

## НЕКОДУЮЧІ ДІЛЯНКИ ДНК ЯК ОБ'ЄКТ ДОСЛІДЖЕННЯ

Вінницький національний технічний університет

### **Анотація**

*Розглянуто актуальність дослідження та комп'ютерного моделювання некодуючих ділянок ДНК, показано важливість виявлення інформативних ознак і паттернів в некодуючій послідовності ДНК для науки і медицини, обґрунтовано потребу в розробленні інформаційних інструментів обчислювального аналізу послідовностей ДНК, сформовано перелік інформаційних ресурсів для проведення досліджень в цій області.*

**Ключові слова:** ДНК; некодуюча нуклеотидна послідовність; обчислювальний аналіз; інформаційні ресурси; біоінформатика.

### **Abstract**

*The paper reveals the relevance of research and computer modeling of non-coding sections of DNA, it shows the importance of identifying informative features and patterns in non-coding DNA sequences for scientific and medical issues, and it substantiates the need for the development of information tools for computational analysis of DNA sequences. We also formed a list of information resources for conducting research in the area.*

**Keywords:** DNA; non-coding nucleotide sequence; computational analysis; information resources; bioinformatics.

### **Вступ**

Некодуючі ділянки ДНК (НКД ДНК) – це частини ДНК організму, які не кодують білкові послідовності. Вони складають 98–99 % загальної послідовності молекули ДНК [1]. Деякі некодуючі ДНК є джерелом синтезу некодуючих молекул РНК різних типів [2, 3]. Інші функціональні ділянки некодуючої ДНК включають регуляторні послідовності, які контролюють експресію генів; області кріплення гістонів та хроматинів, джерела реплікації ДНК, центромери і теломери [4–10]. Деякі некодуючі області вважають нефункціональними, такі як інтрони, псевдогени, міжгенна ДНК та вставні фрагменти транспозонів і вірусів. Дослідження останнього десятиліття показують, що некодуючі ділянки впливають на схильність організму до різних хвороб, зокрема до онкологічних [5].

Задачею нашого дослідження є обґрунтування необхідності вивчення некодуючих ділянок ДНК та формування набору ресурсів, які можуть бути інформативною базою.

### **Результати дослідження**

У зв'язку з великими обсягами і слабкою структурованістю НКД ДНК є складним об'єктом дослідження. Навіть виявлені та встановлені нуклеотидні послідовності НКД ДНК вимагають значних обчислювальних потужностей для їх комп'ютерного аналізу. Тому необхідно визначити підстави та потребу в таких дослідженнях. В результаті нашого аналізу сучасної наукової літератури [1–10] було визначено, що актуальними задачами обчислювального аналізу НКД ДНК є такі:

- ідентифікація та характеристика регуляторних елементів послідовності, які впливають на механізми регуляції генів;
- ідентифікація ділянок, відповідальних за розвиток захворювань та передбачення їхнього потенційного функціонального впливу, що має вирішальне значення для розуміння генетичної основи захворювань і розробки цільової (в т. ч. генної) терапії;
- виявлення еволюційного зв'язку між видами і вищими таксонами, відслідковування процесів паралельного перенесення генів;
- виявлення ділянок, відповідальних за зв'язування з хроматином, що забезпечує тривимірну структуру ДНК;
- виявлення ділянок, які виступають матрицями для синтезу різних типів некодуючих РНК. Ці молекули РНК беруть участь в різних біохімічних процесах, зокрема в посттранскрипційній регуляції генів, організації хроматину та підтримці стабільності геному, що також впливає на розвиток патологічних станів у клітинах і тканинах.

Вибір методів вивчення повинен відштовхуватись від конкретного завдання і мети дослідження, доступних типів даних і природи ділянок НКД ДНК, які аналізуються. Для одержання повного розуміння функціональності елементів і регуляторних ролей некодуєчих ділянок ДНК можна використати комбінацію методів, вибір яких є окремою науковою задачею.

Тому актуальною науково-прикладною задачею є розроблення інформаційної системи обчислювального аналізу некодуєчих ділянок ДНК, яка забезпечує вибір методів аналізу залежно від запиту користувача. Проміжними дослідницькими етапами в процесі розроблення такої системи можуть бути:

- проведення ґрунтового огляду і аналізу методів та систем обчислювальних досліджень НКД ДНК та розроблення їх класифікації;
- розроблення критеріїв пошуку функціональних НКД ДНК, шаблонів та правил для пошуку;
- розроблення архітектури системи обчислювального аналізу та дослідження її на тестових зразках.

Серед методів дослідження, які доцільно при цьому використовувати, можна навести такі: науковий пошук, методи біоінформатики, математичне моделювання, математична статистика, методи виявлення регуляторних елементів, аналізу консервативних послідовностей, епігеномного аналізу даних, інтегративного аналізу, функціональної анотації, методи об'єктно-орієнтованого і проблемно-орієнтованого програмування.

