A guide to naming human non-coding RNA genes

Ruth L Seal1,2,*, Ling-Ling Chen3, Sam Griffiths-Jones4, Todd M Lowe5, Michael B Mathews6, Dawn O’Reilly7, Andrew J Pierce8, Peter F Stadler9,10,11,12,13, Igor Ulitsky14, Sandra L Wolin15,16 & Elspeth A Bruford1,2

Abstract

Research on non-coding RNA (ncRNA) is a rapidly expanding field. Providing an official gene symbol and name to ncRNA genes brings order to otherwise potential chaos as it allows unambiguous communication about each gene. The HUGO Gene Nomenclature Committee (HGNC, www.genenames.org) is the only group with the authority to approve symbols for human genes. The HGNC works with specialist advisors for different classes of ncRNA to ensure that ncRNA nomenclature is accurate and informative, where possible. Here, we review each major class of ncRNA that is currently annotated in the human genome and describe how each class is assigned a standardised nomenclature.

Keywords gene nomenclature; gene symbols; non-coding RNA

Introduction

The HUGO Gene Nomenclature Committee (HGNC) works under the auspices of Human Genome Organisation (HUGO) and is the only worldwide authority that assigns standardised symbols and names to human genes (Braschi et al., 2019). A unique symbol for every gene is essential to enable unambiguous scientific communication, and approved symbols should be used ubiquitously in research papers, conference talks and posters, and biomedical databases. The HGNC endeavours to approve symbols for all classes of genes that are supported by gene annotation projects and began working on non-coding RNA (ncRNA) nomenclature in the mid-1980s with the approval of initial gene symbols for mitochondrial transfer RNA (tRNA) genes. Since then, we have worked closely with experts in the ncRNA field to develop symbols for many different kinds of ncRNA genes.

The number of genes that the HGNC has named per ncRNA class is shown in Fig 1, and ranges in number from over 4,500 long ncRNA (lncRNA) genes and over 1,900 microRNA genes, to just four genes in the vault and Y RNA classes. Every gene symbol has a Symbol Report on our website, www.genenames.org, which displays the gene symbol, gene name, chromosomal location and also includes links to key resources such as Ensembl (Zerbino et al., 2018), NCBI Gene (O’Leary et al., 2016) and GeneCards (Stelzer et al., 2016). We collaborate directly with these biomedical databases and, importantly, these databases always use our gene symbols as the primary symbol for the gene. Due to the relative completeness of the HGNC ncRNA gene set, our data have been chosen as the canonical human dataset in the RNAcentral database (The RNAcentral Consortium, 2019), an RNA sequence database resource. For microRNAs, we work with the specialist resource miRBase (Kozomara et al., 2019), and for tRNAs, we work with the specialist resource GtRNAdb (Chan & Lowe, 2016). We display links to these resources from the relevant Symbol Report. Where available, for lncRNAs we provide specialist links to LNCipedia (Volders et al., 2019), a key lncRNA resource that displays HGNC gene symbols (Box 1).

For each class of ncRNA, we host curated gene group pages on www.genenames.org—a list of URLs for these is shown in Table 1.

The aim of this paper was to provide an overview for each of the main types of ncRNA that we have named, as well as a guide to how we name them. Each section has been written in collaboration with our specialist advisors for each ncRNA class: Sam Griffiths-
Jones of the University of Manchester for microRNAs, Todd Lowe of the University of California, Santa Cruz for tRNAs, Dawn O’Reilly of the University of Oxford for small nuclear RNAs (snRNAs), Peter Stadler of the University of Leipzig for small nucleolar and vault RNAs, Andrew Pierce currently at AstraZeneca, Cambridge for ribosomal RNAs (rRNAs), Sandra Wolin of the NIH for Y RNAs, Michael Mathews of Rutgers New Jersey Medical School for small NF90 (ILF3) associated RNAs, and Igor Ulitsky of the Weizmann Institute of Science and Ling-Ling Chen of the Shanghai Institute for Biochemistry and Cell Biology for long non-coding RNAs. We finish by outlining recommendations for the nomenclature of circular and circular intronic RNAs, which are currently lacking official nomenclature.

MicroRNAs

MicroRNAs are transcripts of ~22 nucleotides that mediate the post-transcriptional regulation of genes via direct binding to messenger RNA (mRNA) molecules. In animal cells, microRNA (miRNA) genes are usually transcribed as long primary transcripts (pri-miRNAs), which are processed by the Drosha microprocessor complex into precursor hairpin stem-loop sequences (pre-miRNAs). These hairpins are exported from the nucleus to the cytoplasm, where the stem-loop is cleaved by the Dicer enzyme to produce a ~22 nt duplex. One strand of the duplex associates with an Argonaute (AGO) protein and this microRNA ribonucleoprotein complex (miRNP) binds to sites in mRNAs that are complementary to the miRNA sequence, usually in the 3' untranslated region (UTR). The Ago-miRNP complex then recruits other proteins, which typically mediate either the degradation or translational repression of the mRNA [for a review, see (Bartel, 2018)]. Approximately 60% of all human genes produce mRNAs that can be bound by miRNAs (Friedman et al., 2009), so these small RNAs provide regulation for diverse biological processes across all tissue types and stages of life. As such, miRNA genes have been implicated in many human diseases including rheumatoid arthritis (Guggino et al., 2018), deafness (Mencia et al., 2009), stroke (Panagal et al., 2019), psoriasis (Yan et al., 2019), cirrhosis (Fernández-Ramos et al., 2018) and several forms of cancer (Kwok et al., 2017).

The name “microRNA” to reflect the small size of the active RNA molecule was agreed upon and first used by three Caenorhabditis elegans research groups that published in the same 2001 issue of Science (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee & Ambros, 2001). Once the field of miRNA research started to expand, experts came together to publish guidelines on how to name these transcripts across species (Ambros et al., 2003), and the miRNA Registry was founded to ensure that the same symbols were not mistakenly used by different research groups for different

Box 1. Useful resources for non-coding RNA genes used by the HGNC

| RNA resource      | Resource URL           | Description                                                                 |
|-------------------|------------------------|------------------------------------------------------------------------------|
| RNAcentral        | https://rnacentral.org/| Centralised database of non-coding RNA sequences collated from expert non-coding RNA member databases, model organism databases and sequence accession databases |
| miRBase           | http://www.mirbase.org/| Searchable database of microRNA sequences and annotations. Also hosts the miRBase registry where researchers can submit prospective new microRNAs |
| GtRNAdb           | http://gtcrndb.ucsc.edu/| The genomic tRNA database, which contains predicted tRNA genes by the tRNAscan-SE program for many different species |
| snoRNABase        | https://www-snoRNA.biotoul.fr/| Database of human snoRNA genes; useful resource but no longer being updated |
| LNCipedia         | https://lncipedia.org/  | Database of human long non-coding RNA sequences and manually curated IncRNA articles |
| Ensembl           | http://www.ensembl.org/| Genome browser for vertebrate genomes that hosts the GENCODE annotation models for non-coding RNA genes for mouse and human genes |
| NCBI Gene         | https://www.ncbi.nlm.nih.gov/gene/| Integrated annotation and related information for many different genomes. Incudes RefSeq manual annotation of human and mouse non-coding RNA genes |
miRNAs (Griffiths-Jones, 2004). The miRNA Registry evolved into the dedicated online miRNA resource miRBase, which has continued to be responsible for providing unique identifiers for miRNAs as well as acting as a database of sequences and curated publications (Kozomara et al., 2019). Researchers submit hairpin and mature microRNA sequences to miRBase, which are then publicly assigned new symbols after manuscript acceptance. miRBase assigns each microRNA stem-loop sequence a symbol in the format “mir-” and each mature miRNA a symbol in the format “miR-#” followed by a unique sequential number that reflects order of submission to the database. The HGNC then approves a gene symbol for human microRNA genes in the format MIR#; for example, MIR10A and MIR10B. The HGNC does not accept any direct requests for miRNA gene symbols, and all requests must go to miRBase first (please see http://www.mirbase.org/registry.shtml).

In accordance with miRBase, the HGNC provides one gene symbol per miRNA gene, even though miRNAs are sometimes processed from the same transcripts as proteins or other miRNAs, and therefore might not be considered separate genes in the canonical sense. For example, many miRNAs are hosted in the introns, or less frequently the exons, of protein coding genes or long non-coding RNA genes (Fig 2 and Box 2). The HGNC has curated gene group pages listing these host genes (Table 1), and the naming conventions for non-coding miRNA host genes are discussed in the long non-coding RNA section below.

