



HAL
open science

Multiple information carried by RNAs: total eclipse or a light at the end of the tunnel?

Baptiste Bogard, Claire Francastel, Florent Hubé

► To cite this version:

Baptiste Bogard, Claire Francastel, Florent Hubé. Multiple information carried by RNAs: total eclipse or a light at the end of the tunnel?. *RNA Biology*, 2020, pp.1-14. 10.1080/15476286.2020.1783868 . hal-02902851

HAL Id: hal-02902851

<https://u-paris.hal.science/hal-02902851>

Submitted on 8 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multiple information carried by RNAs: total eclipse or a light at the end of the tunnel?

Baptiste Bogard¹, Claire Francastel^{1,*} and Florent Hubé^{1,*}

¹ Université de Paris, Epigenetics and Cell Fate, CNRS, F-75013 Paris, France

* Author to whom correspondence should be addressed:

Dr. Florent Hubé

UMR7216 - Epigénétique et Destin Cellulaire

Université Paris 7 Diderot

Bâtiment Lamarck - 4ème étage

Case Courrier 7042

35, rue Hélène Brion

75013 PARIS

Tél: +33 1 57 27 89 32

Fax: +33 1 57 27 89 10

E-mail: florent.hube@univ-paris-diderot.fr

ORCID 0000-0001-7560-2530

E-mail: claire.francastel@univ-paris-diderot.fr

ORCID 0000-0002-6353-4320

Abstract:

The findings that an RNA is not necessarily either coding or non-coding, or that a precursor RNA can produce different types of mature RNAs, whether coding or non-coding, long or short, have challenged the dichotomous view of the RNA world almost 15 years ago. Since then, and despite an increasing number of studies, the diversity of information that can be conveyed by RNAs is rarely searched for, and when it is known, it remains largely overlooked in further functional studies. Here, we provide an update with prominent examples of multiple functions that are carried by the same RNA or are produced by the same precursor RNA, to emphasize their biological relevance in most living organisms. An important consequence is that the overall function of their locus of origin results from the balance between various RNA species with distinct functions and fates. The consideration of the molecular basis of this multiplicity of information is obviously crucial for downstream functional studies when the targeted functional molecule is often not the one that is believed.

Keywords: Bifunctional RNA; Non-coding RNA; Protein-coding; Dual function RNA; Splicing; circRNA; small ORF

36 Introduction

37 Almost three decades ago, unprecedented research efforts to sequence the human genome and
38 identify the genes that it contains led to the striking evidence that the vast majority of the genome
39 does not contain information to make proteins. Even more surprising, most of these so-called non-
40 coding sequences were shown to be competent for transcription and non-coding transcripts to
41 constitute the bulk of the human transcriptome [1,2].

42 By definition, a non-coding RNA (ncRNA) is an RNA molecule that is not translated into a
43 protein product, and is thus distinguished from messenger RNAs (mRNA). Abundant and
44 functionally important classes of ncRNAs include structural RNAs like transfer RNAs (tRNAs) and
45 ribosomal RNAs (rRNAs), small regulatory RNAs such as microRNAs (miRNAs), small interfering
46 RNAs (siRNAs), PIWI-interacting RNAs (piRNAs), small nuclear RNAs (snRNAs), small
47 nucleolar RNAs (snoRNAs), small cajal body RNAs (scaRNAs) [3]. We should also mention an
48 extensive list of so-called long ncRNAs (lncRNA) whose functions remain largely unclear for most
49 of them [4]. Nonetheless, ncRNAs have been assigned functions in most aspects of cell biology
50 including transcription, chromatin remodelling, splicing, nuclear import and chromosome
51 architecture, and to function as scaffolding, guiding, signalling or decoy for small regulatory RNAs
52 or proteins [3,5]. Not surprisingly, they are often deregulated in human diseases, notably in a range
53 of cancers or inherited disorders [6,7], although they have been more rarely causally linked to the
54 emergence of disease phenotypes.

55 Since the release of the first draft, the annotation of the human genome has been quite
56 dynamic and new data continues to enrich the transcriptome every day. Even recently, the number
57 of protein-coding genes has been revised downwards [8]. It is important to remember that the
58 transcriptional potential of eukaryotic genomes is very pervasive and widely intertwined [1,2]. For
59 example, independent transcription units can be hosted in introns of larger ones or overlap in the
60 antisense orientation. In addition, the use of alternative promoters or termination sites is well-
61 known to regulate transcriptional output in a tissue- or stage-specific manner through the production
62 of various isoforms. In that respect, the multifunctionality of a given genomic sequence, whereby a
63 single locus can release more than one type of transcripts, *i.e.* coding, non-coding, long or short,
64 sense or antisense, has been well documented [9-11] and is not the topic of this review. Importantly,
65 this diversity is mostly regulated at the level of transcription depending on cellular context or
66 environmental cues. An important consequence is that the transcripts thus produced may not co-
67 exist in the same cellular context.

68 However, difficulties in genomic annotation are also complicated by mounting evidence that
69 multifunctionality can also apply to transcripts themselves. Indeed, certain precursor RNAs can
70 release more than one class of transcripts, *i.e.* coding and non-coding or two non-coding RNAs, and

71 more strikingly, that some mature transcripts can perform more than one function [reviewed in [12]
72 and **Figure 1**]. The first example of an RNA with dual function was probably the Steroid Receptor
73 RNA activator (SRA) [13,14], and the term ‘bifunctional’ was then coined by Marcel Dinger and
74 John Mattick [15]. Ever since, this duality of information conveyed by some of the transcripts has
75 actually been reported in almost all organisms, from lower eukaryotes through plants to mammals
76 [see **Table 1** for prominent examples]. Hence, the view that an RNA should be either coding or
77 non-coding is rather binary and the duality in RNA functions is far from being anecdotal.
78 Unfortunately, and as often, this awareness is somewhat mitigated by the multiplication of various
79 denominations such as “dual-function RNAs” [16,17], "coding and non-coding RNAs" [cncRNAs,
80 [16,18,19]] or “long non-coding chimeric RNA” [lncRNA, [20]].

81 We propose here to go over remarkable contexts where any attempt at categorization into
82 coding or non-coding is likely to be reductive. We will go over bifunctional precursor RNAs (pre-
83 RNA) that can release coding and non-coding, or two non-coding, functions depending on post-
84 transcriptional maturation processes. We will also mention single RNA molecules that can operate
85 at least two functions [for reviews see [12,15,21,22]]. This includes messenger RNAs (mRNA)
86 shown to operate as functional RNAs, and certain lncRNAs initially classified as non-coding but
87 shown to release small peptides or re-annotated as coding RNAs. Following the logic, any
88 functional ncRNA that serves as a precursor to a smaller regulatory RNA should be considered as
89 bifunctional (**Figure 1**). Examples will include certain snoRNAs that function as precursors of
90 miRNAs. Likewise, snoRNAs being exclusively produced from intron splicing, at least in humans,
91 their host precursor is a bifunctional RNA since its splicing releases a mature mRNA and a snoRNA
92 with distinct mechanisms of action and locations.

93 Understanding the combination of functions and fates that can be carried by a transcript, being
94 it a mature RNA or a pre-RNA, is not only futile nor is it intended to add confusion by naming new
95 categories of RNAs. On a practical note, the annotation of transcripts with coding, non-coding or
96 mixed potential is important for genome or transcriptome manipulations for which the impact may
97 be wider than expected or simply not the desired one. This knowledge is also crucial for genomic
98 studies that aim to understand the determinants of human traits or predisposition to disease.

99

100 **1/ When a precursor-RNA produces two or more transcripts with distinct fates**

101 We already mentioned cases where the dual coding/non-coding function is carried by the
102 products of a gene and its related pseudogene (reviewed in [12,23]), yet it potentially involves
103 thousands of pairs of gene/pseudogene transcripts that remain to be characterized. In a more
104 remarkable way, this duality of functions can also be released by the same transcription unit, *i.e.*
105 carried by the same pre-RNA, after steps of post-transcriptional maturation.

106

107 ***Intron retention and the production of lncRNAs***

108 We have already discussed alternative splicing (AS) as a means used by organisms to enhance
109 their proteome, through the production of diverse mRNAs which fate is to be translated into protein
110 variants, but also their transcriptome, by producing both coding and non-coding RNAs in fewer
111 cases [12,24].

112 The first example was the SRA RNA first identified as a ncRNA with trans-activating
113 functions of the activity of hormone receptor complexes [25]. Further characterization identified
114 dozens of transcripts that classically differ in their initiation and termination sites or by their exon
115 content through exon skipping [13,14,26]. However, the striking finding was that these transcripts
116 also differed in their coding capacity, with the most remarkable disruption of the ORF being
117 through the retention of the first intron. In sum, fully spliced isoforms encode a SRAP protein
118 whereas intron-retaining isoforms form the SRA ncRNA [13,14,27]. Although it is yet unknown
119 how AS of SRA intron 1 is regulated and how the intron-retaining isoforms escape surveillance
120 machineries like the Nonsense-Mediated Decay (NMD), the overall function of the *SRA1* locus
121 results from the balance between coding and non-coding isoforms, at least in a muscle context
122 [12,27]. Whereas SRAP was shown to have an antagonistic impact on the function of its cognate
123 SRA RNA, a switch towards non-coding isoforms accompanies muscle differentiation and
124 accelerates reprogramming of non-muscle cells towards the muscle fate [27]. Evidence for the
125 functional importance of this switch is the finding that it does not occur in cells from patients with
126 splicing defects causing the Myotonic Dystrophy type 1 (DM1) [27]. Importantly, because coding
127 and non-coding isoforms co-exist in the cell, addressing their individual functions required to
128 mutate the ORF without affecting the secondary structure of the ncRNA or to destroy the secondary
129 structures without affecting the ORF [27]. Even truer than for classical transcription units, the
130 extensive knowledge of the variety of mature transcripts than can be produced through post-
131 transcriptional maturation of a pre-RNA is key to prevent false assumptions from non-targeted
132 functional approaches.

