
Sleeping Beauty and the Beast (of pervasive transcription)

DOMENICO LIBRI

Institut Jacques Monod, CNRS-Université Paris Diderot, 75205 Paris cedex 13, France

In a 2003 movie, *Good Bye Lenin*, the protagonist's mother, living in East Berlin in 1989, enters a coma just before the fall of the Berlin wall and the reunification of Germany. When she wakes up, her son makes all possible efforts to conceal the changes, as the shock would jeopardize the mother's fragile health. To this end, he builds up an old East Germany world around her—up to the restoration of a lost East Germany pickles brand using jars recovered from garbage.

A transcription scientist that would have slept through the last fifteen years or so would probably be disconcerted to a similar extent when going back to the bench. This modern *Sleeping Beauty* would probably feel fuzzy discovering the complexity and diversity of the new transcription world (and, of course, its contradictions) pretty much as if jumping from an old East Germany grocery store to a modern supermarket. She/he would probably trash results confusing gene expression and transcriptional activity, and would certainly lose her/his positivistic faith in the accuracy of Mother Nature when dealing with transcription initiation. *Sleeping Beauty* might regret the old comfortable drawings depicting regions of transcription as definite boxes driven by well-defined promoters firing to deliver a well-defined message, and would definitely forget the notion that RNA is just an intermediate in the flow of genetic information from the DNA to the protein. But she/he would also dig out from the trash many “artifacts” and “crazy” stories, including the unexplained association of intragenic sequence variants to human diseases—East Germany pickles.

The fall of the Berlin wall in this fantasy is the discovery that transcription occurs almost everywhere in the genome, overlapping annotated transcription units in the sense or antisense direction, initiating in a rather promiscuous manner outside or inside genes. And many of these transcription events are hidden by the fact that their RNA product—the traditional landmark of transcription—is often undetectable because degraded by nuclear or cytoplasmic pathways. The discovery of pervasive transcription owes much to the development of high throughput approaches for analyzing the transcriptome, and in particular to tiling microarrays. The first reports describing a vast repertoire of transcripts derived

from non-annotated regions (or antisense to annotated regions) were published in the early 2000s, from the Gingeras, Rotman, Shoemaker labs in the human system, the Ecker, White, and Steinmetz labs in other eucaryotes (including yeast) and the Church lab in bacteria. In 2005 we published what is probably the first report of hidden transcription, i.e., transcription generating RNAs that are not sufficiently stable to be detectable above current sensitivity thresholds. This was a fortunate encounter of unexplained results from three teams. Analyzing the transcriptome of a nuclear RNA degradation pathway mutant (*rrp6Δ*), A. Jacquier had noticed unexpected signals deriving from intergenic regions. We had observed that RNA Pol II ChIP signals from “negative” controls (i.e., “non-transcribed” intergenic regions) were sensitive to mutations in transcription initiation factors. The Séraphin lab was working on a poly(A) polymerase of unknown function. Together, these data were the starting point for demonstrating the existence of a class of RNAs that we called Cryptic Unstable Transcripts or CUTs: these are produced outside canonical transcription units and degraded by the nuclear exosome after polyadenylation by a non-canonical poly(A) polymerase. The Ares and the Tollervey labs had similar data that were published shortly after. Since then, the coverage of the genome by the annotation of non-coding, stable and unstable RNAs has increased continuously. High-resolution genome-wide maps of non-coding RNAs in yeast have been provided by the Jacquier, Steinmetz, Morillon, and, more recently, the Cramer and the Corden labs. Very few “blank” spaces are left in the *Saccharomyces cerevisiae* genome, with the suspicion that they have not yet been filled in by the right study. Key findings from these studies have drastically modified the way we used to think about transcription. One of these is that by default promoter firing appears to occur in a bi-directional manner, and in many cases the two directions are not equivalent in terms of functional significance (e.g., a protein coding gene in one direction and a non-coding RNA without established function in the other). Although work from many teams, including the Weissman, the Proudfoot, and the Buratowski labs, have demonstrated that transcription in the most informative direction (e.g., a protein coding

Corresponding author: libri.domenico@ijm.univ-paris-diderot.fr

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.050948.115>. Freely available online through the RNA Open Access option.