Recently, there have been a few ideas published on how to “improve” miRNA nomenclature, including correcting the identifiers of particular miRNA genes to show evolutionary relationships (e.g. Desvignes et al., 2015; Fromm et al., 2015; Budak et al., 2016). As nomenclature advisors, we understand the desire to perfect nomenclature systems once more information becomes available. At the same time, experience has taught us that such revised systems are often not fully adopted and may cause considerable confusion in the community. It can therefore be more appropriate to find other ways to represent relationships between genes, in order to maintain stable gene symbols. The HGNC has recently curated gene groups to show

| Gene group name                  | Gene group URL                  | Description                                                                                      |
|----------------------------------|---------------------------------|---------------------------------------------------------------------------------------------------|
| Non-coding RNAs                  | https://www.genenames.org/data/ | Overview page of all non-coding RNAs in the HGNC project. Can be used as a starting point to   |
|                                  | negroup/#!/group/475           | browse through all types of named ncRNAs                                                        |
| MicroRNAs                        | https://www.genenames.org/data/ | Starting page for all microRNAs, which are split into curated human families where possible.    |
|                                  | negroup/#!/group/476           | MicroRNAs not in a defined family are listed on the first page                                   |
| MicroRNA host genes              | https://www.genenames.org/data/ | A curated list of microRNA host genes, which is split into protein coding and non-coding         |
|                                  | negroup/#!/group/1690          | subgroups                                                                                       |
| Transfer RNAs                    | https://www.genenames.org/data/ | Starting page for all transfer RNA genes, with subgroups “Mitochondrially encoded transfer       |
|                                  | negroup/#!/group/478           | RNAs” and “Cytoplasmic transfer RNAs” (this page also has the subsets “Cytoplasmic transfer      |
|                                  |                                 | RNA pseudogenes” and “Low confidence cytoplasmic transfer RNAs”)                                 |
| Small nuclear RNAs               | https://www.genenames.org/data/ | Lists all canonical small nuclear RNA genes; variant snRNA genes are shown as a subgroup        |
|                                  | negroup/#!/group/1819          |                                                                                                  |
| Small nucleolar RNAs             | https://www.genenames.org/data/ | Starting page for snoRNAs with the subgroups “Small Cajal body-specific RNAs”, “Small         |
|                                  | negroup/#!/group/844           | nucleolar RNAs, C/D box” and “Small nucleolar RNAs, HACA box”                                  |
| Small nucleolar RNA host genes   | https://www.genenames.org/data/ | A curated list of snoRNA host genes, which is split into protein coding and non-coding           |
|                                  | negroup/#!/group/1838          | subgroups                                                                                       |
| Ribosomal RNAs                   | https://www.genenames.org/data/ | Starting page for all ribosomal RNAs, split into the major subgroups “Mitochondrially encoded   |
|                                  | negroup/#!/group/848           | ribosomal RNAs” and “Cytoplasmic ribosomal RNAs”, which is further split into                     |
|                                  |                                 | subtypes of rRNAs                                                                               |
| Vault RNAs                       | https://www.genenames.org/data/ | Full list of vault RNA genes                                                                      |
|                                  | negroup/#!/group/852           |                                                                                                  |
| Y RNAs                           | https://www.genenames.org/data/ | Full list of Y RNA genes                                                                          |
|                                  | negroup/#!/group/853           |                                                                                                  |
| Small NF90 (ILF3) associated RNAs| https://www.genenames.org/data/ | Full list of SNAR genes                                                                           |
|                                  | negroup/#!/group/1624          |                                                                                                  |
| Long non-coding RNAs             | https://www.genenames.org/data/ | Starting page for all long non-coding RNA gene. Divided into subgroups: Long intergenic          |
|                                  | negroup/#!/group/788           | non-protein coding RNAs, MicroRNA non-coding host genes, Overlapping transcripts,                |
|                                  |                                 | Intronic transcripts, Antisense RNAs, Divergent transcripts, Small nucleolar RNA non-coding      |
|                                  |                                 | host genes, Long non-coding RNAs with non-systematic symbols, Long non-coding RNAs with           |
|                                  |                                 | FAM root symbol                                                                                  |

© 2020 The Authors The EMBO Journal 39: e103777 | 2020 3 of 18
paralogous relationships between human miRNA genes, based on the family groups at miRBase and information in publications. For example, the “MicroRNA MIR1/206 family” contains the family members MIR1-1, MIR1-2 and MIR206. The miRNA symbol miR-206 has already been used in over 600 papers so it would be unhelpful to try to alter this symbol. However, the MIR206 Symbol Report now provides a link to the curated MicroRNA MIR1/206 family gene group page, where there are also associated publications and a link through to the corresponding miRBase Family MIPF0000038 page, which lists orthologous and paralogous miRNAs in different species. Where possible, the miRNA Symbol Reports on genenames.org also display the mouse and rat miRNA orthologs, with links to the relevant gene report on the Mouse Genomic Database (http://www.informatics.jax.org/) and Rat Genome Database (https://rgd.mcw.edu/), see Box 2.

Transfer RNAs

Transfer RNA was the first type of non-coding RNA to be characterised over 60 years ago (Hoagland et al., 1958). The term “transfer” (Smith et al., 1959) represents the function of this RNA in transferring amino acids from the cytosol of the cell to the ribosome where the amino acids are bonded together to form a peptide according to the sequence of the mRNA being translated. Typical tRNAs vary in size from 73 to 93 nucleotides (Rich & RajBhandary, 1976) and have a distinctive cloverleaf secondary structure that folds into an L-shaped tertiary structure (Kim et al., 1973). At one end of the L is the CCA acceptor site where the tRNA binds to the relevant amino acid (Hou, 2010) and at the other end is a loop that contains the three-nucleotide anticodon which precisely pairs to the codons of mRNA (Kim et al., 1973). The first two nucleotides of the anticodon form Watson-Crick base pairs with the corresponding mRNA codon, while the third nucleotide can form “wobble” pairing which allows one tRNA to recognise more than one mRNA codon. Post-transcriptional modifications at the “wobble” position can influence binding to a particular mRNA codon (Agris et al., 2018).

Transfer RNA genes share characteristics that make it possible to predict them from genomic sequence. The Genomic tRNA Database (GtRNAdb) (Chan & Lowe, 2016) contains predicted tRNA gene sets for thousands of species across Eukaryota, Archaea and Bacteria, including a set of 429 high confidence tRNA genes for the most current human reference genome, GRCh38. tRNA gene predictions are made using the tRNAscan-SE analysis pipeline (Lowe & Chan, 2016), which uses probabilistic tRNA primary sequence and secondary structure “covariance models” to determine the gene loci and the functional identity (i.e. tRNA isotype and anticodon) for each putative tRNA gene. The predicted tRNA genes then undergo further analysis by comparison with isotype-specific covariance models to give confirmation of isotype classification. The GtRNAdb assigns a unique ID to each tRNA gene in the format tRNA-[three letter amino acid code]-[anticodon]-[GtRNAdb gene identifier], e.g. tRNA-Ala-AGC-1-1. (Note the “GtRNAdb gene identifier” is actually made up of two numbers, the first is a “transcript ID”, the second a “locus ID”, such that multiple gene loci producing identical tRNA transcripts share the same transcript ID, but each have a different locus numbers; e.g., Ala-AGC-1-1 and Ala-AGC-1-2 are two different gene loci producing identical mature tRNAs, whereas Ala-AGC-2-1 and Ala-AGC-3-1 are genes that each produce different tRNA transcripts.) The HGNC assigns a slightly condensed but equivalent tRNA gene symbol in the format tRNA-[three letter amino acid code]-[anticodon], e.g. TRA-AGC1-1 (Fig 3). tRNAscan-SE analysis also predicts tRNA pseudogenes and candidate genes that include atypical tRNA features and may not be transcribed.
and/or may not be capable of ribosomal translation. To reflect these different sets, the HGNC displays the gene groups “Cytosolic transfer RNAs”, “Low confidence cytosolic transfer” RNAs and “Transfer RNA pseudogenes on genenames.org” (Table 1).

The human mitochondrial genome contains 22 tRNA genes (Anderson et al., 1981) that encode tRNAs with both canonical and non-canonical cloverleaf structures which enable translation within mitochondrial ribosomes in the mitochondria. While pathological mutations in cytosolic tRNA genes have not yet been discovered, mutations in mitochondrial tRNA genes cause a variety of well-studied mitochondrial diseases such as MELAS (mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes) and MERRF (myoclonic epilepsy with ragged red fibres) (Suzuki & Nagao, 2011; Abbott et al., 2014). Mitochondrial tRNA genes were named in collaboration with the MitoMap resource (Lott et al., 2013); gene symbols are of the format “MT-T + one letter amino acid code”; e.g., MT- TA represents the mitochondrial tRNA gene that recruits alanine. Most amino acids are decoded by just one human mitochondrial tRNA, but there are two mitochondrial leucine and serine tRNA genes—these gene symbols therefore include numbers to distinguish the individual loci: MT-TL1, MT-TL2, MT-TS1 and MT-TS2.

Small nuclear RNAs

Small nuclear RNAs are abundant transcripts of around 150 nucleotides that end in a 3’ stem loop (Matera et al., 2007). While the name of this RNA class is based on cellular location, each individual snRNA has a “U” identifier that stems from the historical name “U-RNA” which was derived from early observations of their high uridine content (Hodnett & Busch, 1968). The U-RNAs were numbered according to their apparent abundance when discovered (Chen & Moore, 2015). Some of these were subsequently found to be small nucleolar RNAs (snoRNAs) resulting in the following numbering for the snRNAs: U1, U2, U4, U5, U6, U7, U11 and U12.

Most snRNAs are involved in the splicing of introns from premRNA as part of either the major or minor spliceosome. The major spliceosome features U1, U2, U4, U5 and U6 snRNPs, plus many other non-snRNP proteins, and performs splicing of U2-type introns.
Figure 3. An annotated tRNA gene symbol explaining what each part of the approved gene symbol represents.

Here, the U1 and U2 snRNPs assemble on introns and are joined by the preassembled U4/U6.U5 tri-snRNP. This is followed by a series of rearrangements resulting in the formation of the U2/U6 catalytic core and the splicing reaction (Anokhina et al., 2013), and finally release of the spliced RNA and disassembly of the spliceosome. The minor spliceosome splices U12-type introns, which make up < 0.5% of introns in the genome (Turunen et al., 2013). It contains the same U5 snRNA as the major spliceosome, but in contrast consists of the snRNAs U11, U12, U4atac and U6atac, which are functional analogs of the major spliceosome U1, U2, U4 and U6 snRNAs. Minor spliceosome snRNAs can fold into similar structures to their equivalent major spliceosome snRNAs, but display limited sequence similarity to them (Will & Lührmann, 2005). The term “atac” in U4atac and U6atac refers to the AT/AC splice sites found in the first U12-type introns to be discovered (Tarn & Steitz, 1996). Instead of splicing, the variant U1 small nuclear”. The snRNA vU1.8, encoded by RNVU1-8, has been shown to be capable of processing the 3’ end of premRNAs expressed from a subset of target genes (O’Reilly et al., 2013). Moreover, snRNAs encoded by RNVU1-3, RNVU1-8 and RNVU1-20 are implicated in stem cell maintenance and neuromuscular disease (Vazquez-Arango et al., 2016).