133 This was also probably the first example providing evidence that AS of intron is not
134 necessarily a faulty mechanism, although it is less widely used than in plants [28]. Indeed, intron
135 retention (IR) is usually thought to be an error in the splicing mechanism that leads to RNA
136 degradation by triggering the NMD pathway [29]. How some intron-retaining RNAs do escape
137 NMD is currently unclear, but the stable ncRNAs thus produced may operate functions as RNA
138 molecules, or at least have an impact on cellular processes. It was suggested that IR could also
139 impact translation efficiency if the retained intron contains miRNA binding sites [30,31], or if it
140 introduces a uORF (upstream of the ORF) or a rigid structure upstream of the start codon [reviewed

141 in [32]]. Since then, hundreds of mRNA with IR have been bioinformatically predicted [24,33] or
142 identified as stable entities in various cell types [14,34] and reviewed in [12,21,35]. In human cells,
143 IR was proposed to fine-tune gene transcriptomes [36] and, more importantly, to orchestrate the
144 establishment of lineage-affiliated expression programs that accompany cell differentiation [35,37].

145 Another example where AS serves to coordinate gene expression programs in response to
146 environmental cues is during the stress response. Upon UV irradiation, the rate of transcriptional
147 elongation is reduced and associated with a shift in expression from long mRNAs to shorter
148 isoforms through the incorporation of an alternative last exon (ALE) [38]. This AS mechanism
149 differs from IR but, as shown for the Activating Signal Cointegrator 1 Complex Subunit 3
150 (ASCC3), it also promotes an expression switch from a long coding mRNA to a shorter isoform that
151 functions as a nuclear ncRNA [38]. As shown for SRA/SRAP [27], long and short ASCC3 isoforms
152 have opposite effects on transcription recovery after DNA damage. This duality of information
153 carried by short and long ALE isoforms may also apply to other UV-induced ALE transcripts with
154 yet unknown contribution to the DNA damage response [38].

155 It is also worth mentioning cases where deregulated AS caused by alterations of splicing
156 mechanisms in pathological conditions promotes the formation of a ncRNA instead of a coding
157 mRNA, with dramatic consequences for cellular phenotypes. Alterations of AS are hallmarks of
158 cancer cells and, beyond the consequences for the formation of protein isoforms with gain or loss of
159 function, they can also lead to the release of functional ncRNA. For example, the locus
160 *PNUTS/PPP1R10* encodes a protein phosphatase 1 (PP1) regulatory subunit also known as PNUTS
161 (Phosphatase Nuclear Targeting Subunit) involved in cell cycle progression, DNA repair and
162 apoptosis by regulating the activity of the protein phosphatase 1 (PP1). In cancer cells, the
163 unmasking of an alternative splice site located in an exon disrupts the ORF and releases a lncRNA.
164 The lncRNA-PNUTS acts as a sponge for a key determinant of the epithelial phenotype, miR-205,
165 thereby promoting Epithelial-Mesenchymal Transition (EMT) and tumour progression [39].

166 As a whole, these examples underline the importance of AS as a developmental switch, which
167 allows a single transcription unit to produce multiple transcripts with distinct functions and fates
168 depending on the needs of the cell. Yet, it is important to remind that these transcripts can co-exist
169 in the cell and that the overall function results from the balance between the levels of the different
170 isoforms and the distinct functions they perform. Hence, addressing the function of bifunctional
171 RNAs requires to uncouple the often-antagonistic coding and non-coding functions. Inferring their
172 function(s) without knowing the full range of transcripts produced in a given context and their
173 intertwining is likely to lead to false mechanistic assumptions and release of poor biomarkers of
174 pathophysiological situations.

175

176

177 *Splicing of introns and the release of short ncRNAs*

178 In contrast to generating RNAs with new functionality when retained in a mRNA, introns can
179 also function as precursors of smaller ncRNAs when they are excised from the pre-mRNA. This is
180 the case for all known mammalian snoRNAs [40], for non-canonical miRNAs like mirtrons or
181 simtrons [41], or yet unknown categories of intron-derived small RNAs. We collectively named
182 these transcripts SID for Short Intron-Derived ncRNAs [42].

183 SnoRNAs are abundant, short, nucleoli-residing, small ncRNA, best known for guiding
184 post-transcriptional modifications of other ncRNAs such as rRNAs, tRNAs and snRNAs [40]. They
185 constitute so far the majority of known SIDs. Unlike in yeast and plants where snoRNAs can have
186 their own promoter, all mammalian snoRNAs are embedded within introns, usually on a one *per*
187 intron basis. They are transcribed from their host genes as portions of the pre-RNA, and the
188 functional snoRNAs are then produced by exonucleolytic trimming after splicing [40,43] (**Figure**
189 **1**). Curiously, snoRNA host transcripts were never referred to as RNAs with dual function, although
190 they clearly are since most host genes generate protein products in addition to their snoRNAs.
191 Maybe is it because snoRNAs were initially identified in introns of ribosomal protein genes or
192 genes encoding translation factors or nucleolar proteins, thus both contributing to the same general
193 biosynthetic process. However, this is no longer a general rule with the identification of more
194 widespread functions in pre-mRNA processing [44] and the identification of non-rRNA targets of
195 snoRNAs [45], stressing the need to distinguish coding from non-coding functions carried by the
196 same snoRNA host transcript in functional studies. More striking is the identification of non-
197 protein-coding hosting transcripts. Indeed, a non-negligible fraction of human snoRNAs [about
198 22% (**Table 2**); calculated from UCSC main table "Genes and Gene Predictions" intersected with
199 DASHR collection [46]] lies within small introns of ncRNAs, with the remarkable case of so-called
200 snoRNA host genes (SNHG) that can shelter many snoRNAs in the same transcript [47]. Prominent
201 examples include SNORD115 and SNORD116 clusters of dozens of snoRNAs produced from the
202 same SNHG14 ncRNA. SNHG ncRNAs do not have clear functions (**Supplementary Table 1**)
203 other than being dedicated to the production of snoRNAs, although their deregulation has been
204 associated with carcinogenesis [47,48]. A prominent example is the growth arrest-specific 5
205 (GAS5) lncRNA, which carries almost one snoRNA in each of its 11 introns, and functions as a
206 tumor suppressor [49]. Human GAS5 was proposed to also carry snoRNA-independent functions as
207 a ribo-suppressor of glucocorticoid receptors binding to their genomic targets, as a decoy for
208 miRNA or even as a small peptide-encoding ncRNA, although these features lack conservation
209 among species [49,50]. Hence, human GAS5 is typically a lncRNA with multiple functions
210 although several questions remain to be formally addressed like the existence of small peptides

211 from GAS5 translatable small ORFs (sORF) *in vivo*, and more importantly, the decoupling of the
212 relative contribution of snoRNAs or ribo-mimic functions of GAS5 to understand which elements
213 confer the tumor suppressor function of the *GAS5* gene.

214 The second most represented class of SIDs is the class of non-canonical miRNAs such as
215 simtrons and mirtrons, which are produced independently of the microprocessor DiGeorge
216 syndrome Critical Region 8 (DGCR8) enzyme [41,51-53]. The release of simtrons is concomitant to
217 splicing and involves cleavage by the RNase III enzyme Drosha recruited through the U1 general
218 splicing factor directly on the pre-mRNA [41,54]. Mirtrons are also processed by Drosha but post-
219 splicing, and thus require debranching of the intron-lariat by Debranching RNA enzyme 1 (DBR1)
220 [51,52]. Hence, simtrons and mirtrons, and a few others described recently [42], are dependent on
221 both the transcription of their host genes and on intron cleavage/splicing. Cases of intron-derived
222 miRNAs have been reported as targeting their own host RNA in regulatory feedback loops [55],
223 suggesting that duos of SIDs/host transcripts are involved in the same biological processes or
224 functions, although it remains poorly documented. In *Drosophila*, a recent report showed how a
225 mirtron, its host and target transcripts form a complex network of regulatory loops to control
226 synaptic homeostasis and neural activity [56]. We can speculate that this type of regulation is far
227 from being anecdotal since nearly 500 mirtrons have been identified in humans [57]. Since miRNAs
228 have many potential target transcripts whatever their origin and biogenesis, they are likely to have
229 more pleiotropic functions than their host transcript, stressing that the evaluation of their release in
230 certain normal or pathological cellular contexts must be discriminated from the expression levels of
231 their host transcript.

232 Another remarkable case of small ncRNAs that are directly produced through splicing of a
233 precursor pre-RNA is the class of circular RNAs, which, by definition, stand out from others by
234 their circular structure. After the splicing reaction, an intron lariat is produced, which normal fate is
235 to be debranched and degraded, although they can be trimmed by an exonuclease to produce an
236 intronic circular RNA (ciRNA) [58]. Although the function of these RNAs is not yet fully
237 understood, 300 ciRNAs have been predicted in humans [59]. In addition, pre-RNAs can also
238 generate covalently-closed circular RNAs (circRNAs) through back-splicing, a reverse splicing
239 process where the donor splice site reacts with an upstream acceptor splice site [see definitions in
240 [60,61]]. CircRNAs can be composed of exons (EcircRNAs), introns (IcircRNAs), or both
241 (EIcircRNAs). Recent studies have shown that this type of ncRNAs could act as miRNA sponges or
242 transcriptional regulators [61]. Similar to the above-mentioned SIDs, circRNAs coexist with the
243 mature long RNA, producing a duo of RNA molecules whose cooperation or antagonist functions
244 remain to be addressed. With the advent of high throughput sequencing, about 100,000 circular
245 RNAs have been identified and characterized in humans [62]. According to circBase and a circRNA

246 expression resource of 20 human tissues [62,63], circRNAs are essentially tissue-specific and only a
247 small subset is expressed in a given cell type, generally at high levels and are often more abundant
248 than their linear host RNAs [62,63]. Although the functions of most circRNAs as decoys and
249 whether they are translated or influence the functions of their linear counterparts is still debated,
250 their abundance and restriction to a given cell type is strikingly reminiscent of that of miRNAs [46],
251 pointing to promising roles in gene regulatory networks.

