Review

The RNA infrastructure: dark matter of the eukaryotic cell?

Lesley J. Collins and David Penny

Allan Wilson Centre for Molecular Ecology and Evolution and Institute of Molecular BioSciences, Private Bag 11222, Massey University, 4442 Palmerston North, New Zealand

Eukaryotes express many functional non-protein-coding RNAs (ncRNAs) that participate in the processing and regulation of other RNA molecules. By focusing on connections between RNA-based processes, common patterns emerge that form a network-like RNA infrastructure. Owing to the intracellular movement of RNA during its processing (both between nuclear compartments and between the nucleus and cytoplasm), the RNA infrastructure contains both spatial and temporal connections. As research moves away from being protein-centric and focuses more on genomics, it is timely to explore these often 'hidden' aspects of the eukaryotic cell. The general and ancestral nature of most basic RNA-processing steps places a new focus on the generality of the spatial and temporal steps in RNA processing.

Eukaryotic RNA processing forms a network

Cell biologists are comfortable with the knowledge that many functional RNAs are required in getting from DNA sequences to proteins. This modern RNA world, as we know it, involves many non-protein-coding RNA (ncRNA)-based macromolecules, which process and regulate other RNA molecules via cleavage, nucleotide modification, transcription and destruction. Eukaryotic RNA biology represents not simply a linear progression of events, but rather forms a network of processes integrated over space and time, which we term the RNA infrastructure. The sheer volume of information concerning each RNA-processing step can obscure the overall picture of RNA biology, making it difficult to recognize important connections. As our understanding of the dynamic nature of the cellular environment increases, it can also be difficult to discern where one process ends and another begins.

Numerous subtypes of ncRNAs participate in the RNA infrastructure, including rRNA, mRNA, tRNA, small nuclear RNA (snRNA; see Glossary), snoRNA (small nucleolar RNA), several classes of regulatory RNAs involved in RNA interference (RNAi) and many long ncRNAs. These subtypes and their associated processes are covered by several reviews (including Refs [1,2]); therefore, instead of a detailed summary of each process, we aim to convey an overall basic model of how different ncRNAs are linked within the framework of the eukaryotic RNA infrastructure (Figure 1). We address three main points: (i) RNAs are processed and regulated by other RNAs (usually in conjunction with proteins); (ii) a network of interactions exists between classes of RNA; and (iii) it is now necessary to understand which RNA processes are ancestral to all modern eukaryotes and which classes evolved only later within eukaryotes. An understanding of these three points provides an integrated overview of RNA processing and regulation in eukaryotes.

Expanding the RNA-processing cascade

A good example of how RNA processes are connected comes from examining transcription-to-translation processes that form an RNA-processing cascade [3]. The processing of the three core RNAs (mRNA, tRNA and rRNA) includes

Glossary

Antisense RNA: sense RNA is expressed from the coding strand of DNA (the sense strand). RNA expressed from the complementary strand (the antisense strand) is collectively termed antisense RNA.

Epigenetics and chromatin modification: epigenetics studies heritable DNA traits other than sequence differences that are linked to phenotypic changes, especially in cellular and organismal development. Chromatin-mediated control of gene expression can be influenced by chemical modifications (e.g. methylation and acetylation). Epigenetic studies often analyse chromatin-modification patterns to investigate which genes are active in specific cell types.

Heterogeneous ribonucleoproteins particles (hnRNPs): complexes of newly synthesized mRNA bound to specific proteins (including protein K and polypyrimidine-tract-binding protein [PTB]). The presence of these proteins prevents export of the unprocessed mRNA to the cytoplasm before splicing.

Long ncRNAs: typically, these are RNAs longer than 200 nt with no protein-coding capacity. Many long ncRNAs are involved in gene regulation and chromatin regulation.

RNA editing: a process by which RNAs are chemically modified within the cell. This can result in either a change the protein coding of a sequence or to introduce a frameshift mutation or a premature stop codon. One example is adenosine to inosine (A to I) RNA editing. Another example is uracil insertion or deletion as found in kinetoplastid protist mitochondria.

Small nuclear RNAs (snRNAs): the Lsm class of small nuclear RNAs comprises the U6 and $U6_{atac}$ snRNA. They are transcribed by RNA Pol III and do not undergo a cytoplasmic assembly step. The Sm class of small nuclear RNAs are the U1, U2, U4, U5, U11, U12, $U4_{atac}$ snRNAs. They are transcribed by RNA Pol II, and biogenesis involves a cytoplasmic stage before re-import to the nucleus.

Small nucleolar RNAs (snoRNAs): these RNAs guide methylation and pseudouridylation of pre-rRNAs. Subclasses include H/ACA snoRNAs: (contain H box and ACA box motifs), C-box snoRNAs (contain C box and D box motifs) and ScaRNAs (modify snRNAs).

Small regulatory RNAs: microRNAs (miRNAs; typically 21–23 nt) are single-stranded RNAs involved in gene expression by base pairing to specific mRNAs. Piwi-interacting RNAs (piRNAs; 27–30 nt) interact with Piwi proteins for chromatin regulation and transposon silencing. Small interfering or silencing RNAs (siRNAs; 20–25 nt) are double-stranded RNAs that use the RNA interference pathway to regulate a specific gene containing a complementary sequence.

***Trans*-splicing**: in contrast to *cis*-splicing (where exons within the same mRNA transcript are spliced together), SL-*trans*-splicing (a more accurate name for this type of splicing in eukaryotes) trims the 5′ ends of mRNA transcripts and replaces them with the SL-leader RNA (100nt sn-type RNA). SL-*trans*-splicing is observed in nematodes, flatworms, cnidarians, ascidians, rotifers and euglenozoans (for which it is the predominant form of splicing).

