

СЕМЕНОВ Андрій

Вінницький національний технічний університет

<https://orcid.org/0000-0001-9580-6602>

e-mail: semenov.a.o@vntu.edu.ua

КУЗНЯК Владислав

Вінницький національний технічний університет

<https://orcid.org/0009-0001-1775-420X>

e-mail: kuzniakvl@gmail.com

GRN-ІНФОРМОВАНИЙ CELLFLOW: ПОКРАЩЕННЯ РЕКОНСТРУКЦІЇ ТРАЄКТОРІЙ КЛІТИННИХ СТАНІВ ЗА ДОПОМОГОЮ БІОЛОГІЧНИХ РЕГУЛЯТОРНИХ МЕРЕЖ

У цій статті представлено результати дослідження інтеграції інформації про генно-регуляторні мережі (GRN) у комп'ютерні моделі для відновлення клітинних траєкторій, зокрема з акцентом на покращення фреймворку CellFlow. Генно-регуляторні мережі описують взаємопов'язану систему генів та регуляторних елементів — включно з транскрипційними факторами та сигнальними шляхами — які координують експресію генів і забезпечують правильну регуляцію процесів підтримання клітинної ідентичності, керування диференціацією та адаптивних відповідей на зміни середовища. Карта таких взаємозв'язків формує основу для розуміння того, як патерни активності генів визначають клітинну поведінку та переходи між станами. У цьому дослідженні ми представляємо GRN-інформований CellFlow — розширення наявної моделі CellFlow, що явно враховує регуляторні взаємодії між генами. На відміну від початкового підходу, який розглядає гени як незалежні ознаки, запропонований метод інтегрує відомі ген-генні залежності для керування реконструкцією траєкторій клітинних станів. Для цього було сформовано GRN-матрицю на основі даних клітин рибальської зебри: транскрипційні фактори визначалися за допомогою анотацій UniProtKB, а взаємодії між ними та іншими генами оцінювалися з використанням кореляційного аналізу та алгоритму GRNBoost2. Побудований лапласіан мережі застосовано як регуляризатор у процесі навчання моделі, що дозволило врахувати структуровані залежності між генами. Результати експериментів показали, що інтеграція GRN дещо знижує показники функції втрати порівняно з класичною конфігурацією CellFlow, проте підвищує біологічну інтерпретованість реконструйованих траєкторій. Це свідчить про перспективність поєднання структурованої мережевої інформації з алгоритмічними підходами до інференції клітинних траєкторій.

Ключові слова: траєкторії клітинних станів, одноклітинний аналіз (single-cell), ген-генні взаємодії, GRN (генно-регуляторні мережі), CellFlow, нейронна мережа.

SEMENOV Andrii, KUZNIAK Vladyslav

Vinnitsia National Technical University

GRN-INFORMED CELLFLOW: ENHANCING CELL STATE TRAJECTORY INFERENCE WITH BIOLOGICAL REGULATORY NETWORKS

This paper presents a study on integrating gene regulatory network (GRN) information into computational models for reconstructing cellular trajectories, with a particular focus on enhancing the CellFlow framework. Gene regulatory networks describe interconnected systems of genes and regulatory elements — including transcription factors and signaling pathways — that coordinate gene expression and ensure proper regulation of processes such as cell identity maintenance, lineage differentiation, and adaptive responses to environmental changes. Mapping these interactions provides a foundation for understanding how gene activity patterns shape cellular behavior and transitions between states. In this work, we introduce GRN-informed CellFlow, an extension of the original CellFlow model that explicitly incorporates regulatory dependencies between genes. Unlike the baseline approach, which treats genes as independent features, the proposed method integrates known gene–gene relationships to guide trajectory reconstruction. To achieve this, we constructed a GRN matrix using zebrafish cell data: transcription factors were identified via UniProtKB annotations, while their interactions with target genes were inferred using correlation analysis and the GRNBoost2 algorithm. The resulting network Laplacian was employed as a regularizer during model training, enabling CellFlow to account for structured dependencies between genes. Experimental results showed that GRN integration slightly worsens the loss function compared to the classical CellFlow configuration, yet improves the biological interpretability of reconstructed trajectories. These findings highlight the potential of combining structured network information with algorithmic approaches to cell trajectory inference.