252 Transfer RNAs are also first transcribed as a precursor (pre-tRNA) and subjected to splicing
253 [64] in all species. We can mention the striking example of the pre-tRNA^{Trp} which produces a
254 functional C/D box snoRNA after splicing of its intron (**Figure 2**) in the archae *Haloflex volcanii*
255 [65]. Interestingly, the associated small nucleolar ribonucleoprotein complex (snoRNP) serves as a
256 guide for the 2'-O-methylation (2'OMe) of its own host transcript, which is a post-transcriptional
257 editing typically involved in the proper functioning of tRNAs [66]. However, the deletion of the
258 intron or the absence of the tRNA^{Trp} editing have no impact on the viability of *H. volcanii* in normal
259 conditions, although it was suggested that it may manifest itself under certain stress or competitive
260 conditions [67].

261 As a whole, certain pre-RNA can produce two types of transcripts, *i.e.* coding mRNAs and
262 small regulatory SIDs or circular RNAs. The case of the above-mentioned SRA is even more
263 remarkable as it can release a coding mRNA and a SID, or a mRNA and a long ncRNA that retains
264 the intron containing the SID [42]. Likewise, certain tRNA genes can release a tRNA and a small
265 regulatory RNA. Hence, it appears that precursor RNAs are not always mere transient conveyors of
266 genetic information as they contain multiple pieces of information and can generate a panel of
267 mature RNAs. In that respect, introns represent an obvious contributor to transcriptional
268 diversification as their retention in, or excision from, the pre-RNA will greatly increase the output
269 of a single transcription unit. It is also not surprising that more examples of bifunctional pre-RNA
270 are found in higher eukaryotes, especially mammals, since the number of introns tends to increase
271 with the increased developmental complexity of organisms [68].

272

273 ***Splicing of exons and the release of short ncRNAs***

274 In addition to AS of introns, exon skipping is also well known to create normal or
275 pathological post-transcriptional diversity. The *H19* locus is fascinating and probably the first
276 example of a ncRNA hosting a small regulatory RNA in an exon. *H19* is transcribed in a 2.3 kb
277 long lncRNA, exclusively from the allele inherited from the mother, which acts as a trans-regulator
278 of an imprinted gene network involved in the control of fetal and early postnatal growth in mice
279 [69]. In fact, transcription within the *H19* locus is extremely complex. In addition to the *H19*
280 transcript, two transcripts in the antisense orientation have been described, also from the maternal

281 allele: a coding transcript HOTS (H19 opposite tumor suppressor, 6 kb long) [70] and a very long
282 ncRNA named 91H (120 kb long) [71]. Adding to the complexity, the H19 transcript itself also
283 contains a highly conserved stem-loop structure embedded within its first exon and shown to release
284 hsa-miR-675 [72,73]. H19 is highly expressed in most fetal tissues and in the placenta, but the
285 processing of miR-675 seems to be restricted to the latter where it operates growth suppressing
286 functions [72]. This tight control may be overcome in physiopathological conditions to allow rapid
287 inhibition of cell proliferation [72]. However, whether H19 functionality resides solely in its role as
288 an abundant pri-miRNA, or if H19 operates independent functions as suggested by the disparity
289 between H19 and miR-675 expression and findings of miRNAs sponge activity in muscle cells [74],
290 is still debated. There are probably many other such examples since exonic miRNAs represent 3.9%
291 of total miRNAs, even if it may represent less than a hundred of associated transcripts and hence,
292 candidate bifunctional RNAs [75]. Yet, and in contrast to miRNAs produced through splicing of
293 introns discussed above, it remains to be tested whether and in which contexts exon skipping would
294 promote the release of a miRNA and a functional RNA knowing that i) about half of the exonic
295 miRNA are embedded within exons of spliced lncRNAs whose functions are often elusive and ii)
296 Drosha processing of an exonic miRNA is likely to impair the production of the spliced host mRNA
297 with coding functions [75].

298

299 *Precursor RNAs producing distinct non-coding functions*

300 It is commonly thought that most lncRNAs are expressed at low levels and are poorly
301 conserved at the sequence level across species. However, certain classes of lncRNAs are clearly as
302 abundant as mRNAs transcribed from housekeeping genes, with a common denominator of
303 escaping degradation and accumulating as surprisingly stable transcripts since they lack terminal
304 structures typical of Pol II transcripts including the polyA tail [76]. The reason is obvious for
305 circular RNAs that have no end or snoRNAs generated through splicing, but this is more
306 remarkable for longer transcripts. Prime examples include MALAT1 (NEAT2) and MEN ϵ/β
307 (NEAT1) transcribed from adjacent loci. Metastasis Associated Lung Adenocarcinoma Transcript 1
308 RNA (MALAT1) is a lncRNA originally described as being upregulated in cancer cells [77], in fact
309 among the most abundant long ncRNAs in mouse and human tissues. This rarely spliced transcript
310 is also not poly-adenylated owing to its 3' end being cleaved by the endonuclease RNase P at the
311 level of an evolutionary conserved tRNA-like structure, which simultaneously generates the non-
312 polyadenylated long MALAT1 transcript and a tRNA-like small ncRNA [77,78]. The thus-
313 generated 3' end of MALAT1 folds into a triple helix that prevents exonucleolytic degradation,
314 which is further processed to generate a mature 61-nt transcript known as mascRNA (MALAT1-
315 associated small cytoplasmic RNA). MALAT1 is retained in nuclear speckles where it associates

316 with serine/arginine-rich (SR) proteins and splicing factors [79]. The knock-down of this lncRNA
317 impacts the localization of splicing factors in nuclear speckles, and thus, MALAT1 is thought to
318 play a role in the regulation of splicing [79] although it was proposed to regulate gene expression
319 via multiple mechanism [76]. In clear contrast, mascRNA is exported to the cytoplasm although it is
320 unlikely that it reads the genetic code as its anticodon loop is poorly conserved. A dichotomy of the
321 immunoregulatory functions of MALAT1-mascRNA system has been reported following selective
322 ablation of mascRNA in monocytes [80], but the exact biological function of mascRNA remains
323 undefined. Nevertheless, the primary MALAT1 transcript is the first example of a pre-RNA
324 processed into two mature ncRNAs with distinct fates other than via the intron splicing process
325 [78].

326 The multiple endocrine neoplasia- β locus (MEN1) is also able to generate lncRNAs (MEN- ϵ
327 and - β) that are essential organizational components of nuclear paraspeckles [81]. Whereas MEN- ϵ
328 is subjected to canonical cleavage/polyadenylation, the mature 3' end of the longer isoform MEN- β
329 is generated via the same mechanism as for MALAT1 together with a tRNA-like RNA (menRNA).
330 Although the latter is structurally unstable in most mouse and human cells, it is stable in other
331 species for unclear reasons and unknown functional consequences, but suggesting that MEN- β
332 could operate dual functions in these species [81,82].

333 Strikingly, more than 100 loci in vertebrate genomes present a MALAT1 3'-end triple helix
334 structure and its immediate downstream tRNA-like structure [83]. Their functional dichotomy
335 remains to be demonstrated but they deserve further attention when the possible pathogenic
336 relevance of their non-coding transcripts is addressed.

337 To add to the most recent sources of regulatory small RNAs, the finding that small ncRNAs
338 can themselves be further processed into smaller RNA species is a direct consequence of the rapid
339 progress in RNA sequencing and bioinformatic analysis [84]. More than 25,000 fragments derived
340 from tRNAs (tRFs) [85-87] or hundreds from snoRNAs (sdRNAs) [88-91], around 14 to 40nt long
341 depending on species, were indeed found in RNAseq data and first thought to be degradation
342 products. There is now mounting evidence that their biogenesis is conserved and that they are
343 processed to generate stably accumulating fragments in a non-random and regulated manner. These
344 small RNAs fragments share similar features with miRNAs and were actually found to be
345 associated with Argonaute (AGO) proteins suggestive of a role in translational repression, although
346 they may interfere with translation through other mechanisms [92]. The processing of a subset of
347 tRNAs into tRFs was actually thought to provide a rapid response to downregulate RNA translation
348 during the stress response [93]. An interesting finding is that tRFs target endogenous retroviruses
349 and inhibit their retrotransposition through a miRNA-like silencing of transposon reverse
350 transcription as demonstrated in mice [94]. Because all organisms have tRNAs, it is possible that

351 this is in fact a highly conserved mechanism to control the deleterious mobility of transposons in
352 contexts where they escape epigenetic silencing, at early embryonic development stages for
353 example when the epigenetic landmarks of parental genomes are reset [94]. At these stages, the
354 functions of tRNAs and their derived tRFs are clearly distinct. As for sdRNAs, examples of
355 processed snoRNAs have been described for both H/ACA and C/D boxes snoRNAs [89,95,96]. As
356 reviewed in [90], there is a significant overlap between snoRNA and miRNA processing enzymes,
357 including AGO and DICER, their functional binding partners and even their subcellular
358 localizations, which renders the distinction between their respective activities somewhat tricky.
359 There are a few reports that do provide support for the functions of both snoRNA and miRNA co-
360 existing within the same molecule. Yet, one has to admit that this is essentially based on the use of
361 mimics for the sdRNA-miRNA-like fragment and evidence that the miRNA-like and its parental
362 snoRNA are enriched in distinct subcellular locations, whereas loss-of-function experiments are
363 likely to affect both types of molecules altogether. Nonetheless, target genes for these sdRNAs have
364 been predicted and validated for a few of them, with prominent examples pertaining to the
365 regulation of the p53/Mdm2 (Mouse double minute 2 homolog) feedback loop central to tumor
366 suppression [[97], reviewed in [88]].