Corresponding author: Collins, L.J. (l.j.collins@massey.ac.nz)

0168-9525/$ – see front matter doi:10.1016/j.tig.2008.12.003 Available online 24 January 2009

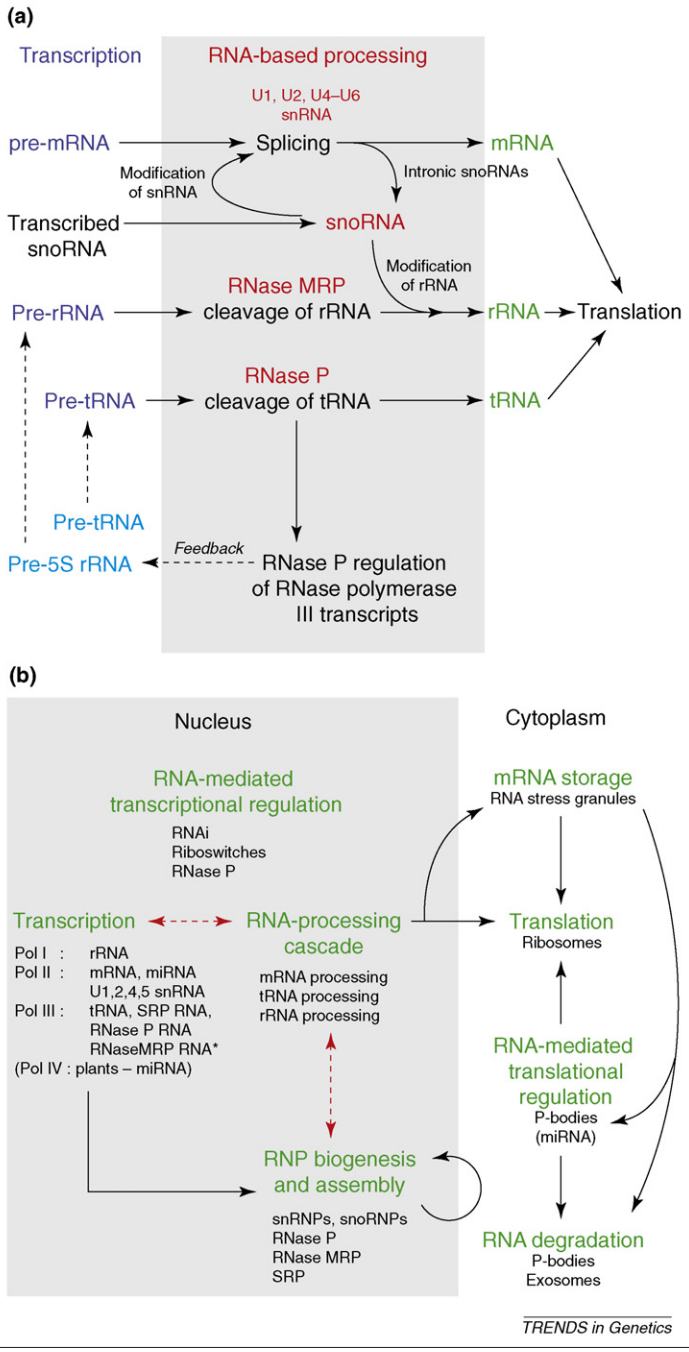

Figure 1. The eukaryotic RNA infrastructure. Although many RNA-processing reactions are linked between transcription and translation, current studies enable us to expand these linkages to include RNA-regulation and feedback processes in an RNA-infrastructure network. **(a)** A generalized RNA-processing cascade (adapted, with permission, from Ref. [3]) showing ncRNA-based processing from transcription to translation, concentrating on the RNA-based processing of mRNA, rRNA and tRNA transcripts. This depicts only the central section of the overall RNA-infrastructure network shown in **(b)**, where these RNA processes are shown in relation to the nucleus and cytoplasm. We expect that this generic model will differ in detail between eukaryotic lineages; however, the overall concept will largely apply to all eukaryotes. Colour code: blue, transcribed RNAs; red, processing ncRNAs; green, processed transcripts; blue text, cellular compartments; green text, major groupings of reactions. Red dashed arrows indicate that processes within each group interact in either direction. Solid arrows indicate directional processing. *MRP RNA is transcribed by Pol III in humans but by Pol II in *S. cerevisiae*.

RNA-based mechanisms of RNA excision and modification: rRNA by RNase MRP and snoRNAs; tRNA by RNaseP; and mRNAs by the snRNAs present within the spliceosome. By examining the connections between these processes in greater detail, a recurring trend of RNAs processing other RNAs emerges (Figure 1a).

mRNA processing

During splicing, mRNA interacts with the spliceosome, a massive complex comprising five snRNAs and ~200 proteins [4]. Each of the five snRNAs binds specific proteins to form small nuclear ribonucleoprotein (snRNP) complexes that interact both with pre-mRNA and with each other during each splicing cycle. Moreover, increasing evidence indicates that snRNAs carry out the actual cleavage reactions [5], rather than merely participating in splice-site recognition, thus increasing the similarity of mRNA processing with that of tRNA and rRNA. mRNA splicing affects the processing of other RNAs; for example, in vertebrates most snoRNAs

(involved in rRNA modification) occur in introns and are therefore released during splicing.

Tight links exist between different mRNA machineries. In humans, transcription by RNA polymerase (Pol) II and mRNA splicing occur in close proximity [6] and this coupling might protect the newly synthesized mRNA from degradation [7] before transcription termination [8,9]. mRNA 3′-end cleavage and polyadenylation is promoted by the U2 small nuclear RNA auxiliary factor U2AF65 (for a review, see Ref. [10]), a protein that has a prominent role in U2 snRNP-associated splicing and recognizes the polypyrimidine tract early in spliceosome assembly [10]. In addition, cleavage factor I (CF-I_m) physically links the splicing of the terminal intron and polyadenylation by interacting with U2 small nuclear riboprotein auxiliary factor (U2AF, comprising U2AF65 + U2AF35), poly(A) polymerase (PAP) and factor interacting with PAP (hFip1). Recent literature has revealed many additional linkages between transcription, 5′ capping, splicing, termination and polyadenylation.

Splicing also influences downstream mRNA processes including RNA localization, translational yield and mRNA decay [11]. The exon junction complex (EJC), a set of proteins deposited on the 5′ end of the exon during splicing [12], remains bound to the spliced mRNA as it is exported to the cytoplasm and interacts transiently with factors that connect the mRNA to downstream RNA processing [13,14]. This complex provides a major link between mRNA splicing and mRNA export and has a potential role in limiting RNA degradation. The EJC also contains several proteins involved in nonsense-mediated decay (NMD) of mRNA transcripts that are incorrectly spliced or terminated. Thus, the EJC identifies the previous location of introns in the processed transcript [11] and therefore enables detection of incorrect splicing events. Subsequent mRNA decay (by NMD) thus prevents translation of any such aberrant proteins.

