

Интеграция филостратиграфического и филотранскриптомного анализа в исследовании дифференциально экспрессирующихся генов в раковых тканях

Иванов Р.А.^{1*}, Лашин С.А.^{1, 2}, Афонников Д.А.^{1, 2}, Матушкин Ю.Г.^{1, 2}

¹ Институт цитологии и генетики СО РАН, Новосибирск, Россия

² Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

* ivanovromanart@bionet.nsc.ru

Ключевые слова: филостратиграфия; филотранскриптомика; дифференциальная экспрессия генов; карциномы

Мотивация и цель: Карциномы представляют собой один из наиболее распространенных видов рака, поражающих различные органы человеческого тела. Понимание эволюционного происхождения и его связи с функциональной ролью генов, связанных с этими заболеваниями, важно не только для осмысления природы рака, но и для разработки новых подходов к его диагностике и лечению. В нашем исследовании мы использовали филостратиграфический и филотранскриптомный анализ для изучения и сравнения эволюционных возрастов генов на различных стадиях развития карцином у человека. Филостратиграфический анализ – метод, направленный на определение эволюционного происхождения генов путем анализа присутствия их ортологов – генов, кодирующих гомологичные белки, разошедшиеся в результате видообразования в геномах разных видов. Такой подход не только выявляет ключевые моменты в эволюции генома, когда наблюдался резкий рост числа новых генов, но и помогает идентифицировать гены, уникальные для специфических организмов. Филотранскриптомика же является методом, объединяющим информацию об эволюционном возрасте генов, полученной с помощью филостратиграфии, и данные об уровне экспрессии генов. Этот подход позволяет исследовать взаимосвязь между давностью возникновения генов и изменениями в их активности в контексте различных физиологических состояний, адаптивных ответов или этапов развития организмов или органов. Целью исследования является выявление генов, дифференциально экспрессирующихся в раковых тканях, эволюционный анализ их возраста и изучение возраста с активностью этих генов в развитии опухолей.

Методы и алгоритмы: Для достижения поставленной цели был использован комплексный подход, сочетающий анализ дифференциальной экспрессии генов, филостратиграфический анализ и филотранскриптомный анализ. Для оценки уровня экспрессии генов и выявления дифференциально экспрессирующихся генов использовался алгоритм DESeq2. Филостратиграфический анализ проводился с использованием веб-сервиса OrthoWeb, который позволяет определить эволюционное происхождение генов в единицах индекса филотранскриптомного возраста (Phylostratigraphy Age Index – PAI) на основе анализа присутствия их ортологов в других видах. В качестве порога

идентичности был выбран уровень идентичности белковых последовательностей 0.6, рассчитанный алгоритмом Смита–Ватермана. При помощи Orthoweb также были выполнены оценки индекса изменчивости dN/dS (Divergency Index – DI) при сравнении последовательностей генов с их ортологами в семействе Гоминид. Для оценки уровня экспрессии генов и выявления дифференциально экспрессирующихся генов использовался алгоритм DESeq2. Филотранскриптомный анализ проводился с помощью инструмента myTAI, что позволило оценить динамику экспрессии генов в контексте их эволюционного возраста на различных патологических стадиях карцином в единицах индексов возраста транскриптома (Transcriptome Age Index – TAI) и индекса изменчивости транскриптома (Transcriptome Divergency Index – TDI).

Результаты: В этой работе мы использовали данные по экспрессии генов в тканях печеночной аденокарциномы, светлоклеточной аденокарциномы почек, аденокарциномы кишечника, карциномы молочных желез, аденокарциномы простаты, карциномы щитовидной железы, карциномы уротелия мочевого пузыря, эндометриоидной карциномы тела матки и из соответствующих им образцов здоровых тканей из базы данных TCGA. В каждом из исследуемых типов раковых тканей были выделены опухоли четырех патологических стадий по мере прогрессии опухоли. Филостратиграфический анализ выявил два пика представленности генов: на ранних этапах возникновения многоклеточности и на этапе возникновения позвоночных (PAI = 3 и 6). Филотранскриптомный анализ показал изменения в возрасте транскриптома на разных стадиях развития некоторых опухолей, где в ранних и поздней патологических стадиях развития карцином идет повышение возраста транскриптома в раковых тканях, а в промежуточных – понижение, с особыми отличиями в аденокарциномах кишечника, где наблюдается противоположная картина. Также было обнаружено, что здоровые ткани имеют значимо более высокий возраст транскриптома по сравнению с раковыми, что может свидетельствовать о повышенной активации древних генов при онкологических заболеваниях.