367 In sum, tRNAs and snoRNAs are among the best-known and best-studied small ncRNAs.
368 However, they also hide unexpected functions through their processing into smaller ncRNAs that
369 add to the complexity of regulatory networks of genomes functions. With the resulting processed
370 small RNA fragments operating miRNA functions, an adverse consequence is that the dual
371 function, or more precisely the individual function of the tRNA or snoRNA entities in the processes
372 under consideration, is largely overlooked compared to that of the miRNA-like fragments.
373 Nevertheless, their abnormal levels is an index of a pathological situation, and because small RNAs
374 are easily detectable in body fluids, they are useful biomarkers for human diseases [88,98]. In
375 addition, it could be considered that if their physiological function in most healthy physiological
376 contexts may be marginal, the stress-induced maturation that they have in common [93,99] is
377 actually a mechanism that promotes the production of distinct entities to orchestrate protective
378 cellular functions.

379

380

381 **2/ When a single transcript performs multiple functions**

382 *Long non coding RNAs with small ORFs*

383 Remarkably enough, some of the RNAs that are currently considered as "true" ncRNAs may
384 actually have the potential to be translated [100]. In recent years, the last criterion for distinguishing
385 coding from non-coding RNAs has been shaken up with findings that lncRNAs can be associated

386 with polysomes [101]. In fact, large-scale ribosome profiling estimated that nearly 40% of known
387 human lncRNAs are associated with ribosomes in the cytoplasm [102]. However, whether certain
388 lncRNAs are actually translated or this is used as a strategy to regulate their abundance in the cell is
389 still far from being clear. Today, the real challenge is to identify these short peptides translated from
390 lncRNAs *in vivo* and determine whether they have a function, or an impact, in a biological process.
391 Some of them have been studied in depth and their associated functions in the cell have been
392 described. This is the case for Myoregulin (MLN) and Dwarf open reading frame (DWORF)
393 micropeptides. MLN and DWORF have 46 and 34 amino acids and are encoded by the lncRNAs
394 LINC00948 and LOC100507537, respectively [103,104]. They are both involved in muscle
395 relaxation mediated by the re-import of Ca^{2+} into the sarcoplasmic reticulum by the
396 sarco/endoplasmic reticulum Ca^{2+} -ATPase (SERCA) pump (**Figure 3**). MLN acts as a SERCA
397 pump inhibitor and therefore prevents muscle relaxation, whereas DWORF increases SERCA pump
398 activity by displacing SERCA inhibitors and thus facilitates muscle relaxation [103,104]. We can
399 also mention the most recent example of the lncRNA LINC00116 which encodes a highly
400 conserved 56-amino-acid mitochondrial micropeptide, Mitoregulin, which has been implicated in
401 the regulation of respiratory efficiency in cardiac and skeletal muscle cells [105]. Micropeptides
402 have so far evaded annotation efforts because of the technical difficulties to identify micropeptides,
403 but also probably because the expression of at least a subset of these bifunctional micropeptide-
404 encoding lncRNAs is restricted to specific cell types [22,106]. Consequently, their host transcripts,
405 as many others, may have just been misannotated as lncRNAs, awaiting the identification of
406 associated peptides. Many of the corresponding transcripts have now been reclassified as coding
407 transcripts in the latest version of GENCODE (v33) with an NM_ nomenclature. They have not
408 been formally studied in terms of function, localization and downregulation in conditions where
409 their coding portion is deleted, so the dual function of such transcripts is still debated. Of course,
410 this would require being able to uncouple the RNA functionality from its coding potential as it has
411 already been performed in other instances [12,107].

412 Among the examples cited in Table 1, LINC00961 has been reclassified as a protein-coding
413 gene because it was shown to encode a 90-amino-acid micropeptide called Small Regulatory
414 Polypeptide of Amino-Acid Response (SPAAR) involved in muscle regeneration [108]. In the
415 context of angiogenesis, SPAAR and LINC00961 have a pro- and an anti-angiogenic role,
416 respectively. Whereas SPAAR interacts with the pro-angiogenic Spectrin Repeat Containing
417 Nuclear Envelope Protein 1 (SYNE1) protein involved in the connection of organelles to actin
418 filament and endothelial cell migration, LINC00961 binds to and negatively regulates the function
419 of Thymosin beta 4 ($\text{T}\beta 4$) protein involved in the rearrangement of actin filament network and the
420 induction of angiogenesis [109]. Thus, LINC00961 can be considered as a bifunctional RNA for

421 which both the lncRNA and its encoded micropeptide contribute to the regulation of angiogenesis.
422 Of note, it would be of interest to know whether a switch of expression of the two entities occurs
423 when angiogenesis is needed and what is the mechanism underlying this regulation. In any case, the
424 example of the pair LINC00961/SPAAR supports the idea that some micropeptide-encoding
425 lncRNAs hide multiple functions that have not yet been uncovered in the appropriate cellular
426 context.

427 In a more large scale study, van Heesch *et al.* have established the translome of the human
428 heart where they found that 129 lncRNAs over the 783 expressed were also translated into
429 micropeptides [110]. Surprisingly, among these, 27 have assigned non-coding functions like the
430 lncRNAs NEAT1 [111] and JPX (Just Proximal to XIST) [112], and 4 are known SNHG such as
431 the already mentioned GAS5. Moreover, 22 micropeptides encoded from lncRNAs were localized
432 to mitochondria and associate with mitochondrial processes, although further investigation is
433 needed to fully dissect the role of these micropeptides in mitochondrial processes [110]. However,
434 this study stresses the fact that a number of translated lncRNAs are likely to possess both coding
435 and non-coding roles, and that this previously unrecognized biology of lncRNAs is likely to prompt
436 a multitude of follow-up studies.

437 In sum, the discovery of translatable sORF in so-called lncRNAs also goes against a rigid and
438 dichotomic classification of RNA molecules into strictly coding or non-coding [100,104,113].
439 Instead, it lends support to a model where translation, just as previously reported for transcription,
440 might also be a rather pervasive mechanism [114].

441

442 ***Messenger RNAs that perform non-coding functions***

443 Just like ncRNAs can produce peptides, some RNAs have been assigned functions distinct
444 from their role as intermediates in protein synthesis, and are also typically bifunctional. This is the
445 case for SRA or p53 (tumor protein 53) transcripts [27,107]. This is not surprising because mRNAs
446 have stable secondary structures just like ncRNA do [12,115], at least predicted with RNAfold
447 (<http://unafold.rna.albany.edu/>), which allow transcripts to interact with proteins, and in some
448 reported cases, to regulate their function or localization. As an example, SRA mRNA interacts with
449 MyoD, the master transcription factor of myogenic differentiation, leading to MyoD transcriptional
450 activation and activated muscle differentiation/reprogramming [27]. Likewise, p53 mRNA interacts
451 with the E3 ubiquitin-protein ligase mdm2 and prevents the latter from promoting the degradation
452 of the p53 protein. Hence, it is tempting to speculate that other mRNAs could actually operate dual
453 functions in a given cell type or in a particular cellular context as it is the case for p53 mRNA in
454 response to stress [107,116].

455 It is interesting to note that the ncRNAs and their associated protein(s) are mainly involved in
456 the same biological processes, or even in the same pathways. Indeed, if we focus on the functions
457 attributed to the examples of ncRNA/protein pairs mentioned in **Table 1**, 14 out of the 18 operate in
458 the same pathways, although it may be in different space-time frames. They can also be involved in
459 the regulation of one another as shown for the duo SRAP/SRA RNA [27]. The molecular basis of
460 this kind of switch of activity is not clear, but it could occur in specific circumstances when the
461 integrity of the genome is threatened and the cell's defence strategies need to be strengthened or
462 modulated.

463

464 **3/ A cascade of events: protein-coding genes that also produce “ncRNAs” with a coding** 465 **capacity**

466 As mentioned above, circRNAs originate from back-splicing of a pre-RNA, and were
467 originally described as ncRNAs with, to date, mainly a miRNA sponge effect [117]. However,
468 hundreds of circRNAs were recently shown to be associated with ribosomes [118], although they
469 lack the 5' cap owing to their splicing origin. Nevertheless, N-6 methyladenosine (m6A) RNA
470 editing or short internal ribosome entry sites-like (IRES-like) elements were suggested to serve as
471 translation start sites [119,120]. As an example, the circRNA originating from the back-splicing of
472 *ZNF609* exon 2 has been involved in the proliferation of myoblasts [121]. It also encodes a peptide
473 owing to the presence of an IRES-like sequence [121]. Further research is still needed to determine
474 the possible functions of this circRNA-encoded peptide, and in a way that would discriminate them
475 from the functions of the circRNA. Likewise, the *CTNBI* (β -catenin) gene produces a circRNA
476 called circ β -catenin, which in turn codes for a new β -catenin shorter isoform [122]. This isoform
477 serves as a decoy and competitively prevents GSK3 β -mediated degradation of the full-length β -
478 catenin protein, the overall consequence being the activation of the Wnt/ β -catenin pathway involved
479 in cancer progression [122]. In fact, the production of smaller protein isoforms through back-
480 splicing and circularization of selected exons from their linear mRNA transcripts may be a more
481 widespread mechanism with the demonstration that hundreds of cirRNAs do harbour a canonical
482 start codon [121]. Whether these shorter isoforms would all compete with the longer isoforms
483 produced from the same locus deserves attention. Despite unknowns regarding the cellular contexts
484 that control the competition between canonical splicing and back-splicing and lead to variable
485 expression patterns of circRNAs and linear RNAs [123], abnormal levels of cirRNAs have been
486 implicated in tumorigenesis, possibly through the production of small protein isoforms [124].

487

488 **4/ From multitasking to multifunctional RNAs throughout evolution?**

489 One of the most preserved processes during evolution, from lower to higher eukaryotes, *i.e.*
490 from yeast to human, is probably the splicing of pre-mRNAs, whereby introns are removed from
491 the original transcript leading to the juxtaposition of coding exons. Splicing has also been described
492 in prokaryotes, although this is a rare event, mainly occurring in tRNAs [125]. In mammals,
493 splicing is performed by the core spliceosome complex composed of snRNAs, the main ones being
494 snRNA U1, U2, U4, U5 and U6, and of about 50 proteins, among which the PRP8 protein (pre-
495 mRNA processing factor 8) is the largest and most highly conserved protein of the spliceosome
496 [126] (**Figure 4**). In bacteria, the whole mechanism is driven by a single molecule, the intron itself,
497 which belongs to the self-catalytic Group II introns [127]. In that case, Group II introns combine the
498 structure and function of snRNAs into six RNA domains (I to VI) with an ORF that encodes the
499 intron-encoded protein (IEP) equivalent to eukaryotic PRP8 [128] (**Figure 4**). Hence, group II
500 introns from bacteria are inherently all bifunctional RNAs. Yet, one could consider that this is
501 multitasking since the same RNA molecule, *i.e.* the intron, carries and not releases *per se*, both
502 RNU RNAs and IEP protein.