Table 1. Dynamic eukaryotic cellular compartments and associated RNA-processing functions[a]

Compartment	N or C[b]	General information	Main RNA-processing functions	Refs
Nucleolus	N	rRNA repeats form nucleolar organizing sites that become nucleoli	rRNA transcription and processing, U6 snRNA methylation and pseudouridylation, tRNA processing by RNase P	[70]
Cajal bodies	N	Can switch from moving freely between chromatin to being tethered. Often associated with, or are even engulfed in nucleoli Coilin is used as a marker protein	Pol II transcription of Sm-class snRNAs, TMG-cap modification of pre-snoRNAs[c], site of action for scaRNAs (a class of snoRNA that modifies snRNAs), hypermethylation of telomerase RNA (teRNA), recycling of tri-snRNPs, accumulation of mature snRNPs	[71,72]
Gemini of Cajal bodies (Gems)	N	Often found paired or juxtaposed to Cajal bodies. SMN is used as a marker protein, but it can also be found in the nucleoplasm.	snRNP assembly	[27,73]
Speckles	N	Contain a high concentration of splicing components and polyA$^+$RNA Also called SC35 or spliceosomal factor compartments Paraspeckles are distinct from speckles but are often found adjacent to them in the interchromatin nucleoplasmic space	Splicing-factor storage	[74,75]
Spliceosomes	N	An RNA- and protein-containing complex that forms around an intron, cleaves it at specific sites and then ligates the exons to form mature mRNA	Recognition of splice sites, mRNA splicing, EJC deposition upon mRNA	[4,8,53]
Stress granules (SGs)	C	Rapidly induce mRNA translation following environmental stress Not found in *S. cerevisiae*, but similar structures are observed in *S. pombe*	mRNA storage in the cytoplasm	[44,47]
Processing bodies (P-bodies)	C	Spheroid particles that increase in size and number in response to stress. They contain the 5′ → 3′ mRNA decay machinery of the NMD pathway and the RNA-induced silencing complex	mRNA storage and decay	[44,45]
Nuclear exosome	N	Consists of a core of nine polypeptides and several other exosome-associated proteins. Found in prokaryotes and eukaryotes	3′ processing of precursor 5.8S rRNA and snoRNAs; U1, U4 and U5 3′ processing, degradation of aberrant pre-RNAs	[76]
Cytoplasmic exosome	C	Cytoplasmic and nuclear exosomes share ten common proteins but differ in the presence of others Involved in nonsense-mediated decay (degrades mRNAs with premature stop codons), non-stop decay (degrades mRNAs lacking a stop codon) and no-go decay (degrades 5′ fragments of mRNA on which translation has stalled)	3′ turnover of normal mRNAs, decay of aberrant RNAs before translation, degradation of mRNAs containing specific A +T-rich sequences	[76]

[a]Not all compartments are found in all eukaryotic species.
[b]Abbreviations: C, cytoplasmic; N, nuclear.
[c]During snoRNA biogenesis, the original m7G-cap is modified to a TMG cap in the Cajal body compared with Sm-class snRNA biogenesis, where this takes place in the cytoplasm.

rRNA processing

The 18S, 5.8S and 28S rRNA transcripts are generated by RNA Pol I in the nucleolus, where they also undergo early processing reactions. Extensive modification of the pre-rRNAs includes ribose 2′ hydroxyl group methylation (guided by C/D box snoRNAs) and uracil to pseudouridine conversion (guided by H/ACA snoRNAs) [15]. The strong network of interactions in the RNA infrastructure is quite apparent in vertebrates because the snoRNAs are mostly found within introns. However, yeast models (primarily *Saccharomyces cerevisiae*) traditionally have been used for unravelling the complex biogenesis of ribosomes, and recent studies link Pol-I-mediated rRNA transcription to downstream rRNA processing and maturation. Pol I, elongation factors and rRNA sequence elements seem to optimize transcription elongation and coordinate interactions (including those with snoRNAs) with the pre-rRNA for correct rRNA processing and ribosome assembly [16]. In addition, a recent study indicates that ribosomal protein production and pre-rRNA processing are linked by the CURI complex (named for the proteins contained within it, casein kinase 2 subunit [Ckb2]–U3 snRNA-associated protein 22, β[Utp22]–RRNA-processing protein 7 [Rrp7]–RRNA-processing protein 3 [Ifh1, also known as Rrp3]). This complex also associates with two transcription factors, repressor-activator protein 1 (Rap1) and forkhead-like protein 1 (Fhl1), in addition to having the already mentioned transcription factor, Ifh1, within the complex [17]. The crucial point here is that rRNA processing provides feedback from the later stages of processing to the transcription process that began it.

tRNA processing

tRNAs show similar patterns of transcription, excision by a ribozyme (RNase P), base modification and (sometimes) splicing. They are transcribed by Pol III and subsequently undergo several processing steps including the removal of the 5′-leader and 3′-trailing sequences, addition of CCA, base modifications and (in some species) intron removal (for a review, see Ref. [18]). Studies of tRNA maturation in *S. cerevisiae* indicate that, after nuclear processing events, tRNA is transported to the cytoplasm for splicing and completion of maturation [19]. Moreover, a close relationship exists between tRNA 5′ processing and tRNA splicing [20]. In addition, in yeast, tRNA intracellular movement is bidirectional; mature tRNAs can shuttle between the nucleus and the cytoplasm in both directions [19]. However, in vertebrates, tRNAs are spliced in the nucleus and not in the cytoplasm, thus highlighting variations in the spatial regulation of tRNA maturation between lineages.

Intra-cellular transport in the RNA infrastructure

Recent findings have shed new light not only on RNA processing in general but also on the spatial location of the different processing steps and how RNAs move through nuclear bodies (e.g. Cajal bodies, Gemini of Cajal bodies [Gems] and nucleoli) and, sometimes, into cytoplasmic bodies (e.g. P-bodies and RNA granules) (Table 1). This RNA movement is precise and coordinated with the different steps in RNA processing. Hence, the RNA infrastructure is networked through the movement of RNPs during their cycles of biogenesis, maturation, catalysis and degradation [21,22]. Indeed, RNA movement between sub-compartments is a general mechanism, ensuring the association of key components during maturation of RNA–protein complexes [23].

The biogenesis of snRNAs and snoRNAs (Figure 2) can be used as an example. Different stages of processing take place within different nuclear sub-compartments; however, Sm-class snRNAs undergo cytoplasmic-orientated stages before re-import back to the nucleus. By contrast, the human Lsm-class snRNAs (U6 and $U6_{atac}$ snRNAs) are transcribed by Pol III, possess a different cap and never leave the nucleus, although they are processed and assembled into snRNPs and multi-snRNP complexes within different nuclear sub-compartments [22]. A certain degree of U6 snRNA nuclear export and re-import might occur in *S. cerevisiae*, but whether this process occurs in all yeasts remains unclear [22]. Compartments such as Cajal bodies are not only repositories for the nuclear RNP biogenesis. Stanek *et al.* [24] recently showed that mature snRNPs travel through Cajal bodies, sometimes moving from one Cajal body to another. Because the pool of snRNPs in human cells consists largely of mature rather than newly synthesized snRNPs, they suggest that the Cajal body is used as a 'recycling center' to enable proper tri-snRNP reassembly [24].

