doi 10.18699/bgrs2026-270

Предсказание и анализ клеточных траекторий дифференцировки от стволовых клеток до легочных предшественников

Витрещак А.Г.[1]*, Любецкий В.А.[1]
[1] *Институт проблем передачи информации им. А.А. Харкевича РАН, Москва, Россия*
* *l_veter@mail.ru*

Ключевые слова: биоинформатика; транскриптомика одиночных клеток; клеточные траектории дифференцировки; стволовые клетки; легочные предшественники

Мотивация и цель: Клеточная дифференцировка является фундаментальным биологическим процессом и в индивидуальном развитии формирует организм. Специализация клетки меняется в эмбриональном развитии, при регенерации, репрограммировании и многих болезнях, например, раке. Транскриптомика одиночных клеток выявляет клеточную гетерогенность с высоким разрешением, что позволяет реконструировать клеточные траектории. Дифференцировка стволовых клеток в легочные предшественники человека исследуется для понимания механизмов развития и заболеваний легких, которое имеют большой потенциал для регенеративных методов лечения. Протоколы in vitro направленной дифференцировки имитируют этапы органогенеза легких in vivo: дефинитивной энтодермы (DE), энтодермы передней кишки (FE), легочных предшественников (PLP) и легочных органоидов. Особый интерес представляют исследования ранних стадий развития, на которых, в частности, формируется FE, поскольку она дает начало нескольким органам, включая легкие, печень, поджелудочную и щитовидную железы. Продолжая нашу работу [1], мы использовали данные scRNA-seq направленной дифференцировки in vitro от эмбриональных стволовых клеток человека (HESC) до легочных предшественников с целью определения клеточных траекторий, анализа типов клеток, определения драйверные генов.
Методы и алгоритмы: Для предварительных вычислений, а также кластеризации, аннотации, выведения клеточных траекторий и нахождения макросостояний мы использовали усовершенствованную нами экосистему Scanpy/CellRank2/AnnData, в которой вычисляются переходная матрица вероятностей, псеводовремя, макросостояния, судьбы клеток и драйверные гены. Количество макросостояний обычно рассчитывается по критериям minChi и crisspnes. Однако они недопредсказывают число макросостояний, что видно, например, на данных scRNA-seq кроветворения. И пользователю приходится устанавливать его вручную. Чтобы определить число макросостояний, мы разработали два формальных правила: вычисление первого индекса интервала, на котором вычисляется minChi, и использование спектра переходной матрицы вероятностей, а именно, его точки перегиба. Эти правила успешно тестировались нами на данных scRNA-seq кроветворения человека и др. Мы адаптировали модель Harmony для анализа гетерогенного (HESC) и гомогенных (CBRAD) образцов, что позволило улучшить пакетную коррекцию и устранить все артефакты кластеризации, присутствовавшие в предыдущей работе [1].
Результаты: Мы использовали данные scRNA-seq протокола направленной дифференцировки in vitro от HESC к PLP [1] для построения клеточных траекторий. Данные были получены для четырех образцов: A1-CBRAD5, G2-CBRAD5, G11-CBRAD5 и HESC. Первые три (CBRAD5) соответствуют экспериментальному протоколу дифференцировки in vitro и взяты на 10-й день дифференцировки. В HESC дифференцировки клеток не происходило, данные scRNA-seq также получены на 10-й день. Нами найдены 15 кластеров, пять из которых соответствуют стволовым клеткам (H1-H5), десять – дифференцированным клеткам (R1, R2, C5-C11). Кластеры стволовых клеток расположены отдельным суперкластером от суперкластера, соответствующеего образцам CBRAD5. Наш анализ клеточных траекторий и макросостояний показал, что дифференцировка начинается от кластера стволовых клеток H4, далее проходит через слабо дифференцированные клетки (R1, R2, C8) и через некоторые промежуточные этапы до терминальных макросостояний (C10 и C7), которые располагаются наиболее удаленно по ходу псевдовремени. Прослеживается бифуркация: разделение клеточной дифференцировки на две ветки. Легочными предшественниками являются кластеры C9 и расположенные рядом кластеры C5_1 и C5_2, в которых экспрессируется маркерный ген NKX2-1. Известно, что эпителиальные клетки легких, полученные из iPSC, происходят от клеток NKX2-1+, [2]. Удивительно, что в кластере C7, включающем терминальное макросостояние с высоким индексом стабильности (SI), экспрессируются маркерные гены APOA2, TTR и др., секретируемые в печени. По нашему мнению, наличие отдельного кластера клеток печени со стабильным терминальным макросостоянием указывает, на устойчивую дифференцировку в клетки печени при легочном протоколе и не требует экзогенной стимуляции FGF2. Учитывая сходство протоколов легких и печени в фазе дефинитивной энтодермы, мы предполагаем, что ранние сигналы дифференцировки в клетки печени начинаются на этой фазе. Известно, что ген Isl1 является ключевым регулятором гена Nkx2-1 на ранней стадии развития FE и легочная траектория дифференцировки начинается во время этой фазы и, предположительно, позже, чем траектория дифференцировки в клетки печени (Рис. 1).