Keywords: cell state trajectories, single-cell analysis, gene–gene interactions, GRN (gene regulatory networks), CellFlow, neural network.

Стаття надійшла до редакції / Received 01.08.2025

Прийнята до друку / Accepted 26.08.2025

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Розвиток технологій секвенування РНК на рівні окремих клітин (single-cell RNA sequencing, scRNA-seq) значно розширив можливості дослідження клітинної гетерогенності, процесів розвитку та диференціації. Тепер можливо отримати детальні профілі експресії генів для кожної окремої клітини, що дозволяє

відстежувати її стан та потенційні траєкторії розвитку. Аналіз таких даних дозволяє виявляти нові клітинні підтипи, досліджувати еволюцію клітинних популяцій та вивчати механізми реакції клітин на різні сигнали або стани захворювання.

Існуючі методи для передбачення траєкторій розвитку клітини, наприклад CellFlow [1], зазвичай розглядають експресію генів як незалежні ознаки. Таке припущення спрощує модель та дозволяє ефективно обробляти великі обсяги даних, але ігнорування взаємозв'язків між генами створює серйозні обмеження. Насправді гени не функціонують ізольовано: їхня активність визначається складними каскадами регуляції, де один ген або транскрипційний фактор може активувати або пригнічувати експресію десятків інших генів. Ігнорування таких зв'язків може призвести до того, що передбачені траєкторії клітинного розвитку будуть статистично коректні, але не відображатимуть справжню біологічну динаміку.

Генетичні регуляторні мережі (Gene Regulatory Networks, GRN) описують складну взаємодію між генами, транскрипційними факторами та іншими регуляторними елементами, які спільно визначають клітинну ідентичність, стан клітини та її здатність до диференціації. Вони є фундаментом для розуміння механізмів, що керують клітинною поведінкою, і дозволяють ідентифікувати критичні вузли та шляхи, які визначають ключові переходи між клітинними станами. Інтеграція знань про GRN у моделі прогнозування дозволяє враховувати біологічно значущі залежності між генами, підвищуючи точність та інтерпретованість передбачень.

Запропонований підхід полягає у безпосередньому включенні GRN у процес тренування моделі CellFlow. Це дозволяє моделі не просто передбачати зміни експресії генів у часі, а й враховувати їх взаємозалежності, що відображають реальні механізми регуляції клітинного розвитку. Інтеграція GRN забезпечує кілька ключових переваг:

1. Більш точне прогнозування траєкторій клітинного розвитку: модель враховує, що зміни в експресії одного гена можуть спричинити каскадні ефекти в інших генах.
2. Покращена інтерпретованість: результати можна аналізувати у контексті відомих біологічних шляхів і регуляторних механізмів, що дозволяє краще розуміти, які взаємодії визначають клітинні стани та переходи.

Таким чином, поєднання методів моделювання траєкторій клітинного розвитку з інформацією про GRN створює потужний інструмент для дослідження динаміки клітин на рівні окремих генів і забезпечує більш точні, біологічно обґрунтовані та інтерпретовані результати, ніж підходи, які ігнорують регуляторні взаємозв'язки.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Класичні підходи до виведення траєкторій розвитку клітин (Trajectory Inference, TI), які реконструюють динамічні процеси шляхом впорядкування клітин у псевдочасі, були детально систематизовані в оглядах, таких як [2]. Ці методи надали фундамент для розуміння клітинної гетерогенності, але їх точність і біологічна інтерпретованість часто обмежувалися через ігнорування складних регуляторних взаємодій між генами. У відповідь на це були розроблені більш сучасні підходи. Наприклад, PAGA [3] використовує графову абстракцію для побудови інтерактивної карти клітинних станів, а методи, засновані на RNA velocity, зокрема scVelo [4], дозволяють визначати напрямки переходів між клітинними станами на основі динаміки сплайсингу РНК.

