Thinking outside the TATA box

Howy Jacobs

When I was doing my PhD research, in the middle of the last century, the nature and complexity of the RNA transcripts of the genome was a topic of lively debate, into which my own work became drawn. The complexity of the mRNA population in a given tissue, cell-type or developmental stage—considered as transcripts that were polyadenylated and associated with ribosomes—was estimated to represent some 40–50,000 distinct mRNAs of typical size, both in plants and vertebrates. The advent of high-throughput sequencing has pretty much confirmed these numbers, although they are derived from fewer genes because of alternative splicing, and variable sites of transcription initiation or polyadenylation.

It was already clear back then that the mRNA populations of different cell types were overlapping but not identical, providing a key plank in support of the idea that an invariant genome could specify different developmental fates via the synthesis of varying subsets of proteins encoded by that genome [1]. As soon as introns were discovered, we also had a clear—if retrospectively only partial—explanation for why the set of RNA molecules in the nucleus was so much more complex and varied than in the cytoplasm. Eukaryotic gene regulation has been viewed ever since as just the lac-operon on a vast scale, with the additional bells and whistles of the many different stages in the life history of each given mRNA.

Over the years I have always taught my students to treat this paradigm with caution, if not skepticism: an unproven assumption that is probably too neat and tidy to be the whole or even the major explanation for cell specification and homeostasis. Unfortunately, I also taught them to be skeptical of whatever I told them, so like most practitioners of molecular biology, they have largely ended up believing that differential mRNA transcription, shaded here and there by a bit

of alternative RNA processing and translation, is a sufficient explanation for how genes execute the developmental programme and respond to the environment.

But times are slowly changing. So-called noncoding RNAs have been rediscovered in the past decade or so, starting with miRNAs, then other small RNAs, and subsequently a vast zoo of long non-coding RNAs. Most biologists now accept that there are many tens of thousands of them, that they are differentially expressed, and that they contribute meaningfully to cell phenotype, as modifiers of gene expression acting at many levels [2]. Despite this, we remain to a large extent stuck in a mental groove. I think it is now time to expand our thinking a bit further.

The problem, to me, is that we have considered only one set of closely related processes as being the target of action of these molecules, namely protein synthesis. To a large degree this is conditioned by what we already know about: the direct functions of rRNAs and tRNAs in translation, and of snRNAs and snoRNAs in mRNA and rRNA processing. Now, armed with many specific examples, we have come to think of ‘noncoding’ RNAs as agents that determine the patterns of which mRNAs are made, stabilized and translated, often acting at the chromatin level. But we have lost sight of another possibility, which the exponents of the ‘RNA world’ hypothesis already taught us 25 years ago: that RNA molecules may have functions in cells that are not connected with gene expression at all. These certainly include enzymatic activities that can influence the structure or behaviour of other RNAs. But even this may be too limited. Another set of possible functions is staring us in the face and has been almost completely ignored.

We already know that proteins are highly versatile molecules that perform a vast array of enzymatic and structural roles. We have also become comfortable with the idea that post-translational modifications can

dramatically alter those properties—protein phosphorylation, acetylation, ubiquitylation, proteolytic cleavage and a host of other covalent alterations create an enormous functional diversity. We also know many examples of cofactors that are required for, or modify the action of a protein. Yet, we seem to have largely ignored the possibility that RNA molecules could also, and directly, influence protein activities, via noncovalent interactions and scaffolding functions. Anyone purifying a protein will usually take steps to exclude RNA, considering it as an annoying contaminant. But could this not be a classic case of throwing out the baby with the bathwater? Are we missing a whole class of biological actions that are generically determined, epigenetically responsive and finely regulated? Only that they are effected by RNA, not protein, or by both acting together.

Modifier RNAs may not bind tightly to proteins they regulate. Indeed, to fulfil a regulatory role by mass action they should mostly associate with their protein partners only weakly. So our benchmark cannot be something like biotin. What can we do to explore such a hypothesis? Fortunately, the tools of high-throughput genomics provide an obvious and easy answer. Protein biologists, having painstakingly designed and executed their most careful and gentle purifications, need only add a bucket-load of proteinase K to their precious materials, then process them as RNA preps. Deep sequencing should then reveal what, if anything, has been co-purified, applying appropriate controls.

References

1. Wilson EB (1925) *The Cell in Development and Inheritance*, Macmillan.
2. Mercer TR, Mattick JS (2013) *Nat Struct Mol Biol* 20: 300

DOI 10.1002/embr.201338175