

MCCMB 2025

Сборник тезисов

12-й Московской
конференции по
вычислительной
молекулярной биологии
(MCCMB)



Проблема интерпретации данных SAMap

А.И. Агламазова, Г.А. Шиловский, В.А. Любецкий

Институт проблем передачи информации aglamazova.an@iitp.ru

Метод SAMap представляет собой современный инструмент для сопоставления данных одноядерной (snRNA-seq) и одноячейковой (scRNA-seq) транскриптомики, который активно используется для создания межвидовых клеточных атласов и изучения эволюционных процессов. Его применение позволяет объединять данные из различных экспериментов, сопоставлять клеточные популяции между видами и выявлять молекулярные механизмы, лежащие в основе клеточной гетерогенности. Однако интерпретация результатов, полученных с помощью SAMap, ограничена рядом трудностей, включая зависимость от качества данных, сложность аннотации кластеров, трудности работы с редкими клеточными типами, ограниченность межвидового анализа и значительные вычислительные затраты.

Одной из ключевых проблем является недостаточная точность визуализации и интерпретации результатов сопоставления. Перекрывание кластеров в многомерном пространстве затрудняет точное определение клеточных типов, особенно в условиях высокой гетерогенности данных. Кроме того, редкие клеточные популяции часто теряются при кластеризации или объединяются с более крупными группами, что приводит к потере биологически значимой информации. Эти проблемы усиливаются при анализе данных из эволюционно удалённых видов, где различия в геномных последовательностях и экспрессии генов могут снижать точность сопоставления клеточных типов.

Предлагаемая комбинация методов SAMap и PHATE (Potential of Heat-diffusion for Affinity-based Transition Embedding) способна значительно улучшить интерпретацию результатов. PHATE — это метод редукции размерности и визуализации, который сохраняет как локальные, так и глобальные структуры

данных, делая его особенно полезным для анализа траекторий клеточных состояний и сложных многомерных отношений между клетками. На основе сочетания SAMar и PHATE нами разработан pipeline, который:

PHATE создает латентное пространство, которое используется SAMar для сопоставления клеточных типов, что помогает смягчить влияние шума и улучшить точность анализа.

PHATE обеспечивает гладкое и интуитивно понятное представление сопоставленных клеточных типов, особенно в задачах, связанных с траекториями клеточных изменений.

PHATE выделяет локальные структуры данных, что позволяет эффективно работать с редкими клеточными популяциями, которые традиционно теряются при стандартных методах обработки.

Комбинация SAMar и PHATE повышает биологическую обоснованность и воспроизводимость результатов. Оптимизация алгоритма визуализации UMAP, включая использование PHATE, уменьшает перекрытие кластеров, что упрощает определение клеточных типов, в том числе в далеко отстоящих на эволюционном древе таксонах Metazoa.

Предложенная интеграция может улучшить анализ данных из эволюционно удалённых видов, поскольку PHATE сохраняет глобальные структуры, что делает результаты сопоставлений SAMar более интерпретируемыми.

Метод SAMar в сочетании с PHATE открывает новые возможности для изучения клеточных типов и их эволюции, улучшая как точность сопоставления, так и удобство интерпретации. Эти подходы частично решают проблему определения родственных клеточных типов в наборе кластеризаций разных видов.

Исследование выполнено за счёт гранта Российского научного фонда № 24-44-00099, <https://rscf.ru/project/24-44-00099/>

Список литературы

1. Tarashansky, Alexander J., et al. "Mapping single-cell atlases throughout Metazoa unravels cell type evolution." *Elife* 10 (2021)
2. Moon, Kevin R., et al. "Visualizing structure and transitions in

high-dimensional biological data." Nature Biotechnology" 37.12 (2019)

3. X. Zhang, Z. Li, J. Chen, W. Yang, X. He, P. Wu, F. Chen, Z. Zhou, Ch. Ren, Y. Shan, X. Wen, V.A. Lyubetsky, L.Yu. Rusin, X. Chen, J.-R. Yang. Stereotyped subclones revealed by high-density single-cell lineage tracing support robust development. Advanced Science, Apr 30 2025, Art. 2406208.

OmiXAI: ансамблевый XAI-пайплайн интерпретации моделей глубинного обучения для омиксных данных

Алаева А.И., Лаптева А., Михайловская Н.Н., Мальков В.В.,
Боревский А.О., Попцова М.С.

*Международная лаборатория Биоинформатики ИИИиЦН
ФКН НИУ ВШЭ, Россия, 101000, г. Москва, ул. Мясницкая, д. 20
a.ameli.ig@gmail.com
mpoptsova@hse.ru*

Методы глубинного обучения в настоящее время становятся основным инструментом для анализа геномных данных. Однако качество работы таких моделей напрямую зависит от объема и характера данных, доступных для обучения. В последние годы наблюдается тенденция к использованию мультиомиксных данных, включающих различные уровни биологической информации: геномные, транскриптомные, эпигеномные, протеомные, метаболомные и различные другие. При достижении высокой точности модели, обученной на подобных данных, возникает важный вопрос: какие именно признаки и биомолекулярные факторы лежат в основе её предсказательной способности?

Методы объяснимого искусственного интеллекта (XAI, Explainable AI) помогают выявить самые важные признаки. Все XAI-подходы можно условно разделить на две категории: зависящие от модели (model-aware) и независящие от модели (model-agnostic). Вторые обычно основаны на комбинаторных нарушениях входных признаков и требуют значительных