

## STRESZCZENIE

Długie niekodujące RNA (lncRNA, ang. long noncoding RNAs) stanowią intensywnie badaną, lecz wciąż stosunkowo słabo scharakteryzowaną klasę cząsteczek RNA licznie reprezentowanych w transkryptomie człowieka. Coraz więcej badań wskazuje na ich jednoznacznie znaczący udział w takich procesach komórkowych, jak transkrypcja, splicing, translacja, regulacja cyklu komórkowego czy apoptozy. Powiązano je również z szeregiem chorób człowieka. Mimo wszystko, wciąż tylko niewielka część tych cząsteczek została scharakteryzowana funkcjonalnie; analizy ewolucyjne lncRNA mogą w sposób znaczący przyczynić się do poprawy tej sytuacji, poprzez na przykład wskazanie tych lncRNA, które z ewolucyjnego punktu widzenia są najbardziej obiecującym przedmiotem badań. Niestety, w porównaniu z genami kodującymi białko, badania nad ewolucją lncRNA są utrudnione, chociażby ze względu na fakt, że nie kodują białka, a więc ograniczone jest znaczenie znanych nam skądinąd mechanizmów ewolucyjnych, które najlepiej opisują zmiany zachodzące w regionach kodujących genów. W związku z tym popularne oprogramowanie, które wykorzystywane jest w badaniach nad ewolucyjnymi związkami między genami (ortologi, paralogi) często nie znajduje zastosowania w przypadku lncRNA.

W odpowiedzi na te ograniczenia, przygotowano narzędzie lncEvo, potok analityczny wykonujący trzy następujące po sobie kroki: składanie transkryptomu ab initio w oparciu o dane RNA-Seq; identyfikacja lncRNA; badanie zależności ewolucyjnych dot. lncRNA pomiędzy dwoma wybranymi gatunkami. Ostatecznie, przygotowana została baza danych SyntDB, przechowująca informacje o zależnościach ewolucyjnych pomiędzy lncRNA człowieka oraz lncRNA zidentyfikowanymi u szeregu innych ssaków naczelnych; jest to odpowiedź na rosnące zapotrzebowanie na kolekcje ortologicznych oraz ewolucyjnie zakonserwowanych lncRNA.

## SUMMARY

Long non-coding RNAs (lncRNAs) are a class of intensely studied, yet enigmatic molecules that make up a substantial portion of the human transcriptome. A rapidly growing number of studies highlight their essential biological roles in processes such as transcription, splicing, translation, the cell cycle and apoptosis, protein localization, imprinting or stem cell pluripotency. They have also been implicated in human diseases. Still only a small fraction of currently known lncRNAs have been functionally characterized. To better understand the biology of lncRNAs, it is essential to perform a more in-depth study of their evolution in order to identify the most relevant lncRNAs for functional studies. In contrast to protein-encoding transcripts, however, these molecules do not show the strong sequence conservation that usually results from purifying selection; therefore, software that is typically used to resolve the evolutionary relationships of protein-encoding genes and transcripts is not applicable to the study of lncRNAs.

To address this issue, we developed lncEvo, a computational pipeline to perform three sequential steps: transcriptome assembly from RNA-Seq data; prediction of lncRNAs; and conservation study — a genome-wide comparison of lncRNA transcriptomes between two species of interest.

Keeping in mind the very limited availability of orthologous relationships for human lncRNAs and the importance of such data in functional studies, we searched for lncRNA orthologues across eleven primate species and built an online database dubbed SyntDB.