Such sub-nuclear compartment-based recycling has not yet been characterized for snoRNPs (H/ACA and CD box). Although the assembly of C/D box snoRNPs seems to occur co-transcriptionally and is triggered by the general splicing factor intron-binding protein 60 (IBP60), the mechanisms of intra-nuclear and intra-cellular snoRNP trafficking are not completely characterized [25]. From what is known of their biogenesis, C/D box snoRNAs move through the nuclear compartments in a similar way to Sm-class snRNAs; however, they do not have a cytoplasmic stage.

Some RNA-associated proteins are involved in the transport of RNAs around the nucleus and cytoplasm, spanning the processes of transcription through modification and RNP assembly to degradation; such proteins provide important links between subsequent RNA-processing stages. One such protein is phosphorylated adaptor for RNA export (PHAX), which facilitates snRNA transport from the nucleus to the cytoplasm and intra-nuclear transport of snoRNAs. PHAX provides an important connection between these RNA-processing pathways while it undergoes cycles of phosphorylation and de-phosphorylation (Figure 2). Another important RNA-escorting macromolecule is the survival of motor neuron (SMN) protein complex, which is found in the nucleoplasm and in Gems [26]. The SMN complex scrutinizes cellular RNAs to ensure that Sm cores (highly reactive RNA-binding Sm proteins) are only assembled on proper snRNAs [27]; the SMN protein Gemin5 can distinguish snRNAs from other cellular RNAs to enable snRNP biogenesis [27]. The SMN complex also participates in other biogenesis processes, including those for hnRNPs and microRNAs (miRNAs) [26].

It is becoming increasingly clear that cellular movement during processing, especially across membranes, is a form of spatial regulation that prevents inappropriate binding interactions and enables control at many points. As a first

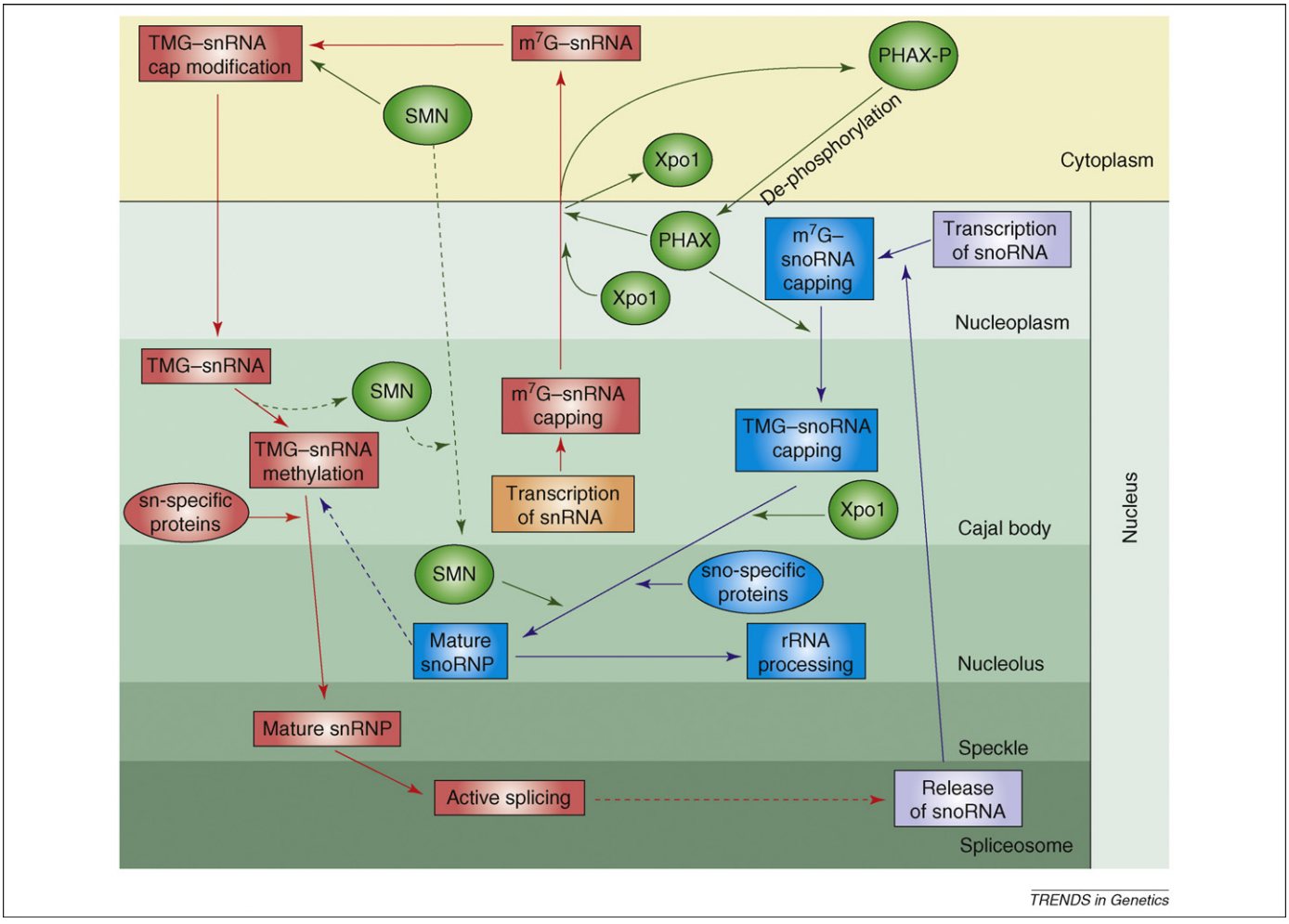

Figure 2. Network connectivity of the Sm-class snRNA and snoRNA biogenesis pathways. The complexity of the RNA infrastructure can be illustrated in this small example of two RNP biogenesis pathways (Sm-class snRNAs and snoRNAs). Sm-class snRNAs are transcribed in the Cajal body, capped with m7G, then exported to the cytoplasm with a dephosphorylated PHAX protein and Exportin1 (Xpo1). After binding of the SMN complex and modification of the m7G cap to a trimethylguanosine (TMG) cap, the TMG–snRNA–SMN complex is re-imported to the Cajal body and then the TMG–snRNA is modified by the scaRNA subclass of snoRNAs (at this time, the exact point of SMN complex dissociation from the snRNA is not known). After assembly with snRNA-specific proteins, the mature snRNP moves to speckles, then on to the spliceosome for active splicing. This splicing can then release snoRNAs or, alternatively, they can be transcribed in the nucleoplasm. The m7G-capped snoRNAs use PHAX to move to the Cajal body where their cap is modified to TMG (a difference in location from the snRNAs). TMG-capped snoRNAs are transported to the nucleolus using Xpo1 where they become mature snRNPs. Connections between the snRNA and snoRNA biogenesis pathways occur by the use of PHAX and Xpo1 for transport at different stages of maturation, the scaRNA (snoRNA-subclass) modification of some snRNAs and the way in which splicing via snRNAs can release snoRNAs. This illustration shows only a few of the many proteins involved in snoRNA and snRNA biogenesis. The actual situation will be even more complex. Colour code: red, the snRNA pathway; blue, snoRNA pathway (starting points of each pathway are noted by lighter colours); green, proteins or protein complexes shared between pathways. The green dashed line between the SMN complex in the cytoplasm and in the nucleolus indicates that this transport might not be direct but is simplified for this illustration. Red and blue dashed lines indicate processing links between the pathways.

step to exploring the generality of these processes in eukaryotes, it will be important to study how universal the PHAX and SMN complexes are in eukaryotes.

