

How noncoding RNAs began to leave the junkyard



The research community focused on noncoding RNAs keeps growing. Skepticism about the field has some history.

By Vivien Marx

Junk. In the view of some, that's what noncoding RNAs (ncRNAs) are – genes that are transcribed but not translated into proteins. With one of his ncRNA papers, University of Queensland researcher Tim Mercer recalls that two reviewers said, “this is good” and the third said, “this is all junk; noncoding RNAs aren't functional.” Debates over ncRNAs, in Mercer's view, have generally moved from ‘it's all junk’ to ‘which ones are functional?’ and ‘what are they doing?’ Researchers are mapping out the future of the field, which is the theme of [a second story](#) in this issue. Scientists in the ncRNA field have faced skepticism and worked to dispel it. Here are some individual histories.

In 1869, Friedrich Miescher isolated ‘nuclein’ and “floated the idea” it might be genetic material, note the authors of a new book, *RNA: The Epicenter of Genetic Information*¹ by John Mattick, professor of RNA biology at the University of New South Wales, and Paulo Amaral from Insper, a university in São Paulo, Brazil. Nuclein later turned out to be DNA. Not long thereafter, ribonucleic acid (RNA) was isolated. But, they write, nucleic acids barely feature in the history of biochemistry before 1940. Proteins were thought to have greater chemical diversity than the “assumed monotony of nucleic acids,” which were considered to be “merely structural or metabolic entities.”

In the 1960s

“Nick, why don't you earn an honest living and work on RNA?” is what Nicholas Delihias recalls hearing from Aaron Bendich, who led a nucleoside lab at the Sloan-Kettering Institute for Cancer Research. Delihias, now an RNA molecular biologist at Stony Brook University, was a postdoctoral fellow whom Bendich mentored. Delihias respected Bendich as an out-of-the-box thinker and took the mentor's words to heart. “I owe my research career to him,” says Delihias.



To some, noncoding RNAs are junk. Over time they have left the proverbial junkyard.

The conversation with Bendich took place sometime in 1961, says Delihias. At the time, many researchers “thought RNA does not do much of anything,” he says. It holds ribosomal proteins together and acts as ribosome scaffolding. This was also the year Sydney Brenner, François Jacob and Matthew Meselson published about an “intermediate information carrier” they called messenger RNA². It was the molecule that moves information from DNA in the nucleus to the protein-making machinery in the cytoplasm. Delihias remembers the excitement about this finding.

These historic mRNA experiments happened in a lab next to his, says Caltech RNA biologist Mitch Guttman, and “the whole crew came out to Caltech to use Meselson's techniques to explore the elusive ‘transient’ intermediate.” The experiments included infection of *Escherichia coli* with a bacteriophage, which switched off bacterial protein synthesis and turned on phage protein

production. The bacteria were first grown in medium with heavy isotopes, infected, then grown in medium with light isotopes. Density gradient centrifugation of purified ribosomes led them to find labeled, newly synthesized RNA and protein.

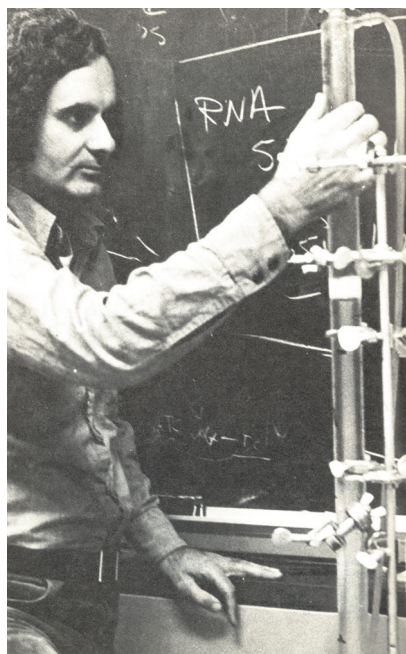
Delihias took note of other thinkers like Bendich. In 1969, Roy Britten from the Carnegie Institution and Eric Davidson at Rockefeller University presented a theory of the genomic regulatory machinery systems in ‘higher organisms’ and suggested “a sizeable portion of the functional genes in differentiated cell types may be regulatory genes.”³ They discuss “activator RNAs” and “integrator genes.” Says Guttman, “in many ways this is precisely what we now know many lncRNAs do,” referring to long noncoding RNAs. Before Davidson's passing in 2015, Guttman frequently talked with him.