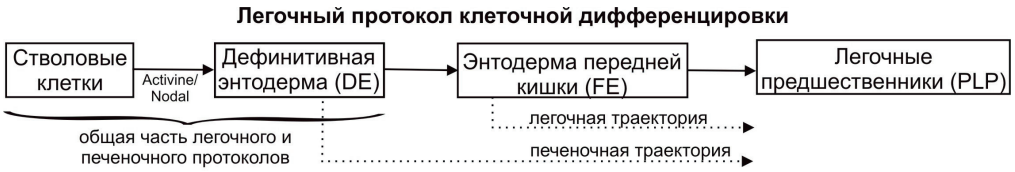


Рисунок 1. Общая схема дифференцировки по легочному протоколу.

Выводы: Показано наличие бифуркации: двух путей дифференцировки от стволовых клеток до легочных предшественников. Аннотированы кластеры, принадлежащие различным типам клеток и степень их зрелости. Для терминальных макросостояний найдены драйверные гены. При легочном протоколе направленной дифференцировки обнаружена значительная фракция клеток печени; мы предполагаем, что ранние сигналы дифференцировки в клетки печени появляются на фазе дефинитивной энтодермы.
Финансирование: Исследование поддержано грантом РНФ № 24-44-00099.

Prediction and analysis of cellular differentiation trajectories from stem sells to lung progenitors

Vitreschak A.G.[1]*, Lyubetsky V.A.[1]
[1] *Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Moscow, Russia*
** l_veter@mail.ru*

Key words: bioinformatics; scRNA-seq; cellular differentiation trajectories; stem sells; lung precursors