**SnorRNAs**

Small nucleolar RNAs are transcripts of around 60–170 nucleotides that can be divided into three major classes: C/D box snorRNAs...
(SNORDs), H/ACA box snoRNAs (SNORAs) and small Cajal body-specific RNAs (scARNAs). Although some are transcribed from independent promoters, most snoRNAs are encoded within the introns of either protein coding or long non-coding “host” genes (see Table 1 for details on accessing gene groups listing these). C/D box snoRNAs are named after their two conserved box motifs: C (sequence: RUGAUGA) and D (sequence: CUGA) (Tyc & Steitz, 1989); these snoRNAs primarily function in the nucleolus within small nucleolar ribonucleoprotein (snoRNP) complexes to direct target site-specific 2'-O-methylation of rRNAs (Kiss-László et al., 1996). H/ACA box snoRNAs share a common secondary structure and contain the AnAnnRNA sequence known as the “hinge” or “H” box and the trinucleotide “ACA” box (Ganot et al., 1997a, 1997b). H/ACA snoRNAs also function with snoRNP complexes in the nucleolus to guide modification of rRNAs, but in this case the modification is pseudouridylation (Kiss et al., 1997a, 1997b). Small Cajal body-specific RNAs function in the Cajal body, a nuclear organelle named after its discoverer Santiago Ramón y Cajal (Gall et al., 1999). ScARNAs contain either H/ACA boxes, C/D boxes or a mixture of both types, and function as guides for the same type of RNA modifications as the nucleolar snoRNAs—guiding RNP complexes to catalyse pseudouridylation or 2'-O-methylation—but for modification of snRNAs instead of rRNAs. The major difference in sequence between scARNAs and snoRNAs is thought to be the presence of Cajal body targeting sequences, the CAB box in H/ACA scARNAs (Richard et al., 2003) or the G.U/UG wobble stems in C/D scARNAs (Marnef et al., 2014). Some scARNAs show no sequence complementarity to either rRNAs or snRNAs, suggesting they have an alternative function to the canonical scARNAs described above. For example, there have been recent reports of scARNAs involved in diverse functions such as activation of enzymes, or regulation of alternative splicing and mRNA levels (Falaleeva et al., 2017).

When snoRNAs were first discovered, they were initially not distinguished from other snRNAs and were therefore assigned “U” numbers, e.g. U3, U8 and U13 (Tyc & Steitz, 1989), which are the identifiers still in use for snRNAs (see small nuclear RNAs section above). Once the H/ACA and C/D boxes were identified, a convention of using the root ACA# (Kiss et al., 2004) or HB-I# for human H/ACA box snoRNAs and HB-II# for C/D box snoRNAs (Cavaillé et al., 2000) was established, which then formed a “rival” nomenclature to the U# system that was still in use. Originally, scARNAs were not discernible from other snoRNAs by symbol; e.g., the first identified scARNA was referred to as U85 (Jády & Kiss, 2001) and another as ACA26 (Tycowski et al., 2009). In 2007, the HGNC worked with snoRNABase (Lestrade & Weber, 2006) to devise a standardised, easily recognisable nomenclature for all three types of snoRNA: SNORD# for “small nucleolar RNA, C/D box” genes; SNORA# for “small nucleolar RNA, H/ACA box” genes; and SCARNA# for “small Cajal body-specific RNA” genes. Unfortunately, the snoRNABase resource, although still valuable, is no longer being updated. The HGNC now works with the Stadler Bioinformatics Leipzig group to assign symbols to newly identified snoRNA genes (Vjorjani et al., 2016), and as such, the HGNC snoRNA gene group pages (Table 1) provide an up-to-date list of canonical human snoRNA and scARNA genes.

A potential issue for nomenclature is that snoRNAs and scARNAs cannot always be distinguished unambiguously without evidence of localisation. Thus, ncRNAs of these classes are by default named as SNORD# or SNORD# unless evidence for Cajal body specificity is available. Some snoRNAs are a source of miRNA-like small RNAs; in a few cases, these small RNAs function in post-transcriptional gene silencing like miRNAs (Scott & Ono, 2011). Interestingly, H/ACA snoRNAs are processed by Dicer, while small RNAs derived from box C/D snoRNA appear to use a different processing pathway (Langenberger et al., 2013). At present, HGNC does not provide a nomenclature for the small RNAs derived from snoRNA and scARNAs.

Ribosomal RNAs

The ribosome is responsible for the synthesis of peptides using mRNA as a template. The term “ribosome” was coined by Richard B. Roberts to provide a more user-friendly version of “ribonucleoprotein particles of the microsome fraction” (Roberts, 1958). The ribosome, its subunits and rRNAs have all been assigned unique identifiers in Svedberg units based on their sedimentation rate in a centrifuge—the eukaryotic ribosome is referred to as the 80S ribosome and comprises a large (60S) subunit that contains 28S, 5S and 5.8S rRNA and a small (40S) subunit that contains 18S rRNA. Both subunits also contain a large number of ribosomal proteins (Khattar et al., 2015). The 28S rRNA forms the core of the large subunit and contains the catalytic peptidyl transferase centre (Polacek & Mankin, 2005) that forms bonds between amino acids to create peptides, meaning that the ribosome is also a ribozyme. 5S rRNA is necessary for translation (Ciganda & Williams, 2011) although its exact role is unclear, while 5.8S rRNA appears to have a role in ribosome translocation (Abou Elela & Nazar, 1997). 18S rRNA is at the core of the small subunit and binds directly to mRNA during translation initiation (Martin et al., 2016) and translation elongation (Tranque et al., 1998; Demeshkina et al., 2000).

Cytoplasmic rRNAs are transcribed from multicopy gene clusters (Fig. 4)—the 5S rRNA cluster on chromosome 1q42.13 (Sørensen & Frederiksen, 1991) transcribed by RNA polymerase III and the 45S rRNA clusters that encode 18S, 5.8S and 28S rRNA on the p arms of the five human acrocentric chromosomes in the cytogenetically visible nucleolar organising regions (NORs) (Gonzalez & Sylvester, 2001) transcribed by RNA polymerase I. There is great variation in the number of rRNA repeats within all of these clusters both within and between different individuals. The 5S cluster is the only one in which individual genes have been annotated on the current GRCh38 human reference genome, although currently there are only 17 annotated 5S rRNA genes, while the average individual has around 98 SS genes (Stults et al., 2008). The HGNC has named the 17 annotated genes RNASS1-RNASS17. The 45S rRNA genes have a highly repetitive nature, which has made accurate sequence assembly difficult, and as a result, no individual 45S rRNA gene is present within the NORs on the GRCh38 reference genome. The number of 45S rRNA genes per cluster differs between individuals and varies from a single gene to more than 140 repeated genes, which are usually arranged in a head-to-tail orientation (Stults et al., 2008). The HGNC has approved a gene symbol for each of the acrocentric 45S rRNA clusters: RNR1 (13p12), RNR2 (14p12), RNR3 (15p12), RNR4 (21p12) and RNR5 (22p12; Fig 4). The 45S rRNA repeats are post-transcriptionally processed into the r18S, 5.8S and 28S by a series of cleavage events. The HGNC has reserved the stem symbols
Y RNAs

Y RNAs are small transcripts of ~100 nucleotides with distinctive secondary structures that are largely bound by the Ro60 protein, which is similar in structure to the TEP1 protein that binds vault RNAs (Bateman & Kickhoefer, 2003), hinting at an evolutionary relationship between these two classes of ncRNAs. These RNAs were first identified in RNP complexes that were immunoprecipitated with anti-Ro60 antibodies from patients with systemic lupus erythematosus (Hendrick et al., 1981; Lerner et al., 1981) and were designated “Y” RNAs because they are mostly cytoplasmic, in contrast to the U class of small nuclear RNAs (Lerner et al., 1981).

The human genome encodes 4 active Y RNA genes, which are all located on 7q36.1 and are transcribed by RNA polymerase III (Wolin & Steitz, 1983; Marra et al., 1994, 1996). While the transcripts are referred to as Y1, Y3, Y4, and Y5, the equivalent approved gene symbols are RNY1, RNY3, RNY4 and RNY5 for “RNA, Ro60-associated Y#”. Note there is no Y2 as this symbol was used for a short transcript that was subsequently found to be a truncated form of Y1 (Hendrick et al., 1981; Wolin & Steitz, 1983).

All Y RNAs contain a stem, formed by base pairing of the 5' and 3' ends, that includes the Ro60 binding site (Wolin & Steitz, 1984; Pruijn et al., 1991; Green et al., 1998). At the other end of this stem are one or more internal loops and stem loops that interact with other proteins to generate specialised RNPs (Sim et al., 2012; Chen et al., 2013). Y RNAs can influence the subcellular location of Ro60 (Sim et al., 2009, 2012) and may regulate the ability of Ro60 to bind misfolded RNAs (Stein et al., 2005; Fuchs et al., 2006; Wolin et al., 2013), a function supported by work in bacteria (Chen et al., 2007, 2013). There have also been reports of a Ro60-independent function for mammalian Y RNAs in DNA replication (Christov et al., 2006; Krude et al., 2009), although mouse cell lines depleted of Y RNAs show no growth defects (Sim et al., 2009, 2012; Reed et al., 2013).

SNARs

Small NF90 (ILF3)-associated RNAs (snaRs) were first identified following immunoprecipitation of ribonucleoproteins with antibodies against NF90, an abundant protein isoform expressed from the ILF3 gene (Parrott & Mathews, 2007). The snaR transcripts are around 117 nucleotides, show highest expression in immortalised cell lines and tests and are transcribed by RNA polymerase III (Parrott & Mathews, 2007). snaR genes are specific to great apes and evolved from an Alu repeat element followed by genomic duplication (Parrott & Mathews, 2009). Bioinformatic analysis identified nine subsets of snaRs based on sequence similarity and the HGNC...
agreed on the root symbol SNAR, for “small NF90 (ILF3)-associated RNA”, followed by a unique letter for each subset and a unique number for each gene in a subset, e.g. SNAR-A1, SNAR-R2 and SNAR-C3. SnARs are the least well-characterised category of small RNAs named by the HGNC and their function remains to be determined but snAR-A transcripts bind to ribosomes, suggesting these RNAs could have a role in translational control (Parrott & Mathews, 2009).

