function.

2.2.3 Subcellular Localization

Matured mRNAs after processing are transported to cytoplasm, while most of ncRNAs are known to be localized in nuclei. Some ncRNAs are localized both in nuclei and cytoplasm (Imamura et al. 2004). Only one ncRNA has been reported to be exclusively localized in cytoplasm (Louro et al. 2009). The long ncRNAs mainly reside in nuclei, suggesting their involvement in transcription.

2.2.4 Transcription of Long ncRNAs

Many of long ncRNAs represent tissue-specific pattern of expression. This suggests that the expression of these long ncRNAs should be strictly regulated and transcribed mostly by RNA polymerase II. Analysis of 1,600 ncRNAs showed that most of long ncRNAs are similar to authentic RNA polymerase II transcript as follows (Guttman et al. 2009; Khalil et al. 2009). First, these long ncRNAs contain trimethyl marks of histone H3-lysine (K) 4 at their promoter regions and trimethyl marks of histone H3-K36 along the length of the transcribed region, which are observed in usual transcripts by RNA polymerase II. These trimethyl marks are designated as “chromatin signature (a K4-K36 domain)” (Guttman et al. 2009). Second, the long ncRNAs generally possess the 5′CAP (7-methyl-guanosine cap) structure at the 5′ edge and also poly (A) tail at their 3′ end as well (Guttman et al. 2009). Third, the long ncRNAs have well-defined transcription factor binding sites like NF-kB in their promoter regions (Martone et al. 2003). These data strongly support that the transcription of the long ncRNAs is performed by RNA polymerase II (Martone et al. 2003). However, it has not been well identified which type of transcription factor could induce the long ncRNA transcription. Thus, regulation of transcription of the long ncRNA still remains uncovered.

2.3 Long ncRNAs Regulate Transcription

Divergent mechanisms of the transcriptional regulation by the long ncRNAs have been reported. At this section, the transcriptional regulations are attempted to categorize into three types (1) the regulation at the basic transcription factors including RNA polymerase II; (2) the regulation at the histone modification; (3) the regulation at the DNA methylation. The predominant type of the regulations appears to be mediated through the histone modification.

2.3.1 Transcriptional Regulation Through Targeting Basic Transcription Factors and RNA Polymerase II by Long ncRNAs

Direct interaction of long ncRNAs with basic core machinery is one of efficient mechanisms of transcriptional repression.

2.3.1.1 Alu RNA

SINE retrotransposon elements including Alu repeats generate numerous species of long ncRNAs (Maraia et al. 1993). It has been reported that Alu RNAs and SINE B2 RNAs exert transcriptional repression under the heat-shock condition (Allen et al. 2004; Espinoza et al. 2007; Mariner et al. 2008). SINE B2 and Alu RNA directly target RNA polymerase II. Furthermore, Alu RNA possesses a regulatory domain for function of RNA polymerase II (Mariner et al. 2008). Biochemical experiments demonstrated that Alu RNAs inhibit association of RNA polymerase II to the promoter DNA and represses the transcription (Mariner et al. 2008: see Chap. 6). SINE B2 turns out to have similar repressive effect on the transcription as well (Mariner et al. 2008). These data suggest that the repetitive sequence that occupies the half of the human genome could be transcribed, and their transcripts, the long ncRNAs, exert transcriptional repression. This presents the biological significance of the repetitive sequence in the human genome.

2.3.1.2 Dehydrofolate Reductase ncRNA

In quiescent mammalian cells, expression of dehydrofolate reductase (DHFR) is repressed. It has been reported that a transcript of a minor promoter located upstream of a major promoter is involved in the repression of DHFR (Martianov et al. 2007). In the quiescent cells, the transcript of the minor promoter was found to inhibit transcriptional initiation from the major promoter through direct binding to TFIIB of the preinitiation complex (Fig. 2.2). The alternative promoters within the

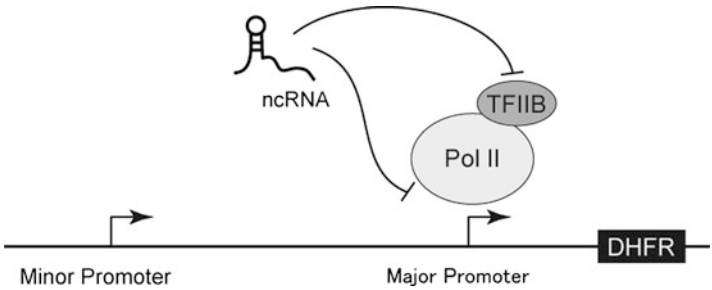

Fig. 2.2 Transcriptional repression of dehydrofolate reductase (DHFR) gene by the ncRNA transcribed from the minor promoter of the DHFR gene. The DHFR ncRNA represses the DHFR gene expression by blocking the preinitiation complex through targeting TFIIB and RNA polymerase II

same gene have been observed in various loci. It could be a general mechanism that the transcripts from the alternative promoters have a regulatory role in transcription of the promoter.

2.3.2 Transcriptional Regulation Through Histone Modification by the Long ncRNAs

The regulation of transcription by long ncRNA has been reported to be performed mainly through histone modification or DNA methylation. Some long ncRNAs activate transcription, while others repress it.

2.3.2.1 Steroid Receptor RNA Activator

Nuclear receptor (NR) forms a super family consisting of more than 50 members in the human genome and is the transcription factor that regulates divergent biological functions such as homeostasis and cellular differentiation and growth (Glass and Rosenfeld 2000). NR activates transcription through exchange of corepressor for coactivator upon specific binding of low molecular weight lipophilic compounds designated as ligands. The corepressor and coactivator were all supposed to be protein molecule. However, steroid receptor RNA activator (SRA) had been reported as a first example of the NR coactivator of RNA molecule (Hatchell et al. 2006; Lanz et al. 1999). SRA was found to activate various NR, for example, steroid hormone receptors such as glucocorticoid and estrogen receptors, retinoic acid, thyroid hormone, and vitamin D receptors. It has been suggested that SRA should activate transcription through recruitment of steroid receptor coactivator 1 (SRC1) and SRC1 with histone acetyltransferase (HAT) activity, and release of histone deacetylase (HDAC).

2.3.2.2 Embryonic Ventral Forebrain-2

During early development, 3.8-kb long ncRNA, embryonic ventral forebrain-2 (Evf2) is transcribed from intergene region between loci Dlx-5 and Dlx-6 (Bond et al. 2009; Feng et al. 2006). The Dlx gene related to Distalless gene (dll) homeodomain protein family of Drosophila plays a pivotal role in neuronal development. The dll gene forms a bigene cluster of Dlx5/6 and Dlx1/2. There are well-conserved enhancer regions, ei and eii, located between Dlx5 and Dlx6. Evf2 is transcribed from the ei and eii enhancer regions and binds the Dlx2 protein and activates the transcription of Dlx5/6 gene. The Evf2 ncRNA exerts transcriptional activation through the protein–protein interaction as follows (Fig. 2.3a). Dlx5/6 regions in their repression status are methylated at the CpG repeat, which is bound by MeCP2 and HDAC, while Evf2 activates them through removing MeCP2 and release of HDAC from the CpG repeat (Bond et al. 2009).

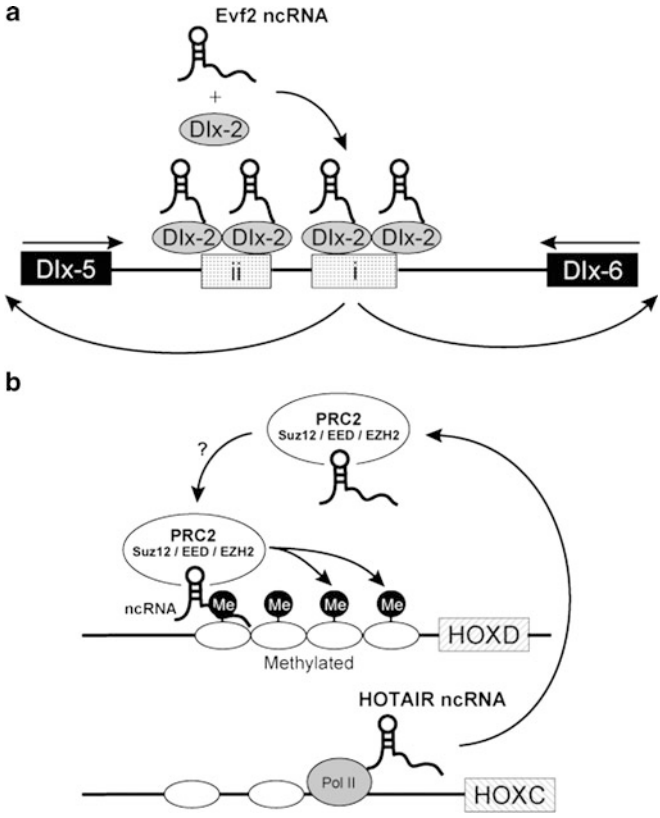

Fig. 2.3 The long ncRNAs involving in transcriptional regulation through chromosomal modification (**a**) Evf2 activates transcription by removing the methylase MeCP2 on CpG regions and releasing HDAC activity from the target gene. (**b**) HOTAIR activates transcription by binding PRC2 and histone methylation of HOXD locus

2.3.2.3 HOX Antisense Intergenic RNA

HOX gene clusters are essential for formation of body axis and segments during embryogenesis. In the human genome, four clusters of HOX genes have been identified, that is, HOXA (chromosome 7), HOXB (chromosome 17), HOXC (chromosome 12), and HOXD (chromosome 7). The tilling array analysis of these four clusters showed 231 novel ncRNAs and a highly conserved ncRNA in vertebrates, the HOX antisense intergenic RNA (HOTAIR) (Rinn et al. 2007). HOTAIR is a 2.2-kb ncRNA transcribed from noncoding region of HOXC cluster and recruited to HOXD locus upon binding the Polycomb repressive complex (PRC) 2. PRC2 possesses the H3K27 histone methyl transferase (HMTase) EZH2, Suz12, and EED as the components of the complex and induces histone methylation to repress expression of the gene. Then, HOTAIR represses the transcription of HOXD by recruitment of PRC2 and trimethylated histone H3-K27 (Fig. 2.3b). PRC2 is also involved in the X-chromosome inactivation (discussed later, see Chap. 3), suggesting that the complex has versatile epigenetic functions to mediate the transcriptional regulation by the long ncRNAs.

2.3.2.4 Cyclin D1

Recently, our group reported that an RNA-binding protein TLS (Translocated in liposarcoma) inhibits histone acetyltransferase (HAT) activity of CBP and p300 (Wang et al. 2008b). The HAT inhibitor, TLS, turns out to have specific target genes, cyclin D1 and E1, and represses the expression of cyclin D1 upon binding the RNA containing the GGUG-consensus sequence (Lerga et al. 2001). Expression of cyclin D1 gene has been repressed with treatment of ionizing radiation (IR) and the DNA damaging reagents (Miyakawa and Matsushime 2001). Our quest for any alteration of level of transcript after the IR treatment has demonstrated the increase of ncRNAs from the cyclin D1 promoter. These ncRNA [promoter (p)-ncRNA] transcribed from the cyclin D1 promoter was found to have the GGUG consensus sequence.

Binding of pncRNAs to TLS induces its recruitment to CBP/p300, major HAT activity in animal cells, and inhibition of their HAT activity (Fig. 2.4). Together with these data, it is suggested that expression of cyclin D1 gene could be repressed by pncRNAs through binding to TLS. This should be a mechanism like autorepression: a transcript from a gene represses its expression itself. We present the mechanism as an ncRNA-dependent transcriptional repression and have been pursuing the fact that the similar promoter-derived ncRNAs repress expression of other genes in the human genome. This could be a genome-wide network of cellular transcription repression.

2.3.3 DNA Methylation

An antisense RNA is known to induce gene-silencing through DNA methylation. This tells us tight relations between ncRNAs and DNA methylations.

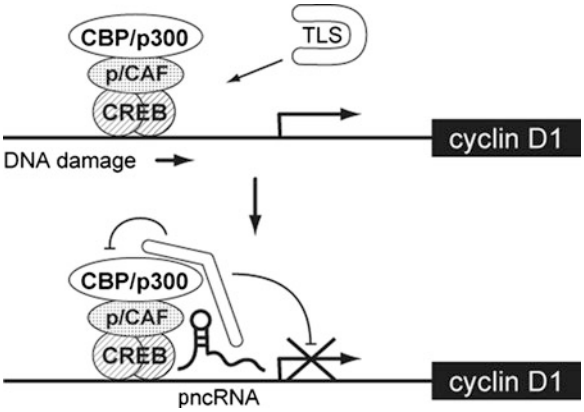

Fig. 2.4 The cyclin D1 pncRNA-dependent transcriptional repression Genotoxic factors like ionizing irradiation and DNA damaging reagents induce the pncRNA transcription. The pncRNAs bind TLS and inhibit the HAT activity of CBP/p300 to exert repressive effect on the cyclin D1 expression

2.3.3.1 P15AS

Antisense RNA of the tumor-suppressor gene p15 repressed the expression of p15 itself (Yu et al. 2008). In leukemia cells, the expression of p15 was reduced, while the level of antisense RNA of p15 was increased. The detailed analysis of the p15-antisense RNA using the leukemia cells showed that the antisense RNA induces methylation of the p15 locus DNA and its heterochromatinization to exert transcriptional repression. In the human genome, antisense RNAs of the 70% of coding genes are supposed to be expressed (Katayama et al. 2005). Taken together, these antisense RNAs might have regulatory role in gene expression.

2.3.3.2 Khps1

Khps1 is an antisense RNA transcribed from T-DMR (tissue-dependent differentially methylated region) of Sphk1 (sphingosine kinase-1). Overexpression of Khps1 stimulates demethylation of the CpG island of T-DMR but the methylation of its non-CG region (Imamura et al. 2004). The modulation of the methylation status of Sphk1 locus has been found to regulate expression of this locus. These data show tight relations between long ncRNA functions and DNA methylations.

2.4 ncRNAs as a Sensor for Cellular Signals

There have been reported divergent long ncRNAs transcribed from numerous regions of the human genome. Expression of long ncRNAs is supposed to be regulated by various “signals”, and suggested to have a role in “sensor” toward

the signals. Actually, we have found that the cyclin D1-pncRNA could work as a sensor for genotoxic signal of ionizing radiation (Wang et al. 2008b).

X-chromosome inactivation employs the ncRNA, the 1.6-kb RepA that is transcribed from the fragment of the Xist locus as an antisense RNA (Zhao et al. 2008). The reduction of expression of Tsix that is a full-length antisense RNA of Xist has a function as a signal. RepA as the sensor receives the reduction of the Tsix expression as the signal, recruits PRC2 to the Xist locus, and induces X chromosome inactivation. During embryonic development, HOTAIR also functions as a sensor and exerts gene silencing effect upon recruitment of PRC2 (Rinn et al. 2007). The long ncRNAs with the function of the sensors have been found to require histone-modifying enzymes. These observations suggest that long ncRNAs function as a sensor for various biological signals and execute regulation of gene expression through histone modification.

2.5 Mechanisms of Transcriptions of Long ncRNAs

Majority of long ncRNAs have been shown to be transcribed through RNA polymerase II, although some long ncRNAs are generated by RNA polymerase III (Dieci et al. 2007; Liu et al. 1995; Nguyen et al. 2001). Although the prevailing analyses of RNA polymerase II indicate that its major function is the precise initiation and elongation of protein-coding genes, early studies showed that RNA polymerase II possesses the ability to catalyze randomly initiated transcription from a calf thymus DNA or other crude DNA fractions as a template (Barbiroli et al. 1977; Legraverend and Glazer 1980; Reinberg and Roeder 1987). Indeed, RNA polymerase was shown to initiate transcription from nicked, gaped, and edge of DNA molecules in a sequence-independent manner (Sekimizu et al. 1979). This led to the notion that RNA polymerase II has potential to generate divergent transcripts from numerous and discrete sites in the genome.

Biochemical approaches using nuclei of the rat livers indicated that RNA polymerase I resides in nucleolus and is involved in generating ribosomal RNAs, while RNA polymerase II is located in nucleus (Roeder and Rutter 1970). RNA polymerase II was found to synthesize the "DNA-like RNA" that is the RNA having a base composition similar to that of total cellular DNA and predicted to work on transcription of the protein-coding genes (Roeder and Rutter 1970). Extensive biochemical and molecular biological studies have demonstrated that RNA polymerase II comprises multiple components, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, and that precise initiation of the transcription requires the RNA polymerase II with its essential components, that is, the holoenzyme of RNA polymerase II (Roeder 1991; Weake and Workman 2010). This shows that RNA polymerase II alone could not initiate specific and precise transcription and that for specific transcription RNA polymerase II needs to form the holoenzyme with general transcription factors like TFIIB and TFIID, while RNA polymerase II is able to catalyze a random transcription reaction with induction by some protein fractions as described below.

The fractions of Ehrlich ascites tumor cells (SII) and of HeLa cell (TFIIS) were shown to stimulate nonspecific transcription by RNA polymerase II (Reinberg and Roeder 1987; Sekimizu et al. 1979). These data give rise to a clue to understanding heterogeneously initiated transcription of ncRNAs from divergent sites of the human genome. Biochemical assay with nuclei of the mouse ascitic carcinoma Krebs II cells and RNA polymerase II with endogenous DNA as templates revealed strong activity of the transcription (Shenkin and Burdon 1966). Indeed, using 0.84 ml of the nuclear fraction, the yield of [^{3}H] RNA was achieved to range from 0.175 to 0.50 mg, indicating that significant percentage of the mouse genome is potentially transcribed at least in the experimental condition. Taken together with these data, the genome has the potential to be transcribed to create divergent RNA species. Yet unidentified protein factor will be shown to stimulate RNA polymerase II to make the great numbers of the long ncRNAs that have been identified recently.

2.6 Perspectives

The mechanisms of the transcriptional regulations discussed in this review indeed appear to be heterogeneous. Majority of the long ncRNAs utilizes histone modification to regulate transcription but not all. One common element for the transcriptional regulation by long ncRNAs is RNA–protein interaction through RNA-binding proteins. Formation of the RNA–protein complexes is one of key events of the long ncRNA-dependent transcriptional regulation. More generally, ncRNAs require their specific binding proteins in order to exert their biological functions, suggesting that identification of an RNA-binding protein specific to an unknown ncRNA should indicate its biological significance. Why are so many long ncRNAs generated in living cells? It should be informative for understanding the diversity of the long ncRNAs to elucidate mechanisms of the transcription of the long ncRNAs themselves. Considering that 90% of the genome is transcribed, the genomic DNA sequence intrinsically possesses the ability to be transcribed. It is likely that the protein-coding genes are evolutionally selected to acquire high efficiency of transcription (Fig. 2.5). The transcription mechanisms of long ncRNAs are supposed to be a primitive one compared to that of messenger RNAs of protein-coding genes, and a prototypic to the more refined RNA polymerase II transcription mechanism. To know more about the transcription of long ncRNAs will facilitate elucidation of the transcription of the coding genes in eukaryote. Employing the long ncRNAs as a regulator for transcription might be a way to salvage junks of the genome, long ncRNAs. Intense investigation of the long ncRNA transcription would lead to a crucial clue to understanding the origin of the long ncRNAs and also a whole structure of the human genome.

Acknowledgments The author thanks Ms. R. Tanji for the preparation of the manuscript and Dr. C.K.Glass for critical discussion. This work was supported by Takeda Science Foundation, the

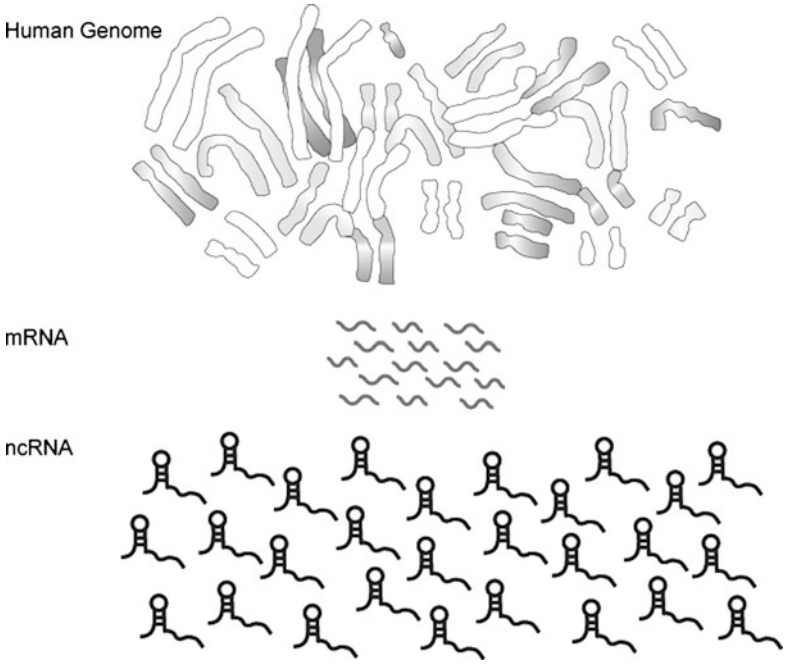

Fig. 2.5 Quantitative models of genomic DNA, protein-coding messenger RNAs, and long ncRNAs

Naito foundations, Astellas Foundation for Research on Metabolic Disorders Foundation, and also by Grant-in-Aid for Scientific Research (B: nos22390057) and Grant-in-aid for "Support Project of Strategic Research Center in Private Universities" from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to Saitama Medical University Research Center for Genomic Medicine.

References

Allen TA, Von Kaenel S, Goodrich JA, Kugel JF (2004) The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat Struct Mol Biol 11:816–821

Ambros V (2001) microRNAs: tiny regulators with great potential. Cell 107:823–826

Barbiroli B, Monti MG, Moruzzi MS, Mezzetti G (1977) Functional modification of liver form-B RNA polymerase activity by a protein fraction from rats accustomed to controlled feeding schedules. Biochim Biophys Acta 479:69–79

Bond AM, Vangompel MJ, Sametsky EA, Clark MF, Savage JC, Disterhoft JF, Kohtz JD (2009) Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. Nat Neurosci 12:1020–1027

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C et al (2005) The transcriptional landscape of the mammalian genome. Science 309:1559–1563

Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A (2007) The expanding RNA polymerase III transcriptome. Trends Genet 23:614–622

Espinoza CA, Goodrich JA, Kugel JF (2007) Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. RNA 13:583–596

Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD (2006) The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes Dev 20:1470–1484

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811

Glass CK, Rosenfeld MG (2000) The coregulator exchange in transcriptional functions of nuclear receptors. Genes Dev 14:121–141

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP et al (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458:223–227

Hatchell EC, Colley SM, Beveridge DJ, Epis MR, Stuart LM, Giles KM, Redfern AD, Miles LE, Barker A, MacDonald LM et al (2006) SLIRP, a small SRA binding protein, is a nuclear receptor corepressor. Mol Cell 22:657–668

Imamura T, Yamamoto S, Ohgane J, Hattori N, Tanaka S, Shiota K (2004) Non-coding RNA directed DNA demethylation of Sphk1 CpG island. Biochem Biophys Res Commun 322:593–600

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL et al (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316:1484–1488

Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J et al (2005) Antisense transcription in the mammalian transcriptome. Science 309:1564–1566

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A et al (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci USA 106:11667–11672

Kurokawa R, Rosenfeld MG, Glass CK (2009) Transcriptional regulation through noncoding RNAs and epigenetic modifications. RNA Biol 6:233–236

Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell 97:17–27

Legraverend M, Glazer RI (1980) Characterization of a non-histone chromosomal protein which stimulates RNA polymerase II. Biochim Biophys Acta 607:92–101

Lerga A, Hallier M, Delva L, Orvain C, Gallais I, Marie J, Moreau-Gachelin F (2001) Identification of an RNA binding specificity for the potential splicing factor TLS. J Biol Chem 276:6807–6816

Liu WM, Chu WM, Choudary PV, Schmid CW (1995) Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. Nucleic Acids Res 23:1758–1765

Louro R, Smirnova AS, Verjovski-Almeida S (2009) Long intronic noncoding RNA transcription: expression noise or expression choice? Genomics 93:291–298

Maraia RJ, Driscoll CT, Bilyeu T, Hsu K, Darlington GJ (1993) Multiple dispersed loci produce small cytoplasmic Alu RNA. Mol Cell Biol 13:4233–4241

Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol Cell 29:499–509

Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature 445:666–670

Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, Luscombe NM, Rinn JL, Nelson FK, Miller P, Gerstein M et al (2003) Distribution of NF-kappaB-binding sites across human chromosome 22. Proc Natl Acad Sci USA 100:12247–12252

Miyakawa Y, Matsushime H (2001) Rapid downregulation of cyclin D1 mRNA and protein levels by ultraviolet irradiation in murine macrophage cells. Biochem Biophys Res Commun 284:71–76

Nguyen VT, Kiss T, Michels AA, Bensaude O (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature 414:322–325

Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. Cell 136:629–641

Reinberg D, Roeder RG (1987) Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains. J Biol Chem 262:3331–3337

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E et al (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129:1311–1323

Roeder RG (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. Trends Biochem Sci 16:402–408

Roeder RG, Rutter WJ (1970) Specific nucleolar and nucleoplasmic RNA polymerases. Proc Natl Acad Sci USA 65:675–682

Sekimizu K, Nakanishi Y, Mizuno D, Natori S (1979) Purification and preparation of antibody to RNA polymerase II stimulatory factors from Ehrlich ascites tumor cells. Biochemistry 18:1582–1588

Shenkin A, Burdon RH (1966) Asymmetric transcription of deoxyribonucleic acid by deoxyribonucleic acid-dependent ribonucleic acid polymerase of Krebs II ascites-tumour cells. Biochem J 98:5C–7C

Wang P, Yin S, Zhang Z, Xin D, Hu L, Kong X, Hurst LD (2008a) Evidence for common short natural trans sense–antisense pairing between transcripts from protein coding genes. Genome Biol 9:R169

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R (2008b) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature 454:126–130

Weake VM, Workman JL (2010) Inducible gene expression: diverse regulatory mechanisms. Nat Rev Genet 11:426–437

Willingham AT, Gingeras TR (2006) TUF love for "junk" DNA. Cell 125:1215–1220

Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature 451:202–206

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322:750–756

Chapter 3
Long Noncoding RNAs and X Chromosome Inactivation

Cristina Gontan, Iris Jonkers, and Joost Gribnau

Abstract In female somatic cells, one of the two X chromosomes is inactivated to equalize the dose of sex-linked gene products between female and male cells. X chromosome inactivation (XCI) is initiated very early during development and requires *Xist*, which is a noncoding X-linked gene. Upon initiation of XCI, *Xist*-RNA spreads along the X chromosome *in cis*, and *Xist* spreading is required for the recruitment of different chromatin remodeling complexes involved in the establishment and maintenance of the inactive X chromosome. Because XCI acts chromosomewise, *Xist*-mediated silencing has served as an important paradigm to study the function of noncoding RNAs (ncRNA) in gene silencing. In this chapter, we describe the current knowledge about the structure and function of *Xist*. We also discuss the important *cis*- and *trans*-regulatory elements and proteins in the initiation, establishment, and maintenance of XCI. In addition, we highlight new findings with other ncRNAs involved in gene repression and discuss these findings in relation to *Xist*-mediated gene silencing.

3.1 Introduction

The evolution of mammalian sex chromosomes started about 150 million years ago by mutations in the *Sox3* gene that resulted in the new male sex determining gene *Sry* (Graves 2006). It is thought that after the birth of *Sry*, genes involved in male

C. Gontan and J. Gribnau (✉)
Department of Reproduction and Development, Erasmus MC, University Medical Center, Room Ee 09-71, PO Box 2040, 3000 CA Rotterdam, The Netherlands
e-mail: j.gribnau@erasmusmc.nl

I. Jonkers
Department of Reproduction and Development, Erasmus MC, University Medical Center, Room Ee 09-71, PO Box 2040, 3000 CA Rotterdam, The Netherlands
and
Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3_3,

fertility evolved in close vicinity of *Sry* and that the accumulation of this block of heterologous genes blocked homologous recombination, which led to the degeneration of the Y chromosome. The loss of the ancestral genes on the new Y chromosome was compensated by a twofold upregulation of these genes on the remaining single X chromosome in male cells (Nguyen and Disteche 2006). However, this would have led to the overexpression of these genes in female cells, and to compensate for this, a silencing process coevolved in the female that ensured downregulation of the expression of X-linked genes. Currently, this silencing process, called X chromosome inactivation (XCI), entails *cis* inactivation of almost the whole X chromosome in most eutherians. XCI occurs early in the development of the female embryo, in mice already after the 4-cell stage (Mak et al. 2004; Okamoto et al. 2004; Okamoto and Heard 2006). Cells in the early mouse embryo always inactivate the paternally inherited X chromosome (Xp) and leave the maternally inherited X chromosome (Xm) active, which is referred to as imprinted XCI (Takagi and Sasaki 1975; West et al. 1977).

In the mouse, at 3.5 days postcoitum (dpc), imprinted XCI is reversed in the inner cell mass (ICM) of the blastocyst, resulting in reactivation of the Xp and subsequent initiation of random XCI around 5.5 dpc, whereas imprinted XCI is maintained in the extraembryonic tissue (Rastan 1982; Mak et al. 2004). Random XCI is also initiated upon differentiation of female embryonic stem (ES) cells derived from the ICM, providing a convenient model system to study XCI in vitro (Chaumeil et al. 2002; Navarro et al. 2008). Also in other eutherian species, including human, XCI is random and initiated early in embryonic development. However, it is unclear whether imprinted XCI is present in other eutherian species besides the mouse. Unlike imprinted XCI, in random XCI, both X chromosomes have an equal chance to be inactivated, causing ~50% of the cells to have an active Xp and ~50% of the cells to have an active Xm (Lyon 1961). Only one of the two X chromosomes should be inactivated because inactivation of all Xs, or even leaving both Xs active, is lethal to the cell (Marahrens et al. 1997; Lee 2002). Therefore, the number of X chromosomes present in the cell must be determined in the developing embryo. When a female cell has established that two X chromosomes are present, XCI is initiated on one of the two X chromosomes. Once random XCI is completed, the process is irreversible, and after each cell division, the inactivated X (Xi) will be clonally propagated, meaning that the same X remains inactivated in all daughter cells (Plath et al. 2002).

In the last few decades, several *cis*- and *trans*-acting factors involved in the regulation of the XCI process have been identified. The two main regulatory factors involved in XCI are *Xist* and *Tsix* (Penny et al. 1996; Marahrens et al. 1997; Lee et al. 1999), both located in a small region on the X chromosome, called the X-inactivation center (*Xic*, Fig. 3.1). *Xist* and *Tsix* encode functional ncRNAs. *Xist* expression and RNA spreading *in cis* is necessary for XCI to occur while *Tsix* represses expression of *Xist in cis*. Together, these two genes determine whether XCI occurs *in cis* on the X chromosome. Other elements, proteins, or genes that are involved in regulation of XCI are DXPas34, *Xite*, *RepA*, RNF12,

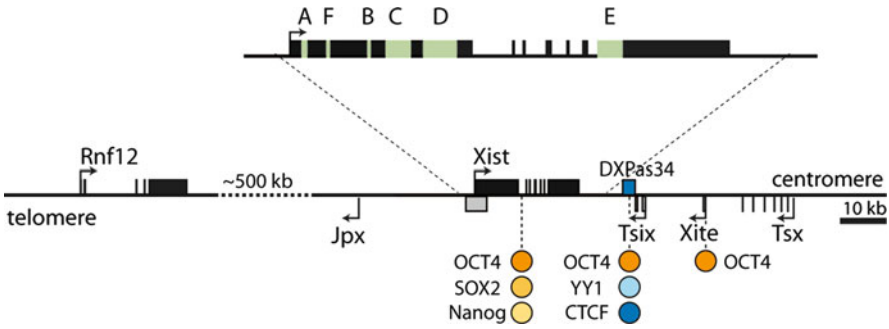

Fig. 3.1 Important players in XCI. Schematic representation of part of the X inactivation center including the *Xist*, *Tsix*, and *Rnf12*. Also shown is the localization of different repeats in *Xist* and the binding sites of different *trans*-acting factors involved in inhibiting XCI

OCT4, SOX2, NANOG, CTCF, and YY1, which seem to regulate *Xist* and/or *Tsix* expression and function, directly or indirectly, as described below.

3.2 *Cis*-Regulatory Factors in XCI

The most important player in XCI is *Xist*, which is located on the X chromosome and encodes a 17 kb long noncoding RNA, which is spliced and polyadenylated (Borsani et al. 1991; Brockdorff et al. 1991; Brown et al. 1991). Prior to XCI, *Xist* expression is low and the transcript is unstable. However, upon initiation of XCI, *Xist* expression is upregulated on the future inactive X (Xi) and spreads along the X chromosome *in cis*, thereby directly or indirectly attracting chromatin modifiers involved in the chromosome-wide silencing process (Brockdorff et al. 1992; Brown et al. 1992). Many experiments have shown the importance of *Xist* in the XCI process. For instance, deletion of *Xist* from one X chromosome in XX female ES cells causes complete skewing of XCI toward the wild type X chromosome, while XY male ES cells are not affected (Penny et al. 1996). This is not a consequence of secondary selection in benefit of female cells inactivating the wild type X chromosome after completion of XCI, but the wild type X chromosome is always inactivated when *Xist* is deleted on one allele in female XX embryos (primary nonrandom XCI) (Marahrens et al. 1997, 1998; Gribnau et al. 2005). Furthermore, ectopic expression and spreading of *Xist* is enough to initiate chromosome inactivation, even on an autosome (Lee et al. 1996; Herzing et al. 1997; Lee and Jaenisch 1997). Silencing, at least partially, of a chromosome from which *Xist* is transcribed is irreversible after 3 days of differentiation in ES cells, as has been shown using an inducible *Xist* transgene. However, when *Xist* RNA is removed beforehand, the silenced state of genes is reversed (Wutz and Jaenisch 2000). Importantly, the expression level of *Xist* is one of the factors that determines skewing of XCI, as

has been shown by changing the *Xist* transcription level on one of the two alleles by introducing a mutation or a deletion in the *Xist* promoter (Newall et al. 2001; Nesterova et al. 2003).

Xist contains different repeat sequences A–F, of which the A repeat is involved in gene silencing. Recent studies indicated that the A repeats form two stem loop structures, each containing four repeats, which attract the chromatin modifier complex PRC2 involved in gene silencing (Wutz et al. 2002; Maenner et al. 2010). The other sequences including repeats B–F play a redundant role in the proper localization of *Xist* to the X chromosome (Wutz et al. 2002). Comparison of the *Xist* genomic sequence across different eutherian species indicates that the *Xist* gene evolved very quickly and only revealed conservation of the promoter region and the different repeat structures (Nesterova et al. 2001). Recently, another smaller 1.6 kb ncRNA transcript, *RepA*, which partially overlaps with *Xist* and includes the A repeat, has been implicated to play a role in the initiation of XCI by locally attracting PRC2 prior to *Xist* spreading (Zhao et al. 2008). However, a clear function for RepA in the XCI process still needs to be established (Table 3.1).

Tsix is located 15 kb downstream from *Xist* and is transcribed in antisense direction of *Xist*. *Tsix* encodes a continuous antisense RNA of approximately 40 kb that spans all of *Xist*. Multiple transcription start sites for *Tsix* have been identified, and approximately 50% of the *Tsix* transcripts are spliced into various small isoforms of which the 3′ ends have an overlap with the promoter of *Xist*

Table 3.1 Noncoding RNAs and gene silencing

ncRNA	Size (kb)	Silencing	Distance	Dicer	G9A recruitment	PRC2 recruitment	References
Xist	17	*cis*	X chr.	?	n.d.	Yes	Maenner et al. (2010), Zhao et al. (2008), Ogawa et al. (2008), Kanellopoulou et al. (2009), and Nesterova et al. (2008)
Air	108	*cis*	250 kb	n.d.	Yes	n.d.	Nagano et al. (2008)
Kcnq1ot1	91	*cis*	400/780 kb	No	Yes	Yes	Pandey et al. (2008) and Redrup et al. (2009)
HOT*AIR*	2.2	*trans*	–	n.d.	n.d.	Yes	Rinn et al. (2007) and Gupta et al. (2010)

This table summarizes the features associated with different mammalian ncRNAs involved in *in cis* and *in trans* gene silencing
n.d. not determined
? = conflicting results

(Sado et al. 2001; Shibata and Lee 2003). *Tsix* is transcribed in male and female undifferentiated ES cells at a level about 10 to 100 times more than *Xist*, and during establishment of XCI from the allele that is to remain active in male and female differentiating ES cells. After completion of XCI, *Tsix* is downregulated (Lee et al. 1999; Shibata and Lee 2003).

Tsix is generally regarded as the major inhibitor of *Xist* and therefore as an important factor in XCI regulation. However, careful examination of the literature shows that overall antisense transcription through the *Xist* locus determines inhibition of *Xist*. For example, the loss of the major promoter of *Tsix* has no significant effect on the counting or initiation processes of XCI (Cohen et al. 2007). However, deletion of DXPas34, a CpG island located downstream of the *Tsix* transcription start site (TSS) from which antisense transcription is also initiated (Fig. 3.1), significantly decreases antisense transcription through the *Xist* locus and causes primary nonrandom inactivation of the targeted allele in female XX ES cells (Debrand et al. 1999; Vigneau et al. 2006; Cohen et al. 2007). Interestingly, the methylation status of DXPas34 coincides perfectly with the antisense transcription through *Xist*. The CpG island is hypomethylated when actively transcribed and hypermethylated when antisense transcription is downregulated (Prissette et al. 2001; Boumil et al. 2006). Antisense transcription is also initiated in a region ~10 kb upstream of *Tsix*, called *Xite*. *Xite* expression and the methylation pattern during XCI is similar to that of *Tsix*, and deletion of Xite results in reduced antisense transcription through the *Xist* locus and skewing of XCI toward inactivation of the targeted allele (Ogawa and Lee 2003; Stavropoulos et al. 2005, Boumil et al. 2006), implying a similar role for *Xite* in inhibition of *Xist* function as DXPas34 and *Tsix*. Furthermore, direct inhibition of antisense transcription by insertion of a polyA site between *Xist* and DXPas34 also causes primary nonrandom XCI in female ES cells and inappropriate XCI in male ES cells. Even more so, overexpression of antisense transcription on one allele results in primary nonrandom inactivation of the wild type allele (Luikenhuis et al. 2001). Finally, a 65 kb deletion encompassing not only *Tsix* but also Xite and DXPas34, thus abrogating all antisense transcription, shows not only complete primary nonrandom XCI of the targeted allele but also severe cell death in X0 and XY cells containing the deletion, invoked by improper XCI (Clerc and Avner 1998; Morey et al. 2004). Thus, inhibition of *Xist* seems to correlate with an increase in antisense transcription through the *Xist* locus.

3.3 *Xist* Versus *Tsix*

How does *Tsix* inhibit *Xist* expression? Different hypotheses have been proposed. First, *Tsix* may function by forming a double-stranded RNA heteroduplex with *Xist*, resulting in repressive small interfering RNA (siRNA), which functionally silences *Xist in cis* (Ogawa et al. 2008). However, overexpression of *Tsix* cDNA, which includes the homologous region with *Xist* on an allele with abrogated endogenous *Tsix* transcription by insertion of a polyA signal, does not restore *Xist* inhibition

(Shibata and Lee 2004), arguing against RNA interference (RNAi)-based inhibition of *Xist*. Also, Dicer knockout mice and ES cells that have an impaired RNAi machinery exhibit correct XCI, although *Xist* is ectopically upregulated at later stages due to loss of DNA methylation at the *Xist* promoter (Nesterova et al. 2008; Kanellopoulou et al. 2009).

Secondly, *Tsix* and *Xite* might form a three-dimensional chromatin structure via DNA looping that enhances *Tsix* and *Xite* antisense transcription but excludes the *Xist* promoter and thereby inhibits *Xist* expression *in cis*. A chromosome-conformation-capture (3C) study has shown that *Tsix* and *Xite* interact over a long distance, while the *Xist* promoter seems to colocalize with the *Jpx* promoter when *Xist* is transcribed (Tsai et al. 2008). DXPas34 is a likely candidate for looping because deletion of DXPas34 causes a severely skewed phenotype in female ES cells and XCI in male ES cells (Debrand 1999; Vigneau et al. 2006; Cohen et al. 2007). Moreover, DXPas34 is bound by CTCF, a protein that is often implicated in the looping of DNA (Chao et al. 2002). However, the DXPas34 deletion does not significantly change the three-dimensional chromatin structure in male ES cells. Furthermore, it is hard to determine whether a specific three-dimensional chromatin conformation *in cis* is the cause or the consequence of the transcription profile of that allele (Tsai et al. 2008).

Finally, antisense transcription through the *Xist* locus may inhibit *Xist* upregulation through a transcription interference mechanism. How antisense transcription-based inhibition of *Xist* works mechanistically has not been shown but one can envision that promoter polymerase initiation complexes (PICs) will have more difficulty forming on a promoter when an elongation complex transcribing in the antisense direction coe*Xist*s at the locus (Shearwin et al. 2005). Furthermore, RNA polymerase II complexes of *Xist* and *Tsix* may collide during transcription elongation, causing a premature halt of *Xist* transcription and less *Xist* accumulation. Evidence for involvement of such a mechanism comes from studies that indicate a bimodal pattern of both *Xist* and *Tsix* transcripts, being highest at the transcription start site and gradually decreasing along the template (Shibata and Lee 2003; Marks et al. 2009). Alternatively, inhibition of *Xist* might be caused by alteration of the chromatin state of the *Xist* locus by the *Tsix* transcript. It has been postulated that *Tsix* transcription induces heterochromatin formation at the *Xist* promoter by *Tsix*-mediated recruitment of histone modifiers (Sado et al. 2005; Navarro et al. 2006). Recently, EED, a component of the PRC2 Polycomb complex, has been shown to work synergistically with *Tsix* in silencing *Xist* (Shibata et al. 2008). Furthermore, loss of antisense transcription through the *Xist* promoter causes reduction of CpG methylation and repressive histone modification marks, indicating that transcription from the *Xist* promoter is enhanced (Ohhata et al. 2008). However, findings of Sun et al. (2006) argue against this hypothesis by showing that activation of *Xist* on the future Xi is characterized by a transient heterochromatic state at the *Xist* promoter, perhaps induced by the silencing capacity of *Xist* itself and thus contradicting a functional role of chromatin modifications in the inhibition of *Xist* by *Tsix*. In conclusion, most evidence points toward a transcription or *Tsix* RNA-mediated mechanism of repression of *Xist* by *Tsix*, but the exact mechanism has yet to be established.

3.4 Trans-Regulatory Factors and Initiation of XCI

In the recent years, several trans-acting factors regulating XCI have been identified. Most of these factors are involved in suppression of XCI (XCI-inhibitors), either by repressing *Xist* or activating *Tsix*. Among the proteins involved in *Tsix* regulation are the insulator protein CTCF, and also the transcription factor Yin Yang 1 (YY1), for which several tandemly organized binding sites have been identified in the *DXpas34* region, which is involved in *Tsix* regulation and in the *Xite* promoter. Knockout studies involving *Yy1*, or partial ablation of *Yy1* and *Ctcf* through RNAi-mediated repression, revealed downregulation of *Tsix* expression and concomitant upregulation of *Xist* expression, supporting a role for YY1 and CTCF in activation of *Tsix* expression (Donohoe et al. 2007).

The pluripotency factors SOX2, Nanog, and OCT4 have also been shown to be involved in the regulation of XCI by the silencing of *Xist* (Donohoe et al. 2007, 2009; Navarro et al. 2008). A binding site for all three factors has been identified in intron 1 of *Xist*, and binding of these factors is involved in the direct suppression of *Xist*. Interestingly, OCT4 and Sox2 also bind in the *Xite* enhancer, and OCT4 together with YY1 is recruited to *Tsix* downstream of the transcription start site and is involved in transcription activation of both *Xite* and *Tsix*. These factors therefore affect *Xist* expression through both *Tsix*-dependent and -independent pathways, indicating that different mechanisms act jointly in setting up the threshold that has to be overcome by *Xist*.

Autosomally encoded factors such as SOX2, OCT4, and Nanog play an important role in XCI. However, it can be excluded that sex-specific initiation of XCI is determined by these factors only because the concentration of these factors, if not regulated by (a) sex-chromosomal factor(s), will most likely be the same in male and female cells. Key to the XCI initiation process is therefore the presence of one or more X-encoded XCI-activators that are differentially expressed between male and female cells. Recently, the E3 ubiquitin ligase RNF12 has been identified as a dose-dependent X-linked activator of XCI (Jonkers et al. 2009). Additional copies of *Rnf12* resulted in ectopic initiation of XCI in transgenic male cells and initiation of XCI on both X chromosomes in a high percentage of female cells. RNF12 may act through activation of *Xist* or suppression of *Tsix*, although the exact mechanism remains elusive so far. Also, *Rnf12* cannot be the only XCI-activator because $Rnf12^{+/-}$ female cells still induce XCI, albeit in a severely reduced percentage of cells, indicating that other X-encoded genes are involved in initiation of XCI (Jonkers et al. 2009).

Different mechanisms for counting the number of X chromosomes and initiation of XCI have been proposed. Most of these models explain XCI as a mutually exclusive process leading to one single Xi per female cell, for instance, through the protection of one X chromosome by an autosomally encoded blocking factor or p*Air*ing and cross communication of both X chromosomes in female cells (Wutz and Gribnau 2007; Jonkers et al. 2009; Starmer and Magnuson 2009). However, recent studies indicate that XCI is more likely to be a stochastic process and that in

female cells, both X chromosomes have a probability to initiate XCI (Monkhorst et al. 2008; Barakat et al. 2010). The probability to initiate XCI is determined by the nuclear concentration of the different XCI-activators and -inhibitors (Monkhorst et al. 2008, 2009). XCI-inhibitors set the threshold by suppression of *Xist* and activation of *Tsix*, which has to be overcome by the action of the XCI-activators. Only in female cells, the nuclear concentration of the XCI-activators is sufficient to boost enough *Xist* transcription, allowing spreading and initiation of XCI *in cis*. Because the XCI-activators are X-linked, initiation of XCI on one X results in rapid downregulation of the XCI-activator genes *in cis*, preventing initiation of XCI on the second X chromosome. Nonetheless, XCI can still be initiated on the remaining active X chromosome until enough XCI-activator protein is degraded after inactivation, which would lead to a female cell with two inactive X chromosomes. Indeed, a small percentage of female cells initiating XCI on both X chromosomes is found during the XCI process, and as expected when XCI-inhibitors are downregulated, or the XCI-activator *Rnf12* is upregulated, this percentage of XiXi cells increases significantly. These results indicate that the regulation of XCI is determined by a tightly regulated balance of X-encoded activators and autosomally encoded inhibitors of XCI.

3.5 Establishment of the Inactive X

The first step in silencing the X chromosome is the spread of *Xist* RNA *in cis* over the X chromosome. Several redundant repeats of *Xist* are important for the localization of *Xist* RNA to the Xi (Wutz et al. 2002). Spreading of *Xist* causes depletion of RNA polymerase II and other components of the transcription machinery on the Xi within one day, and abrogates transcription of repeat and intergenic sequences, independently of the A-repeat (Chaumeil et al. 2006). However, silencing of X-linked genes is mediated by the A-repeat within *Xist* RNA and starts after 1–2 days, continuing until gene silencing is more or less completed after approximately 7 days of differentiation (Chaumeil et al. 2006; Lin et al. 2007). Silencing of genes is hypothesized to be associated with the relocation of active genes at the outer rim of the X chromosome territory toward the silent Xi territory invoked by the A-repeats (Chaumeil et al. 2006; Clemson et al. 2006; Lin et al. 2007).

After depletion of the transcription machinery from the Xi territory, the Xi chromatin is changed drastically (Fig. 3.2a, b). First, histone 3 lysine 27 trimethylation (H3K27me3) is acquired by transient localization to the Xi of the Polycomb repressor complex 2 (PRC2), which comprises protein subunits EED, EzH2, RbAp47/48, and Suz12, of which EzH2 has histone methyltransferase activity (Wang et al. 2001; Plath et al. 2003; Silva et al. 2003; Cao and Zhang 2004; de la Cruz et al. 2005). PRC2 is recruited by *Xist* RNA, as has been shown by either deletion of *EED* or conditional deletion of *Xist*, which both cause loss of H3K27me3 (Wang et al. 2001; Plath et al. 2003, 2004). PRC2 subunit EzH2 has been identified as the protein that targets the PRC2 complex to the A-repeat of *Xist*

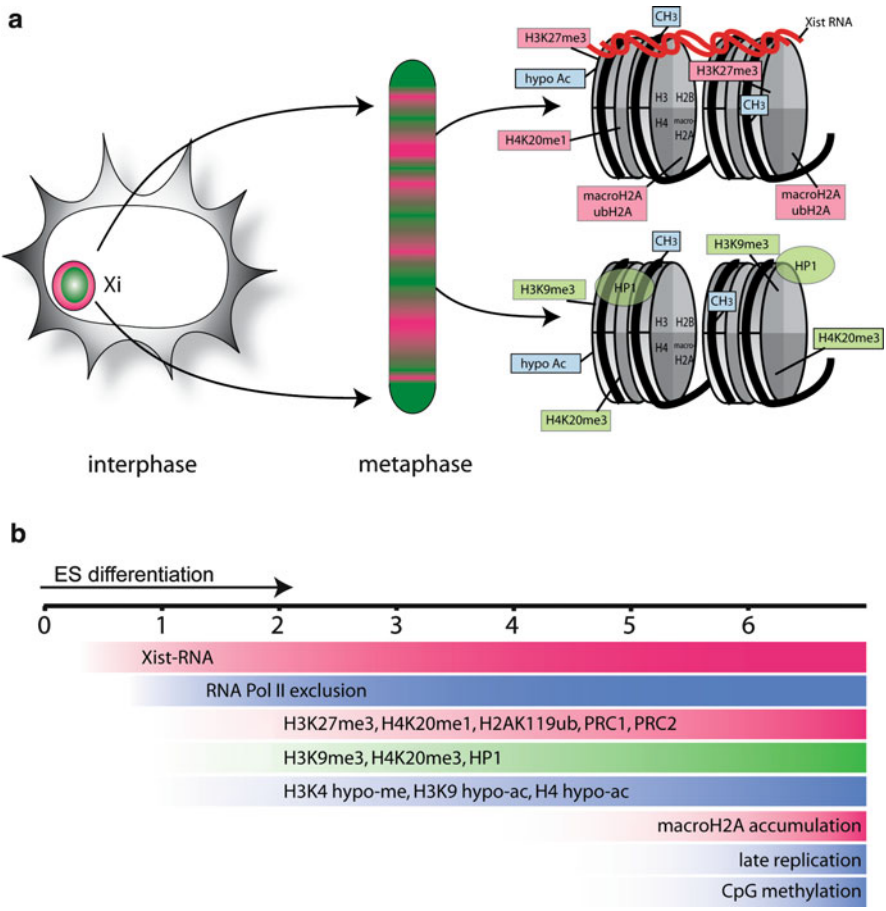

Fig. 3.2 The landscape of chromatin modifications on the inactive X. *(***a**) On the left, the Xi in interphase is shown consisting of two distinct regions of heterochromatin, in *pink* and *green*. *Xist* RNA association, and H3K27me3, and ubH2A accumulation, among others, characterize the pink chromatin, whereas histone marks such as H3K9me3 and recruitment of HP1 characterize the green chromatin. The different chromatin states form a banded pattern on the inactive X chromosome in metaphase. On the right, the specific histone marks and other epigenetic features are depicted for the *Xist* associated pink chromatin (*top*) and green chromatin (*bottom*). (**b**) A large number of epigenetic changes are associated with the XCI process. The temporal changes, when induced by differentiation of female ES cells, are depicted along the timescale (days) and separated in color (*pink* or *green*) depending on which heterochromatin state the modification is associated with (as described in **a**). Changes associated with both heterochromatin states are shown in *blue*

RNA (Zhao et al. 2008), although a more recent study indicated that SUZ12 may play a more important role in targeting PRC2 to *Xist* (Kanhere et al. 2010). Although PRC2 seems to be important for binding *Xist* to the Xi, it is not likely to be the only protein complex doing so because loss of PRC2 does not seem to affect random XCI in the embryo proper (Wang et al. 2001; Plath et al. 2003).

Apart from histone methylation, most cells also show accumulation of H2A lysine 119 ubiquitination (ubH2A) on the Xi after the onset of XCI, which is established by the Ring1A/B subunit of Polycomb repressor complex 1 (PRC1) (de Napoles et al. 2004; Fang et al. 2004; Plath et al. 2004). *Ring1A* and *Ring1B* have redundant functions in ubiquitination (de Napoles et al. 2004; Leeb and Wutz 2007), and only deletion of both *Ring1* genes results in loss of ubH2A on the Xi (de Napoles et al. 2004). PRC1 recruitment to the Xi follows PRC2 recruitment, but is not solely mediated by H3K27me3, as has been shown in *EED*-deficient ES cells, but also by the 3′ end of *Xist* RNA, either directly through interaction with *Xist* or by indirect interaction with an *Xist* binding protein (Plath et al. 2004; Schoeftner et al. 2006). A potential candidate for targeting of the PRC1 complex to *Xist* RNA is the Polycomb homolog CBX7, which shows a high affinity for H3K27me3 and for RNA (Bernstein et al. 2006) and has been shown to interact with the Ring1 protein (Gil et al. 2004).

Another histone methylation mark associated with silenced chromatin, histone 3 lysine 9 trimethylation (H3K9me3), accumulates on the Xi just after H3K27me3 (Heard et al. 2001; Boggs et al. 2002; Mermoud et al. 2002; Peters et al. 2002; Rougeulle et al. 2004). H3K9me3 is most likely put in place by HMTase *Suv39*, and maintained by HP1, which is enriched on the Xi (Chadwick and Willard 2003, 2004), but other histone methyltransferases (HMTases) might also play a role.

H3K9me3 accumulation appears more or less simultaneous with the loss of acetylation of histone H3 and H4 (H3K9Ac and H4K5Ac, H4K8Ac and H4K12Ac, respectively) and trimethylation of histone H3 lysine 4 (H3K4me3) and histone H3 lysine 36 (H3K36me3), which are all hallmarks of euchromatin (Jeppesen and Turner 1993; Belyaev et al. 1996; Boggs et al. 1996, 2002; Keohane et al. 1996; Heard et al. 2001; Chaumeil et al. 2002; Chadwick and Willard 2003). Probably, a set of histone modifiers, including histone deacetylases (HDACs) and histone demethylases (HDMs), are attracted by H3K27me3 and *Xist* and colocalize with the Xi to direct the chromatin toward a heterochromatic state. Among the late epigenetic changes are macroH2A incorporation (Costanzi and Pehrson 1998; Mermoud et al. 1999), CpG island methylation, and late replication (Priest et al. 1967; Mohandas et al. 1981; Norris et al. 1991). MacroH2A is a H2A variant with a large C-terminal domain (Nusinow et al. 2007) that replaces H2A histones on the Xi after approximately 7 days of differentiation, forming a macrochromatin body (MCB) in a significant proportion of the cells (Costanzi and Pehrson 1998; Rasmussen et al. 2001). *Xist* expression is sufficient for initiation of H2A replacement by macroH2A and MCB formation (Rasmussen et al. 2001), and conditional deletion of *Xist* leads to loss of the MCB (Csankovszki et al. 1999). CpG methylation is also a late Xi mark and is put in place by de novo methyltransferase 3A (DNMT3A) (Hansen 2003) and maintained by DNMT1 (Sado et al. 2000).

Recently, several other factors have been shown to be involved in the maintenance phase of XCI. First, the DNA binding hinge-domain protein SmcHD1 plays a role in DNA methylation of the Xi. Loss of SmcHD1 results in depletion of DNA methylation at the X-linked CpG islands and reactivation of the Xi (Blewitt et al. 2008). It was postulated that SmcHD1 targets DNMT3A to the Xi, although no

direct evidence in that direction was presented. Second, ATRX, encoded by an X-linked gene, has been shown to be involved in XCI. ATRX is a chromatin remodeler and a member of the SWI/SNF2 helicase family, which is enriched at the Xi, and the accumulation of ATRX can be regarded as a late mark of the Xi (Baumann and De la Fuente 2008). Interestingly, ATRX does repress not only X-linked genes on the Xi but also pseudo-autosomal genes that have translocated to an autosome, implicating that a (former) X chromosomal sequence is required to attract ATRX to a gene (Levy et al. 2008). Also, SATB1, which has been implicated in nuclear organization and involved in many forms cancers, has been identified as an important factor in *Xist*-mediated gene silencing (Agrelo et al. 2009). Expression of SATB1 allows *Xist*-mediated gene silencing even after the developmental window where *Xist* silencing is normally restricted to, indicating that SATB1 plays a key role in the establishment of the Xi. SAF-A is another factor involved in nuclear organization which plays an important role together with the tritorax protein Ashl, in the establishment of the Xi (Pullirsch et al. 2010). Both proteins, together with macroH2A, are involved in chromosome-wide histone H4 hypoacetylation, Interestingly, recruitment of most of the mentioned factors including components of PRC1 and PRC2, and SATB1, Ashl1, and SAF-A is not dependent on the A repeat of *Xist*, which is required for *Xist* mediated silencing of the Xi. This suggests that chromatin changes evoked by these proteins and protein complexes provide a repressive nuclear compartment, which may be required for subsequent gene silencing on the Xi mediated by the *Xist* A repeat. The recent discovery of these factors indicates that silencing of the Xi is more complex than initially thought and involves multiple factors, of which many are probably not yet revealed.

All these features of the Xi are important to lock-in the silenced state of the X chromosome. Together, they ensure that the Xi is nearly impossible to reactivate. The redundancy of the Xi hallmarks is demonstrated by conditional deletion of *Xist* after establishment of XCI, which causes loss of the macroH2A (Csankovszki et al. 1999) but still only leads to minor reactivation of the Xi, even when it is combined with loss of DNA methylation and inhibition of hypoacetylation (Csankovszki et al. 2001; Hernández-Muñoz et al. 2005).

3.6 *Xist* Spreading, Xi Organization, and Nuclear Organization

After *Xist* is upregulated on one of the two X chromosomes, it starts to spread *in cis* over the entire chromosome (Clemson et al. 1996; Hall and Lawrence 2003). *Xist* RNA is restricted to the inactivated X chromosome and does not localize to neighboring autosomes (Brown et al. 1992; Jonkers et al. 2008). Furthermore, studies on X:autosome translocations show that endogenously expressed *Xist* preferentially binds the X chromosomal part of the chromosome (Duthie et al. 1999; Keohane et al. 1999; Popova et al. 2006), and spreading into the autosome seems to be correlated with the density of LINE repeats (Popova et al. 2006).

This observation has led to the LINE repeat hypothesis (Lyon 1998), in which it is stated that spreading of *Xist* is mediated by binding to LINE repeats. Indeed, LINE repeats are enriched twofold on the human X chromosome compared to autosomes, and the distribution of LINE repeats seems to correlate with the degree of XCI on the X chromosome (Boyle et al. 1990; Bailey et al. 2000; Ross et al. 2005). Also, computational studies of the DNA sequence surrounding genes escaping XCI compared to silenced X-chromosomal genes indicate a depletion of LINE repeats around escaping genes (Carrel et al. 2006; Wang et al. 2006).

Not all computational studies on the DNA sequence of the X chromosome find a clear correlation between LINE repeats and XCI (Chureau et al. 2002; Ke and Collins 2003). Also, *Xist* RNA does not spread over the X chromosome homogenously but appears to have a banded pattern when detected on a metaphase Xi and an open circle shape at the periphery of the Xi in interphase cells (Fig. 3.2a, left) (Duthie et al. 1999; Chadwick and Willard 2004; Smith et al. 2004). Curiously, this *Xist* RNA localization pattern does not seem to correspond to the density of underlying LINE repeats, but rather to the gene density on the X chromosome (Smith et al. 2004; Clemson et al. 2006). The banded pattern on the metaphase Xi of *Xist* RNA and gene rich regions can also be observed with histone marks H3K27me3, macroH2A and ubH2A, while histone marks H4K20me3 and H3K9me3 are enriched on the gene-poor regions of the Xi metaphase chromosome (Fig. 3.2a, b) (Gilbert et al. 2000; Chadwick and Willard 2004; Smith et al. 2004; Chadwick 2007). Therefore, the importance of the DNA sequence in the silencing process remains elusive as a direct interaction of LINE repeats or another specific DNA motive with histone marks and/or *Xist* RNA has not yet been reported.

