Game of tails: Understanding the role of 3'-end processing of long noncoding RNAs during zebrafish development

The development of each of us starts at the time of fertilization, from a single cell formed after association of female and male gamete cells. At this stage, we inherit from our parents genetic material that is a special instruction determining what we will be like. In the later stages, this single cell, called zygote, undergoes multiple division, migration and precise differentiation to produce internal organs and finally a fully functional human body. Interestingly, despite the apparent differences in appearance and function, each cell in our body contains exactly the same genetic material. This unusual phenomenon is associated with strict control of the regions of our genome that are active at a given moment and place. These regions, called genes, contain the information needed for synthesis of all the particles essential for our body to function. For a long time, it was assumed that protein-coding genes (mRNAs) responsible for encoding proteins, an executive and building blocks of the cell, were the major part of the human genome. At the beginning of the 21st century, however, it turned out that these genes constitute only less than 2% of our DNA, while the rest of the so-called non-coding fraction can also be active, producing RNA molecules with important regulatory functions. The most numerous and most diverse class of non-coding protein genes are long non-coding RNAs (lncRNAs). Unlike mRNA, the activity of lncRNAs, and therefore their appearance and disappearance, is limited to the appropriate time points or to a specific cell type. This well-defined activity therefore proves that lncRNAs may play an important role in early embryogenesis. Indeed, many of the long non-coding RNAs have proven to be biologically important, for example in cell differentiation or organ formation, but despite exhaustive efforts, most of them still have unknown function. Additionally, the mechanisms determining their specific and strictly defined expression patterns are still unknown. 3'-end processing, especially the polyadenylation process, seems to play an important role in this aspect. This process relies on adding tails to RNA molecules consisting of repeated adenosine bases that stabilize the RNA molecule and protect it from degradation. The length and the exact composition of the poly(A) tails thus determine the "viability" of RNA in the cell. The process of mRNAs polyadenylation is well understood, but this aspect is unfortunately very often overlooked for lncRNAs, especially at early stages of embryonic development. For ethical reasons, it is impossible to study the above-mentioned mechanism in human embryos. Therefore, in order to understand human biology, animal organisms are used, which in many cases have provided important information about physiological and pathological processes. Zebrafish turned out to be extremely useful for research related to early organismal development. An important aspect that makes it a competitive model is the fact that the fertilization and embryogenesis of zebrafish takes place outside the mother's body, which greatly facilitates the acquisition of material for research, allows for constant monitoring of its developmental stages and significantly improves conducted functional research.

To understand the role of 3'-end processing of long non-coding RNAs in the early stage of zebrafish development, firstly we plan to establish an innovative method allowing for transcriptomic analysis focused specifically on lncRNAs. Taking into account the fact that lncRNAs are present in the cell in a very small amount, this method will take advantage of the CLS approach and specifically increase the abundance of these RNAs in the sample. CLS uses specially designed molecular probes capable of capturing target RNA from a pool of total RNA. In addition, the method of preparing samples for third generation sequencing will be optimized to allow the analysis of both polyadenylated and non-polyadenylated lncRNAs and at the same time to allow for an in-depth assessment of the length and composition of the identified poly(A) tails. The next step will be to identify factors potentially associated with differential poly(A) tail status, including the level of cellular abundance or specific expression patterns in the cell. Long non-coding RNAs diversify more rapidly during evolution than mRNAs, so only a small fraction of them is conserved between distant organisms such as humans and fish. It seems that these conserved lncRNAs may play an important biological function. The analysis of polyadenylation status will also be subjected to comparing evolutionarily stable lncRNAs to those that evolve more rapidly. Ultimately, based on the obtained results, lncRNAs targets will be selected for experimental validation. By alteration of polyadenylation process of selected long non-coding RNAs, their potential impact on the early embryonic development of zebrafish will be investigated. The obtained results will not only allow for a broader understanding of the lncRNA biology, such as factors regulating their presence in the cell or determining their function at early development. Most importantly, this knowledge may be crucial for designing RNA therapies aimed at treating diseases related to dysregulation and abnormal lncRNA function, including cancer. Unfortunately, these solutions are impossible to implement without a prior in-depth understanding of lncRNA biology, in which proposed research is expected to significantly help.