Разом із розвитком методів TI активно зростає інтерес до моделювання генних регуляторних мереж (GRN), які дозволяють враховувати взаємозв'язки між генами та транскрипційними факторами. Методи на кшталт SCENIC [5] спрямовані на ідентифікацію регулонів і ключових регуляторів, що визначають клітинну ідентичність, забезпечуючи більш глибоке розуміння механізмів, які лежать в основі клітинної диференціації та переходів між станами. Інтеграція знань про GRN у моделі TI відкриває перспективи для побудови більш точних, інтерпретованих і біологічно обґрунтованих траєкторій розвитку клітин.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою роботи є дослідження поєднання підходів до виведення траєкторій розвитку клітин шляхом вдосконалення моделі CellFlow для "направлення" процесу навчання, надаючи моделі інформацію про генні регуляторні мережі. Це уможливить обмежити простір можливих траєкторій лише біологічно правдоподібними шляхами, що значно підвищить точність та механістичну обґрунтованість отриманих моделей клітинної динаміки.

Завдання досліджень:

- побудувати матрицю, яка відображатиме регуляторну мережу генів (Gene Regulatory Network, GRN) на основі наявного набору даних, використовуючи інформацію про кореляції та взаємодії між генами;
- інтегрувати GRN у процес тренування моделі для корекції траєкторій клітинної диференціації, зокрема у випадках, коли профілі експресії генів не відображають взаємодії, передбаченої GRN.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

У цьому дослідженні для реконструкції клітинних траєкторій було використано підхід на основі CellFlow, який застосовує Continuous Normalizing Flows (CNF) [6, 7] для моделювання динаміки клітинних станів у латентному просторі. Основна ідея CellFlow полягає в тому, щоб представити розподіл клітинних станів у вигляді диференційованого рівняння, що описує еволюцію щільності ймовірності $p(z, t)$ у часі t , де $z \in \mathbb{R}^G$ – простір експресії генів. Архітектура CellFlow складається з глибоких нейронних мереж, які параметризують скорочення або розширення щільності через поле швидкості (velocity field) та дозволяють навчатися шляхом оптимізації дифеоморфізму, трансформуючи простий початковий розподіл (наприклад, стандартний гаусівський) у складний розподіл кінцевих клітинних станів. Такий підхід забезпечує високу гнучкість у моделюванні нелінійних взаємодій клітинних станів, дозволяючи ефективно реконструювати як безперервні, так і розгалужені траєкторії розвитку.

Для інтеграції біологічної інформації про регуляторні взаємодії між генами було розроблено GRN-інформований модуль. Першим кроком стало створення матриці GRN, що описує потенційні взаємодії між транскрипційними факторами (TF) та усіма генами у даних. Для цього на основі датасету секвенованих клітин риби зебри (zebrafish scRNA-seq) [8] було визначено гени, які можуть належати до транскрипційних факторів, використовуючи інформацію з бази UniProtKB [9], що містить анотації щодо функцій білків і ролі генів у регуляції транскрипції. Наступним кроком стала оцінка потенційних взаємодій між визначеними транскрипційними факторами та усіма іншими генами. Для цього застосовувалися два підходи: кореляційний аналіз та алгоритм GRNBoost2 [10], який на основі градієнтного бустінгу дозволяє інферувати регуляторні взаємозв'язки, забезпечуючи більш точну і стабільну оцінку важливості кожної взаємодії в мережі. Приклад генів визначених як фактори транскрипції показаний на рис. 1.

```

1  spi1b
2  spi1b
3  runx1t1
4  tcf7l2
5  spi1b
6  runx1t1
7  tcf7l2
8  tcf12
9  esrrga
10 tshz1
11 foxp2
12 zbtb20
13 lhx6
14 tbx16
15 tcf7l1b
    
```

Рис. 1. Приклад генів визначених як фактори транскрипції

Отримані значення взаємодій були використані для побудови лапласіану графу залежностей на основі GRN, який представляє регуляторну мережу у вигляді матриці регуляризації. Лапласіан відіграє роль регуляризатора у функції втрат, який вводиться при тренуванні моделі CellFlow для прогнозування латентного вектора μ . Завдяки цьому модель під час оптимізації навчалася не лише відновлювати розподіл клітинних станів, але й враховувати структуровані залежності між генами, закладені у GRN. Приклад результатів моделі CellFlow без інтеграції регуляторної функції на короткому інтервалі навчання показаний на рис. 2.