503 Even in ancient living organisms like viruses, the RNA contains much more than just the
504 genetic information translated into functional proteins since they also carry essential elements for
505 their replication [129]. In addition, their untranslated extremities also seem to have a structure
506 similar to that of an intrinsic tRNA [130].

507 Prokaryotic genomes are dominated by coding sequences, although dozens of regulatory
508 ncRNAs are present and expressed. Conversely, eukaryotic genomes are mostly non-coding, which
509 has led to the suggestion that the evolution of organisms towards increasing complexity has led to
510 the emergence of numerous ncRNAs with various functions, the expression of which is thought to
511 be restricted to different cell types and contexts [68,131]. The emergence of non-protein regulatory
512 molecules could have then favoured the formation of increasingly complex regulatory networks,
513 allowing for increasing sophistication in the way genomes are regulated in space and time to
514 achieve more complex organismal developmental processes, in particular related to the
515 development of new cognitive functions [68,132].

516 In fact, all the mentioned observations suggest that the multi-functionality of RNAs could in
517 fact result from a combination of the first two hypotheses with the following scenario: RNAs were
518 essential and omnipresent during the emergence of life, and were gradually supplanted by other
519 more effective and stable macromolecules, including DNA and proteins. Although they remained
520 present, they became available for additional functions as living organisms became more complex
521 and required increasingly sophisticated regulatory networks. If this scenario holds true, some RNAs
522 must have appeared recently during evolution. This is indeed the case in bacteria: most bacterial
523 regulatory RNAs are expressed exclusively in some species and are absent in others, despite their

524 phylogenetic proximity. Thus, RNAs are essential macromolecule in the emergence, evolution and
525 complexification of living organisms, but bifunctional RNAs probably allow high reactivity to
526 changes in the environmental conditions to which host organisms are subjected and for rapid and
527 effective micro-evolution.

528

529 **Conclusion**

530 It is clear now that the flow of genetic information does not simply go from DNA to proteins
531 through RNA intermediates, or from DNA to a vast repertoire of long and short non-coding RNAs.
532 We went over specific examples where a same genetic locus or pre-RNA can produce coding and
533 non-coding transcript, long or short, and shake the original belief that portions of the genome are
534 either coding or non-coding. Whether these are just discrete cases or the manifestation of a more
535 pervasive phenomenon remains to be properly evaluated. Nonetheless, one has to admit that the
536 information contained in a given portion of the genome can be multiple. As mentioned, all the
537 components produced may co-exist in the cell in a tightly controlled equilibrium and potentially in
538 the same regulatory loops or pathways. There are also cases where a switch of functions operates
539 during development or differentiation for instance. We have also mentioned in many places that
540 cellular stress is a widespread mechanism in most species that can cause a change between distinct
541 RNA entities. Why switching from a mRNA to a ncRNA or a long to small ncRNA may have to do
542 with the need to produce regulatory molecules in specific sub-cellular locations, with increased or
543 decreased stability or number of partners and targets. With splicing being mainly co-transcriptional,
544 it certainly offers a versatile system to rapidly switch gears in a developmental process or in
545 response to genomic insults. Although our vision of the mechanisms and actors behind these
546 changes in the nature of the RNAs produced by a genetic locus is still fragmentary, it is clear that
547 alterations of the equilibrium between these components is altered in pathological situations.

548 In contrast to the concept of pervasive transcription born over a decade ago, or of pervasive
549 translation proposed fairly recently, the concept of multiple information carried by genes and RNAs
550 only slowly attracts attention. This probably reflects the difficulties in searching for and testing this
551 multiplicity of information when non-targeted functional assays are likely to be misleading. This
552 would be however important in contexts where it appears that both entities from the same
553 transcription unit (mRNA and ncRNA or two ncRNAs) often operate in the same biological
554 processes, or even in the same biological pathways. If the balance between these entities contributes
555 to a global function and its dynamics influences cell fate, then, knowing which molecules are at
556 play would become important for functional experiments. Although it may still seem eccentric to
557 want to know if the function of a genomic locus is through proteins, long or short RNAs, or a

558 dynamic combination of these, it may become less anecdotal in disease situations and in the hunting
559 of diagnostic biomarkers or druggable targets.

560 The concept of bifunctional RNA is gaining momentum with an increasing number of
561 reported cases. We would like to speculate that studying these “chimeras” between coding and non-
562 coding RNA will help to decipher the evolutionary links between these two groups of molecules.

563

564

565 **Acknowledgments**

566 B.B. was supported by the Association Française Contre les Myopathies (AFM-Téléthon)
567 under fellowship #21363, C.F. is supported by the Institut National de la Santé et de la Recherche
568 Médicale (INSERM) and F.H. is supported by the Centre National de la Recherche Scientifique
569 (CNRS). This work was supported by AFM (Association Française contre les Myopathies), under
570 Grant #22341.

571

572

573 **Disclosure of interest**

574 The authors declare no conflict of interest.

575

576 **Author Contributions**

577 B.B., C.F. and F.H. wrote the paper.

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

References

1. Amaral PP, Dinger ME, Mercer TR et al. The eukaryotic genome as an RNA machine. *Science*. 2008;319:1787-1789.
2. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57-74.
3. Hombach S Kretz M. Non-coding RNAs: Classification, Biology and Functioning. *Adv Exp Med Biol*. 2016;937:3-17.
4. St Laurent G, Wahlestedt C, Kapranov P. The Landscape of long noncoding RNA classification. *Trends Genet*. 2015;31:239-251.
5. Kopp F Mendell JT. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*. 2018;172:393-407.
6. Cipolla GA, de Oliveira JC, Salviano-Silva A et al. Long Non-Coding RNAs in Multifactorial Diseases: Another Layer of Complexity. *Noncoding RNA*. 2018;4.
7. Bao Z, Yang Z, Huang Z et al. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res*. 2019;47:D1034-D1037.
8. Abascal F, Juan D, Jungreis I et al. Loose ends: almost one in five human genes still have unresolved coding status. *Nucleic Acids Res*. 2018;46:7070-7084.
9. Carninci P, Kasukawa T, Katayama S et al. The transcriptional landscape of the mammalian genome. *Science*. 2005;309:1559-1563.
10. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet*. 2007;8:413-423.
11. Gerstein MB, Bruce C, Rozowsky JS et al. What is a gene, post-ENCODE? History and updated definition. *Genome Res*. 2007;17:669-681.
12. Ulveling D, Francastel C, Hube F. When one is better than two: RNA with dual functions. *Biochimie*. 2011;93:633-644.
13. Chooniedass-Kothari S, Emberley E, Hamedani MK et al. The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Lett*. 2004;566:43-47.
14. Hube F, Guo J, Chooniedass-Kothari S et al. Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines. *DNA Cell Biol*. 2006;25:418-428.
15. Dinger ME, Pang KC, Mercer TR et al. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput Biol*. 2008;4:e1000176.
16. Kumari P Sampath K. cncRNAs: Bi-functional RNAs with protein coding and non-coding functions. *Semin Cell Dev Biol*. 2015;47-48:40-51.

- 616 17. Raina M, King A, Bianco C et al. Dual-Function RNAs. *Microbiol Spectr.* 2018;6.
- 617 18. Dhamija S Menon MB. Non-coding transcript variants of protein-coding genes - what are
618 they good for? *RNA Biol.* 2018;1-7.
- 619 19. Sampath K Ephrussi A. CncRNAs: RNAs with both coding and non-coding roles in
620 development. *Development.* 2016;143:1234-1241.
- 621 20. Qin F, Zhang Y, Liu J et al. SLC45A3-ELK4 functions as a long non-coding chimeric RNA.
622 *Cancer Lett.* 2017;404:53-61.
- 623 21. Hube F Francastel C. Mammalian introns: when the junk generates molecular diversity. *Int J*
624 *Mol Sci.* 2015;16:4429-4452.
- 625 22. Hube F Francastel C. Coding and Non-coding RNAs, the Frontier Has Never Been So
626 Blurred. *Front Genet.* 2018;9:140.
- 627 23. Muro EM, Mah N, Andrade-Navarro MA. Functional evidence of post-transcriptional
628 regulation by pseudogenes. *Biochimie.* 2011;93:1916-1921.
- 629 24. Ulveling D, Francastel C, Hube F. Identification of potentially new bifunctional RNA based
630 on genome-wide data-mining of alternative splicing events. *Biochimie.* 2011;93:2024-2027.
- 631 25. Lanz RB, McKenna NJ, Onate SA et al. A steroid receptor coactivator, SRA, functions as an
632 RNA and is present in an SRC-1 complex. *Cell.* 1999;97:17-27.
- 633 26. Emberley E, Huang GJ, Hamedani MK et al. Identification of new human coding steroid
634 receptor RNA activator isoforms. *Biochem Biophys Res Commun.* 2003;301:509-515.
- 635 27. Hube F, Velasco G, Rollin J et al. Steroid receptor RNA activator protein binds to and
636 counteracts SRA RNA-mediated activation of MyoD and muscle differentiation. *Nucleic*
637 *Acids Res.* 2011;39:513-525.
- 638 28. Chaudhary S, Khokhar W, Jabre I et al. Alternative Splicing and Protein Diversity: Plants
639 Versus Animals. *Front Plant Sci.* 2019;10:708.
- 640 29. Barash Y, Calarco JA, Gao W et al. Deciphering the splicing code. *Nature.* 2010;465:53-59.
- 641 30. Schmitz U, Pinello N, Jia F et al. Intron retention enhances gene regulatory complexity in
642 vertebrates. *Genome Biol.* 2017;18:216.
- 643 31. Thiele A, Nagamine Y, Hauschildt S et al. AU-rich elements and alternative splicing in the
644 beta-catenin 3'UTR can influence the human beta-catenin mRNA stability. *Exp Cell Res.*
645 2006;312:2367-2378.
- 646 32. Jacob AG Smith CWJ. Intron retention as a component of regulated gene expression
647 programs. *Hum Genet.* 2017;136:1043-1057.
- 648 33. Middleton R, Gao D, Thomas A et al. IRFinder: assessing the impact of intron retention on
649 mammalian gene expression. *Genome Biol.* 2017;18:51.
- 650 34. Edwards CR, Ritchie W, Wong JJ et al. A dynamic intron retention program in the
651 mammalian megakaryocyte and erythrocyte lineages. *Blood.* 2016;127:e24-e34.