Motivation and Aim: Cellular differentiation is one of the fundamental processes in biology and plays an important role in the organism development. Cell transitions can occur during embryonic development as well as regeneration, reprogramming and cancer. Single-cell RNA sequencing uncover cellular heterogeneity at high resolution that allows us to reconstruct differentiation trajectories. Research in the field of stem cells-derived differentiation to human lung progenitors is essential for understanding lung development and disease and has potential for future regenerative therapies. in vitro lung-directed differentiation protocols are known to mimic the stages of the lung organogenesis in vivo: definitive endoderm (DE), foregut endoderm (FE), lung progenitors (PLP) and lung organoids. Of particular interest are studies in the earlier stages of development, in particular, the formation of FE, as it gives rise to several organs, including lung, liver, pancreas and thyroid. Continuing our previous work [1], we used scRNA-seq data on in vitro directed differentiation from human embryonic stem cells (HESC) to lung progenitors to infer cellular differentiation trajectories, analyze cell types and find driver genes.
Methods and Algorithms: We used the Scanpy/CellRank2/AnnData ecosystem for preliminary calculations, as well as clustering, annotation, inference of cellular trajectories, and finding macrostates. Palantir was used to calculate a pseudotime. CellRank2 [2] uses various methods (kernels) to calculate a cell-transition matrix derived from a Markov chain. Estimator (GPCCA) analyzes the cell-transition matrix to infer macrostates as well as fate probabilities and lineage driver genes. CellRank2 calculates the number of macrostates using the minChi criterion with the crispness ξ. However, this criterion underestimates the number of macrostates, for example, for hematopoietis scRNA-seq data, and the user has to set their number manually. To determine the number of macrostates, we developed two formal rules: calculating the first index of the interval on which the minChi is calculated and using the spectrum of the transition probability matrix, namely, the inflection point. Both rules were successfully tested on human hematopoiesis scRNA-seq data. The Harmony model was adapted to analyze heterogeneous (HESC) and homogeneous (CBRAD) samples, which allowed us to improve the batch-correction and eliminate all clustering artifacts present in the previous work [1].
Results: We used scRNA-seq data on in vitro directed differentiation from HESC to PLP to infer cellular differentiation trajectories. Data were obtained for four samples: A1-CBRAD5, G2-CBRAD5, G11-CBRAD5, and HESC. The first three (CBRAD) correspond to the experimental in vitro differentiation protocol and were taken on the 10th day of differentiation. In the last HESC sample, cell differentiation did not occur, scRNA-seq data were also obtained on day 10. 15 clusters were found, five of which correspond to stem cells (H1-H5), ten to differentiated cells (R1, R2, C5-C11). The stem cell clusters are located in a separate supercluster relative to the supercluster corresponding to the CBRAD5 samples. Analysis of cellular trajectories and macrostates showed that differentiation starts from the H4 stem cell cluster, then through weekly differentiated cells (R1, R2, C8), some intermediate states to terminal macrostates (C10 and C7), which are located most distantly along the pseudotime. The division of cellular differentiation into two branches (bifurcation) is visually traced. The lung precursors are clusters C9, and nearby clusters C5_1 and C5_2, in which the marker gene NKX2-1 is expressed. It was shown earlier that iPSC-derived lung epithelial cells originate from lung NKX2-1+ progenitors [3]. Surprisingly, the C7 cluster, which contains the terminal macrostate with a high stability index (SI), expresses the marker genes APOA2, TTR and others secreted in the liver. The presence of a separate liver cluster with stable terminal macrostate clearly indicates that liver differentiation is resistant to the lung-directed protocol and do not require exogenous stimulation by FGF2. Given the similarity of lung and liver protocols in the definitive endoderm phase, we assume that early signals of differentiation into liver cells begin in this phase. Isl1 is

shown to be a key regulator of Nkx2-1 in the early foregut endoderm phase and the lung trajectory emergences during this phase, presumably later than the liver trajectory (Fig. 1).

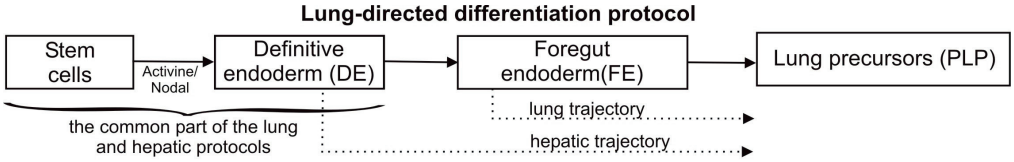


Figure 1. Scheme of the lung-directed differentiation protocol.

Conclusion: It was shown that there are two paths of differentiation from stem cells to lung progenitors (bifurcation). Clusters corresponding to various cell types and cell maturity were annotated. Driver genes were found for terminal macrostates. Significant population of liver cells was found during lung-directed differentiation protocol. We presume that early signals of differentiation into liver cells emerge in the definitive endoderm phase.
Funding: The study is supported by a grant from RSF No. 24-44-00099.

Список литературы/References

1. Zhang X. et al. Stereotyped Subclones Revealed by High-Density Single-Cell Lineage Tracing Support Robust Development. *Adv Sci (Weinh)* 2025;12(30):e2406208. doi 10.1002/advs.202406208.
2. Hawkins F. et al. Prospective isolation of NKX2-1-expressing human lung progenitors derived from pluripotent stem cells. *J Clin Invest* 2017;127(6):2277-2294. doi 10.1172/JCI89950