Together, these data suggest a three-dimensional organization of the Xi, in which the gene-poor regions enriched in histone marks H4K20me3 and H3K9me3 are more internally located and the gene rich-regions, enriched in *Xist* RNA, H3K27me3, macroH2A, and ubH2A are present on the outer rim of the Xi territory (Chadwick and Willard 2004; Chaumeil et al. 2006; Clemson et al. 2006). Overall, the Xi becomes more spherical but retains a similar volume to the Xa (Eils et al. 1996). This Xi organization corresponds to DNA-FISH analysis of escaping and silenced X chromosomal genes, which shows that all analyzed genes are localized at the periphery of the Xi territory, but that active genes seem to "loop-out" of the chromosome territory (Dietzel et al. 1999; Chaumeil et al. 2006; Clemson et al. 2006). Early during the XCI process, *Xist* accumulation results in transcriptionally silent compartment devoid of RNA polymerase II and enriched for heterochromatin marks (Chaumeil et al. 2006). Interestingly, at this stage, only repetitive DNA is repressed and located within this silent compartment, and subsequent silencing of X-linked genes is accompanied by a shift in the localization of these genes toward a more internal localization. This change in localization and silencing of X-linked genes requires the presence of the *Xist* A repeat, in contrast to the RNA polymerase II excluded silent compartment that is also formed without the A repeat. Whether relocalization of X-linked genes upon XCI is the consequence of the XCI process itself or is directly involved in enforcing gene inactivation remains to be determined.

The Xi might not only have an intrinsic three-dimensional organization but is also specifically positioned within the nucleus. After inactivation, the Xi is preferentially located either at the periphery of the nucleus (Bourgeois et al. 1985; Belmont et al. 1986) or near the perinucleolar region (Bourgeois et al. 1985; Zhang et al. 2007). The specific positioning of the Xi could be mediated by the components of nuclear matrix. For instance, nuclear matrix scaffold protein SAF-A colocalizes with the Xi, which seems to be dependent on the RNA binding domain of the protein (Helbig and Fackelmayer 2003, Fackelmayer 2005). Furthermore, cells expressing mutated LaminA show depletion of heterochromatic marks H3K27me3 and H3K9me3 at the Xi, and the peripheral localization of the Xi is lost (Shumaker et al. 2006). These results indicate that the localization of the Xi in the nuclear periphery is either a consequence of its heterochromatic state or affects the heterochromatic state of the Xi (Shumaker et al. 2006; Fedorova and Zink 2008). However, the perinucleolar localization of the Xi is less easy to comprehend, especially because the Xi seems to preferentially colocalize with the perinucleolar region during S phase (Zhang et al. 2007). The S phase-specific localization is dependent on *Xist*, as autosomes containing an *Xist* transgene are also repositioned to the perinucleolar region in S phase, and conditional *Xist* knockout cells loose the preferential perinucleolar localization of the Xi. Interestingly, heterochromatin replication occurs late during S phase, at which point replication can only be observed around nucleoli and at the periphery of the nucleus (O'Keefe et al. 1992; Kennedy et al. 2000). Thus, perhaps, heterochromatin characterized by H3K27me3 needs a specialized nuclear compartment for replication and/or maintenance of the silenced state after replication.

3.7 Other Functional ncRNAs

The discovery of *Xist* provided a powerful model system to study the role and function of long ncRNA's. Besides *Xist*, several other ncRNAs have been described to be involved in gene silencing *in cis* and *in trans*, and several parallels can be drawn between the action of these RNAs. *Air* and *Kcnq1ot1* are two well-studied imprinted genes, both encoding noncoding transcripts involved in silencing *in cis*. *Air* encodes a 108-kb long unspliced transcript, which is transcribed antisense to the protein coding gene *Igf2r* (Lyle et al. 2000). *Air* expression is exclusively paternal, whereas *Igf2r* is maternally expressed. Besides *Air*-mediated silencing of the overlapping *Igf2r* gene, silencing also involves genes, including *Slc22a3*, located more than 200 kb away from *Air*, suggesting a direct role for the *Air* transcript in long range gene silencing. *In cis* silencing by *Air* involves the recruitment of G9A, required for H3K9 mono and dimethylation, and similar to *Xist* spreading (although less robust) the *Air* RNA appears to form a silent nuclear domain that envelops the paternal *Slc22a3* locus. Interestingly, G9A appears to be needed for the silencing of *Slc22a3* but not for the repression of *Igf2r* (Nagano et al. 2008). This finding indicates that different mechanisms may be involved in the regulation of antisense transcribed overlapping genes (*Air*/*Igf2r*) and long range gene silencing *in cis*

(*Air*/*Slc22a3*). This is reminiscent of findings obtained with the regulation of the *Xist*/*Tsix* locus and silencing of X-linked genes *in cis*, which also supports the presence of different mechanisms involved in these processes.

Expression of *Kcnq1ot1* is also imprinted, and the 91 kb paternally expressed gene is transcribed antisense to, and partially overlaps with *Kcnq1* (Fitzpatrick et al. 2002; Pandey et al. 2008). *Kcnq1ot1* is involved in the regulation of a cluster of imprinted genes on mouse chromosome 7 (Mancini-Dinardo et al. 2006). *In cis* silencing of *Kcnq1ot1* spans a region of 400 kb in the embryo and 780 kb in the placenta and involves recruitment of several chromatin modifiers including G9A and PRC2 (Pandey et al. 2008; Redrup et al. 2009). Similar to *Xist* and *Air*, RNA FISH studies indicate that *Kcnq1ot1* appears to form a silent nuclear domain, which is larger in the placenta than in the embryo. Interestingly, the *Kcnq1* domain is also found in close proximity to the nucleolus in a high percentage of cells, suggesting a lineage-specific localization of this locus with the nucleolus (Pandey et al. 2008). Whether there is a functional role for this localization close to the nucleolus and whether the localization is dependent on spreading of *Kcnq1ot1* remain to be determined.

In contrast to *Kcnq1ot1* and *Air*, which invoke silencing *in cis*, HOTAIR has recently been identified as an ncRNA involved in silencing *in trans* (Rinn et al. 2007). HOTAIR is a 2.2-kb RNA expressed from the HOXC locus, which represses transcription of different Hox genes in the HOXD locus, which is located on a different chromosome. HOTAIR associates with PRC2, which mediates silencing of HOXD genes *in trans*, through H3k27me3 of target genes. HOTAIR expression is increased in several tumors, and loss of HOTAIR expression inhibits cancer invasiveness (Gupta et al. 2010). Interestingly, overexpression of HOTAIR results in genome-wide changes in the targeting of PRC2 and increased cancer invasiveness. This indicates that HOTAIR plays a much broader role in targeting of PRC2, besides regulation of the HOXD locus.

Recently, a genome-wide study indicated the presence of more than 3,300 large ncRNAs (Guttman et al. 2009; Khalil et al. 2009). About 20% of these ncRNAs associate with PRC2 and other chromatin modification complexes, indicating that findings with *Xist* and other well-studied ncRNAs including *Air*, *Kcnq1ot1*, and HOT*AIR* may be extrapolated to explain the function of these newly identified ncRNAs.

3.8 Conclusion

The discovery of *Xist* exemplified the importance of ncRNAs in cellular function. *Xist* was the first identified mammalian large ncRNA involved in gene silencing, providing a powerful model system to study RNA-mediated gene silencing. New advances in RNA sequencing indicate that many more ncRNAs will soon be identified as functional RNAs, and unraveling the role of *Xist* in XCI will help in understanding the function of these ncRNAs in the regulation of gene expression.

Nevertheless, despite a lot of progress in understanding the role of *Xist* in XCI, many questions remain. For instance, how is the binding specificity of RNA binding proteins and complexes generated, which proteins are involved in fixing *Xist* to the chromatin, and why are so many proteins implicated in XCI dispensable for the XCI process? We are hopeful that the quickly advancing technology allows these questions to be addressed in the near future.

References

Agrelo R, Souabni A, Novatchkova M, Haslinger C, Leeb M, Komnenovic V, Kishimoto H, Gresh L, Kohwi-Shigematsu T, Kenner L, Wutz A (2009) SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells. Dev Cell 16:507–516

Bailey JA, Carrel L, Chakravarti A, Eichler EE (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proc Natl Acad Sci USA 97:6634–6639

Barakat TS, Jonkers I, Monkhorst K, Gribnau J (2010) X-changing information on X inactivation. Exp Cell Res 316:679–687

Baumann C, De La Fuente R (2008) ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells. Chromosoma 118:209–222

Belmont AS, Bignone F, Ts'o PO (1986) The relative intranuclear positions of Barr bodies in XXX non-transformed human fibroblasts. Exp Cell Res 165:165–179

Belyaev N, Keohane AM, Turner BM (1996) Differential underacetylation of histones H2A, H3 and H4 on the inactive X chromosome in human female cells. Hum Genet 97:573–578

Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol 26:2560–2569

Blewitt ME, Gendrel AV, Pang Z, Sparrow DB, Whitelaw N, Craig JM, Apedaile A, Hilton DJ, Dunwoodie SL, Brockdorff N, Kay GF, Whitelaw E (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat Genet 40:663–669

Boggs BA, Connors B, Sobel RE, Chinault AC, Allis CD (1996) Reduced levels of histone H3 acetylation on the inactive X chromosome in human females. Chromosoma 105:303–309

Boggs BA, Cheung P, Heard E, Spector DL, Chinault AC, Allis CD (2002) Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nat Genet 30:73–76

Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C et al (1991) Characterization of a murine gene expressed from the inactive X chromosome. Nature 351:325–329

Boumil RM, Ogawa Y, Sun BK, Huynh KD, Lee JT (2006) Differential methylation of Xite and CTCF sites in Tsix mirrors the pattern of X-inactivation choice in mice. Mol Cell Biol 26:2109–2117

Bourgeois CA, Laquerriere F, Hemon D, Hubert J, Bouteille M (1985) New data on the in-situ position of the inactive X chromosome in the interphase nucleus of human fibroblasts. Hum Genet 69:122–129

Boyle AL, Ballard SG, Ward DC (1990) Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence in situ hybridization. Proc Natl Acad Sci USA 87:7757–7761

Brockdorff N, Ashworth A, Kay GF, Cooper P, Smith S, McCabe VM, Norris DP, Penny GD, Patel D, Rastan S (1991) Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature 351:329–331

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71:515–526

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349:38–44

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71:527–542

Cao R, Zhang Y (2004) SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED–EZH2 complex. Mol Cell 15:57–67

Carrel L, Park C, Tyekucheva S, Dunn J, Chiaromonte F, Makova KD (2006) Genomic environment predicts expression patterns on the human inactive X chromosome. PLoS Genet 2:e151

Chadwick BP (2007) Variation in Xi chromatin organization and correlation of the H3K27me3 chromatin territories to transcribed sequences by microarray analysis. Chromosoma 116:147–157

Chadwick BP, Willard HF (2003) Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. Hum Mol Genet 12:2167–2178

Chadwick BP, Willard HF (2004) Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. Proc Natl Acad Sci USA 101:17450–17455

Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT (2002) CTCF, a candidate trans-acting factor for X-inactivation choice. Science 295:345–347

Chaumeil J, Okamoto I, Guggiari M, Heard E (2002) Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells. Cytogenet Genome Res 99:75–84

Chaumeil J, Le Baccon P, Wutz A, Heard E (2006) A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev 20:2223–2237

Chureau C, Prissette M, Bourdet A, Barbe V, Cattolico L, Jones L, Eggen A, Avner P, Duret L (2002) Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. Genome Res 12:894–908

Clemson CM, McNeil JA, Willard HF, Lawrence JB (1996) XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol 132:259–275

Clemson CM, Hall LL, Byron M, McNeil J, Lawrence JB (2006) The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. Proc Natl Acad Sci USA 103:7688–7693

Clerc P, Avner P (1998) Role of the region 3′ to Xist exon 6 in the counting process of X-chromosome inactivation. Nat Genet 19:249–253

Cohen DE, Davidow LS, Erwin JA, Xu N, Warshawsky D, Lee JT (2007) The DXPas34 repeat regulates random and imprinted X inactivation. Dev Cell 12:57–71

Costanzi C, Pehrson JR (1998) Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature 393:599–601

Csankovszki G, Panning B, Bates B, Pehrson JR, Jaenisch R (1999) Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat Genet 22:323–324

Csankovszki G, Nagy A, Jaenisch R (2001) Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J Cell Biol 153:773–784

de la Cruz CC, Fang J, Plath K, Worringer KA, Nusinow DA, Zhang Y, Panning B (2005) Developmental regulation of Suz 12 localization. Chromosoma 114:183–192

de Napoles M, Mermoud JE, Wakao R, Tang YA, Endoh M, Appanah R, Nesterova TB, Silva J, Otte AP, Vidal M, Koseki H, Brockdorff N (2004) Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev Cell 7:663–676

Debrand E, Chureau C, Arnaud D, Avner P, Heard E (1999) Functional analysis of the DXPas34 locus, a 3′ regulator of Xist expression. Mol Cell Biol 19(12):8513–8525

Dietzel S, Schiebel K, Little G, Edelmann P, Rappold GA, Eils R, Cremer C, Cremer T (1999) The 3D positioning of ANT2 and ANT3 genes within female X chromosome territories correlates with gene activity. Exp Cell Res 252:363–375

Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT (2007) Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. Mol Cell 25:43–56

Donohoe ME, Silva SS, Pinter SF, Xu N, Lee JT (2009) The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature 460:128–132

Duthie SM, Nesterova TB, Formstone EJ, Keohane AM, Turner BM, Zakian SM, Brockdorff N (1999) Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. Hum Mol Genet 8:195–204

Eils R, Dietzel S, Bertin E, Schröck E, Speicher MR, Ried T, Robert-Nicoud M, Cremer C, Cremer T (1996) Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. J Cell Biol 135:1427–1440

Fackelmayer FO (2005) A stable proteinaceous structure in the territory of inactive X chromosomes. J Biol Chem 280:1720–1723

Fang J, Chen T, Chadwick B, Li E, Zhang Y (2004) Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. J Biol Chem 279:52812–52815

Fedorova E, Zink D (2008) Nuclear architecture and gene regulation. Biochim Biophys Acta 1783:2174–2184

Fitzpatrick GV, Soloway PD, Higgins MJ (2002) Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet 32:426–431

Gil J, Bernard D, Martinez D, Beach D (2004) Polycomb CBX7 has a unifying role in cellular lifespan. Nat Cell Biol 6:67–72

Gilbert SL, Pehrson JR, Sharp PA (2000) XIST RNA associates with specific regions of the inactive X chromatin. J Biol Chem 275:36491–36494

Graves JA (2006) Sex chromosome specialization and degeneration in mammals. Cell 124:901–914

Gribnau J, Luikenhuis S, Hochedlinger K, Monkhorst K, Jaenisch R (2005) X chromosome choice occurs independently of asynchronous replication timing. J Cell Biol 168:365–373

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464:1071–1076

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458:223–227

Hall LL, Lawrence JB (2003) The cell biology of a novel chromosomal RNA: chromosome painting by XIST/Xist RNA initiates a remodeling cascade. Semin Cell Dev Biol 14:369–378

Hansen RS (2003) X inactivation-specific methylation of LINE-1 elements by DNMT3B: implications for the Lyon repeat hypothesis. Hum Mol Genet 12:2559–2567

Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL (2001) Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell 107:727–738

Helbig R, Fackelmayer FO (2003) Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. Chromosoma 112:173–182

Hernández-Muñoz I, Lund AH, van der Stoop P, Boutsma E, Muijrers I, Verhoeven E, Nusinow DA, Panning B, Marahrens Y, van Lohuizen M (2005) Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. Proc Natl Acad Sci USA 102:7635–7640

Herzing LB, Romer JT, Horn JM, Ashworth A (1997) Xist has properties of the X-chromosome inactivation centre. Nature 386:272–275

Jeppesen P, Turner BM (1993) The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell 74:281–289

Jonkers I, Monkhorst K, Rentmeester E, Grootegoed JA, Grosveld F, Gribnau J (2008) Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol Cell Biol 28:5583–5594

Jonkers I, Barakat TS, Achame EM, Monkhorst K, Kenter A, Rentmeester E, Grosveld F, Grootegoed JA, Gribnau J (2009) RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. Cell 139:999–1011

Kanellopoulou C, Muljo SA, Dimitrov SD, Chen X, Colin C, Plath K, Livingston DM (2009) X chromosome inactivation in the absence of Dicer. Proc Natl Acad Sci USA 106:1122–1127

Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW, Pombo A, Fisher AG, Young RA, Jenner RG (2010) Short RNAs are transcribed from repressed Polycomb target genes and interact with Polycomb repressive complex-2. Mol Cell 38:675–688

Ke X, Collins A (2003) CpG islands in human X-inactivation. Ann Hum Genet 67:242–249

Kennedy BK, Barbie DA, Classon M, Dyson N, Harlow E (2000) Nuclear organization of DNA replication in primary mammalian cells. Genes Dev 14:2855–2868

Keohane AM, O'Neill LP, Belyaev ND, Lavender JS, Turner BM (1996) X-Inactivation and histone H4 acetylation in embryonic stem cells. Dev Biol 180:618–630

Keohane AM, Barlow AL, Waters J, Bourn D, Turner BM (1999) H4 acetylation, XIST RNA and replication timing are coincident and define x;autosome boundaries in two abnormal X chromosomes. Hum Mol Genet 8:377–383

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci USA 106:11667–11672

Lee JT (2002) Homozygous Tsix mutant mice reveal a sex-ratio distortion and revert to random X-inactivation. Nat Genet 32:195–200

Lee JT, Jaenisch R (1997) Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature 386:275–279

Lee JT, Strauss WM, Dausman JA, Jaenisch R (1996) A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell 86:83–94

Lee JT, Davidow LS, Warshawsky D (1999) Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet 21:400–404

Leeb M, Wutz A (2007) Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. J Cell Biol 178:219–229

Levy MA, Fernandes AD, Tremblay DC, Seah C, Berube NG (2008) The SWI/SNF protein ATRX co-regulates pseudoautosomal genes that have translocated to autosomes in the mouse genome. BMC Genomics 9:468

Lin H, Gupta V, VerMilyea MD, Falciani F, Lee JT, O'Neill LP, Turner BM (2007) Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes. PLoS Biol 5:e326

Luikenhuis S, Wutz A, Jaenisch R (2001) Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells. Mol Cell Biol 21:8512–8520

Lyle R, Watanabe D, te Vruchte D, Lerchner W, Smrzka OW, Wutz A, Schageman J, Hahner L, Davies C, Barlow DP (2000) The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. Nat Genet 25:19–21

Lyon MF (1961) Gene action in the X-chromosome of the mouse (*Mus musculus* L.). Nature 190:372–373

Lyon MF (1998) X-chromosome inactivation: a repeat hypothesis. Cytogenet Cell Genet 80:133–137

Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, Dubois A, Sanglier-Cianférani S, Van Dorsselaer A, Clerc P, Avner P, Visvikis A, Branlant C (2010) 2-D structure of the A region of Xist RNA and its implication for PRC2 association. PLoS Biol 8:e1000276

Mak W, Nesterova TB, de Napoles M, Appanah R, Yamanaka S, Otte AP, Brockdorff N (2004) Reactivation of the paternal X chromosome in early mouse embryos. Science 303:666–669

Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM (2006) Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev 20:1268–1282

Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R (1997) Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev 11:156–166

Marahrens Y, Loring J, Jaenisch R (1998) Role of the Xist gene in X chromosome choosing. Cell 92:657–664

Marks H, Chow JC, Denissov S, Françoijs KJ, Brockdorff N, Heard E, Stunnenberg HG (2009) High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res 19:1361–1373

Mermoud JE, Costanzi C, Pehrson JR, Brockdorff N (1999) Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J Cell Biol 147:1399–1408

Mermoud JE, Popova B, Peters AH, Jenuwein T, Brockdorff N (2002) Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation. Curr Biol 12:247–251

Mohandas T, Sparkes RS, Shapiro LJ (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science 211:393–396

Monkhorst K, Jonkers I, Rentmeester E, Grosveld F, Gribnau J (2008) X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. Cell 132:410–421

Monkhorst K, de Hoon B, Jonkers I, Mulugeta Achame E, Monkhorst W, Hoogerbrugge J, Rentmeester E, Westerhoff HV, Grosveld F, Grootegoed JA, Gribnau J (2009) The probability to initiate X chromosome inactivation is determined by the X to autosomal ratio and X chromosome specific allelic properties. PLoS ONE 4:e5616

Morey C, Navarro P, Debrand E, Avner P, Rougeulle C, Clerc P (2004) The region 3′ to Xist mediates X chromosome counting and H3 Lys-4 dimethylation within the Xist gene. EMBO J 23:594–604

Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 322:1717–1720

Navarro P, Page DR, Avner P, Rougeulle C (2006) Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. Genes Dev 20:2787–2792

Navarro P, Chambers I, Karwacki-Neisius V, Chureau C, Morey C, Rougeulle C, Avner P (2008) Molecular coupling of Xist regulation and pluripotency. Science 321:1693–1695

Nesterova TB, Slobodyanyuk SY, Elisaphenko EA, Shevchenko AI, Johnston C, Pavlova ME, Rogozin IB, Kolesnikov NN, Brockdorff N, Zakian SM (2001) Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res 11:833–849

Nesterova TB, Johnston CM, Appanah R, Newall AE, Godwin J, Alexiou M, Brockdorff N (2003) Skewing X chromosome choice by modulating sense transcription across the Xist locus. Genes Dev 17:2177–2190

Nesterova TB, Popova BC, Cobb BS, Norton S, Senner CE, Tang YA, Spruce T, Rodriguez TA, Sado T, Merkenschlager M, Brockdorff N (2008) Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. Epigenetics Chromatin 1:2

Newall AE, Duthie S, Formstone E, Nesterova T, Alexiou M, Johnston C, Caparros ML, Brockdorff N (2001) Primary non-random X inactivation associated with disruption of Xist promoter regulation. Hum Mol Genet 10:581–589

Nguyen DK, Disteche CM (2006) Dosage compensation of the active X chromosome in mammals. Nat Genet 38:47–53

Norris DP, Brockdorff N, Rastan S (1991) Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. Mamm Genome 1:78–83

Nusinow DA, Sharp JA, Morris A, Salas S, Plath K, Panning B (2007) The histone domain of macroH2A1 contains several dispersed elements that are each sufficient to direct enrichment on the inactive X chromosome. J Mol Biol 371:11–18

Ogawa Y, Lee JT (2003) Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. Mol Cell 11:731–743

Ogawa Y, Sun BK, Lee JT (2008) Intersection of the RNA interference and X-inactivation pathways. Science 320:1336–1341

Ohhata T, Hoki Y, Sasaki H, Sado T (2008) Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. Development 135:227–235

Okamoto I, Heard E (2006) The dynamics of imprinted X inactivation during preimplantation development in mice. Cytogenet Genome Res 113:318–324

Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E (2004) Epigenetic dynamics of imprinted X inactivation during early mouse development. Science 303:644–649

O'Keefe RT, Henderson SC, Spector DL (1992) Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. J Cell Biol 116:1095–1110

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32:232–246

Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N (1996) Requirement for Xist in X chromosome inactivation. Nature 379:131–137

Peters AH, Mermoud JE, O'Carroll D, Pagani M, Schweizer D, Brockdorff N, Jenuwein T (2002) Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nat Genet 30:77–80

Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B (2002) Xist RNA and the mechanism of X chromosome inactivation. Annu Rev Genet 36:233–278

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science 300:131–135

Plath K, Talbot D, Hamer KM, Otte AP, Yang TP, Jaenisch R, Panning B (2004) Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. J Cell Biol 167:1025–1035

Popova BC, Tada T, Takagi N, Brockdorff N, Nesterova TB (2006) Attenuated spread of X-inactivation in an X;autosome translocation. Proc Natl Acad Sci USA 103:7706–7711

Priest JH, Heady JE, Priest RE (1967) Delayed onset of replication of human X chromosomes. J Cell Biol 35:483–487

Prissette M, El-Maarri O, Arnaud D, Walter J, Avner P (2001) Methylation profiles of DXPas34 during the onset of X-inactivation. Hum Mol Genet 10:31–38

Pullirsch D, Härtel R, Kishimoto H, Leeb M, Steiner G, Wutz A (2010) The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. Development 137:935–943

Rasmussen TP, Wutz AP, Pehrson JR, Jaenisch RR (2001) Expression of Xist RNA is sufficient to initiate macrochromatin body formation. Chromosoma 110:411–420

Rastan S (1982) Timing of X-chromosome inactivation in postimplantation mouse embryos. J Embryol Exp Morphol 71:11–24

Redrup L, Branco MR, Perdeaux ER, Krueger C, Lewis A, Santos F, Nagano T, Cobb BS, Fraser P, Reik W (2009) The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. Development 136:525–530

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129:1311–1323

Ross MT et al (2005) The DNA sequence of the human X chromosome. Nature 434:325–337

Rougeulle C, Chaumeil J, Sarma K, Allis CD, Reinberg D, Avner P, Heard E (2004) Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. Mol Cell Biol 24:5475–5484

Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E (2000) X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev Biol 225:294–303

Sado T, Wang Z, Sasaki H, Li E (2001) Regulation of imprinted X-chromosome inactivation in mice by Tsix. Development 128:1275–1286

Sado T, Hoki Y, Sasaki H (2005) Tsix silences Xist through modification of chromatin structure. Dev Cell 9:159–165

Schoeftner S, Sengupta AK, Kubicek S, Mechtler K, Spahn L, Koseki H, Jenuwein T, Wutz A (2006) Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J 25:3110–3122

Shearwin KE, Callen BP, Egan JB (2005) Transcriptional interference – a crash course. Trends Genet 21:339–345

Shibata S, Lee JT (2003) Characterization and quantitation of differential Tsix transcripts: implications for Tsix function. Hum Mol Genet 12:125–136

Shibata S, Lee JT (2004) Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice. Curr Biol 14:1747–1754

Shibata S, Yokota T, Wutz A (2008) Synergy of Eed and Tsix in the repression of Xist gene and X-chromosome inactivation. EMBO J 27:1816–1826

Shumaker DK, Dechat T, Kohlmaier A, Adam SA, Bozovsky MR, Erdos MR, Eriksson M, Goldman AE, Khuon S, Collins FS, Jenuwein T, Goldman RD (2006) Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. Proc Natl Acad Sci USA 103:8703–8708

Silva J, Mak W, Zvetkova I, Appanah R, Nesterova TB, Webster Z, Peters AHFM, Jenuwein T, Otte AP, Brockdorff N (2003) Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell 4:481–495

Smith KP, Byron M, Clemson CM, Lawrence JB (2004) Ubiquitinated proteins including uH2A on the human and mouse inactive X chromosome: enrichment in gene rich bands. Chromosoma 113:324–335

Starmer J, Magnuson T (2009) A new model for random X chromosome inactivation. Development 136:1–10

Stavropoulos N, Rowntree RK, Lee JT (2005) Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation. Mol Cell Biol 25:2757–2769

Sun BK, Deaton AM, Lee JT (2006) A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol Cell 21:617–628

Takagi N, Sasaki M (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature 256:640–642

Tsai CL, Rowntree RK, Cohen DE, Lee JT (2008) Higher order chromatin structure at the X-inactivation center via looping DNA. Dev Biol 319:416–425

Vigneau S, Augui S, Navarro P, Avner P, Clerc P (2006) An essential role for the DXPas34 tandem repeat and Tsix transcription in the counting process of X chromosome inactivation. Proc Natl Acad Sci USA 103(19):7390–7395. Epub 2006 Apr 28

Wang J, Mager J, Chen Y, Schneider E, Jc C, Nagy A, Magnuson T (2001) Imprinted X inactivation maintained by a mouse Polycomb group gene. Nat Genet 28:371–375

Wang Z, Willard HF, Mukherjee S, Furey TS (2006) Evidence of influence of genomic DNA sequence on human X chromosome inactivation. PLoS Comput Biol 2:e113

West JD, Frels WI, Chapman VM, Papaioannou VE (1977) Preferential expression of the maternally derived X chromosome in the mouse yolk sac. Cell 12:873–882

Wutz A, Gribnau J (2007) X inactivation Xplained. Curr Opin Genet Dev 17:387–393

Wutz A, Jaenisch R (2000) A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell 5:695–705

Wutz A, Rasmussen TP, Jaenisch R (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet 30:167–174

Zhang LF, Huynh KD, Lee JT (2007) Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell 129:693–706

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322:750–756

Chapter 4
TERRA: Long Noncoding RNA at Eukaryotic Telomeres

Rajika Arora, Catherine M. C. Brun, and Claus M. Azzalin

Abstract Telomeres protect the ends of linear eukaryotic chromosomes from being recognized as DNA double-stranded breaks, thereby maintaining the stability of our genome. The highly heterochromatic nature of telomeres had, for a long time, reinforced the idea that telomeres were transcriptionally silent. Since a few years, however, we know that DNA-dependent RNA polymerase II transcribes telomeric DNA into *TE*lomeric *R*epeat-containing *RNA* (TERRA) molecules in a large variety of eukaryotes. In this chapter, we summarize the current knowledge of telomere structure and function and extensively review data accumulated on TERRA biogenesis and regulation. We also discuss putative functions of TERRA in preserving telomere stability and propose future directions for research encompassing this novel and exciting aspect of telomere biology.

4.1 Eukaryotic Telomeres

In 1938, almost 20 years before James D. Watson and Francis Crick described the double helix structure of DNA, Hermann J. Muller made the seminal discovery that the ends of linear eukaryotic chromosomes behaved differently from the remainder of the chromosomes. By exposing flies to ionizing radiations, he obtained mutants carrying diverse chromosomal aberrations, such as inversions, deletions, and translocations, encompassing different genomic regions but sparing chromosome termini. He surmised that, although chromosomes appeared as homogenous cytological entities, their extremities, which he named telomeres (from the Greek nouns telos "end" and meros "part"), might exert the unique function of "sealing" (Muller 1938).

R. Arora, C.M.C. Brun, and C.M. Azzalin (✉)
Institute of Biochemistry, ETHZ-Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland
e-mail: claus.azzalin@bc.biol.ethz.ch

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3_4,

Further evidence for a protective role of telomeres was obtained in the early forties when Barbara McClintock observed that forced breakage within chromosomes resulted in chromosomal fusions, while intact chromosome ends failed to fuse. She also demonstrated that cells harboring fused chromosomes accumulated "anaphase bridges" during mitosis and that breakage of these DNA bridges generated new protective ("healed") chromosome termini, inducing a complete cessation of the breakage-fusion-bridge cycle (McClintock 1941).

Starting in the late seventies, Elizabeth H. Blackburn isolated and defined the DNA sequence of *Tetrahymena* telomeres (Yao et al. 1979, 1981). Together with Jack W. Szostak, she also experimentally confirmed the protective role of telomeres, as previously hypothesized by Muller and McClintock, by showing that telomeric repeats from *Tetrahymena* stabilized linear DNA transformed into yeast cells (Szostak and Blackburn 1982; Shampay et al. 1984). Soon after, Blackburn and Carol W. Greider identified telomerase, a specialized enzyme capable of extending single-stranded (ss) telomeric DNA molecules by addition of newly synthesized telomeric repeats, thus providing further clues on how telomeres are maintained (Greider and Blackburn 1985). In recognition of their work in telomere biology, Blackburn, Szostak, and Greider were awarded the Nobel Prize in Physiology and Medicine in 2009.

Although we have come a long way since their initial description, telomeres continue to intrigue scientists. Recent research in this field has essentially focused on characterizing details of telomere structure and function. A new twist occurred in 2007, when the longstanding dogma that telomeres are transcriptionally silent was overturned by the discovery that noncoding (nc) RNA molecules, named *TE*lomeric *R*epeat-containing *RNA* (TERRA), were found to emanate from and associate with telomeres (Azzalin et al. 2007; Schoeftner and Blasco 2008). In this chapter, we provide an overview of telomere structure, function, and biology and extensively review the current knowledge about TERRA biogenesis, regulation, and potential functions (Table 4.1).

4.1.1 Saccharomyces cerevisiae Telomeres

Structurally, telomeres are made of ribonucleoprotein complexes containing both DNA and protein components. In budding yeast, the DNA component consists of tandem arrays of short 5′ to 3′ G-rich repeats with the consensus sequence $TG_{1–3}$ or $TG_{2–3}(TG)_{1–6}$, varying in size between ~250 and 400 base pairs (bp) (Wang and Zakian 1990; Vega et al. 2003; Teixeira and Gilson 2005). The G-rich strand is also referred to as the G-strand and the complementary one as the C-strand. The double-stranded (ds) repeat sequence terminates with a single-strand G-rich 3′ overhang, the G-overhang, which ranges in length between 12 and 14 nucleotides (nt) during most of the cell cycle, while it reaches 30 nt or more during S-phase. Upon completion of telomere replication, the G-tail is thought to be shortened by end resection activities (Wellinger et al. 1993; Larrivee et al. 2004).

Table 4.1 TERRA regulators

Factor (species)	Function	Experimental approach	Effect on TERRA	Reference
TRF2 (Hs)	Inhibits NHEJ and ATM activation at telomeres	siRNA-mediated knockdown	p53-dependent increase in TERRA steady-state levels	Caslini et al. (2009)
		Overexpression of a dominant negative allele (TRF2ΔB)	Decrease in TERRA telomeric foci	Deng et al. (2009)
TRF2 (Mm)	Inhibits NHEJ and ATM activation at telomeres	Overexpression	Decrease in TERRA telomeric foci	Benetti et al. (2008)
TRF1 (Hs)	Facilitates DNA replication at telomeres	siRNA-mediated knockdown	Decrease in TERRA telomeric foci	Schoeftner and Blasco (2008)
UPF1 (Hs)	Effector of NMD pathway; maintenance of genome stability	shRNA-mediated knockdown	Increase in telomere-associated TERRA without apparent effect on TERRA half-life	Azzalin et al. (2007)
hEST1A/SMG6 (Hs)	Effector of NMD pathway; maintenance of telomere integrity	shRNA-mediated knockdown	Increase in telomere-associated TERRA without apparent effect on TERRA half-life	Azzalin et al. (2007)
SMG1 (Hs)	Effector of NMD pathway; maintenance of genome stability	shRNA-mediated knockdown	Increase in telomere-associated TERRA without apparent effect on TERRA half-life	Azzalin et al. (2007)
Suv3-9h and Suv4-20h (Mm)	Histone metilation	Gene deletion	Increase in TERRA steady-state levels	Schoeftner and Blasco (2008)
Terc (Mm)	Telomerase-mediated telomeric repeat synthesis	Gene deletion	Decrease in TERRA steady-state levels	Schoeftner and Blasco (2008)

(*continued*)

Table 4.1 (continued)

Factor (species)	Function	Experimental approach	Effect on TERRA	Reference
Rb proteins (Mm)	Tumor suppressors	Gene deletion	Increase in TERRA steady-state levels	Gonzalez-Suarez et al. (2009)
Lamins (Mm)	Nuclear organization and regulation of gene expression	Gene deletion	Decrease in TERRA steady-state levels	Gonzalez-Suarez et al. (2009)
Atrx (Mm)	Chromatin remodeling	Gene deletion	Increase in TERRA steady-state levels	Goldberg et al. (2010)
Dnmt1 and Dnmt3b (Hs)	Methylation of cytosines within CpG dinucleotides	Gene deletion	Increase in TERRA steady-state levels	Nergadze et al. (2009)
Dnmt1 or Dnmt3a/b (Mm)	Methylation of cytosines within CpG dinucleotides	Gene deletion	Decrease in TERRA steady-state levels	Schoeftner and Blasco (2008)
Dicer (Mm)	Effector of the RNAi pathway	Gene deletion and complementation	Decrease in TERRA steady-state levels	Schoeftner and Blasco (2008)
Dicer (Mm)	Effector of the RNAi pathway	Gene deletion and complementation	Increase in TERRA steady-state levels	Zhang et al. (2009)
MLL (Hs)	H3K4 methyltransferase	siRNA-mediated knockdown	Decrease in TERRA steady-state levels	Caslini et al. (2009)

The bulk of telomeric repeats are synthesized together with the rest of the genome by the conventional DNA replication machineries. Nevertheless, the full replication of linear DNA molecules, such as linear eukaryotic chromosomes, poses several challenges altogether referred to as "the end replication problem." To begin with, DNA polymerases are only capable of adding nucleotides in a 5′ to 3′ direction. They extend the 3′-hydroxyl end of a short RNA primer deposited by a specialized RNA polymerase, called primase, at specific genomic loci. This RNA primer is then degraded and replaced by DNA synthesized through extension of an upstream primer. However, during replication of the 3′ chromosomal terminus by lagging-strand synthesis, removal of the RNA primer leaves a gap behind the newly synthesized DNA, causing loss of the corresponding sequences. In addition, the 5′-end-containing parental strand is recessed and cannot provide a template for the synthesis of a 3′ overhang, implying that generation of the G-overhang on the leading strand telomere requires postreplicative nucleolytic events. Taken together, sequence loss from chromosome ends is hence expected to occur at every replication cycle both at the lagging and leading strand telomeres, ultimately leading to loss of entire stretches of telomeric DNA as well as to potential loss of genetic information (Watson 1972; Olovnikov 1973; Bianchi and Shore 2008). Several mechanisms have indeed evolved to counteract such sequence attrition, the most common, evolutionary conserved one being represented by telomerase. The core telomerase holoenzyme essentially consists of a reverse transcriptase catalytic subunit and an RNA moiety comprising a short region complementary to the G-rich telomeric repeats, which is used as template during reverse transcription (Bianchi and Shore 2008; Artandi and DePinho 2010).

Budding yeast core telomerase consists of the *E*ver *S*horter *T*elomeres 2 (*EST2*) gene product, which codes for the reverse transcriptase enzyme Est2p (Lundblad and Szostak 1989; Lundblad and Blackburn 1990; Lingner et al. 1997) and an RNA template encoded by the *Tel*omerase *C*omponent 1 (*TLC1*) gene (Singer and Gottschling 1994). Other telomerase subunits include Est1p and Est3p (Hughes et al. 2000; Seto et al. 2002). Yeast telomerase-deficient deletion mutant strains (such as *tlc1Δ, est2Δ, est1Δ, and est3Δ*) are initially viable but progressively lose telomeric DNA, ultimately entering cellular senescence (Vega et al. 2003). A plethora of other proteins [for example, the Ku70p–Ku80p heterodimer (Peterson et al. 2001) and the Sm protein complex (Seto et al. 1999)] have also been found to associate with telomerase and modulate its activity in vivo.

*R*epressor/*a*ctivator site-binding *p*rotein (Rap1p) binds directly to ds telomeric repeats, regulates telomere length homeostasis, and prevents telomere fusions (Fisher and Zakian 2005). An increase in average telomere length is observed in cells that harbor a truncation in the C-terminus of Rap1p (Kyrion et al. 1992). *R*ap1-*i*nteracting *f*actors Rif1p and Rif2p are also involved in telomere length maintenance, and mutations in these genes result in moderate telomere elongation. Because telomere length is directly proportional to the number of DNA-bound Rap1p molecules, it has been speculated that Rap1p regulates telomere length by establishing a counting mechanism (Marcand et al. 1997).

The budding yeast G-overhang is protected by the binding of *C*ell *d*ivision *c*ontrol protein 13 (Cdc13p); its functional impairment causes degradation of the telomeric C-strand and cell-cycle arrest (Garvik et al. 1995; Lin and Zakian 1996; Bourns et al. 1998). In addition, Cdc13p forms a complex with two other ss DNA-binding proteins: *S*uppressor of Cdc *t*hirtee*n* (Stn1p) and *Te*lomeric pathways with St*n*1 (Ten1p) (Gao et al. 2007). This multiprotein complex (referred to as the CST complex) maintains yeast telomere length and integrity by regulating telomerase activity at chromosome ends. In particular, Cdc13p interacts with Est1p and promotes recruitment of telomerase to telomeres (Nugent et al. 1996). Stn1, on the other hand, inhibits the Cdc13p-mediated recruitment of telomerase (Grandin et al. 2000). All three proteins have also been implicated in facilitating the recruitment of DNA polymerase α, which mediates synthesis of the C-strand (Qi and Zakian 2000; Chandra et al. 2001; Petreaca et al. 2006)

4.1.2 Schizosaccharomyces pombe Telomeres

Telomeres in *S. pombe* consist of ~300 bp long tandemly repeated GGTTA-$CA_{0-1}C_{0-1}G_{0-6}$ sequences, with the GGTTAC repetitive unit being the most commonly found (Sugawara 1989). The ≥30 nt long G-overhang of wild type fission yeast can only be detected during S-phase by native gel analysis (Tomita et al. 2003). The catalytic subunit of telomerase is encoded by the gene $trt1^+$ (Nakamura et al. 1997) and the RNA subunit by $ter1^+$ (Leonardi et al. 2008; Webb and Zakian 2008).

The protein Taz1p binds to ds telomeric repeats and shares homology with the Myb domain of human *T*elomere *R*epeat binding *F*actors 1 and 2 (TRF1 and TRF2; see Sect. 4.1.5) (Cooper et al. 1997). *taz1* gene deletion results in a dramatic telomerase-mediated elongation of telomeres and replication fork stalling within the telomeric tract, indicating that Taz1p facilitates telomere replication (Cooper et al. 1997; Miller et al. 2006). In addition, Taz1p also prevents unwanted DNA repair activities, including *n*on*h*omologous *e*nd *j*oining (NHEJ) and *h*omology-*d*irected *r*epair (HDR), from acting at telomeres (Ferreira and Cooper 2001). Rap1p and Rif1p are interacting partners of Taz1p, and telomeric fusions are observed in *rap1Δ* but not *rif1Δ* mutants (Miller et al. 2005). Also, Rap1p, but not Rif1p, is required to regulate 3′ overhang formation in conjunction with Taz1p (Miller et al. 2005). In addition, the ss DNA-binding protein *R*eplication *P*rotein *A* (RPA) appears to play a role in telomere maintenance by acting synergistically with Taz1p (Kibe et al. 2007).

*P*rotection *o*f *t*elomeres (SpPot1p) binds to the ss G-overhang via its *O*ligonucleotide/oligosaccharide *B*inding (OB)-fold domain and interacts with Tpz1p (TPP1 homolog in *Schizosaccharomyces pombe*) to prevent telomeric fusions (Miyoshi et al. 2008). Tpz1p also interacts with Poz1p (Pot1-associated in *Schizosaccharomyces pombe*) and *C*oiled-*c*oil protein *q*uantitatively enriched (Ccq1p; Miyoshi et al. 2008). Poz1p bridges the Pot1p–Tpz1p complex to the Taz1p–Rap1p

complex (Miyoshi et al. 2008). It has been speculated that Pot1p exists in two alternative complex conformations at telomeres with different lengths. The increased association of Pot1p with ds DNA of longer telomeres, mediated by Poz1p's interactions with Taz1p and Rap1p, supposedly restrains telomerase action (Miyoshi et al. 2008). When telomeres shorten, Pot1p dissociates from the Taz1p/Rap1p complexes and thus facilitates telomerase activity (Miyoshi et al. 2008). Ccq1p is recruited to telomeres by Taz1p, and *ccq1* deletion results in telomere shortening as well as subtelomeric rearrangements (Tomita and Cooper 2008).

4.1.3 Plant Telomeres

In most plants, the telomeric DNA consists of TTTAGGG tandem repeats. In *Arabidopsis thaliana*, telomere length usually varies between 2 and 9 kilobases (kb), with size heterogeneity observed among different telomeres within the same cell as well as among different cell types (Richards and Ausubel 1988). Longer telomeres, up to 150 kb in length, are observed in tobacco plants (Fajkus et al. 1995). The G-overhang in *Arabidopsis* varies in length between ~20 and 30 nt and is only found on half of the telomeres in seedlings and less than 35% of the telomeres in leaves (Riha et al. 2000).

Plant telomerase expression correlates with cellular proliferation capacity (Riha and Shippen 2003). High levels of telomerase expression are observed in undifferentiated cells of meristematic tissues, root tips, flowers, calli, and embryos, while little to no activity is detected in differentiated tissues (Fitzgerald et al. 1996; Killan et al. 1998; Riha and Shippen 2003). Typically, very short telomeres are lengthened by telomerase until they regain a favorable size range, while longer telomeres are shortened at a rate of 200–500 bp per generation due to the end replication problem (Shakirov and Shippen 2004). Although they accumulate widespread cytogenetic damage upon ablation of telomerase activity, *Arabidopsis* mutant strains are able to survive for up to ten generations, after which they arrest in a vegetative state (Riha et al. 2001). While the catalytic subunit of telomerase has been isolated from many plant species, the identity of the RNA component remains a mystery (Shippen 2006). The protein Ku80 also regulates telomere length since *ku80* mutant *Arabidopsis* strains undergo progressive telomere lengthening, possibly due to enhanced access of the free telomeric 3′-end to telomerase (Gallego et al. 2003; Zellinger and Riha 2007).

Only few telomeric proteins have been isolated in plants based on their affinity for ss G-rich or ds telomeric DNA repeats or based on their sequence homology to mammalian and yeast telomeric polypeptides. Interestingly, different telomere-binding proteins also bind to the telomeric-like sequence AAACCCTAA found in the promoter region of a number of plant genes (Shippen 2006). The protein *Nicotiana tabacum G*-strand-specific single-stranded *T*elomere-*B*inding *P*rotein (NtGTBP1) binds to the telomeric G-overhang and is similar to the proteins *T*hirteen *c*omplementation *g*ene 1 (Tcg1p) and *G*-strand *b*inding *p*rotein 2 (Gbp2p), which have been

shown to rescue the G2/M cell-cycle arrest induced upon Cdc13 impairment in *S. cerevisiae* (Zellinger and Riha 2007). NgTRF1, which shares homology with the Myb domain of mammalian TRF1 and TRF2 (see Sect. 4.2.5) as well as of Taz1p, binds ds telomeric repeats in vitro and is a negative regulator of telomere length (Yang et al. 2004). The *A. thaliana* protein *S*ingle-stranded *TE*lomere-binding *P*rotein 1 (STEP1) binds ss telomeric repeats and inhibits telomerase activity in vitro (Kwon and Chung 2004). A homology search for the OB-fold domain of Pot1 retrieved two proteins in *Arabidopsis*: AtPot1 and AtPot2 (Shakirov et al. 2005). AtPot1 physically associates with telomerase and positively regulates telomere length (Surovtseva et al. 2007).

4.1.4 *Drosophila melanogaster Telomeres*

Drosophila telomeric sequences consist of a mixed array of variably 5′ truncated retrotransposons. Three telomeric retrotransposons have been identified in flies: *He*T-A, *T*ART, and *T*AHRE (HTT). Telomeres comprise multiple copies of HTTs at their terminal ends, while their most proximal parts consist of complex subterminal repeat arrays, termed *T*elomere *A*ssociated *S*equence (TAS). Like in other eukaryotes, the telomeric terminus is capped by a multiprotein complex, although capping of *Drosophila* telomeres does not require sequence-specific binding, as demonstrated by the fact that chromosomes devoid of retrotransposons can also be correctly capped (Rong 2008b). *H*eterochromatin *P*rotein 1 (HP1) binds to telomeres either by direct interaction with histone H3 dimethylated at lysine 9 (H3K9me2) via its chromodomain or directly with telomeric DNA via its hinge domain (Vermaak and Malik 2009). In the absence of HP1, *Drosophila* telomeres undergo fusion events (Fanti et al. 1998). Other proteins found to be deposited at *Drosophila* telomeres include the *Ub*iquitin *c*onjugating enzyme (UbcD1) (Cenci et al. 1997), a putative transcription factor named *W*ith*O*ut *C*hildren (WOC) (Raffa et al. 2005), the H2A.Z histone variant (Rong 2008a) and HipHop (Gao et al. 2010).

Drosophila cells do not possess telomerase activity. Instead, telomeres are maintained by two telomerase-independent pathways: gene conversion and retrotransposition. In the gene conversion pathway, the 3′ end of one chromosome terminus invades another chromosome, and the sequence of the invaded strand is used as a template to extend the invading 3′ end (Mikhailovsky et al. 1999; Kahn et al. 2000). Second strand synthesis followed by ligation results in the extension of the invading chromosome end (Mason et al. 2008). The retrotransposition pathway relies on reverse transcriptase and Gag proteins encoded by HTTs. Mature Gag proteins bind to cytoplasmic RNA molecules emanating from the retrotransposons and shuttle them back into the nucleus in close proximity to telomeres. The reverse transcriptase then carries out first-strand synthesis by extending the 3′ OH of the chromosome terminus using the RNA moiety as a template (Mason et al. 2008).

In addition to larger HTT transcripts, small RNAs from HTT loci, ranging from 26 to 30 nt in length, have also been identified in the germline (Saito et al. 2006) and

during embryogenesis (Aravin et al. 2003). These RNA species, termed *r*epeat *a*ssociated *s*mall *i*nterfering RNAs (rasiRNAs), are bound by the *P*-element *i*nduced *wi*mpy testis (PIWI) protein family members PIWI, *AGO*naute3 (AGO3), and *AUB*ergine (AUB) and have been implicated in a feedback loop regulating HTT transcript levels (Brennecke et al. 2007). A recent study revealed that the protein encoded by the gene *PRO*liferation *D*isrupter (PROD) regulates the cellular levels of transcripts originating from HeT-A elements (Torok et al. 2007). Also, hinge domain-mediated binding of HP1 to telomeric DNA increases transcription from Het-A and TART retrotransposons (Perrini et al. 2004). On the other hand, flies heterozygous for HP1 mutant alleles defective in H3K9me2 binding exhibit elongated telomeres and increased transcript levels both from TART and HeT-A (Perrini et al. 2004).

4.1.5 Mammalian Telomeres

As an example for mammalian telomeres, a sketch illustrating the core molecular components of a human telomere is depicted in Fig. 4.1. In mammalian cells, telomeres consist of ds (TTAGGG)n sequences, with length varying approximately from 5 up to 50 kb amongst different organisms, cell types, and different chromosome ends within the same cell. The G-overhang ranges in length between 50 and 500 nt. The telomerase reverse transcriptase TERT and the RNA moiety TR represent the minimal components required for telomerase activity in vitro (Bianchi and Shore 2008). Accessory telomerase-associated factors include the Est1p-like proteins hEST1A and hEST1B (Reichenbach et al. 2003; Snow et al. 2003), the

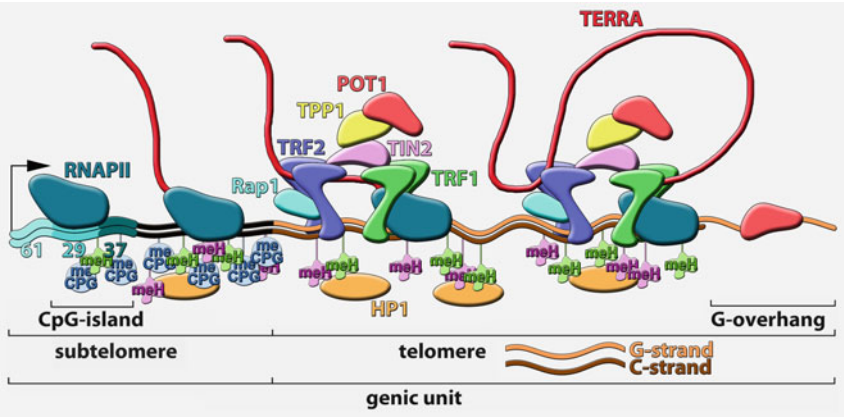

Fig. 4.1 Cartoon depicting the core molecular composition of human telomeres. The terminal chromosomal region comprising telomeric repeats and TERRA CpG island promoters (61-29-37 repeats) constitute a complete genic unit. Methylated histone variants are indicated by meH. The *black arrow* shows the direction of transcription from the TERRA promoter

ATPases reptin and pontin (Venteicher et al. 2008), dyskerin (Mitchell et al. 1999a, b), and *T*elomerase *CA*jal *B*ody protein 1 (TCAB1) (Venteicher et al. 2009).

In healthy human individuals, telomerase is normally expressed in proliferative tissues such as bone marrow, skin, gastrointestinal epithelium, testis, activated lymphocytes, and germ cells. On the contrary, in most human somatic cells, telomerase activity is virtually absent due to silencing of the hTERT promoter (Kim et al. 1994). This leads to progressive shortening of telomeres upon cell division (Harley et al. 1990), essentially due to the above-described end replication problem (see Sect. 4.1.1). Ultimately, telomeres reach a "critical length," which evokes a persistent DNA damage response, triggering p53- and pRb-dependent cellular senescence and/or death (Shay et al. 1991; Bianchi and Shore 2008). Telomeres are hence envisaged as cellular clocks that set the lifespan of normal somatic cells. Induction of cellular senescence appears to be important both in organismal aging and in counteracting cancerogenesis and cancer progression (Campisi and Yaswen 2009). Indeed, immortal cancer cells rely on reactivation of telomere-lengthening mechanisms to maintain their telomeres at a constant length indefinitely. While telomerase reactivation is the most common cancer-associated telomere lengthening mechanism (Shay and Bacchetti 1997), a limited number of human mesenchymal tumor cells maintain telomere length through the homologous recombination-based *A*lternative *L*engthening of *T*elomere (ALT) pathway (Bryan et al. 1997; Cesare and Reddel 2010).

A multiprotein complex named shelterin binds mammalian telomeres and is essential for both telomere protection and telomere length maintenance (Palm and de Lange 2008). Specificity of binding for telomeric sequences is conferred via the subunits TRF1 and TRF2 (Fig. 4.1), which directly interact with ds telomeric repeats. Both proteins recruit *T*RF1-*I*nteracting *N*uclear protein 2 (TIN2) (Kim et al. 1999) through direct protein–protein interactions. TRF2 also recruits hRap1, the human homolog of yeast Rap1p (Li et al. 2000). Finally, human POT1 is tethered to telomeres via TPP1, which, in turn, associates with telomeres through interaction with TIN2 (O'Connor et al. 2006). In addition, in vitro studies showed that POT1 binds specifically to ss G-rich telomeric DNA repeats (Lei et al. 2004), giving rise to the accepted idea that POT1 associates with telomeres also through direct interactions with the G-overhang (Fig. 4.1).

Because telomeres closely resemble DNA ends generated at sites of DNA ds breaks, a major question in the field of telomere biology is why telomeres are not perceived and processed as DNA damage and how major repair pathways, such as NHEJ and HDR, are prevented from acting at functional telomeres. One current model proposes that the G-overhang loops towards and invades the ds telomeric repeat tract, giving rise to a lasso-like structure known as the T-loop. T-loop formation is speculated to be a major mechanism by which the ss telomeric terminus is hidden from the DNA damage detection and repair machineries (Palm and de Lange 2008; Bianchi and Shore 2008; Riethman 2008).

In addition, TRF2, Rap1, POT1, and TPP1 have been identified as crucial players in protecting telomeres from unwanted DNA repair in ways that do not necessarily depend on T-loop formation. In vivo ablation of TRF2 leads to

accumulation of DNA damage factors such as p*53* *B*inding *P*rotein 1 (53BP1) and gamma (γ)-H2AX at telomeres, giving rise to the so-called *T*elomere dysfunction-*I*nduced *F*oci (TIF) (Takai et al. 2003). TIF formation is accompanied by cell death or occurrence of covalent fusions among different telomeres mediated by DNA ligase IV-dependent NHEJ (van Steensel et al. 1998; Smogorzewska et al. 2002). Loss of TRF2 also activates the *A*taxia *T*elangiectasia *M*utated (ATM) kinase-mediated DNA damage signaling pathway (Celli and de Lange 2005). This, together with the fact that TRF2 physically interacts with ATM (Karlseder et al. 2004), has fostered a model where TRF2 functions at telomeres by preventing activation of ATM (and downstream DNA damage detection and repair activities) locally. Therefore, in the absence of functional TRF2, telomeres are recognized as ds breaks. Because TRF2 recruits Rap1 to telomeres, it is possible that not only TRF2 but also Rap1 might prevent NHEJ at telomeres, an idea further supported by the observation that artificial tethering of hRap1 to TRF2-depleted telomeres in HeLa-S3 cell lines is sufficient to prevent end-to-end fusions (Sarthy et al. 2009). Nevertheless, recent analysis of mouse knock-out models showed that Rap1 deletion does not cause telomere-damage or telomeric fusions, disproving the aforementioned hypothesis at least for murine model systems (Sfeir et al. 2010).

POT1 also contributes to prevent activation of DNA damage machineries at telomeres, presumably by inhibiting the ATM-alternative *A*taxia *T*elangiectasia and *R*ad3-related (ATR) signaling pathway (Denchi and de Lange 2007). It is hypothesized that POT1 competes with RPA binding to telomeric ss DNA, therefore preventing RPA-mediated activation of the ATR signaling pathway (Denchi and de Lange 2007).

Some shelterin proteins have also been implicated in regulating telomere length and telomerase activity. Overexpression of TRF1 results in a continuous shortening of telomeres upon successive population doublings (van Steensel and de Lange 1997). Conversely, expression of a dominant negative mutant form of TRF1 induces telomere elongation (van Steensel and de Lange 1997). Changes in telomere length upon deregulation of TRF1 levels have been hypothesized to be due to its impact on telomere replication rather than on telomerase action. Consistently, TRF1 does not affect telomerase activity in vitro (Smogorzewska et al. 2000), and TRF1 dysfunction leads to severe replication fork stalling within telomeric tracts during replication (Sfeir et al. 2009). A crosstalk between TRF1 and the helicases *BL*oo*M* syndrome helicase (BLM) and *R*egulator of *T*elomere *EL*ongation helicase 1 (RTEL1) has been shown to be essential in order to assure processive replication of telomeric repeats (Sfeir et al. 2009).

Ectopic overexpression of POT1 in telomerase-positive human cancer cells results in an increase in average telomere length (Colgin et al. 2003). Such an increase is not observed when exogenous POT1 is expressed in telomerase-negative human cells, while lengthening of the shorter telomeres is observed upon concomitant ectopic expression of POT1 and hTERT in the same cells (Colgin et al. 2003). Expression of various truncated versions of POT1 also results in telomere elongation, possibly due to the dominant negative effects exerted on the recruitment of endogenous POT1 to telomeres (Loayza and De Lange 2003; Liu et al. 2004;

Kendellen et al. 2009). Finally, short hairpin (sh) RNA-mediated knockdown of POT1 or TPP1 gives rise to elongated telomeres (Ye et al. 2004). Together, these studies suggest that POT1 (and TPP1) are negative regulators of telomerase activity, although they might also promote telomerase function in some contexts, as suggested by the fact that TPP1 and POT1 increase the processivity of telomerase by slowing its dissociation rate as well as aiding in the translocation process in vitro (Latrick and Cech 2010).

Besides shelterin proteins, which are exclusively found at telomeres, telomeric chromatin is also enriched for heterochromatin marks such as histone H4 trimethylated at lysine 20 (H4K20m3) and histone H3 trimethylated at lysine 9 (H3K9m3) (see Fig. 4.1), the latter functioning as a platform for HP1 recruitment (Lachner et al. 2001; Garcia-Cao et al. 2004; Blasco 2007). The histone methyltransferases SUV39H1-H2 and SUV4-20H1-2 directly generate these histone modifications (Garcia-Cao et al. 2004; Schotta et al. 2004). Also, subtelomeric DNA is highly methylated at CpG dinucleotides by the concerted action of the *DN*A *Me*T*h*yltransferases DNMT1, DNMT3a, and DNMT3b (Gonzalo et al. 2006). Finally, both subtelomeric and telomeric regions display a low density of acetylated histones H3 and H4 that are generally enriched at euchromatic loci (Fraga et al. 2005).