Long non-coding RNAs

Before the human genome was sequenced, a small number of functional non-coding transcripts had been identified that could not be placed into any of the categories described so far in this paper: 7SK (encoded by RN7SK) (Zieve & Penman, 1976; Diribarne & Bensaude, 2009) and 7SL (encoded by three human loci: RNA7SL1, RN7SL2 and RN7SL3) (Walker et al., 1974; Walter & Blobel, 1982) in the 1970s, and H19 (Brannan et al., 1990), BCYRN1 (BC200) (Tiedge et al., 1993), and XIST (Brown et al., 1991, 1992) in the early 1990s. 7SK, 7SL and BCYRN1 are all transcribed by RNA polymerase III and function via forming complexes with proteins: 7SK is an RNA scaffold in a complex that regulates the P-TEFb transcription factor (Diribarne & Bensaude, 2009); 7SL is the RNA component of the signal recognition particle that targets proteins with a signal peptide to the endoplasmic reticulum (Walter & Blobel, 1982); BCYRN1 inhibits translation via binding to eIF4A and PABP (Muddashetty et al., 2002; Lin et al., 2008). In contrast, H19 and XIST, like protein coding transcripts, are transcribed by RNA polymerase II. While XIST has a defined molecular function in binding to and silencing the inactive X chromosome (Chow et al., 2005), the exact molecular function of H19 is still not clear—it has been associated with many types of cancer and regulates several target genes by post-transcriptional mechanisms (Gabory et al., 2010).

Large-scale studies made possible following the release of the sequenced human genome in 2001 (Lander et al., 2001) revealed the existence of large numbers of transcripts that appear to be untranscribed and, like those above, do not belong to previously defined classes of non-coding RNAs (Kapranov et al., 2002; Bertone et al., 2004; Cheng et al., 2005). These were initially referred to as mRNA-like ncRNAs because they are generally transcribed by RNA polymerase II, and are capped, spliced and polyadenylated like protein-coding mRNAs (Erdmann et al., 1999; Lottin et al., 2002; Bomfmunenerer et al., 2005; Széll et al., 2008). A 2007 study on non-coding transcripts in human muscle first used the term “long” to refer to transcripts of over 200 nucleotides (Kapranov et al., 2007), and this classification became widespread with the term “long non-coding RNA” (or “long non-coding RNA”) appearing in the title of 18 papers in a 2010 PubMed search, increasing to 123 papers by the year 2013 and 1,517 papers in 2018. Although the term “long non-coding RNA” (abbreviated to IncRNA) does not truly represent a class of non-coding RNA, it has become a useful shorthand for such transcripts of varied/unknown function and is entrenched in the scientific literature.

Functional studies have been performed for a relatively small subset of IncRNAs. The modes of action that have been described can be grouped into several different categories (see Chen, 2016) for a comprehensive review of IncRNA by category):
human. Although the HGNC discourages the use of punctuation in gene symbols, hyphens may sometimes be used in lncRNA gene symbols. Where known, the gene name should represent the normal function of the lncRNA gene; e.g., the full name of NEAT1 is “nuclear paraspeckle assembly transcript 1”, and the full name of NRON is “non-coding repressor of NFAT”. We appreciate that for lncRNA genes this is not always possible, and we do permit reference to expression, e.g. BMNCR for “bone marrow associated non-coding RNA” (Li et al, 2018a), and, in some cases, disease where the association of the lncRNA with the disease is based on more than a change in expression, e.g. PRINS for “psoriasis associated non-protein coding RNA induced by stress” (Sonkoly et al, 2005), NBAT1 for “neuroblastoma associated transcript 1” (Pandey et al, 2014). Again, such cases should be discussed individually with the HGNC prior to publication.

In addition to providing a unique symbol for each named lncRNA gene, the HGNC records other alternative symbols used by different research groups, which we refer to as alias symbols. For example, the lncRNA gene LINC00261 was first approved by the HGNC in 2012; this unique symbol first appeared in a publication in 2013 (Cao et al, 2013) and has since appeared in more than 25 publications. A different symbol, ALIEN, was used in a 2015 publication (Kurian et al, 2015) with no reference to the approved symbol and the symbol DEANR1 appeared the same year (Jiang et al, 2015) with reference to the approved symbol in the paper but not in the title or abstract. Using or referencing the approved symbols in the title or abstract allows all papers on a particular gene to be found easily and ensures that key information will not be missed. The HGNC symbol report for LINC00261 shows all alias symbols so that searching our database with any of the published symbols will retrieve the correct gene in our database and in other major biomedical databases such as NCBI Gene and Ensembl. Although the HGNC endeavours to record alias symbols, there is always the possibility that these may be missed and valuable data on genes lost to future interested parties if the approved symbol is not referenced anywhere else in the publication.

Where possible, we coordinate with the Mouse Genomic Nomenclature Committee to assign the equivalent symbol for orthologous human and mouse lncRNA genes. For example, the mouse ortholog of the human lncRNA gene NEAT1 has the symbol Neat1, while the

| Human symbol | Human gene name | Mouse symbol | References |
|--------------|----------------|--------------|------------|
| AIRN         | antisense of IGF2R non-protein coding RNA | Airn | Yotova et al (2008) |
| DANCR        | Differentiation antagonising non-protein coding RNA | Dncr | Chalei et al (2014) |
| EOT          | Eosinophil granule ontology transcript | Eot | Rose and Stadler (2011) |
| FENDRR       | FOXF1 adjacent non-coding developmental regulatory RNA | Fendrr | Grote et al (2013) |
| FIRRE        | Firre intergenic repeating RNA element | Firre | Hacisuleyman et al (2014) |
| GASS         | Growth arrest specific 5 | Gas5 | Smith and Steitz (1998) |
| HOTAIR       | HOX transcript antisense RNA | Hotair | Rinn et al (2007) |
| HOTITP       | HOXA distal transcript antisense RNA | Hotitp | Wang et al (2011) |
| KCNQ1OT1     | KCNQ1 opposite strand/antisense transcript 1 | Kcng1ot1 | Gicquel et al (2003) |
| MALAT1       | Metastasis associated lung adenocarcinoma transcript 1 | Malat1 | Ji et al (2003) |
| MEG3         | Maternally expressed 3 | Meg3 | Miyoshi et al (2000) |
| MIR17HG      | mir-17-92a-1 cluster host gene | Mir17hg | Dews et al (2010) |
| NEAT1        | Nuclear paraspeckle assembly transcript 1 | Neat1 | Clemson et al (2009) |
| NRON         | Non-coding repressor of NFAT | Nron | Willingham et al (2005) |
| PANTR1       | POU3F3 adjacent non-coding transcript 1 | Pantr1 | Clark and Blackshaw (2017) |
| PAUPAR       | PAX6 upstream antisense RNA | Paupar | Vance et al (2014) |
| PEG13        | Paternally expressed 13 | Peg13 | Court et al (2014) |
| PVT1         | PVT1 oncogene | Pvt1 | Carramusa et al (2007) |
| RN7SK        | RNA component of 7SK nuclear ribonucleoprotein | Rn7sk | Driscoll et al (1994) |
| SNHG3        | Small nucleolar RNA host gene 3 | Snhg3 | Pelczar and Filipowicz (1998) |
| SOX1-OT      | SOX1 overlapping transcript | Sox1ot | Ahmad et al (2017) |
| TSIX         | TSIX transcript, XIST antisense RNA | Tsix | Lee et al (1999) |
| TUG1         | Taurine up-regulated 1 | Tug1 | Young et al (2005) |
| TUNAR        | TCL1 upstream neural differentiation-associated RNA | Tunar | Ulitsky et al (2011) |
| XIST         | X inactive specific transcript | Xist | Brown et al (1991) |
| ZFAS1        | ZNFX1 antisense RNA 1 | Zfas1 | Askarian-Amiri et al (2011) |
Box 3

1 Intergenic lncRNA genes are assigned the root symbol “LINC” for “long intergenic non-protein coding RNA” followed by a unique 5-digit number, e.g. LINC01018 (Fig 5A).

A lncRNA gene is considered intergenic (meaning between protein coding genes in this context) if it does not overlap a protein coding gene on either strand, does not share a bidirectional promoter with a protein coding gene and is not a host gene for a microRNA or snoRNA.

2 Antisense lncRNA genes are named using the format [protein coding gene symbol] with the suffix -AS and a sequential number, e.g. FAS-AS1 for “FAS antisense RNA 1” (Fig 5B).

A lncRNA gene is considered antisense if it overlaps the genomic coordinates of a protein coding gene on the opposite strand. There does not need to be exon–exon overlap. These symbols are not intended to imply that there is a regulatory role between the protein coding and lncRNA gene. If the lncRNA is antisense to more than one protein coding gene, the symbol of the most 5’ protein coding gene will be chosen as the basis of the lncRNA gene symbol, unless there is exon–exon overlap between a more 3’ protein coding gene, which would be chosen in preference.

3 Divergent transcripts that are transcribed from a bidirectional promoter in the opposite direction to a protein coding gene are named using the format [protein coding gene symbol] with the suffix -DT, e.g. ABCF1-DT for “ABCF1 divergent transcript” (Fig 5C).

A lncRNA is considered divergent if it is within 300–500 nucleotides of the 5’ end of a protein coding gene on the other strand. Usually evidence of bidirectional transcription can be seen with cap analysis gene expression tags, although this is not a requirement. If a protein coding gene has multiple transcription start sites, the lncRNA will be named as a divergent transcript only if it shares the 5’ most promoter; otherwise, it will overlap the genomic span of the protein coding gene and be considered antisense.