© 2015 Libri This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

gene) is generally preferred, “reverse direction” firing is far from been negligible and generally leads to the production of non-coding, non-functional RNAs that may overlap neighboring transcription units. Another important—and largely unexpected—established principle is that the control of initiation is inherently leaky. Paradoxically, it could be argued that more information is required to prevent transcription initiation than to promote it. This is maybe best illustrated by the demonstration, for instance by the Winston and the Workman labs, that disfavoring nucleosome re-assembly in the wake of transcribing polymerase leads to frequent intragenic transcription initiation at positions where initiation is not presumed to be selected for function. The nucleosomal structure of chromatin is an obstacle for the initiation of transcription. Transcription leading to the production of “pervasive transcripts” (such as CUTs) might require poor sequence informational content to initiate once a breach in such a repressive chromatin structure is found. The third important lesson from these years of hidden transcription studies is that steady state RNA levels are only partial benchmarks of transcriptional activity. Degradation pathways, both nuclear and cytoplasmic, have clear and major roles in shaping the transcriptome. Work from many labs, including ours, has demonstrated that the RNA landscape in the absence of key degradation enzymes, such as the exosome, Xrn1 and nonsense mediated decay factors is drastically different.

Studies on hidden transcription, pioneered in yeast, have extended to metazoans. A seminal work published in 2008 from the Jensen lab has revealed the occurrence of hidden RNA production in humans (PROMPTs, promoter proximal transcripts), detected, as in yeast, when degradation is defective. Parallel studies from the Sharp and Lis labs demonstrated that promoters of mouse and human genes fire bidirectionally, extending the analogy with the fungal system.

Pervasive transcription is dangerous for the stability of the transcriptome: if concurrent transcription is generally tolerated in regions of elongation, transcription through promoter regions generally inhibits their function—a phenomenon known as transcription interference. Failing the establishment of a robust control on initiation, the cell has developed systems to confine transcription that has started inappropriately, which, both in yeast and the metazoan systems, rely heavily on transcription termination and, to a lesser extent, on the degradation of the RNA produced.

But the Beast of pervasive transcription is probably so only at first sight, as in the tale. Some of the non-canonical tran-

scription events have been shown to have a regulatory function, via the non-coding RNA produced or the act of transcription itself. The share of non-canonical transcriptional events that have a function is forever growing. Maybe the most striking example is the accumulation of DNA variants mapping to regions that do not contain annotated genes and that have been linked to human diseases by genome-wide association studies (GWAS). These sites have been calculated to represent 90% of GWAS hits: one of the possible explanations for these variants is that some of these mutations alter the sequence of a non-coding RNAs that might be involved in regulation.

The nature of the regulatory event linked to non-canonical transcription generally appears to be different in the different systems (e.g., yeast versus human cells): if in human cells (and many other organisms including *S. pombe*) many non-coding RNAs have been convincingly shown to be involved in the control of gene expression, in *S. cerevisiae* it is essentially the act of transcription itself, and not the non-coding RNA produced, that functions in regulation.

However, it is very likely that the functional cases will remain a small minority of the bulk of pervasive transcription, although this is still the subject of intense debate.

Thus, the question that haunts the transcription field remains: what is the ultimate advantage of such a “consumeristic” strategy in producing in excess to discard a large fraction rather than more strictly targeting the production to the cellular needs. Many hypotheses have been formulated. Extended offer of (often superfluous) goods stimulates consumer demand, sometimes helping to identify successful products. This leads to economically justifying the waste. Extended expression of the information written on the DNA might be a more efficient way to evolve new functions, as mutations arising in transcribed regions have higher chances of being exposed to natural selection than mutations in silent regions. In this perspective, the bulk of pervasive transcripts might represent an important depository of RNA variants that have gone one step further on the path to function. Developing systems to increase the specificity of transcription would not only restrict such a source of variability, but might also have a high evolutionary cost, especially if the intrinsic propensity of the polymerase to initiate transcription is mechanistically linked to essential features of the transcription process (e.g., the stability of the elongation complex).

But overall, the real appearance of the Prince that hides behind the Beast of pervasive transcription remains quite mysterious, and is a challenge for many future studies.