RNA regulation: connecting components of the RNA infrastructure

RNA storage, degradation and regulation (including via RNAi, riboswitches and RNA editing) are all linked to the processes discussed in the earlier sections, but the networking aspects of the RNA infrastructure go beyond transcription-to-translation processes. Regulation by antisense RNAs and long ncRNAs highlight that RNA regulation extends beyond single cells, to the developmental stages of multicellular organisms (Box 1), including epigenetics and chromatin modification. However, there is still much to learn about the regulation of RNA by RNA in the single-cell environment.

Regulation of RNA by RNA

Although best known for its role in regulating mRNA levels, RNAi (involving miRNAs, small interfering RNAs [siRNAs] and piwi-interacting RNAs [piRNAs]; reviewed in Refs [28,29]), is also directly involved in other cellular processes including chromatin-mediated gene silencing and DNA rearrangements. Indeed, the impact of RNAi upon total cellular regulation and its conservation throughout eukaryotes is only now coming to light. A recent report [30] showed that a single miRNA could directly or indirectly downregulate the production of thousands of genes. Although RNAi seems to be a general regulatory mechanism in eukaryotes, the timing and location of miRNA expression varies even among vertebrates [31] owing to changes in miRNA copy number, genomic context (either exclusively intergenic, or intronic and intergenic) or both.

Box 1. ncRNA and multicellularity

It was once thought that small regulatory RNAs evolved 'for' developmental regulation of multicellular eukaryotes. However, under a Darwinian mechanism of evolution, a new function cannot be selected 'to be useful' at some unknown time in the future. Rather, existing functions can be 'co-opted' or 'recruited' into new roles by processes such as the duplication of existing genes followed by the divergence of the duplicates, a Darwinian process that is as valid for RNAs as it is for proteins.

Long ncRNAs and miRNAs are prominent in multicellular gene regulation and a wide variety of mechanisms exist by which ncRNAs can regulate chromatin (for a review, see Ref. [65]). One of the best characterized mechanisms is X chromosome inactivation, in which X-linked genes of female mammals are downregulated to maintain expression at a comparable level with those in the male. The actual X chromosome to be silenced is randomly selected in each cell and this suppressed chromosome is selected through the actions of the Xist sense ncRNA and its antisense counterpart Tsix [65].

Long ncRNAs might also have crucial roles in cell differentiation [66]. HOX gene expression that specifies positional identity of cells (e.g. head–tail, dorsal–ventral) is under complex epigenetic regulation. Not only are many newly identified ncRNAs expressed differentially along developmental axes of the body (in a manner similar to HOX gene expression), but one of them, the long ncRNA HOX antisense intergenic RNA (HOTAIR) acts *in trans* (in contrast to Xist which operates *in cis*) to regulate chromatin [66]. Importantly, this result demonstrates that silencing of crucial developmental genes can also be controlled at a distance.

Recent studies have provided examples of cell-type-specific or even cancer-specific miRNA expression patterns [67]. For example, Dicer inactivation in mouse skin caused the development of cyst-like structures [68], indicating the importance of miRNA regulation in the development of these cells. A more recent study [69] of human corneal epithelial cells showed that one corneal specific miRNA, miR-184, can interfere with the ability of the more broadly expressed miR-205 to maintain SH2-containing phosphoinositide 5′-phosphatase 2 (SHIP2) expression, thereby providing an example of one miRNA regulating another. Hence, it is clear that the networking of RNA-mediated RNA regulation, the key concept of the RNA infrastructure, extends beyond the classic transcription-to-translation pathways.

In animals, double-stranded precursor 'pri-miRNAs' are processed to pre-miRNAs in the nucleus by the RNase III Drosha, before they are exported to the cytoplasm by exportin-5 where they are processed further by Dicer [32,33]. However, an alternative pathway was recently identified [34] in the fruit fly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*, in which some debranched introns (termed 'mirtrons') mimic the structural features of pre-miRNAs to enter the miRNA-processing pathway in the absence of Drosha-mediated cleavage. The very fact that alternative strategies for miRNA processing exist highlights the main point that RNA-mediated RNA regulation is a crucial part of the RNA infrastructure.

RNase P, which has a positive effect on Pol III promoter activity, provides another example of RNA-based transcriptional regulation [35,36]. RNase P associates with the chromatin of tRNA and 5S rRNA genes, which contain the type-1 Pol III promoter sequences, but not with the U6snRNA and 7SL-RNA that harbour type-3 Pol III promoter sequences. Transcription of these type-1 Pol-III-transcribed ncRNAs declines sharply in extracts depleted of active RNase P [35]. RNase P might also participate in the splicing-independent maturation of snoRNAs, as recently demonstrated in yeast [37], thus linking the production of tRNAs and rRNA.

Riboswitches (for a review, see Ref. [38]) perform yet another class of RNA-based feedback whereby binding of a small metabolite, such as a cofactor, alters the tertiary structure of RNA and changes the expression of the mRNA. Although well known in prokaryotes, their role in eukaryotic gene regulation is only now becoming clear. Thiamine pyrophosphate (TPP) riboswitches, which regulate (by negative feedback) to TPP synthesis (a derivative of vitamin B1), are present in the 5′ introns of three genes from the filamentous fungus *Neurospora crassa* [39]. This mechanism confirms results from another fungus *Aspergillus oryzae* [40] and from plants [38]. In these examples, the binding of TPP alters the availability of different splice sites. When TPP concentration is low, the mRNA is translated correctly but, at high TPP concentration, binding triggers the production of an alternatively spliced mRNA that is not translated. The authors suggest that there could be other undiscovered classes of riboswitches in eukaryotes, but also that other control mechanisms that affect RNA structure, such as temperature, will probably be uncovered [39]. Regardless of future discoveries, it is important to note that this mode of regulation occurs at the RNA level in the absence of additional protein factors [39].