4.2 TERRA: TElomeric Repeat-Containing RNA

4.2.1 TERRA Discovery and Biogenesis

Due to their repressive chromatin state and low gene density, chromosome ends were considered for a very long time to be transcriptionally silent genomic loci. This longstanding dogma was overturned by the discovery of TERRA, a nuclear localized RNA deriving from the active transcription of telomeric sequences (Azzalin et al. 2007; Schoeftner and Blasco 2008). To date, telomere transcription has been reported in humans, rodents, birds, budding yeast, and zebra fish, indicating an extensive evolutionary conservation of this cellular feature (Solovei et al. 1994; Azzalin et al. 2007; Luke et al. 2008; Schoeftner and Blasco 2008).

TERRA comprises heterogeneously long (100 to up to more than 9,000 bases in mammals; see Fig. 4.2) molecules that are transcribed using the telomeric C-strand as template, thus generating RNA species comprising G-rich RNA repeats (UUAGGG in mammals). Transcripts from the complementary strand are undetectable with standard hybridization-based techniques, suggesting that telomere transcription occurs only using the C-rich telomeric strand as a template or that RNAs derived from transcription of the G-rich strand are rapidly degraded (Azzalin et al. 2007; Schoeftner and Blasco 2008).

Experimental evidence demonstrates that DNA-dependent RNA polymerase II (RNAPII) plays a major role in TERRA biogenesis. Treatment of human and mouse cells with the specific RNAPII inhibitor α-amanitin leads to a substantial decrease

in total TERRA steady-state levels within a few hours (Schoeftner and Blasco 2008; Azzalin and Lingner 2008). However, because some TERRA molecules are still detectable even after prolonged α-amanitin treatments, one cannot exclude that RNA polymerases other than RNAPII could participate in telomere transcription. Indeed, mass spectrometric analysis of purified human telomeric chromatin identified subunits of all three RNA polymerases (RNAPI, II, and III) (Déjardin and Kingston 2009). Further strengthening the idea of a major role for RNAPII in telomere transcription is the observation that RNAPII associates with mammalian telomeres in vivo as well as with TRF1 (Schoeftner and Blasco 2008; Fig. 4.1). In addition, at least a fraction of TERRA is 3′-end polyadenylated (Schoeftner and Blasco 2008; Azzalin and Lingner 2008) as the majority of RNAPII products. The UUAGGG sequence present in mammalian TERRA molecules does not resemble canonical polyadenylation signals, thus rendering unclear which factors promote

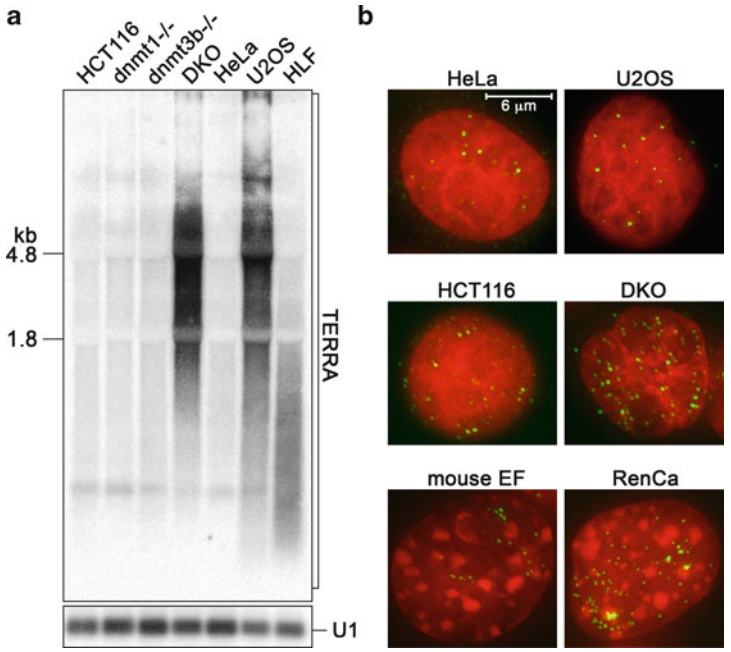

Fig. 4.2 Detection of TERRA in different mammalian cells. (**a**) Northern blot hybridization of nuclear RNA prepared from the indicated cell lines using radioactively labeled TERRA probes. HCT116: telomerase-positive human colon carcinoma cell line; Dnmt1−/−, Dnmt3b−/−, and DKO: HCT116-derived cell lines singly knocked-out for the indicated DNA methyltransferases or concomitantly knocked-out for both enzymes (double KO – DKO); *HeLa* telomerase-positive human cervical cancer cell line, *U2OS* ALT human osteosarcoma cell line, *HLF* human lung primary fibroblasts. Hybridization with small nuclear RNA U1 probes was used to demonstrate equal RNA loading. Molecular weights are *on the left*. (**b**) TERRA detection by RNA fluorescence in situ hybridization using fluorescently labeled TERRA probes. TERRA is shown in *green* while DAPI-stained DNA is shown in *red*. *Mouse EF* mouse ear primary fibroblasts, *RenCa* mouse renal carcinoma cell line

TERRA polyadenylation. However, in budding yeast, TERRA bears sequences that strongly resemble the canonical U-rich 3′-end processing signal that is known to be polyadenylated by *P*oly-*A* *p*olymerase 1 (Pap1) (Luke et al. 2008). Interestingly, *pap1* deletion leads to disappearance of TERRA molecules, indicating that the poly-A tail could stabilize TERRA (Luke et al. 2008).

Northern blot and RT-PCR experiments demonstrated that individual TERRA molecules contain both a telomeric and a subtelomeric RNA tract, indicating that TERRA transcription starts within subtelomeres (Azzalin et al. 2007). This hypothesis was confirmed by the discovery of subtelomeric promoter regions dedicated to the transcription of TERRA from several human chromosome ends. TERRA promoters are located ~250 bp away from the subtelomere-to-telomere transition and contain three repetitive DNA tracts: the most centromere-proximal tract comprises tandemly repeated 61 bp units, the middle tract comprises 29 bp tandem repeats, and the most distal tract comprises tandemly repeated 37 bp units (Fig. 4.1). These repetitive DNA elements have been referred to as "61-29-37 repeats" and are found immediately upstream of transcription start sites of several TERRA molecules (Nergadze et al. 2009). In addition, the 29 bp and 37 bp repeats display a high content in CpG dinucleotides, similar to the large majority of mammalian RNAPII-associated promoter regions. Indeed, total and phosphorylation-activated RNAPII was found to associate with TERRA promoter DNA in vivo (Nergadze et al. 2009).

BLAST and DNA fluorescence in situ hybridization (FISH) analyses localized 61-29-37 repeats at 20 different subtelomeres in human cells: 1p, 2p, 3q, 4p, 5p, 6p, 8p, 9p, 9q, 10q, 11p, 12p, 15q, 16p, 17p, 19p, 20p, 21q, Xq, and Yq (Nergadze et al. 2009). Among the remaining human subtelomeres, at least two (11q and Xp/Yp) are also transcribed (Azzalin et al. 2007), implying that different promoter types might contribute to the biogenesis of total human TERRA. However, it is also possible that ill-defined subtelomeric sequences available in the databases might have led to an underestimation of the actual number of human subtelomeres carrying 61-29-37 repeats.

4.2.2 TERRA Localization

RNA FISH analysis using fluorescently labeled telomeric probes revealed that TERRA forms discrete foci in the nucleus of mammalian cells during interphase (Azzalin et al. 2007; for some examples, see Fig. 4.2). RNA FISH combined with indirect immunofluorescence analysis using antibodies against telomeric proteins showed that most TERRA molecular foci colocalize with telomeres (Azzalin et al. 2007; Schoeftner and Blasco 2008). The number of TERRA foci varies among different tested cell lines, with 3–7 detectable TERRA foci in human cervical cancer cells and human primary lung fibroblasts and 20–40 foci in human osteosarcoma cells and murine renal cancer cells (Azzalin et al. 2007). Importantly, TERRA foci are also detected at the physical tips of chromosomes during mitosis, when transcription is paused, suggesting that at least a fraction of TERRA remains stably

bound to telomeres even in the absence of ongoing transcription (Azzalin et al. 2007). Thus, we infer that posttranscriptional mechanisms might have been established during evolution to retain TERRA at chromosome ends. It is important to note that not all telomeres colocalize with detectable TERRA, nor do all TERRA foci localize to telomeres (Azzalin et al. 2007). While the absence of detectable TERRA molecules at several telomeres might reflect different levels of transcriptional activity at individual telomeres, the nature of the TERRA-associated nontelomeric loci remains to be elucidated.

Some hints about this last issue derive from the observation that, in immortalized mouse embryonic fibroblasts, the most intense TERRA foci partially overlap with the *X I*nactive-*S*pecific *T*ranscript (XIST) RNA, which coats the inactive X chromosome (Schoeftner and Blasco 2008). In addition, TERRA is enriched on both mouse sex chromosomes in a developmentally specific manner. In male- and female-derived mouse embryonic stem cells, TERRA accumulates at both sex chromosomes. Upon cellular differentiation, TERRA undergoes a change in localization and associates only with the telomeres of the heterochromatic (inactive) sex chromosome of each sex (Ogawa et al. 2008). It will be important to determine whether such a phenomenon is peculiar only to mouse cells and whether TERRA molecules marking the sex chromosomes are transcribed from these same chromosomes or from other telomeres. Indeed, it remains unknown whether TERRA localizes to telomeric heterochromatin *in cis* or *in trans*. In the aforementioned study (Ogawa et al. 2008), TERRA foci were not detected at telomeres of autosomes, possibly due to less sensitive oligonucleotide probes as compared to those used to detect TERRA foci in other studies (Azzalin and Lingner 2008; Schoeftner and Blasco 2008).

Different sets of data are also starting to unravel potential pathways regulating TERRA localization. The human *S*uppressors with *M*orphogenetic defects in *G*enitalia (SMG) proteins *UP F*rameshift 1 (UPF1), hEST1A/SMG6, and SMG1 are best characterized as effectors of *N*onsense-*M*ediated mRNA *D*ecay (NMD), an evolutionary conserved, cytoplasmic RNA quality control mechanism, which recognizes and immediately degrades faulty mRNA molecules carrying premature termination codons (Nicholson et al. 2010). In addition, the three SMG proteins have also been independently implicated in different DNA metabolism pathways (including S-phase progression, DNA damage detection and/or repair, telomere capping, telomerase regulation, and apoptosis), which seem not to depend on their function in cytoplasmic RNA surveillance (Reichenbach et al. 2003; Snow et al. 2003; Brumbaugh et al. 2004; Azzalin and Lingner 2006; Redon et al. 2007; Oliveira et al. 2008). ShRNA-mediated downregulation of these three factors results in an increase in the number of telomere-associated TERRA foci, without affecting total TERRA steady-state levels or half-life, and in sudden loss of entire telomeric tracts (Azzalin et al. 2007). A direct role for these factors in negatively regulating TERRA localization to telomeres and in maintaining telomere integrity is substantiated by the fact that these polypeptides localize, although at low levels and probably in a transient manner, to telomeric heterochromatin in vivo (Azzalin and Lingner 2006). It remains to be determined whether these proteins perform their telomeric

functions in a complex, as it is the case for NMD, and whether the telomeric defects observed upon depletion of these factors are causally linked to TERRA mislocalization. Although it is tempting to speculate that UPF1, hEST1A/SMG6, and SMG1 actively displace TERRA molecules from telomeres, one cannot exclude that increased TERRA binding to telomeres might be an indirect consequence of the telomere damage occurring in these settings.

Another observation worth noting comes from overexpression experiments, with a mutant version of TRF2 harboring a deletion of the N-terminal basic domain (TRF2ΔB) (Deng et al. 2009). TRF2ΔB, ectopically expressed in U2OS cells, retains its ability to associate with telomeric DNA and to protect telomeres from NHEJ, although it promotes excision of T-loop-sized telomeric circles by homologous recombination. This leads to dramatic loss of telomeric DNA and eventually senescence (Wang et al. 2004). Interestingly, TERRA transcripts no longer form discrete foci in cells overexpressing TRF2ΔB (Deng et al. 2009). However, TERRA steady-state levels were not measured in these cells, thus making it unclear whether TRF2ΔB expression affects only TERRA localization or global TERRA levels.

4.2.3 TERRA and Interaction with Proteins

In a recent study, biotinylated TERRA-like RNA oligonucleotides were used to purify putative TERRA-binding factors from human nuclear extracts (Deng et al. 2009). This screening identified, among other factors, the shelterin proteins TRF1 and TRF2. RNA immunoprecipitation assays performed using antibodies against endogenous or ectopically expressed epitope-tagged shelterin proteins confirmed that TRF1 and TRF2 are able to interact with cellular TERRA (Deng et al. 2009; see Fig. 4.1). On the contrary, hRap1, POT1, and TPP1 appear not to bind TERRA in these assays (Deng et al. 2009). Furthermore, immunoprecipitation and electrophoretic mobility shift assays performed using recombinant TRF2 deletion mutants revealed that the basic aminoterminal GAR domain of TRF2, which is implicated in telomere stability and in the recruitment of the *O*rigin *R*ecognition *C*omplex (ORC) to telomeres, and, to a lesser extent, the carboxyterminal DNA-binding myb/SANT domain are involved in TERRA binding (Deng et al. 2009).

Other human nuclear proteins that were found to interact with TERRA include the telomerase-interacting partners hEST1A and dyskerin, the ORC subunits ORC1, ORC2, and ORC4, the *Me*thyl *C*pG-binding *P*rotein (MeCP2), and proteins involved in DNA metabolism such as DNA-dependent *P*rotein *K*inase *c*atalytic *s*ubunit (DNA PKcs), the *P*oly(*A*DP-*R*ibose) *P*olymerase 1 (PARP1) enzyme, BLM helicase, topoisomerase I, the ss DNA-binding protein RPA1, and *M*ediator of DNA *D*amage *C*heckpoint protein 1 (MDC1). In addition, the chromatin modifier COREST and several *h*eteroge*n*eous *R*ibo*N*ucleo*P*roteins (hnRNPs) were also identified in this study (Deng et al. 2009). In particular, hnRNPA1 seems to represent a very good candidate for a TERRA-interaction partner at telomeres: it possesses two

*R*NA-*R*ecognition *M*otifs (RRMs) and binds with very high specificity to ss G-rich telomeric DNA and RNA in vitro (McKay and Cooke 1992; Ishikawa et al. 1993). In addition, hnRNPA1 physically associates with telomeres, where it is thought to positively regulate telomerase-mediated telomere elongation (LaBranche et al. 1998; Zhang et al. 2006). It will be worth establishing loss-of-function experiments in order to test whether the RRMs of telomere-bound hnRNPA1 mediate TERRA localization to telomeres.

4.2.4 Regulation of TERRA Levels

The heterochromatic state of telomeres seems to impact on TERRA transcript steady-state levels. Treatment of human cells with trichostatin A, an inhibitor of classes I and II histone deacetylases, results in an increase in TERRA levels (Azzalin et al. 2007). Similarly, in cell lines derived from mice deficient for the histone methyltransferases Suv3-9h and Suv4-20h, the cellular levels of TERRA are elevated as compared to cells from wild-type mice (Schoeftner and Blasco 2008). These results suggest that TERRA is epigenetically regulated and that an open chromatin structure favors TERRA transcription. On the contrary, in mouse cells knocked-out for the telomerase RNA component Terc, TERRA levels are found to decrease (Schoeftner and Blasco 2008). The same cells also exhibit shortened telomeres, decreased levels of telomeric and subtelomeric methylated H3K9 and H4K20, and increased levels of acetylated histones H3 and H4 (Benetti et al. 2007a), suggesting that telomere shortening promotes the establishment of an "open" chromatin structure at both telomeric and subtelomeric regions and negatively regulates TERRA cellular levels. The apparent contradiction in terms of TERRA regulation as deduced by these different scenarios might be ascribed to multiple functions associated to the knocked-out polypeptides. For example, so far unknown functions of mouse telomerase RNA could be directly responsible for maintaining elevated TERRA levels.

The mouse *R*etino*b*lastoma (Rb) family of proteins consists of the three factors Rb1, RbL1, and RbL2. In humans, the Rb family includes pRb1, p107, and p130 (Longworth and Dyson 2010). Members of the Rb family function as tumor suppressors and affect gene expression by regulating the activity of the E2F family of transcription factors (Gonzalo and Blasco 2005). They also impact on gene expression by recruiting chromatin-modifying factors such as the histone methyltransferases Suv4-20h1 and Suv4-20h2 (Gonzalo et al. 2005). Three different Rb-deficient *M*ouse *E*mbryonic *F*ibroblast (MEF) cellular systems (Rb−/−, Rb−/− p107−/− and Rb−/−p107−/−p130−/−) display increased TERRA levels as compared to wild type controls (Gonzalez-Suarez et al. 2009). Importantly, concomitant deletion of all three Rb members leads to decreased H4K20 trimethylation at constitutive heterochromatin loci, and this change in histone modification is not due to a decrease in transcript levels of Suv4-20h1 and Suv4-20h2, although it could be rescued by overexpressing a full-length EGFP-tagged Suv4-20h

(Gonzalo et al. 2005; Benetti et al. 2007b). In addition to changes in histone methylation, Rb triple KO cells also exhibit decreased methylated cytosine levels (Gonzalo et al. 2005). Thus, alteration of TERRA transcript levels in Rb-deficient cells may be due to their impact on chromatin compaction through histone modification as well as DNA methylation.

The nuclear lamina consists of a dense fibrillar network that lines the inside of the nuclear envelope. Its two major components are class V intermediate filaments, called lamins, and the lamin-binding proteins. In vertebrates, lamins include A/C- or B-type *lamin*s (LMNA/C and LMNB), which are involved in nuclear organization and regulation of gene expression (Towbin et al. 2009). LMNA-deficient MEFs display a decrease in global H4K20me3 and markedly reduced TERRA levels (Gonzalez-Suarez et al. 2009), although it still remains to be determined whether this decrease is due to impaired TERRA transcription or augmented degradation. Interestingly, as pointed out by the authors of this study (Gonzalez-Suarez et al. 2009), Rb levels are also decreased in LMNA-deficient cells, indicating that TERRA upregulation observed upon Rb deletion might require intact lamins. Furthermore, the misregulation of TERRA levels in LMNA-deficient cells is accompanied by an altered positioning of telomeres within the nucleus, with an apparent nonrandom redistribution towards the nuclear periphery (Gonzalez-Suarez et al. 2009). Although the nuclear periphery is considered to be essentially transcriptionally silent (Kumaran et al. 2008), recent studies showed that, in mammals, transcription of a transgene can be induced even upon targeting to the nuclear periphery (Kumaran and Spector 2008). It is thus possible that changes in nuclear localization of telomeres are responsible, at least in part, for the reduced levels of TERRA transcripts. In this light, it will be interesting to analyze the rate of transcription of transgenic telomeres experimentally tethered to different compartments of the nucleus.

In a more recent study, the deposition of the histone variant H3.3 onto chromatin was indirectly linked to TERRA regulation. Histone H3.3 is specifically enriched at transcriptionally active gene promoters and at regulatory elements in pluripotent cells (Elsaesser et al. 2010). Chip-seq analysis additionally revealed H3.3 enrichment at telomeres (Goldberg et al. 2010). Deposition of H3.3 is generally mediated by the protein Hira, which acts in conjunction with the chromatin remodeler *C*hromodomain-*H*elicase-*D*NA-binding protein 1 (CHD1) (Elsaesser et al. 2010). Surprisingly, H3.3 deposition at telomeres was found to be Hira independent but dependent on the new H3.3-interacting partner *A*lpha *t*halassemia/mental *r*etardation syndrome *X*-linked (Atrx), which is also enriched at telomeric loci (Goldberg et al. 2010). In Atrx-deficient mouse ES cells, a 1.7-fold increase in TERRA levels is observed, independent of changes in H3K4 and H3K9 trimethylation levels (Goldberg et al. 2010), although the density of other heterochromatin marks still needs to be measured. An attractive, yet to be tested hypothesis is that the TERRA deregulation observed in Atrx-deficient cells might be due to improper H3.3 deposition at telomeres.

Dnmt-mediated methylation of cytosines at promoter CpG-dinucleotides is generally associated with transcriptional gene silencing (Esteller 2007). The 29 bp

and 37 bp repeats comprised in the identified TERRA promoters are methylated in different human cell lines, including cancer cells (HeLa, HCT116 and U20S) and primary lung fibroblasts (Nergadze et al. 2009). In a HCT116-derived cell line deficient for both Dnmt1 and Dnmt3b (double KO – DKO), DNA methylation at TERRA promoters is absent, while single deletion of only one of the two Dnmts does not perturb methylation levels substantially (Nergadze et al. 2009; Fig. 4.2). Thus, at least in the HCT116 cellular background, Dnmt1 and Dnmt3b cooperatively sustain methylation of TERRA promoter CpG dinucleotides. Importantly, the decreased methylation at TERRA promoters is accompanied by a dramatic increase in TERRA transcripts and by augmented binding of phosphorylation-activated RNAPII to TERRA promoters, suggesting that CpG methylation negatively regulates transcriptional activity of TERRA promoters (Nergadze et al. 2009). Similarly, treatment of different human cultured cells with 5-azacytidine, an inhibitor of Dnmts, results in increased TERRA levels (Nergadze et al. 2009). Interestingly, in peripheral blood mononuclear cells derived from infants, the levels of CpG methylation at subtelomeric CpG islands located on chromosomes 2p, 4p, and 18p are similar to those observed in cells derived from 69- to 89-year-old adults, while a significant decrease in telomere length is observed in the latter cells (Ng et al. 2009). The presence of 61-29-37 promoter sequences at 2p and 4p subtelomeres suggests that telomere shortening associated with aging does not affect TERRA promoter methylation state (Nergadze et al. 2009). It would be interesting to directly measure TERRA promoter CpG methylation and TERRA transcript levels in individuals of different ages.

Another intriguing connection between TERRA and methylation of subtelomeric regions emerged from the comparison of telomerase-positive cancer cells with cancer cells that resort to the ALT pathway to maintain telomere length. ALT cells exhibit overall decreased and more variable density of methylated CpG dinucleotides at subtelomeric loci as compared to telomerase positive cancer cells (Ng et al. 2009). Consistently, higher total TERRA levels are also observed in ALT cell lines (Ng et al. 2009; and Fig. 4.2). Importantly, telomeres within the same ALT cells are very heterogeneous in size (Cesare and Reddel 2010), suggesting that the methylation of each individual telomere might correlate with different telomere lengths as well as with different transcription rates of TERRA from individual promoters.

In another study, higher TERRA transcript levels were observed in primary cells derived from patients affected by *I*mmunodeficiency, *C*entromere instability, and *F*acial anomalies (ICF) syndrome, which, at the molecular level, is characterized by hypomethylated subtelomeric DNA arising from mutations in Dnmt3b (Yehezkel et al. 2008), further strengthening the idea that CpG methylation represses TERRA transcription. An apparent conundrum is nevertheless posed by the observation that mouse cells knocked-out for Dnmt1 or Dnmt3a/b display decreased TERRA levels as compared to wild type animal cells (Schoeftner and Blasco 2008). It is possible that different regulatory mechanisms exist amongst mice and humans, thus resulting in these contrasting observations.

As already mentioned, TERRA has been the object of studies also in nonmammalian eukaryotes. In *S. cerevisiae*, TERRA is kept at very low levels by the 5′ to 3′ RNA exonuclease *R*ibonucleic *a*cid *t*rafficking (Rat1p). Indeed, while TERRA is almost undetectable in wild type yeasts, it can be easily detected in a *rat1-1* mutant background using northern blot hybridization or RT-PCR (Luke et al. 2008). Interestingly, *rat1-1* telomeres are approximately 150 bp shorter than wild type counterparts. This telomere shortening is not incremented by concomitant deletion of telomerase components and is rescued by over-expression of RNase H, which specifically digests RNA molecules engaged in DNA/RNA hybrids (Luke et al. 2008). Altogether, these results suggest that increased TERRA expression might inhibit telomerase activity at telomeres, possibly by forming RNA–DNA hybrids with ss C-rich telomeric DNA exposed during DNA replication. Alternatively, TERRA might prevent telomerase action at telomeres by directly inhibiting its activity (see Sect. 4.2.5).

Alteration of different shelterin components seems also to impact on TERRA levels. Overexpression of TRF2 in mouse cells results in telomere shortening accompanied by a decrease in histone H3 and H4 abundance at telomeres and increased nucleosomal spacing (Benetti et al. 2008). Despite this open chromatin structure, a decrease in TERRA levels is observed (Benetti et al. 2008). On the other hand, depletion of TRF2 leads to an increase in TERRA transcript in a p53-dependent manner (Caslini et al. 2009). Although these observations seem to point to TRF2 as a negative regulator of TERRA abundance, one has to keep in mind that TRF2 depletion leads to accumulation of DNA damage at telomeres, raising the possibility that TERRA upregulation might be part of a physiological cellular response to telomere-specific DNA damage events.

SiRNA-mediated knockdown of TRF1 in human ALT cells has been shown to downregulate TERRA transcript levels. In addition, TRF1 physically associates with RNAPII, although TRF1 depletion does not result in impaired RNAPII recruitment to telomeres (Schoeftner and Blasco 2008). Thus, the impact of TRF1 depletion on TERRA levels seems not to depend on RNAPII recruitment to telomeric DNA but rather on a yet unidentified mechanism. On the contrary, conditional deletion of TRF1 in mouse embryonic fibroblasts gives rise to a severe block of replication fork progression through the telomeric tract without affecting TERRA steady-state levels (Sfeir et al. 2009). The differences between these two studies could again reflect differences between mice and humans in terms of TERRA regulatory circuits.

Recently, a connection between TERRA and small RNA species was uncovered. Three major classes of small RNA molecules have been discovered so far (1) short interfering RNA (siRNA); (2) micro RNA (miRNA); and (3) PIWI interacting RNAs (piRNA) (Jinek and Doudna 2009). In mammalian cells, the siRNA pathway mediates gene silencing prevalently by degrading target mRNAs, while miRNAs have been implicated in regulating gene expression through translational inhibition. The protein Dicer, an endoribonuclease belonging to the RNAseIII family, mediates the cleavage of ds RNA and pre-miRNA molecules into siRNA and miRNA duplexes that are 20–25 nt long (Jinek and Doudna 2009). On the other hand,

piRNAs are 24–31 nt long, and their biogenesis, which is Dicer independent, and function remain poorly defined. Interestingly, in mouse ES cells with compromised Dicer activity, dot blot analysis showed a decrease in TERRA levels (Schoeftner and Blasco 2008). On the contrary, in an independent study, northern blot analysis of RNA from Dicer-deficient mouse ES cell showed elevated TERRA levels (Zhang et al. 2009). In a third separate study, 23–24 nt long TERRA-like RNA species were identified in mouse ES cells as well as in human somatic cells, although at much lower levels (Cao et al. 2009), and no change in this RNA species was observed in Dicer-deficient cells. Functional ablation of the H3K4 methyltransferase *M*yeloid/*L*ymphoid *L*eukemia, alternatively named *M*ixed *L*ineage *L*eukemia (MLL), which promotes deposition of the euchromatic mark H3K4me3, leads to a twofold increase in TERRA-like small RNA (Cao et al. 2009) and concomitant decrease in long TERRA species in different cells (Caslini et al. 2009). This is accompanied by a decrease in H3K4me3 and, surprisingly, an increase in H3K9me3 density at telomeres (Caslini et al. 2009). In addition, diminished binding of RNAPII to telomeres was also observed (Cao et al. 2009). Altogether, these studies raise some important questions about TERRA RNA metabolism and the role of TERRA-like small RNAs in telomere biology. In particular, are the 23–24 nt long TERRA transcripts generated via the degradation of the longer TERRA molecules or via alternative pathways? Have the two transcript families independent roles? Do the smaller RNA species regulate transcription of the longer TERRA molecules or vice versa?

4.2.5 TERRA-Associated Functions

The characterization of TERRA-associated putative functions still remains a major challenge for current and future research. Given the exclusive localization of TERRA to the nucleus and the cytoplasmic localization of the siRNA machinery, one would predict that TERRA is unlikely to be knocked-down using canonical siRNA-mediated approaches. Unexpectedly, however, transfection of human cancer cells with siRNA molecules against TERRA UUAGGG repeats resulted in a 40% reduction in TERRA levels and in a substantial decrease in the number of TERRA nuclear foci (Deng et al. 2009). TERRA siRNA transfection was accompanied by loss in cell viability, TIF formation, telomeric aberrations, diminished recruitment of ORC to telomeres, and decreased density of telomere-bound di- and trimethylated H3K9 (Deng et al. 2009). Although this set of data makes it tempting to speculate that TERRA could play fundamental roles in maintaining telomere integrity and in telomeric heterochromatin establishment, it still remains possible that at least some of the observed phenotypes could result from secondary effects exerted by the TERRA-like siRNA molecules, independently of TERRA downregulation and mislocalization. Indeed, transfection of short (TTAGGG)*n* oligonucleotides in human cells generates a severe DNA damage response at telomeres, perhaps by sequestering POT1 molecules from the G-overhang (Milyavsky et al. 2001). It is

also conceivable that the observed phenotypes (including TERRA down-regulation itself) might derive from a so far unforeseen function of TERRA-like small RNA molecules at telomeres rather than from direct downregulation of TERRA through noncanonical nuclear RNAi machineries. In addition, transient transfection of random siRNA molecules into mammalian cells induced heterochromatinization of telomeres and upregulation of TERRA transcripts (Ho et al. 2008), rendering even more problematic to unequivocally interpret results obtained using siRNA-based approaches.

It has been proposed that TERRA negatively regulates telomerase-mediated telomere elongation. Indeed, telomerase-positive cells exhibit higher methylation levels of TERRA promoters and lower TERRA levels as compared to telomerase-negative cell lines (Ng et al. 2009). Similarly, TERRA levels are diminished in high-grade tumor cells as compared to cells from low-grade tumors, and low TERRA levels are observed during mouse embryonic development when telomerase is highly active (Schoeftner and Blasco 2008). In addition, RNA oligonucleotides comprising the TERRA-like sequence $(UUAGGG)_3$ inhibit telomerase activity in vitro (Schoeftner and Blasco 2008; Redon et al. 2010). Finally, *rat1Δ* yeast strains display high TERRA levels and short telomeres (Luke et al. 2008; see Sect. 4.2.4), and forced transcription of a yeast chromosome end leads to shortening of its telomeric tract (Sandell et al. 1994). Because telomeric TERRA repeats are complementary to the template region of telomerase RNA, it is likely that TERRA-mediated inhibition of telomerase occurs through competitive base-pairing. Indeed, short UUAGGG RNA sequences seem to prevent telomerase action at telomeres also in vivo, when overexpressed from transgenic retroviral promoters integrated in the genome (Bisoffi et al. 1998).

4.3 Conclusions and Future Directions

The discovery of TERRA is fuelling research in a previously unforeseen aspect of telomere biology and promises to generate new and exciting data, thus adding to the complexity and pleiotropic nature of telomeres. One crucial aspect that urgently needs to be clarified is what functions TERRA and/or transcription exert at telomeres. As already mentioned, proper TERRA binding to telomeres seems to be essential for telomere integrity, telomere replication, and heterochromatin deposition. Independent loss- or gain-of-function systems need to be developed in order to dissect TERRA roles in these different aspects of telomere biology.

How eukaryotic cells assure proficient TERRA transcription also needs to be further characterized. With the isolation of TERRA promoters, the way has been paved for the identification of TERRA transcription factors and for the characterization of the roles played by such factors in TERRA biogenesis and in maintaining correct telomere structure and functions. Also, a possible involvement of different transcription machineries in TERRA biogenesis remains to be carefully tested.

Finally, once the molecular details of TERRA biogenesis and functions are elucidated, it will be essential to place TERRA in the wider context of telomere-associated functions during cellular senescence, organismal aging, and cancerogenesis. A direct involvement of TERRA in these crucial aspects of human biology could, in the long term, open the way for new therapeutic approaches for curing age-associated diseases and cancer. The landing on "TERRA," the Latin noun for planet Earth, has indeed marked the beginning of a new era in telomere biology.

Acknowledgments We thank Harry Wischnewski for contributing to Fig. 4.2. We also thank Raghav Chawla and Amadou Bah for critical reading of the manuscript. CMA laboratory is supported by grants from ETHZ (ETH-15 08-1 and ETH-03 08-3), the Swiss National Science Foundation (3100A0-120090 and PP00P3-123356), the European Research Council (BFTERRA), and Fondazione Cariplo (2008-2507). CMCB is supported by a fellowship from Boehringer Ingelheim Fonds.

References

Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T (2003) The small RNA profile during *Drosophila melanogaster* development. Dev Cell 5:337–350

Artandi SE, DePinho RA (2010) Telomeres and telomerase in cancer. Carcinogenesis 31:9–18

Azzalin CM, Lingner J (2006) The human RNA surveillance factor UPF1 is required for S phase progression and genome stability. Curr Biol 16:433–439

Azzalin CM, Lingner J (2008) Telomeres: the silence is broken. Cell Cycle 7:1161–1165

Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318:798–801

Benetti R, Garcia-Cao M, Blasco MA (2007a) Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. Nat Genet 39:243–250

Benetti R, Gonzalo S, Jaco I, Schotta G, Klatt P, Jenuwein T, Blasco MA (2007b) Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. J Cell Biol 178:925–936

Benetti R, Schoeftner S, Munoz P, Blasco MA (2008) Role of TRF2 in the assembly of telomeric chromatin. Cell Cycle 7:3461–3468

Bianchi A, Shore D (2008) How telomerase reaches its end: mechanism of telomerase regulation by the telomeric complex. Mol Cell 31:153–165

Bisoffi M, Chakerian AE, Fore ML, Bryant JE, Hernandez JP, Moyzis RK, Griffith JK (1998) Inhibition of human telomerase by a retrovirus expressing telomeric antisense RNA. Eur J Cancer 34:1242–1249

Blasco MA (2007) The epigenetic regulation of mammalian telomeres. Nat Rev Genet 8:299–309

Bourns BD, Alexander MK, Smith AM, Zakian VA (1998) Sir proteins, Rif proteins, and Cdc13p bind *Saccharomyces* telomeres in vivo. Mol Cell Biol 18:5600–5608

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell 128:1089–1103

Brumbaugh KM, Otterness DM, Geisen C, Oliveira V, Brognard J, Li X, Lejeune F, Tibbetts RS, Maquat LE, Abraham RT (2004) The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. Mol Cell 14:585–598

Bryan TM, Marusic L, Bacchetti S, Namba M, Reddel RR (1997) The telomere lengthening mechanism in telomerase-negative immortal human cells does not involve the telomerase RNA subunit. Hum Mol Genet 6:921–926

Campisi J, Yaswen P (2009) Aging and cancer cell biology, 2009. Aging Cell 8:221–225

Cao F, Li X, Hiew S, Brady H, Liu Y, Dou Y (2009) Dicer independent small RNAs associate with telomeric heterochromatin. RNA 15:1274–1281

Caslini C, Connelly JA, Serna A, Broccoli D, Hess JL (2009) MLL associates with telomeres and regulates telomeric repeat-containing RNA transcription. Mol Cell Biol 29:4519–4526

Celli GB, de Lange T (2005) DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. Nat Cell Biol 7:712–718

Cenci G, Rawson RB, Belloni G, Castrillon DH, Tudor M, Petrucci R, Goldberg ML, Wasserman SA, Gatti M (1997) UbcD1, a *Drosophila* ubiquitin-conjugating enzyme required for proper telomere behavior. Genes Dev 11:863–875

Cesare AJ, Reddel RR (2010) Alternative lengthening of telomeres: models, mechanisms and implications. Nat Rev Genet 11(5):319–330

Chandra A, Hughes TR, Nugent CI, Lundblad V (2001) Cdc13 both positively and negatively regulates telomere replication. Genes Dev 15:404–414

Colgin LM, Baran K, Baumann P, Cech TR, Reddel RR (2003) Human POT1 facilitates telomere elongation by telomerase. Curr Biol 13:942–946

Cooper JP, Nimmo ER, Allshire RC, Cech TR (1997) Regulation of telomere length and function by a Myb-domain protein in fission yeast. Nature 385:744–747

Déjardin J, Kingston RE (2009) Purification of proteins associated with specific genomic Loci. Cell 136:175–186

Denchi EL, de Lange T (2007) Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. Nature 448:1068–1071

Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM (2009) TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. Mol Cell 35:403–413

Elsaesser SJ, Goldberg AD, Allis CD (2010) New functions for an old variant: no substitute for histone H3.3. Curr Opin Genet Dev 20(2):110–117

Esteller M (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet 8:286–298

Fajkus J, Kovarik A, Kralovics R, Bezdek M (1995) Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. Mol Gen Genet 247:633–638

Fanti L, Giovinazzo G, Berloco M, Pimpinelli S (1998) The heterochromatin protein 1 prevents telomere fusions in *Drosophila*. Mol Cell 2:527–538

Ferreira MG, Cooper JP (2001) The fission yeast Taz1 protein protects chromosomes from Ku-dependent end-to-end fusions. Mol Cell 7:55–63

Fisher TS, Zakian VA (2005) Ku: a multifunctional protein involved in telomere maintenance. DNA Repair (Amst) 4:1215–1226

Fitzgerald MS, McKnight TD, Shippen DE (1996) Characterization and developmental patterns of telomerase expression in plants. Proc Natl Acad Sci USA 93:14422–14427

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, Iyer NG, Pérez-Rosado A, Calvo E, Lopez JA, Cano A, Calasanz MJ, Colomer D, Piris MA, Ahn N, Imhof A, Caldas C, Jenuwein T, Esteller M (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet 37:391–400

Gallego ME, Jalut N, White CI (2003) Telomerase dependence of telomere lengthening in Ku80 mutant *Arabidopsis*. Plant Cell 15:782–789

Gao H, Cervantes RB, Mandell EK, Otero JH, Lundblad V (2007) RPA-like proteins mediate yeast telomere function. Nat Struct Mol Biol 14:208–214

Gao G, Walser JC, Beaucher ML, Morciano P, Wesolowska N, Chen J, Rong YS (2010) HipHop interacts with HOAP and HP1 to protect *Drosophila* telomeres in a sequence-independent manner. EMBO J 29(4):819–829

Garcia-Cao M, O'Sullivan R, Peters AH, Jenuwein T, Blasco MA (2004) Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. Nat Genet 36:94–99

Garvik B, Carson M, Hartwell L (1995) Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint. Mol Cell Biol 15:6128–6138

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, Wen D, Chapgier A, DeKelver RC, Miller JC, Lee YL, Boydston EA, Holmes MC, Gregory PD, Greally JM, Rafii S, Yang C, Scambler PJ, Garrick D, Gibbons RJ, Higgs DR, Cristea IM, Urnov FD, Zheng D, Allis CD (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140:678–691

Gonzalez-Suarez I, Redwood AB, Perkins SM, Vermolen B, Lichtensztejin D, Grotsky DA, Morgado-Palacin L, Gapud EJ, Sleckman BP, Sullivan T, Sage J, Stewart CL, Mai S, Gonzalo S (2009) Novel roles for A-type lamins in telomere biology and the DNA damage response pathway. EMBO J 28:2414–2427

Gonzalo S, Blasco MA (2005) Role of Rb family in the epigenetic definition of chromatin. Cell Cycle 4:752–755

Gonzalo S, Garcia-Cao M, Fraga MF, Schotta G, Peters AH, Cotter SE, Eguia R, Dean DC, Esteller M, Jenuwein T, Blasco MA (2005) Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. Nat Cell Biol 7:420–428

Gonzalo S, Jaco I, Fraga MF, Chen T, Li E, Esteller M, Blasco MA (2006) DNA methyltransferases control telomere length and telomere recombination in mammalian cells. Nat Cell Biol 8:416–424

Grandin N, Damon C, Charbonneau M (2000) Cdc13 cooperates with the yeast Ku proteins and Stn1 to regulate telomerase recruitment. Mol Cell Biol 20:8397–8408

Greider CW, Blackburn EH (1985) Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. Cell 43:405–413

Harley CB, Futcher AB, Greider CW (1990) Telomeres shorten during ageing of human fibroblasts. Nature 345:458–460

Ho CY, Murnane JP, Yeung AK, Ng HK, Lo AW (2008) Telomeres acquire distinct heterochromatin characteristics during siRNA-induced RNA interference in mouse cells. Curr Biol 18:183–187

Hughes TR, Evans SK, Weilbaecher RG, Lundblad V (2000) The Est3 protein is a subunit of yeast telomerase. Curr Biol 10:809–812

Ishikawa F, Matunis MJ, Dreyfuss G, Cech TR (1993) Nuclear proteins that bind the pre-mRNA 3′ splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. Mol Cell Biol 13:4301–4310

Jinek M, Doudna JA (2009) A three-dimensional view of the molecular machinery of RNA interference. Nature 457:405–412

Kahn T, Savitsky M, Georgiev P (2000) Attachment of HeT-A sequences to chromosomal termini in *Drosophila melanogaster* may occur by different mechanisms. Mol Cell Biol 20:7634–7642

Karlseder J, Hoke K, Mirzoeva OK, Bakkenist C, Kastan MB, Petrini JH, de Lange T (2004) The telomeric protein TRF2 binds the ATM kinase and can inhibit the ATM-dependent DNA damage response. PLoS Biol 2:E240

Kendellen MF, Barrientos KS, Counter CM (2009) POT1 association with TRF2 regulates telomere length. Mol Cell Biol 29:5611–5619

Kibe T, Ono Y, Sato K, Ueno M (2007) Fission yeast Taz1 and RPA are synergistically required to prevent rapid telomere loss. Mol Biol Cell 18:2378–2387

Killan A, Heller K, Kleinhofs A (1998) Development patterns of telomerase activity in barley and maize. Plant Mol Biol 37:621–628

Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL, Shay JW (1994) Specific association of human telomerase activity with immortal cells and cancer. Science 266:2011–2015

Kumaran RI, Spector DL (2008) A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. J Cell Biol 180:51–65

Kumaran RI, Thakar R, Spector DL (2008) Chromatin dynamics and gene positioning. Cell 132:929–934

Kwon C, Chung IK (2004) Interaction of an *Arabidopsis* RNA-binding protein with plant single-stranded telomeric DNA modulates telomerase activity. J Biol Chem 279:12812–12818

Kyrion G, Boakye KA, Lustig AJ (1992) C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in *Saccharomyces cerevisiae*. Mol Cell Biol 12:5159–5173

LaBranche H, Dupuis S, Ben-David Y, Bani MR, Wellinger RJ, Chabot B (1998) Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. Nat Genet 19:199–202

Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410:116–120

Larrivee M, LeBel C, Wellinger RJ (2004) The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. Genes Dev 18:1391–1396

Latrick CM, Cech TR (2010) POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. EMBO J 29(5):924–933

Lei M, Podell ER, Cech TR (2004) Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. Nat Struct Mol Biol 11:1223–1229

Leonardi J, Box JA, Bunch JT, Baumann P (2008) TER1, the RNA subunit of fission yeast telomerase. Nat Struct Mol Biol 15:26–33

Li B, Oestreich S, de Lange T (2000) Identification of human Rap1: implications for telomere evolution. Cell 101:471–483

Lin JJ, Zakian VA (1996) The *Saccharomyces* CDC13 protein is a single-strand TG1-3 telomeric DNA-binding protein in vitro that affects telomere behavior in vivo. Proc Natl Acad Sci USA 93:13760–13765

Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR (1997) Reverse transcriptase motifs in the catalytic subunit of telomerase. Science 276:561–567

Liu D, Safari A, O'Connor MS, Chan DW, Laegeler A, Qin J, Songyang Z (2004) PTOP interacts with POT1 and regulates its localization to telomeres. Nat Cell Biol 6:673–680

Loayza D, De Lange T (2003) POT1 as a terminal transducer of TRF1 telomere length control. Nature 423:1013–1018

Longworth MS, Dyson NJ (2010) pRb, a local chromatin organizer with global possibilities. Chromosoma 119:1–11

Luke B, Panza A, Redon S, Iglesias N, Li Z, Lingner J (2008) The Rat1p 5′ to 3′ exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. Mol Cell 32:465–477

Lundblad V, Blackburn EH (1990) RNA-dependent polymerase motifs in EST1: tentative identification of a protein component of an essential yeast telomerase. Cell 60:529–530

Lundblad V, Szostak JW (1989) A mutant with a defect in telomere elongation leads to senescence in yeast. Cell 57:633–643

Marcand S, Wotton D, Gilson E, Shore D (1997) Rap1p and telomere length regulation in yeast. Ciba Found Symp 211:76–93, discussion 93–103

Mason JM, Frydrychova RC, Biessmann H (2008) *Drosophila* telomeres: an exception providing new insights. Bioessays 30:25–37

McClintock B (1941) The stability of broken ends of chromosomes in *Zea mays*. Genetics 26:234–282

McKay SJ, Cooke H (1992) hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGGn. Nucleic Acids Res 20:6461–6464

Mikhailovsky S, Belenkaya T, Georgiev P (1999) Broken chromosomal ends can be elongated by conversion in *Drosophila melanogaster*. Chromosoma 108:114–120

Miller KM, Ferreira MG, Cooper JP (2005) Taz1, Rap1 and Rif1 act both interdependently and independently to maintain telomeres. EMBO J 24:3128–3135

Miller KM, Rog O, Cooper JP (2006) Semi-conservative DNA replication through telomeres requires Taz1. Nature 440:824–828

Milyavsky M, Mimran A, Senderovich S, Zurer I, Erez N, Shats I, Goldfinger N, Cohen I, Rotter V (2001) Activation of p53 protein by telomeric (TTAGGG)n repeats. Nucleic Acids Res 29:5207–5215

Mitchell JR, Cheng J, Collins K (1999a) A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3′ end. Mol Cell Biol 19:567–576

Mitchell JR, Wood E, Collins K (1999b) A telomerase component is defective in the human disease dyskeratosis congenita. Nature 402:551–555

Miyoshi T, Kanoh J, Saito M, Ishikawa F (2008) Fission yeast Pot1-Tpp 1 protects telomeres and regulates telomere length. Science 320:1341–1344

Muller HJ (1938) The remaking of chromosomes. Collecting Net 13:181–189

Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, Cech TR (1997) Telomerase catalytic subunit homologs from fission yeast and human. Science 277:955–959

Nergadze SG, Farnung BO, Wischnewski H, Khoriauli L, Vitelli V, Chawla R, Giulotto E, Azzalin CM (2009) CpG-island promoters drive transcription of human telomeres. RNA 15:2186–2194

Ng LJ, Cropley JE, Pickett HA, Reddel RR, Suter CM (2009) Telomerase activity is associated with an increase in DNA methylation at the proximal subtelomere and a reduction in telomeric transcription. Nucleic Acids Res 37:1152–1159

Nicholson P, Yepiskoposyan H, Metze S, Zamudio Orozco R, Kleinschmidt N, Muhlemann O (2010) Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cell Mol Life Sci 67:677–700

Nugent CI, Hughes TR, Lue NF, Lundblad V (1996) Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. Science 274:249–252

O'Connor MS, Safari A, Xin H, Liu D, Songyang Z (2006) A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly. Proc Natl Acad Sci USA 103:11874–11879

Ogawa Y, Sun BK, Lee JT (2008) Intersection of the RNA interference and X-inactivation pathways. Science 320:1336–1341

Oliveira V, Romanow WJ, Geisen C, Otterness DM, Mercurio F, Wang HG, Dalton WS, Abraham RT (2008) A protective role for the human SMG-1 kinase against tumor necrosis factor-alpha-induced apoptosis. J Biol Chem 283:13174–13184

Olovnikov AM (1973) A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. J Theor Biol 41:181–190

Palm W, de Lange T (2008) How shelterin protects mammalian telomeres. Annu Rev Genet 42:301–334

Perrini B, Piacentini L, Fanti L, Altieri F, Chichiarelli S, Berloco M, Turano C, Ferraro A, Pimpinelli S (2004) HP1 controls telomere capping, telomere elongation, and telomere silencing by two different mechanisms in *Drosophila*. Mol Cell 15:467–476

Peterson SE, Stellwagen AE, Diede SJ, Singer MS, Haimberger ZW, Johnson CO, Tzoneva M, Gottschling DE (2001) The function of a stem-loop in telomerase RNA is linked to the DNA repair protein Ku. Nat Genet 27:64–67

Petreaca RC, Chiu HC, Eckelhoefer HA, Chuang C, Xu L, Nugent CI (2006) Chromosome end protection plasticity revealed by Stn1p and Ten1p bypass of Cdc13p. Nat Cell Biol 8:748–755

Qi H, Zakian VA (2000) The *Saccharomyces* telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein. Genes Dev 14:1777–1788

Raffa GD, Cenci G, Siriaco G, Goldberg ML, Gatti M (2005) The putative *Drosophila* transcription factor woc is required to prevent telomeric fusions. Mol Cell 20:821–831

Redon S, Reichenbach P, Lingner J (2007) Protein RNA and protein protein interactions mediate association of human EST1A/SMG6 with telomerase. Nucleic Acids Res 35:7011–7022

Redon S, Reichenbach P, Lingner J (2010) The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. Nucleic Acids Res 38(17):5797–5806

Reichenbach P, Höss M, Azzalin CM, Nabholz M, Bucher P, Lingner J (2003) A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed. Curr Biol 13:568–574

Richards EJ, Ausubel FM (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. Cell 53:127–136

Riethman H (2008) Human telomere structure and biology. Annu Rev Genomics Hum Genet 9:1–19

Riha K, Shippen DE (2003) Telomere structure, function and maintenance in *Arabidopsis*. Chromosome Res 11:263–275

Riha K, McKnight TD, Fajkus J, Vyskot B, Shippen DE (2000) Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. Plant J 23:633–641

Riha K, McKnight TD, Griffing LR, Shippen DE (2001) Living with genome instability: plant responses to telomere dysfunction. Science 291:1797–1800

Rong YS (2008a) Loss of the histone variant H2A.Z restores capping to checkpoint-defective telomeres in *Drosophila*. Genetics 180:1869–1875

Rong YS (2008b) Telomere capping in *Drosophila*: dealing with chromosome ends that most resemble DNA breaks. Chromosoma 117:235–242

Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. Genes Dev 20:2214–2222

Sandell LL, Gottschling DE, Zakian VA (1994) Transcription of a yeast telomere alleviates telomere position effect without affecting chromosome stability. Proc Natl Acad Sci USA 91:12061–12065

Sarthy J, Bae NS, Scrafford J, Baumann P (2009) Human RAP1 inhibits non-homologous end joining at telomeres. EMBO J 28:3390–3399

Schoeftner S, Blasco MA (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat Cell Biol 10:228–236

Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev 18:1251–1262

Seto AG, Zaug AJ, Sobel SG, Wolin SL, Cech TR (1999) *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle. Nature 401:177–180

Seto AG, Livengood AJ, Tzfati Y, Blackburn EH, Cech TR (2002) A bulged stem tethers Est1p to telomerase RNA in budding yeast. Genes Dev 16:2800–2812

Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, de Lange T (2009) Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. Cell 138:90–103

Sfeir A, Kabir S, van Overbeek M, Celli GB, de Lange T (2010) Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. Science 327:1657–1661

Shakirov EV, Shippen DE (2004) Length regulation and dynamics of individual telomere tracts in wild-type *Arabidopsis*. Plant Cell 16:1959–1967

Shakirov EV, Surovtseva YV, Osbun N, Shippen DE (2005) The *Arabidopsis* Pot1 and Pot2 proteins function in telomere length homeostasis and chromosome end protection. Mol Cell Biol 25:7725–7733

Shampay J, Szostak JW, Blackburn EH (1984) DNA sequences of telomeres maintained in yeast. Nature 310:154–157

Shay JW, Bacchetti S (1997) A survey of telomerase activity in human cancer. Eur J Cancer 33:787–791

Shay JW, Pereira-Smith OM, Wright WE (1991) A role for both RB and p53 in the regulation of human cellular senescence. Exp Cell Res 196:33–39
Shippen DE (2006) Plant telomeres. In: deLange T, Lundblad V, Blackburn EH (eds) Telomeres. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, p 525
Singer MS, Gottschling DE (1994) TLC1: template RNA component of *Saccharomyces cerevisiae* telomerase. Science 266:404–409
Smogorzewska A, van Steensel B, Bianchi A, Oelmann S, Schaefer MR, Schnapp G, de Lange T (2000) Control of human telomere length by TRF1 and TRF2. Mol Cell Biol 20:1659–1668
Smogorzewska A, Karlseder J, Holtgreve-Grez H, Jauch A, de Lange T (2002) DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2. Curr Biol 12:1635–1644
Snow BE, Erdmann N, Cruickshank J, Goldman H, Gill RM, Robinson MO, Harrington L (2003) Functional conservation of the telomerase protein Est1p in humans. Curr Biol 13:698–704
Solovei I, Gaginskaya ER, Macgregor HC (1994) The arrangement and transcription of telomere DNA sequences at the ends of lampbrush chromosomes of birds. Chromosome Res 2:460–470
Sugawara N (1989) DNA sequences at the telomeres of the fission yeast *S. pombe*. Ph.D. thesis, Harvard University, Cambridge, MA
Surovtseva YV, Shakirov EV, Vespa L, Osbun N, Song X, Shippen DE (2007) *Arabidopsis* POT1 associates with the telomerase RNP and is required for telomere maintenance. EMBO J 26:3653–3661
Szostak JW, Blackburn EH (1982) Cloning yeast telomeres on linear plasmid vectors. Cell 29:245–255
Takai H, Smogorzewska A, de Lange T (2003) DNA damage foci at dysfunctional telomeres. Curr Biol 13:1549–1556
Teixeira MT, Gilson E (2005) Telomere maintenance, function and evolution: the yeast paradigm. Chromosome Res 13:535–548
Tomita K, Cooper JP (2008) Fission yeast Ccq1 is telomerase recruiter and local checkpoint controller. Genes Dev 22:3461–3474
Tomita K, Matsuura A, Caspari T, Carr AM, Akamatsu Y, Iwasaki H, Mizuno K, Ohta K, Uritani M, Ushimaru T, Yoshinaga K, Ueno M (2003) Competition between the Rad50 complex and the Ku heterodimer reveals a role for Exo1 in processing double-strand breaks but not telomeres. Mol Cell Biol 23:5186–5197
Torok T, Benitez C, Takacs S, Biessmann H (2007) The protein encoded by the gene proliferation disrupter (prod) is associated with the telomeric retrotransposon array in *Drosophila melanogaster*. Chromosoma 116:185–195
Towbin BD, Meister P, Gasser SM (2009) The nuclear envelope-a scaffold for silencing? Curr Opin Genet Dev 19:180–186
van Steensel B, de Lange T (1997) Control of telomere length by the human telomeric protein TRF1. Nature 385:740–743
van Steensel B, Smogorzewska A, de Lange T (1998) TRF2 protects human telomeres from end-to-end fusions. Cell 92:401–413
Vega LR, Mateyak MK, Zakian VA (2003) Getting to the end: telomerase access in yeast and humans. Nat Rev Mol Cell Biol 4:948–959
Venteicher AS, Meng Z, Mason PJ, Veenstra TD, Artandi SE (2008) Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. Cell 132:945–957
Venteicher AS, Abreu EB, Meng Z, McCann KE, Terns RM, Veenstra TD, Terns MP, Artandi SE (2009) A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. Science 323:644–648
Vermaak D, Malik HS (2009) Multiple roles for heterochromatin protein 1 genes in *Drosophila*. Annu Rev Genet 43:467–492
Wang SS, Zakian VA (1990) Sequencing of *Saccharomyces* telomeres cloned using T4 DNA polymerase reveals two domains. Mol Cell Biol 10:4415–4419

Wang RC, Smogorzewska A, de Lange T (2004) Homologous recombination generates T-loop-sized deletions at human telomeres. Cell 119:355–368

Watson JD (1972) Origin of concatemeric T7 DNA. Nat New Biol 239:197–201

Webb CJ, Zakian VA (2008) Identification and characterization of the *Schizosaccharomyces pombe* TER1 telomerase RNA. Nat Struct Mol Biol 15:34–42

Wellinger RJ, Wolf AJ, Zakian VA (1993) *Saccharomyces* telomeres acquire single-strand TG1-3 tails late in S phase. Cell 72:51–60

Yang SW, Kim SK, Kim WT (2004) Perturbation of NgTRF1 expression induces apoptosis-like cell death in tobacco BY-2 cells and implicates NgTRF1 in the control of telomere length and stability. Plant Cell 16:3370–3385

Yao MC, Blackburn E, Gall JG (1979) Amplification of the rRNA genes in *Tetrahymena*. Cold Spring Harb Symp Quant Biol 43(Pt 2):1293–1296

Yao MC, Blackburn E, Gall J (1981) Tandemly repeated C-C-C-C-A-A hexanucleotide of *Tetrahymena* rDNA is present elsewhere in the genome and may be related to the alteration of the somatic genome. J Cell Biol 90:515–520

Ye JZ, Hockemeyer D, Krutchinsky AN, Loayza D, Hooper SM, Chait BT, de Lange T (2004) POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. Genes Dev 18:1649–1654

Yehezkel S, Segev Y, Viegas-Pequignot E, Skorecki K, Selig S (2008) Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. Hum Mol Genet 17:2776–2789

Zellinger B, Riha K (2007) Composition of plant telomeres. Biochim Biophys Acta 1769:399–409

Zhang QS, Manche L, Xu RM, Krainer AR (2006) hnRNP A1 associates with telomere ends and stimulates telomerase activity. RNA 12:1116–1128

Zhang LF, Ogawa Y, Ahn JY, Namekawa SH, Silva SS, Lee JT (2009) Telomeric RNAs mark sex chromosomes in stem cells. Genetics 182:685–698

Chapter 5
Transcription of Satellite DNAs in Mammals

Claire Vourc'h and Giuseppe Biamonti

Abstract Centromeric and pericentric regions have long been regarded as transcriptionally inert portions of chromosomes. A number of studies in the past 10 years disproved this dogma and provided convincing evidence that centromeric and pericentric sequences are transcriptionally active in several biological contexts.

In this chapter, we provide a comprehensive picture of the various contexts (cell growth and differentiation, stress, effect of chromatin organization) in which these sequences are expressed in mouse and human cells and discuss the possible functional implications of centromeric and pericentric sequences activation and/or of the resulting noncoding RNAs. Moreover, we provide an overview of the molecular mechanisms underlying the activation of centromeric and pericentromeric sequences as well as the structural features of encoded RNAs.

5.1 Introduction

In eukaryotic cells, correct segregation and inheritance of genetic information rely on the activity of specialized chromosomal regions called centromeres, which ensure that during mitosis, each daughter cell receives one copy of each chromosome. Defects in chromosome segregation are associated with human disease. Defects in meiosis lead to aneuploid embryos and cause genetic syndromes while mitotic errors contribute to tumor formation. One major centromeric function is to dictate the site of assembly of the kinetochore, a critical structure that mediates

C. Vourc'h
Université Joseph Fourier-Grenoble; INSERM U823; Institut Albert Bonniot, La Tronche BP170, 38042 Grenoble cedex 9, France
e-mail: claire.vourch@ujf-grenoble.fr

G. Biamonti (✉)
Istituto di Genetica Molecolare CNR, Via Abbiategrasso 207, 27100 Pavia, Italy
e-mail: biamonti@igm.cnr.it

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3_5,

binding of chromosomes to the spindle, monitors bipolar attachment, and pulls chromosomes to the poles during anaphase. The centromere region also contributes to sister chromatid cohesion function via a second centromeric domain, namely the pericentric heterochromatin structure, which surrounds the kinetochore. Although centromeres have been identified more than a century ago as the primary constriction of condensed metaphase chromosomes, their molecular characterization was hampered for a long time by its unusual enrichment in highly repetitive satellite DNA sequences. Two types of repetitive DNA sequences are usually associated with centromeres: major satellite repeats that are located pericentrically (PCT) and the minor satellite repeats that coincide with the centric (CT) constriction. The poor evolutionary conservation of all these elements underscore the fact that in most eukaryotes, including mammals, centromere identity and function is not simply specified by DNA sequence. This finding has led to the concept of an "epigenetic" component in centromere function that can be inherited throughout multiple divisions. An increasing number of epigenetic marks have been uncovered that are associated with the constitutive heterochromatic state of centromeric regions, which remain condensed during the entire cell cycle. Clear examples of epigenetic control come from the analysis of neocentromeres, where new centromeres are formed on noncentromeric DNA, and inactivation of one centromere in dicentric chromosomes. Current centromere models indicate that, once formed, centromeres are specified epigenetically and maintained at the same locus, cell division after cell division. Centromeric protein A (CENPA) has emerged as the best candidate to carry the epigenetic centromere mark, while specific histone modifications and heterochromatin protein 1 (HP1) are distinguishing features of pericentric heterochromatin.