- 652 35. Wong JJ, Au AY, Ritchie W et al. Intron retention in mRNA: No longer nonsense: Known
653 and putative roles of intron retention in normal and disease biology. *Bioessays*. 2016;38:41-
654 49.
- 655 36. Braunschweig U, Barbosa-Morais NL, Pan Q et al. Widespread intron retention in mammals
656 functionally tunes transcriptomes. *Genome Res*. 2014;24:1774-1786.
- 657 37. Wong JJ, Ritchie W, Ebner OA et al. Orchestrated intron retention regulates normal
658 granulocyte differentiation. *Cell*. 2013;154:583-595.
- 659 38. Williamson L, Saponaro M, Boeing S et al. UV Irradiation Induces a Non-coding RNA that
660 Functionally Opposes the Protein Encoded by the Same Gene. *Cell*. 2017;168:843-855.
- 661 39. Grelet S, Link LA, Howley B et al. A regulated PNUTS mRNA to lncRNA splice switch
662 mediates EMT and tumour progression. *Nat Cell Biol*. 2017;19:1105-1115.
- 663 40. Kiss T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular
664 functions. *Cell*. 2002;109:145-148.
- 665 41. Havens MA, Reich AA, Duelli DM et al. Biogenesis of mammalian microRNAs by a non-
666 canonical processing pathway. *Nucleic Acids Res*. 2012;40:4626-4640.
- 667 42. Hube F, Ulveling D, Sureau A et al. Short intron-derived ncRNAs. *Nucleic Acids Res*.
668 2017;45:4768-4781.
- 669 43. Dieci G, Preti M, Montanini B. Eukaryotic snoRNAs: a paradigm for gene expression
670 flexibility. *Genomics*. 2009;94:83-88.
- 671 44. Falaleeva M, Pages A, Matuszek Z et al. Dual function of C/D box small nucleolar RNAs in
672 rRNA modification and alternative pre-mRNA splicing. *Proc Natl Acad Sci U S A*.
673 2016;113:E1625-E1634.
- 674 45. Dudnakova T, Dunn-Davies H, Peters R et al. Mapping targets for small nucleolar RNAs in
675 yeast. *Wellcome Open Res*. 2018;3:120.
- 676 46. Leung YY, Kuksa PP, mlie-Wolf A et al. DASHR: database of small human noncoding
677 RNAs. *Nucleic Acids Res*. 2016;44:D216-D222.
- 678 47. Eddy SR. Noncoding RNA genes. *Curr Opin Genet Dev*. 1999;9:695-699.
- 679 48. Smith CM Steitz JA. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host
680 gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features
681 of snoRNA host genes. *Mol Cell Biol*. 1998;18:6897-6909.
- 682 49. Goustin AS, Thepsuwan P, Kosir MA et al. The Growth-Arrest-Specific (GAS)-5 Long
683 Non-Coding RNA: A Fascinating lncRNA Widely Expressed in Cancers. *Noncoding RNA*.
684 2019;5.
- 685 50. Smith CM Steitz JA. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host
686 gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features
687 of snoRNA host genes. *Mol Cell Biol*. 1998;18:6897-6909.
- 688 51. Berezikov E, Chung WJ, Willis J et al. Mammalian mirtron genes. *Mol Cell*. 2007;28:328-
689 336.

- 690 52. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing.
691 Nature. 2007;448:83-86.
- 692 53. Da Fonseca BHR, Domingues DS, Paschoal AR. mirtronDB: a mirtron knowledge base.
693 Bioinformatics. 2019;35:3873-3874.
- 694 54. Janas MM, Khaled M, Schubert S et al. Feed-Forward Microprocessing and Splicing
695 Activities at a MicroRNA-Containing Intron. PLoS Genet. 2011;7:e1002330.
- 696 55. Gao X, Qiao Y, Han D et al. Enemy or partner: relationship between intronic micromnas and
697 their host genes. IUBMB Life. 2012;64:835-840.
- 698 56. Amourda C Saunders TE. The mirtron miR-1010 functions in concert with its host gene
699 SKIP to balance elevation of nAcRbeta2. Sci Rep. 2020;10:1688.
- 700 57. Svanborg K, Gottlieb C, Bendvold E et al. Variation in, and inter-relationship between,
701 prostaglandin levels and other semen parameters in normal men. Int J Androl. 1989;12:411-
702 419.
- 703 58. Wilusz JE. A 360 degrees view of circular RNAs: From biogenesis to functions. Wiley
704 Interdiscip Rev RNA. 2018;9:e1478.
- 705 59. Talhouarne GJS Gall JG. Lariat intronic RNAs in the cytoplasm of vertebrate cells. Proc
706 Natl Acad Sci U S A. 2018;115:E7970-E7977.
- 707 60. Bogard B, Francastel C, Hube F. A new method for the identification of thousands of
708 circular RNAs. Non-coding RNA Investigation. 2018;2:1-5.
- 709 61. Jeck WR Sharpless NE. Detecting and characterizing circular RNAs. Nat Biotechnol.
710 2014;32:453-461.
- 711 62. Glazar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. RNA.
712 2014;20:1666-1670.
- 713 63. Maass PG, Glazar P, Memczak S et al. A map of human circular RNAs in clinically relevant
714 tissues. J Mol Med (Berl). 2017;95:1179-1189.
- 715 64. Abelson J, Trotta CR, Li H. tRNA splicing. J Biol Chem. 1998;273:12685-12688.
- 716 65. Singh SK, Gurha P, Tran EJ et al. Sequential 2'-O-methylation of archaeal pre-tRNA^{Trp}
717 nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of
718 pre-tRNA. J Biol Chem. 2004;279:47661-47671.
- 719 66. Pan T. Modifications and functional genomics of human transfer RNA. Cell Res.
720 2018;28:395-404.
- 721 67. Joardar A, Gurha P, Skariah G et al. Box C/D RNA-guided 2'-O methylations and the intron
722 of tRNA^{Trp} are not essential for the viability of *Haloferax volcanii*. J Bacteriol.
723 2008;190:7308-7313.
- 724 68. Kapusta A Feschotte C. Volatile evolution of long noncoding RNA repertoires: mechanisms
725 and biological implications. Trends Genet. 2014;30:439-452.
- 726 69. Gabory A, Jammes H, Dandolo L. The H19 locus: role of an imprinted non-coding RNA in
727 growth and development. Bioessays. 2010;32:473-480.

- 728 70. Onyango P, Feinberg AP. A nucleolar protein, H19 opposite tumor suppressor (HOTS), is a
729 tumor growth inhibitor encoded by a human imprinted H19 antisense transcript. *Proc Natl*
730 *Acad Sci U S A*. 2011;108:16759-16764.
- 731 71. Berteaux N, Aptel N, Cathala G et al. A novel H19 antisense RNA overexpressed in breast
732 cancer contributes to paternal IGF2 expression. *Mol Cell Biol*. 2008;28:6731-6745.
- 733 72. Keniry A, Oxley D, Monnier P et al. The H19 lincRNA is a developmental reservoir of
734 miR-675 that suppresses growth and Igf1r. *Nat Cell Biol*. 2012;14:659-665.
- 735 73. Cai X, Cullen BR. The imprinted H19 noncoding RNA is a primary microRNA precursor.
736 *RNA*. 2007;13:313-316.
- 737 74. Dykes IM, Emanuelli C. Transcriptional and Post-transcriptional Gene Regulation by Long
738 Non-coding RNA. *Genomics Proteomics Bioinformatics*. 2017;15:177-186.
- 739 75. Liu B, Shyr Y, Cai J et al. Interplay between miRNAs and host genes and their role in
740 cancer. *Brief Funct Genomics*. 2018;18:255-266.
- 741 76. Wilusz JE. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability.
742 *Biochim Biophys Acta*. 2016;1859:128-138.
- 743 77. Zhang X, Hamblin MH, Yin KJ. The long noncoding RNA Malat1: Its physiological and
744 pathophysiological functions. *RNA Biol*. 2017;14:1705-1714.
- 745 78. Wilusz JE, Freier SM, Spector DL. 3' end processing of a long nuclear-retained noncoding
746 RNA yields a tRNA-like cytoplasmic RNA. *Cell*. 2008;135:919-932.
- 747 79. Tripathi V, Ellis JD, Shen Z et al. The nuclear-retained noncoding RNA MALAT1 regulates
748 alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*.
749 2010;39:925-938.
- 750 80. Gast M, Schroen B, Voigt A et al. Long noncoding RNA MALAT1-derived mascRNA is
751 involved in cardiovascular innate immunity. *J Mol Cell Biol*. 2016;8:178-181.
- 752 81. Sunwoo H, Dinger ME, Wilusz JE et al. MEN epsilon/beta nuclear-retained non-coding
753 RNAs are up-regulated upon muscle differentiation and are essential components of
754 paraspeckles. *Genome Res*. 2009;19:347-359.
- 755 82. Brown JA, Valenstein ML, Yario TA et al. Formation of triple-helical structures by the 3'-
756 end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc Natl Acad Sci U S A*.
757 2012;109:19202-19207.
- 758 83. Zhang B, Mao YS, Diermeier SD et al. Identification and Characterization of a Class of
759 MALAT1-like Genomic Loci. *Cell Rep*. 2017;19:1723-1738.
- 760 84. Pages A, Dotu I, Pallares-Albanell J et al. The discovery potential of RNA processing
761 profiles. *Nucleic Acids Res*. 2018;46:e15.
- 762 85. Lee YS, Shibata Y, Malhotra A et al. A novel class of small RNAs: tRNA-derived RNA
763 fragments (tRFs). *Genes Dev*. 2009;23:2639-2649.
- 764 86. Keam SP, Hutvagner G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an
765 Ancient RNA in the Regulation of Gene Expression. *Life (Basel)*. 2015;5:1638-1651.