Several longer ncRNAs directly target transcription (for a review, see Ref. [41]), including transcriptional co-activator for several steroid-hormone receptors (SRA), neuron-restrictive silencer element dsRNA (NRSE), heat shock RNA-1 (HSR1) and 7SK RNA. In particular, 7SK RNA is transcribed by Pol III and represses Pol II transcript elongation [41]. Another instance is in the human dihydrofolate reductase (DHFR) gene in which a regulatory transcript from a minor promoter interferes with the expression of the main transcript [42]. With the ongoing discovery [43] of ncRNAs in an increasingly wider range of eukaryotes, we expect that additional direct transcriptional regulators will be identified.

Translational regulation by cytoplasmic RNA granules

Some mRNAs are translated immediately after nuclear export but others can be stored in cytoplasmic RNA granules until developmental or environmental cues signal their translation [44]. These cytoplasmic RNA granules (for a review, see Ref. [44]) include processing bodies and stress granules, and compartments found in germ cells (polar and germinal granules) and neurons (neuronal granules).

Processing bodies (P-bodies or GW bodies, named for the presence of GW182) are cytoplasmic compartments (Table 1) that control post-transcriptional processes, including mRNA degradation, NMD, translational repression and RNA-mediated gene silencing (for a review, see Ref. [45]). mRNA degradation is initiated by the deadenylation (shortening) of the 3′ polyA-tail followed by decapping [46]. Current research indicates that the presence of a polyA-tail affects whether mRNAs are decapped, stored or returned to translation [46]. Stress granules are a class of cytoplasmic RNA granule that are usually only observed during a stress response (whereas P-bodies are present continuously [47]). During stress conditions, mRNAs encoding housekeeping proteins are redirected

from ribosomes to stress granules in synchrony with stress-induced translational arrest [47]. mRNAs within stress granules are not degraded so they are available for rapid re-initiation after recovery from stress [47].

We are only just beginning to unravel the mechanisms that regulate the movement of RNA between cytoplasmic bodies, especially between stress granules, P-bodies and polysomes [44] and, hence, the links between mRNA translation and mRNA decay. Indeed, the RNA infrastructure connects RNA processing (including ncRNA-mediated RNA processing) beyond nuclear compartments and into the cytoplasm.

Variation in the RNA infrastructure throughout eukaryotes

So far, we have discussed a basic model of the RNA infrastructure as a composite of features mostly from research on mammals, yeasts and plants. However, we expect to observe variations on this basic model for individual species or lineages. For example, the SMN protein, a crucial component of snRNP biogenesis in humans and fission yeast *Schizosaccharomyces pombe* [48], is not found in the budding yeast *S. cerevisiae* [49]. Similarly, RNase MRP, 7SL and 7SK RNAs are transcribed by Pol III in vertebrates but by Pol II in *S. cerevisiae* [50]. Differences also exist in how processes are linked: the co-transcriptional nature of splicing is less pronounced in *S. cerevisiae* than in other eukaryotes [51].

Which features of the RNA infrastructure are conserved throughout eukaryotes and how have they evolved? If we follow the general principle that basic properties of a mechanism found in genomes of all deep lineages of eukaryotes are ancestral, then we can infer that such features were present in the last common ancestor of eukaryotes [52–54]. Unfortunately, to accomplish this goal we have to move away from 'model' organisms and study RNA processing throughout protists. However, because these are non-model organisms, we have a long way to go.

For example, Trypanosomes, a distant group of excavate protists, predominantly use *trans*-splicing, a process in which a polycistronic transcript is cleaved into mRNAs and activated by the addition of a spliced leader SL-RNA [55]. Thus, in comparison to 'higher' eukaryotes, these organisms display some differences in their basic RNA infrastructure (i.e. pre-mRNA transcription is uncoupled from *trans*-splicing and is not linked to capping) [56]. SL-RNA transcription also occurs in a restricted region next to the nucleolus, separated from most of the pre-mRNA transcription [56]. However, splicing factors are found not only in trypanosome 'speckles', which seem similar to those in metazoa [57] but also in the cytoplasm, in which SL-RNP assembly occurs [55]. Similarly to metazoa, trypanosome Sm-proteins bind snRNAs and the SL-RNA. Likewise, U4 and U5 snRNAs shuttle from the nucleus to the cytoplasm. In another variation, the SMN gene has not been identified in any of the trypanosome genome projects [55] and the trypanosome SMN-binding homologous proteins do not contain SMN-binding sites [55]. By contrast, the trypanosome PHAX protein has been identified and it might have the same role in RNA transport as observed in metazoa [55]. It remains unclear whether *trans*- and *cis*-splicing require different types of spliceosomes. A preliminary finding of a single set of snRNPs in trypanosomes that might perform both *trans*- and *cis*-splicing, along with the presence of splicing factors that are shared by both splicing reactions, indicates that the same machinery might perform both types of splicing [58]. Although the splicing mechanism and the proteins responsible exhibit some differences, it seems that the general RNA infrastructure found in metazoa is also present in trypanosomes.

Trypanosomes do not represent all excavates and, indeed, the flagellate diplomonad *Giardia lamblia* is unusual in that it does not contain visible nucleoli, although homologs of nucleolar proteins are present [59]. The exact localization of RNA processing in this species is not yet clear, but the standard enzymes are present and, therefore, it is predicted that the fundamental nucleolar processes also occur (e.g. the rRNA genes display the same organization, and the basic splicing apparatus is present [59]). Although only a small section of *Giardia* RNA biology is characterized [53,60], ongoing work continues to reveal additional details about this parasitic protist [61].

The impact of RNA-based processing on the evolution of multicellularity is of particular interest (Box 1). miRNAs characterized from the single-celled alga *Chlamydomonas reinhardtii* are similar to those found in higher plants [62], indicating that RNAi, once thought to be primarily associated with multicellularity and developmental regulation, is likely to have been a feature of single-celled eukaryotes before the advent of multicellularity. Thus, variations on the basic RNA infrastructure merit additional attention; hopefully, such research will increase the focus on key aspects of the infrastructure that are ancestral to modern eukaryotes, and will help to distinguish the variations present along different lineages.

Concluding remarks

RNA-based reactions are highly concentrated in RNA processing and regulation; they definitely are not randomly spread throughout the standard categories of metabolism. Although often investigated individually, RNA processing is strongly inter-connected (Figure 1), especially in the way components move throughout the cell during RNA processing and maturation (Figure 2). This review focuses on intra-cellular processes, but ncRNA-mediated signalling pathways in important extracellular and developmental processes are increasingly being recognized [1,63], and the networking aspect of the RNA infrastructure easily encompasses these emerging functions.