A number of studies in the last 10 years have shown that the epigenetic status at centromeres is controlled by the concerted action of several mechanisms involving nucleosome remodeling, the histone variant CenH3, and histone modifications. More recently, a role of transcription and RNA in establishing the correct centric and pericentric chromatin status has emerged. In this chapter, we give a comprehensive view of the different aspects controlling the expression of pericentic and centric RNAs and discuss the role of unscheduled expression on the centromere function.

5.2 General Organization of Centromeric and Pericentric Regions in Mouse and Human Chromosomes

In mouse and human cells, centromeric and pericentric regions are formed by tandem repeats of DNA sequences, also known as "satellite" DNA. Beyond this generic appellation, the term "satellite" embraces different types of DNA repeats with different sequences. The repetitive units of centromeric (CT) and pericentromeric (PCT) regions of human chromosomes are given in Table 5.1.

Table 5.1 Sequence of human and mouse centromeric and pericentric repetitive units

Mouse
Centromeric – minor satellite 120 bp
From Wong and Rattner (1988)
GGAAAATGATAAAAACCACACTGTAGAACATATTAGATGAGTGAGTT ACACTGAAAAACACA
TTCGTTGGAAACGGGATTTGTAGAACAGTGTATATCAATGAGTTACAA TGAGAAACAT
Pericentric – major satellite (234 bp)
From Manuelidis (1982)
GGACCTGGAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATACACAC TTTAGGAC
GTGAAATATGGCGAGGAAAACTGAAAAAGGTGGAAAATTTAGAAATGTCCACTG TAGGACGTG
GAATATGGCAAGAAAACTGAAAATCATGGAAAATGAGAACATCCACTTGACGAC TTGAAAAAT
GACGAAATCACTAAAAAACGTGAAAAATGAGAAATGCACACTGAA
Human
Centromeric – alphoid (171 bp)
From Vissel and Choo (1987)
CTTCTGTCTAGTTTTTATATGAAGATATTCCCGTTTCCAACCAAGGCCTCAAAGCG GTCCAAATATC
CACAAGCTGATTCTACAAAAAGAGTGTTTCAAAACTGCTCTATGAAAAGGAAGGT TCAACTCTGTG
AGTTGAATGTATACATCACAAAGAAGTTTCTGAGAATG
Pericentric – satellite I, II, III
From Prosser et al. (1986)
Satellite 1: alternating arrays of A (17bp) and B (25 bp) motives
A: ACATAAAATAT(G/C)AAAGT
B: AC(AT/CC)CAAATATA(G/T)ATT(A/T)TAT(A/T)CTGT
or ACCCAA(AGT/GCC)AT(AT/GC)ATT(A/C)TATACTGT
Satellite 2: Poorly conserved 5 bp repeat GGAAT
Satellite 3: CAACCCGA(A/G)T$(GGAAT)_n$

5.2.1 Mouse Chromosomes

The repetitive units composing centromeric and pericentric regions are known as minor and major satellite (Sat) sequences, respectively, and represent 0.5% and 5% of the genome, respectively. Minor satellite units are 122 bp long. They probably result from a head-to-tail duplication of a 60 bp motif containing the 17 bp CENPB box (Wong and Rattner 1988). Different from what observed in the fission yeast *Schizoccharomyces pombe* (*S. pombe*), in mouse, all chromosomes are acrocentric with pericentric heterochromatin present only on q arms. For this reason, "pericentric heterochromatin" is also referred to as "juxtacentric heterochromatin." Major satellites are 234 bp long A/T-rich sequences that are present on all mouse chromosomes. In interphase nuclei, they often appear as large chromocenters.

5.2.2 *Human Chromosomes*

In humans, centromeric regions are acro-(chromosomes 13, 14, 15, 21 and 22), meta-, or submetracentric. Centromeres are composed of diverged alphoid AT-rich 171-bp motives repeated in head-to-tail fashion, to form a higher-order unit that, in turn, is reiterated several times to generate 500 kb to 1.5 Mb arrays (Vissel and Choo 1987). As in mouse, pericentric regions are positioned on long q chromosome arms, juxtaposed to the arrays of alphoid sequences. Important interchromosomal and interindividual differences exist concerning the size of pericentric regions. In chromosomes 1, 9, 16, and Y, pericentric regions are particularly large, of the order of several megabases. Pericentric sequences are composed of three different types of repetitive elements, called Sat 1, 2, and 3. Sat 1 sequences are AT rich and are formed by an alternance of 17 and 25 bp monomers (Prosser et al. 1986). Sat 1 arrays are restricted to chromosomes 3 and 4 and to acrocentric chromosomes (Meyne et al. 1994; Tagarro et al. 1994). Sat 2 and 3, which represent 2% and 1.5% of the genome, respectively, are more abundant than Sat 1 sequences (0.5%) (Jones et al. 1974; Mitchell et al. 1979; Jeanpierre 1994). They are formed by runs of the GGAAT motif (more divergent in the case of Sat 2), organized in a head-to-tail orientation (Mitchell et al. 1979; Frommer et al. 1982; Prosser et al. 1986). These repeats are interspersed with termination sequences, CATCATCGA(A/G)T in the case of Sat 2 array and CAACCCGA(A/G)T in Sat 3. Sat 2 sequences are abundant in chromosomes 1 and 16 but are present also in chromosomes 2, 7, 10, 15, 17, and 22, although to a lower extent. Sat 3 sequences are mainly found in pericentric regions in chromosomes 1, 5, 9, 10, 17, 20, and Y (Frommer et al. 1988; Tagarro et al. 1994).

5.3 Expression of Centromeric and Pericentric Sequences in Cell Lines and Tissues

A detailed analysis of the various contexts in which expression of centromeric and pericentromeric sequences occurs strongly suggests that transcription of these chromatin regions may have a role in normal and pathological cells. However, the complexity, the high level of polymorphism, and the repetitive nature of these genomic regions have so far hampered a detailed characterization of their transcriptional activity. Moreover, due to the high risk of contamination with genomic DNA, the sequences contained in gene expression databases are often expurgated of sequences of repetitive origin, hampering in silico analysis of expression profile. The different contexts in which transcription of CT and PCT regions have been described as well as the size and orientation of transcripts, when known, are given in Table 5.2.

Table 5.2 Different contexts in which expression of centromeric and pericentric sequences has been observed in the literature in human and mouse cells

		CT/PCT	H/M	Size and orientation	Orientation of RNA strands	
Normal and cancer cells and tissues						
Embryonic tissues	Whole embryo CNS	PCT	M	Smear	A-rich (11.5 dpc) T-rich (11.5–15.5 dpc)	Rudert and Gronemeyer (1993)
Extra embryonic tissue	Placenta	CT	H	N.D.	Both strands	Eymery et al. (2009b)
Adult tissues	Liver and testis	PCT	M	Smear 1–4 kb	N.D.	Rudert and Gronemeyer (1993)
	Testis	PCT	H	Smear 1 to >20 kb	C-rich	Jehan et al. (2007) and Eymery et al. (2009b)
Tumor tissues	Lung	PCT and CT	H	N.D.	Both strands	Eymery et al. (2009b)
Cancer cells	HeLa	PCT	H	N.D.	G-rich >>>C-rich	Valgardsdottir et al. (2007) and Eymery et al. (2009b)
Differentiation, senescence and aging, cell cycle						
Differentiation	Myoblast C2C12	PCT	M	N.D.	N.D.	Terranova et al. (2005)
	Idem	CT	M	Smear less than 1 kb to several kb in NDC 120 nt in DC	Both strands	Bouzinba-Segard et al. (2006)
	P19	PCT	M	1 kb (only in NDC)	N.D.	Rudert et al. (1995)
Senescence and aging	Cardiac muscle	PCT	M	Smear 6–12 kb	N.D.	Gaubatz and Cutler (1990)
	Progeria cells	PCT	H	N.D.		Shumaker et al. (2006)
Cell cycle	C127	PCT	M	G1 to mid S phase Smear less than 200 nt to more than 6 kb	N.D.	Lu and Gilbert (2007)
	NIH3T3/MEL	CT	M	Mitosis ≈200 nt Barely detectable in G1 to a distinct peak in G2/M	N.D.	Ferri et al. (2009)

(*continued*)

Table 5.2 (continued)

		CT/PCT	H/M	Size and orientation	Orientation of RNA strands	
Environmental stress, chromatin remodeling, and posttranscriptional control						
Environmental stress[a]	Seemingly ubiquitous in normal and cancer cells	PCT	H	More than 10 kb Smear less than 2–5 kb	G-rich>>>C-rich	Jolly et al. (2004) Rizzi et al. (2004) Valgardsdottir et al. (2007)
DICER KO	DT40 cell hybrids	CT and PCT	H	Long RNA 20 nt to several kb	Both strands	Fukagawa et al. (2004)
	ES	Idem	M	Smear with major signals at 25–30 nt and 150 nt in Dicer expressing cells	Resistant to Rnase ONE	Kanellopoulou et al. (2005)
SUV39 KO	ES	CT and PCT	M	N.D.	Resistant to RNAse ONE	Martens et al. (2005)
KDM2A KD	HeLa NIH3T3	PCT	H M	N.D.	N.D.	Frescas et al. (2008)
NP95 KD	NIH3T3	PCT	M	N.D.	N.D.	Papait et al. (2007)
Azacytidin	MEL	CT	M	120 nt	Both strands	Bouzinba-Segard et al. (2006)
	HeLa	PCT	H	N.D	Both strands	Eymery et al. (2009b)
TSA	ES	CT and PCT	M	N.D	N.D.	Kanellopoulou et al. (2005)
ZFPIP/Zfp462 KD	P19	PCT and CT	M	N.D	Both strands	Massé et al. (2010)

H Human, *M* Mouse, *DC* Differentiated cells, *NDC* Nondifferentiated cells, *N.D.* not determined
[a]Only data concerning the heat-stress response are presented

5.3.1 Development and Cell Differentiation

Several reports have described the accumulation of centromeric and PCT transcripts in the course of mouse development. In mouse embryos, PCT RNAs display a complex pattern of expression. They are detected by in situ approaches in the whole embryo 11.5–15.5 dpc (day postcoitum), in the central nervous system (CNS) at 12.5 dpc and in scattered cells from the CNS at 15.5 dpc, while in adult tissues, they have been detected in liver and testis (Rudert et al. 1995). Testis-specific expression of PCT sequences has been reported also in human (Jehan et al. 2007; Eymery et al. 2009b), suggesting a functional role of these RNAs in spermatogenesis. Moreover, in human, a sporadic expression of CT sequences is also detected in placenta, ovary, and liver tissues (Eymery et al. 2009b). Interestingly, in a few cases, both PCT strands are sequentially expressed, leading to the production of A- and T-rich transcripts, presumably with different functions. For example, in the mouse brain, the expression of T-rich PCT sequences precedes the accumulation of A-rich transcripts. On the contrary, only A-rich and T-rich PCT transcripts are expressed in mouse and human testis, respectively (Rudert et al. 1995), indicating that the transcriptional activity of these chromosomal regions is differentially regulated, in a cell type and in a differentiation specific manner. In vitro analysis further supports the hypothesis that the level of CT and PCT transcripts is modulated during cell differentiation. Thus, accumulation of CT (Bouzinba-Segard et al. 2006) and of PCT (Terranova et al. 2005) RNAs occurs upon terminal differentiation of C2C12 myoblasts and upon DMSO-induced erythroid differentiation of mouse erythroleukemic MEL cells (Bouzinba-Segard et al. 2006). These experimental systems may offer the opportunity to explore the underlying regulatory mechanisms.

5.3.2 Response to Environmental Stimuli

In the last 10 years, a number of studies have shown that heat shock induces the transcriptional activation of a particular PCT region of the human genome, namely the 9q12 band mainly composed of long arrays of Sat 3 with the general formula (GGAAT)n CAAC(C/A)CGAGT with $n > 1$. This activation depends on the activity of heat shock factor 1 (HSF1) that binds to the Sat 3 sequence and drives the production of long noncoding Sat 3 RNAs corresponding to the G-rich strand of the repeat (Jolly et al. 2004; Rizzi et al. 2004). Although initial Northern blot analysis revealed Sat 3 RNAs only after thermal stress, more sensitive quantitative RT-PCR has recently detected a basal expression of Sat 3 sequences even in unstressed cells (Valgardsdottir et al. 2007). The level of Sat 3 RNAs drastically increases during heat shock and during the first 3 h of recovery from heat shock and is still higher than in unstressed cells 1 day later. In situ hybridization proved that these ncRNAs are exclusively nuclear and remain in close proximity of Sat 3 DNA

arrays, giving rise to nuclear Stress Bodies (reviewed in Biamonti 2004; Biamonti and Vourc'h in press). Within nSBs, ncRNAs recruit a number of RNA binding proteins involved in pre-mRNA processing, including splicing regulators SF2/ASF, SRp30c, and Sam68 (Denegri et al. 2001; Chiodi et al. 2004; Metz et al. 2004). Notably, transcriptional activation of Sat 3 sequences is induced by a large number of stressing agents other than heat shock, such as DNA damaging agents (methyl methane sulfonate – MMS), inhibitors of DNA replication (etoposide and aphidicolin), heavy metals (cadmium), H2O2, UV-light, 8-hydroxyquinoline (8HQ), zinc sulfate (ZnSO4), ibuprofen, proteasomal inhibitors (MG132, lactacystin), protein biosynthesis inhibitors (puromycin), and hyperosmotic stress. All these treatments induce both Sat 3 RNAs and the formation nSBs (Valgardsdottir et al. 2007; Sengupta et al. 2009). However, the extent of induction and the number of nSBs detectable in the cells depend on the nature of the stressing agent, on the severity of the stress treatment, and on the cell type. Contrary to most of stressing agents that act through HSF1, transcription of Sat 3 sequences in response to hyperosmotic stress depends on TonEBP (Tonicity Enhancer Binding Protein) that controls the expression of genes involved in the ability of the cells to survive high osmotic pressure as, for instance, in the kidney (Valgardsdottir et al. 2007). Thus, it appears that induction of Sat 3 RNAs and the formation of nSBs may be part of a general cellular response to stress through the activity of at least two independent pathways identified by HSF1 and TonEBP. However, the function of this event is still to be clarified, particularly in view of its restriction to primates that contain Sat 3 sequences in their genome. Indeed, there is no evidence so far that Sat 3 RNAs confers any advantage to expressing cells, in terms of ability to cope with stressing conditions. Probably, such an advantage has to be searched in cells and tissues that express Sat 3 sequences under physiological conditions, as in testis (see ahead). Finally, it is worth noting that heat shock triggers transcription of other human PCT regions on chromosomes other than 9, even though the extent of induction is more limited (Rizzi et al. 2004; Eymery et al. 2010). The mechanism underlying the specificity of HSF1 binding to chromosome 9 is still unclear and deserves further investigation. Secondary HSF1 binding sites are also present within PCT regions enriched in both Sat 2 and Sat 3 repeats. Formation of secondary nSBs is increased by upregulation of HSF1 expression (Eymery et al. 2010).

In vivo, an accumulation of PCT transcripts has only been reported in the heart of aging mouse, possibly associated to mitochondria-induced oxidative stress (Gaubatz and Cutler 1990). In human, nothing is presently known about the conditions allowing a stress-induced expression of PTC sequences in vivo in normal and tumor cells in response to heat shock, hypoxia, or inflammation.

5.3.3 Regulation During the Cell Cycle

The stress-induced expression of PCT sequences occurs both in dividing and nondividing cells. This does not imply that the expression of these chromatin

regions is completely independent of the cell cycle. Indeed, formation of nSBs, which is a marker of Sat 3 DNA transcription, occurs very rapidly in late S-phase while is delayed in early G1, as if the accessibility of these sequences to transcription can change during the cell cycle. It is still unknown whether this reflects a different epigenetic organization of the Sat 3 arrays (Weighardt et al. 1999).

Contrary to stress-induced transcription of PCT sequences, physiological expression of CT and PCT and sequences appears to be connected with cell growth and cell cycle. Moreover, in mouse, both CT and PCT sequences display a precise pattern of expression during the cell cycle, suggesting a link with the higher-order organization of these chromatin regions.

PCT sequences are expressed throughout G1, reaching a peak at the G1/S transition. Accumulation of PCT sequences requires the passage through the G1/S restriction point and Cdk activation. Accumulation of PCT RNAs is then substantially reduced after replication of these heterochromatic regions. A second peak of accumulation is observed as the cells enter mitosis, followed by transcriptional repression at the metaphase–anaphase transition (Lu and Gilbert et al. 2007) Thus, constitutive and stress-induced transcription of PCT sequences appears to follow complementary kinetics during the cell cycle, with stress-induced expression occurring in late S phase and constitutive expression in early S and M phases.

Differences also exist between the expression profiles of CT and PCT sequences. CT transcripts are barely detectable in G1, and their expression increases in S phase to remain relatively constant in late S phase. A comparative analysis of the expression profile of CT and PCT sequences during S phase progression in the same cells is still missing. However, one can speculate that RNAs expressed in early S phase could play a role in DNA replication, while RNAs that are expressed later could be involved in the propagation and/or stabilization of heterochromatic epigenetic marks.

Interestingly, CT and PCT RNAs also accumulate in mitosis (Bouzinba-Segard et al. 2006; Lu and Gilbert 2007) in parallel with the dissociation of HP1 from PCT regions. The different size of PCT transcripts in S and M phases could suggest that they fulfill different roles. Moreover, since no accumulation of PCT transcripts is observed in *S. pombe*, a functional explanation should come from a precise comparative analysis between mitotic processes in fission yeast and mammals.

At first sight, expression of CT and PCT sequences in dividing cells may appear paradoxal with regard to the observation that the expression of these sequences increases during cell differentiation (Terranova et al. 2005; Bouzinba-Segard et al. 2006), senescence (Enukashvily et al. 2007), or aging (Goldman et al. 2004), namely conditions associated with cell cycle slow-down or withdrawal. Although the possibility exists that accumulation of CT and PCT RNAs could result from read-through transcription facilitated by a loss of heterochromatic epigenetic marks, we favor an alternative model whereby accumulation of CT and/or PCT transcripts occurring during cell differentiation represents an active mechanism to modulate the organization of heterochromatic regions. In this perspective, CT and PTC transcripts could fulfill common functions in senescent cells, aging cells, and stressed cells, known to be associated with major heterochromatin remodeling events.

5.4 Molecular Mechanisms Underlying Expression of CT and PCT Sequences

The molecular mechanisms underlying the constitutive expression of CT and PCT sequences are still largely unexplored. In particular, nothing is known about the promoter regions controlling the expression of these sequences. In mouse and human cells, only four transcription factors have been formally characterized for their role in the transcriptional activation of PCT sequences in response to heat shock, osmotic-pressure, and steroid-treatments. Evidence, however, exist that the expression of CT and PCT repeats is also epigenetically regulated, involving changes of DNA methylation and posttranslational histone modifications.

5.4.1 Transcription Factors

In mouse and human cells, the expression of major satellite sequences is driven by RNA pol II (Jolly et al. 2004; Rizzi et al. 2004; Lu and Gilbert 2007), and RNA pol II inhibitors strongly decrease the level of PCT RNAs.

The first transcription factor shown to control the expression of PCT sequences was the retinoic acid receptor (RAR) (Rudert and Gronemeyer 1993). Interestingly, the expression pattern of PCT sequences is either up or downregulated upon cell differentiation with retinoic acid (RA) depending on the considered cell type. Thus, RA treatment downregulates PTC expression in P19 embryonic carcinoma cells (Rudert et al. 1995), while upregulation is observed in embryonic stem (ES) (Martens et al. 2005). This suggests that distinct cofactors are probably present in different cell types to define the activating or repressive activity of RAR (Rudert et al. 1995).

Three transcription factors have been identified so far that control the expression of PTC Sat 3 sequences in human cells in response to stress. Two are members of the family of Heat Shock Factors, namely HSF1 and HSF2 (reviewed by Morimoto 1998), involved in the developmental and heat-induced regulation of heat shock genes. The third one is TonEBP, which is activated in response to high concentrations of osmolytes, such as sorbitol. Upon suitable stress treatments, these factors relocate to nSBs (see Sect. 5.2.2) primarily assembled on the PCT q12 region of human chromosome 9 that is mainly composed of long tandem arrays of Sat3 sequences. The assembly of nSBs requires the DNA Binding Domain of HSF1 (Jolly et al. 2002), and direct binding of HSF1 to the 9q12 locus has been demonstrated by in vitro recapitulation of the first step of nSBs formation using human chromosome spreads and purified HSF1 (Jolly 2002). However, given the limited resolution of in situ approaches combined with our scanty knowledge of the organization of PCT regions, particularly in the human genome, the binding sites of HSF1 and TonEBP within these regions are still to be defined. Canonical HSF1 binding sites consist of multiple adjacent and inverse iterations of the

pentanucleotide motif 5′-nGAAn-3′ (Fernandes et al. 1994). In particular, three such motifs form the 5′-nGAAnnTTCnnGAAn-3′ Heat Shock Element (HSE) in the promoters of hsp genes. Canonical HSF1 binding sites are not present in the prototypical Sat III element in pHuR98 plasmid (accession number GenBank: X06137.1). However, in an in vitro EMSA assay, this sequence is specifically recognized and bound by HSF1. Even less is known in the case of TonEBP, although in silico analysis of the sat III sequence picked up motifs matching the consensus binding site (TGGAAANN(C/T)N(C/T)) of this factor (Valgardsdottir et al. 2008).

Unlike HSF1, HSF2 does not possess an intrinsic DNA binding capacity to PCT sequences (Alastalo et al. 2003), and its presence within nSBs requires heterotrimerization with HSF1. HSF2 knockdown does not alter the stress-induced relocalization of HSF1 to nSBs but increases the expression of PCT sequences in heat-shocked cells, indicating that the extent of stress-induced activation of PCT sequences is regulated through HSF1–HSF2 heterotrimerization (Sandqvist et al. 2009). Interestingly, overexpression of HSF2 in unstressed cells leads to the constitutive expression of PCT sequences in HeLa cells. Consistent with this, PTC sequences are constitutively transcribed in testis, where HSF2 is abundantly expressed and HSF1–HSF2 heterotrimers are present (Sandqvist et al. 2009).

5.4.2 Epigenetic Regulation

PCT regions are enriched in methylated CpG and trimethylation of histones H3 on K9 and H4 on K20, two hallmarks of transcriptional repression. In the last few years, the epigenetic status of PCT and CT sequences and the expression level of enzymes controlling DNA and histone methylation have been under investigation through siRNA-mediated downregulation of specific DNA methyl transferases (DNMTs) and histone methyl transferases (HMTs) and by exploiting drugs with a broad impact on the epigenome.

5.4.2.1 DNA Methylation

In mouse and human cells, two DNMTs, DNMT1 and DNMT3B, are involved in maintenance and de novo methylation of DNA, respectively. Loss of Dnmt1 and/or Dnmt3b genes causes severe hypomethylation of pericentric heterochromatin. Methylated cytosines represent docking sites for both Histone Deacetylases (HDACs) (Nan et al. 1998a) and Histone Methyl Transferases (HMTs) (Fuks et al. 2003) through their binding to the methyl CpG-binding protein (MECP2). Thus, drugs or pathological conditions that lead to DNA hypomethylation could favor transcriptional derepression of PCT sequences (reviewed in Nan et al. 1998b).

In mouse erythroleukemic (Bouzinba-Segard et al. 2006) and human HeLa cells (Eymery et al. 2009b), increased expression of CT (mouse) and PCT (human) sequences is indeed observed upon azacytidin treatment. Moreover, a constitutive expression of human PCT sequences at the 1q12 locus occurs in A431 epithelial carcinoma cells and in senescent embryonic lung MRC5 cells, whose genomes are globally highly hypomethylated (Enukashvily et al. 2007). However, no clear evidence exists so far that demethylation has a causal effect on the increased expression of CT and PCT sequences. Indeed, no increased expression of CT and PCT sequences is detectable in mouse ES cells (Lehnertz et al. 2003; Martens et al. 2005) and human cells (Eymery et al. 2009b) deficient for DNMT3b and/or DNMT1, despite a lower level of DNA methylation. Similarly, absence of strong constitutive expression of PCT sequences is also observed in cells from patients with ICF syndrome (Immunodeficiency, Centromeric instability, and Facial anomalies), a pathology characterized by a severe demethylation of PCT regions (Ehrlich 2003). These somehow contradictory impacts of drug- and genetically induced DNA demethylation on expression of CT and PCT sequences clearly suggest the existence of compensatory mechanisms to maintain a transcriptional repression of these sequences in spite of epigenetic reprogramming (Eymery et al. 2009a, b). Alternatively, in addition to epigenetic status, still unidentified transcription factors may be involved in the activation of PTC sequences.

At last, while increased expression of PCT and CT sequences is observed in azacytidin treated cells, accumulation of PCT specific transcripts in heat-shocked cells is not accompanied by DNA demethylation. Thus, demethylation of PCT sequences is not a prerequisite for transcriptional activation, and the expression of these sequences may follow different signaling pathways in response to heat shock and DNA hypomethylation (Eymery et al. 2009a, b).

5.4.2.2 Histone Modifications

A large body of data in the last decade unveiled a connection between epigenetic marks and transcriptional activity. In particular, silent portions of the genome are usually associated with H3-K9me3, H4K20me3 (enriched in constitutive heterochromatin), and H3K27me3 (enriched in facultative heterochromatin). The level of heterochromatic marks depends on the activity of major enzymatic systems transducing H3K9 trimethylation (SUV39H1, SUV39H2), H4K20 trimethylation (SUV420H1, SUV420H2), and H3K27 trimethylation (EZH1, EZH2). The level of these marks and enzymes has been analyzed in different cellular contexts relative to the expression of CT and PCT sequences. Likewise, the impact of downregulation of these enzymes on the expression of CT and PCT sequences has also been evaluated.

A strong evidence that loss of repressive epigenetic histone marks could facilitate transcription of PTC sequences derived from the analysis of primary fibroblasts from patients with Hutchinson Gilford Progeria Syndrome (HGPS). HGPS is caused by constitutive expression of a truncated form of pre-lamin A, whose accumulation

produces aberrant nuclear shape, reduced resistance to mechanical stress, and defects in heterochromatin-specific histone modifications. Moreover, HPGS fibroblasts are characterized by a reduction of Heterochromatin Protein 1 α (HP1α), by a concomitant reduction or complete loss of H3K9me3 (reviewed in Misteli and Scaffidi 2005) and by a loss of H3K27me3 on the inactive X chromosome, which is partially compensated by the increase of H4K20me3 (Shumaker et al. 2006). HPGS also shows downregulation of the methyltransferases EZH2 and SUV39H1/2. This profound reorganization of the epigenome is accompanied by the upregulation of PCT RNAs (Shumaker et al. 2006).

The status of CT and PCT sequences expression with regard to histone methyltransferases (HMTs) has been also evaluated in embryonic fibroblasts (MEF), trophoblasts stem (TS) cells, and in mouse embryonic Stem cells treated or not with retinoic acids (RA). All of the major HMTases genes are broadly expressed in ES and MEFs cells, whereas the expression of Suv39H1/2, Glp1, Eset, and Suv4-20h2 is downregulated in TS cells, which express higher levels of both PCT and CT transcripts (Martens et al. 2005). This pattern clearly suggests that transcription of PCT and CT regions may be facilitated by reduction of H3K9me3 and H4K20me3. In addition, a higher level of CT and PCT RNAs is observed in suv39h double knockout ES cells, suggesting an inverse correlation between the expression level of SUV39H and H3K9me3, and that of CT and PCT sequences (Martens et al. 2005).

However, loss of heterochromatic marks and expression of CT and PCT sequences is not always associated with downregulation of the corresponding HMTs, revealing the complexity of the control mechanisms underlying histone demethylation and derepression of CT and PCT sequences. In mouse, for example, the expression of PCT sequences, which occurs in S phase, precedes replication of pericentric heterochromatin and is actually downregulated during heterochromatin replication in mid/late S-phase. This behavior represents a strong argument against the hypothesis that transcription is facilitated by a replication-dependent dilution of H3K9me3. On the other hand, in mouse, cell-cycle-dependent transcription of PCT sequences is not affected in suv39 mutants, indicating that this phenomenon is independent of H3K9me3 and HP1 (Lu and Gilbert 2007 and reviewed in Lu and Gilbert 2008). In human lung tumors, for example, expression of CT and PCT sequences correlates with a global loss of H3K27me3 without the downregulation of Ezh1 and Ezh2 genes (Eymery et al. 2009b). Moreover, terminal muscle differentiation of mouse C2C12 cells induced by RA is accompanied not only by increased level of histones H3K9 and H4K20 trimethylation across PCT regions but also, paradoxically, by increased levels of CT and PCT transcripts (Terranova et al. 2005), once again illustrating the complexity of the regulatory circuits involved. At last, since histone methylation on specific lysines is associated with gene activation rather than inactivation, depletion of specific histone demethylases may promote gene activity. This is the case of KDM2A, a Heterochromatin Protein 1 (HP1)-interacting protein that promotes HP1 localization to chromatin and demethylates H3K36me2. H3K36me2 is associated with positive regulation of pol II-mediated transcription and is negatively regulated by KDM2A. In mouse

NIH3T3 cells, knock down of KDM2A is associated with increased expression of PCT sequences, while in human HeLa cells, an expression of alphoid sequences occurs (Frescas et al. 2008). The reason why CT and PCT sequences are differentially targeted in mouse and human cells remains to be determined. However, this represents a new illustration of the impact that epigenetic organization may exert on the expression of CT and/or PCT sequences.

Histone acetylation, which is associated with gene activation, also appears to be involved in the transcription of PTC sequences. Indeed, in heat-shocked cells transcription of Sat 3 sequences is accompanied by the acetylation of histones in nSBs. Although the implication of specific HATs in the transactivation process has not yet been demonstrated, the presence of the histone acetyl transferase CBP in nSBs supports the notion that core histone de novo acetylation at the 9q12 locus has a causal role in this process (Jolly et al. 2004). The global level of histone acetylation is controlled by the balance between histone acetyl transferases (HATs) and histone deacetylases (HDACs). Therefore, treatment with Trichostatin A (TSA), a potent inhibitor of classes I and II HDACs, increases global histone acetylation. In mouse ES cells, TSA treatment results in a pronounced increase of both PCT and CT sequences' expression (Kanellopoulou et al. 2005). A transcriptional activation of mouse PCT sequences, and not of CT sequences, is also observed in mouse NIH3T3 cells depleted of Np95, a cell-cycle-regulated nuclear histone-binding protein that recruits HDAC-1 to target promoters. NP95 ablation causes a strong reduction in pericentric heterochromatin and is associated with hyperacetylation of histone H4 (Papait et al. 2007).

In contrast to data suggesting that increased acetylation of PCT or CT sequences leads to their transcriptional activation, no effect on the basal level of PCT or CT RNAs is observed in human HeLa cells upon treatment with TSA or Butyrate (inhibitor of class I and class II HDAC, but HDCA6). Again, these contrasting results suggest that, depending on the origin of the cells, embryonic or differentiated, global changes in histone acetylation levels are not necessarily sufficient to trigger the activation of satellite sequences and that acetylation of PCT regions must be combined to other transcriptional regulatory mechanisms. More surprising is the fact that in HeLa cells, TSA treatment also prevents the formation of HSF1 foci and the expression of PCT sequences, in response to heat shock (Rizzi et al. 2004). The reason for this inhibitory effect remains to be clarified. From these different observations, it is clear that multiple signaling pathways impacting CT and/or PCT chromatin structure may lead to transcriptional activation of these sequences. This is, for example, the case of ZFPIP/Zfp462, a zinc finger nuclear factor necessary for correct cell division during early embryonic developmental steps of vertebrates. Recently, ZFPIP/Zfp462 has been shown to play a role in chromatin integrity and survival of mouse P19 pluripotent cells. It has been hypothesized that ZFPIP/Zfp462 acts as a platform for other factors such as Pbx1, Meis, or Prep proteins involved in pericentric chromatin assembly in P19 cells. Cells deficient for this factor exhibit a complete destructuration of pericentromeric domains, associated with a redistribution of the HP1α proteins and with increased expression of CT and PCT sequences (Massé et al. 2010).

5.5 Role of Pericentric and Centromeric Functions

In *S. pombe*, transcripts of pericentric origin, generated by the RNAi machinery, play a role in the establishment and maintenance of chromatin organization in PCT regions (Reviewed in Verdel et al. 2009; Grewal and Elgin 2007). In this chapter, we give a short overview of the role of mouse and human satellite transcripts in pericentromeric and centromeric structure and function, both of which are epigenetically defined. We will also discuss the role of CT and PCT RNAs in the control of gene expression both at a global nuclear and more local chromosomal level. Heterochromatic PCT regions represent centers of repressive chromatin and in *Drosophila* are associated with the phenomenon of position effect variegation (PEV) that downregulates the expression of genes through juxtaposition with heterochromatin. Moreover, chromatin remodeling events associated with the transcriptional activation of constitutive heterochromatic regions, as during the stress-induced formation of nSBs, are likely to impact on the functional organization of the cell nucleus. Indeed, transient trapping of specific transcription factors and chromatin remodeling activities in nSBs could contribute to shutdown or global reprogramming gene expression. Similarly, transient sequestration of specific RNA binding proteins in nSBs may affect splicing decisions in other nuclear districts and/or orientate the splicing profile of genes relevant for the cell response to stress. Finally, a recent report involves CT and PCT transcripts in trans-splicing events, indicating the ability of these ncRNAs to control gene expression at a posttranscriptional level.

5.5.1 Structural Components of Pericentric and Centromeric Structure and Function

In *S. pombe*, 20–30 nt transcripts of PCT origin are generated by the RNAi machinery and targeted to PCT regions. Two RNAi complexes, the RNA-Induced Transcriptional Silencing complex (RITS), which contains a siRNA bound to an Argonaute protein, and the RNA-Directed RNA polymerase Complex (RDRC), are critical components to the deposition of H3K9me and heterochromatin marks. These small RNAs play a role in the establishment and maintenance of chromatin organization in PCT regions. RNA molecules appear to be essential constituents of heterochromatin in higher eukaryotes as well (Muchardt et al. 2002; Maison et al. 2002). However, the nature of these RNAs and the mechanisms through which they are targeted to the pericentric regions is still unclear. Homologs of Chp1 and Tas 3, two components of the RITS complex, do not exist in mammals, and small RNAs (21 nt-long) have not been formally implicated as essential actors in heterochromatin structure.

In mouse, 200 nt-long PCT transcripts accumulate throughout G1 phase and colocalize with early replicating DNA at the G1/S transition (Lu and Gilbert 2007). After mid-S-phase, coincident with the time of chromocenter replication, their level starts to decrease, as recently described in *S. pombe* (Chen et al. 2008). The general

pattern of expression of PCT sequences during the cell cycle suggests that these transcripts may assist the reassembly of heterochromatin after replication (Lu and Gilbert 2007). Finally, residual PTC RNAs in late mitosis could stabilize heterochromatin, after cohesin removal and, as in S phase, assist the reassembly of heterochromatin (Lu and Gilbert 2007). As mentioned earlier, a structural role in heterochromatin complexes is suggested by the observation that at least PCT RNAs generated in response to stress remain in close association with nSBs assembly sites, even when HSF1 and RNA pol II are no longer detected at this foci (Jolly et al. 2004). In stressed cells, Sat III transcripts, which are more stable (longer half-life) than PCT RNAs physiologically expressed in S phase, could be necessary to the reformation or stabilization of pericentric heterochromatin following heat shock.

By analogy to what is described in *S. pombe*, the involvement of Dicer in pericentric structure and function would represent a strong argument in favor of a role for PCT transcripts in heterochromatin structure in mammals. In chicken–human somatic hybrid cell lines containing only human chromosome 21, small transcripts of CT and PCT origin are indeed detected. Interestingly, upon downregulation of Dicer, long CT and PCT transcripts, ranging in size from 20 nt to several kb, start to accumulate, suggesting an evolutionarily conserved role of the gene-silencing pathway in controlling the expression of CT and PCT sequences (Fukagawa et al. 2004). Notably, in chicken cells, Dicer-deficiency is accompanied by mitotic defects due mainly to premature sister chromatid separation rather than from alterations of centromeres, as indicated by the unperturbed distribution of centromeric proteins, CENPA and CENPC (Fukagawa et al. 2004).

In mammals, a role of RNAi machinery in the posttranscriptional maturation of CT and PCT sequences could be restricted to undifferentiated cells. In mouse ES cells, small dsRNA from CT and PCT origin are detected. As in chicken, they accumulate as long dsRNA in the absence of Dicer. These observations suggest that CT and PCT sequences are initially transcribed as long stable transcripts, which are further processed by the RNAi machinery (Kanellopoulou et al. 2005). Small dsRNA of CT origin have not been detected in mouse-differentiated cells (Segard et al. 2006). Likewise, in human HeLa cells, no accumulation of CT and PCT transcripts is observed upon knock down of Dicer (Eymery et al. 2009b). However, the presence of unstable or low abundance 20–30 nt RNAs in these cells, originating from CT or PCT regions, cannot be presently ruled out. In support of this possibility, an inverse correlation between the level of Dicer and PCT RNAs has been reported in differentiating myogenic mouse cells (Terranova et al. 2005). Even though prevailing models involve short RNAs, long RNA molecules may also play a role in the establishment of higher order heterochromatin organization. The best example is provided by the long Xist RNA that controls X chromosome inactivation in mammals (reviewed in Masui and Heard 2006). One can hypothesize that long CT or PCT RNAs are directly involved in heterochromatin organization and that processing by Dicer would be important to momentarily dissociate these RNAs from chromatin.

The fact that CT and PCT transcripts do not accumulate in human HeLa cells after downregulation of Dicer 1 and 2 activities (Eymery et al. 2009b) clearly

suggests that, at least in these cells, the level of these molecules is mainly determined at the transcriptional level. It is worth noticing however that, at least in human HeLa cells, PCT transcripts could represent a target, rather than a source, of small RNAs since the two strands are transcribed at a very different level, with C-rich transcripts being almost undetectable (Valgardsdottir et al. 2007). Indeed, in human cells, contrary to yeast, no RNA-dependent RNA polymerase (RdRP) enzyme exists for dsRNA synthesis from single-strand transcripts. Further analysis of both undifferentiated and differentiated cells is needed to delineate the exact cellular context involving the RNAi machinery, as well as its role in controlling the expression of CT and PCT sequences in mammalian cells.

No evidence exists so far that implicates the RNAi machinery in the formation of the kinetochore complex. Knock down of Dicer affects sister chromatid cohesion but does not impact on CENPC and INCENP distribution on either interphase or mitotic centromeres (Fukagawa et al. 2004). In contrast, several publications raise the exciting possibility that long single-strand RNAs could participate in the epigenetic process that ensures centromeric stability and inheritance. In maize, transcripts of CT origin have been found to promote DNA binding of CENP-C, a protein of the inner kinetochore, which has a key role in centromere recognition and maintenance (Du et al. 2010). It has been hypothesized that CENP-C would be first recruited to kinetochore through protein/protein interaction and that DNA binding of CENP-C would in turn be facilitated by the presence of centromeric RNA (Du et al. 2010). In human, single-strand RNA of CT origin and the RNA binding domain of CENP-C have also been identified for their role in CENP-C targeting to centromeres (Wong et al. 2007).

During mitosis, the large multifunctional kinetochore complex ensures the connection of chromosomes to microtubules and regulates the timing of anaphase (reviewed in Allshire and Karpen 2008; Ugarković 2009). This process also involves the chromosomal passenger complex (CPC) composed of Aurora B kinase and its regulatory subunits Inner centromere protein (INCENP), Survinin and Borealin. Both CENP-C and the proteins of the CPC, INCENP and Survinin, are dissociated from human mitotic chromosomes treated with single-strand RNA-specific RNAse. Conversely, in the presence of RNAse inhibitors, CT RNA is capable of partially restoring the relocalization of CENPC and INCENP, in the reconstitution assays mentioned above (Wong et al. 2007). Recent evidence also suggests that single-stranded CT RNAs, which accumulates in G2/M in mouse cells, are necessary to the formation of functional kinetochores during mitosis. Indeed, these transcripts are associated with proteins of the CPC complex in the G2/M phase of the cell cycle (Ferri et al. 2009). The assembly of Aurora B/Survivin complex and the enzymatic activity of Aurora B kinase are both enhanced by the presence of CT RNA (Ferri et al. 2009). The implication of CT transcripts in mitosis is further supported by experiments where enforced expression of single-strand CT transcripts leads to increasing number of anomalies in mitotic cells including misaligned chromosomes and abnormal chromosome numbers. This is probably a consequence of the sequestration of important components of centromeres, by ectopic CT RNA (Bouzinba-Segard et al. 2007).

5.5.2 Position Effect

In addition to storing genetic information, chromosomes have a crucial role in organizing the nuclear functions, and chromosomal territories define active/repressed nuclear domains to orchestrate gene expression. Large constitutive heterochromatic blocks, usually found at PCT regions and formed by tandem arrays of repetitive DNA, specify nuclear domains that exert negative effects on gene expression. A large body of data supports a model whereby the recruitment of a gene within such a domain may be part of a mechanism aimed at preventing its expression (reviewed in Fisher and Merkenschlager 2002). Due to the complex architectural organization of the nucleus, to be embedded in such repressive domains, genes do not need to be physically adjacent to heterochromatic blocks and may be actually located on distinct chromosomes. Thus, one can predict that any epigenetic reorganization of heterochromatic PCT regions may impact on gene expression of specific sets of genes.

In yeast and Drosophila, for which position effects have been best described, the mechanisms underlying gene repression involve spreading of repressive epigenetic marks (reviewed in Talbert and Henikoff 2006). In human T lymphocytes, similar mechanisms have recently been described as a result of chromosome translocation events involving the PCT regions of chromosome 1 and 2 (Fournier et al. 2010). In these cells, sequences in the 2p region and adjacent to the translocation site are characterized by increased levels of repressive histone modifications, including H4K20me3 and H3K9me3. This event is accompanied by the transcriptional repression of specific genes and by the repositioning of these chromosomal regions at the periphery of the nucleus (Fournier et al. 2010). An opportunity to verify the effect of pericentric heterochromatin on gene expression may be offered by the drastic epigenetic reprogramming of the 9q12 domain elicited by thermal stress. It is tempting to speculate that transcriptional activation of PCT sequences on 9q12 could favor the formation of an open chromatin conformation on nearby genes. However, no significant difference in the transcriptional activity of the genes located in the vicinity of the 9q12 region has been detected between unstressed and heat-shocked cells (Eymery et al. 2010). It is possible that position effects may occur under different conditions or impact the expression of genomic regions through "*trans*" acting mechanisms involving repositioning in nuclear districts close to PTC sequences. The major obstacle to the validation of this hypothesis remains the identification of genes associated with the 9q12 domain, in both unstressed and stressed cells.

5.5.3 Sequestration of Transcription and Splicing Factors

An alternative model, not mutually exclusive with the "position effect" model, predicts that gene expression programs may be directly influenced by the transcriptional activation of the PCT regions in 9q12. In heat-shocked cells, for example, the

massive recruitment to nSBs of factors involved in transcription, chromatin organization, and pre-mRNA processing may play a global negative effect on gene expression by sequestering transcription factors, histone acetylases, and at the same time, orientating splicing decisions through sequestration of pre-mRNA processing factors (reviewed in Jolly and Lakhotia 2006). In support of this possibility, a global reduction of euchromatic marks occurs after stress. Although this is likely to occur independently of nSBs, the interaction of histone acetylases with Sat3 sequences may certainly contribute to this phenomenon. Interestingly, whereas HSF1 is very rapidly displaced from Sat3 arrays during the recovery from stress, other proteins and enzymes remain associated with nSBs for longer intervals. This is, for instance, the case of the histone acetylase CREB after puromycin treatment (Sengupta et al. 2009) and of pre-mRNA processing factors (Weighardt et al. 1999). Moreover, Sat3 RNAs are relatively stable molecules and remain associated with the 9q12 region more than 1 day. It is, therefore, conceivable that the interaction of splicing regulators with these ncRNAs may somehow influence the splicing profile of genes relevant for recovery from stress.

5.5.4 Stabilization of Specific Protein Encoding Transcripts

The tissue-specific expression of CT and PCT sequences along with the preferential transcription of only one strand strongly argue in favor of a role of these RNAs in the tissue-specific control of gene expression. A striking illustration of this possibility is the discovery that a PCT RNA encoded by chromosome Y controls the stability of a protein-coding testis-specific transcript. This ncRNA trans-splices with CDC2L2 mRNA from chromosome 1p36.3 locus to generate a testis-specific chimeric beta sv13 isoform that contains a 67-nt 5′UTR provided by a PCT transcript. Within the 5′UTR, a 5′-CCAAT-3′ motif is present that may control translation of the β sv13 isoform in testis (Jehan et al. 2007). It is worth noticing that this is the example of trans-splicing between transcripts encoded by Y and autosomal chromosomes.

5.6 Conclusion

In the last 10 years, CT and PCT transcripts have been implicated in a large variety of cellular functions such as the transmission of epigenetic information, cell differentiation, and the cell defense to stress. However, in most cases, the exact function of CT and PCT transcripts and the molecular mechanisms underlying their expression remain elusive.

There are several open questions that need to be addressed in the near future. First, it is necessary to define the conditions underlying the expression of CT and PCT chromosomal regions in different organisms and in different growth conditions or differentiation states. For example, so far, stress-induced expression of PCT

sequences has been reported only in human cells. It would be nice to understand whether or not a similar process occurs in other species, despite differences in the primary sequence of PCT Repeats.

It is plausible that, at least in some cases, a loss of repressive epigenetic marks may favor unscheduled read-through transcription. Thus, the analysis of the conditions that lead to the expression of CT and PCT sequences could help the identification of new actors involved in remodeling of PCT and CT chromatin regions. From this viewpoint, CT and PCT transcripts could be regarded as molecular markers of extensive epigenetic remodeling events occurring during cell differentiation or cell proliferation or under pathological conditions. Intriguingly, the kinetics of expression, the RNA size, and the sense of transcription of CT and PCT sequences strongly suggest the existence of dedicated control mechanisms. A major goal will be the identification of the regulatory regions and of the cognate factors controlling the expression of CT and PCT sequences. Third, there is a clear need to characterize the transcripts that accumulate in the different cellular contexts as, in many cases, not even the sense of transcription is known.

Finally, the functional implication of CT and PCT RNAs is still largely unknown.

According to prevailing models, PCT RNAs would be more involved in determining the structure of pericentric chromatin, while transcripts of CT origin could control the organization of centromeric regions, as defined by the presence of CENPA. However, it is easy to anticipate that the situation is more complex than what predicted by this simplistic model. In mouse, for example, enforced expression of CT transcripts results in a dramatic redistribution of both Aurora-B and HP1 localization, an epigenetic mark associated with PCT heterochromatin (Bouzinba-Segard et al. 2006). Clarifying the respective role of CT and PCT transcripts should help to better define the role of their respective encoding regions.

The next decade should bring important clues about the structural and functional characterization of CT and PCT regions and transcripts. No doubt that these studies will bring important discoveries concerning the role of CT and PCT RNAs in the epigenetic control of gene expression and in the transmission of epigenetic information through cell divisions.

Acknowledgments Work in the team of Claire Vourc'h is supported by grants from by ARC (#5113) and by ANR (PCV08_324703). G. Biamonti is supported by grants from AIRC, Cariplo Foundation, and from European Union (EURASNET) Network of Excellence on Alternative Splicing (EURASNET).

References

Alastalo TP, Hellesuo M, Sandqvist A, Hietakangas V, Kallio M, Sistonen L (2003) Formation of nuclear stress granules involves HSF2 and coincides with the nucleolar localization of Hsp70. J Cell Sci 116:3557–3570

Allshire RC, Karpen GH (2008) Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet 9:923–937

Biamonti G (2004) Nuclear stress bodies: a heterochromatin affair? Nat Rev Mol Cell Biol 5:493–498

Biamonti G, Vourc'h C (2010) Nuclear Stress Bodies Cold Spring Harb Perspect Biol. 2:a000695

Bouzinba-Segard H, Guais A, Francastel C (2006) Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proc Natl Acad Sci USA 103:8709–8714

Chen ES, Zhang K, Nicolas E, Cam HP, Zofall M, Grewal SI (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature 451:734–737

Chiodi I, Corioni M, Giordano M, Valgardsdottir R, Ghigna C, Cobianchi F, Xu RM, Riva S, Biamonti G (2004) RNA recognition motif 2 directs the recruitment of SF2/ASF to nuclear stress bodies. Nucleic Acids Res 32:4127–4136

Denegri M, Chiodi I, Corioni M, Cobianchi F, Riva S, Biamonti G (2001) Stress-induced nuclear bodies are sites of accumulation of pre-mRNA processing factors. Mol Biol Cell 12:3502–3514

Du Y, Topp CN, Dawe RK (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. PLoS Genet 6:e1000835

Ehrlich M (2003) The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clin Immunol 109:17–28

Enukashvily NI, Donev R, Waisertreiger IS, Podgornaya OI (2007) Human chromosome 1 satellite 3 DNA is decondensed, demethylated and transcribed in senescent cells and in A431 epithelial carcinoma cells. Cytogenet Genome Res 118:42–54

Eymery A, Callanan M, Vourc'h C (2009a) The secret message of heterochromatin: new insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription. Int J Dev Biol 53:259–268

Eymery A, Horard B, El Atifi-Borel M, Fourel G, Berger F, Vitte AL, Van den Broeck A, Brambilla E, Fournier A, Callanan M, Gazzeri S, Khochbin S, Rousseaux S, Gilson E, Vourc'h C (2009b) A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. Nucleic Acids Res 19:6340–6354

Eymery A, Souchier C, Vourc'h C, Jolly C (2010) Heat shock factor 1 binds to and transcribes satellite II and III sequences at several pericentromeric regions in heat-shocked cells. Exp Cell Res 316:1845–1855

Fernandes M, Xiao H, Lis JT (1994) Fine structure analyses of the Drosophila and Saccharomyces heat shock factor-heat shock element interactions. Nucleic Acids Res 22:167–173

Ferri F, Bouzinba-Segard H, Velasco G, Hubé F, Francastel C (2009) Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res 37:5071–5080

Fisher AG, Merkenschlager M (2002) Gene silencing, cell fate and nuclear organisation. Curr Opin Genet Dev 12:193–197

Fournier A, McLeer-Florin A, Lefebvre C, Duley S, Barki L, Ribeyron J, Alboukadel K, Hamaidia S, Granjon A, Gressin R, Lajmanovich A, Bonnefoix T, Chauvelier S, Debernardi A, Rousseaux S, de Fraipont F, Figeac M, Kerckaert JP, De Vos J, Usson Y, Delaval K, Grichine A, Vourc'h C, Khochbin S, Feil R, Leroux D, Callanan MB (2010) 1q12 chromosome translocations form aberrant heterochromatic foci associated with changes in nuclear architecture and gene expression in B cell lymphoma. EMBO Mol Med 2:159–171

Frescas D, Guardavaccaro D, Kuchay SM, Kato H, Poleshko A, Basrur V, Elenitoba-Johnson KS, Katz RA, Pagano M (2008) KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. Cell Cycle 7:3539–3547

Frommer M, Prosser J, Tkachuk D, Reisner AH, Vincent PC (1982) Simple repeated sequences in human satellite DNA. Nucleic Acids Res 10:547–563

Frommer M, Paul C, Vincent PC (1988) Localisation of satellite DNA sequences on human metaphase chromosomes using bromodeoxyuridine. Chromosoma 97:11

Fukagawa T, Nogami M, Yoshikawa M, Ikeno M, Okazaki T, Takami Y, Nakayama T, Oshimura M (2004) Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat Cell Biol 6:784–791

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem 278:4035–4040

Gaubatz JW, Cutler RG (1990) Mouse satellite DNA is transcribed in senescent cardiac muscle. J Biol Chem 265:17753–17758

Goldman RD, Shumaker DK, Erdos MR, Eriksson M, Goldman AE, Gordon LB, Gruenbaum Y, Khuon S, Mendez M, Varga R, Collins FS (2004) Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson-Gilford progeria syndrome. Proc Natl Acad Sci USA 101:8963–8968

Grewal SI, Elgin SC (2007) Transcription and RNA interference in the formation of heterochromatin. Nature 447:399–406

Jehan Z, Vallinayagam S, Tiwari S, Pradhan S, Singh L, Suresh A, Reddy HM, Ahuja YR, Jesudasan RA (2007) Novel noncoding RNA from human Y distal heterochromatic block (Yq12) generates testis-specific chimeric CDC2L2. Genome Res 17:433–440

Jolly C, Lakhotia SC (2006) Human sat III and Drosophila hsr omega transcripts: a common paradigm for regulation of nuclear RNA processing in stressed cells. Nucleic Acids Res 34:5508–5514

Jolly C, Konecny L, Grady DL, Kutskova YA, Cotto JJ, Morimoto RI, Vourc'h C (2002) In vivo binding of active heat shock transcription factor 1 to human chromosome 9 heterochromatin during stress. J Cell Biol 156:775–781

Jolly C, Metz A, Govin J, Vigneron M, Turner BM, Khochbin S, Vourc'h C (2004) Stress-induced transcription of satellite III repeats. J Cell Biol 164:25–33

Jones KW, Purdom IF, Prosser J, Corneo G (1974) The chromosomal localisation of human satellite DNA I. Chromosoma 49:161–171

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev 19:489–501

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol 13:1192–1200

Lu J, Gilbert DM (2007) Proliferation-dependent and cell cycle regulated transcription of mouse pericentric heterochromatin. J Cell Biol 179:411–21

Lu J, Gilbert DM (2008) Cell cycle regulated transcription of heterochromatin in mammals vs. fission yeast: functional conservation or coincidence? Cell Cycle 7:1907–1910

Maison C, Bailly D, Peters AH, Quivy JP, Roche D, Taddei A, Lachner M, Jenuwein T, Almouzni G (2002) Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat Genet 30:329–334

Manuelidis L (1982) Nucleotide sequence definition of a major human repeated DNA, the Hind III 1.9 kb family. Nucleic Acids Res 10(10):3211–3219

Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J 24:800–801

Massé J, Laurent A, Nicol B, Guerrier D, Pellerin I, Deschamps S (2010) Involvement of ZFPIP/Zfp462 in chromatin integrity and survival of P19 pluripotent cells. Exp Cell Res 316:1190–1201

Masui O, Heard E (2006) RNA and protein actors in X-chromosome inactivation. Cold Spring Harb Symp Quant Biol 71:419–428

Metz A, Soret J, Vourc'h C, Tazi J, Jolly C (2004) A key role for stress-induced satellite III transcripts in the relocalization of splicing factors into nuclear stress granules. J Cell Sci 117:4551–4558

Meyne J, Goodwin EH, Moyzis RK (1994) Chromosome localization and orientation of the simple sequence repeat of human satellite I DNA. Chromosoma 103:99

Misteli T, Scaffidi P (2005) Genome instability in progeria: when repair gets old. Nat Med 7:718–719

Mitchell AR, Beauchamp RS, Bostock CJ (1979) A study of sequence homologies in four satellite DNAs of man. J Mol Biol 135:127–149
Morimoto RI (1998) Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes Dev 12:3788–3796
Muchardt C, Guilleme M, Seeler JS, Trouche D, Dejean A, Yaniv M (2002) Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. EMBO Rep 3:975–981
Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998a) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393:386–389
Nan X, Cross S, Bird A (1998b) Gene silencing by methyl-CpG-binding proteins. Novartis Found Symp 214:6–16
Papait R, Pistore C, Negri D, Pecoraro D, Cantarini L, Bonapace IM (2007) Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. Mol Biol Cell 18:1098–1106
Prosser J, Frommer M, Paul C, Vincent PC (1986) Sequence relationships of three human satellite DNAs. J Mol Biol 187:145–155
Rizzi N, Denegri M, Chiodi I, Corioni M, Valgardsdottir R, Cobianchi F, Riva S, Biamonti G (2004) Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. Mol Biol Cell 15:543–551
Rudert F, Gronemeyer H (1993) Retinoic acid-response elements with a highly repetitive structure isolated by immuno-selection from genomic DNA. J Steroid Biochem Mol Biol 46:121–133
Rudert F, Bronner S, Garnier JM, Dolle P (1995) Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mamm Genome 6:76–83
Sandqvist A, Björk JK, Akerfelt M, Chitikova Z, Grichine A, Vourc'h C, Jolly C, Salminen TA, Nymalm Y, Sistonen L (2009) Heterotrimerization of heat-shock factors 1 and 2 provides a transcriptional switch in response to distinct stimuli. Mol Biol Cell 5:1340–1347
Sengupta S, Parihar R, Ganesh S (2009) Satellite III non-coding RNAs show distinct and stress-specific patterns of induction. Biochem Biophys Res Commun 382:102–107
Shumaker DK, Dechat T, Kohlmaier A, Adam SA, Bozovsky MR, Erdos MR, Eriksson M, Goldman AE, Khuon S, Collins FS, Jenuwein T, Goldman RD (2006) Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. Proc Natl Acad Sci USA 103:8703–8708
Tagarro I, Wiegant J, Raap AK, González-Aguilera JJ, Fernández-Peralta AM (1994) Assignment of human satellite 1 DNA as revealed by fluorescent in situ hybridization with oligonucleotides. Hum Genet 93:125–128
Talbert PB, Henikoff S (2006) Spreading of silent chromatin: inaction at a distance. Nat Rev Genet 7:793–803
Terranova R, Sauer S, Merkenschlager M, Fisher AG (2005) The reorganisation of constitutive heterochromatin in differentiating muscle requires HDAC activity. Exp Cell Res 310:344–356
Ugarković Đ (2009) Centromere-competent DNA: structure and evolution. Prog Mol Subcell Biol 48:53–76
Valgardsdottir R, Chiodi I, Giordano M, Rossi A, Bazzini S, Ghigna C, Riva S, Biamonti G (2007) Transcription of Satellite III non-coding RNAs is a general stress response in human cells. Nucleic Acids Res 36:423–434
Verdel A, Vavasseur A, Le Gorrec M, Touat-Todeschini L (2009) Common themes in siRNA-mediated epigenetic silencing pathways. Int J Dev Biol 53:245–257
Vissel JB, Choo KH (1987) Human alpha satellite DNA – consensus sequence and conserved regions. Nucleic Acids Res 15:6751–6752
Weighardt F, Cobianchi F, Cartegni L, Chiodi I, Villa A, Riva S, Biamonti G (1999) A novel hnRNP protein (HAP/SAF-B) enters a subset of hnRNP complexes and relocates in nuclear granules in response to heat shock. J Cell Sci 112:1465–1476

Wong AK, Rattner JB (1988) Sequence organization and cytological localization of the minor satellite of mouse. Nucleic Acids Res 16:11645–11661

Wong LH, Brettingham-Moore KH, Chan L, Quach JM, Anderson MA, Northrop EL, Hannan R, Saffery R, Shaw ML, Williams E, Choo KH (2007) Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res 17:1146–1160

Chapter 6
Multiple Roles of *Alu*-Related Noncoding RNAs

Audrey Berger and Katharina Strub

Abstract Repetitive *Alu* and *Alu*-related elements are present in primates, tree shrews (*Scandentia*), and rodents and have expanded to 1.3 million copies in the human genome by nonautonomous retrotransposition. Pol III transcription from these elements occurs at low levels under normal conditions but increases transiently after stress, indicating a function of *Alu* RNAs in cellular stress response. *Alu* RNAs assemble with cellular proteins into ribonucleoprotein complexes and can be processed into the smaller sc*Alu* RNAs. *Alu* and *Alu*-related RNAs play a role in regulating transcription and translation. They provide a source for the biogenesis of miRNAs and, embedded into mRNAs, can be targeted by miRNAs. When present as inverted repeats in mRNAs, they become substrates of the editing enzymes, and their modification causes the nuclear retention of these mRNAs. Certain *Alu* elements evolved into unique transcription units with specific expression profiles producing RNAs with highly specific cellular functions.