Says Mattick, Davidson had experimentally observed ‘informational RNAs’ in amphibian

and sea urchin embryos. In their 1969 paper, when Davidson and Britten highlighted the large genomes of 'higher' organisms, they noted diversity of transcripts in the nucleus and the abundance of repetitive sequence transcribed in a cell-specific fashion. Their ideas were, says Mattick, influenced by François Jacob and Jacques Monod, who characterized *E. coli*'s lac operon and their concepts of genetic operators, regulators and structural components. Britten and Davidson's ideas were also influenced by Barbara McClintock's model of gene regulation, says Mattick. Although some researchers supported such regulatory concepts, the scientific community more generally was unreceptive and the protein-based regulatory schema dominated. RNA regulation was unnecessary and, as Mattick and Amaral note in their book, seen as "flights of fancy."

Hello 1980s

Delahas joined the faculty at Stony Brook after postdoctoral fellowships. During one stint a department chairman told him it was pointless to work with RNA, but Delahas remained fascinated by RNA structure and function. In the 1980s, his lab provided the first experimental evidence of an RNA gene and RNA transcript regulating another gene's expression⁴. It is "a very good feeling" and a privilege to have



In the 1980s, Nicholas Delahas and his lab at Stony Brook University provided the first experimental evidence of an RNA gene and RNA transcript regulating another gene's expression.

done so, he says. In the early 1980s Delahas's colleague Masayori Inoue approached him asking for help finding *E. coli* *micF*'s transcript. In their long collaboration, Delahas and his team did just that. The work involved isolating and characterizing ³²P-labeled *micF* RNA, a 93-nucleotide ncRNA. They amplified the *micF* gene by building a plasmid with multiple gene copies, purified the RNA from the bacteria and pulled the ³²P-labeled *micF* from other RNAs using gel electrophoresis. They characterized the gene and its promoter and determined experimentally that the ncRNA inhibits translation of a target messenger RNA in response to environmental stress. *micF* base-pairs and forms a duplex with a target mRNA, namely *ompF*, thereby regulating gene expression. The finding predates finding the first eukaryotic regulatory ncRNA, a microRNA (miRNA).

As John Rinn at University of Colorado Boulder says, a number of women in 1980s discovered lncRNAs before those ncRNAs "had a name," notably, Shirley Tilghman, Denise Barlow and Carolyn Brown. "Newfangled" lncRNA is just a "bigger version of what they discovered." As the 1990s unfolded, important RNAs were being found in the genome, he says.

It's the 1990s

RNA is a trespasser: it "trespasses in what was once thought to be protein's province," note Marvin Wickens and Kathy Takayama from the biochemistry department at the University of Wisconsin Madison in their 1994 comment about ncRNAs in the nematode *Caenorhabditis elegans*⁵. The trespasser RNA regulates the *lin-14* gene in *C. elegans*, which encodes a nuclear protein important early in development and is then repressed. Two RNAs encoded by the *lin-4* gene appeared to repress expression of *lin-14* mRNA after the mRNA is made and processed. Since that early work, says Victor Ambros, who is now at University of Massachusetts Medical School and was lead author of one of the papers, it's become clear "the short one is the repressor" and that it is processed from a longer precursor.

lin-4 had been identified in Sydney Brenner's lab. That team and others had studied how mutations of this gene sent worm development awry. Repression of *lin-14* involves acting on sequences of the *lin-14* mRNA between the termination codon and the poly(A) tail, which was, Wickens and Takayama point out, "thought to be barren." It holds "the tantalizing possibility that a new family of regulatory RNAs awaits discovery." At the time, says Wickens, he was "pretty confident" other RNAs would be found, but his sense was that many "thought