Cellular localization studies of both RNA and proteins are crucial for understanding the integration of the RNA infrastructure. These interactions form an 'underworld' network of the cell in which RNA can be viewed as the 'dark matter' [64] that links many basic cellular functions. Detailed analyses of individual RNA-processing machineries (including how RNA-based regulation can affect transcription) are crucial, but a mechanistic view of how RNA moves within the cell during each process is also needed. This information will enable researchers to move to a more integrated view of eukaryote cellular biology.

From an evolutionary point of view, it is important to explore the consequences of component sharing between different RNA-processing machineries. Some characteristics might relate to the adaptive cellular response to environmental conditions (e.g. the storage of heat-shock mRNAs in cytoplasmic RNA stress granules). Other linkages could be more ancient, reflecting the ancestral condition of the eukaryotic cell. Many researchers understandably focus on the mammalian, plant or fungal RNA infrastructure. However, in the long run, an understanding of the ancestral eukaryotic RNA-processing machineries and the specific modifications, losses and innovations in different lineages will address many unanswered questions. Such an understanding will be possible soon; new research continues to uncover the basic underlying biology of the ancestral eukaryotic cell [52] and ancestral eukaryotic RNA-processing machineries [53]. Detailed investigations of individual species and general integrative studies are needed to achieve this goal. Genomic sequencing of additional protists and other free-living unicellular eukaryotes should also be performed as a foundation to such studies.

It is not yet possible to determine if the RNA infrastructure is as dense a network as other biochemical pathways or transcriptional networks; current efforts still only touch the tip of the proverbial iceberg in uncovering the full extent of RNA biology present in eukaryotes [43]. The differences already observed in RNA infrastructure among eukaryotes indicate that it is dynamic and evolving and certainly not a fixed product of an ancient RNA world. Before an understanding of the origin of the eukaryotic nucleus can be achieved, it will be important to investigate RNA infrastructure in greater detail to catch sight of the additional surprises that await us from the 'dark matter' of the eukaryotic cell.

Acknowledgements

Our work is funded by the Allan Wilson Center for Molecular Ecology and Evolution and the Institute of Molecular BioSciences at Massey University, New Zealand.

References

1 Amaral, P.P. *et al.* (2008) The eukaryotic genome as an RNA machine. *Science* 319, 1787–1789
2 Costa, F.F. (2007) Non-coding RNAs: lost in translation? *Gene* 386, 1–10
3 Woodhams, M.D. *et al.* (2007) RNase MRP and the RNA processing cascade in the eukaryotic ancestor. *BMC Evol. Biol.* 7 (Suppl 1), S13
4 Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12, 5–14
5 Valadkhan, S. (2005) snRNAs as the catalysts of pre-mRNA splicing. *Curr. Opin. Chem. Biol.* 9, 603–608
6 Kornblihtt, A.R. *et al.* (2004) Multiple links between transcription and splicing. *RNA* 10, 1489–1498
7 Hicks, M.J. *et al.* (2006) Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. *PLoS Biol.* 4, e147
8 Gornemann, J. *et al.* (2005) Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol. Cell* 19, 53–63
9 Lacadie, S.A. and Rosbash, M. (2005) Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5′ss base pairing in yeast. *Mol. Cell* 19, 65–75
10 Millevoi, S. *et al.* (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3′ end processing machineries. *EMBO J.* 25, 4854–4864
11 Chang, Y.F. *et al.* (2007) The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* 76, 51–74
12 Nojima, T. *et al.* (2007) The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export. *J. Biol. Chem.* 282, 15645–15651
13 Shibuya, T. *et al.* (2006) Mutational analysis of human eIF4AIII identifies regions necessary for exon junction complex formation and nonsense-mediated mRNA decay. *RNA* 12, 360–374
14 Tange, T.O. *et al.* (2004) The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.* 16, 279–284
15 Moss, T. *et al.* (2007) A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell. Mol. Life Sci.* 64, 29–49
16 Schneider, D.A. *et al.* (2007) Transcription elongation by RNA polymerase I is linked to efficient rRNA processing and ribosome assembly. *Mol. Cell* 26, 217–229
17 Rudra, D. *et al.* (2007) Potential interface between ribosomal protein production and pre-rRNA processing. *Mol. Cell. Biol.* 27, 4815–4824
18 Hopper, A.K. and Phizicky, E.M. (2003) tRNA transfers to the limelight. *Genes Dev.* 17, 162–180
19 Yoshihisa, T. *et al.* (2007) Cytoplasmic splicing of tRNA in *Saccharomyces cerevisiae*. *Genes Cells* 12, 285–297
20 Hiley, S.L. *et al.* (2005) Global analysis of yeast RNA processing identifies new targets of RNase III and uncovers a link between tRNA 5′ end processing and tRNA splicing. *Nucleic Acids Res.* 33, 3048–3056
21 Herbert, A. and Rich, A. (1999) RNA processing and the evolution of eukaryotes. *Nat. Genet.* 21, 265–269
22 Hopper, A.K. (2006) Cellular dynamics of small RNAs. *Crit. Rev. Biochem. Mol. Biol.* 41, 3–19
23 Handwerger, K.E. and Gall, J.G. (2006) Subnuclear organelles: new insights into form and function. *Trends Cell Biol.* 16, 19–26
24 Stanek, D. *et al.* (2008) Spliceosomal small nuclear ribonucleoprotein particles repeatedly cycle through Cajal bodies. *Mol. Biol. Cell* 19, 2534–2543
25 Matera, A.G. *et al.* (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Genet.* 8, 209–220
26 Gubitz, A.K. *et al.* (2004) The SMN complex. *Exp. Cell Res.* 296, 51–56
27 Battle, D.J. *et al.* (2006) The Gemin5 protein of the SMN complex identifies snRNAs. *Mol. Cell* 23, 273–279
28 Munroe, S.H. and Zhu, J. (2006) Overlapping transcripts, double-stranded RNA and antisense regulation: a genomic perspective. *Cell. Mol. Life Sci.* 63, 2102–2118
29 Hartig, J.V. *et al.* (2007) piRNAs–the ancient hunters of genome invaders. *Genes Dev.* 21, 1707–1713
30 Selbach, M. *et al.* (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63
31 Ason, B. *et al.* (2006) Differences in vertebrate microRNA expression. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14385–14389
32 Zhang, B. *et al.* (2007) MicroRNAs and their regulatory roles in animals and plants. *J. Cell. Physiol.* 210, 279–289
33 Grimson, A. *et al.* (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455, 1193–1197
34 Ruby, J.G. *et al.* (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature* 448, 83–86
35 Reiner, R. *et al.* (2006) A role for the catalytic ribonucleoprotein RNase P in RNA polymerase III transcription. *Genes Dev.* 20, 1621–1635
36 Jarrous, N. and Reiner, R. (2007) Human RNase P: a tRNA-processing enzyme and transcription factor. *Nucleic Acids Res.* 35, 3519–3524
37 Coughlin, D.J. *et al.* (2008) Genome-wide search for yeast RNase P substrates reveals role in maturation of intron-encoded box C/D small nucleolar RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 105, 12218–12223
38 Montange, R.K. and Batey, R.T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.* 37, 117–133
39 Cheah, M.T. *et al.* (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447, 391–393
40 Kubodera, T. *et al.* (2003) Thiamine-regulated gene expression of *Aspergillus oryzae* thiA requires splicing of the intron containing a riboswitch-like domain in the 5′-UTR. *FEBS Lett.* 555, 516–520
41 Goodrich, J.A. and Kugel, J.F. (2006) Non-coding-RNA regulators of RNA polymerase II transcription. *Nat. Rev. Mol. Cell Biol.* 7, 612–616
42 Martianov, I. *et al.* (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445, 666–670