6.1 Introduction

Alu elements are the most abundant repetitive elements in the human genome and belong to the short interspersed elements (SINE). Nonautonomous retrotransposition allowed these elements to propagate successfully in primate genomes. Amplification occurred in sequential waves, and *Alu* elements are currently present at more than one million copies in the human genome, representing more than 10% of its content (Lander et al. 2001). If these parasites of our genome were initially considered as junk DNA, it has now become evident that they play crucial and diverse roles in regulating gene expression. In addition, *Alu* elements have a major

A. Berger and K. Strub (✉)
Department of Cell Biology, University of Geneva, 30 quai Ernest Ansermet, 1211 Geneva 4, Switzerland
e-mail: Katharina.Strub@unige.ch

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3_6,

impact on the architecture of the human genome through homologous recombination resulting in duplications and deletions (for review, see Batzer and Deininger 2002). The modern *Alu* elements are approximately 300 base pairs (bp) in length and are composed of two similar but nonidentical *Alu* monomers linked by an A-rich sequence (Fig. 6.1, Deininger et al. 1981). *Alu* RNAs are expressed from *Alu* elements, but can also be found embedded into large transcripts such as pre-mRNAs and mRNAs when transcribed as part of protein-coding genes. *Alu* RNAs located in introns promote alternative splicing by providing splice donor and acceptor sites. Thus, they contribute substantially to increase the diversity of the human proteome (for review, see Keren et al. 2010). In the following, we will summarize our knowledge on *Alu* RNA biogenesis and *Alu* RNA functions in modulating gene expression at the transcriptional and at the posttranscriptional level.

6.2 *Alu*: Birth and Evolution

Alu and *Alu*-related elements derive from the 7SL RNA gene, which encodes the RNA moiety of the signal recognition particle (SRP, Ullu and Tschudi 1984). SRP is a cytoplasmic ribonucleoprotein particle involved in cotranslational translocation of proteins into the endoplasmic reticulum. *Alu* and *Alu*-related elements are found in primates, rodents, and tree shrews, and their common ancestor is the monomeric FLAM-A *Alu* element (free left *Alu* monomer subtype A). In primates, an additional element derived from the 7SL RNA gene, the fossil *Alu* monomer (FAM), is considered to be the ancestor of the free right *Alu* monomer (FRAM, Kriegs et al. 2007). A fusion between a FRAM and a FLAM element gave rise to the dimeric *Alu* element (Quentin 1992). The modern *Alu* element comprises therefore two copies of the 7SL-derived *Alu* sequences, named the left and the right arm, linked by an A-rich sequence. The two arms have similar, but nonidentical, sequences (Fig. 6.1).

In addition to dimeric *Alu* elements, primate genomes also contain truncated *Alu* elements and unique *Alu*-derived transcription units with tissue-specific expression patterns such as the BC200, the 21A, and the NMD29 RNA genes (Fig. 6.1, Pagano et al. 2007; Tiedge et al. 1993). Interestingly, the type II family of repetitive sequences in *Galago crassicaudatus* contains a tRNA-like region followed by a sequence resembling the *Alu* right arm (Daniels and Deininger 1983, 1985). Two *Alu*-related SINE families comprising a tRNA-like region followed by an *Alu*-like region were also described in tree shrew (*Tupaia belangeri*, Nishihara et al. 2002).

In rodents, the repetitive monomeric elements B1 and 4.5S are present. The B1 element is probably found in all rodents and represents 2.7% of the mouse genome (Vassetzky et al. 2003; Waterston et al. 2002). B1 elements can be classified into six subfamilies (Quentin 1989). They comprise an *Alu* domain of approximately 130 bp followed by an A-rich region of variable length (Jelinek and Schmid 1982). They resemble the *Alu* left monomer, but they contain a deletion (7, 9, or 10 bp) and a tandem duplication (20 or 29 bp, Fig. 6.1, Quentin 1994; Rogers 1985; Veniaminova et al. 2007). The 4.5S element is present in mice, Chinese hamsters,

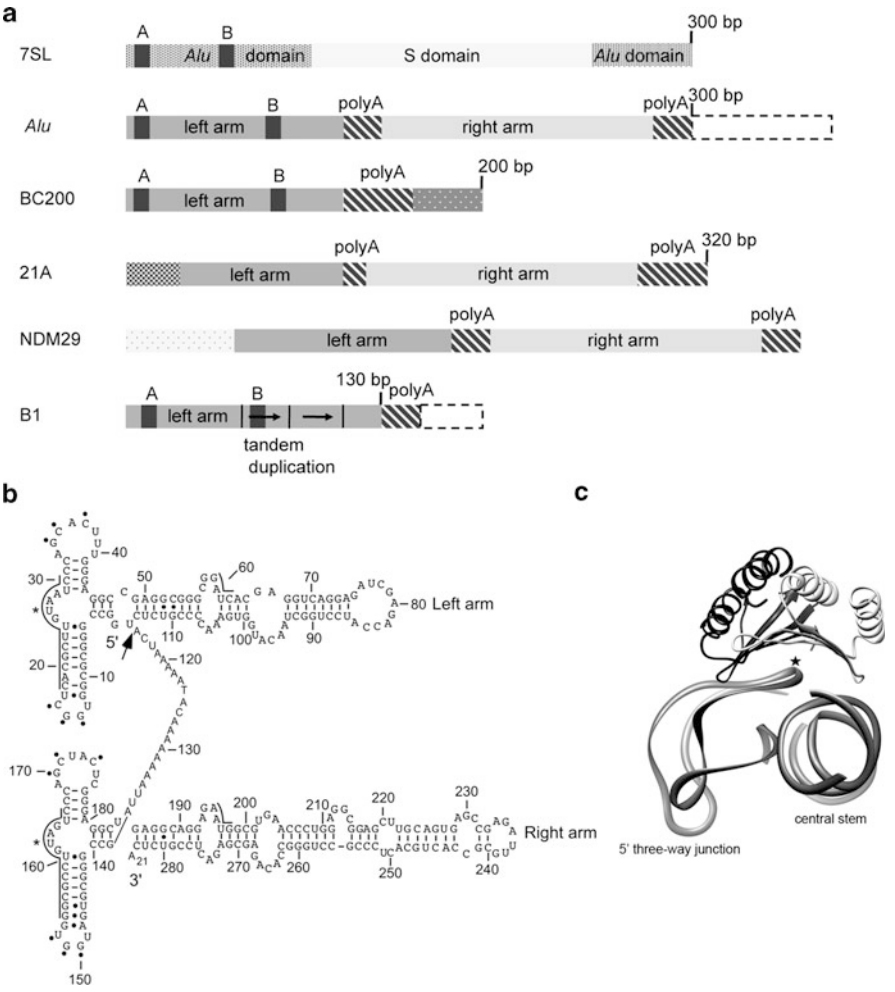

Fig. 6.1 Structures of Alu elements, Alu RNA, and SRP9/14 bound to Alu RNA. (**a**) The Alu domain of the 7SL RNA gene is composed of a 5′ and a 3′ portion interrupted by the S domain. The modern Alu element is dimeric including two copies of the 7SL Alu domain, named the left (*dark grey*) and the right arm (*light grey*). Their sequences are similar but not identical and they are connected by a poly(A) linker. The left arm contains the internal promoter elements, box A and box B. The right arm is followed by a poly(A) tail and genomic sequences of variable length extending to the transcription termination site (*dashed lines*). The BC200 RNA gene is a monomeric Alu element followed by a specific sequence at the 3′ end (*white dot shaded*). The 21A RNA gene represents a partially conserved dimeric AluJb RNA gene. The 5′ sequences are strongly degenerated (*double-hatched rectangle*). The NDM29 RNA gene is composed of a unique sequence at the 5′ end (*black dot shaded*) followed by a conserved AluJb sequence. The B1 element is represented by a left arm monomer comprising a duplication in tandem followed by a poly(A) tail and genomic sequences of variable length (*dashed lines*). (**b**) The secondary structure of the Alu RNA present in the intron 4 of the alpha-fetoprotein (nt 5,069–5,372) as predicted in analogy to the one of the Alu domain of 7SL RNA. *Lines*: SRP9/14 binding sites; *stars*: U-turns;

and rats (Harada and Kato 1980; Harada et al. 1979; Haynes et al. 1981; Leinwand et al. 1982). They have a length of about 100 bp and share sequence similarities with the first part of the B1 element. The sequences of 4.5S elements are highly conserved between Chinese hamster and mice (Harada and Kato 1980; Haynes et al. 1981).

Alu elements spread through the genome by retrotransposition. The rates of retrotransposition have not remained constant throughout evolution. The greatest burst in activity was detected around 40 Mya (Britten 1994; Lander et al. 2001; Shen et al. 1991), whereas today, the activity is relatively low with one insertion event for 20 births (Cordaux et al. 2006). To proliferate, *Alu* elements have to be actively transcribed and retrotransposed. Active *Alu* elements are referred to as source genes. According to the oldest model (for a review, see Deininger et al. 1992), only few source genes, the so-called master genes, were actively retrotransposed during a certain time period, giving rise to a subfamily of *Alu* sequences with diagnostic mutations passed on from the master gene. These diagnostic mutations allowed the classification of *Alu* elements into three major groups. The most ancient group is called *Alu*J subfamily, the *Alu*S elements constitute an intermediate (Jurka and Milosavljevic 1991; Jurka and Smith 1988), and the *Alu*Y elements represent the youngest subfamily (Batzer et al. 1996). Due to the availability of the human genome sequence, it became clear over the past years that the evolution of *Alu* elements might have been more complex than previously anticipated (Price et al. 2004). Based on the sequence information, a more complex tree of *Alu* elements has been built, containing 213 subfamilies grouped into *Alu*J, *Alu*S, and *Alu*Y subfamilies. While in certain subfamilies *Alu* elements proliferated according to the master gene theory, a plethora of source genes were active in other subfamilies. Most likely, the numbers of source genes is still underestimated today, and it is conceivable that it may reach thousands of copies (Cordaux et al. 2004; Price et al. 2004; Styles and Brookfield 2009).

Alu elements are nonautonomous for their retrotransposition. They use in trans the enzymes encoded by the repetitive elements LINE-1 (L1, Dewannieux et al. 2003; Hagan et al. 2003; Jurka 1997). The size of the L1 element is 6 kilobase pair (kb), and it contains an internal polymerase II (Pol II) promoter and possesses two open reading frames coding for ORF1p and ORF2p proteins (Dombroski et al. 1991; Scott et al. 1987). ORF1p is probably not essential but enhances *Alu* retrotransposition (Wallace et al. 2008). The ORF2p protein has endonuclease and reverse transcriptase activity. It cleaves the genomic DNA at the consensus site TTAAAA (Feng et al. 1996; Jurka 1997) and mediates the integration of *Alu*

Fig. 6.1 (Continued) dots: Base pairs between the two loops; *arrow*: Processing site for scAlu RNA. (**c**) Structure model of human SRP9/14 bound to a small 7SL Alu RNA. SRP9 (*light grey*) and SRP14 (*black*) are structurally homologues polypeptides that form together a six-stranded β-sheet, which represents the RNA binding site. Alu RNA: The 5′ three-way junction is folded into a compact structure and the central stem flips back by almost 180° to align alongside the 5′ domain underneath the curved β-sheet of the protein. *Star*: U-turn. The molecular graphics image was produced using the UCSF Chimera package

elements through target-primed reverse transcription (TPRT, Luan et al. 1993; Mathias et al. 1991). The remaining steps of the process have not been elucidated, but they lead to the insertion of a double-strand DNA flanked by direct repeats. Notably, *Alu* elements can also use other mechanisms of retrotransposition (Callinan et al. 2005; Srikanta et al. 2009a, b).

6.3 Cellular Levels, Maturation, and Localization of *Alu* RNAs

Despite the high abundance of *Alu* elements in the primate genomes, *Alu* RNAs are expressed at very low levels in most cells and tissues (Liu et al. 1994; Paulson and Schmid 1986; Shaikh et al. 1997; Sinnett et al. 1992). The abundance of *Alu* RNAs was estimated to be around 100–1,000 copies in HeLa cells (Liu et al. 1994). *Alu* RNAs are expressed from different genomic loci, and members of the young *Alu* subfamilies are more frequently expressed (Shaikh et al. 1997; Sinnett et al. 1992). The average half-life of *Alu* RNA is quite short in the order of 0.5 h (Chu et al. 1995; Li and Schmid 2004). B1 RNAs in rodents are also short-lived (Li and Schmid 2004).

A specific 10- to 20-fold increase of *Alu* RNA levels is observed in human cells following different types of stress-inducing treatments such as inhibition of protein synthesis with cycloheximide or puromycin, heat shock, and viral infection (Jang and Latchman 1989; Liu et al. 1995; Panning and Smiley 1993, 1994; Russanova et al. 1995). This increase is transient, and it starts 30 min to 1 h after cycloheximide treatment and declines approximately 7 h later (Allen et al. 2004; Carey et al. 1986; Li et al. 1999; Liu et al. 1995). A transient increase of B1 RNA levels was also observed in mouse liver, kidney, and spleen after a severe hyperthermic shock (Li et al. 1999), as well as in mouse testes, which already contain high levels of B1 RNAs compared to other tissues. Cellular *Alu* and B1 RNA levels are also increased in human hepatocellular carcinomas and in murine plasmocytoma tumors (Kramerov et al. 1990; Tang et al. 2005) and after treatment with DNA damaging agents (Hagan and Rudin 2007; Rudin and Thompson 2001). Thus, *Alu* RNAs may have important functions in the cellular response to stress and to malignant transformation.

Alu RNAs can be processed into small cytoplasmic *Alu* RNAs (sc*Alu* RNAs) comprising only the left arm (Fig. 6.1, Chu et al. 1995; Maraia et al. 1993; Matera et al. 1990), while *Alu* RNA comprising the right arm fails to accumulate stably (Li and Schmid 2004). When compared to *Alu* RNAs, sc*Alu* RNAs are more stable with a half-life of about 3 h (Chu et al. 1995; Li and Schmid 2004; Sarrowa et al. 1997) and more abundant with about 10^3 at 10^4 copies in HeLa cells. They are generated from *Alu* RNAs of different loci and more frequently from *Alu* RNAs expressed from *Alu* elements of young subfamilies. This might be explained by (1) the higher expression levels of *Alu* elements of young subfamilies and (2) the decreased ability

of the *Alu* right arm RNA derived from such elements to bind the protein SRP9/14 (see below). In the absence of protein binding, RNA processing into scAlu RNA might be facilitated (Maraia et al. 1993; Sarrowa et al. 1997; Shaikh et al. 1997). Furthermore, overexpression of *Alu* RNAs does not increase the amount of sc*Alu* RNA, suggesting an additional level of regulation (Chu et al. 1995; Liu et al. 1994, 1995; Russanova et al. 1995). B1 RNA is also processed into scB1 RNA, and processing is known to occur in the nucleus (Adeniyi-Jones and Zasloff 1985; Maraia 1991; Maraia et al. 1992).

Whereas sc*Alu* RNA was localized to the cytoplasm, the cellular localization of *Alu* RNA is still controversial. Using biochemical fractionation experiments, some reports showed that *Alu* RNA accumulates in the cytoplasm under standard conditions (Liu et al. 1994; Sinnett et al. 1992) and after stress (Liu et al. 1995; Panning and Smiley 1993, 1994). In another report, the nuclear RNA fraction was studied and it showed an increase in *Alu* RNA after stress (Mariner et al. 2008). In localization studies using *Alu* RNA comprising MS2 binding sites at its 3′ end in conjunction with the GFP-NLS-MS2 fusion protein, the RNA was found in Cajal bodies (CBs) and, less frequently, in promyelocytic leukemia bodies (PML, Goodier et al. 2010). *Alu* RNA expressed in *Xenopus laevis* oocytes showed a nuclear accumulation (Perlino et al. 1985). Further studies are required to clarify this issue. Notably, the *Alu* portion of 7SL RNA plays a role in its export from the nucleus (He et al. 1994), suggesting that *Alu* RNA might be competent for export to the cytoplasm.

6.4 Transcriptional Regulation of *Alu* RNA Expression

Alu elements are transcribed by Pol III and contain two internal promoter regions, named box A and box B, which are also present in tRNA promoters (Paolella et al. 1983; Perez-Stable et al. 1984). Box B was shown to be essential for *Alu* transcription, while box A determines the transcription start site (Fuhrman et al. 1981; Paolella et al. 1983; Perez-Stable et al. 1984; Perez-Stable and Shen 1986). While box A and box B are sufficient to drive efficient transcription of *Alu* elements in vitro (for example see Elder et al. 1981; Perez-Stable et al. 1984), their expression in vivo requires 5′ and 3′ flanking sequences (Shaikh et al. 1997). The flanking sequences are of genomic origin and most likely contribute to enhance transcription by providing additional upstream promoter elements. The expression of the 7SL RNA gene also depends on flanking sequences (Bredow et al. 1990; Ullu and Weiner 1985), and when these flanking sequences were inserted upstream and downstream of an *Alu* element, respectively, its expression increased a 100-fold in human kidney cells (Chu et al. 1995). In addition, other types of upstream Pol III promoters such as the U6 RNA gene promoter were shown to enhance *Alu* RNA expression in several cell lines (Roy et al. 2000). For the specific *Alu* element, EPL *Alu*, a 5′ proximal binding site for the transcription factor AP-1 was important for

transcription as well as another nonidentified upstream element (Chesnokov and Schmid 1996).

Alu and B1 elements usually lack the canonical TTTT stretch necessary for Pol III termination (Elder et al. 1981; Fuhrman et al. 1981). Transcription will therefore terminate at the most proximal TTTT stretch in the genomic environment, generating RNAs of different sizes ranging between 300 and 500 nucleotides (nt) and with variable 3′ sequences (Chang and Maraia 1993; Liu et al. 1994, 1995; Maraia et al. 1993; Matera et al. 1990; Russanova et al. 1995). Other types of terminator sequences can also be used by Pol III to terminate *Alu* RNA synthesis (Hess et al. 1985). Furthermore, cellular *Alu* RNA levels change in response to mutations affecting the place or the nature of the terminator sequence, suggesting that the 3′ sequences influence the steady-state levels of *Alu* RNAs, possibly, by changing the stability and the processing of the newly synthesized *Alu* RNA (Aleman et al. 2000). Notably, the La protein was shown to be involved in both termination and transcription activation. The La protein binds to the oligo (U) stretch present at the 3′ ends of Pol III transcripts such as B1 and *Alu* RNAs (Chang et al. 1996; Maraia et al. 1988). It acts as a transcription activator by facilitating the release of B1 and *Alu* RNAs from Pol III upon transcription termination (Maraia 1996; Maraia et al. 1994) and protects B1 transcripts from 3′ processing (Maraia et al. 1994). Recycling of transcription complexes by La is dependent on the sequences flanking the terminator (Goodier and Maraia 1998).

Viral proteins may increase *Alu* element expression. Specifically, the viral proteins ICP27 (Herpes Simplex Virus) and Tat (Human Immunodeficiency Virus) raise the expression by enhancing the activity of the Pol III transcription factor TFIIIC, which binds to the internal box B element (Jang et al. 1992; Jang and Latchman 1992). However, the same mechanism does not explain the increased cellular levels of *Alu* RNA after adenovirus infection. In this case, as well as after heat shock and cycloheximide treatment, the increase correlated with greater chromatin accessibility of genomic *Alu* elements, and upon recovery from stress, chromatin reclosing occurred concomitantly with a reduction in *Alu* RNA levels (Kim et al. 2001; Li et al. 2000; Panning and Smiley 1993; Russanova et al. 1995).

Reversible methylation of cytosine in CpG dinucleotides is a known mechanism for transcriptional repression. Methylation can abrogate transcription factor binding and induces a repressive chromatin structure. One third of CpG dinucleotides in the human genome resides in *Alu* elements (Hellmann-Blumberg et al. 1993; Schmid 1991) and methylation is therefore an important mechanism for their transcriptional repression. Accordingly, the inhibition of methylation with 5-aza-2′-deoxycytidine results in the derepression of *Alu* element transcription (Liu et al. 1994), and hypomethylation of *Alu* Y elements correlates with abundant *Alu* RNA expression in K562 cells (Li et al. 2000). Methylation also inhibits in vitro transcription of different *Alu* elements by interfering with binding of transcription factors (Kochanek et al. 1993, 1995). Several reports show that hypomethylation of *Alu* elements correlates with high levels of *Alu* RNA in cancer cells (Cho et al. 2007; Choi et al. 2007; Daskalos et al. 2009; Xiang et al. 2010). Two tumor suppressor proteins p53 and Rb (retinoblastoma susceptibility gene product) suppress transcription of *Alu*

elements in vivo and in vitro, most likely via their interaction with the Pol III transcription factor TFIIIB (Chesnokov et al. 1996; Chu et al. 1997). The repression by p53 is alleviated in the presence of flanking sequences enhancing transcription (Chesnokov et al. 1996). Inactivation of tumor suppressor proteins and hypomethylation of CpG dinucleotides might therefore both contribute to increased cellular *Alu* RNA levels in cancer cells.

Whereas *Alu* elements are generally heavily methylated in somatic tissues and oocytes, they are hypomethylated in testis, and as mentioned before, the expression levels of B1 RNA are higher in mouse testis than in other tissues (Hellmann-Blumberg et al. 1993; Kochanek et al. 1993; Rubin et al. 1994; Schmid 1991). This might be due to the presence of specific *Alu*-binding protein (SABP), a protein extracted from human sperm chromatin, which binds to *Alu* sequences and protects them from being methylated in vitro (Chesnokov and Schmid 1995).

Methylated CpG dinucleotides provide hotspots for TpG or CpA transitions (Bird 1980). Hence, internal promoter elements (A and B-boxes) are expected to be more degenerate in older than in younger *Alu* elements. It is therefore likely that two different mechanisms of transcriptional repression are used: The transcription of old elements is repressed by frequent mutations in the promoter elements, whereas transcription of young elements is strongly diminished by heavy methylation (Britten et al. 1988; Jurka and Milosavljevic 1991; Labuda and Striker 1989; Liu and Schmid 1993; Schmid 1991).

In conclusion, the expression of *Alu* RNAs is tightly regulated by different factors and at different levels. *Alu* RNA levels are kept low at normal conditions, a mechanism, which contributes to maintain retrotransposition frequencies low, avoiding harmful damage to the organism. Certain conditions such as stress induce the expression levels transiently. Remarkably, different *Alu* elements do not have the same expression profile, and stress conditions do not activate the expression of all *Alu* elements, indicating a regulation of *Alu* RNA expression at the local rather than at the global level (Li and Schmid 2001).

6.5 *Alu* RNA in Transcription Regulation

The first hint for a role of *Alu* RNA in transcription regulation came from the observation that, in human kidney cells, the increase in *Alu* RNA levels after stress could be linked to a decrease in the levels of four mRNAs (Mariner et al. 2008). In addition, synthetic *Alu* RNA inhibited effectively transcription in a minimally reconstituted transcription reaction composed of the transcription factors TATA-box-binding protein (TBP), TFIIB, and TFIIF, as well as Pol II. Further in vitro and in vivo experiments demonstrated that the *Alu* RNA and Pol II were simultaneously present at the inactivated promoters. The inhibitory activity was assigned to the A-rich linker between the arms and to the L region of the right arm (nt 193–209; 257–272 in Fig. 6.1, Mariner et al. 2008). Interestingly, B1 and left arm *Alu* RNAs lacked inhibitory activity, although they bound Pol II (Allen et al. 2004; Espinoza

et al. 2004; Mariner et al. 2008). Using cross-linking and footprinting experiments in transcription reactions in vitro, it was established that *Alu* RNA inhibits transcription by preventing direct contacts between the two large subunits of Pol II and the promoter, probably by inducing conformational changes in the complex. In agreement with this model, *Alu* RNA is no longer able to prevent transcription, once the closed transcription complex is formed on the promoter (Yakovchuk et al. 2009). In the same complex, TBP binding to the promoter is maintained and TFIIB binding is enhanced (Fig. 6.2).

Recent in vitro studies identified an additional piece of the puzzle. TFIIF selectively destabilized the B1/sc*Alu* RNA-Pol II complex, whereas it had no effect on the *Alu* RNA-Pol II complex (Wagner et al. 2010). Moreover, when the *Alu* RNA repressor region was added to B1 RNA, the complex with Pol II was no longer destabilized by TFIIF. Clearly, TFIIF appears to provide selectivity in this mechanism of transcriptional regulation by distinguishing between RNAs with and without a repressor domain.

It should be mentioned that B2 RNA, a small murine ncRNA, which is transcribed from the repetitive B2 element originating from a tRNA gene, inhibits transcription by a similar mechanism (Espinoza et al. 2004, 2007). Since both RNAs act directly on Pol II and in minimal transcription systems, they are expected to function as general transcriptional repressors. It will therefore be interesting to identify other target genes that are inhibited by *Alu* RNA to obtain information about the scope of this regulatory mechanism and to reveal factors that may restrict *Alu* RNA repression to specific genes.

6.6 *Alu* RNA and *Alu* RNP in Translation Regulation

Most cellular RNAs work together with protein partners to accomplish their functions. An important binding partner of *Alu* RNA is SRP9/14. In SRP, SRP9/14 binds to the *Alu* portion of 7SL RNA and confers elongation arrest activity to SRP, which is required to slow down specifically the elongation of ER-targeted nascent chains in vitro and in vivo (Lakkaraju et al. 2008; Siegel and Walter 1988). *Alu* RNAs have preserved many structural elements of the *Alu* portion of 7SL RNA and can therefore be folded into a similar secondary structure (Fig. 6.1). The similarity in structure was confirmed by the binding of SRP9/14 to *Alu*, sc*Alu*, and B1 RNAs in vitro and in vivo (Bovia et al. 1995; Chang et al. 1994; Hsu et al. 1995). Competition experiments in vitro revealed that the affinity of the protein is decreased up to 40-fold for *Alu* RNAs as compared to the *Alu* domain of 7SL RNA (Bennett et al. 2008; Bovia et al. 1997; Sarrowa et al. 1997). However, the dissociation constants stay very low with values in the subnanomolar and nanomolar range for sc*Alu* and *Alu* right arm RNAs, respectively, and binding remains therefore highly specific. Remarkably, in primate species, SRP9/14 expression was uncoupled from the expression of other SRP proteins. The protein is present in 20-fold excess over SRP, is predominantly found in the cytoplasm, and can bind to

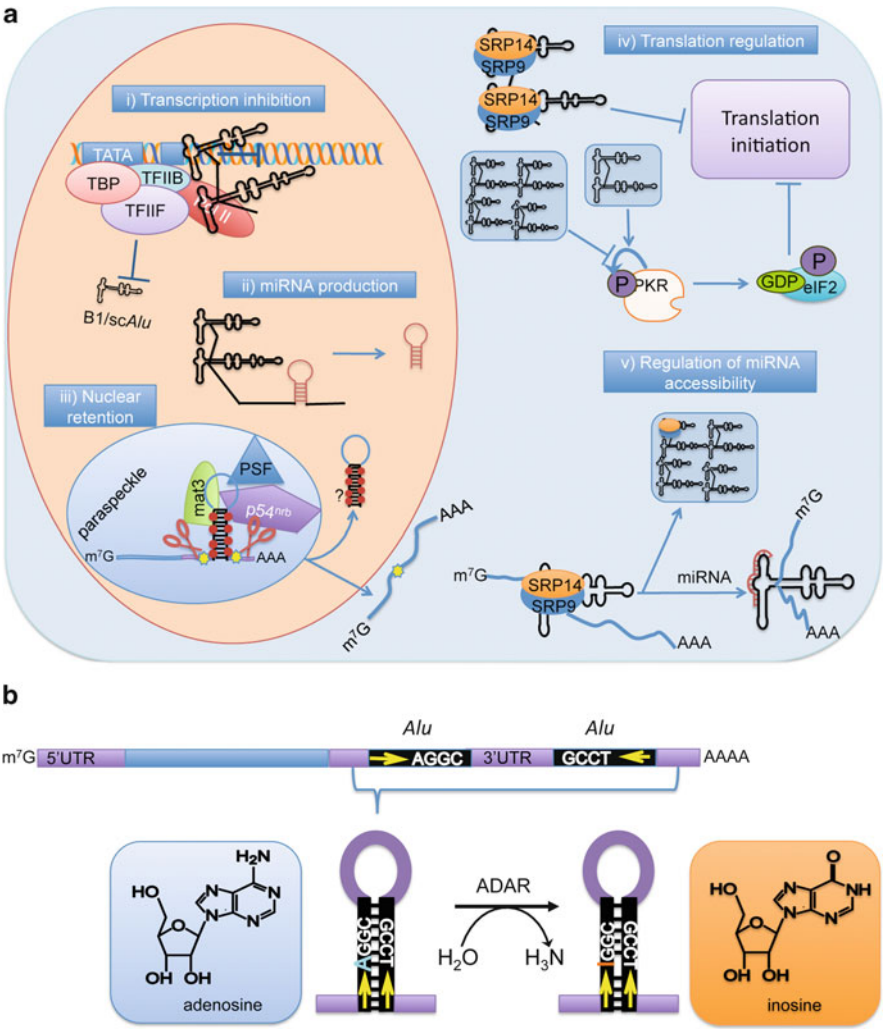

Fig. 6.2 Proposed roles for Alu RNAs in regulating gene expression. (**a**) Alu RNAs may regulate gene expression in several ways (1) They inhibit transcription by blocking contacts between Pol II and the promoter. The transcription factor TFIIF selectively prevents the stable association of B1/scAlu RNAs with Pol II. (2) miRNA sequences present in the 3′ flanking regions of Alu elements can be transcribed simultaneously by Pol III. (3) IRAlus (*hairpin*) are main targets of editing (*red dots*) and edited mRNAs are retained in the nucleus. Stress conditions induce the cleavage of IRAlus allowing the transcript to be exported to the cytoplasm after its religation (*yellow star*). It is not known whether IRAlus are further degraded PSPIα not shown. (4) SRP9/14 bound to Alu, scAlu and right arm Alu RNAs inhibits translation initiation, most likely, by binding to soluble initiation factor(s). Alu RNA binding to the kinase PKR regulates translation initiation. At low levels of Alu RNA, the kinase is activated and phosphorylates the initiation factor eIF2 resulting in the inhibition of initiation. At high levels, PKR is inhibited and translation initiation proceeds normally. (5) Alu RNAs embedded in mRNAs are protected from miRNAs by binding to SRP9/14. Increased Alu RNA levels in response to stress compete for SRP9/14 binding, thereby rendering Alu RNAs in mRNAs accessible for miRNAs. (**b**) Editing of an mRNA containing an IRAlu in its 3′ UTR. IRAlu form intramolecular duplexes recognized by ADAR enzymes. Deamination of adenosine by ADAR produces inosine. *Blue*: Coding region

synthetic *Alu* RNA added to cell extracts (Bovia et al. 1995; Chang et al. 1995). Primate cells therefore have a large pool of SRP9/14 for binding to *Alu* RNAs.

The crystal structure of the human *Alu* RNA-SRP9/14 complex revealed significant changes in the RNA structure upon SRP9/14 binding (Weichenrieder et al. 2000). In the RNA–protein complex, the three-way junction at the 5′ end is folded into a compact structure in which a central U-turn bends the helices to bring the two loops in close proximity, allowing the formation of three base pairs. The U-turn is also the major binding site for the protein (Fig. 6.1). Formation of the high affinity complex is completed when the central stem flips back by 180° and thereby aligns alongside the 5′ domain to make contacts with the positively charged curved structure of the protein (Weichenrieder et al. 2000). SRP9/14 recognizes mostly the shape of the negatively charged phosphate backbone of the RNA.

Another protein, which binds *Alu* and *Alu*-related RNAs, is the cytoplasmic poly (A)-binding protein (PABP, Khanam et al. 2006, 2007; Kondrashov et al. 2005; Muddashetty et al. 2002; Mullin et al. 2004). sc*Alu* and *Alu* RNPs are likely to contain other proteins, as indicated by their Svedberg values of 8.5S and 11S, respectively (Bovia et al. 1995; Liu et al. 1994).

Purified *Alu* RNPs composed of dimeric *Alu*, as well as with only the right arm or the left arm *Alu* RNAs in complex with SRP9/14 specifically inhibited translation in vitro (Fig. 6.2), whereas SRP9/14 had no effect in the same concentration range. Since all RNPs inhibited translation, the activity is most likely located in the protein-bound RNA portion of both arms. These *Alu* RNPs did not affect nascent chain elongation; they inhibited initiation of translation and were not stably associated with ribosomes in the translation reactions. They may therefore inhibit translation by sequestering an initiation factor (Hasler and Strub 2006).

In contrast, *Alu* RNAs alone stimulate translation in vitro and in vivo (Chu et al. 1998; Hasler and Strub 2006; Rubin et al. 2002). In reticulocyte and wheat germ lysates, the translation of several reporter mRNAs was stimulated by the addition of *Alu* RNA at final concentrations of 50–300 nM. The stimulatory effect was on translation initiation and was not equal for all mRNAs. This suggests that the mechanism, by which translation is stimulated, may include an "mRNA-specific component." Consistent with this interpretation, no overall increase in protein synthesis could be detected, if wheat germ lysate was programmed with total poly(A)$^+$ RNA of HeLa cells. The stimulatory activity could be assigned completely to the right arm of *Alu* RNA (Hasler and Strub 2006). In contrast, at concentrations equal or higher than 0.5 μM, *Alu* and B1 RNAs inhibited protein synthesis in vitro by sequestering PABP, which is required for efficient translation initiation and appears to be limiting in cell-free translation systems (Kondrashov et al. 2005).

Changes in *Alu* RNA expression levels can modulate translation via binding of *Alu* RNA to protein kinase R (PKR). PKR is a cellular kinase, which upon activation inhibits protein synthesis via phosphorylation of eIF2α. PKR activity is controlled by double-stranded (ds)RNA-binding and serves as a defense mechanism against viral infections (for review, see Garcia et al. 2007). *Alu* RNA can bind and activate PKR at low concentrations. At high concentrations, *Alu* RNA inhibits

PKR activity, thereby stimulating protein synthesis (Fig. 6.2, Chu et al. 1998; Williams 1999). Subsequent experiments indicated that the *Alu*-dependent regulation of PKR activity does not seem to be the only mechanism by which *Alu* RNA may activate translation. In cells derived from a PKR knockout mouse, *Alu* and B1 RNA expression specifically shortened the lag time for the synthesis of the reporter protein without increasing protein synthesis in general, thus favoring the translation of newly synthesized mRNAs (Rubin et al. 2002). Similar results were obtained in human cells. The stimulating activity could be assigned to the *Alu* right arm as found in the cell-free translation assays. This mechanism would stimulate specifically protein synthesis from the mRNAs that become newly expressed in response to stress such as heat shock and viral infections (Rubin et al. 2002).

In another study, however, it was shown that B1 and BC200 RNAs inhibit the synthesis of a reporter protein (Kondrashov et al. 2005) in vivo, and the inhibitory activity was assigned, at least partially, to the poly(A)-tract. In contrast to the experiments described above, in which the RNAs were expressed from transfected plasmids, in these experiments, cells were directly transfected with the *Alu*-related and the reporter RNAs. As a consequence, these RNAs might assemble into different cellular complexes, which may explain the difference in the experimental outcome as compared to the experiments described above.

In summary, we can say that *Alu* RNA may influence protein synthesis by different mechanisms, and not surprisingly, the critical factor that determines its activity is its assembly with cellular proteins, specifically SRP9/14 and PABP. Clearly, we need to know when and where *Alu* RNAs assemble with SRP9/14 and PABP and whether other proteins also bind *Alu* RNA. Moreover, further insights into the mechanisms of translation activation and repression will be required to understand the scope of these regulatory activities.

6.7 *Alu* Elements as Source and Target of miRNAs

Alu elements contribute to the synthesis of miRNAs and are also targets of miRNAs; these two issues will be discussed in the following.

miRNAs may be transcribed as part of an *Alu* element transcription unit, when located between the *Alu* element and the termination signal of Pol III (Fig. 6.2). This mechanism was proposed to account for the production of miRNAs from the miRNA cluster located on chromosome 19 (C19MC), which contains more than 40 miRNAs interspersed with *Alu* elements. It was shown that a functional miRNA was expressed from a plasmid containing a C19MC *Alu* element followed by the miRNA sequence when transfecting into A549 cells (Borchert et al. 2006). Moreover, a CHIP assay carried out with Hela and 293T cells showed the presence of Pol III about 300 nt upstream of three miRNAs belonging to the C19MC cluster. Whereas these studies provided a proof of principle, further studies on the same cluster in a physiological context gave different results. In placental JEG3 cells, where C19MC miRNAs are normally expressed, introns of one or several newly

identified noncoding RNAs were found to be at the origin of C19MC miRNA expression. The treatment with α-amanitin abolished the presence of the miRNA precursor, and Pol II was shown to be present at the C19MC cluster consistent with the interpretation that expression of these noncoding RNAs is Pol II-dependent. Moreover, in JPEG3 cells depleted of Drosha/DGCR8, an increase in intron-containing RNA species comprising the miRNAs was observed, as expected, if these introns are the precursor of the miRNAs (Bortolin-Cavaille et al. 2009).

Another recent study provided new evidence for *Alu*-driven expression of miRNAs by Pol III. Additional 60 putative miRNAs were identified within a region of 200 bp downstream of an *Alu* element and before a Pol III termination sequence using bioinformatic analyses (Gu et al. 2009). High-throughput sequencing confirmed the expression of 24 miRNAs among them, and the expression levels of three miRNAs were increased, when *Alu* expression in Hela and 293T cells was induced by stress such as treatment with cycloheximide and heat shock, providing evidence for their coordinated expression.

Alu elements can also be targets of miRNAs. Two in silico screens revealed the existence of about 30 and 53 miRNAs, respectively, with 5′ seed complementarities against *Alu* sequences embedded in the 3′ UTR of mRNAs (Daskalova et al. 2006; Smalheiser and Torvik 2006). Notably, the different target sequences of the miRNAs are highly conserved between *Alu* elements of different subfamilies. In addition, miRNA-targeted mRNAs are enriched in *Alu* sequences in sense orientation as compared to antisense orientation (Lehnert et al. 2009), suggesting a mechanism of regulation specific to sense *Alu* elements. It was proposed that SRP9/14 binding to the *Alu* sequences in sense orientation might prevent miRNA-induced degradation and translational repression. A hypothesis was postulated in which increased *Alu* RNA levels during stress would compete for SRP9/14-binding rendering the embedded *Alu* sequences accessible for miRNAs (Fig. 6.2, Daskalova et al. 2006).

Analysis of copy number and length variation of the miRNA family on chromosome 19 in primates indicated that the expansion of miRNAs in this cluster occurred through segmental duplications facilitated by *Alu* element expansion (Zhang et al. 2008). As mentioned before, miRNAs target *Alu* RNAs embedded into mRNAs in sense orientation. A recent in silico study revealed a statistically significant overrepresentation of miRNAs in the C19MC cluster that target sense *Alu* RNAs (Lehnert et al. 2009). In conclusion of these studies, a model was proposed in which *Alu* elements facilitate expansion of miRNA segments on one hand, whereas miRNAs derived from the amplified region target *Alu* RNA for degradation on the other, thereby reducing the duplication activities of the *Alu* elements. This dual relationship would protect the genome against excessive proliferation of *Alu* elements, which is ultimately deleterious for the organism. Interestingly, *Alu* RNAs as well as miRNAs from the C19MC cluster are highly expressed in placenta and testis. It is conceivable that C19MC-derived miRNAs may play a role in *Alu* surveillance in these tissues (Lehnert et al. 2009). This hypothesis is also in agreement with the notion that "smart" retrotransposons are able to control their amplification to conserve the viability of the host genome (for review, see Deininger and Batzer 2002).

6.8 *Alu* RNA Editing and Nuclear Retention

A comparison of sequences available in human EST and cDNA libraries with genomic sequences revealed that a great majority of editing sites in the human transcriptome are located in RNA duplexes formed by two *Alu* sequences present in inverted orientation in the same transcript (IR*Alu*, Fig. 6.2, Blow et al. 2004; Kim et al. 2004; Levanon et al. 2004; Morse et al. 2002). These inverted repeats may be present in introns, UTRs, and noncoding poly(A) RNAs (Blow et al. 2004; Chen et al. 2008; Levanon et al. 2004; Morse et al. 2002). More than a thousand transcripts are potentially edited (Athanasiadis et al. 2004; Kim et al. 2004; Levanon et al. 2004).

Editing consists in changing one nucleotide within a double-stranded RNA region. In the case of *Alu* sequences, a member of the enzyme family adenosine-deaminase acting on RNA (ADAR) converts adenosine into inosine (A-to-I, Fig. 6.2). A-to-I conversions in the coding region may change the genetic code since inosine is read as guanosine by the translation machinery (for review, see Bass 2002). The optimum distance between two *Alu* sequences for efficient editing is 300–400 nt, although this requirement is not very stringent (Athanasiadis et al. 2004; Blow et al. 2004). Tissues with frequently edited RNAs are thymus, brain, pancreas, spleen, trachea, kidney, and prostate (Athanasiadis et al. 2004; Kim et al. 2004; Levanon et al. 2004). Notably, human cancer cells have low editing frequencies, and mRNA levels of three ADAR family members were found to be reduced in brain tumors as compared to normal tissues. In addition, overexpression of these enzymes in glioblastoma-derived cells decreased the proliferation rate (Paz et al. 2007), suggesting that editing is incompatible with fast growth.

Interestingly, recent studies suggest a novel mechanism for regulating gene expression involving RNA editing. When IR*Alu* was inserted downstream of a reporter gene (EGFP), the transcript became heavily edited and accumulated in the nucleus (Chen et al. 2008), resulting in a decrease in the expression level of the protein. Hyperedited mRNAs are associated in the nucleus with $p54^{nrb}$ (Chen et al. 2008; Prasanth et al. 2005; Zhang and Carmichael 2001), PSF, matrin 3 (Zhang and Carmichael 2001), and PSP1α (Fig. 6.2, Prasanth et al. 2005). $p54^{nrb}$ shows a high affinity for inosine-containing mRNAs (Zhang and Carmichael 2001). This protein as well as PSF and PSP1α are known to localize in paraspeckles, newly identified nuclear structures (Fox et al. 2002). However, binding to $p54^{nrb}$ is not sufficient for nuclear retention of these mRNAs. In embryonic stem cells, where editing occurs, IR*Alu*-containing mRNAs bound to $p54^{nrb}$ are efficiently transported to the cytoplasm. IR*Alu* retention probably requires the presence of paraspeckles, the formation of which is dependent on hNEAT, a noncoding RNA. hNEAT RNA is absent in stem cells, and the formation of paraspeckles is therefore abrogated. Consistently, depletion of hNEAT from HeLa cells lead to the dissolution of paraspeckles and to an increase of IR*Alu*-containing mRNAs in the cytoplasm (Chen and Carmichael 2009).

Hence, editing of *Alu* sequences can be used as a mechanism to regulate expression of mRNAs with embedded *Alu* sequences. Editing-mediated repression

might be an important regulatory mechanism in brain tissue and in response to physiological signals and stress as suggested by the high expression levels of ADAR enzymes in brain and after stress (for example, see Paul and Bass 1998; Rabinovici et al. 2001; Yang et al. 2003) and by the over representation of IR*Alu* in the mRNAs of zinc finger transcription factors and apoptosis-related proteins (Chen et al. 2008).

The murine gene of the cationic amino acid transporter 2 (*mCAT2*) produces two mRNAs, mCAT2 and CTN-RNA (*mCAT2* transcribed nuclear RNA), which share the same coding region but differ in their 5′ and 3′ UTRs. Due to the presence of inverted B1 repeats, the CTN-RNA has the possibility to form an intramolecular duplex. CTN-RNA is edited by ADAR and is specifically retained in the nucleus. Upon stress, CTN-RNA is cleaved in its 3′ UTR and the truncated transcript is exported to the cytoplasm where it will produce the protein mCAT2 (Prasanth et al. 2005). Thus, nuclear retention and cleavage may provide a mechanism ensuring the fast production of the mCAT2 protein in response to various stress conditions. The mCAT2 protein is a plasma membrane transporter involved in the cellular uptake of arginine, which is a precursor for nitric oxide (NO) production. NO production is increased in response to stress and induces a cellular defense mechanism to overcome stress (for review, see Lowenstein and Padalko 2004).

To understand whether editing followed by the truncation of the mRNA is a widespread mechanism, a large-scale sequence analysis was undertaken to identify "noncanonical" introns (Osenberg et al. 2009). Noncanonical introns were defined as gaps in the alignment of mRNA sequences with the human genome sequence that cannot be explained by splicing. Presumably, these gaps would be the result of a cleavage in the edited RNA followed by ligation of the cleaved 5′ and 3′ fragments of the RNA (Fig. 6.2). Over 500 noncanonical introns containing IR*Alu* were found and were named putatively cleaved segments (PCSs). Consistent with its high editing activity, PCSs containing mRNAs are overrepresented in brain tissue. The mRNAs produced from two PCS-containing genes were chosen to validate the hypothesis. Sequencing of the cDNAs demonstrated that they lacked most of the IR*Alu* portion and that the putative cleavage sites were found at diverse positions in the *Alu* sequences producing mRNAs of heterogeneous sizes. Further studies are required to determine to what extent PCS-containing RNAs use nuclear retention as a regulatory mechanism. Removing segments from mRNAs might affect the metabolism of these transcripts in many different ways such as removing miRNA target sites and protein-binding sites important for translational regulation and RNA stability.

It should also be mentioned that not all edited IR*Alu*-containing transcripts are retained in the nucleus. At least two transcripts with IR*Alus* were found to be associated with polyribosomes in the cytoplasm (Hundley et al. 2008). In addition to nuclear retention, editing itself might affect mRNAs in different ways. Editing could be a way to mark aberrant, nonfunctional transcripts in order to keep them in the nucleus or to degrade them (Kim et al. 2004; Zhang and Carmichael 2001). It may also regulate alternative splicing (Athanasiadis et al. 2004; Morse et al. 2002; Rueter et al. 1999), exonization (Lev-Maor et al. 2007), stability of dsRNA

(Levanon et al. 2004; Serra et al. 2004), and could act as a protection against RNAi (Scadden and Smith 2001; Tonkin and Bass 2003).

6.9 Unique *Alu*-Derived Transcription Units

Insertion of *Alu* elements into specific loci may result in the expression of unique noncoding RNAs such as BC200, NDM29, and 21A RNAs with specific expression patterns and functions.

The BC200 RNA gene arose from an *Alu* left arm element after the divergence of anthropoids from prosimians (Kuryshev et al. 2001; Martignetti and Brosius 1993). It has a length of about 200 bp and can be divided into three different domains. Its 5′ part (120 bp) is homologous to the *Alu* left monomer and contains box A and box B of the Pol III promoter and the binding site of SRP9/14. The central part is composed of an A-rich region. The 3′ end contains a unique sequence of 42 nt followed by a cytosine-rich tract and the Pol III termination sequence (Fig. 6.1, Tiedge et al. 1993; Watson and Sutcliffe 1987).

Although more than 200 pseudogenes exist, the primate-specific BC200 RNA is expressed by Pol III specifically in neurons from a single gene and it localizes to the somatodendritic region (Cheng et al. 1997; Kuryshev et al. 2001; Martignetti and Brosius 1993; Skryabin et al. 1998; Tiedge et al. 1993; Watson and Sutcliffe 1987).

A weak expression of BC200 RNA was also detected in testis (Tiedge et al. 1993). Moreover, it is expressed in lung and breast carcinomas (Chen et al. 1997) and was therefore proposed as a marker for diagnosis and progression of breast cancer (Iacoangeli et al. 2004).

The dissection of the physiological functions of BC200 RNA has been hampered by the absence of an in vivo model. However, BC200 RNA has often been compared to the murine BC1 RNA. Although the genes have different origins, the BC1 gene originates from a tRNA gene, BC200, and BC1 RNAs share the same neuron-specific expression, the same subcellular distribution, and most of their partner proteins suggesting that they are functional analogs (DeChiara and Brosius 1987; Tiedge et al. 1991, 1993). BC1 knockout mice are healthy and show no major neurological abnormalities (Skryabin et al. 2003). Studies in these mice revealed that BC1 RNA contributes to the neuronal excitation–repression balance in dendrites (Zhong et al. 2009). A group of mRNAs, which is stimulated by the metabotropic glutamate receptor group I for protein synthesis, is in turn translationally repressed by BC1 RNA.

BC1 RNA represses translation by preventing the formation of the 48S complex (Wang et al. 2002). BC1 and BC200 RNAs inhibit translation in vitro and in vivo by sequestering PABP (Kondrashov et al. 2005). Further studies on the mechanism revealed that BC1 and BC200 RNAs bind directly to the initiation factor 4A and inhibit its unwinding activity while concomitantly stimulating its ATPase activity. BC1 RNA did not inhibit translation of mRNAs without secondary structures in their 5′ UTR. About 80% of the translation inhibitory activities of BC1 and BC200

RNAs are explained by the inactivation of 4A, and only 20% can be attributed to the binding of PABP (Lin et al. 2008; Wang et al. 2002, 2005). In summary, it is conceivable that BC1 and BC200 RNAs inhibit translation initiation by forming an inactive complex of initiation factors comprising 4A, 4G, and PABP (Fig. 6.3). It remains uncertain whether this complex may also contain mRNA bound to 4G via the cap-binding protein 4E. BC1 RNA also binds 4B, which synergistically stimulates 4A binding. 4B is phosphorylated by kinases of several signaling pathways such as the kinase S6 activated by the TOR pathway. Its phosphorylation may serve to modulate the inhibitory activity of BC1 RNA, and possibly also of BC200 RNA, in translation.

SRP9/14-binding sites are well conserved in BC200 RNA (Fig. 6.3), and the protein indeed binds to BC200 RNA in vitro and in vivo (Bovia et al. 1997; Kremerskothen et al. 1998). SRP9/14 may play a role in the stabilization and/or the nuclear export of BC200 RNA (Fig. 6.3) as described before for *Alu* and 7SL RNAs. In analogy to *Alu* RNP (Hasler and Strub 2006), the inhibitory activity of BC200 RNA in translation may be enhanced by SRP9/14 binding.

Another conceivable role for BC200 and BC1 RNAs is the transport of mRNAs to dendrites (Fig. 6.3). It is established that some neuronal mRNAs are transported in a translationally inactive state to dendrites where they will become activated. This transport, which involves large RNPs, is mediated by kinesin motors and microtubules (as rewieved by Bramham and Wells 2007). Interestingly, BC200 RNA and BC1 RNA have partial complementarities with the dendritic mRNAs *Arc, α-CaMKII, and MAP1B* (Zalfa et al. 2003). Moreover, BC200 RNA was demonstrated to bind to Purα, fragile X mental retardation 1 protein (FMRP), and SYNCRIP (Duning et al. 2008; Johnson et al. 2006; Zalfa et al. 2003). These proteins localize together with the Arc and α-CaMKII mRNAs to large RNA granules in mouse brain, which are transported in a kinesin- and microtubule-dependent manner to dendrites (Kanai et al. 2004). The Purα protein was shown to bind to an annealed complex of BC200 RNA and MAP1B mRNA in vitro (Johnson et al. 2006) and might therefore link BC200 RNA to microtubules, as it was shown for BC1 RNA (Ohashi et al. 2000).

In addition, because of its binding to the *N*-terminal domain of FMRP, BC200 RNA might prevent FMRP from entering the nucleus by sequestering the nuclear localization signal present in this region (Fig. 6.3, Zalfa et al. 2005). BC200 RNA might therefore interfere with FMRP-mediated putative nuclear functions such as mRNA export (Kim et al. 2009).

Thus, BC200 RNA is tightly linked to the translation of specific mRNAs in dendrites and might also play a role in mRNA transport. In the future, studies on the role of BC200 RNA in neuronal plasticity will be facilitated by the availability of transgenic mice expressing BC200 RNA in a pattern similar to the one observed in primates (Khanam et al. 2007).

Recent work on two unique *Alu* RNAs, NMD29 RNA (neuronal differentiation marker 29 RNA) and 21A RNA, offers a new perspective on the potential roles of specific *Alu* RNAs. These *Alu* elements were discovered using the proximal sequence element (PSE) sequence of the *H1* gene promoter to search the human

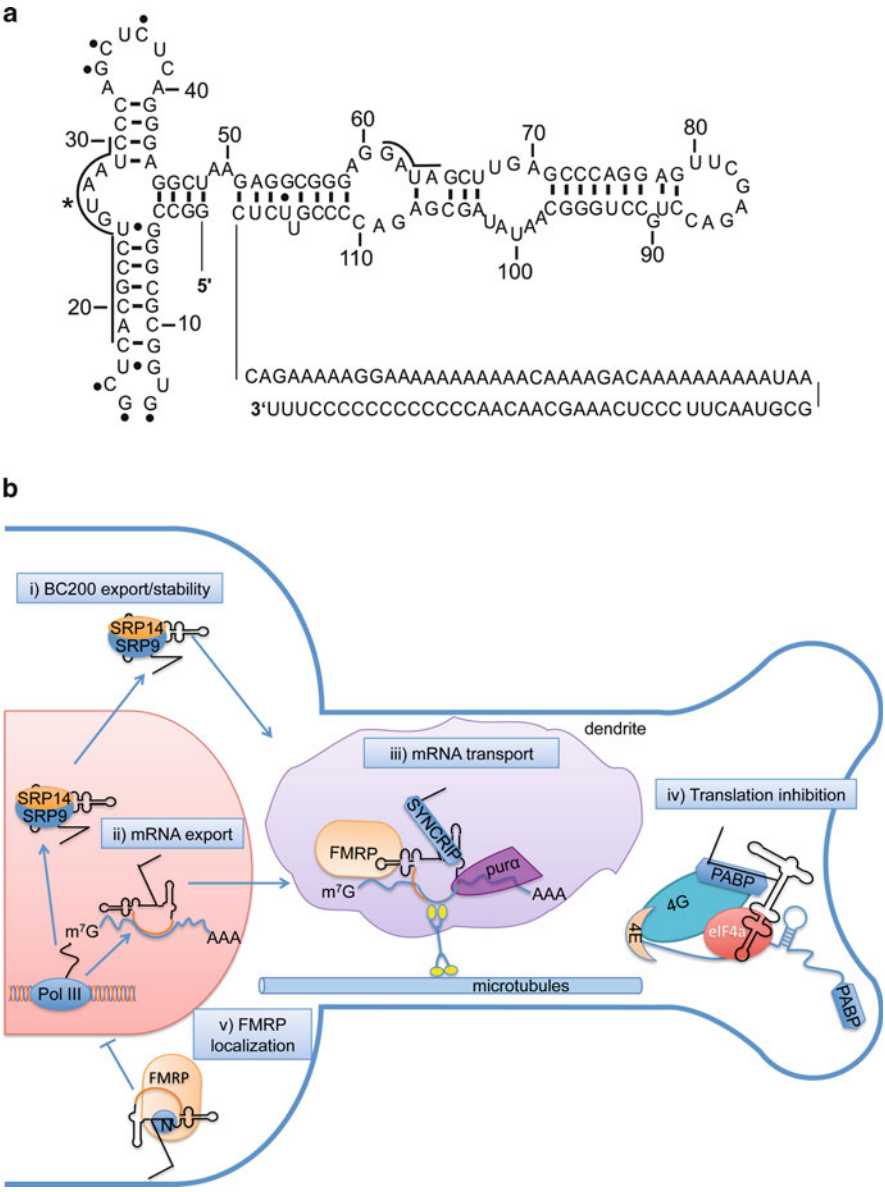

Fig. 6.3 Hypothetical model for BC200 RNA functions in neurons. (**a**) Secondary structure of BC200 RNA as predicted in analogy to the Alu domain of 7SL RNA. The A-rich and the 3′ unique sequences are shown unfolded. *Lines*: SRP9/14 binding sites. *Dots*: Base pairs between the loops; *star*: U-turn in RNA structure. (**b**) Proposed functions of BC200 RNA. (1) After synthesis by Pol III, SRP9/14 binds BC200 RNA (in *black*) in the nucleus to ensure proper folding and, possibly, nuclear export. (2) BC200 RNA partially anneals (*orange*) to complementary sequences in specific mRNAs and facilitates their export from the nucleus. (3) In the cytoplasm, BC200 RNA participates in the transport of mRNAs to dendrites together with other proteins in large granules

genome for new Pol III transcription units (Pagano et al. 2007). The transcription of these two *Alu* elements by Pol III is dependent on two upstream sequence elements, DSE and PSE, and on the TATA box. They do not contain box A and box B. The neuronal NDM29 transcription unit contains a unique sequence at its 5′ end followed by an *Alu* element belonging to the *Alu*Jb subfamily (Fig. 6.1), and NMD29 RNA levels are over 100-fold higher in the neuronal SHSY5Y and SKNBe cell lines as compared to HeLa and HEK 293T cell lines. NMD29 RNA was found to play a role in cell proliferation. Increasing the levels of NMD29 RNA in neuroblastoma cell lines slowed down cell proliferation and promoted cell differentiation as indicated by the appearance of differentiation markers. Furthermore, cells from neuroblastoma cell lines with increased NMD29 RNA levels had a lower potential to form tumors in mice. NMD29 RNA is found in the cytoplasm suggesting that it might affect cell proliferation via a posttranscriptional regulatory mechanism, which remains to be elucidated (Castelnuovo et al. 2010).

The 21A RNA transcription unit also belongs to the *Alu*Jb subfamily (Fig. 6.1). The 21A RNA shows partially sequence complementarities to regions in three introns of the centromere protein F gene (*CEN-F*). The centromere protein F is essential for kinetochore function and chromosome segregation. When expression levels of 21A RNA are increased, the levels of the *CENP-F* mRNA and of the protein are reduced, resulting in a slowdown of cell proliferation. In addition, the high proliferation rates of tumor cells correlate with low 21A RNA levels, consistent with a role of 21A RNA in growth control. Although the mechanism remains to be elucidated, it was suggested that 21A RNA might affect cell proliferation via an antisense mechanism targeting CEN-F mRNA (Pagano et al. 2007).

6.10 Conclusion

Over the past 20 years, it has become evident that *Alu* RNAs and *Alu*-like RNAs have been recruited for multiple functions in primate and, to a lesser extent, in rodent species. It is quite likely that still more functions will be discovered in the future for *Alu* RNA and for unique *Alu*-containing noncoding RNAs with specific expression profiles. At this point, still substantial work is required to understand the mechanisms by which *Alu* and *Alu*-containing RNAs accomplish their functions and the putative regulatory circuits that may, in turn, modulate or control *Alu* RNA functions. One way to approach these questions is to get an extensive view on "the life cycle" of *Alu* RNAs by identifying their partner proteins and the cellular

Fig. 6.3 (Continued) (*purple*). The transport is kinesin- and microtubule-dependent. (4) At dendrites, BC200 RNA inhibits translation of mRNAs containing secondary structures in the 5′ UTR by blocking the unwinding function of eIF4A and by binding PABP. (5) BC200 RNA binding to FMRP interferes with the nuclear localization and, thus, also with nuclear functions of FMRP

structures in which they fulfill their activities. *Alu* element proliferation has not only shaped our genome, but *Alu* RNAs are also associated with human disease (Castelnuovo et al. 2010; Kiesel et al. 2010; Mus et al. 2007; Pagano et al. 2007), and the elucidation of the molecular mechanisms is expected to improve our understanding of these diseases.