4 Intronic transcripts that are transcribed entirely from within an intron of a protein coding gene on the same strand are named using the format [protein coding gene symbol] with the suffix -IT1 for “AOAH intronic transcript” (Fig 5D).

This category accounts for a small number of our named lncRNA genes and is applied sparingly because we have found that future evidence may reveal that these loci are alternative exons or rare intron degradation intermediates of the protein coding locus.

5 Overlapping transcripts that overlap a protein coding gene on the same strand are named using the format [protein coding gene symbol] with the suffix -OT and a sequential number, e.g. C5-OT1 for “C5 3’ UTR overlapping transcript 1” (Fig 5E).

As for the intronic transcripts above, this category is applied with caution because experience has shown that such lncRNA genes may eventually be merged into the protein coding locus when further transcriptional evidence becomes available.

6 Host genes for microRNAs or snoRNAs. The small RNA may be in an exon or intron but must be on the same strand as the lncRNA (Fig 5F).

LncRNA genes that host a microRNA gene are named using the format [microRNA gene symbol]-HG, e.g. MIR122HG for “MIR122 host gene”. Where there are several microRNA genes hosted by the same lncRNA gene, the lncRNA is named after the 5’ most microRNA. If the lncRNA gene hosts a cluster, this is shown in the gene name; e.g., MIR17HG has the full gene name “mir-17-92a-1 cluster host gene”. MIR200CHG hosts the microRNA genes MIR200C and MIR141; this is shown in the full gene name “MIR200C and MIR141 host gene”.

mouse ortholog of human XIST has the symbol Xist (see Table 2 for further selected examples). However, it is not always straightforward to determine orthology between human and mouse lncRNA genes (Ulitsky, 2016). We require the two genes to be at a conserved syntenic location and to have detectable sequence similarity.

As mentioned above, a relatively small fraction of the predicted total number of lncRNA genes have been cited in publications. In addition to naming published lncRNA genes, the HGNC names genes that have been annotated by the RefSeq (O’Leary et al, 2016) and GENCODE (Frankish et al, 2019) projects. These projects initially annotated lncRNA genes based on EST, cDNA and mRNA data, which provided a set of relatively high stringency, but not necessarily full-length, transcripts. Both projects have since started to incorporate long read RNA-Seq data, e.g. (Lagarde et al, 2017). Genes are annotated as lncRNAs where there is sufficient transcriptional support for a locus, but there is not sufficient evidence of protein coding potential. Assessment of protein coding potential includes assessing cross-species conservation of a putative open reading frame (ORF), length of a putative ORF, presence/absence of encoded features such as protein domains, ribosome profiling data and evidence of peptides via mass spectrometry. Due to the constant emergence of new data, there is a certain amount of flux between the protein coding and lncRNA gene sets. Some protein coding genes have subsequently been reannotated as lncRNA genes; e.g., the gene formerly known as C6orf48 was reannotated as an lncRNA gene and therefore renamed by the HGNC as SNHG32. Equally, some lncRNA genes have been reannotated as protein coding genes either due to a reassessment based on new metrics such as phyloCSF (Lin et al, 2011) or based on emerging evidence from new publications. For example, LINCO00383 was reannotated as protein coding gene CLEC20A because the ORF is conserved and exhibits a C-type lectin domain, while LINCO1420 was reannotated as protein coding and renamed NBD1 based on published data (D’Lima et al, 2017).

Feedback from conferences and research groups informed us that the lncRNA community finds genomic context with respect to protein coding genes a useful metric when considering lncRNA genes on a genomic scale. Therefore, working with the lncRNA annotation classification used by the GENCODE group, we devised a nomenclature system using the following categories (see Fig 5):

LncRNA genes that host a snoRNA gene are named using the root symbol SNHG for snoRNAs host gene followed by a unique number, e.g. SNHG1. This lncRNA hosts seven different snoRNA genes, so an early decision was taken to not include reference to individual snoRNA genes at the gene symbol level.

Please see the previous sections on snoRNA and microRNA genes above for more information on these small RNAs and their host genes.

In future, the HGNC will explore the possible annotation and naming of sno-lncRNAs, a new class of transcript with a snoRNA at each end (Yin et al, 2012; Xing et al, 2017). These are processed from the introns of snoRNA host genes that host more than one snoRNA within an intron. We will also explore transcripts derived from snoRNA host genes that have a 5’ snoRNA and a poly(A) tail, which have been referred to as 5’ snoRNA capped and 3’ polyadenylated (SPAs) (Wu et al, 2016; Lykke-Andersen et al, 2018).

The HGNC names genes and not alternative transcripts, so we assign only one name per lncRNA gene and do not provide separate symbols for non-coding transcripts that are part of protein coding loci. Please note that the symbols in the above scheme do not mean that the lncRNA genes they represent have no function—the
symbols are systematically applied where no other informative data are available at the time of naming. The HGNC will only change such symbols once future information is available where there is a consensus from groups working on these genes to do so. In some cases, our systematic symbols are already becoming well used in the literature, e.g. LINC00473, MIR17HG, LOXL1-AS1. We have already named over 4,300 lncRNA genes, but we are still a long way from naming all annotated lncRNA genes; we are currently working on naming a dataset of intergenic lncRNA genes that are consistently annotated by both the GENCODE and RefSeq projects.

Circular RNAs

Circular RNAs (circRNAs) and circular intronic RNAs (ciRNAs) are both produced during the splicing of pre-mRNA—the major difference being that circRNAs are derived from exonic sequence, while ciRNAs are derived from intronic sequence. Currently, there are no approved symbols for circRNAs or ciRNAs; this may be a future task for the HGNC if a consensus is found in the community. CircRNAs are the result of back-splicing of exons from pre-mRNA, which creates a circRNA joined to itself by a 3',5'-phosphodiester bond (Wu et al., 2017; Li et al., 2018b). Although most of these RNAs are expressed at low levels, there are examples where the circRNA is more highly expressed than the spliced linear mRNA (Salzman et al., 2013). Recent studies have suggested roles for circRNAs in competitive regulation of pre-mRNA splicing (Ashwal-Fluss et al., 2014; Zhang et al., 2014), competitive binding to microRNAs (Hansen et al., 2013; Memczak et al., 2013), regulation of RNA polymerase II (Li et al., 2015) and involvement in innate immunity (Liu et al., 2019a). CiRNAs are derived from spliced-out intron lariats that have escaped cleavage by the debranching enzyme. These RNAs have 2',5'-phosphodiester bonds between their 5' ends and the intronic branching site creating a circular structure. Sequence analysis shows that the generation of ciRNAs is not random but depends on the presence of a consensus RNA motif containing a seven nucleotide GU-rich element near the 5' splice site and an 11 nucleotide C-rich element near the intron branch point of the parent mRNA (Zhang et al., 2013). Knockdown of ciRNAs has been shown to reduce expression of the genes from which they are derived (Zhang et al., 2013).

While there is no current standardised system for naming circRNAs or ciRNAs, we suggest the following nomenclature schemes:

A **Intergenic**

| Protein-coding gene | lncRNA gene | snoRNA or microRNA gene |
|----------------------|-------------|-------------------------|
| UBE2QL1              | LINC01018   | NSUN2                   |

B **Antisense**

| Protein-coding gene | lncRNA gene |
|----------------------|-------------|
| FAS                  | FAS-AS1     |

C **Divergent**

| Protein-coding gene | lncRNA gene |
|----------------------|-------------|
| ABCF1                | ABCF1-DT    |

D **Intronic**

| Protein-coding gene | lncRNA gene |
|----------------------|-------------|
| AOAH                | AOAH-IT1    |

E **Overlapping**

| Protein-coding gene | lncRNA gene |
|----------------------|-------------|
| C5                   | C5-OT1      |

F **RNA host gene**

| Protein-coding gene | lncRNA gene | snoRNA or microRNA gene |
|----------------------|-------------|-------------------------|
| SNHG7                | SNORA17A    | SNORA17B                |

Figure 5. LncRNA naming schema for lncRNA genes with no published information at the time of naming.

A LncRNAs that are intergenic with respect to protein coding genes are assigned the root symbol LINC# followed by a 5-digit number.
B LncRNAs that are antisense to the genomic span of a protein coding gene are assigned the symbol format [protein coding gene symbol]-AS#.
C LncRNAs that are divergent to (share a bidirectional promoter with) a protein coding gene are assigned the symbol format [protein coding gene symbol]-DT.
D LncRNAs that are contained within an intron of a protein coding gene on the same strand are assigned the symbol format [protein coding gene symbol]-IT#.
E LncRNAs that overlap a protein coding gene on the same strand are assigned the symbol format [protein coding gene coding symbol]-OT#.
F LncRNAs that contain microRNA or snoRNA genes within introns or exons are named as host genes. See the main text for details on how these microRNA host genes and snoRNA host genes are named.
For circRNAs:

- circ[gene symbol]-n where the gene symbol represents the unspliced “host” gene and n is an iterative five digit number; e.g., the first circRNA named for the host gene PARN would be circPARN-00001

For ciRNAs:

- ci[gene symbol]-n where the gene symbol represents the unspliced “host” gene and n is an iterative five digit number; e.g., the first ciRNA named for the host gene PARN would be ciPARN-00001

There are currently huge numbers of circRNAs listed in public databases such as CIRCpedia (Dong et al, 2018), circBank (Liu et al, 2019b) and circBase (Glazar et al, 2014) all using different identifiers. We call on the community to come together to discuss standards creating a consensus set of circRNAs and ciRNAs that could be given standardised nomenclature in the future.

Conclusion

In summary, the HGNC works directly with specialist advisors in the ncRNA field to ensure that appropriate and informative gene symbols are approved for ncRNA genes. We urge all ncRNA researchers to use, or at least mention, HGNC-approved gene symbols. This will ensure that ncRNA genes are correctly cited and will prevent confusion in the field. To discuss any aspect of ncRNA nomenclature, please contact the HGNC via our email address hgnc@genenames.org.