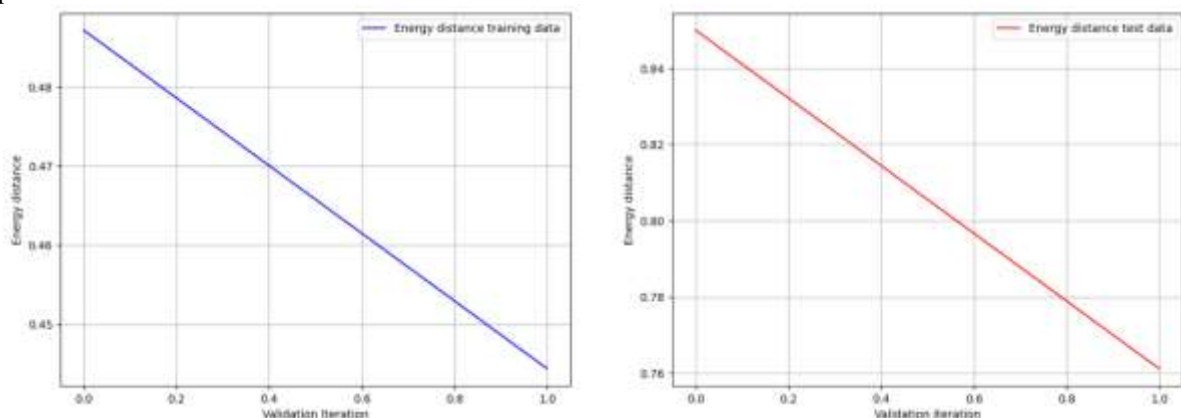


Рис. 2. Приклад результатів моделі CellFlow без інтеграції регуляторної функції на короткому інтервалі навчання

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

У цьому дослідженні розроблено підхід GRN-інформованої інференції клітинних траєкторій на основі CellFlow, який інтегрує інформацію про регуляторні взаємодії між генами для покращення реконструкції траєкторій із даних однохвилинного секвенування РНК (scRNA-seq). Показано, що класичний CellFlow, хоча й ефективний у моделюванні нелінійних траєкторій клітинних станів, не враховує залежності між генами, що обмежує біологічну інтерпретованість його прогнозів.

Щоб подолати це обмеження, було створено матрицю GRN на основі даних клітин рибальської зебри. Транскрипційні фактори визначалися за допомогою анотацій UniProtKB, а взаємодії між TF та іншими генами оцінювалися за допомогою кореляційного аналізу та алгоритму GRNBoost2. Побудований лапласіан регуляторної мережі було використано як регуляризатор під час тренування моделі, що дозволило CellFlow враховувати структуровані взаємозв'язки між генами.

Для оцінки впливу GRN-інформованої регуляризації було проведено експеримент, у якому цей підхід застосовувався під час навчання моделі. Паралельно виконувалося навчання на стандартних налаштуваннях, відповідно до оригінальної процедури CellFlow. Порівняння результатів показало, що модель із інтегрованою інформацією про GRN продемонструвала дещо гірші показники функції втрати порівняно з базовою конфігурацією без GRN. Це свідчить про те, що поточна реалізація інтеграції ген-генних взаємодій ще не забезпечує покращення числової оптимізації, хоча вона може залишатися потенційно корисною для підвищення інтерпретованості реконструйованих траєкторій.

У подальших дослідженнях доцільно розглянути зміну архітектури нейронної мережі на графову, щоб краще вивчити закономірності взаємодії генів під час навчання.