Acknowledgments We would like to thank Dr. F. Stutz for her comments on the manuscript. This work was supported by grants from the Swiss National Science Foundation and the Canton of Geneva. A.B. was a recipient of a fellowship from the Roche Research Foundation.

References

Adeniyi-Jones S, Zasloff M (1985) Transcription, processing and nuclear transport of a B1 Alu RNA species complementary to an intron of the murine alpha-fetoprotein gene. Nature 317:81–84

Aleman C, Roy-Engel AM, Shaikh TH, Deininger PL (2000) Cis-acting influences on Alu RNA levels. Nucleic Acids Res 28:4755–4761

Allen TA, Von Kaenel S, Goodrich JA, Kugel JF (2004) The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat Struct Mol Biol 11:816–821

Athanasiadis A, Rich A, Maas S (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol 2:e391

Bass BL (2002) RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem 71:817–846

Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379

Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E (1996) Standardized nomenclature for Alu repeats. J Mol Evol 42:3–6

Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE (2008) Active Alu retrotransposons in the human genome. Genome Res 18:1875–1883

Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res 8:1499–1504

Blow M, Futreal PA, Wooster R, Stratton MR (2004) A survey of RNA editing in human brain. Genome Res 14:2379–2387

Borchert GM, Lanier W, Davidson BL (2006) RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol 13:1097–1101

Bortolin-Cavaille ML, Dance M, Weber M, Cavaille J (2009) C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. Nucleic Acids Res 37:3464–3473

Bovia F, Fornallaz M, Leffers H, Strub K (1995) The SRP9/14 subunit of the signal recognition particle (SRP) is present in more than 20-fold excess over SRP in primate cells and exists primarily free but also in complex with small cytoplasmic Alu RNAs. Mol Biol Cell 6:471–484

Bovia F, Wolff N, Ryser S, Strub K (1997) The SRP9/14 subunit of the human signal recognition particle binds to a variety of Alu-like RNAs and with higher affinity than its mouse homolog. Nucleic Acids Res 25:318–326

Bramham CR, Wells DG (2007) Dendritic mRNA: transport, translation and function. Nat Rev Neurosci 8:776–789

Bredow S, Surig D, Muller J, Kleinert H, Benecke BJ (1990) Activating-transcription-factor (ATF) regulates human 7S L RNA transcription by RNA polymerase III in vivo and in vitro. Nucleic Acids Res 18:6779–6784

Britten RJ (1994) Evidence that most human Alu sequences were inserted in a process that ceased about 30 million years ago. Proc Natl Acad Sci USA 91:6148–6150

Britten RJ, Baron WF, Stout DB, Davidson EH (1988) Sources and evolution of human Alu repeated sequences. Proc Natl Acad Sci USA 85:4770–4774

Callinan PA, Wang J, Herke SW, Garber RK, Liang P, Batzer MA (2005) Alu retrotransposition-mediated deletion. J Mol Biol 348:791–800

Carey MF, Singh K, Botchan M, Cozzarelli NR (1986) Induction of specific transcription by RNA polymerase III in transformed cells. Mol Cell Biol 6:3068–3076

Castelnuovo M, Massone S, Tasso R, Fiorino G, Gatti M, Robello M, Gatta E, Berger A, Strub K, Florio T, et al (2010) An Alu-like RNA promotes cell differentiation and reduces malignancy of human neuroblastoma cells. FASEB J 24:4033–4046

Chang DY, Maraia RJ (1993) A cellular protein binds B1 and Alu small cytoplasmic RNAs in vitro. J Biol Chem 268:6423–6428

Chang DY, Nelson B, Bilyeu T, Hsu K, Darlington GJ, Maraia RJ (1994) A human Alu RNA-binding protein whose expression is associated with accumulation of small cytoplasmic Alu RNA. Mol Cell Biol 14:3949–3959

Chang DY, Sasaki-Tozawa N, Green LK, Maraia RJ (1995) A trinucleotide repeat-associated increase in the level of Alu RNA-binding protein occurred during the same period as the major Alu amplification that accompanied anthropoid evolution. Mol Cell Biol 15:2109–2116

Chang DY, Hsu K, Maraia RJ (1996) Monomeric scAlu and nascent dimeric Alu RNAs induced by adenovirus are assembled into SRP9/14-containing RNPs in HeLa cells. Nucleic Acids Res 24:4165–4170

Chen LL, Carmichael GG (2009) Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell 35:467–478

Chen W, Bocker W, Brosius J, Tiedge H (1997) Expression of neural BC200 RNA in human tumours. J Pathol 183:345–351

Chen LL, DeCerbo JN, Carmichael GG (2008) Alu element-mediated gene silencing. EMBO J 27:1694–1705

Cheng JG, Tiedge H, Brosius J (1997) Expression of dendritic BC200 RNA, component of a 11.4S ribonucleoprotein particle, is conserved in humans and simians. Neurosci Lett 224:206–210

Chesnokov IN, Schmid CW (1995) Specific Alu binding protein from human sperm chromatin prevents DNA methylation. J Biol Chem 270:18539–18542

Chesnokov I, Schmid CW (1996) Flanking sequences of an Alu source stimulate transcription in vitro by interacting with sequence-specific transcription factors. J Mol Evol 42:30–36

Chesnokov I, Chu WM, Botchan MR, Schmid CW (1996) p53 inhibits RNA polymerase III-directed transcription in a promoter-dependent manner. Mol Cell Biol 16:7084–7088

Cho NY, Kim BH, Choi M, Yoo EJ, Moon KC, Cho YM, Kim D, Kang GH (2007) Hypermethylation of CpG island loci and hypomethylation of LINE-1 and Alu repeats in prostate adenocarcinoma and their relationship to clinicopathological features. J Pathol 211:269–277

Choi IS, Estecio MR, Nagano Y, Kim do H, White JA, Yao JC, Issa JP, Rashid A (2007) Hypomethylation of LINE-1 and Alu in well-differentiated neuroendocrine tumors (pancreatic endocrine tumors and carcinoid tumors). Mod Pathol 20:802–810

Chu WM, Liu WM, Schmid CW (1995) RNA polymerase III promoter and terminator elements affect Alu RNA expression. Nucleic Acids Res 23:1750–1757

Chu WM, Wang Z, Roeder RG, Schmid CW (1997) RNA polymerase III transcription repressed by Rb through its interactions with TFIIIB and TFIIIC2. J Biol Chem 272:14755–14761

Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW (1998) Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. Mol Cell Biol 18:58–68

Cordaux R, Hedges DJ, Batzer MA (2004) Retrotransposition of Alu elements: how many sources? Trends Genet 20:464–467
Cordaux R, Hedges DJ, Herke SW, Batzer MA (2006) Estimating the retrotransposition rate of human Alu elements. Gene 373:134–137
Daniels GR, Deininger PL (1983) A second major class of Alu family repeated DNA sequences in a primate genome. Nucleic Acids Res 11:7595–7610
Daniels GR, Deininger PL (1985) Repeat sequence families derived from mammalian tRNA genes. Nature 317:819–822
Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, Kotsinas A, Gorgoulis V, Field JK, Liloglou T (2009) Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. Int J Cancer 124:81–87
Daskalova E, Baev V, Rusinov V, Minkov I (2006) 3′UTR-located ALU elements: donors of potential miRNA target sites and mediators of network miRNA-based regulatory interactions. Evol Bioinform Online 2:103–120
DeChiara TM, Brosius J (1987) Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. Proc Natl Acad Sci USA 84:2624–2628
Deininger PL, Batzer MA (2002) Mammalian retroelements. Genome Res 12:1455–1465
Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J Mol Biol 151:17–33
Deininger PL, Batzer MA, Hutchison CA 3rd, Edgell MH (1992) Master genes in mammalian repetitive DNA amplification. Trends Genet 8:307–311
Dewannieux M, Esnault C, Heidmann T (2003) LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 35:41–48
Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr (1991) Isolation of an active human transposable element. Science 254:1805–1808
Duning K, Buck F, Barnekow A, Kremerskothen J (2008) SYNCRIP, a component of dendritically localized mRNPs, binds to the translation regulator BC200 RNA. J Neurochem 105:351–359
Elder JT, Pan J, Duncan CH, Weissman SM (1981) Transcriptional analysis of interspersed repetitive polymerase III transcription units in human DNA. Nucleic Acids Res 9:1171–1189
Espinoza CA, Allen TA, Hieb AR, Kugel JF, Goodrich JA (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nat Struct Mol Biol 11:822–829
Espinoza CA, Goodrich JA, Kugel JF (2007) Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. RNA 13:583–596
Feng Q, Moran JV, Kazazian HH Jr, Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905–916
Fox AH, Lam YW, Leung AK, Lyon CE, Andersen J, Mann M, Lamond AI (2002) Paraspeckles: a novel nuclear domain. Curr Biol 12:13–25
Fuhrman SA, Deininger PL, LaPorte P, Friedmann T, Geiduschek EP (1981) Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. Nucleic Acids Res 9:6439–6456
Garcia MA, Meurs EF, Esteban M (2007) The dsRNA protein kinase PKR: virus and cell control. Biochimie 89:799–811
Goodier JL, Maraia RJ (1998) Terminator-specific recycling of a B1-Alu transcription complex by RNA polymerase III is mediated by the RNA terminus-binding protein La. J Biol Chem 273:26110–26116
Goodier JL, Mandal PK, Zhang L, Kazazian HH Jr (2010) Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. Hum Mol Genet 19:1712–1725
Gu TJ, Yi X, Zhao XW, Zhao Y, Yin JQ (2009) Alu-directed transcriptional regulation of some novel miRNAs. BMC Genomics 10:563
Hagan CR, Rudin CM (2007) DNA cleavage and Trp53 differentially affect SINE transcription. Genes Chromosom Cancer 46:248–260

Hagan CR, Sheffield RF, Rudin CM (2003) Human Alu element retrotransposition induced by genotoxic stress. Nat Genet 35:219–220

Harada F, Kato N (1980) Nucleotide sequences of 4.5S RNAs associated with poly(A)-containing RNAs of mouse and hamster cells. Nucleic Acids Res 8:1273–1285

Harada F, Kato N, Hoshino H (1979) Series of 4.5S RNAs associated with poly(A)-containing RNAs of rodent cells. Nucleic Acids Res 7:909–917

Hasler J, Strub K (2006) Alu RNP and Alu RNA regulate translation initiation in vitro. Nucleic Acids Res 34:2374–2385

Haynes SR, Toomey TP, Leinwand L, Jelinek WR (1981) The Chinese hamster Alu-equivalent sequence: a conserved highly repetitious, interspersed deoxyribonucleic acid sequence in mammals has a structure suggestive of a transposable element. Mol Cell Biol 1:573–583

He XP, Bataille N, Fried HM (1994) Nuclear export of signal recognition particle RNA is a facilitated process that involves the Alu sequence domain. J Cell Sci 107:903–912

Hellmann-Blumberg U, Hintz MF, Gatewood JM, Schmid CW (1993) Developmental differences in methylation of human Alu repeats. Mol Cell Biol 13:4523–4530

Hess J, Perez-Stable C, Wu GJ, Weir B, Tinoco I Jr, Shen CK (1985) End-to-end transcription of an Alu family repeat. A new type of polymerase-III-dependent terminator and its evolutionary implication. J Mol Biol 184:7–21

Hsu K, Chang DY, Maraia RJ (1995) Human signal recognition particle (SRP) Alu-associated protein also binds Alu interspersed repeat sequence RNAs. Characterization of human SRP9. J Biol Chem 270:10179–10186

Hundley HA, Krauchuk AA, Bass BL (2008) *C. elegans* and *H. sapiens* mRNAs with edited 3′ UTRs are present on polysomes. RNA 14:2050–2060

Iacoangeli A, Lin Y, Morley EJ, Muslimov IA, Bianchi R, Reilly J, Weedon J, Diallo R, Bocker W, Tiedge H (2004) BC200 RNA in invasive and preinvasive breast cancer. Carcinogenesis 25:2125–2133

Jang KL, Latchman DS (1989) HSV infection induces increased transcription of Alu repeated sequences by RNA polymerase III. FEBS Lett 258:255–258

Jang KL, Latchman DS (1992) The herpes simplex virus immediate-early protein ICP27 stimulates the transcription of cellular Alu repeated sequences by increasing the activity of transcription factor TFIIIC. Biochem J 284:667–673

Jang KL, Collins MK, Latchman DS (1992) The human immunodeficiency virus tat protein increases the transcription of human Alu repeated sequences by increasing the activity of the cellular transcription factor TFIIIC. J Acquir Immune Defic Syndr 5:1142–1147

Jelinek WR, Schmid CW (1982) Repetitive sequences in eukaryotic DNA and their expression. Annu Rev Biochem 51:813–844

Johnson EM, Kinoshita Y, Weinreb DB, Wortman MJ, Simon R, Khalili K, Winckler B, Gordon J (2006) Role of Pur alpha in targeting mRNA to sites of translation in hippocampal neuronal dendrites. J Neurosci Res 83:929–943

Jurka J (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA 94:1872–1877

Jurka J, Milosavljevic A (1991) Reconstruction and analysis of human Alu genes. J Mol Evol 32:105–121

Jurka J, Smith T (1988) A fundamental division in the Alu family of repeated sequences. Proc Natl Acad Sci USA 85:4775–4778

Kanai Y, Dohmae N, Hirokawa N (2004) Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. Neuron 43:513–525

Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 11:345–355

Khanam T, Muddashetty RS, Kahvejian A, Sonenberg N, Brosius J (2006) Poly(A)-binding protein binds to A-rich sequences via RNA-binding domains 1+2 and 3+4. RNA Biol 3:170–177

Khanam T, Rozhdestvensky TS, Bundman M, Galiveti CR, Handel S, Sukonina V, Jordan U, Brosius J, Skryabin BV (2007) Two primate-specific small non-protein-coding RNAs in

transgenic mice: neuronal expression, subcellular localization and binding partners. Nucleic Acids Res 35:529–539
Kiesel P, Gibson TJ, Ciesielczyk B, Bodemer M, Kaup FJ, Bodemer W, Zischler H, Zerr I (2010) Transcription of Alu DNA elements in blood cells of sporadic Creutzfeldt-Jakob disease (sCJD). Prion 4(2):87–93
Kim C, Rubin CM, Schmid CW (2001) Genome-wide chromatin remodeling modulates the Alu heat shock response. Gene 276:127–133
Kim DD, Kim TT, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A (2004) Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res 14:1719–1725
Kim M, Bellini M, Ceman S (2009) Fragile X mental retardation protein FMRP binds mRNAs in the nucleus. Mol Cell Biol 29:214–228
Kochanek S, Renz D, Doerfler W (1993) DNA methylation in the Alu sequences of diploid and haploid primary human cells. EMBO J 12:1141–1151
Kochanek S, Renz D, Doerfler W (1995) Transcriptional silencing of human Alu sequences and inhibition of protein binding in the box B regulatory elements by 5′-CG-3′ methylation. FEBS Lett 360:115–120
Kondrashov AV, Kiefmann M, Ebnet K, Khanam T, Muddashetty RS, Brosius J (2005) Inhibitory effect of naked neural BC1 RNA or BC200 RNA on eukaryotic in vitro translation systems is reversed by poly(A)-binding protein (PABP). J Mol Biol 353:88–103
Kramerov DA, Tillib SV, Shumyatsky GP, Georgiev GP (1990) The most abundant nascent poly (A) + RNAs are transcribed by RNA polymerase III in murine tumor cells. Nucleic Acids Res 18:4499–4506
Kremerskothen J, Zopf D, Walter P, Cheng JG, Nettermann M, Niewerth U, Maraia RJ, Brosius J (1998) Heterodimer SRP9/14 is an integral part of the neural BC200 RNP in primate brain. Neurosci Lett 245:123–126
Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J (2007) Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. Trends Genet 23:158–161
Kuryshev VY, Skryabin BV, Kremerskothen J, Jurka J, Brosius J (2001) Birth of a gene: locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropoidea lineage. J Mol Biol 309:1049–1066
Labuda D, Striker G (1989) Sequence conservation in Alu evolution. Nucleic Acids Res 17:2477–2491
Lakkaraju AK, Mary C, Scherrer A, Johnson AE, Strub K (2008) SRP keeps polypeptides translocation-competent by slowing translation to match limiting ER-targeting sites. Cell 133:440–451
Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
Lehnert S, Van Loo P, Thilakarathne PJ, Marynen P, Verbeke G, Schuit FC (2009) Evidence for co-evolution between human microRNAs and Alu-repeats. PLoS One 4:e4456
Leinwand LA, Wydro RM, Nadal-Ginard B (1982) Small RNA molecules related to the Alu family of repetitive DNA sequences. Mol Cell Biol 2:1320–1330
Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D et al (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol 22:1001–1005
Lev-Maor G, Sorek R, Levanon EY, Paz N, Eisenberg E, Ast G (2007) RNA-editing-mediated exon evolution. Genome Biol 8:R29
Li TH, Schmid CW (2001) Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. Gene 276:135–141
Li TH, Schmid CW (2004) Alu's dimeric consensus sequence destabilizes its transcripts. Gene 324:191–200
Li T, Spearow J, Rubin CM, Schmid CW (1999) Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. Gene 239:367–372

Li TH, Kim C, Rubin CM, Schmid CW (2000) K562 cells implicate increased chromatin accessibility in Alu transcriptional activation. Nucleic Acids Res 28:3031–3039

Lin D, Pestova TV, Hellen CU, Tiedge H (2008) Translational control by a small RNA: dendritic BC1 RNA targets the eukaryotic initiation factor 4A helicase mechanism. Mol Cell Biol 28:3008–3019

Liu WM, Schmid CW (1993) Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation. Nucleic Acids Res 21:1351–1359

Liu WM, Maraia RJ, Rubin CM, Schmid CW (1994) Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. Nucleic Acids Res 22:1087–1095

Liu WM, Chu WM, Choudary PV, Schmid CW (1995) Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. Nucleic Acids Res 23:1758–1765

Lowenstein CJ, Padalko E (2004) iNOS (NOS2) at a glance. J Cell Sci 117:2865–2867

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72:595–605

Maraia RJ (1991) The subset of mouse B1 (Alu-equivalent) sequences expressed as small processed cytoplasmic transcripts. Nucleic Acids Res 19:5695–5702

Maraia RJ (1996) Transcription termination factor La is also an initiation factor for RNA polymerase III. Proc Natl Acad Sci USA 93:3383–3387

Maraia R, Zasloff M, Plotz P, Adeniyi-Jones S (1988) Pathway of B1-Alu expression in microinjected oocytes: *Xenopus laevis* proteins associated with nuclear precursor and processed cytoplasmic RNAs. Mol Cell Biol 8:4433–4440

Maraia RJ, Chang DY, Wolffe AP, Vorce RL, Hsu K (1992) The RNA polymerase III terminator used by a B1-Alu element can modulate 3′ processing of the intermediate RNA product. Mol Cell Biol 12:1500–1506

Maraia RJ, Driscoll CT, Bilyeu T, Hsu K, Darlington GJ (1993) Multiple dispersed loci produce small cytoplasmic Alu RNA. Mol Cell Biol 13:4233–4241

Maraia RJ, Kenan DJ, Keene JD (1994) Eukaryotic transcription termination factor La mediates transcript release and facilitates reinitiation by RNA polymerase III. Mol Cell Biol 14:2147–2158

Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol Cell 29:499–509

Martignetti JA, Brosius J (1993) BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. Proc Natl Acad Sci USA 90:11563–11567

Matera AG, Hellmann U, Schmid CW (1990) A transpositionally and transcriptionally competent Alu subfamily. Mol Cell Biol 10:5424–5432

Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A (1991) Reverse transcriptase encoded by a human transposable element. Science 254:1808–1810

Morse DP, Aruscavage PJ, Bass BL (2002) RNA hairpins in noncoding regions of human brain and *Caenorhabditis elegans* mRNA are edited by adenosine deaminases that act on RNA. Proc Natl Acad Sci USA 99:7906–7911

Muddashetty R, Khanam T, Kondrashov A, Bundman M, Iacoangeli A, Kremerskothen J, Duning K, Barnekow A, Huttenhofer A, Tiedge H et al (2002) Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. J Mol Biol 321:433–445

Mullin C, Duning K, Barnekow A, Richter D, Kremerskothen J, Mohr E (2004) Interaction of rat poly(A)-binding protein with poly(A)- and non-poly(A) sequences is preferentially mediated by RNA recognition motifs 3+4. FEBS Lett 576:437–441

Mus E, Hof PR, Tiedge H (2007) Dendritic BC200 RNA in aging and in Alzheimer's disease. Proc Natl Acad Sci USA 104:10679–10684

Nishihara H, Terai Y, Okada N (2002) Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. Mol Biol Evol 19:1964–1972

Ohashi S, Kobayashi S, Omori A, Ohara S, Omae A, Muramatsu T, Li Y, Anzai K (2000) The single-stranded DNA- and RNA-binding proteins pur alpha and pur beta link BC1 RNA to microtubules through binding to the dendrite-targeting RNA motifs. J Neurochem 75:1781–1790

Osenberg S, Dominissini D, Rechavi G, Eisenberg E (2009) Widespread cleavage of A-to-I hyperediting substrates. RNA 15:1632–1639

Pagano A, Castelnuovo M, Tortelli F, Ferrari R, Dieci G, Cancedda R (2007) New small nuclear RNA gene-like transcriptional units as sources of regulatory transcripts. PLoS Genet 3:e1

Panning B, Smiley JR (1993) Activation of RNA polymerase III transcription of human Alu repetitive elements by adenovirus type 5: requirement for the E1b 58-kilodalton protein and the products of E4 open reading frames 3 and 6. Mol Cell Biol 13:3231–3244

Panning B, Smiley JR (1994) Activation of RNA polymerase III transcription of human Alu elements by herpes simplex virus. Virology 202:408–417

Paolella G, Lucero MA, Murphy MH, Baralle FE (1983) The Alu family repeat promoter has a tRNA-like bipartite structure. EMBO J 2:691–696

Paul MS, Bass BL (1998) Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. EMBO J 17:1120–1127

Paulson KE, Schmid CW (1986) Transcriptional inactivity of Alu repeats in HeLa cells. Nucleic Acids Res 14:6145–6158

Paz N, Levanon EY, Amariglio N, Heimberger AB, Ram Z, Constantini S, Barbash ZS, Adamsky K, Safran M, Hirschberg A et al (2007) Altered adenosine-to-inosine RNA editing in human cancer. Genome Res 17:1586–1595

Perez-Stable C, Shen CK (1986) Competitive and cooperative functioning of the anterior and posterior promoter elements of an Alu family repeat. Mol Cell Biol 6:2041–2052

Perez-Stable C, Ayres TM, Shen CK (1984) Distinctive sequence organization and functional programming of an Alu repeat promoter. Proc Natl Acad Sci USA 81:5291–5295

Perlino E, Paonessa G, Ciliberto G (1985) Alu sequences transcription in *X. laevis* oocytes: nuclear-cytoplasmic partitioning and evidence for 3′ end processing reactions. Nucleic Acids Res 13:8359–8377

Prasanth KV, Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL (2005) Regulating gene expression through RNA nuclear retention. Cell 123:249–263

Price AL, Eskin E, Pevzner PA (2004) Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. Genome Res 14:2245–2252

Quentin Y (1989) Successive waves of fixation of B1 variants in rodent lineage history. J Mol Evol 28:299–305

Quentin Y (1992) Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. Nucleic Acids Res 20:487–493

Quentin Y (1994) A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. Nucleic Acids Res 22:2222–2227

Rabinovici R, Kabir K, Chen M, Su Y, Zhang D, Luo X, Yang JH (2001) ADAR1 is involved in the development of microvascular lung injury. Circ Res 88:1066–1071

Rogers JH (1985) The origin and evolution of retroposons. Int Rev Cytol 93:187–279

Roy AM, West NC, Rao A, Adhikari P, Aleman C, Barnes AP, Deininger PL (2000) Upstream flanking sequences and transcription of SINEs. J Mol Biol 302:17–25

Rubin CM, VandeVoort CA, Teplitz RL, Schmid CW (1994) Alu repeated DNAs are differentially methylated in primate germ cells. Nucleic Acids Res 22:5121–5127

Rubin CM, Kimura RH, Schmid CW (2002) Selective stimulation of translational expression by Alu RNA. Nucleic Acids Res 30:3253–3261

Rudin CM, Thompson CB (2001) Transcriptional activation of short interspersed elements by DNA-damaging agents. Genes Chromosom Cancer 30:64–71

Rueter SM, Dawson TR, Emeson RB (1999) Regulation of alternative splicing by RNA editing. Nature 399:75–80

Russanova VR, Driscoll CT, Howard BH (1995) Adenovirus type 2 preferentially stimulates polymerase III transcription of Alu elements by relieving repression: a potential role for chromatin. Mol Cell Biol 15:4282–4290

Sarrowa J, Chang DY, Maraia RJ (1997) The decline in human Alu retroposition was accompanied by an asymmetric decrease in SRP9/14 binding to dimeric Alu RNA and increased expression of small cytoplasmic Alu RNA. Mol Cell Biol 17:1144–1151

Scadden AD, Smith CW (2001) RNAi is antagonized by A–>I hyper-editing. EMBO Rep 2:1107–1111

Schmid CW (1991) Human Alu subfamilies and their methylation revealed by blot hybridization. Nucleic Acids Res 19:5613–5617

Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP, Cooley T, Heath P, Smith KD, Margolet L (1987) Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics 1:113–125

Serra MJ, Smolter PE, Westhof E (2004) Pronounced instability of tandem IU base pairs in RNA. Nucleic Acids Res 32:1824–1828

Shaikh TH, Roy AM, Kim J, Batzer MA, Deininger PL (1997) cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. J Mol Biol 271:222–234

Shen MR, Batzer MA, Deininger PL (1991) Evolution of the master Alu gene(s). J Mol Evol 33:311–320

Siegel V, Walter P (1988) Each of the activities of signal recognition particle (SRP) is contained within a distinct domain: analysis of biochemical mutants of SRP. Cell 52:39–49

Sinnett D, Richer C, Deragon JM, Labuda D (1992) Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. J Mol Biol 226:689–706

Skryabin BV, Kremerskothen J, Vassilacopoulou D, Disotell TR, Kapitonov VV, Jurka J, Brosius J (1998) The BC200 RNA gene and its neural expression are conserved in Anthropoidea (Primates). J Mol Evol 47:677–685

Skryabin BV, Sukonina V, Jordan U, Lewejohann L, Sachser N, Muslimov I, Tiedge H, Brosius J (2003) Neuronal untranslated BC1 RNA: targeted gene elimination in mice. Mol Cell Biol 23:6435–6441

Smalheiser NR, Torvik VI (2006) Alu elements within human mRNAs are probable microRNA targets. Trends Genet 22:532–536

Srikanta D, Sen SK, Conlin EM, Batzer MA (2009a) Internal priming: an opportunistic pathway for L1 and Alu retrotransposition in hominins. Gene 448:233–241

Srikanta D, Sen SK, Huang CT, Conlin EM, Rhodes RM, Batzer MA (2009b) An alternative pathway for Alu retrotransposition suggests a role in DNA double-strand break repair. Genomics 93:205–212

Styles P, Brookfield JF (2009) Source gene composition and gene conversion of the AluYh and AluYi lineages of retrotransposons. BMC Evol Biol 9:102

Tang RB, Wang HY, Lu HY, Xiong J, Li HH, Qiu XH, Liu HQ (2005) Increased level of polymerase III transcribed Alu RNA in hepatocellular carcinoma tissue. Mol Carcinog 42:93–96

Tiedge H, Fremeau RT Jr, Weinstock PH, Arancio O, Brosius J (1991) Dendritic location of neural BC1 RNA. Proc Natl Acad Sci USA 88:2093–2097

Tiedge H, Chen W, Brosius J (1993) Primary structure, neural-specific expression, and dendritic location of human BC200 RNA. J Neurosci 13:2382–2390

Tonkin LA, Bass BL (2003) Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. Science 302:1725

Ullu E, Tschudi C (1984) Alu sequences are processed 7SL RNA genes. Nature 312:171–172

Ullu E, Weiner AM (1985) Upstream sequences modulate the internal promoter of the human 7SL RNA gene. Nature 318:371–374

Vassetzky NS, Ten OA, Kramerov DA (2003) B1 and related SINEs in mammalian genomes. Gene 319:149–160

Veniaminova NA, Vassetzky NS, Kramerov DA (2007) B1 SINEs in different rodent families. Genomics 89:678–686

Wagner SD, Kugel JF, Goodrich JA (2010) TFIIF facilitates dissociation of RNA polymerase II from noncoding RNAs that lack a repression domain. Mol Cell Biol 30:91–97

Wallace N, Wagstaff BJ, Deininger PL, Roy-Engel AM (2008) LINE-1 ORF1 protein enhances Alu SINE retrotransposition. Gene 419:1–6

Wang H, Iacoangeli A, Popp S, Muslimov IA, Imataka H, Sonenberg N, Lomakin IB, Tiedge H (2002) Dendritic BC1 RNA: functional role in regulation of translation initiation. J Neurosci 22:10232–10241

Wang H, Iacoangeli A, Lin D, Williams K, Denman RB, Hellen CU, Tiedge H (2005) Dendritic BC1 RNA in translational control mechanisms. J Cell Biol 171:811–821

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

Watson JB, Sutcliffe JG (1987) Primate brain-specific cytoplasmic transcript of the Alu repeat family. Mol Cell Biol 7:3324–3327

Weichenrieder O, Wild K, Strub K, Cusack S (2000) Structure and assembly of the Alu domain of the mammalian signal recognition particle. Nature 408:167–173

Williams BR (1999) PKR; a sentinel kinase for cellular stress. Oncogene 18:6112–6120

Xiang S, Liu Z, Zhang B, Zhou J, Zhu BD, Ji J, Deng D (2010) Methylation status of individual CpG sites within Alu elements in the human genome and Alu hypomethylation in gastric carcinomas. BMC Cancer 10:44

Yakovchuk P, Goodrich JA, Kugel JF (2009) B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. Proc Natl Acad Sci USA 106:5569–5574

Yang JH, Luo X, Nie Y, Su Y, Zhao Q, Kabir K, Zhang D, Rabinovici R (2003) Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation. Immunology 109:15–23

Zalfa F, Giorgi M, Primerano B, Moro A, Di Penta A, Reis S, Oostra B, Bagni C (2003) The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. Cell 112:317–327

Zalfa F, Adinolfi S, Napoli I, Kuhn-Holsken E, Urlaub H, Achsel T, Pastore A, Bagni C (2005) Fragile X mental retardation protein (FMRP) binds specifically to the brain cytoplasmic RNAs BC1/BC200 via a novel RNA-binding motif. J Biol Chem 280:33403–33410

Zhang Z, Carmichael GG (2001) The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell 106:465–475

Zhang R, Wang YQ, Su B (2008) Molecular evolution of a primate-specific microRNA family. Mol Biol Evol 25:1493–1502

Zhong J, Chuang SC, Bianchi R, Zhao W, Lee H, Fenton AA, Wong RK, Tiedge H (2009) BC1 regulation of metabotropic glutamate receptor-mediated neuronal excitability. J Neurosci 29:9977–9986

Chapter 7
roX RNAs and Genome Regulation in *Drosophila Melanogaster*

S. Kiran Koya and Victoria H. Meller

Abstract Organisms with dimorphic sex chromosomes suffer a potentially lethal imbalance in gene expression in one sex. Addressing this fundamental problem can be considered the first, and most essential, aspect of sexual differentiation. In the model organisms *Drosophila, Caenorhabditis elegans*, and mouse, expression from X-linked genes is modulated by selective recruitment of chromatin-modifying complexes to X chromatin. In both flies and mammals, large noncoding RNAs have a central role in recruitment and activity of these complexes. This review will summarize current knowledge of the function of the noncoding *roX* genes in this process in *Drosophila*. Identification of an autosomal function for the *roX* RNAs raises intriguing questions about the origin of the modern dosage compensation system in flies.

7.1 Introduction

7.1.1 Genome Regulation and Large, Noncoding RNAs

Control of gene expression is central to life in all organisms. In addition to local gene regulation, many eukaryotes rely on coordinated control large chromatin domains. These clusters of coregulated genes can be as large as an entire chromosome. While the mechanisms that coordinate control of groups of genes are often poorly understood, the frequent association of large, noncoding RNAs (lncRNAs) with this process suggests that RNA is extremely well suited for regional chromatin regulation. The most dramatic example of this is sex chromosome dosage compensation in flies and mammals. Many diploid species, such as *Caenorhabditis elegans, Drosophila*, and mammals, have dimorphic sex chromosomes. Females and

S.K. Koya and V.H. Meller (✉)
Department of Biological Sciences, Wayne State University, 5047 Gullen Mall, Detroit, MI 48202, USA
e-mail: sk_koya@wayne.edu; meller@biology.biosci.wayne.edu

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3_7,

C. elegans hermaphrodites have two X chromosomes, but males have a single X chromosome (XY or XO). In these species, the Y chromosome is gene poor, and the genes present on it are expressed only in testes. The resulting imbalance in the ratio of X to autosomal gene expression is potentially lethal to one sex (Gupta et al. 2006; Nguyen and Disteche 2006). Several independently evolved strategies to balance X-linked gene expression between the sexes, a process called dosage compensation, have arisen (Lucchesi et al. 2005). *Drosophila* males increase transcription from their single X chromosome. This increase requires transcript produced from the X-linked *roX* (*RNA on the X 1, – 2*) genes. Mammalian females silence transcription from most genes on one of their two X chromosomes. *Xist* (*X inactive specific transcript*) is a lncRNA produced from the *X inactivation center* (*Xic*) [reviewed by Plath et al. (2002) and Chap. 3]. Production of *Xist* induces silencing of the X chromosome on which it is situated, one of the two X chromosomes present in females. In spite of striking differences in dosage compensation between flies and mammals, both employ lncRNAs that regulate this process and are necessary for identification of X chromatin. This convergence of function suggests that lncRNAs are particularly well suited for the regulation of broad chromatin domains. This review will explore the regulatory role the *roX* transcripts in *Drosophila*.

7.2 Noncoding RNAs in *Drosophila*: A Wealth of Transcripts with Few Known Functions

The *Drosophila melanogaster* genome consists of about 15,000 genes. While only a few hundreds are currently annotated as noncoding RNAs, this group is poorly understood, and annotations of noncoding transcripts lag the rest of the genome (Tweedie et al. 2009). The best-studied noncoding RNAs participate in translation and have well-defined functions that are determined by the structure of the RNA in question. But, in addition, there are numerous long, spliced, and polyadenylated transcripts that appear similar to mRNAs but lack significant open reading frames (Tupy et al. 2005). Identification of these began over two decades ago, but just a handful of the predicted lncRNAs in the fly genome have been studied in any detail. Of these, the *roX* RNAs are perhaps the best understood. The potent regulatory effects of the *roX* genes raise the question of whether some of the many transcripts with no known function may have similar actions in genomic regulation.

7.3 *roX* RNAs and Dosage Compensation

To overcome the potentially lethal imbalance in gene expression caused by hemizygosity of the X chromosome, male flies increase expression from almost all genes on their single X chromosome. This ensures a constant ratio of X to autosomal gene

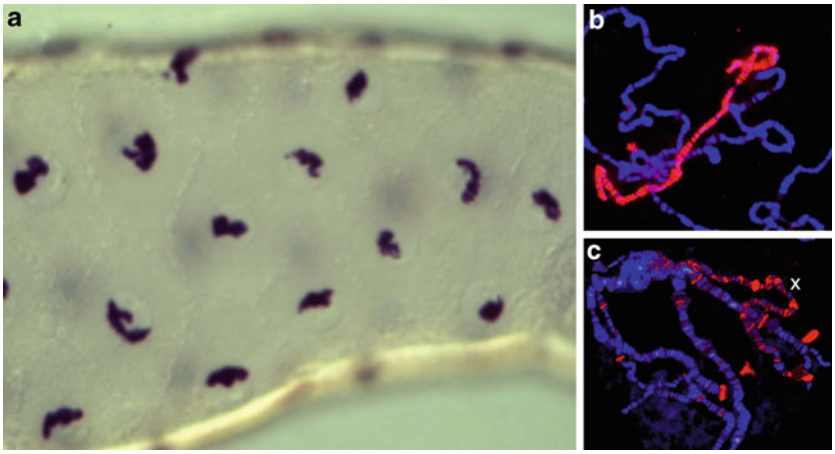

Fig. 7.1 The roX transcripts localize to the X chromosome and are necessary for X chromosome recognition. (**a**) *roX1* coats the X chromosome in a male salivary gland. In situ hybridization to an antisense *roX1* probe is detected by alkaline phosphatase staining (*purple*). (**b**) MSL1 localization in a polytene preparation from a wild type male reveals exclusive localization to the X chromosome. Anti-MSL1 is detected by Texas Red, DNA by DAPI. (**c**) MSL1 localization in *roX1 roX2* males is no longer exclusive to the X chromosome (X) but appears at a number of ectopic autosomal sites

product in both sexes. To achieve this, a complex of protein and *roX* RNA, termed the Male Specific Lethal complex (MSL complex or dosage compensation complex, DCC), is recruited to the X chromosome with exquisite selectivity, as illustrated in Fig. 7.1a, b. The MSL complex alters expression by modifying chromatin within the body of transcribed genes. *roX1* and *roX2* are polyadenylated, noncoding RNAs that are dissimilar in size and sequence (Amrein and Axel 1997; Meller et al. 1997). The major forms of *roX1* are almost 4 kb, but the most abundant form of *roX2* is only 500 bp (Park et al. 2005). In spite of their dramatic difference in size and sequence, *roX1* and *roX2* are redundant for all known functions. Mutation of either *roX* gene alone has no phenotype, but simultaneous mutation of both is male-lethal. Chromosome preparations from *roX1 roX2* males display reduced X localization of the proteins in the MSL complex, and these proteins now can be observed binding at ectopic sites throughout the genome (Fig. 7.1c). In contrast, females mutated for both *roX* genes display no detectable phenotype and are fully viable (Deng et al. 2009; Meller and Rattner 2002).

Although both *roX* genes are situated on the X chromosome, transcripts from autosomal *roX* transgenes will assemble with the MSL complex, bind to the X chromosome, and rescue *roX1 roX2* males (Meller and Rattner 2002). *roX* RNA can therefore travel through the nucleoplasm to regulate a chromosome in trans to its site of transcription. This suggests that the *roX* genes act in a fundamentally different way than *Xist*, whose action appears limited to its chromosome of origin, a feature necessitated by the need to protect one X chromosome from inactivation. However, *roX* also has the ability to direct binding of the MSL

complex to autosomal chromatin in cis to *roX* transgenes (Kageyama et al. 2001; Kelley et al. 1999). While the mechanisms that underlie the ability to recruit the MSL complex in cis remain speculative, all studies to date suggest that recruitment is determined by the ratio of MSL protein to *roX* RNA (Kelley et al. 2008; Oh et al. 2003; Park et al. 2002). High levels of MSL protein are proposed to allow formation of intact complexes as *roX* RNA is being transcribed, favoring localization close to the site of transcription. In contrast, when protein levels are low and *roX* transcription is high, *roX* will be released from its site of synthesis before assembly of the complex, eliminating the preference for local binding. Many questions remain about the precise molecular mechanisms by which *roX* RNAs act, but it is clear that the *roX* genes are central to X chromosome targeting.

7.4 Proteins of the MSL Complex

The *roX* RNAs assemble with five proteins, collectively known as the Male-Specific Lethals (MSLs; reviewed by (Gelbart and Kuroda 2009; Mendjan and Akhtar 2007)). These are MSL1, -2, and -3 (*Male Specific Lethal 1, -2, and -3*), MLE (*Maleless*), and MOF (*Males absent on the first*). All of the MSL proteins are necessary for dosage compensation. Mutation of any one of the *msl* genes causes male lethality as third instar larvae or pupae. In spite of the male-limited role of the MSL complex, most of the MSL proteins are present in both sexes, albeit at lower levels in females (Chang and Kuroda 1998; Lyman et al. 1997). MSL2 is the sole member of the complex whose expression is limited to males (Bashaw and Baker 1995; Kelley et al. 1995; Zhou et al. 1995). Translation of MSL2 mRNA is blocked by the *Sexlethal* protein, which is present only in females (SXL; Cline and Meyer 1996; Gebauer et al. 1998). Ectopic expression of MSL2 in females leads to formation of intact MSL complexes that bind both female X chromosomes, causing female lethality (Kelley et al. 1995). Female lethality is presumably due to elevated expression from both X chromosomes. This supports the idea that maintenance of the correct ratio of X to autosomal gene products is critical for normal development.

MSL2 and MSL1 are essential for all chromatin binding by the rest of the complex (Li et al. 2005; Lyman et al. 1997). MSL1 and MSL2 interact with each other, and this protein–protein interaction has been postulated to form a joint DNA-binding surface, although biochemical studies have yet to confirm this (Copps et al. 1998; Li et al. 2008; Rodriguez et al. 2007). MSL1 also serves as a scaffold for assembly of the other MSL proteins, as carboxy-terminal regions of MSL1 interact with MOF and MSL3 (Morales et al. 2004; Scott et al. 2000). MOF and MSL3 also interact with each other, and both proteins have been reported to bind RNA (Akhtar et al. 2000; Buscaino et al. 2003). MLE is the only member of the complex whose association with the other MSL proteins appears to be indirect (Copps et al. 1998). MLE is an RNA and DNA helicase of the DExH subfamily (Kuroda et al. 1991). MLE association with the polytene X chromosome is RNA dependent (Richter et al. 1996). MLE can be coimmunoprecipitated with *roX2* RNA from SL2 cells

(Smith et al. 2000; Akhtar et al. 2000). In early embryos, *roX1* stability depends on maternally deposited MLE (Meller 2003). Taken together, these observations suggest that MLE is tethered to the MSL complex through *roX* RNA, and, in the absence of MLE, *roX* RNAs are not integrated into the MSL complex, subjecting them to rapid degradation. Localization of MOF and MSL3 has also been reported to be sensitive to RNase treatment (Akhtar et al. 2000; Buscaino et al. 2003). *roX* RNAs are thus believed to play a major role in the assembly of the MSL complex, and their continued presence may be necessary for stable binding to the X chromosome.

While elimination of MSL1 or MSL2 results in loss of all chromatin binding by the remaining member of the complex, elimination of MLE, MSL3, or MOF leaves residual MSL proteins bound at a subset of X-linked sites. These are proposed to be recruitment sites from which the MSL complex can spread to nearby genes (Kelley et al. 1999). Indeed, recent studies have identified a short sequence motif enriched at sites of strong MSL3-independent binding, termed MSL recognition elements (MRE; Alekseyenko et al. 2008; Straub et al. 2008). Current models propose that the complete pattern of MSL binding along the X chromosome involves attraction of the MSL complex to sites containing MREs, followed by spreading into transcribed genes situated nearby (Gelbart and Kuroda 2009). Examination of MSL binding at high resolution revealed enrichment in the body and 3′ ends of actively transcribed genes (Larschan et al. 2007). This pattern is similar to that of the cotranscriptional H3K36me3 mark, and this observation is explained by the finding that the MSL3 chromodomain binds H3K36me3 (Sural et al. 2008). Mutation of conserved residues in the MSL3 chromodomain disrupts normal spreading of the MSL complex into transcribed genes. Taken together, these studies support an elegant model that explains the local distribution of the MSL complex on the X chromosome. What this model fails to explain is how the MSL complex is limited to the X chromosome. All transcribed genes share enrichment for H3K36me3, and MREs are only modestly enriched on the X chromosome. It thus seems unlikely that these are the only factors directing localization of the MSL complex. A recent study determined that during interphase, regions of the X chromosome with high affinity for the MSL complex are closer together than regions with little or no binding (Grimaud and Becker 2009). This suggests that interphase chromosome architecture might be a factor in selective recognition of X chromatin.

MOF is a histone acetyltransferase specific for lysine 16 on H4 (H4Ac16) (Akhtar and Becker 2000; Hilfiker et al. 1997; Smith et al. 2000). Histone acetylation in general is thought to reduce the strength of histone–DNA interactions, making DNA more accessible. The H4Ac16 modification specifically prevents tight packing of nucleosomes, and this may contribute to elevated expression, as well as the slightly decondensed character of the male X chromosome (Shogren-Knaak et al. 2006). The distribution of H4Ac16 enrichment is similar to that of the MSL complex, being more pronounced in the 3′ ends of genes and coding regions than on promoters of transcribed genes (Kind et al. 2008). As chromatin modification by the MSL complex occurs mainly within the body of genes, it is likely that enhanced transcription is due to facilitation of elongation, rather than initiation of

transcription (Smith et al. 2001). Modulation of a general property of RNA pol II, such as speed or processivity, would explain how the MSL complex achieves a uniform two-fold increase in expression of thousands of genes with disparate expression patterns and regulatory regions.

A second chromatin modification is enriched on the male X chromosome and depends on the MSL complex. The JIL-1 kinase is an essential protein required in both sexes, yet partial loss of function alleles affect males more severely than females, and a genetic study suggests a role for JIL-1 in compensation of an X-linked gene (Jin et al. 1999; Lerach et al. 2005). JIL-1 phosphorylates serine 10 on histone 3 (H3pS10), a mark that is associated with open chromatin structure and increased gene expression (Wang et al. 2001). JIL-1 localizes to interband regions on all chromosomes and is enriched on the male X chromosome. On the X, JIL-1 colocalizes with MSL proteins and, under some conditions, it may immunoprecipitate with the MSL complex, suggesting a possible molecular interaction (Jin et al. 2000). However, it remains unclear if JIL-1 enrichment on the X chromosome is due to a direct interaction with the MSL complex or if it is the consequence of MSL complex action, for example, a response to elevated transcription or chromatin modification by the MSL complex.

7.5 Separate Domains of *roX1* Regulate X-Localization and Histone Modification

At least one *roX* transcript is essential for targeting the intact MSL complex to the male X chromosome (Deng and Meller 2006; Meller and Rattner 2002). In *roX1 roX2* males, the proteins of the MSL complex still colocalize but are no longer exclusive to the X chromosome. Although MOF is present at these ectopic autosomal sites in *roX1 roX2* males, H4Ac16 modification at these sites is low, suggesting that *roX* association with the MSL complex is necessary for full MOF activity (Deng and Meller 2006). Interestingly, the 3′ end of *roX1* contains a stem loop that is necessary for *roX1* function, but deletion of this portion of the transcript has a relatively mild effect on X-localization (Stuckenholz et al. 2003). In addition, short repeats in the 3′ end of *roX1* are also present in *roX2* and in the *roX* genes of numerous related species (Franke and Baker 1999; Kelley et al. 2008). The presence of these repeats appears to regulate activity of the complex (Kelley et al. 2008; Park et al. 2007). While it is tempting to speculate that *roX* RNA is the allosteric regulator of MOF, other MSL proteins also influence MOF activity. Interaction of MOF with a subcomplex of MSL1 and MSL3 increases the efficiency and substrate specificity of MOF (Morales et al. 2004). This emphasizes the point that the normal activity of MOF requires assembly of the intact MSL complex.

In contrast to the function of 3′ *roX1* sequences, deletions removing significant portions of the 5′ end affect localization of the complex to the X chromosome. This region comprises almost 1.5 kb and lacks obvious repetitive sequences or secondary structures of high stability. A scanning deletion analysis that removed ~300 bp

portions of *roX1* failed to identify essential elements in the 5′ end (Stuckenholz et al. 2003). However, deletions removing large portions of this region reduce X chromosome binding of the MSL complex, although mutants retaining even a very small portion of the 5′ end support partial dosage compensation (Deng and Meller 2008; Deng et al. 2005). Together, these studies support the idea that separate regions of *roX1* direct MSL complex localization and the chromatin-modifying activity of the complex. This is reminiscent of the distribution of function in *Xist*. Short, tandem stem loops are necessary for *Xist*-mediated chromatin silencing in mice, but painting of the X chromosome is directed by several large segments of *Xist* that may work cooperatively to ensure X recognition (Wutz et al. 2002).

7.6 Ancestral Origins of Complexes that Dosage Compensate Sex Chromosomes

The sex chromosomes of mammals, *C. elegans*, and flies are unrelated to each other in evolutionary origin. Indeed, the de novo origin of differentiated sex chromosomes has occurred repeatedly in different animal lineages (Bull 1985). In accordance with this independent origin, sex chromosome dosage compensation has arisen independently many times. While each system for dosage compensation achieves the goal of maintaining an appropriate ratio of X to autosomal gene products, each has adopted a completely different strategy to do so. All three systems have developed through recruitment of preexisting chromatin regulatory complexes (Table 7.1). For example, the DCC of *C. elegans* is related to, and shares subunits with, the condensin complex that compacts chromosomes and enables normal segregation during mitosis and meiosis (Chan et al. 2004; Csankovszki et al. 2009). In accordance with this, some mutations that disrupt *C. elegans* dosage compensation also disrupt meiosis and mitosis (Hagstrom et al. 2002; Lieb et al. 1996). Silencing of an X chromosome in female mice is a complex process that takes place over several days during early embryogenesis, but an early event is recruitment of the Polycomb group 2 and 1 (Pcg2, Pcg1) complexes (Schoeftner et al. 2006; Zhao et al. 2008). Pcg2 deposits the silencing H3K27 trimethylation mark, and Pcg1 ubiquitinates H2 on K119 (Kohlmaier et al. 2004; Plath et al. 2003). These modifications may contribute to

Table 7.1 Dosage compensation recruits existing chromatin-modifying complexes for novel functions

Organism	Compensation machinery	Compensatory function	Ancestral complex and function
C. elegans	DCC complex	Downregulation of X-linked genes	Condensin mitotic and meiotic chromosome condensation
Mammals	Prc1, Prc2	Inactivation of X chromosome	Prc1, Prc2 developmentally stable repression
Drosophila	MSL complex	Upregulation of X-linked genes	?

the stability of X inactivation. In addition to their role in X inactivation, both complexes continue to function in epigenetic repression throughout the genome during mammalian development (Bernstein et al. 2007).

While the ancestral functions of the proteins that achieve dosage compensation in *Drosophila* remain to be fully defined, homologs of MOF, MSL1, MSL2, and MSL3 have been found in organisms as diverse as yeast and mammals (Eisen et al. 2001; Smith et al. 2005; Marin 2003; Sanjuan and Marin 2001). With the exception of MLE, the mammalian homologs associate with each other, suggesting that the modern MSL complex of flies has an ancient origin. Human MOF (hMof) is notable as it participates in multiple complexes, and these are responsible for the majority of H4KAc16 modification in mammalian cells (Cai et al. 2010; Mendjan et al. 2006; Smith et al. 2005). Although the precise molecular function of hMOF-containing complexes is not clear, depletion of hMof affects DNA repair, possibly by disruption of damage signaling (Gupta et al. 2005; Taipale et al. 2005). MOF also participates in multiple complexes in flies, which may allow it to serve as a general regulator of chromatin at promoters, although this finding remains controversial (Mendjan et al. 2006; Gelbart et al. 2009; Kind et al. 2008). Recent work in our laboratory suggests a different autosomal role for MOF, and other MSL proteins, in flies.

7.7 Regulation of Heterochromatic Genes by *roX* and a Subset of MSL Proteins

In addition to reduced expression of X-linked genes in *roX1 roX2* males, several hundred autosomal genes situated in heterochromatic regions are also misregulated (Deng et al. 2009). Regions containing misregulated genes include the entire 4th chromosome. The 4th chromosome has several peculiarities, including its small size, lack of recombination, and possible evolutionary kinship with the X chromosome (Larsson and Meller 2006; Riddle and Elgin 2006). However, the fact that the 4th chromosome is enriched for heterochromatin is the feature that it shares with the other autosomal genes that depend on *roX* RNA for full expression. Unexpectedly, this feature of heterochromatic gene regulation is limited to males (Deng et al. 2009). Analysis of expression in *msl* mutants revealed that MSL1, MSL3, MLE, and MOF are also required for full expression of heterochromatic and 4th-linked genes in males. However, no misregulation of these autosomal genes is observed in *msl2* mutants, indicating that the intact MSL complex is not involved (Deng et al. 2009). Because MSL2 is the sole member of the MSL complex that is strictly male-limited, it remains unclear how the sex-specificity of heterochromatic gene regulation is maintained. As MSL1 and MSL2 are postulated to work together to target the MSL complex to the X chromosome, it appears likely that MSL2 is dedicated for the recognition of the X chromosome (Li et al. 2008; Rodriguez et al. 2007). The X chromosome is about twofold enriched for MREs in comparison to the autosomes, but interestingly, MREs are depleted from the 4th chromosome (Alekseyenko et al. 2008). This reinforces the idea that although regulation of X-linked and

heterochromatic genes requires overlapping sets of molecules, recognition of these two groups occurs by different mechanisms. To explain these findings, we have proposed that a second complex composed of *roX* RNA and subset of MSL proteins is responsible for the modulation of chromatin at autosomal heterochromatic sites in males.

It is tempting to speculate that regulation of heterochromatic genes reflects an ancestral function of the members of the MSL complex. Heterochromatic genes are situated in a difficult environment and have long been thought to utilize specialized regulatory mechanisms (Yasuhara and Wakimoto 2006). However, the limitation to males suggests a process that coevolved with the modern sex chromosomes of flies. One possibility is that the highly differentiated sex chromosomes create nuclear environments that are sufficiently different to require a dedicated regulatory system in one sex. As the *Drosophila* Y chromosome is large and entirely heterochromatic, it is plausible that it alters the balance of chromatin proteins throughout the nucleus (Weiler and Wakimoto 1995). Taking into account the multiple functions of *roX* RNA in genome regulation, we present a hypothetical model for the origin of *roX*-dependent complexes in Fig. 7.2. In human cells, homologs of MLE and *roX* have

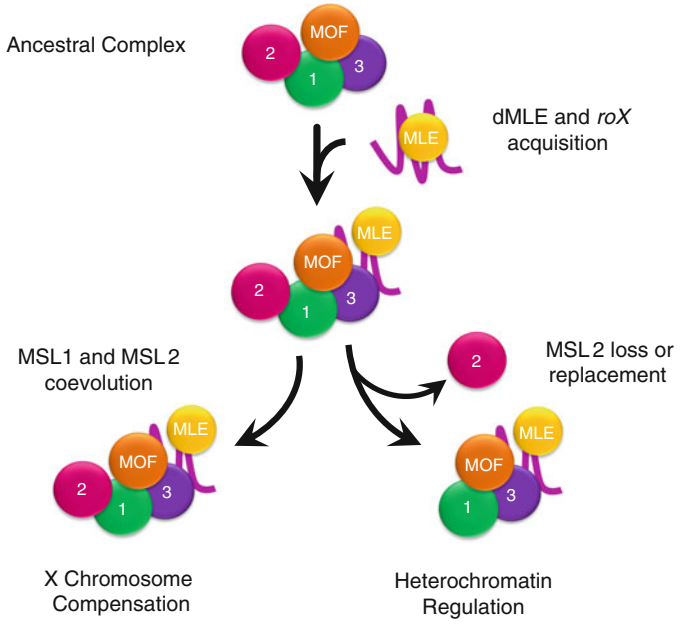

Fig. 7.2 Proposed origin *roX* RNA complexes in *Drosophila*. As homologs of MSL1, MSL2, MSL3, and MOF associate in many organisms, this association may represent the ancestral form of the complex in flies (*top*). We speculate that acquisition of MLE and *roX* was an early event in formation of the modern MSL complex, followed by coevolution of MSL1 and MSL2 (*left*), a feature that may determine X recognition. Loss or replacement of MSL2 may enable complex members to acquire autosomal functions such as regulation of genes in heterochromatic environments (*right*). While the association of *roX* and MSL proteins in the intact MSL complex (*left*) is well established, the presence of a subcomplex lacking MSL2 (*right*) has yet to be established

not been identified in association with the other hMSL proteins (Cai et al. 2010; Smith et al. 2005; Taipale et al. 2005). Because of this, we have chosen to model the acquisition of *roX* and MLE as an early step in the evolution of modern *roX*-containing complexes of flies. Rapid co-evolution of the MSL2 and MSL1 interaction domains may enable recognition of MREs, likely to be an essential function of the MSL complex that dosage compensates the X chromosome (left). MSL2 is the sole member of this complex lacking a heterochromatic role and is thus anticipated to be dedicated to X recognition. We speculate that loss or replacement of MSL2 has enabled the remaining MSL proteins and *roX* RNA to be recruited for a new purpose, regulation of heterochromatic genes in males. While it appears logical that a subset of MSL proteins and the *roX* RNAs form a second complex, the existence of this has yet to be demonstrated. While studies in flies have identified multiple MOF-containing complexes, the technique used, affinity purification followed by mass spectrometric analysis, would not reveal a minor contribution of a subcomplex lacking MSL2 (Mendjan et al. 2006). The mechanism by which autosomal genes are regulated by the *roX* RNAs remains to be fully elucidated.

7.8 Conclusions

The large noncoding *roX* RNAs have a central role in sex chromosome dosage compensation in flies, where they fulfill a role with similarities to that of *Xist* during mammalian dosage compensation. *roX* transcripts assemble with the MSL proteins to form a complex that displays exclusive X chromosome binding. Situation of the *roX* genes on the X chromosome facilitates X recognition through the ability of the *roX* genes to attract the MSL complex in cis. These observations created the impression that the *roX* RNAs were dedicated to identification and modification of the X chromosome, but this idea has been revised by the unexpected discovery of a role for *roX* in expression of heterochromatic genes. It appears that both regulatory systems may be necessitated by the presence of highly differentiated sex chromosomes.