Выводы: Анализ дифференциальной экспрессии генов и их эволюционного возраста в карциномах выявил значительные различия между раковыми и здоровыми тканями, а также специфические изменения в динамике возраста транскриптома на разных стадиях развития опухоли. Эти результаты подчеркивают значимость эволюционного аспекта в изучении рака.

Финансирование: Исследование поддержано бюджетным проектом № FWNR-2022-0006 Министерства науки и высшего образования РФ.

Integration of phylostratigraphic and phylotranscriptomic analysis in the study of differentially expressed genes in cancer tissues

Ivanov R.^{1*}, Lashin S.^{1,2}, Afonnikov D.^{1,2}, Matushkin Yu.^{1,2}

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

* ivanovromanart@bionet.nsc.ru

Key words: phylostratigraphy; phylotranscriptomics; differentially expressed genes; carcinoma

Motivation and Aim: Carcinomas are among the most prevalent cancer types, affecting various human organs. Understanding the evolutionary origin and its correlation with the functional role of genes associated with these diseases is important not only for understanding the nature of cancer, but also for developing new approaches to its diagnosis and treatment. In our study, we employed phylostratigraphic and phylotranscriptomic analyses to investigate the evolutionary ages of genes involved in different stages of human carcinoma development. Phylostratigraphy is a method used to determine the evolutionary origins of genes by analyzing the presence of their orthologs – genes encoding homologous proteins that diverged due to speciation – in the genomes of diverse species. This method reveals key moments in genome evolution characterized by surges in new gene numbers and identifies genes unique to specific organisms. Phylotranscriptomics, on the other hand, merges phylostratigraphic data on gene evolutionary ages with gene expression levels. This integrated approach enables the exploration of how gene age correlates with changes in gene activity across various physiological states, adaptive responses, or developmental stages of organisms or organs. Our study aims to identify genes that are differentially expressed in cancer tissues, analyze their evolutionary ages, and assess the relationship between gene age and activity during tumor development.

Methods and Algorithms: To achieve our goal, we employed a comprehensive approach that integrates differential gene expression analysis, phylostratigraphic analysis, and phylotranscriptomic analysis. For assessing gene expression levels and identifying differentially expressed genes, we utilized the DESeq2 algorithm. To determine the evolutionary origins of genes, we conducted phylostratigraphic analysis using the Orthoweb web service. This service enables the calculation of the Phylostratigraphy Age Index (PAI) by analyzing the presence of gene orthologs across different species, with a protein sequence identity threshold, calculated by Smith-Waterman algorithm, set at 0.6. Additionally, Orthoweb facilitated the estimation of the dN/dS variability index (Divergency Index, DI), comparing divergence of gene sequences with their orthologs within the Hominid family. Furthermore, we performed phylotranscriptome analysis using the myTAI tool. This analysis allowed us to explore the dynamics of gene expression in relation to their evolutionary age across various pathological stages of carcinomas, quantified in terms of the Transcriptome Age Index (TAI) and Transcriptome Divergency Index (TDI).

Results: In this study, we analyzed gene expression data from tissues affected by various carcinomas, including hepatic adenocarcinoma, renal clear cell adenocarcinoma, intestinal adenocarcinoma, breast carcinoma, prostate adenocarcinoma, thyroid carcinoma, urothelial carcinoma of the bladder, and endometrioid carcinoma of the uterine body. We also included samples from corresponding healthy tissues obtained from the TCGA database. For each cancer type, tumors were categorized into four pathological stages to assess progression. Our phylostratigraphic analysis identified two significant peaks of gene representation occurring at the early multicellular and vertebrate evolutionary stages (PAI = 3 and 6). Through phylotranscriptome analysis, we observed variations in the transcriptome age across different tumor development stages. Notably, in both early and late pathological stages of carcinomas, there was an increase in transcriptome age in cancerous tissues, whereas a decrease was noted in intermediate stages. This pattern was particularly distinct in intestinal adenocarcinomas, where the opposite trend was observed. Additionally, our findings suggest that healthy

tissues exhibit a significantly higher transcriptome age compared to their cancerous counterparts, indicating a potential upregulation of ancient genes in cancer.

Conclusion: The analysis of differential gene expression and their evolutionary age in carcinomas revealed significant differences between cancerous and healthy tissues, as well as specific changes in the dynamics of transcriptome age at different stages of tumor development. These results emphasize the importance of the evolutionary aspect in the study of cancer.

Funding: The study was supported by the Budget Project # FWNR-2022-0006 of the Ministry of Science and Higher Education of The Russian Federation.