In 1993, Victor Ambros (left), Rosalind Lee (middle) — both now at UMass Chan Medical School — and Rhonda Feinbaum (right), now at Lahey Hospital & Medical Center, published findings about *lin-4* in *C. elegans*. It was the first eukaryotic noncoding regulatory RNA and the founding member of the class of microRNAs.

it was just a weird *C. elegans* phenomenon." With discovery of regulatory elements in the 3' untranslated regions of many mRNAs, they note, "has come an assault on the factors with which they interact" involving gel shift experiments, ultraviolet crosslinking and screens of expression libraries. Odd, enigmatic RNAs lurk in the literature, they write: RNAs that are polyadenylated and spliced as mRNAs yet "seem to not be translated." They point to the genes *Xist* and *H19* and to sea urchin eggs with RNAs of unknown function. They ask if these RNAs might be "grotesque deviants, one-of-a-kind aberrations, like characters in a Fellini film?" Instead of oddities, they might turn out to "have been our first emissaries from an unexplored and vast RNA world."

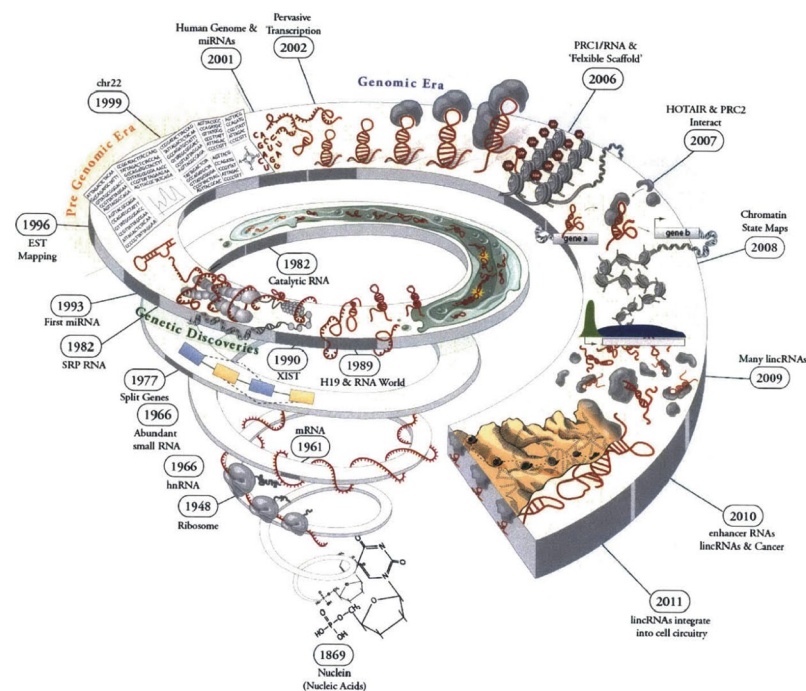
Looking back in 2004 on their work on *lin-4* and *lin-14*, the scientists leafed through their lab notebooks⁶. "We were astonished," write Ambros, then at Dartmouth Medical School, and Rhonda Feinbaum and Rosalind Lee of Harvard Medical School, about how science had changed in the decade since their 1993 *C. elegans* publication. Sequencing was done using 18-inch gels and autoradiography. The worm genome was but a fragmentary collection and software to explore it was accessed "by obtuse line commands to a lethargic central mainframe." Sequence alignments for one of their figures were done by hand and took, seemingly, months. A decade on, the tasks would have been "trivial" with modern software. The most dramatic change was that in 1993 there was little to no interest in *lin-4* or its "little RNA product," that is, "outside of a very small circle of friends."