References

Akhtar A, Becker PB (2000) Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol Cell 5:367–375

Akhtar A, Zink D, Becker PB (2000) Chromodomains are protein-RNA interaction modules. Nature 407:405–409

Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ et al (2008) A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. Cell 134:599–609

Amrein H, Axel R (1997) Genes expressed in neurons of adult male Drosophila. Cell 88:459–469

Bashaw GJ, Baker BS (1995) The msl-2 dosage compensation gene of Drosophila encodes a putative DNA-binding protein whose expression is sex specifically regulated by Sex-lethal. Development 121:3245–3258

Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. Cell 128:669–681

Bull JJ (1985) Sex determining mechanisms: an evolutionary perspective. Experientia 41:1285–1296

Buscaino A, Kocher T, Kind JH, Holz H, Taipale M, Wagner K, Wilm M, Akhtar A (2003) MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex. Mol Cell 11:1265–1277

Cai Y, Jin J, Swanson SK, Cole MD, Choi SH, Florens L, Washburn MP, Conaway JW, Conaway RC (2010) Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. J Biol Chem 285:4268–4272

Chan RC, Severson AF, Meyer BJ (2004) Condensin restructures chromosomes in preparation for meiotic divisions. J Cell Biol 167:613–625

Chang KA, Kuroda MI (1998) Modulation of MSL1 abundance in female Drosophila contributes to the sex specificity of dosage compensation. Genetics 150:699–709

Cline TW, Meyer BJ (1996) Vive la difference: males vs females in flies vs worms. Annu Rev Genet 30:637–702

Copps K, Richman R, Lyman LM, Chang KA, Rampersad-Ammons J, Kuroda MI (1998) Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. EMBO J 17:5409–5417

Csankovszki G, Collette K, Spahl K, Carey J, Snyder M, Petty E, Patel U, Tabuchi T, Liu H, McLeod I et al (2009) Three distinct condensin complexes control C. elegans chromosome dynamics. Curr Biol 19:9–19

Deng X, Meller VH (2006) roX RNAs are required for increased expression of X-linked genes in Drosophila melanogaster males. Genetics 174:1859–1866

Deng X, Meller VH (2008) Molecularly severe roX1 mutations contribute to dosage compensation in Drosophila. Genesis 47:49–54

Deng X, Rattner BP, Souter S, Meller VH (2005) The severity of roX1 mutations is predicted by MSL localization on the X chromosome. Mech Dev 122:1094–1105

Deng X, Koya SK, Kong Y, Meller VH (2009) Coordinated regulation of heterochromatic genes in Drosophila melanogaster males. Genetics 182:481–491

Eisen A, Utley RT, Nourani A, Allard S, Schmidt P, Lane WS, Lucchesi JC, Cote J (2001) The yeast NuA4 and Drosophila MSL complexes contain homologous subunits important for transcription regulation. J Biol Chem 276:3484–3491

Franke A, Baker BS (1999) The roX1 and roX2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. Mol Cell 4:117–122

Gebauer F, Merendino L, Hentze MW, Valcarcel J (1998) The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. RNA 4:142–150

Gelbart ME, Kuroda MI (2009) Drosophila dosage compensation: a complex voyage to the X chromosome. Development 136:1399–1410

Gelbart ME, Larschan E, Peng S, Park PJ, Kuroda MI (2009) Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. Nat Struct Mol Biol 16:825–832

Grimaud C, Becker PB (2009) The dosage compensation complex shapes the conformation of the X chromosome in Drosophila. Genes Dev 23:2490–2495

Gupta A, Sharma GG, Young CS, Agarwal M, Smith ER, Paull TT, Lucchesi JC, Khanna KK, Ludwig T, Pandita TK (2005) Involvement of human MOF in ATM function. Mol Cell Biol 25:5292–5305

Gupta V, Parisi M, Sturgill D, Nuttall R, Doctolero M, Dudko OK, Malley JD, Eastman PS, Oliver B (2006) Global analysis of X-Chromosome compensation. J Biol 5:3

Hagstrom KA, Holmes VF, Cozzarelli NR, Meyer BJ (2002) C. elegans condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. Genes Dev 16:729–742

Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC (1997) mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila. EMBO J 16:2054–2060

Jin Y, Wang Y, Walker DL, Dong H, Conley C, Johansen J, Johansen KM (1999) JIL-1: a novel chromosomal tandem kinase implicated in transcriptional regulation in Drosophila. Mol Cell 4:129–135

Jin Y, Wang Y, Johansen J, Johansen KM (2000) JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. J Cell Biol 149:1005–1010

Kageyama Y, Mengus G, Gilfillan G, Kennedy H, Stuckenholz C, Kelley R, Becker P, Kuroda M (2001) Association and spreading of the *Drosophila* dosage compensation complex from a discrete *roX1* chromatin entry site. Embo J 20(9):2236–2245

Kelley RL, Solovyeva I, Lyman LM, Richman R, Solovyev V, Kuroda MI (1995) Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila. Cell 81:867–877

Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI (1999) Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell 98:513–522

Kelley RL, Lee OK, Shim YK (2008) Transcription rate of noncoding roX1 RNA controls local spreading of the Drosophila MSL chromatin remodeling complex. Mech Dev 125:1009–1019

Kind J, Vaquerizas JM, Gebhardt P, Gentzel M, Luscombe NM, Bertone P, Akhtar A (2008) Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. Cell 133:813–828

Kohlmaier A, Savarese F, Lachner M, Martens J, Jenuwein T, Wutz A (2004) A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS Biol 2: E171

Kuroda MI, Kernan MJ, Kreber R, Ganetzky B, Baker BS (1991) The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila. Cell 66:935–947

Larsson J, Meller VH (2006) Dosage compensation, the origin and the afterlife of sex chromosomes. Chromosome Res 14:417–431

Larschan E, Alekseyenko, A, Gortchakov, A, Peng S, Li B, Yang P, Workman J, Park P, Kuroda M (2007) MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. Mol Cell 28:121–133

Lerach S, Zhang W, Deng H, Bao X, Girton J, Johansen J, Johansen KM (2005) JIL-1 kinase, a member of the male-specific lethal (MSL) complex, is necessary for proper dosage compensation of eye pigmentation in Drosophila. Genesis 43:213–215

Li F, Parry DA, Scott MJ (2005) The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. Mol Cell Biol 25:8913–8924

Li F, Schiemann AH, Scott MJ (2008) Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex. Mol Cell Biol 28:1252–1264

Lieb JD, Capowski EE, Meneely P, Meyer BJ (1996) DPY-26, a link between dosage compensation and meiotic chromosome segregation in the nematode. Science 274:1732–1736

Lucchesi JC, Kelly WG, Panning B (2005) Chromatin remodeling in dosage compensation. Annu Rev Genet 39:615–651

Lyman LM, Copps K, Rastelli L, Kelley RL, Kuroda MI (1997) Drosophila male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association. Genetics 147:1743–1753

Marin I (2003) Evolution of chromatin-remodeling complexes: comparative genomics reveals the ancient origin of "novel" compensasome genes. J Mol Evol 56:527–539

Meller VH (2003) Initiation of dosage compensation in Drosophila embryos depends on expression of the roX RNAs. Mech Dev 120:759–767

Meller VH, Rattner BP (2002) The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J 21:1084–1091
Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL (1997) roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. Cell 88:445–457
Mendjan S, Akhtar A (2007) The right dose for every sex. Chromosoma 116:95–106
Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, Vermeulen M, Buscaino A, Duncan K, Mueller J et al (2006) Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. Mol Cell 21:811–823
Morales V, Straub T, Neumann MF, Mengus G, Akhtar A, Becker PB (2004) Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. EMBO J 23:2258–2268
Nguyen DK, Disteche CM (2006) Dosage compensation of the active X chromosome in mammals. Nat Genet 38:47–53
Oh H, Park Y, Kuroda MI (2003) Local spreading of MSL complexes from roX genes on the Drosophila X chromosome. Genes Dev 17:1334–1339
Park Y, Kelley RL, Oh H, Kuroda MI, Meller VH (2002) Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. Science 298:1620–1623
Park Y, Oh H, Meller VH, Kuroda MI (2005) Variable Splicing of Non-Coding roX2 RNAs Influences Targeting of MSL Dosage Compensation Complexes in Drosophila. RNA Biol 2:157–164
Park SW, Kang Y, Sypula JG, Choi J, Oh H, Park Y (2007) An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys16 acetylation on the Drosophila X chromosome. Genetics 177:1429–1437
Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B (2002) Xist RNA and the mechanism of X chromosome inactivation. Annu Rev Genet 36:233–278
Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science 300:131–135
Richter L, Bone JR, Kuroda MI (1996) RNA-dependent association of the Drosophila maleless protein with the male X chromosome. Genes Cells 1:325–336
Riddle NC, Elgin SC (2006) The dot chromosome of Drosophila: insights into chromatin states and their change over evolutionary time. Chromosome Res 14:405–416
Rodriguez MA, Vermaak D, Bayes JJ, Malik HS (2007) Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in Drosophila. Proc Natl Acad Sci USA 104:15412–15417
Sanjuan R, Marin I (2001) Tracing the origin of the compensasome: evolutionary history of DEAH helicase and MYST acetyltransferase gene families. Mol Biol Evol 18:330–343
Schoeftner S, Sengupta AK, Kubicek S, Mechtler K, Spahn L, Koseki H, Jenuwein T, Wutz A (2006) Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J 25:3110–3122
Scott MJ, Pan LL, Cleland SB, Knox AL, Heinrich J (2000) MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila. EMBO J 19:144–155
Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science 311:844–847
Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC (2000) The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. Mol Cell Biol 20:312–318
Smith ER, Allis CD, Lucchesi JC (2001) Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in Drosophila males. J Biol Chem 276:31483–31486
Smith ER, Cayrou C, Huang R, Lane WS, Cote J, Lucchesi JC (2005) A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. Mol Cell Biol 25:9175–9188

Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB (2008) The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. PLoS Genet 4: e1000302

Stuckenholz C, Meller VH, Kuroda MI (2003) Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in Drosophila melanogaster. Genetics 164:1003–1014

Sural TH, Peng S, Li B, Workman JL, Park PJ, Kuroda MI (2008) The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X chromosome. Nat Struct Mol Biol 15:1318–1325

Taipale M, Rea S, Richter K, Vilar A, Lichter P, Imhof A, Akhtar A (2005) hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. Mol Cell Biol 25:6798–6810

Tupy JL, Bailey AM, Dailey G, Evans-Holm M, Siebel CW, Misra S, Celniker SE, Rubin GM (2005) Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster. Proc Natl Acad Sci USA 102:5495–5500

Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, Millburn G, Osumi-Sutherland D, Schroeder A, Seal R et al (2009) FlyBase: enhancing Drosophila Gene Ontology annotations. Nucleic Acids Res 37:D555–D559

Wang Y, Zhang W, Jin Y, Johansen J, Johansen KM (2001) The JIL-1 tandem kinase mediates histone H3 phosphorylation and is required for maintenance of chromatin structure in Drosophila. Cell 105:433–443

Weiler KS, Wakimoto BT (1995) Heterochromatin and gene expression in Drosophila. Annu Rev Genet 29:577–605

Wutz A, Rasmussen TP, Jaenisch R (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet 30:167–174

Yasuhara JC, Wakimoto BT (2006) Oxymoron no more: the expanding world of heterochromatic genes. Trends Genet 22:330–338

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322:750–756

Zhou S, Yang Y, Scott MJ, Pannuti A, Fehr KC, Eisen A, Koonin EV, Fouts DL, Wrightsman R, Manning JE et al (1995) Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. EMBO J 14:2884–2895

Chapter 8
Transcription of Satellite DNAs in Insects

Željka Pezer, Josip Brajković, Isidoro Feliciello, and Đurđica Ugarković

Abstract The very complex life cycle and extreme diversity of insect life forms require a carefully regulated network of biological processes to switch on and off the right genes at the right time. Chromatin condensation is an important regulatory mechanism of gene silencing as well as gene activation for the hundreds of functional protein genes harbored in heterochromatic regions of different insect species. Being the major heterochromatin constituents, satellite DNAs (satDNAs) serve important roles in heterochromatin regulation in insects in general. Their expression occurs in all developmental stages, being the highest during embryogenesis. satDNA transcripts range from small RNAs, corresponding in size to siRNAs, and piwiRNAs, to large, a few kb long RNAs. The long transcripts are preferentially nonpolyadenylated and remain in the nucleus. The actively regulated expression of satDNAs by *cis* or *trans* elements as well as by environmental stress, rather than constitutive transcription, speaks in favor of their involvement in differentiation, development, and environmental response.

8.1 Satellite DNAs in Insects

Satellite DNAs (satDNAs) are the major DNA component of eukaryotic heterochromatin. These noncoding sequences constitute a considerable part of the genomic DNA in many insect species, which can reach over half of the genomic content

Ž. Pezer, J. Brajković, and Đ. Ugarković (✉)
Ruđer Bošković Institute, Bijenička 54, 10001 Zagreb, P.O. Box 1016, Croatia
e-mail: ugarkov@irb.hr

I. Feliciello
Ruđer Bošković Institute, Bijenička 54, 10001 Zagreb, P.O. Box 1016, Croatia
and
Dipartimento di Medicina Clinica e Sperimentale, Università degli Studi di Napoli Federico II, via Pansini 5, 80131 Napoli, Italy

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3_8,

(Davis and Wyatt 1989). Composed of tandemly reiterated arrays, usually millions of base pairs long, satDNA is located mainly in the pericentromeric and telomeric regions of chromosomes and are the major building elements of functional centromeres in many eukaryotes including insects (Ugarković 2009a).

The complexity of satDNA sequences varies between insect species. Some are very simple, composed of short repeats oriented in a head-to-tail fashion, such as ten *Drosophila melanogaster* satellites that are between 5 and 10 bp long (Lohe and Roberts 1988). The nucleotide sequences of ten simple repeats in *D. melanogaster* conforms to a formula $(AAN)_m(AN)_n$ where N is any nucleotide. In addition to simple sequence monomers, *D. melanogaster* and other related *Drosophila* species share also a complex satellite of 359 bp, known as 1.688 satellite according to CsCl buoyant density. All *D. melanogaster* satellites are AT rich and, together, they make approximately 20% of the genome. Satellite monomers in insect can also exceed 1,000 bp, as revealed in some beetle species (Pons 2004). However, most of insect satDNAs fall into two size classes: one ranging from 140 to 190 bp and the other in the range of 300–400 bp (reviewed in Palomeque and Lorite (2008)). The total length of the satellite tandem arrays varies from less than 100 bp to over 100 Mb between species. Some satellites are restricted to particular chromosomes such as *Drosophila* 359 bp satellite that is located on X chromosome, while most of the insect satellites are spread on all chromosomes of particular species. Satellites present on the same chromosome can be organized into separate arrays or can exhibit interspersed type of organization characteristic for beetle species *Tribolium madens* (Durajlija Žinić et al. 2000). In many insect species, satDNAs encompass both centromeric and pericentromeric regions. However, satellite sequences in centromeric and pericentromeric regions are almost indistinguishable, even in the best studied *Drosophila* centromere (Sun et al. 1997). Moreover, due to the technical difficulties associated with the sequencing and assembling of highly repetitive regions, the detailed structure and organization of heterochromatic and centromeric regions for many insect species is not known.

satDNAs sometimes possess an additional level of organization, called higher order repeats (HORs), which are best described for α satellite of primates and are characteristic for human centromeres (Schueler et al. 2001). For example, the predominant form of α satellite on human chromosome 17 is a 2.7 kbp HOR unit consisting of 16 α satellite monomers (Waye and Willard 1986). In insects, however, HORs are not found very often and are usually in the form of complex dimers and trimers (Palomeque and Lorite 2008).

It has been demonstrated that related insect species share a set of common satDNAs, differentially amplified between species, so that changes in the number of copies produce a species-specific profile of satDNAs (Meštrović et al. 1998; Bruvo-Mađarić et al. 2007). According to the "library" hypothesis (Fry and Salser 1977), some of the satDNAs originally contained in the common ancestor could be amplified during speciation. In each of the descendants, usually one satDNA would exist as a major satellite sequence while the others would remain as low-copy-number satellites. It is proposed that satellite sequences residing within a library exhibit certain structural characteristics that enable them to confer a centromeric

role and therefore could represent a source for the evolution of new centromere (Ugarković 2008, 2009b).

Satellite sequences are known to evolve fast, changing their nucleotide sequence and copy number by following a pattern known as concerted evolution. Changes among repeats create new variants, which are spread and homogenized within the genome by a variety of DNA turnover mechanisms, such as unequal crossing over, gene conversion, replication slippage, and rolling circle replication (Charlesworth et al. 1994). Variants are subsequently fixed within the reproductive group of organisms in a stochastic process known as molecular drive (Dover 2002). Taken together, recombinational mechanisms and molecular drive seem to be the major cause of high turnover of satDNA sequences, resulting in the significant sequence divergence of satDNAs and changes in copy number, even between closely related species (Ugarković and Plohl 2002).

8.2 Functional Elements Within satDNAs

Due to the homogenizing effects of concerted evolution, satDNAs usually display low internal sequence variability. However, comparison of monomer sequences of a given satDNA reveals that some monomer regions are more conserved while others show higher mutation rates (Borstnik et al. 1994; Romanova et al. 1996; Mravinac et al. 2004). Such a nonuniform rate of evolution along the sequence indicates the presence of selective pressure on satDNAs. It can be proposed that selection affects pericentromeric satellite repeats whose transcripts play a role in heterochromatin establishment through RNA interference mechanism (Volpe et al. 2002). In fission yeast, *Schizosaccharomyces pombe*, analysis of siRNAs involved in heterochromatin formation showed that they derive preferentially from the most conserved regions of repeats (Djupedal et al. 2009). This indicates that conservation is more probably due to functional constraints than to frequent events of homologous recombination causing sequence identity. Therefore, conserved regions found in other satDNAs could be functional in the sense that they represent a preferential source of siRNAs that recruit protein complexes responsible to heterochromatin formation.

Due to possible functional constraints on satDNAs, it is not surprising that some characteristics of satDNAs are shared between many eukaryotic organisms (Ugarković 2005). Probably the most common feature of satDNA is its intrinsic curvature. Satellite repeats are generally AT rich, and the periodical distribution of AT tracts causes DNA to bend into a super-helical tertiary structure (Fitzgerald et al. 1994). This sequence-dependent property is thought to be responsible for the tight packing of DNA and proteins in heterochromatin (Ugarković et al. 1992; Fitzgerald et al. 1994). Conserved CENP-B box-like motifs have been identified within satDNA of mammals and insects (Kipling and Warburton 1997; Lorite et al. 2002a; Mravinac et al. 2004). The CENP-B box is a 17 bp motif in human α-satDNA and a binding site for centromere protein B (CENP-B) (Masumoto et al. 1989) whose homologs have been found in many eukaryotes. Not every repeat of

α satellite contains a functional CENP-B box, but they appear at regular intervals in human centromeres and seem to be essential for centromeric chromatin assembly (Ohzeki et al. 2002).

Given their relatively simple sequence and the lack of any significant open-reading frame, previously reported transcription of satDNA has been ascribed to read-through from upstream genes and transposable elements (Diaz et al. 1981; Wu et al. 1986; Gaubatz and Cutler 1990). However, promoter elements and transcription start sites as well as binding motifs for transcription factors have been mapped within some satellites. Putative internal promoters have been reported in the wasp *Diadromus pulchellus* (Renault et al. 1999) where motifs cognate to RNA Pol II and III are present within the satellite monomer sequence. In schistosome satDNA, which encodes an active ribozyme, a functional RNA Pol III promoter is present (Ferbeyre et al. 1998). The sequence of highly conserved satellite 2 found in distant families of salamanders shares structural and functional properties with the typical vertebrate small nuclear RNA (snRNA) promoter (Coats et al. 1994).

The *Drosophila* GAGA transcription factor that binds GA/CT-rich elements in promoters of many *Drosophila* genes and activates transcription by opening chromatin structure was found associated with heterochromatin throughout the cell cycle. It is proposed that GAGA factor directly interacts with a GA/CT-rich subset of satDNA repeats and modifies heterochromatin structure (Raff et al. 1994). Human satellite III has a binding motif for the heat-shock transcription factor 1, which drives its RNA Pol II-dependent transcription in stress conditions (Metz et al. 2004; Chap. 5 in this book). γ-satDNA, the abundant pericentromeric sequence of all murine chromosomes, contains conserved binding sites for ubiquitous transcription factor Yin Yang 1 (YY1) (Shestakova et al. 2004). YY1 belongs to the Polycomb group of proteins involved in gene regulation during development. It has been found to be associated with γ-satDNA in proliferating cells, but the association strongly diminishes during transition to the quiescent state (G_0). It has been proposed that the interaction of YY1 with γ-satDNA could lead to the targeting of proteins required for heterochromatization or to the silencing of euchromatic genes by bringing them in close proximity to pericentromeric heterochromatin.

Even though some time ago satDNAs were considered useless evolutionary remnants, the functional significance of their sequences is becoming ever more clear. The existence of conserved motifs and structural properties as well as emerging evidence on their widespread transcriptional activity prompt us to reexamine these highly abundant eukaryotic sequences.

8.3 Preserved satDNA Repeats in Insects

Beetles of the family Tenebrionidae (Coleoptera) represent convenient models for studying satDNAs. Located in the (peri)centromeric heterochromatin of virtually all chromosomes, satDNAs comprise up to half of the genome size in most of the

coleopteran species analyzed so far (Ugarković and Plohl 2002). They are characterized by low intraspecific variability of the basic repeating units. However, their sequences differ dramatically between species, with some exceptions such as satDNAs from the genus *Pimelia* (Pons et al. 1997).

Members of the genus *Palorus* share a "library" of satDNAs – a collection of common satellite sequences, differentially amplified in each species. Besides the species-specific major satellite, each species contains several low-copy-number satellites, which account for approximately 0.05% of the genome and are dispersed over the whole heterochromatic block, interrupting large arrays of the major satellite. Shared satellites are extremely conserved with respect to sequence, monomer length, and tandem repeat organization (Meštrović et al. 1998). Satellites from some *Palorus* species are widely distributed even beyond the level of the genus, such as PRAT and PSUB, major satellites that comprise 40 and 20% of the respective genome size of *P. ratzeburgii* and *P. subdepressus* (Ugarković et al. 1992; Plohl et al. 1998). These two satellites share no sequence homology but have similar repeat lengths of 142 bp (PRAT) and 144 bp (PSUB). In both sequences, a motif homologous to the human CENP-B box has been identified (Mravinac et al. 2005). PRAT and PSUB are found in a low number of copies in species belonging to the related genera *Tribolium*, *Tenebrio, Pimelia*, and *Latheticus*, subfamily Pimelinae, and family Chrysomelidae (Mravinac et al. 2002, 2005). Despite the fact that some of these species diverged from the *Palorus* group up to 60 Myr ago, PRAT and PSUB sequences remained virtually unchanged, showing no species diagnostic mutations and retaining the nonrandom pattern of variability along the sequence. It is assumed that the extreme preservation of these two satellites over a long period is related to their functional significance (Mravinac et al. 2005; Ugarković 2005).

Ancient satDNAs up to 80 Myr old have also been reported in some species of fish (De la Herrán et al. 2001) and whales (Arnason et al. 1984), and their remarkable preservation is thought to be related to the low mutation rates generally observed in aquatic environments (De la Herrán et al. 2001). However, functional constraints have been implicated in the preservation of salamander satellite 2 sequence – its promoter activity and the self-cleavage ability of its transcripts have remained conserved for 200 Myr (Green et al. 1993).

8.4 Transcription of Insects satDNAs

Insects represent one of the most diverse groups of animals, accounting for more than half of all known living organisms. Despite this, they are poorly represented as model organisms for the study of satDNA function. With the present study, satDNA transcription has been described in species from only four orders: Hymenoptera, Orthoptera, Diptera, and Coleoptera (Table 8.1).

Table 8.1 Transcription of satDNAs in insects

Species	satDNA		satDNA transcription			
	satDNA/repeat unit	Amount in the genome	Size of transcripts	Strand specific	Other features	Reference
Orthoptera						
Dolichopoda baccettii	pDo500 420-510-bp	5%	Monomeric and multimeric	N/A	Ribozyme, self-cleavage activity	Rojas et al. (2000)
Hymenoptera						
Symphyta. Diprionidae						
Diprion pini	Ps 311-bp	1.8–4%	Heterogeneous	Differential expression	Preferential transcription in females	Rouleux-Bonnin et al. (1996)
	Pv 280-bp	0.1–1.3%	Heterogeneous	N/A		Rouleux-Bonnin et al. (1996)
	Ec 148-bp	1.6–7%	Heterogeneous	N/A		Rouleux-Bonnin et al. (1996)
Apocrita. Apidae						
Bombus terrestris	BT 422-bp	2.4–3.2%	Discrete sizes (2.1 kb, 1.55 kb, 1.05 kb, 0.6 kb)	Strand specific in embryos Both strands in images		Rouleux-Bonnin et al. (2004)
Apocrita. Eupelmidae						
Eupelmus orientalis	107-bp	7%	Heterogeneous	N/A	Differential expression between sexes	Renault et al. (1999)
Eupelmus vuilleti	107-bp	25%	Heterogeneous	N/A	Sex-specific	Renault et al. (1999)
Apocrita. Formicidae						
Aphaenogaster subterranea	APSU 162-bp	4–16%	Long	Both strands		Lorite et al. (2002a, b)
Apocrita. Ichneumonidae						
Diadromus collaris	512-bp	5%	Discrete sizes	N/A	Differential expression between sexes	Renault et al. (1999)
Diadromus pulchellus	320-bp	15%	Discrete sizes (1.9 kb, 1.5 kb, 1.2 kb, 0.62 kb, 0.32 kb)	Both strands	Nonpolyadenylated, differential expression between stages Putative promoter elements	Renault et al. (1999)

Diptera						
Drosophila hydei	YSI 600-bp, YLI 693-bp, YLII 77-bp, YLIII 4-bp	3–9% of Y chromosomal DNA	Heterogeneous	Strand specific	Nuclei of primary spermatocytes	Trapitz et al. (1988)
Drosophila melanogaster	1.688 g/cm^3 (four subfamilies: 254, 353, 356 and 359-bp)	4%	siRNA (19–28 nt)	Both strands	In ovaries and testes	Aravin et al. (2003), Usakin et al. (2007)
	AAGAC repeats from satDNA 1.686 g/cm^3	2.4%	Heterogeneous	N/A	On Y chromosome loops of primary spermatocytes	Bonaccorsi et al. (1990)
Coleoptera						
Family Tenebrionidae						
Tribolium castaneum	TCAST 360-bp	17%	Long: ~0.5–5 kb; small: 21–28 nt	Threefold difference	Differential expression between stages partially polyadenylated	Pezer et al. unpublished
Palorus genalis	PGEN 429-bp	30%	Heterogeneous (~0.5–5 kb)	N/A		Pezer and Ugarković unpublished
Palorus ratzeburgii	PRAT 142-bp	40%	Heterogeneous (~0.5–5 kb)	Tenfold difference	Located in nuclei and cytoplasm Partially polyadenylated Putative promoter elements	Pezer and Ugarković (2008)
Palorus subdepressus	PSUB 144-bp	20%	Heterogeneous (~0.5–5 kb)	Both strands	Located in nuclei and cytoplasm Partially polyadenylated Putative promoter elements	Pezer and Ugarković (2009)

8.4.1 Hymenoptera

Transcription of satDNAs seems to be a general phenomenon in Hymenopteran suborder Apocrita, which includes ants, wasps, and bees (Renault et al. 1999; Lorite et al. 2002b). Satellite expression rates vary significantly between sexes, and female sex-specific transcripts are detected. Also, the amount of satellite transcripts differs among the queen, the worker, and the male. The presence of specific transcription factors might influence different rates of satellite expression in males and females, which suggests that transcription is regulated in *trans* rather than by elements inside the satellite sequence itself. The wasp *Diprion pini* has a higher satDNA expression in females, despite lower satDNA sequence content compared to the male genome, indicating that transcription is not constitutive (Rouleux-Bonnin et al. 1996). Gender-specific satDNA transcription in hymenopteran species could be related to sexual differentiation at the chromatin level and it is proposed that long multimeric transcripts probably have a structural role (Rouleux-Bonnin et al. 2004).

In addition to gender-specific expression, satDNAs seem to be differentially transcribed during Hymenopteran insect development, displaying higher expression in earlier stages. In the wasp *Diadromus pulchellus,* single-stranded, nonpolyadenylated transcripts heterogeneous in size but with discrete bands of 1.9 kb and 0.6 kb were detected. satDNA seems to be differentially expressed during developments since the transcripts were found to be more abundant in embryos and larvae than adults (Renault et al. 1999). In addition, putative promoters and transcription initiation site were mapped within the monomer sequence. In the bumble bee *Bombus terrestris*, multimeric transcripts arise from one strand preferentially in embryos opposed to both strands in imagos (Rouleux-Bonnin et al. 2004). It seems therefore that satDNA transcription interferes with development as well as sexual and caste-differentiation in this insect order.

Stage-specific transcription associated with differentiation has also been observed in systems other than insects. The most abundant mouse γ satDNA is differentially expressed in cells of the developing central nervous system as well as in adult liver and testis (Rudert et al. 1995; Chap. 5 in this book). In chicken and zebrafish, transcription of alphoid repeat sequences displays a specific temporal and spatial expression pattern during embryogenesis (Li and Kirby 2003).

8.4.2 Diptera

In Diptera, transcription of satDNA has been reported in primary spermatocytes of *D. melanogaster* and *D. hydei*. (Trapitz et al. 1988; Bonaccorsi et al. 1990). Transcripts of simple satellite sequence AAGAC that are highly heterogeneous in size, ranging from less than 1 kb to 10s of kb, have been found on Y chromosome loops. They do not appear to migrate to cytoplasm and are degraded during the first meiotic prophase. It has been proposed that the transcripts probably act as fertility

factors by providing a structural framework for accumulation of Y-encoded proteins involved in sperm differentiation.

In addition to the transcripts of simple sequence satellite AAGAC, transcripts of 1.688 *D. melanogaster* satellite with complex repeat unit of 359 bp were found in germinal tissues. Transcription proceeds from both DNA strands and is under the control of RNA interference machinery (Usakin et al. 2007). It is revealed that RNAi is necessary to maintain the silenced state of centromeric and pericentromeric 1.688 repeats located specifically on X chromosome. The heterochromatic locus on X chromosome that contains 1.688 satellite is responsible for hybrid female lethality in crosses between *D. simulans* females and *D. melanogaster* males (Ferree and Barbash 2009). It induces mitotic failure in early embryos due to the inability of 1.688 (359 bp) satellite block to form a proper heterochromatin state. Both *Drosophila* species share common satellites that differ in amount and location between the species, but 1.688 (359 bp) satellite is 50 times more abundant in *D. melanogaster* relative to *D. simulans*. It is proposed that hybrid female lethality occurs owing to the absence of the 1.688 satellite-derived small RNAs in the maternal cytoplasm that are required for heterochromatin establishment on 1.688 satellite array.

Transcription of complex telomeric repeats that are characteristic for chromosomal ends in Diptera was demonstrated in *Chironomus thummi* (Martínez-Guitarte et al. 2008). Transcripts are heterogeneous in length and correspond to multimers of the repeat. Moreover, transcription of telomeric repeats is not constitutive and is activated under conditions of environmental stress, such as heat shock.

8.4.3 Orthoptera

satDNAs were studied in the cave cricket genus *Dolichopoda (*Rhaphidophoridae). Three different satellite families were characterized in this species, among which pDo500 satellite is present in all species of the genus. The 500-bp satDNA family is actively expressed in the form of long multimeric transcripts, although the monomeric transcripts are also detected (Rojas et al. 2000). The transcripts act as ribozymes, as they have the ability to adopt hammerhead-like secondary structures and self-cleave in vitro. It is also possible that hammerhead sequences from the pDo500 satellite can *trans*-cleave host transcripts in the cells of *Dolichopoda*. The physiological role of these ribozymes is unknown, but it can be proposed that they may affect certain regulatory mechanisms in the cell. High sequence conservation of their corresponding satellites and active transcription suggests that they are under selective pressure (Rojas et al. 2000).

The hammerhead ribozyme structures associated with transcribed satDNA sequences have also been found in salamanders (Epstein and Gall 1987) and schistostomes (Ferbeyre et al. 1998). All hammerhead ribozymes detected in animal satDNA so far have been shown to self-cleave in *cis* long multimeric satellite transcripts into monomers.

8.4.4 Coleoptera

Within order Coleoptera, expression of satDNAs was investigated in species of genera *Palorus* and *Tribolium* that belong to family Tenebrionidae. Species of the two genera are characterized by the presence of large blocks of pericentromeric heterochromatin on all chromosomes and molecular analyses suggested that these blocks are composed almost exclusively of satDNAs that comprise up to 40% of the whole genome and encompass the regions of functional centromere (Ugarković et al. 1996; Durajlija Žinić et al. 2000).

In beetles *Palorus ratzeburgii*, *Palorus subdepressus*, and *Tribolium castaneum*, the major satDNAs called PRAT, PSUB, and TCAST, respectively, are continuously expressed during larval, pupal, and imago stages. The transcripts are of heterogeneous size, ranging from 0.5 kb to more than 5 kb, and originate from both strands of satDNA, albeit with a difference in expression between the two strands (Pezer and Ugarković 2009). Most of the transcripts are detected in the nucleus and are not polyadenylated. Although transcription from both DNA strands could potentially activate the RNA interference (RNAi) pathway, no processing of long PRAT and PSUB transcripts into small interfering RNAs (siRNA) was detected. However, small RNAs cognate to the major satellite TCAST (Ugarković et al. 1996) have been detected in the red flour beetle *T. castaneum*, (unpublished results). Small RNAs are more abundant in embryos than in later developmental stages, ranging in size between 21 and 26 nt with a predominant size of 24 nt. According to their size, these RNAs could be assigned to small interfering RNAs (siRNAs 21–23 nt) and piwiRNAs (piRNAs, 24–26 nt). piRNAs are characterized as the long class of siRNAs that bind to the Piwi clade of Argonaute proteins (Hamilton et al. 2002; Aravin et al. 2003). It is proposed that both types of small RNAs function as guide molecules during heterochromatin formation. The piwiRNAs, also known as repeat-associated RNAs ranging 23–26 nt in size, are most abundant in testes and early embryos, which may be related to dramatic changes in heterochromatin structure that occur in these stages. In addition to siRNAs and piRNAs, some components of the RNAi machinery have been identified in the sequenced genome of *T. castaneum*, such as Dicer and Argonaute protein families but not the RNA-dependent RNA polymerase (RdRP) gene (Tomoyasu et al. 2008). RdRP transcribes single-stranded RNA from an RNA template and is important for the production of siRNA as well as the amplification of the RNAi effect in fungi, protists, nematodes, and plants. However, it seems to be lacking in insects and vertebrates.

Multiple transcription initiation and termination sites as well as putative RNA Pol II promoter elements were mapped within PRAT and PSUB sequences. Overlapping promoter-like sequences on both DNA strands and the close position of transcription initiation sites suggest bidirectional activity of putative promoters. Presence of 5′ cap structure on portion of PRAT transcripts and susceptibility of transcription to inhibition of Pol II further confirm role of RNA Pol II in transcription of satDNAs. In addition, motifs similar to A and B boxes, associated with RNA

Pol III transcription, are located in PSUB and PRAT satellites. Thus, Pol III or some other polymerase(s) might be responsible for the production of the main nonpolyadenylated fraction of transcripts. Involvement of different RNA polymerases in the production of siRNAs was demonstrated in plants. Noncoding transcripts generated by RNA Pol II in *Arabidopsis* act as a scaffold for the recruitment of two other polymerases, Pol IV and Pol V, which seem to be important for the production of siRNAs (Zheng et al. 2009). The interplay of all three polymerases is required for siRNA-mediated transcriptional gene silencing in *Arabidopsis*.

The heterogeneous transcript size of satDNAs in beetles could be explained by the multiple transcription initiation and termination sites. Read-through transcription from the nearby gene promoters and transposable elements cannot be excluded either, although there is a strong indication that satDNAs are transcribed as autonomous transcription units from own promoters that reside within the satellite sequences (Pezer and Ugarković 2008, 2009). However, a smaller portion of PRAT and PSUB satDNA transcripts is polyadenylated and is found in the cytoplasm. Cytoplasmic localization and the presence of a polyA tail have been reported before for the satellite transcripts of various species. Polyadenylated transcripts of the G + C-rich satDNA of the Bermuda land crab are present in the cytoplasm of different tissues (Varadaraj and Skinner 1994). Satellite 2 is an abundant tandemly repeated sequence distributed in clusters throughout the genome of the newt *Notophthalamus viridescens* and is transcribed on lampbrush chromosomes. However, stable, strand-specific transcripts homologous to satellite 2 are present in the cytoplasm in a variety of different tissues (Epstein et al. 1986). Satellite III DNA is transcribed in response to stress in human cells, generating heterogeneous-sized RNAs that contain a polyA tail but remain in the nucleus (Valgardsdottir et al. 2005; see Chap. 5 in this book). In addition, many eukaryotic long ncRNAs that have regulatory roles are always polyadenylated (Amaral and Mattick 2008). For instance, the polyA tail is part of the mature *Xist* RNA, which mediates X chromosome inactivation in dosage compensation (Lucchesi et al. 2005; see Chap. 3 in this book)

In conclusion, expression of satDNAs in beetles is developmentally regulated and proceeds in the form of long, stable, nonpolyadenylated transcripts that remain mostly in the nucleus where they probably play a structural role in the organization of pericentromeric heterochromatin. A small portion of transcripts is exported to the cytoplasm where they perform an unknown role. In addition, long transcripts are processed into small RNAs, 21–26 nt long, that are proposed to function as guide molecules during heterochromatin formation.

8.5 Heterochromatin Formation in *Drosophila*: Role of Heterochromatic Transcripts

D. melanogaster heterochromatin is prominent in pericentromeric regions and is mostly comprised of satDNA and transposon elements (TE). As in fission yeast *Schizosaccharomyce pombe*, it is associated with histone H3 methylation on lysine 9

(H3K9) by the histone methylase Su(var)3-9 that enables recruitment of heterochromatin protein HP1 necessary to maintain and spread heterochromatic state (Ebert et al. 2006). It has been speculated for a long time whether an endogenous siRNA pathway, similar to those in *S. pombe*, is involved in the formation of heterochromatin in *Drosophila*. Small RNA molecules related to several types of repetitive DNA have been isolated from *D. melanogaster* (Aravin et al. 2003). These repeat-associated RNAs, 23–26 nt in size, are most abundant in testes and early embryos, which may be related to the regulation of transposon activity and the dramatic changes in heterochromatin structure that occur in these stages.

Examination and analysis of small RNA libraries obtained from different developmental stages of fly revealed the presence of TE-derived small RNAs in all stages: in early embryos, most of them correspond to 25 nt long piwiRNAs. They are formed in gonads from long transcripts of TEs and induce silencing of TEs through a feedback regulatory mechanism involving the Piwi subfamily of Argonaute proteins (Brennecke et al. 2007). In other developmental phases, 25 nt piRNAs are partially replaced by a population of 21 nt long RNAs that also derive from long TE transcripts. Due to the limitation of method of high throughput deep sequencing that is restricted to nontandemly repeated DNA, small RNAs that derive from satDNA were not systematically examined. However, siRNA deriving from 1.688 satellite have a size range between 19 and 28 nt and were detected in early embryos as well as in larvae (Aravin et al. 2003). It has been shown that a nuclear pool of TE-derived 21 nt long siRNAs is involved in heterochromatin formation in somatic cells of *Drosophila* and that components of the RNAi pathway participate in heterochromatin process (Fagegaltier et al. 2009). This implicates similarity between mechanisms of heterochromatin formation in *S. pombe* and *Drosophila* and points to the role of pericentromeric transcripts, either satDNA or transposon-derived, in heterochromatin formation. The possible mechanism by which repeat-derived siRNAs could promote heterochromatin formation in *Drosophila* is by tethering complementary nascent transcript of satDNAs and transposons and guiding chromatin modifiers, such as histone methylase Su(var)3-9, that induce H3K9 methylation. Identification of proteins that tether siRNAs to chromatin in *Drosophila* and other animals needs, however, to be elucidated.

In *Drosophila*, distinct heterochromatic loci are the source of primary piRNAs, which target a large number of transposons that are active in the germline and induce degradation of their transcripts. The mechanism of silencing is not well explained but includes Piwi proteins loaded with piRNAs that target and cleave RNA molecules. It is also proposed that piRNAs could promote chromatin modifications and, recently, a role for Rhino protein, one of the HP1-like proteins, in piRNA generation has been established (Klattenhoff et al. 2009). Rhino protein in germ cells of *Drosophila* replaces HP1 and seems to promote the expression of piRNAs and transposon silencing. It is suggested that piRNAs associated with Piwi protein target Rhino to transposon clusters and promote the production of additional piRNAs from the cluster. In this way, a link between the two major transposon defense pathways involving heterochromatin and RNA silencing mechanism exists (Klattenhoff et al. 2009). This could also represent another mechanism of

heterochromatin formation specific for germline. It can be also proposed that numerous satDNA-derived piRNAs present in germ cells could also contribute to heterochromatin establishment and maintenance by similar mechanism that might include Piwi protein and HP1 germline analog Rhino.

8.6 Possible Regulatory Role of satDNAs and Their Transcripts

Although satellite repeats show remarkable restriction in their distribution along chromosomes to pericentromeric and subtelomeric heterochromatin, there are several exceptions involving minor amounts of satellite sequences present in euchromatin. Such examples of limited localization of satellite sequences in euchromatin involve a simple and a complex satellite. Eight tandem repeats of *D. melanogaster* satellite AATAC are found in front of the s38 chorion gene on X chromosome (Spradling et al. 1987). The 359-bp repeats of the 1.688 satellite, located predominantly in pericentromeric heterochromatin of X chromosome, are also found in other positions of the same chromosome (Tartof et al. 1984). In beetle *T. castaneum* 360 bp repeats of abundant centromeric and pericentromeric satellite TCAST are found dispersed in the vicinity of genes on all chromosomes (unpublished results). The discovery of short satellite segments interspersed among the genes in euchromatic portion of genomes suggest possible regulatory role of these sequences, since they are often source of regulatory elements such as promoters and/or transcription factors binding sites (Ugarković 2005, Fig. 8.1). Recently, a regulatory role of 32 bp satellite repeats located in the intron of the major histocompatibility complex gene (MHIIβ) of fish *Salvelinus fontinalis*, on MHIIβ gene expression was demonstrated (Croisetiere et al. 2010). The level of gene expression depends on temperature being higher at lower temperatures as well as on the length of satellite repeats:

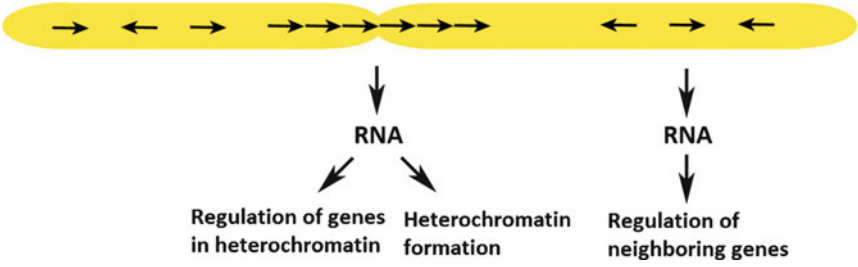

Fig. 8.1 Role of satDNAs and corresponding transcripts in the regulation of genes located in heterochromatin and euchromatin. Transcripts of tandemly repeated satellite repeats, located in (peri)centromeric regions, play a role in heterochromatin formation as well as in the regulation of the genes located in heterochromatin. Transcripts of satDNA repeats dispersed within the euchromatin could play a role in the regulation of the neighboring genes. Transcription of satellite repeats is temperature sensitive, and the role of transcripts in the environmental stress response is proposed

a longer satellite array induces reduced expression. Although the mechanism of *cis*-acting satellite gene regulation is not clear, there is evidence that temperature-sensitive satellites could play an important role in the gene regulation of the adaptive immune response.

Influence of satDNAs and their transcripts on gene regulation could not refer only to genes located in euchromatin but also on heterochromatic genes (Fig. 8.1). It is known that important developmental genes are located in heterochromatin, as revealed for *D. melanogaster,* (Pimpinelli et al. 1985) and that the proximity of heterochromatin is an important regulatory requirement for their function (Dimitri et al. 2009). Heterochromatin is also involved in gene silencing, and this process is developmentally programmed in *Drosophila* and mammals (Lu et al. 1998). Heterochromatin formation in *D. melanogaster* is influenced by transcripts of satDNA elements and transposons present in heterochromatin (see Sect. 8.5 in this Chapter). On the other hand, insect development is very sensitive to changes in the environment, particularly temperature. With a lowering temperature, the length of the development period is prolonged, and at a critical temperature, development ceases altogether. It has been shown that in beetle *T. castaneum,* expression of satDNA is temperature sensitive during embryogenesis, being significantly decreased at low temperatures where development is stopped (unpublished results). It can be proposed that decrease of satDNA expression affects heterochromatin formation during embryogenesis and in this way influences activity of heterochromatin-localized developmental genes. Temperature-sensitive expression of heterochromatic satDNAs also indicates their involvement in the signaling mechanism responsible for insect development, differentiation, and stress response (Fig. 8.1).

8.7 Conclusion

satDNAs are major heterochromatin constituents in many insect species and are found to be transcribed during all developmental stages. Transcripts are heterogeneous in size ranging from long multimers to small interfering RNAs. Their role in heterochromatin establishment and regulation is proposed although the detailed molecular mechanism and proteins involved are not elucidated yet. The satDNA transcription is not constitutive but associated with development and differentiation and is actively regulated by environmental factors such as temperature. It is proposed that satDNAs play a role in regulation of genes positioned within heterochromatin as well as those located in the vicinity of satellite elements in euchromatin. Mechanism of gene regulation is not explained but could be related to the presence of active regulatory elements within satDNAs such as promoters and transcription factor binding sites as well as corresponding satellite transcripts. Further studies are needed in order to explain the complex role of satDNAs and their transcripts in the signaling mechanism responsible for insect development, differentiation, and stress response.

Acknowledgments This work was supported by EU FP6 Marie Curie Transfer of Knowledge Grant MTKD-CT-2006-042248 and grant 00982604 from the Croatian Ministry of Science. Isidoro Feliciello is Marie Curie Fellow at Ruder Boskovic Institute.

References

Amaral PP, Mattick JS (2008) Noncoding RNA in development. Mamm Genome 19:454–492

Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B et al (2003) The small RNA profile during *Drosophila melanogaster* development. Dev Cell 5:337–350

Arnason U, Höglund M, Widegren B (1984) Conservation of highly repetitive DNA in cetaceans. Chromosoma 89:238–242

Bonaccorsi S, Gatti M, Pisano C, Lohe A (1990) Transcription of a satellite DNA on two Y chromosome loops of *Drosophila melanogaster*. Chromosoma 99(4):260–266

Borstnik B, Pumpernik D, Lukman D, Ugarković Đ, Plohl M (1994) Tandemly repeated pentanucleotides in DNA sequences of eukaryotes. Nucleic Acids Res 22(16):3412–3417

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA generating loci as master regulators of transposon activity in Drosophila. Cell 128:1089–1103

Bruvo-Mađarić B, Plohl M, Ugarković Đ (2007) Wide distribution of related satellite DNA families within the genus Pimelia (Tenebrionidae). Genetica 130:35–42

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

Coats SR, Zhang Y, Epstein LM (1994) Transcription of satellite 2 DNA from the newt is driven by a snRNA type of promoter. Nucleic Acids Res 22:4697–4704

Croisetiere S, Bernatchez L, Belhumeur P (2010) Temperature and length-dependent modulation of the MH class IIβ gene expression in brook charr (*Salvelinus fontinalis*) by a *cis*-acting minisatellite. Mol Immun 47:1817–1829

Davis CA, Wyatt GR (1989) Distribution and sequence homogeneity of an abundant satellite DNA in the beetle, *Tenebrio molitor*. Nucleic Acids Res 17(14):5579–5586

De la Herrán R, Fontana F, Lanfredi M, Congiu L, Leis M, Rossi R, Ruiz Rejón C, Ruiz Rejón M, Garrido-Ramos MA (2001) Slow rates of evolution and sequence homogenization in an ancient satellite DNA family of sturgeons. Mol Biol Evol 18:432–436

Diaz MO, Barsacchi-Pilone G, Mahon KA, Gall JG (1981) Transcripts from both DNA strands of a satellite DNA occur on lampbrush chromosome loops of the newt *Notophthalmus*. Cell 24:649–659

Dimitri P, Caizzi R, Giordano E, Carmela Accardo M, Lattanzi G, Biamonti G (2009) Constitutive heterochromatin: a surprising variety of expressed sequences. Chromosoma 118:419–435

Djupedal I, Kos-Braun IC, Mosher RA, Söderholm N, Simmer F, Hardcastle TJ, Fender A, Heidrich N, Kagansky A, Bayne E, Wagner EG, Baulcombe DC, Allshire RC, Ekwall K (2009) Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. EMBO J 28(24):3832–3844

Dover G (2002) Molecular drive. Trends Genet 18:587–589

Durajlija Žinić S, Ugarković Đ, Cornudella L, Plohl M (2000) A novel interspersed type of organization of satellite DNAs in *Tribolium madens* heterochromatin. Chromosome Res 8:201–212

Ebert A, Lein S, Schotta G, Reuter G (2006) Histone modification and the control of heterochromatin gene silencing in *Drosophila*. Chromosome Res 14:377–392

Epstein LM, Mahon KA, Gall JG (1986) Transcription of a satellite DNA in the newt. J Cell Biol 103:1137–1144

Epstein LM, Gall JG (1987) Self-cleaving transcripts of a satellite DNA in a newt. Cell 48:535–543

Fagegaltier D, Bougé AL, Berry B, Poisot E, Sismeiro O, Coppée JY, Théodore L, Voinnet O, Antoniewski C (2009) The endogenous siRNA pathway is involved in heterochromatin formation in Drosophila. Proc Natl Acad Sci USA 106:21258–21263

Ferbeyre G, Smith JM, Cedergren R (1998) Schistosome satellite DNA encodes active hammerhead-ribozymes. Mol Cell Biol 18:3880–3888

Ferree PM, Barbash DA (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. PLoS Biol 7:e1000234

Fitzgerald DJ, Dryden GL, Bronson EC, Williams JS, Anderson JN (1994) Conserved pattern of bending in satellite and nucleosome positioning DNA. J Biol Chem 269:21303–21314

Fry K, Salser W (1977) Nucleotide sequences of HS-α satellite DNA from kangaroo rat Dipodomys ordii and characterisation of similar sequences in other rodents. Cell 12:1069–1084

Gaubatz JW, Cutler RG (1990) Mouse satellite DNA is transcribed in senescent cardiac muscle. J Biol Chem 265:17753–17758

Green B, Pabon-Pena LM, Graham TA, Peach SE, Coats SR, Epstein LM (1993) Conserved sequence and functional domains in satellite 2 from three families of salamanders. Mol Biol Evol 10:732–750

Hamilton A, Voinnet O, Chappell L, Baulcombe D (2002) Two classes of short interfering RNA in RNA silencing. EMBO J 21:4671

Kipling D, Warburton PE (1997) Centromeres, CENP-B and *Tigger* too. Trends Genet 13:141–145

Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A, Seitz H, Zamore PD, Weng Z, Theurkauf WE (2009) The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell 138(6):1137–1149

Li YX, Kirby ML (2003) Coordinated and conserved expression of alphoid repeat and alphoid repeat-tagged coding sequences. Dev Dyn 228:72–81

Lohe A, Roberts P (1988) Evolution of satellite DNA sequences in *Drosophila*. In: Verma RS (ed) Heterochromatin: molecular and structural aspects. Cambridge University Press, Cambridge, pp 148–186

Lorite P, Carrillo JA, Tinaut A, Palomeque T (2002a) Comparative study of satellite DNA in ants of the *Messor* genus. Gene 297(1–2):113–122

Lorite P, Renault S, Rouleux-Bonnin F, Bigot S, Periquet G, Palomeque T (2002b) Genomic organization and transcription of satellite DNA in the ant *Aphaenogaster subterranea* (Hymenoptera, Formicidae). Genome 45:609–616

Lu BY, Ma J, Eissenberg JC (1998) Developmental regulation of heterochromatin-mediated silencing in *Drosophila*. Development 125:2223–2234

Lucchesi JC, Kelly WG, Panning B (2005) Chromatin remodeling in dosage compensation. Annu Rev Genet 39:615–651

Martínez-Guitarte JL, Díez JL, Morcillo G (2008) Transcription and activation under environmental stress of the complex telomeric repeats of *Chironomus thummi*. Chromosome Res 16 (8):1085–1096

Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J Cell Biol 109:1963–1973

Meštrović N, Plohl M, Mravinac B, Ugarković Đ (1998) Evolution of satellite DNAs from the genus *Palorus* – experimental evidence for the 'library' hypothesis. Mol Biol Evol 15:1062–1068

Metz A, Soret J, Vourc'h C, Tazi J, Jolly C (2004) A key role for stress-induced satellite III transcripts in the relocalization of splicing factors into nuclear stress granules. J Cell Sci 117: 4551–4558

Mravinac B, Plohl M, Meštrović N, Ugarković Đ (2002) Sequence of PRAT satellite DNA "frozen" in some coleopteran species. J Mol Evol 54:774–783

Mravinac B, Plohl M, Ugarković Đ (2004) Conserved patterns in the evolution of *Tribolium* satellite DNAs. Gene 332:169–177

Mravinac B, Plohl M, Ugarković Đ (2005) Preservation and high sequence conservation of satellite DNAs suggest functional constraints. J Mol Evol 61:542–550

Ohzeki J, Nakano M, Okada T, Masumoto H (2002) CENP-B box is required for *de novo* centromere chromatin assembly on human alphoid DNA. J Cell Biol 159:765–775

Palomeque T, Lorite P (2008) Satellite DNA in insects: a review. Heredity 100:564–573

Pezer Z, Ugarković Đ (2008) RNA Pol II promotes transcription of centromeric satellite DNA in beetles. PLoS One 3:e1594

Pezer Z, Ugarković Đ (2009) Transcription of pericentromeric heterochromatin in beetles – satellite DNAs as active regulatory elements. Cytogenet Genome Res 124:268–276

Pimpinelli S, Sullivan W, Prout M, Sandler L (1985) On biological functions mapping to the heterochromatin of *Drosophila melanogaster*. Genetics 109:701–724

Plohl M, Meštrović N, Bruvo B, Ugarković Đ (1998) Similarity of structural features and evolution of satellite DNAs from Palorus subdepressus (Coleoptera) and related species. J Mol Evol 46:234–249

Pons J (2004) Cloning and characterization of a transponsable-like repeat in the heterochromatin of the darkling beetle *Misolampus goudoti*. Genome 47:769–774

Pons J, Bruvo B, Juan C, Petitpierre E, Plohl M, Ugarković Đ (1997) Conservation of satellite DNA in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). Gene 205:183–190

Raff JW, Kellum R, Alberts B (1994) The Drosophila GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. EMBO J 13:5977–5983

Renault S, Rouleux-Bonnin F, Periquet G, Bigot Y (1999) Satellite DNA transcription in *Diadromus pulchellus* (Hymenoptera). Insect Biochem Mol Biol 29:103–111

Rojas AA, Vázquez-Tello A, Ferbeyre G, Venanzetti F, Bachmann L, Paquin B et al (2000) Hammerhead-mediated processing of satellite pDo500 family transcripts from *Dolichopoda* cave crickets. Nucleic Acids Res 28:4037–4043

Romanova LY, Deriagin GV, Mashkova TD, Tumeneva IG, Mushegian AR, Kisselev LL, Alexandrov IA (1996) Evidence for selection in evolution of alpha satellite DNA: the central role of CENP-B/pJ alpha binding region. J Mol Biol 261(3):334–340

Rouleux-Bonnin F, Renault S, Bigot Y, Periquet G (1996) Transcription of four satellite DNA subfamilies in Diprion pini (Hymenoptera, Symphyta, Diprionidae). Eur J Biochem 238: 752–759

Rouleux-Bonnin F, Bigot S, Bigot Y (2004) Structural and transcriptional features of *Bombus terrestris* satellite DNA and their potential involvement in the differentiation process. Genome 47:877–888

Rudert F, Bronner S, Garnier J-M, Dollé P (1995) Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mamm Genome 6:76–83

Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. Science 294:109–115

Shestakova EA, Mansuroglu Z, Mokrani H, Ghinea N, Bonnefoy E (2004) Transcription factor YY1 associates with pericentromeric *y*-satellite DNA in cycling but not in quiescent (G_0) cells. Nucleic Acids Res 32:4390–4399

Spradling AC, de Cicco DV, Wakimoto BT, Levine JF, Katfayan LJ, Cooley L (1987) Amplification of the X-linked *Drosophila* chorion gene cluster requires a region upstream from the s38 chorion gene. EMBO J 6:1045–1053

Sun X, Wahlstrom J, Karpen GH (1997) Molecular structure of a functional Drosophila centromere. Cell 91:1007–1019

Tartof KD, Hobbs C, Jones M (1984) A structural basis for variegating position effects. Cell 37:869–878

Tomoyasu Y, Miller SC, Tomita S, Schoppmeier M, Grossman D, Bucher G (2008) Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in *Tribolium*. Genome Biol 9:R10

Trapitz P, Wlaschek M, Bunemann H (1988) Structure and function of Y chromosomal DNA II. Analysis of lampbrush loop associated transcripts in nuclei of primary spermatocytes of

Drosophila hydei by in situ hybridization using asymmetric RNA probes of four different families of repetitive DNA. Chromosoma 96:159–170

Ugarković Đ (2005) Functional elements residing within satellite DNAs. EMBO Rep 6: 1035–1039

Ugarković Đ (2008) Satellite DNA libraries and centromere evolution. Open Evol J 2:1–6

Ugarković Đ (2009a) Centromere: structure and evolution. Springer, Berlin

Ugarković Đ (2009b) Centromere-competent DNA: structure and evolution. Prog Mol Subcell Biol 48:53–76

Ugarković Đ, Plohl M, Lucijanić-Justić V, Borštnik B (1992) Detection of satellite DNA in *Palorus ratzeburgii*: analysis of curvature profiles and comparison with *Tenebrio molitor* satellite DNA. Biochimie 74:1075–1082

Ugarković Đ, Podnar M, Plohl M (1996) Satellite DNA of the red flour beetle *Tribolium castaneum* – comparative study of satellites from the genus *Tribolium*. Mol Biol Evol 13: 1059–1066

Ugarković Đ, Plohl M (2002) Variation in satellite DNA profiles – causes and effects. EMBO J 21(22):5955–5959

Usakin L, Abad J, Vagin VV, de Pablos B, Villasante A et al (2007) Transcription of the 1.688 Satellite DNA family is under the control of RNA interference machinery in *Drosophila melanogaster* ovaries. Genetics 176:1343–1349

Valgardsdottir R, Chiodi I, Giordano M, Cobianchi F, Riva S, Biamonti G (2005) Structural and functional characterization of noncoding repetitive RNAs transcribed in stressed human cells. Mol Biol Cell 16:2597–2604

Varadaraj K, Skinner DM (1994) Cytoplasmic localization of transcripts of a complex G + C-rich crab satellite DNA. Chromosoma 103:423–431

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297:1833–1837

Waye JS, Willard HF (1986) Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome. Moll Cell Biol 6:3156–3165

Wu ZG, Murphy C, Gall JG (1986) A transcribed satellite DNA from the bullfrog *Rana catesbeiana*. Chromosoma 93(4):291–297

Zheng B, Wang Z, Li S, Yu B, Liu JY, Chen X (2009) Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes Dev 23:2850–2860

Chapter 9
Long Nonprotein-Coding RNAs in Plants

Virginie Jouannet and Martin Crespi

Abstract In recent years, nonprotein-coding RNAs (or npcRNAs) have emerged as a major part of the eukaryotic transcriptome. Many new regulatory npcRNAs or riboregulators have been discovered and characterized due to the advent of new genomic approaches. This growing number suggests that npcRNAs could play a more important role than previously believed and significantly contribute to the generation of evolutionary complexity in multicellular organisms. Regulatory npcRNAs range from small RNAs (si/miRNAs) to very large transcripts (or long npcRNAs) and play diverse functions in development and/or environmental stress responses. Small RNAs include an expanding number of 20–40 nt RNAs that function in the regulation of gene expression by affecting mRNA decay and translational inhibition or lead to DNA methylation and gene silencing. They generally involve double-stranded RNA or stem loops and imply transcriptional or posttranscriptional gene silencing (PTGS). RNA silencing besides small interfering RNA and microRNA, gene silencing in plants is also mediated by tasiRNAs (trans-acting siRNAs) and nat-siRNAs (natural antisense mediated siRNAs). In contrast to small RNAs, much less is known about the large and diverse population of long npcRNAs, and only a few have been implicated in diverse functions such as abiotic stress responses, nodulation and flower development, and sex chromosome-specific expression. Moreover, many long npcRNAs act as antisense transcripts or are substrates of the small RNA pathways, thus interfering with a variety of RNA-related metabolisms. An emerging hypothesis is that long npcRNAs, as shown for

V. Jouannet
Department of Stem Cell Biology, University of Heidelberg, INF230, 69120 Heidelberg, Germany and
Centre National de la Recherche Scientifique, Institut des Sciences du Végétal, 91198 Gif-sur-Yvette Cedex, France

M. Crespi (✉)
Centre National de la Recherche Scientifique, Institut des Sciences du Végétal, 91198 Gif-sur-Yvette Cedex, France
e-mail: martin.crespi@isv.cnrs-gif.fr

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3_9,

small si/miRNAs, integrate into ribonucleoprotein particles (RNPs) to modulate their function, localization, or stability to act on target mRNAs. As plants show a remarkable developmental plasticity to adapt their growth to changing environmental conditions, understanding how npcRNAs work may reveal novel mechanisms involved in growth control and differentiation and help to design new tools for biotechnological applications.