Acknowledgements

We would like to thank Dr. Patricia Chan for her input on transfer RNA nomenclature and the HGNC team, both past and present, for their helpful discussions. This work was supported by the National Human Genome Research Institute (NHGRI) grant U24HG003345, Wellcome Trust grant 208349/Z/17/Z and Germany Ministry Education and Research grant no 031AS38A, de.NBI-RBC.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Abbott JA, Franklyn CS, Robey-Bond SM (2014) Transfer RNA and human disease. Front Genet 5: 158
Abou Elela S, Nazar RN (1997) Role of the 5.8S rRNA in ribosome translocation. *Nucleic Acids Res* 25: 1788 – 1794
Agris PF, Eryusal ER, Narendran A, Vare VYP, Vangaveti S, Ranganathan SV (2018) Celebrating wobble decoding: half a century and still much is new. *RNA Biol* 15: 537 – 553
Ahmad A, Strohbeuecker S, Tufarelli C, Sottile V (2017) Expression of a SOX1 overlapping transcript in neural differentiation and cancer models. *Cell Mol Life Sci* 74: 4245 – 4258
Akerman I, Tu Z, Beucher A, Rolando DNY, Sauty-Colace C, Benazra M, Nakic N, Yang J, Wang H, Pasquali L et al (2017) Human pancreatic β cell lncRNAs control cell-specific regulatory networks. *Cell Metab* 25: 400 – 411
Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M et al (2003) A uniform system for microRNA annotation. *RNA* 9: 277 – 279
Amort M, Nachbauer B, Tuzlak S, Kieser A, Schepers A, Villunger A, Polacek N (2015) Expression of the vault RNA protects cells from undergoing apoptosis. *Nat Commun* 6: 7030
Anderson S, Bankier AT, Bartell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F et al (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457 – 465
Anokhina M, Bessonov S, Miao Z, Westhof E, Hartmuth K, Lührmann R (2013) RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J* 32: 2804 – 2818
Arab K, Park YJ, Lindroth AM, Schäfer A, Oakes C, Weichenhan D, Lukanova A, Lundin E, Risch A, Meister M et al (2014) Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GAD64A.* Mol Cell* 55: 604 – 614
Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evanth N, Memczak S, Rajewsky N, Kadener S (2014) circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 56: 55 – 66
Askarian-Amiri ME, Crawford J, French JD, Smart CE, Smith MA, Clark MB, Ru K, Mercer TR, Thompson ER, Lkhani SR et al (2011) SNORD-host RNA Zfas is a regulator of mammalian development and a potential marker for breast cancer. *RNA* 17: 878 – 891
Bartel DP (2018) Metazoan MicroRNAs. *Cell* 173: 20 – 51
Bateman A, Kickhoefer V (2003) The TROVE module: a common element in Telomerase, Ro and Vault ribonucleoproteins. *BMCI Bioinformatics* 4: 49
Bertone P, Stolic V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S et al (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science* 306: 2242 – 2246
Bompineuwer AF, Flamme C, Fried C, Fritzsch G, Hofacker IL, Lehmann J, Missal K, Mosig A, Müller B, Prohaska SJ et al (2005) Evolutionary patterns of non-coding RNAs. *Theory Biosci* 123: 301 – 369
Brannan CI, Dees EC, Ingram RS, Tlghman SM (1990) The product of the H19 gene may function as an RNA. *Mol Cell Biol* 10: 28 – 36
Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedle S, Yates B, Bruford E (2013) Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res* 47: D786 – D792
Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349: 38 – 44
Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71: 527 – 542
Budak H, Bulut R, Kantar M, Alptekin B (2016) MicroRNA nomenclature and the need for a revised naming prescription. *Brief Funct Genomics* 15: 65 – 71
Cao WJ, Wu HL, He BS, Zhang YS, Zhang ZY (2013) Analysis of long non-coding RNA expression profiles in gastric cancer. *World J Gastroenterol* 19: 3658 – 3664
Carramus L, Contino F, Ferro A, Minafra L, Perconti G, Giallongo A, Feo S (2007) The PVT-1 oncogene is a Myc protein target that is overexpressed in transformed cells. *J Cell Physiol* 213: 511 – 518
Cavaillé J, Buiting K, Kieffmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc Natl Acad Sci USA 97: 14311–14316

Cavdar Koc E, Burkhart W, Blackburn K, Moseley A, Spremulli LL (2001) The small subunit of the mammalian mitochondrial ribosome. Identification of the full complement of ribosomal proteins present. J Biol Chem 276: 19363 – 19374

Chalei V, Sansom S, Kong L, Lee S, Montiel J, Vance K (2014) The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. Elife 3: e04530

Chan PP, Lowe TM (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res 44: O184 – D189

Chen X, Wurtmann Ej, Van Batavia J, Zybailov B, Washburn MP, Wolin SL (2007) An orthology of the Ro autoantigen functions in 23S rRNA maturation in D. radiodurans. Genes Dev 21: 1328 – 1339

Chen X, Taylor DW, Fowler CC, Galán JE, Wang HW, Wolin SL (2013) An RNA degradation machine sculpted by Ro autoantigen and noncoding RNA. Cell 153: 166 – 177

Chen W, Moore MJ (2015) Spliceosomes. Curr Biol 25: R181 – R183

Chen LL (2016) Linking long noncoding RNA localization and function. Trends Biochem Sci 41: 761 – 772

Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G et al (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308: 1149 – 1154

Chow JC, Yen Z, Ziesche SM, Brown CJ (2005) Silencing of the mammalian X chromosome. Annu Rev Genomics Hum Genet 6: 69 – 92

Christov CP, Gardiner TJ, Szüts D, Krude T (2006) Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. Mol Cell Biol 26: 6993 – 7004

Ciganda M, Williams N (2011) Eukaryotic 5S rRNA biogenesis. Wiley Interdiscip Rev RNA 2: 523 – 533

Clark BS, Blackshaw S (2017) Understanding the role of IncRNAs in nervous system development. Adv Exp Med Biol 1008: 253 – 282

Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell 33: 717 – 726

Court F, Camprubi C, Garcia CV, Guillaumet-Adkins A, Sparago A, Seruggia D, Sandoval J, Esteller M, Riccio A, Sempere LF, Flatmark K, Hovig E et al (2015) A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. Annu Rev Genet 49: 213 – 242

Fuchs G, Stein Aj, Fu C, Reischm KH, Wolin SL (2006) Structural and biochemical basis for misfolded RNA recognition by the Ro autoantigen. Nat Struct Mol Biol 13: 1002 – 1009

Gabory A, Jammes H, Dandolo L (2010) The H19 locus: role of an imprinted non-coding RNA in growth and development. BioEssays 32: 473 – 480

Gall JC, Bellini M, Wu Z, Murphy C (1999) Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptional complexes. Mol Biol Cell 10: 4385 – 4402

Ganot P, Bortoloni ML, Kiss T (1997a) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell 89: 799 – 809

Ganot P, Caizergues-Ferrer M, Kiss T (1997b) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. Genes Dev 11: 941 – 956

Gicquel C, Gaston V, Mandelbaum J, Siffroi JP, Flahault A, Le Bouc Y (2005) Biochemical basis for misfolded RNA recognition by the Ro autoantigen. Trends Genomics 17: 286 – 293

Griffiths-Jones S (2004) The microRNA registry. Nucleic Acids Res 32: D109 – D111

Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kelis M, Werber M et al (2013) The tissue-specific IncRNA Fendrr is an...
essential regulator of heart and body wall development in the mouse. Dev Cell 24: 206 – 214
Grill MP, Massé E (2019) Mimicry, deception and competition: the life of competing endogenous RNAs. *Wiley Interdiscip Rev RNA* 10: e1525
Gugino G, Orlando V, Saieva L, Ruscelli P, Ciprani P, La Mannna MP, Giacomelli R, Alessandro R, Triolo G, Ciccia F et al (2018) Downregulation of miRNA17-92 cluster marks Vs9v62 T cells from patients with rheumatoid arthritis. *Arthritis Res Ther* 20: 236
Hacisuleyman E, Coff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McLellan H, Hendrickson DC, Sauvageau M, Kelley DR et al (2014) Topological organization of multichromosomal regions by the long intergenic noncoding RNA F ire. *Nat Struct Mol Biol* 21: 198 – 206
Hansen TB, Jensen TL, Clausen BH, Brämsen JB, Finsen B, Damgaard CK, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. *Nature* 495: 316 – 318
Hendrick JP, Wolin SL, Rinke J, Lerner MR, Steitz JA (1981) A soluble ribonuclease intermediate in protein synthesis. *J Biol Chem* 231: 241 – 257
Hodnett JL, Busch H (1968) Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei. *J Biol Chem* 243: 6334 – 6342
Horos R, Büscher M, Klei nendros R, Alleaume MA, Tarafder T, Schwarzl T, Dziuba D, Tischer C, Zielonka EM, Adak A et al (2019) The small non-coding vault RNA-1 acts as a riboregulator of autophagy. *Cell* 176: 1054 – 1067.e12
Hou YM (2010) CCA addition to tRNA: implications for tRNA quality control. *IUBMB Life* 62: 251 – 260
Hung T, Wang Y, Lin MF, Koegel AK, Takayama K, Grant GD, Holingsworth LM, Shah N, Uemichi T, Wang P et al (2011) Extensive and coordinated transcription of competing endogenous RNAs within cell-cycle promoters. *Nat Genet* 43: 621 – 629
Jädy BE, Kiss T (2003) A small nucleolar guide RNA functions both in 2'-O- ribose methylation and pseudouridylation of the US splicingosomal RNA. *EMBO J* 20: 541 – 551
Jeon SH, Lee K, Lee KS, Kunkeaw N, Johnson BH, Holthauzen LM, Gong B, Leelayuwat C, Lee YS (2012) Characterization of the direct physical interaction of nc886, a cellular non-coding RNA, and PKR. *FEBS Lett* 586: 3477 – 3484
Ji P, Dieterich S, Wang W, Böing S, Metzger R, Schneider PM, Tidow N, Jeon SH, Lee K, Lee KS, Kunkeaw N, Johnson BH, Holthauzen LM, Gong B, Leelayuwat C, Lee YS (2012) Characterization of the direct physical interaction of nc886, a cellular non-coding RNA, and PKR. *FEBS Lett* 586: 3477 – 3484
Khanter H, Myasnikov AG, Natchiar SK, Klaholz BP (2015) Structure of the human 80S ribosome. *Nature* 520: 640 – 645
Kickhoefer VA, Liu Y, Kong LB, Snow BE, Stewart PL, Harrington L, Rome LH (2015) The Telomerase/vault-associated protein TEP1 is required for vault RNA stability and its association with the vault particle. *J Cell Biol* 162: 157 – 164
Kim SH, Quigley GJ, Suddath FL, McPherson A, Soden D, Kim JJ, Weinzierl J, Rich A (1973) Three-dimensional structure of yeast phenylalnine transfer RNA: folding of the polynucleotide chain. *Science* 179: 285 – 288
Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP (2010) Noncoding RNA gas 5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 3: ra8
Kiss AM, Jädy BE, Bertrand E, Kiss T (2004) Human box H/ACA pseudouridylation guide RNA machinery. *Mol Cell Biol* 24: 5797 – 5807
Kiss-László Z, Heny S, Bachelerie JP, Caizergues-Ferrer M, Kiss T (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nuclear RNAs. *Cell* 85: 1077 – 1088
Koc EC, Burkhardt W, Blackburn K, Moyer MB, Schlatter DM, Moseley A, Spremulli LL (2003) The large subunit of the mammalian mitochondrial ribosome. Analysis of the complement of ribosomal proteins present. *J Biol Chem* 278: 43958 – 43969
Kong L, Hao Q, Wang Y, Zhou Z, Zhang YX (2015) Regulation of p53 expression and apoptosis by vault RNA 1-5p in cervical cancer cells. *Oncotarget* 6: 28371 – 28388
Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res* 47: D155 – D162
Kretz M, Webster DE, Flockhart R, Lee CS, Zehnder A, Lopez-Pajares V, Qu K, Zheng GX, Chow J, Kim GE et al (2012) Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev* 26: 338 – 343
Krude T, Christov CP, Hyrien O, Marheineke K (2009) Y RNA functions at the initiation step of mammalian chromosomal DNA replication. *J Cell Sci* 122: 2836 – 2845
Kunkeaw N, Moon SH, Lee K, Johnson BH, Tanasanvimon S, Javel M, Painjaitjai C, Chamgramol Y, Wongfeng W, Gong B et al (2013) Cell death/ proliferation roles for nc886, a non-coding RNA, in the protein kinase R pathway in cholangiocarcinoma. *Oncogene* 32: 3722 – 3731
Kurian L, Aguirre A, Sancho-Martinez I, Benner C, Hishida T, Nguyen TB, Reddy P, Nivet E, Krause MN, Nelles DA et al (2015) Identification of novel long noncoding RNAs underlying vertebrate cardiovascular development. *Circulation* 131: 1278 – 1290
Kwok GT, Zhao JT, Weiss J, Mugridge N, Brahmbhatt H, MacDiarmid JA, Robinson BG, Sidhu SB (2017) Translational applications of microRNAs in cancer, and therapeutic implications. *Noncoding RNA Res* 2: 143 – 150
Lagarde J, Uszcynska-Ratajczak B, Carbonelli S, Pérez-Lluch S, Abad A, Davis C, Gingers TR, Frankish A, Harrow J, Guigo R et al (2017) High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat Genet* 49: 1731 – 1740
Lagos-Quiñata M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294: 853 – 858
Lam YW, Trinkle-Mulcahy L, Lamond AI (2001) Structure of the human vault particle. *Science* 294: 853 – 858
Langenberger D, Çakir MV, Hoffmann S, Stadler PF (2013) Dicer-processed small RNAs: rules and exceptions. *J Exp Zool B Mol Dev Evol* 320: 35 – 46
Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294: 858–862
Lee JT, Davidow LS, Warshawsky D (1999) Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet 21: 400–404
Lee RC, Ambros V (2001) An extensive class of small RNAs in Caenorhabditis elegans. Science 294: 862–864
Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, Yu H, Xie Y, Mendell JT (2016) Noncoding RNA NOAR regulates genomic stability by sequestering PUMILO proteins. Cell 164: 69–80
Lerner MR, Boyle JA, Hardin JA, Steitz JA (1981) Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. Science 211: 400–402
Lestrade L, Weber MJ (2006) snoRNA-LBME-db, a comprehensive database of human HACA and C/D box snoRNAs. Nucleic Acids Res 34: D158–D162
Li Z, Chao TC, Chang KY, Lin N, Patil VS, Shimizu C, Head SR, Burns JC, Rana TM (2014) The long noncoding RNA TRHL regulates TNFα expression through its interaction with hRNPL. Proc Natl Acad Sci USA 111: 1002–1007
Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, Zhong G, Yu B, Hu W, Dai L et al (2015) Exon-intron circular RNAs regulate transcription in the nucleus. Nat Struct Mol Biol 22: 256–264
Li C, Xiao Y, Yang M, Su T, Sun X, Guo Q, Huang Y, Luo XH (2018a) Long noncoding RNA Bmncr regulates mesenchymal stem cell fate during skeletal aging. J Clin Invest 128: 5251–5266
Li X, Yang L, Chen LL (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics 27: i275–i282
Liu CX, Li X, Nan F, Jiang S, Gao X, Guo SK, Xue W, Cui Y, Dong K, Ding H et al (2019a) Structure and degradation of circular RNAs regulate PKR activation in innate immunity. Cell 177: 865–880.e21
Liu M, Wang Q, Shen J, Yang BB, Ding X (2019b) CircRNAbank: a comprehensive database for circRNA with standard nomenclature. RNA Biol 16: 899–905
Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, Procaccio V, Wallace DC (2013) mtDNA variation and analysis using mitomap and mitomaster. Curr Protoc Bioinformatics 44: i.23.1–i.23.26
Lottin S, Verroult-Édouard AS, Adriaenssen E, Czeszak X, Lemoine J, Roudbaraki M, Coll J, Hondermarck H, Dugimont T, Curgy JJ (2015) Novel rapidly evolving hominid RNAs bind to regulate neurogenesis. Nucleic Acids Res 44: D733–D745
Lund E, Dahilberg JE (1984) True genes for human U1 small nuclear RNA. Copy number, polymorphism, and methylation. J Biol Chem 259: 2013–2021
Lykke-Andersen S, Ardal BK, Hollensen AK, Damgaard CK, Jensen TH (2018) Box C/D snoRNA autoregulation by a cis-acting snoRNA in the NOP56 pre-mRNA. Mol Cell 72: 99–111.e5
Maraia RJ, Sasaki-Tozawa N, Driscoll CT, Green ED, Darlington GJ (1994) The human Y4 small cytoplasmic RNA gene is controlled by upstream elements and resides on chromosome 7 with all other hY scRNA genes. Nucleic Acids Res 22: 3045–3052
Mariaa R, Sakulich AL, Brinkmann E, Green ED (1996) Gene encoding human Ro-associated autoantigen Y5 RNA. Nucleic Acids Res 24: 3552–3559
Marnef A, Richard P, Pinzón N, Kiss T (2014) Targeting vertebrate intron-encoded box C/D 2′-O-methylation guide RNAs into the Cajal body. Nucleic Acids Res 42: 6616–6629
Martin F, Ménetret JF, Simonetti A, Myasnikov AG, Vicens Q, Prongidi-Fix L, Natchiar SK, Klaholz BP, Eriani G (2016) Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. Nat Commun 7: 2622
Marz M, Mosig A, Stadler BM, Stadler PF (2007) U7 snRNAs: a computational survey. Genomics Proteomics Bioinformatics 5: 187–195
Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol 8: 209–220
Memenčak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maiier L, Mackowiak SD, Gregersen LH, Munschauer M et al (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495: 333–338
Mencía A, Modamio-Høystjer S, Redshaw N, Morín M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T et al (2009) Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet 41: 609–613
Miyoshi N, Wagsatsuma H, Wakana S, Shiroishi T, Nomura M, Aisaka K, Kohda T, Surani MA, Kaneko-Ishino T, Ishino F (2000) Identification of an imprinted gene, Meg3/Gtl2, and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q. Genes Cells 5: 211–220
Muddashetty R, Khanam T, Kondrashov A, Bundman M, Iacoangeli A, Kremerskotten J, Duning K, Barnekow A, Hüttenhofer A, Tiedge H et al (2002) Poly(A)-binding protein is associated with neuronal BC RNA targets the eukaryotic initiation factor 4A helicase mechanism. Mol Cell Biol 28: 3008–3019
Lin MF, Jungreis I, Kellis M (2011) snaR genes: recent descendants of Alu snaR genes? Trends Genet 27: 100–107
Ng SY, Bogu GK, Soh BS, Stanton LW (2013) The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. Mol Cell 51: 349–359
O’Reilly D, Dienstbier M, Cowley SA, Vazquez P, Drozdz M, Taylor S, James WS, Murphy S (2013) Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. Genome Res 23: 281–291
Panagal M, Biruntha M, Vidhyavathth LM, Sivakumaranathan P, Senthilkumar SR, Sekar D (2019) Dissecting the role of miR-21 in different types of stroke. Gene 681: 69–72
Pandey GK, Mitra S, Subhash S, Hertwig F, Kenduri M, Mishra K, Fransson S, Ganeshram A, Mondal T, Bandaru S et al (2014) The risk-associated long noncoding RNA NABT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. Cancer Cell 26: 722–737
Parrott AM, Mathews MB (2007) Novel rapidly evolving hominid RNAs bind nuclear factor 90 and display tissue-restricted distribution. Nucleic Acids Res 35: 6249–6258
Parrott AM, Mathews MB (2009) snRNA genes: recent descendants of Alu involved in the evolution of chorionic gonadotropins. Cold Spring Harb Symp Quant Biol 74: 363–373
Pelczar P, Filipowicz W (1998) The host gene for intronic U17 small nuclear RNAs in mammals has no protein-coding potential and is a member of the 5′-terminal oligopyrimidine gene family. *Mol Cell Biol* 18: 4509–4518