As Aurora Esquela-Kerscher from Eastern Virginia Medical School points out⁷, *lin-4* is “the founding member of the miRNA superfamily.” Many miRNAs have been identified in plant, animal and viral genomes, and they appear to affect diverse cellular processes including proliferation, apoptosis, differentiation, metabolic and immune responses. Studying *lin-4* in *C. elegans* brought a fundamental understanding of miRNAs mechanisms. miRNAs are “more complex than initially predicted,” and they direct important functions in the nucleus and cytoplasm; they modulate genes in positive and negative ways. “Stay tuned – these tiny RNAs likely have bigger surprises in store for us!”

miRNAs and this *lin-4* work shape the work of Megan Linscott, a postdoctoral fellow in the lab of Toni Pak at Loyola University. That lab’s focus is on ncRNAs and their regulatory role throughout the human lifespan. Linscott explores the regulatory switches in puberty and hormone regulation. Once researchers discovered that one miRNA can regulate hundreds of different mRNAs – this began with the work on *lin-4* – “it was a total game changer,” says Linscott. “Suddenly, we had an explanation for how many different parts of a given pathway might be influenced by a single noncoding element.” She read about hydroxymethylation of RNA transcripts, and the concept of multiple layers of regulation, especially on the RNA level, held her interest. She had known about DNA hydroxymethylation, so the idea that RNA could also be modified or that RNA itself could help induce hydroxymethylation “was the ultimate chicken or the egg question; it was impossible not to take interest in the noncoding RNA field.” Size, Linscott says, is one exciting miRNA trait. “Who would have thought a 22-nucleotide sequence could pack such a punch?” In the rat brain, she and her colleagues have shown that specific miRNAs are influenced by aging, hormones and alcohol. The ncRNAs essentially operate as small molecular ‘switches’, which also hints at their potential as therapeutic targets.

Fascination in the aughts

It was the spring of 2008 and she was a postdoctoral fellow at Harvard studying chromatin and histone modifiers, says Maite Huarte, who is now a principal investigator at Cima Universidad de Navarra, the research center of Clínica Universidad de Navarra in Spain. She and her team work on lncRNAs and gene regulation in cancer. At Harvard, she met John Rinn, who was setting up his lab at the Broad Institute of MIT and Harvard. He was part of what was informally called ‘the Broad Group’ with Aviv Regev and Eric Lander, says Rinn. Huarte also



For his 2012 dissertation at MIT, Mitch Guttman considered the history of ncRNAs, to which he has devoted his career.

met Guttman, then a Lander lab PhD student. RNA sequencing had not yet arrived, she says. Tiling arrays were yielding RNA expression data from intergenic parts of the mouse genome from different tissues and from various stages of development. “It totally fascinated me,” says Huarte, when Rinn told her about their results on regulated expression of mRNA-like RNAs from noncoding intergenic regions, which they named long noncoding RNAs.

It was unknown at the time why these RNAs were being transcribed. “There was so much to do, and many exciting possibilities,” she says. When Rinn invited her to work in his lab, “I didn’t hesitate,” says Huarte. One week later she was doing northern blots of lncRNAs at the Broad Institute. “I was particularly interested in how lncRNAs would affect key biological pathways, so I started to investigate their role in the *p53* response.” *p53* is a tumor suppressor gene and is mutated in many tumor types.

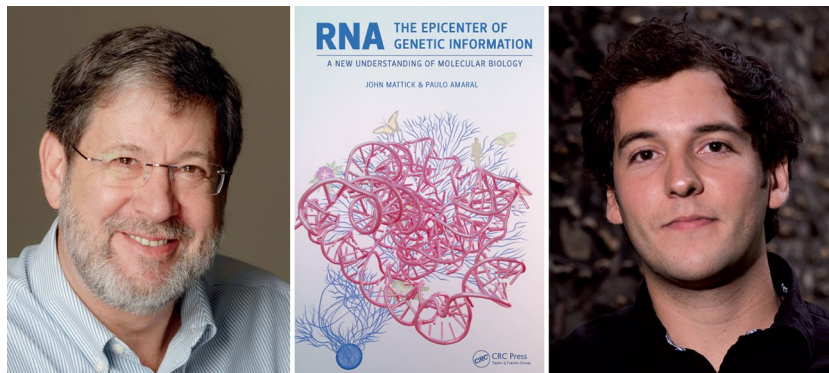
In Huarte’s view, one of the most important reasons ncRNAs have transitioned from the realm of junk to importance is what labs have seen in loss-of-function experiments with lncRNAs and by using RNA interference (RNAi), among other methods. These experiments have been “key to show that lncRNAs have genuine cellular functions.” Few lncRNAs show strong phenotypes when mutated, but many have regulatory roles that can be

