Mgr inż. Monika Kwiatkowska

Genomic characterization of long-noncoding RNAs in the zebrafish genome

Vertebrate genomes produce thousands of long noncoding RNAs (lncRNAs) - transcripts longer than 200 nucleotides with limited protein-coding potential. Despite a growing number of lncRNAs involved in crucial biological processes, over 97% of them remain functionally uncharacterized. Employment of animal models can aid in understanding the biological roles of lncRNAs, but the success of this exploration heavily depends on the quality of genome annotations. While zebrafish (*Danio rerio*) has emerged as a powerful and promising vertebrate model for exploring lncRNA biology, its genome annotation lags far behind that of humans or mice, significantly hindering its application. As part of this project, I aimed to create a comprehensive and accurate catalog of lncRNA genes in the zebrafish genome. To achieve this, I optimized the CapTrap-CLS protocol in zebrafish. CapTrap-seq is a full-length library preparation method that has demonstrated superiority in specifically enriching for 5' and 3'-complete transcripts while simultaneously and efficiently reducing the presence of ribosomal RNA (rRNA) molecules. The implementation of CapTrap-seq in zebrafish led to a more accurate annotation of transcription start sites (TSSs) for biologically relevant genes. Furthermore, to enhance the detection of lowly represented lncRNAs, I combined CapTrap-seq with the Capture Long-read Sequencing (CLS) approach—a targeted RNA sequencing method that integrates RNA capture with long-read Oxford Nanopore Technologies (ONT) sequencing. The RNA capture probes effectively enriched lncRNAs, significantly enhancing their representation in the post-capture libraries and extending the reference annotation for the zebrafish genome. Moreover, targeted RNA sequencing of human-mouse-zebrafish syntenic regions not only revealed new transcript isoforms for potentially functional lncRNAs, but also significantly enhanced the detection of novel genes in the intergenic space. This result underscores the importance of positional conservation as an effective strategy for the discovery of novel, potentially functional lncRNA loci. The improved zebrafish genome annotation developed during my PhD project offers a strong foundation for advancing the biological relevance of zebrafish as a model organism for studying the function of lncRNAs.