43 Washietl, S. *et al.* (2007) Structured RNAs in the ENCODE selected regions of the human genome. *Genome Res.* 17, 852–864
44 Anderson, P. and Kedersha, N. (2006) RNA granules. *J. Cell Biol.* 172, 803–808
45 Jakymiw, A. *et al.* (2007) The role of GW/P-bodies in RNA processing and silencing. *J. Cell Sci.* 120, 1317–1323
46 Brengues, M. and Parker, R. (2007) Accumulation of polyadenylated mRNA, Pab1p, eIF4E, and eIF4G with P-Bodies in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 18, 2592–2602
47 Kedersha, N. *et al.* (2005) Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell Biol.* 169, 871–884
48 Paushkin, S. *et al.* (2000) The survival motor neuron protein of *Schizosacharomyces pombe*. Conservation of survival motor neuron interaction domains in divergent organisms. *J. Biol. Chem.* 275, 23841–23846
49 Yong, J. *et al.* (2004) Why do cells need an assembly machine for RNA-protein complexes? *Trends Cell Biol.* 14, 226–232
50 Haeusler, R.A. and Engelke, D.R. (2006) Spatial organization of transcription by RNA polymerase III. *Nucleic Acids Res.* 34, 4826–4836
51 Tardiff, D.F. *et al.* (2006) A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol. Cell* 24, 917–929
52 Kurland, C.G. *et al.* (2006) Genomics and the irreducible nature of eukaryotic cells. *Science* 312, 1011–1014
53 Collins, L. and Penny, D. (2005) Complex spliceosomal organization ancestral to extant eukaryotes. *Mol. Biol. Evol.* 22, 1053–1066
54 Collins, L. and Penny, D. (2006) Investigating the intron recognition mechanism in eukaryotes. *Mol. Biol. Evol.* 23, 901–910
55 Mandelboim, M. *et al.* (2003) Silencing of Sm proteins in *Trypanosoma brucei* by RNA interference captured a novel cytoplasmic intermediate in spliced leader RNA biogenesis. *J. Biol. Chem.* 278, 51469–51478
56 Dossin Fde, M. and Schenkman, S. (2005) Actively transcribing RNA polymerase II concentrates on spliced leader genes in the nucleus of *Trypanosoma cruzi*. *Eukaryot. Cell* 4, 960–970
57 Tkacz, I.D. *et al.* (2007) Identification of novel snRNA-specific Sm proteins that bind selectively to U2 and U4 snRNAs in Trypanosoma brucei. *RNA* 13, 30–43
58 Liang, X.H. *et al.* (2006) Analysis of spliceosomal complexes in *Trypanosoma brucei* and silencing of two splicing factors Prp31 and Prp43. *Mol. Biochem. Parasitol.* 145, 29–39
59 Ghosh, S. *et al.* (2001) Expression and purification of recombinant *Giardia fibrillarin* and its interaction with small nuclear RNAs. *Protein Expr. Purif.* 21, 40–48
60 Chen, X.S. *et al.* (2007) Combined experimental and computational approach to identify non-protein-coding RNAs in the deep-branching eukaryote *Giardia intestinalis*. *Nucleic Acids Res.* 35, 4619–4628
61 Chen, X.S. *et al.* (2008) Computational identification of four spliceosomal snRNAs from the deep-branching eukaryote *Giardia intestinalis*. *PLoS One* 3, e3106
62 Molnar, A. *et al.* (2007) miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature* 447, 1126–1129
63 Martello, G. *et al.* (2007) MicroRNA control of Nodal signalling. *Nature* 449, 183–188
64 Freyhult, E.K. *et al.* (2007) Exploring genomic dark matter: a critical assessment of the performance of homology search methods on noncoding RNA. *Genome Res.* 17, 117–125
65 Whitehead, J. *et al.* (2008) Regulation of the mammalian epigenome by long noncoding RNAs. *Biochim. Biophys. Acta*, DOI: 10.1016/j.bbagen.2008.10.007
66 Rinn, J.L. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323
67 Prueitt, R.L. *et al.* (2008) Expression of microRNAs and protein-coding genes associated with perineural invasion in prostate cancer. *Prostate* 68, 1152–1164
68 Yi, R. *et al.* (2006) Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat. Genet.* 38, 356–362
69 Yu, J. *et al.* (2008) MicroRNA-184 antagonizes microRNA-205 to maintain SHIP2 levels in epithelia. *Proc. Natl. Acad. Sci. U. S. A.* 105, 19300–19305
70 Hernandez-Verdun, D. (2006) The nucleolus: a model for the organization of nuclear functions. *Histochem. Cell Biol.* 126, 135–148
71 Stanek, D. and Neugebauer, K.M. (2006) The Cajal body: a meeting place for spliceosomal snRNPs in the nuclear maze. *Chromosoma* 115, 343–354
72 Matera, A.G. and Shpargel, K.B. (2006) Pumping RNA: nuclear bodybuilding along the RNP pipeline. *Curr. Opin. Cell Biol.* 18, 317–324
73 Li, L. *et al.* (2006) Dynamic nature of cleavage bodies and their spatial relationship to DDX1 bodies, Cajal bodies, and gems. *Mol. Biol. Cell* 17, 1126–1140
74 Hall, L.L. *et al.* (2006) Molecular anatomy of a speckle. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 288, 664–675
75 Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* 4, 605–612
76 Houseley, J. *et al.* (2006) RNA-quality control by the exosome. *Nat. Rev. Genet.* 7, 529–539