Література

1. Klein, D., Fleck, J. S., Bobrovskiy, D., Zimmermann, L., Becker, S., Palma, A., Dony, L., Tejada-Lapuerta, A., Huguet, G., Lin, H.-C., Azbukina, N., Sanchís-Calleja, F., Uscidda, T., Szalata, A., Gander, M., Regev, A., Treutlein, B., Camp, J. G., & Theis, F. J. (2025). CellFlow enables generative single-cell phenotype modeling with flow matching. *Cold Spring Harbor Laboratory*. <https://doi.org/10.1101/2025.04.11.648220>
2. Deconinck, L., Cannoodt, R., Saelens, W., Deplancke, B., & Saeys, Y. (2021). Recent advances in trajectory inference from single-cell omics data. *Current Opinion in Systems Biology*, 27, 100344. <https://doi.org/10.1016/j.coisb.2021.05.005>
3. Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L., & Theis, F. J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, 20(1). <https://doi.org/10.1186/s13059-019-1663-x>
4. Bergen, V., Lange, M., Peidli, S., Wolf, F. A., & Theis, F. J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38(12), 1408–1414. <https://doi.org/10.1038/s41587-020-0591-3>
5. Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature Methods*, 14(11), 1083–1086. <https://doi.org/10.1038/nmeth.4463>
6. Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., & Le, M. (2022). Flow Matching for Generative Modeling. arXiv. URL: <https://arxiv.org/abs/2210.02747>
7. Tong, A., et al. (2023). Improving and generalizing flow-based generative models with minibatch optimal transport. arXiv. URL: <https://arxiv.org/abs/2302.00482>
8. Saunders, L. M., Srivatsan, S. R., Duran, M., Dorrity, M. W., Ewing, B., Linbo, T. H., Shendure, J., Raible, D. W., Moens, C. B., Kimelman, D., & Trapnell, C. (2023). Embryo-scale reverse genetics at single-cell resolution. *Nature*, 623(7988), 782–791. <https://doi.org/10.1038/s41586-023-06720-2>
9. Hauschild, A.-C., Pastrello, C., Ekaputeri, G. K. A., Bethune-Waddell, D., Abovsky, M., Ahmed, Z., Kotlyar, M., Lu, R., & Jurisica, I. (2022). MirDIP 5.2: tissue context annotation and novel microRNA curation. *Nucleic Acids Research*, 51(D1), D217–D225. <https://doi.org/10.1093/nar/gkac1070>
10. Moerman, T., Aibar Santos, S., Bravo González-Blas, C., Simm, J., Moreau, Y., Aerts, J., & Aerts, S. (2018). GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 35(12), 2159–2161. <https://doi.org/10.1093/bioinformatics/bty916>

References

1. Klein, D., Fleck, J. S., Bobrovskiy, D., Zimmermann, L., Becker, S., Palma, A., Dony, L., Tejada-Lapuerta, A., Huguet, G., Lin, H.-C., Azbukina, N., Sanchís-Calleja, F., Uscidda, T., Szalata, A., Gander, M., Regev, A., Treutlein, B., Camp, J. G., & Theis, F. J. (2025). CellFlow enables generative single-cell phenotype modeling with flow matching. *Cold Spring Harbor Laboratory*. <https://doi.org/10.1101/2025.04.11.648220>
2. Deconinck, L., Cannoodt, R., Saelens, W., Deplancke, B., & Saeys, Y. (2021). Recent advances in trajectory inference from single-cell omics data. *Current Opinion in Systems Biology*, 27, 100344. <https://doi.org/10.1016/j.coisb.2021.05.005>

3. Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L., & Theis, F. J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, 20(1). <https://doi.org/10.1186/s13059-019-1663-x>
4. Bergen, V., Lange, M., Peidli, S., Wolf, F. A., & Theis, F. J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38(12), 1408–1414. <https://doi.org/10.1038/s41587-020-0591-3>
5. Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thi, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature Methods*, 14(11), 1083–1086. <https://doi.org/10.1038/nmeth.4463>
6. Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., & Le, M. (2022). Flow Matching for Generative Modeling. arXiv. URL: <https://arxiv.org/abs/2210.02747>
7. Tong, A., et al. (2023). Improving and generalizing flow-based generative models with minibatch optimal transport. arXiv. URL: <https://arxiv.org/abs/2302.00482>
8. Saunders, L. M., Srivatsan, S. R., Duran, M., Dorrity, M. W., Ewing, B., Linbo, T. H., Shendure, J., Raible, D. W., Moens, C. B., Kimelman, D., & Trapnell, C. (2023). Embryo-scale reverse genetics at single-cell resolution. *Nature*, 623(7988), 782–791. <https://doi.org/10.1038/s41586-023-06720-2>
9. Hauschild, A.-C., Pastrello, C., Ekaputeri, G. K. A., Bethune-Waddell, D., Abovsky, M., Ahmed, Z., Kotlyar, M., Lu, R., & Jurisica, I. (2022). MirDIP 5.2: tissue context annotation and novel microRNA curation. *Nucleic Acids Research*, 51(D1), D217–D225. <https://doi.org/10.1093/nar/gkac1070>
10. Moerman, T., Aibar Santos, S., Bravo González-Blas, C., Simm, J., Moreau, Y., Aerts, J., & Aerts, S. (2018). GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 35(12), 2159–2161. <https://doi.org/10.1093/bioinformatics/bty916>