Polacek N, Mankin AS (2005) The ribosomal peptide transferase center: structure, function, evolution, inhibition. *Crt Rev Biochem Mol Biol* 40: 285–311

Pruijn G, Slobbe RL, van Venrooij WJ (1991) Analysis of protein–RNA interactions within Ro ribonucleoprotein complexes. *Nucleic Acids Res* 19: 5173–5180

Reed JH, Sim S, Wolin SL, Clancy RM, Buyon JP (2013) Ro60 requires Y3 RNA for cell surface exposure and inflammation associated with cardiac manifestations of neonatal lupus. *J Immunol* 191: 110–116

Rich A, RajBhandary UL (1987) Transfer RNA: molecular structure, sequence, and properties. *Annu Rev Biochem* 56: 805–860

Richard P, Darzacq X, Bertrand E, Jädy BE, Verheggen C, Kiss T (2003) A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J* 22: 4283–4293

Rinn J, Kertesz M, Wang J, Squazzo S, Xu X, Brugmann S, Goodnough L, Helms J, Farnham P, Segal E et al (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by Noncoding RNAs. *Cell* 129: 1111–1123

Roberts RB (1958) Introduction. In *Microsomal particles and protein synthesis*, Roberts RB (ed.), pp vii–viii. New York, NY: Pergamon Press Inc

Rose D, Stadler PF (2011) Molecular evolution of the non-coding eosinophil granule ontogeny transcript. *Front Genet* 2: 69

Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO (2013) Cell-type-specific features of circular RNA expression. *PLoS Genet* 9: e1003777

Scott MS, Ono M (2011) From snoRNA to miRNA: dual function regulatory non-coding RNAs. *Biochimie* 93: 1987–1992

Sim S, Weinberg DE, Fuchs G, Choi K, Chung J, Wolin SL (2009) The subcellular distribution of an RNA quality control protein, the Ro autoantigen, is regulated by noncoding Y RNA binding. *Mol Cell Biol* 20: 1555–1564

Sim S, Yao J, Weinberg DE, Niessen S, Yates JR, Wolin SL (2012) The zipcode-binding protein ZBL1 influences the subcellular location of the Ro 60-kDa autoantigen and the noncoding Y3 RNA. *RNA* 18: 100–110

Singh R, Reddy R (1989) Gamma-monomethyl phosphate: a cap structure in splicedosomal U6 small nuclear RNA. *Proc Natl Acad Sci USA* 86: 8280–8283

Smith KC, Cordes E, Schweet RS (1959) Fractionation of transfer ribonucleic acid. *Biochim Biophys Acta* 33: 286–287

Smith C, Steitz J (1998) Classification of gas5 as a multi-small-nuclear-RNA (snRNA) host gene and a member of the 5′-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol Cell Biol* 18: 6897–6909

Sonkoly E, Bata-Csörgo Z, Pivarcsi A, Polyanka H, Kenderessy-Szabo A, Molnar S, Ógrimsson N, Mankin AS (2005) The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing. *EMBO J* 3: 2801–2807

Stultijs DM, Killen MW, Pierce HH, Ulitsky I (2008) Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res* 18: 13–18

Suzuki T, Nagao A (2011) Human mitochondrial tRNAs: biogenes is, function, structural aspects, and diseases. *Annu Rev Genet* 45: 299–329

Széli M, Bata-Csörgo Z, Kemény L (2008) The enigmatic world of mRNA-like ncRNAs: their role in human evolution and in human diseases. *Semin Cancer Biol* 18: 141–148

Tarn WY, Steitz JA (1996) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* 273: 1824–1832

The RNAcentral Consortium (2019) RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res* 47: D221–D229

Tichon A, Gil N, Lubelsky Y, Havkin Solomon T, Lemze D, Itzkovitz S, Shamir A, Bubulja PA et al (2010) The nuclear retarded noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 39: 925–938

Turunen JJ, Niemelä EH, Verma B, Frilander MJ (2013) The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA* 4: 61–76

Tyc K, Steitz JA (1989) U3, U6 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleus. *EMBO J* 8: 3113–3119

Tyczowski KT, Shu MD, Kukoyi A, Steitz JA (2009) A conserved WDDO protein binds the Cajal body localization signal of scaRNP particles. *Mol Cell* 34: 47–57

Ulitsky I, Shukumatova A, Jan CH, Sive H, Bartel DP (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147: 1537–1550

Ulitsky I (2016) Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* 17: 601–614

Ulitsky I (2018) Interactions between short and long noncoding RNAs. *FEBS Lett* 592: 2874–2883

Van Arsdel SW, Weiner AM (1984) Human genes for U2 small nuclear RNA are tandemly repeated. *Mol Cell Biol* 4: 492–499

Vance KW, Sansom SN, Lee S, Chalei V, Kong L, Cooper SE, Oliver PL, Ponting CP (2014) The long non-coding RNA Paupar regulates the expression of both local and distal genes. *EMBO J* 33: 296–311

Vazquez-Arango P, Vowles J, Browne C, Hartfield E, Fernandes HJ, Mandefro B, Soreen D, James W, Wade-Martins R, Cowley SA et al (2016) Variant U1 snRNAs are implicated in human pluripotent stem cell maintenance and neuromuscular disease. *Nucleic Acids Res* 44: 10960–10973
Vazquez-Arango P, O’Reilly D (2018) Variant snRNPs: new players within the spliceosome system. RNA Biol 15: 17–25
Volders PJ, Anckaert J, Verheggen K, Nuyten J, Martens L, Mestdagh P, Vandesompele J (2019) LNCipedia S: towards a reference set of human long non-coding RNAs. Nucleic Acids Res 47: D135–D139
Walker TA, Pace NR, Erikson RL, Erikson E, Behr F (1974) The 7S RNA common to oncornaviruses and normal cells is associated with polyribosomes. Proc Natl Acad Sci USA 71: 3390–3394
Walter P, Blobel G (1982) Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature 299: 691–698
Wang K, Yang Y, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie B, Protacio A, Flynn R, Gupta R et al (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472: 120–U158
Will CL, Lührmann R (2005) Splicing of a rare class of introns by the U12-dependent spliceosome. Biol Chem 386: 713–724
Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogeness JB, Schultz PG (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science 309: 1570–1573
Wolin SL, Steitz JA (1983) Genes for two small cytoplasmic Ro RNAs are adjacent and appear to be single-copy in the human genome. Cell 32: 735–744
Wolin SL, Steitz JA (1984) The Ro small cytoplasmic ribonucleoproteins: identification of the antigenic protein and its binding site on the Ro RNAs. Proc Natl Acad Sci USA 81: 1996–2000
Wolin SL, Belair C, Boccitto M, Chen X, Sim S, Taylor DW, Wang HW (2013) Non-coding Y RNAs as tethers and gates: insights from bacteria. RNA Biol 10: 1602–1608
Wu H, Yin QF, Luo Z, Yao RW, Zheng CC, Zhang J, Xiang JF, Yang L, Chen LL (2016) Unusual processing generates SPA LncRNAs that sequester multiple RNA binding proteins. Mol Cell 64: 534–548
Wu H, Yang L, Chen LL (2017) The diversity of long noncoding RNAs and their generation. Trends Genet 33: 540–552
Xing YH, Yao RW, Zhang Y, Guo CJ, Jiang S, Xu G, Dong R, Yang L, Chen LL (2017) SLERT regulates DDX21 rings associated with Pol I transcription. Cell 169: 664–678.e16
Yan JJ, Qiao M, Li RH, Zhao XT, Wang XY, Sun Q (2019) Downregulation of miR-145-5p contributes to hyperproliferation of keratinocytes and skin inflammation in psoriasis. Br J Dermatol 180: 365–372
Ye XT, Huang H, Huang WP, Hu WL (2018) LncRNA THOR promotes human renal cell carcinoma cell growth. Biochem Biophys Res Commun 501: 661–667
Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, Chen LL (2012) Long noncoding RNAs with snoRNA ends. Mol Cell 48: 219–230
Yotova I, Vlatkovic I, Pauler F, Warczok K, Ambros P, Oshimura M, Theussl H, Gessler M, Wagner E, Barlow D (2008) Identification of the human homolog of the imprinted mouse Air non-coding RNA. Genomics 92: 464–473
Young TL, Matsuda T, Cepko CL (2005) The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. Curr Biol 15: S01–S12
Younis I, Dittmar K, Wang W, Foley SW, Berg MC, Hu KY, Wei Z, Wan L, Dreyfuss G (2013) Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. Elife 2: e00780
Zemmour D, Pratama A, Loughhead SM, Mathis D, Benoist C (2017) A long noncoding RNA, modulates Foxp3 expression and autoimmunity. Proc Natl Acad Sci USA 114: E3472–E3480
Zerbova DR, Achuthan P, Akanni W, Amode MR, Barrett D, Bhai J, Billis K, Cummins C, Gall A, Girn C et al (2018) Ensembl 2018. Nucleic Acids Res 46: D754–D761
Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL (2013) Circular intronic long noncoding RNAs. Mol Cell 51: 792–806
Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L (2014) Complementary sequence-mediated exon circularization. Cell 159: 134–147
Zieve G, Penman S (1976) Small RNA species of the HeLa cell: metabolism and subcellular localization. Cell 8: 19–31

License: This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.