- 766 87. Pliatsika V, Loher P, Magee R et al. MINTbase v2.0: a comprehensive database for tRNA-
767 derived fragments that includes nuclear and mitochondrial fragments from all The Cancer
768 Genome Atlas projects. *Nucleic Acids Res.* 2018;46:D152-D159.
- 769 88. Abel Y Rederstorff M. SnoRNAs and the emerging class of sdRNAs: Multifaceted players
770 in oncogenesis. *Biochimie.* 2019;164:17-21.
- 771 89. Taft RJ, Glazov EA, Lassmann T et al. Small RNAs derived from snoRNAs. *RNA.*
772 2009;15:1233-1240.
- 773 90. Scott MS Ono M. From snoRNA to miRNA: Dual function regulatory non-coding RNAs.
774 *Biochimie.* 2011;93:1987-1992.
- 775 91. Chow RD Chen S. Sno-derived RNAs are prevalent molecular markers of cancer immunity.
776 *Oncogene.* 2018;37:6442-6462.
- 777 92. Kuscu C, Kumar P, Kiran M et al. tRNA fragments (tRFs) guide Ago to regulate gene
778 expression post-transcriptionally in a Dicer-independent manner. *RNA.* 2018;24:1093-1105.
- 779 93. Tao EW, Cheng WY, Li WL et al. tiRNAs: A novel class of small noncoding RNAs that
780 helps cells respond to stressors and plays roles in cancer progression. *J Cell Physiol.*
781 2020;235:683-690.
- 782 94. Schorn AJ, Gutbrod MJ, LeBlanc C et al. LTR-Retrotransposon Control by tRNA-Derived
783 Small RNAs. *Cell.* 2017;170:61-71.
- 784 95. Ender C, Krek A, Friedlander MR et al. A human snoRNA with microRNA-like functions.
785 *Mol Cell.* 2008;32:519-528.
- 786 96. Brameier M, Herwig A, Reinhardt R et al. Human box C/D snoRNAs with miRNA like
787 functions: expanding the range of regulatory RNAs. *Nucleic Acids Res.* 2011;39:675-686.
- 788 97. Yu F, Bracken CP, Pillman KA et al. p53 Represses the Oncogenic Sno-MiR-28 Derived
789 from a SnoRNA. *PLoS One.* 2015;10:e0129190.
- 790 98. Martens-Uzunova ES, Olvedy M, Jenster G. Beyond microRNA--novel RNAs derived from
791 small non-coding RNA and their implication in cancer. *Cancer Lett.* 2013;340:201-211.
- 792 99. Xiao J, Lin H, Luo X et al. miR-605 joins p53 network to form a p53:miR-605:Mdm2
793 positive feedback loop in response to stress. *EMBO J.* 2011;30:524-532.
- 794 100. Ruiz-Orera J, Messegue X, Subirana JA et al. Long non-coding RNAs as a source of new
795 peptides. *Elife.* 2014;3:e03523.
- 796 101. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells
797 reveals the complexity and dynamics of mammalian proteomes. *Cell.* 2011;147:789-802.
- 798 102. Zeng C, Fukunaga T, Hamada M. Identification and analysis of ribosome-associated
799 lncRNAs using ribosome profiling data. *BMC Genomics.* 2018;19:414.
- 800 103. Anderson DM, Anderson KM, Chang CL et al. A micropeptide encoded by a putative long
801 noncoding RNA regulates muscle performance. *Cell.* 2015;160:595-606.

- 802 104. Nelson BR, Makarewich CA, Anderson DM et al. A peptide encoded by a transcript
803 annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*.
804 2016;351:271-275.
- 805 105. Stein CS, Jadiya P, Zhang X et al. Mitoregulin: A lncRNA-Encoded Microprotein that
806 Supports Mitochondrial Supercomplexes and Respiratory Efficiency. *Cell Rep*.
807 2018;23:3710-3720.
- 808 106. Choi SW, Kim HW, Nam JW. The small peptide world in long noncoding RNAs. *Brief*
809 *Bioinform*. 2018;20:1853-1864.
- 810 107. Candeias MM, Malbert-Colas L, Powell DJ et al. P53 mRNA controls p53 activity by
811 managing Mdm2 functions. *Nat Cell Biol*. 2008;10:1098-1105.
- 812 108. Matsumoto A, Pasut A, Matsumoto M et al. mTORC1 and muscle regeneration are
813 regulated by the LINC00961-encoded SPAR polypeptide. *Nature*. 2017;541:228-232.
- 814 109. Spencer HL, Sanders R, Boulberdaa M et al. The LINC00961 transcript and its encoded
815 micropeptide, small regulatory polypeptide of amino acid response, regulate endothelial cell
816 function. *Cardiovascular Research*. 2020;116:1-14.
- 817 110. van HS, Witte F, Schneider-Lunitz V et al. The Translational Landscape of the Human
818 Heart. *Cell*. 2019;178:242-260.
- 819 111. Clemson CM, Hutchinson JN, Sara SA et al. An architectural role for a nuclear noncoding
820 RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell*. 2009;33:717-
821 726.
- 822 112. Vallot C Rougeulle C. Long non-coding RNAs and human X-chromosome regulation: a
823 coat for the active X chromosome. *RNA Biol*. 2013;10:1262-1265.
- 824 113. Chen J, Brunner AD, Cogan JZ et al. Pervasive functional translation of noncanonical
825 human open reading frames. *Science*. 2020;367:1140-1146.
- 826 114. Ingolia NT, Brar GA, Stern-Ginossar N et al. Ribosome profiling reveals pervasive
827 translation outside of annotated protein-coding genes. *Cell Rep*. 2014;8:1365-1379.
- 828 115. Ulveling D, Dinger ME, Francastel C et al. Identification of a dinucleotide signature that
829 discriminates coding from non-coding long RNAs. *Front Genet*. 2014;5:316.
- 830 116. Candeias MM. The can and can't dos of p53 RNA. *Biochimie*. 2011;93:1962-1965.
- 831 117. Panda AC. Circular RNAs Act as miRNA Sponges. *Adv Exp Med Biol*. 2018;1087:67-79.
- 832 118. Schneider T Bindereif A. Circular RNAs: Coding or noncoding? *Cell Res*. 2017;27:724-
833 725.
- 834 119. Yang Y, Fan X, Mao M et al. Extensive translation of circular RNAs driven by N(6)-
835 methyladenosine. *Cell Res*. 2017;27:626-641.
- 836 120. Fan X, Yang Y, Wang Z. Pervasive translation of circular RNAs driven by short IRES-like
837 elements. *BioRxiv*. 2018;23-29.
- 838 121. Legnini I, Di TG, Rossi F et al. Circ-ZNF609 Is a Circular RNA that Can Be Translated and
839 Functions in Myogenesis. *Mol Cell*. 2017;66:22-37.

- 840 122. Liang WC, Wong CW, Liang PP et al. Translation of the circular RNA circbeta-catenin
841 promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biol.*
842 2019;20:84.
- 843 123. Chen LL Yang L. Regulation of circRNA biogenesis. *RNA Biol.* 2015;12:381-388.
- 844 124. Guarnerio J, Bezzi M, Jeong JC et al. Oncogenic Role of Fusion-circRNAs Derived from
845 Cancer-Associated Chromosomal Translocations. *Cell.* 2016;165:289-302.
- 846 125. Reinhold-Hurek B Shub DA. Self-splicing introns in tRNA genes of widely divergent
847 bacteria. *Nature.* 1992;357:173-176.
- 848 126. Nguyen TH, Galej WP, Fica SM et al. CryoEM structures of two spliceosomal complexes:
849 starter and dessert at the spliceosome feast. *Curr Opin Struct Biol.* 2016;36:48-57.
- 850 127. Novikova O Belfort M. Mobile Group II Introns as Ancestral Eukaryotic Elements. *Trends*
851 *Genet.* 2017;33:773-783.
- 852 128. Vosseberg J Snel B. Domestication of self-splicing introns during eukaryogenesis: the rise
853 of the complex spliceosomal machinery. *Biol Direct.* 2017;12:30.
- 854 129. Forterre P. The two ages of the RNA world, and the transition to the DNA world: a story of
855 viruses and cells. *Biochimie.* 2005;87:793-803.
- 856 130. Chujo T, Ishibashi K, Miyashita S et al. Functions of the 5'- and 3'-untranslated regions of
857 tobamovirus RNA. *Virus Res.* 2015;206:82-89.
- 858 131. Mattick JS. The hidden genetic program of complex organisms. *Sci Am.* 2004;291:60-67.
- 859 132. Mattick JS Gagen MJ. The evolution of controlled multitasked gene networks: the role of
860 introns and other noncoding RNAs in the development of complex organisms. *Mol Biol*
861 *Evol.* 2001;18:1611-1630.
- 862 133. Novick RP, Ross HF, Projan SJ et al. Synthesis of staphylococcal virulence factors is
863 controlled by a regulatory RNA molecule. *EMBO J.* 1993;12:3967-3975.
- 864 134. Balaban N Novick RP. Translation of RNAlII, the *Staphylococcus aureus* agr regulatory
865 RNA molecule, can be activated by a 3'-end deletion. *FEMS Microbiol Lett.* 1995;133:155-
866 161.
- 867 135. Vanderpool CK, Balasubramanian D, Lloyd CR. Dual-function RNA regulators in bacteria.
868 *Biochimie.* 2011;93:1943-1949.
- 869 136. Gimpel M, Preis H, Barth E et al. SR1--a small RNA with two remarkably conserved
870 functions. *Nucleic Acids Res.* 2012;40:11659-11672.
- 871 137. Laressergues D, Couzigou JM, Clemente HS et al. Primary transcripts of microRNAs
872 encode regulatory peptides. *Nature.* 2015;520:90-93.
- 873 138. Campalans A, Kondorosi A, Crespi M. Enod40, a short open reading frame-containing
874 mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago*
875 *truncatula*. *Plant Cell.* 2004;16:1047-1059.
- 876 139. Rongo C, Gavis ER, Lehmann R. Localization of oskar RNA regulates oskar translation and
877 requires Oskar protein. *Development.* 1995;121:2737-2746.