assessed by studying how they alter gene expression. Orthogonal methods matter for getting such insight on ncRNAs, she says. “The lack of rigor in some studies has fed the skepticism of some researchers, and we face the challenge of producing the best possible evidence to overcome this prejudice.”

Into the jungle

In 2012, a consortium annotating the human genome – Encyclopedia of DNA Elements (ENCODE) – published a flurry of papers⁸. By identifying, for instance, the human genome’s protein-coding regions and regulatory elements including RNAs, the researchers said they had been able to assign biochemical functions to 80% of the genome, “in particular outside of the well-studied protein-coding regions.” Critics noted⁹ ENCODE played “fast and loose” with the term ‘function’ and had divorced genomic analysis from evolutionary context. Mercer, summarizing the ENCODE critics, says they claimed ENCODE had been “sequencing a whole bunch of junk.” He recalls the debate at times reached a fever pitch.

Mercer first learned about ncRNAs as a PhD student at University of Queensland working with Mattick, then also at that university. When scanning the genome with tiling arrays, “the whole thing was lighting up,” says Mercer. It seemed like “too much noise,”



In a new book, University of New South Wales researcher John Mattick (left) and Insper researcher Paolo Amaral reflect on the history and future of noncoding RNA research.

perhaps a technical problem. RNA sequencing has enabled digging into this ‘noise’ to reveal sections that indicate gene architecture; regulation by transcription factors and epigenetic factors; and the abundances of RNAs, which are often expressed in only some cell types. Matters can now shift, he says, from sweeping statements about junk and transcriptional noise to the practicalities of exploring functionality of ncRNAs. Of late, Mercer’s career has shifted toward more translational work, and he sees potential in RNA-based therapeutics. He is a co-author of a forthcoming community-driven paper in *Nature Reviews Molecular Cell Biology*¹⁰ about definitions, functions and challenges related to lncRNAs.

Fitting in

In the 1960s and 1970s, says Guttman, many RNAs were identified that were not translated into proteins. These RNAs were sometimes grouped together as heteronuclear RNAs (hnRNAs). Plenty of them were well-characterized and shown, for example, to never leave the nucleus nor engage with polyribosomes. Looking back, he says, this seems to be early evidence of what is now seen in assays. At the time, once splicing was discovered, many scientists thought splicing explained how mRNA “turns over” in the nucleus. Ten years of work on heteronuclear RNAs had led to the realization they are “just introns,” he says. “Mystery solved, right?” People weren’t ignorant, he says; they had an explanation for these ncRNAs and the community moved on. Into the 1980s most molecular biologists knew of many types of RNAs: ribosomal RNAs, tRNAs, mRNAs and also small RNAs. The research community knew that genes could have different properties, and “the ambiguities that existed early on started to be filled in.”

When Guttman began his PhD research^{11,12} in Eric Lander’s lab at MIT and the Broad Institute, it was one of just a few labs with the new Solexa sequencer, which later led to Illumina instruments, he says. It was, he says, a “huge inflection point in the history of genomics,” as the research community shifted from genome sequencing with Sanger-based instruments to this new technology. The Lander lab, like others at the time, used the instrument to study chromatin modifications. As Guttman tinkered with the datasets, he came across signatures not annotated as protein-coding genes, which sparked his fascination with ncRNAs.