Важливою інформаційною базою дослідження, крім наукових журналів, монографій та патентів, будуть виступати відкриті бази даних, які містять інформацію про послідовності ДНК різних видів та індивідуальних організмів. Основними серед таких баз даних є: ресурси Національного центру біотехнологічної інформації NCBI [11, 12], бази даних Європейського інституту біоінформатики EMBL-EBI [13], зокрема геномний браузер проєкту Ensembl [14], Банк даних білків RCSB PDB [15], геномний браузер Каліфорнійського університету (UC Santa Cruz) [16], проєкт Європейського нуклеотидного архіву [17], Банк даних ДНК Японії [18] та інші спеціалізовані бази даних і проєкти. Перераховані ресурси містять великі обсяги даних про послідовності ДНК та їх описи у вільному доступі для науковців всього світу.

## Висновки

Обчислювальний аналіз некодуєчих ділянок ДНК має важливе значення для розкриття регуляторної картини геному, розуміння процесів генної регуляції, ідентифікації послідовностей, асоційованих із певними захворюваннями, розшифрування функціональних елементів і механізмів, що лежать в основі організації геному та процесів, на які впливають НКД ДНК. Це підтверджує актуальність досліджень некодуєчих ділянок ДНК та створення інформаційної системи для їх комп'ютерного аналізу. Інформаційними ресурсами при цьому, крім наукової літератури, є відкриті біоінформаційні бази даних, які містять послідовності ДНК різних організмів, доступні для обчислювального аналізу.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Alexander R. P., Fang G., Rozowsky J., Snyder M., Gerstein M. B. Annotating non-coding regions of the genome. *Nature Reviews Genetics*, 2010. Vol. 11 (8). P. 559–571.
2. Washietl S., Will S., Hendrix D. A., et al. Computational analysis of noncoding RNAs. *Wiley Interdisciplinary Reviews: RNA*, 2012 Vol. 3 (6). P. 759–778.
3. Mattick J. S., Amaral P. P., Carninci P., et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature Reviews Molecular Cell Biology*, 2023. Vol. 24. P. 430–447.
4. Doh S. T., Zhang Y., Temple M. H., Cai L. Non-coding sequence retrieval system for comparative genomic analysis of gene regulatory elements. *BMC Bioinformatics*. 2007. Vol. 8 (1). P. 1–10. <https://doi.org/10.1186/1471-2105-8-94>
5. Paul D. S., Soranzo N., Beck S. Functional interpretation of non-coding sequence variation: concepts and challenges. *Bioessays*. 2014. Vol. 36 (2). P. 191–199.
6. Zeng H., Gifford D. K. Predicting the impact of non-coding variants on DNA methylation. *Nucleic acids research*. 2017. Vol. 45 (11). P. e99. <https://doi.org/10.1093/nar/gkx177>
7. Krude H., Mundlos S., Øien N. C., Opitz R., Schuelke M. What can go wrong in the non-coding genome and how to interpret whole genome sequencing data. *Medizinische Genetik*, 2021. Vol. 33 (2). P. 121–131.

8. Ellingford J. M., Ahn J. W., Bagnall R. D. et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Medicine*. 2022. Vol. 14 (1). P. 73. <https://doi.org/10.1186/s13073-022-01073-3>
9. Gaulton K. J., Preissl S., Ren B. Interpreting non-coding disease-associated human variants using single-cell epigenomics. *Nature Reviews Genetics*, 2023. <https://doi.org/10.1038/s41576-023-00598-6>
10. Копилов К. В., Копилова К. В. Дослідження некодуючих ділянок ДНК геному різних видів тварин. *Розведення і генетика тварин*. 2023. Вип. 65. С. 184–190.
11. NCBI GenBank. URL : <https://www.ncbi.nlm.nih.gov/genbank/>
12. NCBI Reference Sequence database <https://www.ncbi.nlm.nih.gov/refseq/>
13. EMBL's European Bioinformatics Institute. URL : <https://www.ebi.ac.uk/>
14. Ensembl genome browser. <https://useast.ensembl.org/index.html>
15. Protein Data Bank RCSB PDB. URL : <https://www.rcsb.org/>
16. University of California Santa Cruz Genome Browser <https://genome.ucsc.edu/>
17. European Nucleotide Archive <https://www.ebi.ac.uk/ena/browser/home>
18. DNA Data Bank of Japan <https://www.ddbj.nig.ac.jp/index-e.html>

**Штофель Дмитро Хуанович** — науковий керівник, канд. техн. наук, доцент, доцент кафедри біомедичної інженерії та оптико-електронних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: [shtofel@vntu.edu.ua](mailto:shtofel@vntu.edu.ua)

**Каплунський Олександр Андрійович** — аспірант кафедри біомедичної інженерії та оптико-електронних систем, факультет інформаційних електронних систем, Вінницький національний технічний університет, Вінниця, e-mail: [okaplunskyi@vntu.edu.ua](mailto:okaplunskyi@vntu.edu.ua)

**Shtofel Dmytro Kh.** — Cand. Sc. (Eng), Associate Professor of the Department of Biomedical Engineering and Optoelectronic Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: [shtofel@vntu.edu.ua](mailto:shtofel@vntu.edu.ua)

**Kaplunskyi Oleksandr A.** — postgraduate student, Department of Biomedical Engineering and Optoelectronic Systems, Vinnytsia National Technical University, Vinnytsia, email: [okaplunskyi@vntu.edu.ua](mailto:okaplunskyi@vntu.edu.ua)