- 878 140. Lim S, Kumari P, Gilligan P et al. Dorsal activity of maternal squint is mediated by a non-
879 coding function of the RNA. *Development*. 2012;139:2903-2915.
- 880 141. Kloc M, Foreman V, Reddy SA. Binary function of mRNA. *Biochimie*. 2011;93:1955-1961.
- 881 142. Peng L, Chen G, Zhu Z et al. Circular RNA ZNF609 functions as a competitive endogenous
882 RNA to regulate AKT3 expression by sponging miR-150-5p in Hirschsprung's disease.
883 *Oncotarget*. 2017;8:808-818.
- 884 143. Young TM, Tsai M, Tian B et al. Cellular mRNA activates transcription elongation by
885 displacing 7SK RNA. *PLoS One*. 2007;2:e1010.
- 886 144. Young TM, Wang Q, Pe'ery T et al. The human I-mfa domain-containing protein, HIC,
887 interacts with cyclin T1 and modulates P-TEFb-dependent transcription. *Mol Cell Biol*.
888 2003;23:6373-6384.
- 889 145. Nagano H, Yamagishi N, Tomida C et al. A novel myogenic function residing in the 5' non-
890 coding region of Insulin receptor substrate-1 (Irs-1) transcript. *BMC Cell Biol*. 2015;16:8.
- 891 146. Long YC, Cheng Z, Copps KD et al. Insulin receptor substrates Irs1 and Irs2 coordinate
892 skeletal muscle growth and metabolism via the Akt and AMPK pathways. *Mol Cell Biol*.
893 2011;31:430-441.
- 894 147. Xu Q, Walker D, Bernardo A et al. Intron-3 retention/splicing controls neuronal expression
895 of apolipoprotein E in the CNS. *J Neurosci*. 2008;28:1452-1459.
- 896 148. Kavela S, Shinde SR, Ratheesh R et al. PNUTS functions as a proto-oncogene by
897 sequestering PTEN. *Cancer Res*. 2013;73:205-214.
- 898 149. Roberts GC Smith CW. Alternative splicing: combinatorial output from the genome. *Curr*
899 *Opin Chem Biol*. 2002;6:375-383.

900
901
902
903
904

We apologize to those whose works have not been cited in this article owing to lack of space

Table 1. Examples of RNAs with dual functions

Organism	Type	Role	Refs
Bacteria			
RNAIII (<i>S. aureus</i>)	RNA	Modulates the production of toxins and enzymes by base-pairing on mRNA targets	[133]
	Protein	Controls the switch between early and late expression of several exotoxins	[134]
SgrS (<i>E. coli</i>)	RNA	Regulate sugar stress (prolonged stress)	[135]
	Protein	Regulate sugar stress (early response)	[135]
SR1 (<i>Bacillus</i>)	RNA	Regulation of arginine catabolism	[136]
	Protein	SR1P binds and stabilises the gapA operon mRNAs	[136]
tmRNA (<i>Bacteria</i>)	RNA	Function as a tRNA by adding alanine to a stalled polypeptide chain	[136]
	Protein	Function as a mRNA by incorporating a degradation tag to the polypeptide chain	[136]
miPEP171b (<i>M. truncatula</i>)	RNA	Downregulates target genes involved in root development	[137]
	Protein	Enhances the accumulation of its corresponding mature miRNA	[137]
miPEP165a (<i>A. thaliana</i>)	RNA	Downregulates target genes involved in root development	[137]
	Protein	Enhances the accumulation of its corresponding mature miRNA	[137]
Plants			
Enod40	RNA	Confers specificity for the recognition of mRNA targets of MtRBP1	[138]
	Protein	Affects MtRBP1 localization during nodule development	[138]
Insects			
Oskar (<i>Drosophila</i>)	RNA	Scaffold for the complexes essential for development of the oocyte (early stage)	[139]
	Protein	Scaffold for the complexes essential for development of the oocyte (late stage)	[139]
Fish			
Sqt (<i>Danio rerio</i>)	RNA	Deliver/sequester maternal factors for dorsal specification	[140]
	Protein	Deliver/sequester maternal factors for dorsal specification	[140]
Amphibians			
VegT (<i>Xenopus</i>)	RNA	Organization of the cyokeratin cytoskeleton	[141]
	Protein	Transcription factor of the T-box family	[141]
Mammals (mainly human and mice)			
p53	RNA	Stimulates p53 translation and prevents p53 protein from Mdm2-induced degradation	[116]
	Protein	Tumor suppressor	[116]
SRA1	RNA	Co-activator of transcription factors	[12]
	Protein	Inhibitor of the co-activation function of SRA RNA	[12]
circ-ZNF609	RNA	Serves as a sponge for miR-150-5p and controls cell proliferation	[142]
	Protein	Control of myoblast proliferation	[121]
ASCC3	RNA	Counteracts the function of the protein-coding isoform	[38]
	Protein	Negative regulator of the host defense response	[38]
LINC00961/SPAR	RNA	has as an anti-angiogenic role by inhibiting the function of Tβ4 protein	[109]
	Protein	Involved in muscle regeneration/has a pro-angiogenic role by binding the SYNE1 protein	[108]
HIC	RNA	Activates P-TEFb by displacing 7SK RNA	[143]
	Protein	Inhibits P-TEFb transcription	[144]
Irs-1	RNA	Inhibits myoblasts differentiation through complementary sequence to Rb mRNA	[145]
	Protein	Coordinates skeletal muscle growth and metabolism	[146]
apoE/apoE-I3	RNA	apoE-I3 controls apoE expression in neurons	[147]
	Protein	Involved in lipid transport system	[147]
PNUTS	RNA	regulates EMT migration and invasion in vitro through its miR-205 interaction	[39]
	Protein	functions as a proto-oncogene by sequestering PTEN	[148]

908 **Table 2. Number of small ncRNAs identified in intronic regions of lncRNAs.** All numbers were
 909 inferred from the DASHR database and were calculated from UCSC main table "Genes and Gene
 910 Predictions" intersected with DASHR collection [46]. *, These intronic snoRNA are included in
 911 snoRNA host-genes (SNHG); miRNA, microRNA; rRNA, ribosomal RNA; scRNA, small
 912 cytoplasmic RNA; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; tRNA, transfer
 913 RNA.
 914

Type of ncRNA	Total number	Within intron of lncRNA	%
miRNA	1,582	198	12.51
rRNA	1,723	204	11.84
scRNA	1,291	170	13.17
snoRNA	402	88 *	21.89
snRNA	4,278	533	12.46
tRNA	623	52	8.35
		Mean:	13.37 ± 4.51

915

916

917 **Figures legends**

918

919 **Figure 1. Diversification of proteomic and transcriptional outputs through constitutive or**
920 **alternative splicing.** More than 95% of introns are rapidly degraded after splicing (top left panel),
921 but some can escape degradation and then represent precursors of short ncRNA [149] (top right
922 panel). Retained introns can also favour the formation of protein isoforms (bottom left panel) or, if
923 it disturbs the ORF, it can promote the formation of a long ncRNAs (bottom right panel). Exons and
924 introns are represented by boxes and lines, respectively. mRNA, messenger RNA; ncRNA, non-
925 coding RNA; H/ACA snoRNA, H/ACA box small nucleolar RNA; C/D snoRNA, C/D small
926 nucleolar RNA.

927

928 **Figure 2. The 2'O-methylation of *H. volcanii* pre-tRNA^{Trp} is guided by its own intron.** Thick
929 arrows indicate the pre-tRNA^{Trp} processing pathway from nucleotide methylation and splicing to
930 the production of a tRNA^{Trp} and the excised intron. RNP, ribonucleoprotein complex; C, cytosine;
931 Cm, methylated cytosine; U, uracil; Um, methylated uracil. Adapted from [65].

932

933 **Figure 3. The micropeptides Myoregulin (MLN) and Dwarf open reading frame (DWORF).**
934 DWORF RNA and MLN RNA were first identified as long non-coding RNAs (lncRNAs; Refseq
935 numbers NR_037902 and BC069675 respectively). It appears that these two lncRNAs can encode
936 DWORF and MLN micropeptides, respectively. Subsequently, their RefSeq category has been
937 revised: NR_037902 became NM_001352129 and BC069675 were replaced by NM_001040109.
938 DWORF and MLN are both involved in muscle contraction. The first one enables muscle relaxation
939 by activating the SERCA calcium pump and thus the re-import of Ca²⁺ into the sarcoplasmic
940 reticulum. Conversely, the second one maintains muscle contraction by preventing the re-import of
941 Ca²⁺ by inhibiting the SERCA pump.

942

943 **Figure 4. The special case of group II self-catalytic introns.** (A) In prokaryotes, group II introns
944 are composed of several domains that confer their self-catalytic activity. Domains I to VI are
945 hypothesized to have evolved into separate activities in eukaryotes, namely snRNAs and the IEP
946 homolog, PRP8 protein. (B) In eukaryotes, splicing of introns requires distinct effector RNAs. Grey
947 arrows represent the 2 steps of the splicing reaction *i.e.* the nucleophilic attack of the branch point A
948 and of the free 3'OH of the exon. 5' and 3' stand for 5'- and 3'-ends of exon; IEP, intron-encoded
949 proteins; ORF, open reading frame. Adapted from Vosseberg [128].