9.1 Introduction

In recent years, RNA researchers have shown a growing interest in a hidden part of the transcriptome: the nonprotein-coding RNAs (npcRNAs). This group of RNAs has a very poor protein-coding potential, but its function is associated with the RNA molecule itself. Although some npcRNAs code for small functional peptides, the bulk of them do not contain long ORFs and consequently, they have eluded bioinformatic searches mainly based on coding capacity. Nonetheless, new bioinformatics and experimental strategies as well as high-throughput sequencing of RNAs, large scale complementary DNA cloning, and microarray analysis have revealed an outstanding number of novel npcRNA candidates in various model organisms from yeast or plants to *Homo sapiens* (Mattick and Makunin 2006; Mercer et al. 2009; Yasuda and Hayashizaki 2008). Apart from the well-known housekeeping npcRNAs such as rRNA, tRNA, snoRNA, and snRNA, many new regulatory npcRNAs or riboregulators have been discovered and characterized (Mercer et al. 2009; Wilusz et al. 2009). NpcRNAs can be transcribed from intergenic regions, but they also include a surprising number of antisense RNA transcripts, pseudogenes, and truncated transcripts in eukaryotes. In fact, the transcriptome is surprisingly complex, with long npcRNAs often overlapping with or interspersed between coding transcripts. This complexity has created a shift in our understanding of gene expression as a DNA sequence can be transcribed in multiple sense and antisense transcripts, intronic npcRNAs, and intergenic or promoter-associated RNAs (Mercer et al. 2009). In *Arabidopsis thaliana*, whole-genome mapping based on the use of tiling arrays revealed that >30% of observed transcription was intergenic and that numerous antisense RNA transcripts exist (Yamada et al. 2003).

Regulatory npcRNAs or riboregulators include npcRNAs that are expressed at certain stage of development, during cell differentiation, or as a response to external stimuli, and can affect transcription or translation of other genes (Mattick and Makunin 2006; Yasuda and Hayashizaki 2008). According to their size, regulatory npcRNAs are classified as small npcRNAs (<40 bp) or long npcRNAs (>40 bp). Certain npcRNAs have been implicated in different regulatory mechanisms in plant development (Brown et al. 2008; Voinnet 2009), in environmental biotic interactions and abiotic stress responses (Ben Amor et al. 2009; Jay et al. 2010; Sunkar 2010), and/or shown to have specific localization at tissular, cellular and subcellular levels (Campalans et al. 2004; Zhan and Lukens 2010). These transcripts are

generally produced by RNA polymerase II and are generally capped and polyadenylated. Although several of these long npcRNAs have been experimentally identified in plants (Ben Amor et al. 2009; Charon et al. 1999; Franco-Zorrilla et al. 2007; Hirsch et al. 2006), few data exist on their activity, subcellular localization, or molecular roles. Globally, npcRNAs have been far less studied in plant than in animals where diverse mechanisms involving npcRNAs in the regulation of gene expression have been discovered (for review see Prasanth and Spector 2007; Voinnet 2009; Wilusz et al. 2009). Long npcRNAs can mediate epigenetic changes by recruiting chromatin remodeling complexes to specific genomic loci, as shown for the *HOTAIR* (*HOX* antisense RNA) that silences transcription across 40 kb of the human *HOXD* locus (Rinn et al. 2007). This process is mediated by the Polycomb chromatin remodeling complex PRC2 and the *HOTAIR* RNA. Homologs of certain members of the PRC complex have been identified in plants: their mutations affect heterochromatin organization and cell proliferation, and lead to spontaneous embryogenesis in plants (Chanvivattana et al. 2004), suggesting a link between heterochromatin regulation and plant development. Nevertheless, npcRNAs linked to the action of these plant *PRC*-like genes have not yet been identified. Other studies on long npcRNAs showed their ability to modulate nuclear activities of different proteins. For example, in the presence of an npcRNA, the Translocated in Liposarcoma (TLS) protein can change its conformation into an active form to inhibit the histone acetyltransferases CBP and P300, and silence the cyclin D1 in human cells (Wang et al. 2008). More recently, the *GAS5* npcRNA (for *GROWTH ARREST SPECIFIC5*) has been shown to bind specifically to the Glucocorticoid Receptor (GR) protein, a transcription factor involved in cell growth, and to inhibit its activity in HeLa cells (Kino et al. 2010). Finally, the transcription of an npcRNA across the promoter region of a downstream protein-coding gene may interfere with its expression pattern (Martens et al. 2004) or induce histone modifications, leading to the repression of transcription initiation (Houseley et al. 2008) or conversely, chromatin remodeling and opening to activate transcription (Hirota et al. 2008).

Several long npcRNAs are processed into small RNAs due to their folding as double-stranded RNA (dsRNA) loops derived from endogenous loci (as the microRNA or microRNA) or due to the action of RNA-dependent RNA polymerases that generate long dsRNAs. These dsRNA structures are processed into small RNA by member(s) of the Dicer family (Vaucheret 2006). Furthermore, introns may themselves code for nonprotein-coding RNAs such as the intronic miRNA or mirtrons (Ying et al. 2010), as also shown in plants for some microRNA (Hirsch et al. 2006; Brown et al. 2008). Interestingly, certain long npcRNAs play cytoplasmic regulatory roles such as inhibiting miRNA activity (Franco-Zorrilla et al. 2007), indirectly affecting microRNA expression (Ben Amor et al. 2009) or acting as *cis*- or *trans*-antisense RNAs leading to small small interfering RNA (Borsani et al. 2005; Katiyar-Agarwal et al. 2007). In contrast to long npcRNAs, much more is known about small npcRNAs, the si/miRNAs, which are key regulators of gene expression (Vaucheret 2006; Voinnet 2009). The small RNAs range from 20 to 40 nucleotides long, are derived from large npcRNA precursors, and play a major role in gene

silencing at transcriptional and posttranscriptional levels. There are many classes of small RNAs acting in the regulation of gene expression by different pathways, including small interfering RNAs (small interfering RNA), microRNAs (microRNA), heterochromatic siRNAs (hc-siRNAs), Piwi-interacting RNAs (piRNAs), *trans*-acting siRNAs (ta-siRNAs), and the naturally occurring antisense siRNAs (nat-siRNAs) (Jamalkandi and Masoudi-Nejad 2009; MacLean et al. 2010). Small si/miRNAs induce mRNA cleavage and translational inhibition through pairing with specific mRNA targets, mainly in the cytoplasm, or lead to transcriptional gene silencing (TGS) RNA silencing, heterochromatin formation, and de novo DNA methylation in the nucleus (Jamalkandi and Masoudi-Nejad 2009; Vaucheret 2006; Verdel et al. 2009). Although heterogeneous in size, sequence, genomic distribution, biogenesis, and action, most of these molecules mediate repressive gene regulation through a mechanism often referred to as RNA silencing or RNA interference (RNAi). Their main role relies on the maintenance of genome integrity and developmental patterning as well as on the generation of novel regulatory mechanisms to help plants to adapt and respond to adverse biotic and abiotic environmental conditions (Ruiz-Ferrer and Voinnet 2009).

In this review, we first discuss a major class of long npcRNAs, the Natural Antisense Transcripts (NATs); second, we present the biogenesis and action of the small si/miRNA derived from long double-stranded RNAs; and finally, we introduce the long npcRNAs that interact with specific RNA-binding proteins to modulate their action or localization. Globally, both long and small npcRNAs integrate cellular RNP networks controlling the final outcome of the transcriptome (Fig. 9.1).

9.2 Natural Antisense Transcripts Include a Major Class of Long ncpRNA in Plants

The study of eukaryotic genomes has revealed a large proportion of overlapping genes: about 22% of all genes overlap in humans (Chen et al. 2004), about 15% in mice (Kiyosawa et al. 2003) and Drosophila (Misra et al. 2002), and 6–9% in plants (Osato et al. 2003; Wang et al. 2005a). Transcription of overlapping gene pairs in a convergent orientation therefore allows the production of antisense transcripts. If these antisense transcripts were first observed in transgenic experiments, it has been clearly shown that Natural Antisense Transcript also occur. Therefore, Natural Antisense Transcript can be defined as endogenous RNA molecules that are transcribed from the opposite DNA strand to other transcripts and whose partial or entire sequences exhibit their complementarity to other transcripts. As in most species, the majority of *Arabidopsis* NATs pairs (72%) overlapped at their 3′ end (Wang et al. 2005a) and, for 99% of them, the overlapping region included exon sequences. In the remaining cases, one of the transcripts is entirely transcribed from intronic regions of the other strand transcript. Both sense and antisense RNAs can encode proteins or be npcRNAs. However, the most prominent form of antisense

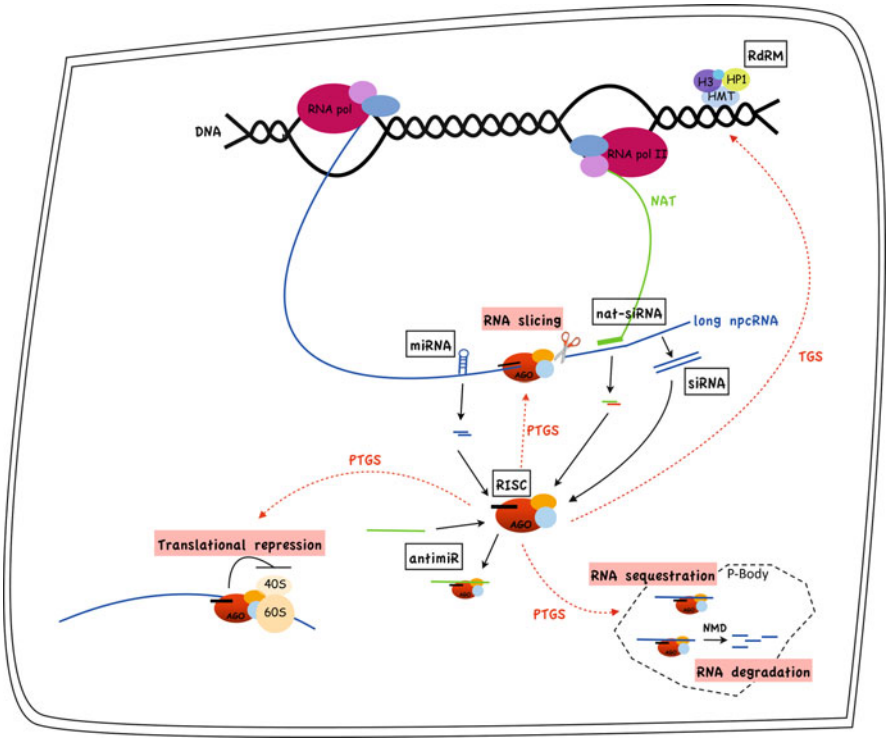

Fig. 9.1 The nonprotein-coding RNA network in plants. Within the cell, npcRNAs may act at different levels (Transcriptional (TGS), Posttranscriptional or translational (PTGS)). Long npcRNAs are produced by RNA polymerase II and can be precursors of small RNAs or antisense RNAs (NATs) of other transcripts. Once incorporated into the RISC effector complex, small RNAs from different pathways of RNA silencing (miRNA, siRNA or nat-siRNA) can act directly on the target mRNA (slicing, RNA degradation, or sequestration through the NMD mechanisms) or on the translation by mechanisms that remain poorly understood. In addition, they can lead to epigenetic changes and DNA methylation (TGS). The regulation of small npcRNA is finely controlled and can be controlled by target mimicry with long npcRNAs (antimiR). Hence, long and short npcRNAs form a network of ribonucleoproteins within the cell that may affect the expression patterns of coding mRNAs

transcription seems to be a protein-coding RNA overlapping with a nonprotein-coding antisense transcript (Faghihi and Wahlestedt 2009). Because both ends of protein-coding genes have a propensity for Natural Antisense Transcript, Natural Antisense Transcript are not evenly distributed across the genome; specifically, antisense transcription is enriched 250 nucleotides upstream of the transcription start site (TSS) (Seila and Sharp 2008) and 1.5 kb downstream of sense genes (Sun et al. 2005).

There are two types of Natural Antisense Transcript. *Cis*-Natural Antisense Transcript are transcribed from the same genomic loci as their sense transcripts but on the opposite DNA strand, in which case the sequence complementarity

between two transcripts is directly related to the overlapping region. By contrast, *trans*-Natural Antisense Transcript originate from genomic regions distinct from those encoding their overlapping sense transcripts. *Cis*-Natural Antisense Transcript usually have a long perfect complementarity between the sense and antisense transcripts, whereas the *trans*-Natural Antisense Transcript often have short and imperfect complementarity (Wang et al. 2005b). In the plant kingdom, several *cis*-Natural Antisense Transcript have been studied. In *Arabidopsis*, 1,340 potential *cis*-Natural Antisense Transcript were identified (Wang et al. 2005a) and the expression of sense and antisense transcripts for 957 *cis*-NATs pairs was confirmed using sequence information of *Arabidopsis* full-length cDNA and massively parallel signature sequencing (MPSS) data. In an independent study, Jen et al. (2005) reported the existence of 1,083 transcript pairs that overlapped in antisense orientation involving 2,147 independent genes. These overlapping genes can be arranged in convergent or divergent manners, although as in other species, the majority of overlapping gene pairs (956 pairs among the 1,083 identified by Jen et al. 2005) are organized with overlapping regions comprising between 1 and 2,820 bp (mean length of 431 bp) in a convergent manner. Although *trans*-Natural Antisense Transcript have been less studied, the existence of 1,320 *trans*-Natural Antisense Transcript pairs was proposed within the *Arabidopsis* genome (Wang et al. 2006). Among them, 658 pairs were supported by either full length cDNA for one transcript, and the remaining 218 pairs were identified solely by comparing annotated gene sequences. One important question is whether these overlapping transcripts exist in the same cell. Interestingly, among the *trans*-Natural Antisense Transcript pairs where in situ hybridization data exist for both transcripts, 67% of them are present in the same cell and with a comparable level of expression. These results suggest that sense and antisense pairing transcripts may interact with one another, particularly to form double-stranded RNA duplexes (dsRNAs). Unlike *cis*-Natural Antisense Transcript pairs where one sense transcript usually has only one antisense partner, one or several potential antisense transcripts are commonly predicted in *trans*-Natural Antisense Transcript pairs. In certain cases, one sense transcript formed different dsRNAs with transcripts derived from the same gene as a result of alternative splicing. Comparison with previously reported *Arabidopsis* *cis*-Natural Antisense Transcript data revealed that 430 transcripts on the *trans*-Natural Antisense Transcript category also had *cis*-Natural Antisense Transcript (Henz et al. 2007), suggesting that antisense transcripts might form complex regulatory networks in *Arabidopsis*.

Long npcRNAs including the antisense npcRNAs have to bypass several RNA-quality control mechanisms occurring in the cell. For example, a genome-wide analysis of exosome substrates in *A. thaliana* revealed, in addition to mRNA and miRNA processing intermediates, hundreds of npcRNAs and antisense RNAs not previously described (Chekanova et al. 2007). The exosome is a macromolecular complex that mediates RNA processing and degradation and is generally essential for viability in eukaryotes. These npcRNAs only detected in exosome mutants include large numbers of antisense RNAs as they are rapidly and actively degraded in wild-type plants. Similarly, the nonsense-mediated mRNA decay (or NMD) is an

mRNA quality control mechanism related to cytoplasmic foci known as P-bodies, which recognizes premature nonsense or stop codons (PTC) within an mRNA (Conti and Izaurralde 2005). After recognition of an incorrectly positioned stop codon, the nonsense-mediated mRNA decay system signals the elimination of the mRNAs through decapping, deadenylation, and exonucleolytic degradation. The UP-frameshift proteins (UPFs) are essential for nonsense-mediated mRNA decay, and three *UPF* genes exist in *A. thaliana*. A genome-wide analysis of these mutants revealed that, in addition to the expected nonsense-mediated mRNA decay substrates, most npcRNAs including large numbers of antisense RNAs are degraded by this pathway, suggesting that one of the most important roles of nonsense-mediated mRNA decay is the genome-wide suppression of aberrant or antisense RNAs (Kurihara et al. 2009). Hence, the steady-state of antisense RNAs and not only its existence or synthesis is important to be considered in relation to their influence on gene expression.

9.3 Long and Short npcRNAs Are Involved in the RNA Silencing Mechanism

As mentioned above, npcRNAs can lead to the generation of dsRNAs and trigger "RNA silencing," a highly conserved process in eukaryotes depending on small RNAs (Fig. 9.2). The basic mechanism is initiated by dsRNA, substrates of the Dicer RNAses, which produces 21–30 nt small RNA duplexes. These small RNAs are then loaded by a member of the ARGONAUTE (AGO) within the RISC effector complex, conferring target specificity to this complex. Plants have evolved numerous RNA silencing pathways, which control multiple aspects of plant development, including its adaptation to the environment, and form the basis of an RNA-based immunity against viruses. The different pathways leading to synthesis of small RNAs from long npcRNA precursors will be described below.

9.3.1 The miRNA Pathway

The first step in the production of an microRNA is the transcription of a long npcRNA (~1 Kb) by RNA polymerase II from an MIR gene distinct from the target gene (Xie et al. 2005). These first transcripts called primary miRNA (pri-miRNA) are capped and polyadenylated, and have the potential to form highly folded structures. In *Arabidopsis*, this structure is recognized by the DICER-LIKE1 protein (DCL1) that will process the pri-miRNA generating first a 70–100 nt precursor miRNA called pre-miRNA, which can be folded into a stem-loop structure. To liberate the miRNA/miRNA* duplex, the activity of DCL1 is then coordinated with the activity of HYPONASTIC LEAVES (HYL1) and SERRATE (SE) within a macromolecular complex in the nucleus (Han et al. 2004; Vazquez et al. 2004) Both strands of the miRNA–miRNA* duplex are methylated by HUA ENHANCER (HEN1), a small RNA methyltransferase that methylates the 2′-

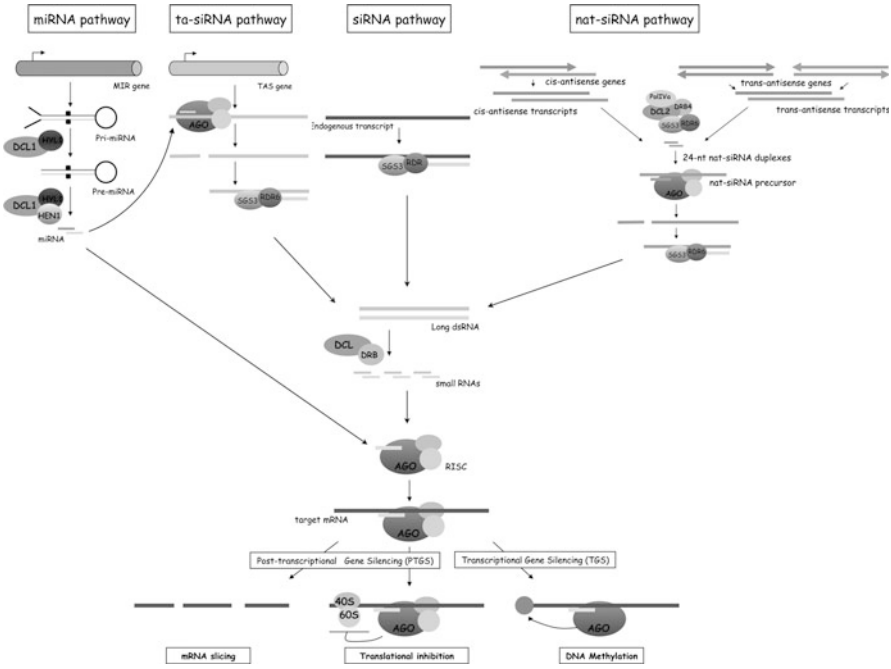

Fig. 9.2 Small RNA pathways in plants. In plants, the different small RNA silencing pathways differ mainly in the way of generation of the small RNA. The basic RNA silencing mechanism is initiated by a long double-stranded RNA (ta-si nat-si and siRNA pathways) or by endogenous loci able to form double-stranded stem-loops (miRNA pathway). In the case of the miRNA pathway, the transcript of an endogenous gene folds forming a stem loop dsRNA. In the siRNA pathway, a single-stranded RNA is targeted by RNA-dependent RNA polymerases and form long double-stranded RNAs to be cut by DICERs. Finally, in the tasiRNA pathway, the transcript of an endogenous long npcRNA is targeted by a specific miRNA and cleaved. The cleaved products become substrates of RNA-dependent RNA polymerases and form the dsRNA. The dsRNA molecule is processed by DICER ribonucleases type III (DCLs) into dsRNA small molecules. One strand of the processed si/miRNAs duplex is incorporated into a multiprotein complex called RISC containing AGO proteins. The presence of this small RNA provides the RISC complex a sequence specificity for the recognition through base complementarity with the target mRNA molecule. Target inhibition can occur at both posttranscriptional (PTGS, through mRNA cleavage and/or translation inhibition) and/or transcriptional levels (TGS, through DNA methylation)

hydroxy termini of miRNA– miRNA* imperfect duplexes (Yu et al. 2005; Li and Ding 2005). Methylation likely protects small RNA from degradation. DCL1-processed miRNAs are characterized by 2-nucleotide 3′-overhangs (Kurihara and Watanabe 2004). One of the active (or mature) miRNA strands with 2-nt 3′-overhangs is more stable (Reinhart et al. 2002; Kasschau et al. 2003) and is loaded onto the RNA-induced silencing complex (Hammond et al. 2000). AGO proteins are the major components of the RISC complex (Vaucheret et al. 2004; Baumberger and Baulcombe 2005). This protein binds to the 3′ miRNA overhang through its PAZ domain (Carmel et al. 2002). Thereafter, the RISC complex is guided by this miRNA strand to the complement mRNA target, possibly through a helicase-

scanning mechanism (Kidner and Martienssen 2005). The target mRNA is then cleaved between the 10th and 11th bases starting from the 5′ end of the miRNA match. The cleaved fragments of the target mRNA are then broken down by 5′-3′ EXORIBONUCLEASE4 (XRN4) most probably in the cytoplasm. In addition, the RISC complex can mediate translational inhibition by an unknown mechanism (Brodersen et al. 2008; Voinnet 2009).

In addition, certain "young" microRNA (e.g., MIR822 and MIR839) showing poor conservation can be processed by DCL4 (Rajagopalan et al. 2006) instead of DCL1. A transcriptomic study identified a novel DCL4 young-processed miRNA gene, MIR869a, due to its accumulation in *dcl4* mutants (Ben Amor et al. 2009). These transcripts may be processed by DCL4 because their precursors adopt an unusually stable secondary structure closer to that of a perfect dsRNA and different from that of conserved miRNA precursors containing several mismatches and processed by DCL1 (Voinnet 2009). Furthermore, it suggests that long dsRNAs may evolve into microRNA by accumulating mismatch mutations along the stem and then becoming better substrates of DCL1 (Fahlgren et al. 2007). In contrast to conserved microRNA, mainly producing miRNA and miRNA*, young miRNAs generally produce several accompanying small interfering RNA from their npcRNA precursors.

Long npcRNAs can also contribute to regulation of the small RNA activity. Indeed, the interaction between an npcRNA and a complementary miRNA can prevent the miRNA interaction with its coding mRNA target. This mechanism, called target mimicry, has been described for the IPS1 npcRNA (INDUCED BY PHOSPHATE STARVATION1) in *Arabidopsis* (Franco-Zorrilla et al. 2007). The IPS1 npcRNA has a strong complementarity with miR399, a microRNA specifically induced in response to phosphate stress starvation. However, the IPS1 and miR399 are not completely complementary, and the pairing with the microRNA is interrupted by a mismatch at the 10th–11th position, the expected site of miR399 cleavage. This interruption causes that the IPS1 npcRNA is not cleavable and likely blocks miR399 action by sequestering the microRNA. Hence, this npcRNA is mimicking a target, preventing the enzymatic cleavage of miR399 on its other mRNA targets (target mimicry). Through co-expression of the miRNA, its mRNA target, and the IPS1 npcRNA, it has been shown that this npcRNA can block microRNA regulation in *Arabidopsis* (Franco-Zorrilla et al. 2007). Furthermore, using related constructs for other microRNA, the IPS npcRNA could be modified to block the action of many other microRNA. It is likely that other long npcRNAs may interfere with microRNA action through this mechanism, but the detection of mismatched npcRNA/miRNA interactions needs to be carefully evaluated to distinguish between potential mimicries and/or nontargets.

9.3.2 The siRNA Pathway

As mentioned above, RNA silencing leads to transcription inhibition Transcriptional Gene Silencing (TGS) or mRNA degradation Post-transcriptional Gene

Silencing (PTGS), and this process is linked with the accumulation of small interfering RNA corresponding to the silenced sequence. Generally, small interfering RNA target the RNAs from which they derive and protect the genome from exogenous DNA or RNA such as transposons, viruses, and transgenes. In plants, co-expression of sense and antisense transgenes, called sense posttranscriptional gene silencing (S-PTGS), or expression of transgenes containing internal repeats (IR-PTGS) was reported to trigger this phenomenon (Béclin et al. 2002). In S-PTGS, RNA-DEPENDENT RNA POLYMERASE 6 (RDR6), one of six RDRs in *Arabidopsis*, recognizes the transgenes transcript with aberrant features (such as lack of 5′ cap and poly-A tail) to generate dsRNAs with the help of the coiled-coil protein SUPPRESSOR OF GENE SILENCING3 (SGS3) (Dalmay et al. 2000a, b; Mourrain et al. 2000). Among the four DICER-like proteins identified in *Arabidopsis*, DCL1 and DCL4 produce 21 nt small RNAs, DCL2, 22 nt-long small interfering RNA and DCL3, 24 nt-long small interfering RNA (Voinnet 2009). As mentioned earlier, DCL1 is involved mainly in microRNA production, DCL2 and DCL4 in viral resistance, DCL3 in transcriptional silencing (TGS)RNA silencing, and DCL4 in posttranscriptional silencing and Trans-acting siRNA production (Henderson et al. 2006). DCL2 and DCL4 are the two enzymes that process small interfering RNA from dsRNA during S-PTGS and IR-PTGS (Dunoyer et al. 2005; Xie et al. 2005). small interfering RNA products are then methylated by HEN1 and incorporated into the RISC complex. Various genetic screens showed the involvement of other proteins in gene silencing, such as NUCLEAR RNA POLYMERASE IVa (NRPD1a), RNA-DEPENDENT RNA POLYMERASE2 (RDR2), SILENCING DEFECTIVE3 (SDE3), and WERNER EXONUCLEASE (WEX) (Dalmay et al. 2001; Glazov et al. 2003; Herr et al. 2005).

The RNA-based immune response against virus infection implies part of the PTGS machinery (AGO1, HEN1, RDR6 and SGS3), suggesting that transgene-derived RNA produced during S-PTGS mimic viral RNAs (Mourrain et al. 2000; Morel et al. 2002). Moreover, the existence of 10 AGOs, 4 DCLs, and 6 RDRs in the plant model *Arabidopsis* (Morel et al. 2002; Schauer et al. 2002; Yu et al. 2005) suggests many possible siRNA pathways to respond to the different viruses. Indeed, *rdr6* mutants show hyper-susceptibility to diverse viruses, but not TMV (Dalmay et al. 2000b, 2001; Mourrain et al. 2000; Qu et al. 2005; Schwach et al. 2005), whereas *rdr1* mutants show hyper-susceptibility only to TMV. This suggests that different posttranscriptional siRNA-mediated pathways can be likely activated in response to different environmental conditions.

9.3.3 The ta-siRNA Pathway

This class of endogenous small RNAs, which seems to be plant-specific, implies elements of the miRNA and siRNA pathways. The TAS genes are long npcRNAs (Peragine et al. 2004; Vazquez et al. 2004; Allen et al. 2005) that are themselves targets of specific microRNA. One of the two single-stranded TAS cleavage

products is then converted to a dsRNA through the action of RDR6 and a coiled-coil protein SUPPRESSOR OF GENE SILENCING3 (SGS3). RDR6-derived TAS dsRNAs are sequentially processed into 21-nt Trans-acting siRNA by DCL4 associated with the dsRNA-binding protein DRB4 (Gasciolli et al. 2005; Yoshikawa et al. 2005; Hiraguri et al. 2005; Adenot et al. 2006). As Trans-acting siRNA, Trans-acting siRNA are methylated by HUA ENHANCER1 (HEN1) (Lie et al. 2005) and show a high level of complementarity with certain endogenous mRNAs. Interestingly, different members of the same gene family can be targeted by either microRNA or ta-siRNAs. For example, members of the same PPR subfamily are targeted by either miR161 or TAS gene ta-siRNA (Rhoades et al. 2002; Allen et al. 2005; Yoshikawa et al. 2005). On the contrary, the TAS3 pathway is unique because TAS gene-derived ta-siRNA biogenesis requires the initial miR390-mediated cleavage of the precursor TAS gene by a specific AGO protein, AGO7. miR390 is uniquely adapted to initiate TAS gene ta-siRNA biogenesis due to its specific association with AGO7 (Fahlgren et al. 2006; Montgomery et al. 2008). The TAS3 pathway plays an essential role in the proper timing and patterning in leaves, by repressing activity on Auxin Response Factor members (ARF2, ARF3, and ARF4) (Adenot et al. 2006; Fahlgren et al. 2006; Hunter et al. 2006). Mutations that impair Trans-acting siRNA production accelerate the juvenile to adult phase transition and cause elongated and curved leaves. Moreover, a recent study has shown the involvement of this pathway in controlling the development of lateral roots (Marin et al. 2010). These results show how a conserved pathway process can be involved in the development of different organs in plants.

9.3.4 The nat-siRNA Pathway

Another class of endogenous siRNAs derivating from pairs of natural *cis*-antisense transcripts was discovered in plants. Antisense overlapping gene pairs of d-pyrroline-5-carboxylate dehydrogenase (*P5CDH*), an intermediate in proline synthesis and catabolism, and a stress-induced gene, *SRO5*, transcribed in antisense orientation, generate two types of siRNAs of 24-nt and 21-nt, the so-called Natural antisense mediated siRNA (Borsani et al. 2005). Upon induction of SRO5 by salt stress, a 24-nt SRO5-P5CDH nat-siRNA is produced that can guide the cleavage of the P5CDH transcript and leads to the synthesis of further 21-nt P5CDH nat-siRNAs. Hence, Natural antisense mediated siRNA downregulate the expression of P5CH by mRNA cleavage after salt stress. This not only leads to proline accumulation, a metabolite linked to salt tolerance, but also causes increased ROS production, a response counteracted by the SRO5 protein. Thus, the SRO5-P5CDH nat-siRNAs together with the P5CDH and SRO5 proteins determine a regulatory loop controlling ROS production and stress responses in *Arabidopsis*. The nat-siRNAs-mediated cross-regulation of P5CDH and SRO5 mRNAs and the functional relationship of these two proteins may suggest a regulatory model that may be applied to other *cis*-antisense gene pairs.

Another example of the generation of Natural antisense mediated siRNA was found in response to the bacterial pathogen *Pseudomonas syringae* (Ps) carrying the effector avrRpt2 (Katiyar-Agarwal et al. 2006). This inductive response requires the cognate host disease resistance (R) gene RPS2 and the NDR1 gene, also required for RPS2-specified resistance. The nat-siRNA precursor transcript ATGB2 is specifically induced by Ps (avrRpt2) and requires DCL1, HYL1, HEN1, RDR6, SGS3, and RNA polymerase IVa to form the nat-siRNA ATGB2 that silences target PPRL. This PPRL gene is probably a negative regulator of the RPS2 signaling pathway, and its silencing by the nat-siRNA ATGB2 plays a positive role in disease resistance. Finally, Natural antisense mediated siRNA were also identified in Rice (Lu et al. 2008), suggesting that this pathway is common in plants. The biogenesis of this 22-nt Natural antisense mediated siRNA again revealed an intricate regulation of endogenous small interfering RNA formation and proposes a new role for DCL2. This specific pathway seems to link specific roles of these npcRNAs in plant adaptation to environmental conditions.

The different small RNA pathways in plants involve dsRNAs originated from different sources (such as endogenous loci, NATs or other long npcRNAs) and specific members of the DCL, RDR, and AGO gene families (Fig. 9.2).

9.4 Biological Roles of npcRNAs in Plants

After npcRNAs were identified, the question of their biological roles was raised. For several years, numerous studies have shown that these npcRNAs play regulatory roles in a broad range of events (Wilusz et al. 2009). In plants, apart from their role in development, adaptation to environmental conditions, and protection against pathogens by RNA silencing, npcRNAs seem important for controlling the circadian clock genes (Crosthwaite 2004); for the epigenetic regulation of transcription, through DNA methylation (Tufarelli et al. 2003, Lewis et al. 2004); for chromatin modification by genomic imprinting (Moore et al. 1997); and for RNA editing (Peters et al. 2003; Kim et al. 2004). In contrast, a role of npcRNAs in alternative splicing in plants has not yet been clearly demonstrated.

9.4.1 Implication of npcRNAs in Circadian Cycle

In most living organisms, biological processes oscillate according to the circadian rhythms. These oscillations imply the existence of a circadian system that controls the biological pathways in response to changes in light and temperature. Many proteins and transcription factors form the endogenous timing mechanism known as circadian clock (for review see Mas and Yanovsky 2009). A recent study in *Arabidopsis* has highlighted that some npcRNAs also follow a circadian rhythm (Hazen et al. 2009). Among the protein-coding genes detected by the arrays, 7% present rhythmic NATs such as antisense transcripts for the core-clock-associated

MYB transcription factors LHY and CCA1 and the PSEUDO RESPONSE REGULATORS (TOC1, PRR3, 5, 7 and 9). Even though the role of these NATs in circadian rhythms is not yet known, their mechanisms may be similar to those described in *Neurospora crassa* for the FREQUENCY gene (FRQ) (Kramer et al. 2003). In addition, Hazen et al. (2009) have also shown that certain microRNA have a cyclic expression: MIR160b, MIR167d, MIR158a, and MIR157a. MIR160 and MIR167 target members of the *AUXIN RESPONSE FACTOR* family (ARF10, 16 and 17 for MIR160; ARF6 and ARF8 for MIR167). MIR157 target members of the *SQUAMOSA BINDING PROTEIN* family, SPL3, SPL4, and SPL5, whereas no target is known for MIR158A. Furthermore, the TAS gene npcRNA that forms tasi-RNAs targeting ARF2, ARF3, and ARF4 genes, whose expression is circadian-clock regulated, could be another link between circadian rhythms and npcRNAs. However, there is lack of a firm demonstration of their role in the regulation of the circadian clock in roots.

9.4.2 Epigenetic Regulation and npcRNA

RNA-directed DNA methylation (RdDM) leads to *de novo* methylation of cytosine residues within the region of sequence identity between the triggering RNA and the target DNA (Aufsatz et al. 2002). As mentioned above, RNA-directed DNA methylation requires a dsRNA formed by RNA DEPENDENT RNA POLYMERASE 2 (RDR2) and processing by DICER LIKE3 (DCL3) into 24-nt small interfering RNA, which are methylated by HEN1 in the Cajal bodies (Yang et al. 2006). These 24nt small interfering RNA are then incorporated into a complex containing AGO4 and a specific RNA polymerase, RNA Pol IV. This complex interacts with DNA methyltransferases such as DRM2 or DRD1 to facilitate DNA cytosine methylation all along siRNA homologous sequences. After this *de novo* methylation, DNA METHYLTRANSFERASE1 (MET1) and CHROMOMETHYLASE3 (CMT3) contribute to maintain the CG and CNG methylation, respectively (Matzke and Birchler 2005). Hence, the generation of dsRNAs and small interfering RNA can lead to epigenetic modifications in chromatin and affect gene expression.

In plants, NATs, through the formation of dsRNAs, can lead to the generation of epigenetic marks. Recently, an antisense RNA has been involved in the epigenetic regulation of flowering. Indeed, the targeted 3′ processing of antisense transcripts at the locus encoding the major flowering repressor FLOWERING LOCUS C (FLC) triggers its silencing in *Arabidopsis* (Liu et al. 2010). FLC is a repressor of several major floral regulators, and vernalization, the regulation of flowering competence through cold exposure of plants, leads to the deposition of epigenetic marks in this locus and activates early flowering (Simpson et al. 2003). Several of the vernalization genes are homologs of the PRC2 complex related to Polycomb genes and linked vernalization-induced chromatin changes to PRC2. This epigenetic control results in FLC transcriptional silencing through the activities of two RNA-binding proteins or RNA-Binding proteins (FCA and FPA), a member of a 3′ RNA processing complex

and a histone demethylase (Liu et al. 2007). A suppressor mutagenesis screen and a detailed analysis of FLC locus transcription revealed the 3′ processing of FLC antisense (but not sense) transcripts. A specific RNA-Binding proteins directs the 3′ processing activities to a proximal antisense polyadenylation site, a targeted processing that triggers local histone demethylation and leads to FLC sense silencing during vernalization (Liu et al. 2010). Hence, the 3′ processing of antisense transcripts may be a general mechanism that triggers chromatin silencing in eukaryotes and heritable changes of gene expression, as well as inducing environmentally driven epigenetic changes.

Another example of epigenetic regulation is the genomic imprinting of a specific locus during seed development. Indeed, transcriptional repression through Polycomb group (PcG) proteins implies the methylation of histone H3 lysine 27 (H3K27), and the deposition of these marks leads to epigenetic inheritance of repressed transcriptional states. One of the *Arabidopsis* Polycomb group complex is composed of MEDEA (MEA), MULTICOPY SUPPRESSOR OF IRA1 (MSI1), and the ESC homolog FERTILIZATION INDEPENDENT ENDOSPERM (FIE; Makaverich et al. 2006). This Polycomb group complex has been detected in flowers and seeds (Köhler et al. 2003). In developing seeds that maternally inherit a mutated *mea* allele, the embryo and the endosperm overproliferate before they eventually abort (Köhler et al. 2004). The only known direct target gene of MEA is the type I MADS-box gene *PHERES1* (*PHE1*). MEA regulates allele-specific expression of *PHE1* by repressing the maternal PHE1 allele. MEA is expressed in the female gametophyte before fertilization and in the embryo and endosperm after fertilization. Therefore, the finding that only the maternal *PHE1* allele is repressed suggests that MEA modifies the maternal allele of *PHE1* before fertilization or shortly thereafter, at a time when the paternal *PHE1* allele is not accessible. Despite the fact that no firm evidence of the implication of NATs in these phenomena has been reported, a study has revealed the existence of antisense transcripts for *FIE* and *MS1a* in *Arabidopsis* (Wang et al. 2005), suggesting a potential link between NATs and genomic imprinting in plants.

9.5 npcRNAs Interacting with Specific RNA-Binding Proteins May Create Cellular Networks

Even though many nuclear RNA-Binding proteins (RBPs) have been identified as having critical roles during development and in epigenetic remodeling of chromatin, it is largely unclear how their action is controlled, primarily due to the difficulty in identifying their RNA partners (Lorkovic 2009). Most RNA-Binding proteins likely have multiple RNA partners such as mRNAs and npcRNAs (e.g., antisense RNAs, various “aberrant” RNAs or mRNA-like npcRNAs) that may compete in the different ribonucleoproteins (RNPs) and interfere with RNA networks where npcRNAs can act as competitors or activators and determine ribonucleoprotein

localization or action. Identification of the RNA-Binding proteins with which each npcRNA is associated is at the core of understanding ribonucleoprotein interaction networks in the cell.

The relocalization of ribonucleoprotein complexes has been linked to the action of npcRNAs. In the fission yeast, the *sme2*/*meiRNA* npcRNA was shown to bind the Mei2p protein, considered as a master regulator of meiosis (Watanabe and Yamamoto 1994). The *mei2* gene encodes an RNA-Binding proteins with three RNA-recognition motifs (RRMs), of which the C-terminal RRM3 is critical for its function. During mitosis, Mei2p remains inactive within the cytoplasm, but under meiosis-inducing conditions (mainly nutrient starvation), Mei2p shuttles from the cytoplasm to the nucleus (Sato et al. 2001; Yamashita et al. 1998). This shuttling has been linked to the binding of Mei2p to the *meiRNA* at the *sme2* locus and the formation of a Mei2p dot structure (Shimada et al. 2003). Formation of this dot may antagonize selective elimination of meiotic mRNAs by sequestering another RNA-Binding proteins, Mmi1p, in this nuclear dot structure (Harigaya et al. 2006). In plants, the *mei2*-like family has undergone a great expansion (Anderson et al. 2004), and the *AMLs* (*Arabidopsis mei2*-like) mainly seem to play a role in meiosis like *mei2* in fission yeast (Kaur et al. 2006). However *sme2mei*-like npcRNAs do not appear to exist in plants, and RNA partners of Mei2p-like RNA-Binding proteins still remain unknown. Indeed, npcRNA sequences can diverge rapidly between closely related species even when playing highly related functions (Mercer et al. 2009).

The npcRNA family *ENOD40* has been involved in the formation of symbiotic nitrogen-fixing nodules in legumes (Charon et al. 1999). Transgenic *Medicago truncatula* plants overexpressing or silenced for *ENOD40* exhibited accelerated nodulation or form only a few and modified nodule-like structures, respectively (Charon et al. 1999; Wan et al. 2007). The *ENOD40* npcRNA is highly structured (Crespi et al. 1994; Girard et al. 2003); however, one must notice that a small peptide has been proposed to be translated from this transcript (Rohrig et al. 2002). Using the yeast three-hybrid system, a constitutively expressed RNA-binding protein, MtRBP1, localized in nuclear speckles, has been identified to interact with the *ENOD40* RNA (Campalans et al. 2004). Immunolocalization experiments and transient assays have demonstrated that the *MtENOD40* npcRNA seems required for the relocalization of MtRBP1, from nuclear speckles to cytoplasmic granules, during nodule organogenesis (Campalans et al. 2004). As nuclear speckles store spliceosomal complexes and act in mRNA processing (Handwerger and Gall 2006), this relocalization event may be linked to changes in mRNA splicing or transport. Besides, nuclear speckles may also supply a stopover and regulatory checkpoint for components traveling with mRNAs through the nuclear pore to the cytoplasm (Handwerger and Gall 2006). Therefore, through interaction with specific RNA-Binding proteins, long npcRNAs may modulate the cellular ribonucleoprotein networks and determine new patterns of gene regulation, similarly as small npcRNAs do through the interaction with the RISC complex.

9.6 Concluding Remarks

The lifestyle of plants requires them to constantly adapt their growth and development to environmental variations. We think that npcRNAs can play a major role in these mechanisms of adaptation because they allow rapid changes in gene expression, acting at different levels (transcriptional, posttranscriptional, and translational). As plant development and growth continues throughout life, they have developed many different pathways, some plant-specific, for the action of npcRNA. Highly related plants may show large variations in their adaptation to environmental conditions despite the global conservation of their coding transcriptome. In contrast, the npcRNA transcriptome rapidly diverges during evolution and within species, suggesting that the action of npcRNA could be a major substrate to adapt gene expression in particular environments. By modifying the spatiotemporal gene expression patterns, npcRNAs may thus play a key role in developmental adaptation and plasticity. Because of the number of npcRNA in plants and their essential role in a wide range of processes, plants are organisms of choice for the study of these molecules and their mechanisms.

References

Adenot X, Elmayan T, Lauressergues D, Boutet S, Bouché N, Gasciolli V, Vaucheret H (2006) DRB4-Dependent TAS3 trans-Acting siRNAs control leaf morphology through AGO7. Curr Biol 16:927–932

Allen E, Xie Z, Gustafson AM, Carrington JC (2005) MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121:207–221

Anderson GH, Alvarez NDG, Gilman C, Jeffares DC, Trainor VCW, Hanson MR, Veit B (2004) Diversification of genes encoding *mei2*-like RNA binding proteins in plants. Plant Mol Biol 54:653–670

Aufsatz W, Mette F, Van der Winden J, Matzke AJM, Matzke M (2002) RNA-directed DNA methylation in Arabidopsis. Proc Natl Acad Sci USA 99:16499–16506

Baumberger N, Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci USA 102: 11928–11933

Béclin C, Boutet S, Waterhouse P, Vaucheret H (2002) A branched pathway for transgene-induced RNA silencing in plants. Curr Biol 12:684–688

Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, Vaucheret H, Thermes C, Crespi M (2009) Novel long non-protein-coding RNAs involved in *Arabidopsis* differentiation and stress responses. Genome Res 19:57–69

Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell 29: 1279–1291

Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O (2008) Widespread Translational Inhibition by Plant miRNAs and siRNAs. Science 320:1185–1190

Brown JW, Marschall DF, Echeverria M (2008) Intronic non-coding RNAs and splicing. Trends Plant Sci 13:335–342

Campalans A, Kondorosi A, Crespi M (2004) *Enod40*, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. Plant Cell 16:1047–1059

Carmel MA, Xuan Z, Zhang MQ, Hannon GJ (2002) The Argonaute family: tentacles that reach into RNAi, developmental control stem cell maintenance, and tumorigenesis. Genes Dev 16:2733–2742

Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon YH, Sung ZR, Goodrich J (2004) Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. Development 131:5263–5276

Charon C, Sousa C, Crespi M, Kondorosi A (1999) Alteration of *enod40* expression modifies *Medicago truncatula* root nodule development induced by *Sinorhizobium meliloti*. Plant Cell 11:1953–1966

Chekanova JA, Gregory BD, Reverdatto SV, Chen H, Kumar R, Hooker T, Yazaki J, Li P, Skiba N, Peng Q, Alonso J, Brukhin V, Grossniklaus U, Ecker JR, Belostotsky DA (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. Cell 131:1340–1353

Chen J, Sun M, Kent WJ, Huang X, Xie H, Wang W, Zhou G, Shi RZ, Rowley JD (2004) Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Res 32:4812–4820

Conti E, Izaurralde E (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Curr Opin Cell Biol 17:316–325

Crespi M, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A (1994) *enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. EMBO J 13:5099–5112

Crosthwaite SK (2004) Circadian clocks and natural antisense RNA. FEBS Lett 567:49–5

Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC (2000a) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101:543–553

Dalmay T, Hamilton A, Mueller E, Baulcombe DC (2000b) Potato virus X amplicons in Arabidopsis mediate genetic and epigenetic gene silencing. Plant Cell 12:369–379

Dalmay T, Horsefield R, Braunstein TH, Baulcombe DC (2001) SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in Arabidopsis. EMBO J 20:2069–2078

Dunoyer P, Himber C, Voinnet O (2005) DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. Nat Genet 37:1356–1360

Faghihi MA, Wahlestedt C (2009) Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol 10:637–643

Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, Alexander AL, Carrington JC (2006) Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in Arabidopsis. Curr Biol 16:939–944

Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One 14:e219

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet 39:1033–1037

Gasciolli V, Mallory AC, Bartel DP, Vaucheret H (2005) Partially redundant functions of Arabidopsis DICER-Like enzymes and a role for DCL4 in producing trans-acting siRNAs. Curr Biol 15:1494–1500

Girard G, Roussis A, Gultyaev AP, Pleij CW, Spaink HP (2003) Structural motifs in the RNA encoded by the early nodulation gene *enod40* of soybean. Nucleic Acids Res 31:5003–5015

Glazov E, Phillips K, Budziszewski GJ, Schöb H, Meins F Jr, Levin JZ (2003) A gene encoding an RNase D exonuclease-like protein is required for post-transcriptional silencing in Arabidopsis. Plant J 35:342–349

Hammond SM, Bernstein E, Beach D, Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature 404:293–296

Han MH, Goud S, Song L, Fedoroff N (2004) The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc Natl Acad Sci USA 101:1093–1098

Handwerger KE, Gall JG (2006) Subnuclear organelles: new insights into form and function. Trends Cell Biol 16:19–26

Harigaya Y, Tanaka H, Yamanaka S, Tanaka K, Watanabe Y, Tsutsumi C, Chikashige Y, Hiraoka Y, Yamashita A, Yamamoto M (2006) Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature 442:45–55

Hazen SP, Naef F, Quisel T, Gendron J, Chen H, Ecker J, Borevitz J, Kay S (2009) Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays. Genome Biol 10:R17

Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE (2006) Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet 38:721–5

Henz SR, Cumbie JS, Kasschau KD, Lohmann JU, Carrington JC, Weigel D, Schmid M (2007) Distinct expression patterns of natural antisense transcripts in *Arabidopsis*. Plant Physiol 144:1247–1255

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. Science 308:118–120

Hiraguri A, Itoh R, Kondo N, Nomura Y, Aizawa D, Murai Y, Koiwa H, Seki M, Shinozaki K, Fukuhara T (2005) Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana. Plant Mol Biol 57:173–188

Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K (2008) Stepwise chromatin remodeling by a cascade of transcription initiation of non-coding RNAs. Nature 456:130–134

Hirsch J, Lefort V, Vankersschaver M, Boualem A, Lucas A, Thermes C, d'Aubenton-Carafa Y, Crespi M (2006) Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts. Plant Physiol 140:1192–1204

Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M (2008) A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol Cell 32:685–695

Hunter C, Willmann MR, Wu G, Yoshikawa M, de la Luz Gutiérrez-Nava M, Poethig RS (2006) Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. Development 133:2973–2981

Jamalkandi SA, Masoudi-Nejad A (2009) Reconstruction of *Arabidopsis thaliana* fully integrated small RNA pathway. Funct Integr Genomics 9:419–432

Jay F, Renou JP, Voinnet O, Navarro L (2010) Biotic stress-associated microRNAs: identification, detection, regulation, and functional analysis. Methods Mol Biol 592:183–202

Jen CH, Michalopoulos I, Westhead DR, Meyer P (2005) Natural antisense transcripts with coding capacity in *Arabidopsis* may have a regulatory role that is not linked to double-stranded RNA degradation. Genome Biol 6:R51

Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, Carrington JC (2003) P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. Dev Cell 4:205–217

Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A Jr, Zhu JK, Staskawicz BJ, Jin H (2006) A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci USA 103:18002–18007

Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H (2007) A novel class of bacteria-induced small RNAs in Arabidopsis. Genes Dev 21:3123–3134

Kaur J, Sebastian J, Siddiqi I (2006) The Arabidopsis-mei2-like genes play a role in meiosis and vegetative growth in Arabidopsis. Plant Cell 18:545–59

Kidner CA, Martienssen RA (2005) The role of ARGONAUTE1 (AGO1) in meristem formation and identity. Dev Biol 280:504–517

Kim J, Krichevsky A, Grad Y, Hayes GD, Kosik KS, Church GM, Ruvkun G (2004) Identification of many microRNAs, that copurify with polyribosomes in mammalian neurons. Proc Natl Acad Sci USA 101:360–365

Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal 3:ra8

Kiyosawa H, Yamanaka I, Osato N, Kondo S, Hayashizaki Y, RIKEN GER Group, GSL Members (2003) Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. Genome Res 13:1324–1334

Köhler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Gruissem W (2003) *Arabidopsis* MSI1 is a component of the MEA/FIE *Polycomb* group complex and required for seed development. EMBO J 22:4804–4814

Köhler C, Page DR, Gagliardini V, Grossniklaus U (2004) The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nat Genet 37:28–30

Kramer C, Loros JJ, Dunlap JC, Crosthwaite SK (2003) Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*. Nature 421:948–952

Kurihara Y, Watanabe Y (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci USA 101:12753–12758

Kurihara Y, Matsui A, Hanada K, Kawashima M, Ishida J, Morosawa T, Tanaka M, Kaminuma E, Mochizuki Y, Matsushima A, Toyoda T, Shinozaki K, Seki M (2009) Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. Proc Natl Acad Sci USA 106:2453–2458

Lewis A, Mitsuya K, Umlauf D, Smith P, Dean W, Walter J, Higgins M, Feil R, Reik W (2004) Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat Genet 36:1291–1295

Li HW, Ding SW (2005) Antiviral silencing in animals. FEBS Lett 579:5965–5973

Lie J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in *Arabidopsis*. Curr Biol 15:1501–1507

Liu F, Quesada V, Crevillén P, Bäurle I, Swiezewski S, Dean C (2007) The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. Mol Cell 28:398–407

Liu F, Marquardt S, Lister C, Swiezewski S, Dean C (2010) Targeted 3′ processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. Science 327:94–97

Lorkovic ZJ (2009) Role of plant RNA-binding proteins in development, stress response and genome organization. Trends Plant Sci 14:229–236

Lu C, Jeong DH, Kulkarni K, Pillay M, Nobuta K, German R, Thatcher SR, Maher C, Zhang L, Ware D, Liu B, Cao X, Meyers BC, Green PJ (2008) Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc Natl Acad Sci USA 105:4951–6

MacLean D, Elina N, Havecker ER, Heimstaedt SB, Studholme DJ, Baulcombe DC (2010) Evidence for large complex networks of plant short silencing RNAs. PLoS One 5:e9901

Makaverich G, Leroy O, Akinci U, Schubert D, Clarenz O, Goodrich J, Grossniklaus U, Köhler C (2006) Different Polycomb group complexes regulate common target genes in Arabidopsis. EMBO Rep 7:947–952

Marin E, Jouannet V, Herz A, Lokerse AS, Weijers D, Vaucheret H, Nussaume L, Crespi MD, Maizel A (2010) miR390, Arabidopsis TAS3 tasiRNA, and their AUXIN RESPONSE FACTOR targets define an autoregulatory network quantitatively regulating lateral root growth. Plant Cell 22:1104–1117

Martens JA, Laprade L, Winston F (2004) Intergenic transcription is required to repress the *Saccharomyces cerevisiae SER3* gene. Nature 429:571–574

Mas P, Yanovsky MJ (2009) Time for circadian rhythms: plants get synchronized. Curr Opin Plant Biol 12:574–579

Mattick JS, Makunin IV (2006) Non-coding RNA. Hum Mol Genet 15:17–29

Matzke MA, Birchler JA (2005) RNAi-mediated pathways in the nucleus. Nat Rev Genet 6:24–35

Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nat Rev Genet 10:155–159

Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P, Huang Y, Kaminker JS, Millburn GH, Prochnik SE, Smith CD, Tupy JL, Whitfied EJ, Bayraktaroglu L, Berman BP, Bettencourt BR, Celniker SE, de Grey AD, Drysdale RA, Harris NL, Richter J, Russo S, Schroeder AJ, Shu SQ, Stapleton M, Yamada C, Ashburner M, Gelbart WM, Rubin GM, Lewis SE (2002) Annotation of the Drosophila melanogaster euchromatic genome: a systemic review. Genome Biol 3:e83

Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC (2008) Specificity of ARGONAUTE 7-miR390 interaction and dual functionality in *TAS3 trans*-acting siRNA formation. Cell 133:1–14

Moore T, Constancia M, Zubair M, Bailleul B, Feil R, Sasaki H, Reik W (1997) Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse Igf2. Proc Natl Acad Sci USA 94:12509–12514

Morel JB, Godon C, Mourrain P, Béclin C, Boutet S, Feuerbach F, Proux F, Vaucheret H (2002) Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus résistance. Plant Cell 14:629–639

Mourrain P, Béclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, Jouette D, Lacombe AM, Nikic S, Picault N, Rémoué K, Sanial M, Vo TA, Vaucheret H (2000) *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell 101:533–542

Osato N, Yamada H, Satoh K, Ooka H, Yamamoto M, Suzuki K, Kawai J, Carninci P, Ohtomo Y, Murakami K, Matsubara K, Kikuchi S, Hayashizaki Y (2003) Antisense transcripts with rice full-length cDNAs. Genome Biol 5:R5

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev 18:2368–2379

Peters NT, Rohrbach JA, Zalewski BA, Byrkett CM, Vaughn JC (2003) RNA editing and régulation of Drosophila 4f-rnp expression by sas-10 antisense read through mRNA transcripts. RNA 9:698–710

Prasanth K, Spector D (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev 21:11–42

Qu F, Ye X, Hou G, Sato S, Clemente TE, Morris TJ (2005) RDR6 has a broad-spectrum but temperature-dependent antiviral defense role in Nicotiana benthamiana. J Virol 79: 15209–15217

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev 20:3407–25

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev 16:1616–1626

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) Prediction of plant microRNA targets. Cell 110:513–520

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129:1311–1323

Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M (2002) Soybean ENOD40 encodes two peptides that bind to sucrose synthase. Proc Natl Acad Sci USA 99:1915–1920

Ruiz-Ferrer V, Voinnet O (2009) Roles of plant small RNAs in biotic stress responses. Annu Rev Plant Biol 60:485–510

Sato M, Shinozaki-Yabana S, Yamashit A, Watanabe Y, Yamamoto M (2001) The fission yeast meiotic regulator Mei2p undergoes nucleocytoplasmic shuttling. FEBS Lett 499:251–255

Schauer SE, Jacobsen SE, Meinke DW, Ray A (2002) DICER-LIKE1: blind men and elephants in *Arabidopsis* development. Trends Plant Sci 7:487–491

Schwach F, Vaistij FE, Jones L, Baulcombe DC (2005) An RNA-dependent RNA polymerase prevents meristem invasion by potato virus X and is required for the activity but not the production of a systemic silencing signal. Plant Physiol 138:1842–1852

Seila AC, Sharp PA (2008) Small RNAs tell big stories in Whistler. Nat Cell Biol 10:630–633

Shimada T, Yamashita A, Yamamoto M (2003) The fission yeast meiotic regulator Mei2p forms a dot structure in the horse-tail nucleus in association with the sme2 locus on chromosome II. Mol Biol Cell 14:2461–2469

Simpson GG, Dijkwel PP, Quesada V, Henderson I, Dean C (2003) FY is an RNA 3′ end-processing factor that interacts with FCA to control the Arabidopsis floral transition. Cell 113:777–787

Sun M, Hurst LD, Carmichael GG, Chen J (2005) Evidence for a preferential targeting of 3′-UTRs by cis-encoded natural antisense transcripts. Nucleic Acids Res 33:5533–5543

Sunkar R (2010) MicroRNAs with macro-effects on plant stress responses. Seminars in Cell & Developmental Biology 21:805–811

Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet 34:157–164

Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes Dev 20:759–771

Vaucheret H, Vazquez F, Crété P, Bartel DP (2004) The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev 18:1187–1197

Vazquez F, Gasciolli V, Crété P, Vaucheret H (2004) The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol 14:346–351

Verdel A, Vavasseur A, le Gorrec M and Touat-Todeschini L (2009) Common themes in siRNA-mediated epigenetic silencing pathways. Int. J. Dev. Biol. 53:245–257

Voinnet O (2009) Origin, biogenesis, and activity of plant microRNAs. Cell 136:669–687

Wan X, Hontelez J, Lillo A, Guarnerio C, Van de Peut D, Fedorova E, Bisseling T and Franssen H (2007) Medicago truncatula ENOD40-1 and ENOD40-2 are both involved in nodule initiation and bacteroid development. Journal of Experimental Botany 58:2033–2041

Wang XJ, Gaasterland T, Chua NH (2005a) Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. Genome Biol 6:R30

Wang H, Chua NH, Wang XJ (2005b) Prediction of trans-antisense transcripts in *Arabidopsis thaliana*. Genome Biol 7:R92

Wang H, Chua NH and Wang XJ (2006) Prediction of trans-antisense transcripts in *Arabidopsis thaliana*. Genome Biol. 7:R92

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature 454:126–130

Watanabe Y, Yamamoto M (1994) *S. pombe* mei2p encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. Cell 78:487–498

Wilusz JE, Sunwoo H, Spector DL (2009) Long non-coding RNAs: functional surprises from the RNA world. Genes Dev 23:1494–1504

Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC (2005) Expression of Arabidopsis MIRNA genes. Plant Physiol 138:2145–2154

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MMH, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Cha Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF,

Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema K, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302:842–846

Yamashita A, Watanabe Y, Nukina N, Yamamoto M (1998) RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. Cell 95:115–123

Yang Z, Ebright YW, Yu B, Chen X (2006) HEN1 recognizes 21-24nt small RNA duplexes and deposits a methyl group onto the 20 OH of the 30 terminal nucleotide. Nucleic Acids Res 34:667–675

Yasuda J, Hayashizaki Y (2008) The RNA continent. Adv Cancer Res 99:77–112

Ying SY, Chang CP, and Lin SL (2010) Intron-Mediated RNA Interference, Intronic MicroRNAs, and Applications. M. Sioud (ed.), RNA Therapeutics, Methods in Molecular Biology 629

Yoshikawa M, Peragine A, Park MY, and Poethig RS (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev. 19:2164–2175

Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X (2005) Methylation as a crucial step in plant microRNA biogenesis. Science 307:932–935

Zhan S, Lukens L (2010) Identification of novel miRNAs and miRNA dependent developmental shifts of gene expression in *Arabidopsis thaliana*. PloS One 5:e10157

Index

Đ. Ugarković (ed.), *Long Non-Coding RNAs*, Progress in Molecular and Subcellular Biology 51, DOI 10.1007/978-3-642-16502-3,