What was lacking then and what is lacking still, he says, is a theory that would allow fitting RNA into the larger scheme of regulation. “Because all of the examples we knew of were kind of one-offs,” says Guttman. Small nuclear RNAs, for example, will base-pair with introns at splice sites to guide the splicing machinery. “How do you generalize beyond splicing?” he asks. Small nucleolar RNAs base-pair with 45S pre-ribosomal RNA; that’s another “one-off.” *Xist*, a lncRNA, silences one of the two X chromosomes in female mammals’ X chromosomes, and it presents another extrapolation challenge, says Guttman. The evidence about *Xist* is generally accepted, he says, but it remains seemingly exceptional.

Mattick prefers the term ‘exemplary’ to exceptional. The *lac* operon was and is taught as the “exemplar” of gene regulation, he says. The ncRNA *Xist* was thought an exception. But given the way it interacts with chromatin structure and affects gene expression, he says, it’s an “exemplar” of a ncRNA. As a post-doctoral fellow at Baylor College of Medicine “fresh off the boat,” says Mattick, he recalls an after-hours pub chat with another Baylor postdoc in late 1977. Mattick was working on

structure and function of the fatty acid synthase complex. “Everyone was trying to clone their favorite gene,” he says. ‘His’ was nearly 20 kilobases long. It took multiple rounds of sucrose density gradient purification “and it was really tough,” he says.

Mattick remembers work from Phil Sharp and Rich Roberts, who showed in electron micrographs that adenovirus RNAs hybridizes to the genome with some sections looped out, or spliced. The loops are introns, which was not generally appreciated: introns were thought to be junk. Protein-coding genes stayed in scientific focus. In that evening conversation at Baylor, Mattick recalls his fellow postdoc calling introns “junk.” They debated this, and many other debates have followed since. ncRNAs are Mattick’s research focus. Over his career, his focus has stayed on the architecture and role of regulatory RNAs.

The history of gene regulation up to the present, says Mattick, is shaped by protein-centric thinking interrupted by work of outstanding figures. His list of ‘heroes’ includes Britten, Davidson and McClintock. These scientists integrated disparate information on RNAs, and “got very close to the truth,” he says. In his view, other explanations of regulation, such as the notion that cellular processes can be explained by the combinatorics of transcription factors or other regulatory proteins, are not valid. It’s true, findings about ncRNAs can appear to some as “just so stories,” says Mattick. Doubts may remain and some aspects remain challenging to prove. Says Mattick, “the field is evolving.”

Vivien Marx ✉

Nature Methods.

✉ e-mail: v.marx@us.nature.com

Published online: 6 October 2022

References

1. Mattick, J. Amaral, P. *RNA: The Epicenter of Genetic Information* (CRC, 2022).
2. Brenner, S., Jacob, F. & Meselson, M. *Nature* **190**, 576–581 (1961).
3. Britten, R. J. & Davidson, E. H. *Science* **165**, 349–357 (1969).
4. Andersen, J., Forst, S. A., Zhao, K., Inouye, M. & Delihai, N. *J. Biol. Chem.* **264**, 17961–17970 (1989).
5. Wickens, M. & Takayama, K. *Nature* **367**, 17–18 (1994).
6. Lee, R., Feinbaum, R. & Ambros, V. *Cell* **116**, S89–S92 (2004).
7. Esquela-Kerscher, A. *Cell Cycle* **13**, 1060–1061 (2014).
8. ENCODE Project Consortium. *Nature* **489**, 47–74 (2012).
9. Graur, D. et al. *Genome Biol. Evol.* **5**, 578–590 (2013).
10. Mattick, J. S. et al. *Nat. Rev. Mol. Cell Biol.* (in the press).
11. Guttman, M. *Functional Large Non-coding RNAs in Mammals*. PhD thesis, MIT (2012).
12. Guttman, M. & Rinn, J. L. *Nature* **482**, 339–346 (2012).