doi 10.18699/bgrs2026-169

Использование метода анализа активности мотивов для идентификации транскрипционных факторов, управляющих онкологической трансформацией и формированием подтипов раковых опухолей

Кропивницкая Н.[1], Калужский И.[2], Мещеряков Г.[3], Макеев В.[4]*
[1] *Лаборатория системной биологии и вычислительной генетики ИОГен РАН им. Н.И. Вавилова, Москва, Россия*
[2] *Лаборатория мультиомики Центра живых систем МФТИ, Москва, Россия*
[3] *Институт белка РАН, Пущино, Россия*
[4] *Cancer Research UK, National Biomarker Centre, Manchester, UK*
[5] *НМИЦ онкологии им. Н.Н. Блохина Минздрава РФ, Москва, Россия*
* *vsevolod.makeev@gmail.com*

Ключевые слова: активность транскрипционного фактора; MARADONER; загрузка мотивов; гепатоцеллюлярная карцинома; мелкоклеточный рак легкого; генно-регуляторные сети

Мотивация и цель: Одним из основных процессов, определяющих молекулярный фенотип и идентичность клеток, является регуляция транскрипции генов через связывание белков-транскрипционных факторов (ТФ) с регуляторными районами генома, промоторами и энхансерами. Участки ДНК, связывающиеся с ТФ, содержат вхождения «мотивов» – характерных нуклеотидных паттернов, узнаваемых ДНК-связывающим доменом ТФ. При этом, ТФ, относящиеся к общим структурным семействам, обычно распознают достаточно похожие мотивы. Сегодня ДНК-мотивы связывания известны для большинства ТФ человека [1]. Результаты экспериментов по профилированию транскриптома (такие как RNA-Seq) содержат информацию о представленности транскриптов различных генов в изучаемых клетках. Для генома человека расположение промоторов генов картировано достаточно хорошо, и можно поставить задачу анализа активности мотивов (Motif Activity Response Analysis, MARA) [2]: определить, какие ДНК-мотивы (и какие соответствующие им ТФ) вносят вклад в изменение активности транскрипции в конкретных наборах образцов, используя наличие мотивов в промоторах активных и неактивных генов.
Методы и алгоритмы: Для анализа использовался инструмент MARADONER [3], реализующий байесовскую модель множественной линейной регрессии. В качестве входных данных использовались профили экспрессии, прошедшие TMM-нормализацию и батч-коррекцию, а также матрица загрузок мотивов на промоторы, рассчитанная методом sumoccupancy [4] для промоторных областей генов (–500/+2000 п.н.). Такой подход позволяет учитывать вклад кластеров сайтов связывания и повышает устойчивость модели к техническому шуму. На этапе постобработки значимость выявленных активностей оценивалась с помощью многофакторного дисперсионного анализа (ANOVA) с поправкой на множественное тестирование (FDR). Анализ позволил идентифицировать ключевые регуляторы, связанные как со стадиями опухоли, так и с подтипами. Полученные результаты интегрировались с данными экспрессии.
Результаты: В данной работе рассмотрено два примера использования MARADONER: (1) выборка индивидуальных опухолей гепатоцеллюлярной карциномы (ГЦК), соответствующих клеточных линий и здоровых образцов печени и (2) выборка ксенографтов мелкоклеточного рака лёгкого (МКРЛ) [5]. Оба вида рака характеризуются высокой степенью гетерогенности, агрессивным клиническим течением и сложной сетью регуляции транскрипции.
В интактных образцах печени наблюдается высокая активность кластеров мотивов, распознаваемых ключевыми регуляторами дифференцировки печени, такими как HNF1A и HNF4A (рис. 1А). Напротив, в опухолевых образцах и клеточных линиях данные кластеры демонстрируют пониженную активность. Примечательно, что при сопоставимых профилях активности мотивов для факторов HNF1A и HNF4A в образцах группы Solid Normal Tissue, уровни экспрессии соответствующих им генов в этой группе различаются (рис. 1Б), причем экспрессия понижена в образцах с высокой активностью мотивов. Отсутствие прямой корреляции между экспрессией гена ТФ и активностью его мотива указывает на то, что функциональное состояние регулятора определяется не только его концентрацией в клетке, но и эпигенетическим контекстом, посттрансляционными модификациями (например, фосфорилированием) или доступностью специфических кофакторов. Противоположная динамика характерна для кластеров ТФ, ассоциированных с контролем клеточной пролиферации. Повышенная активность мотивов семейств YY1 (инициация транскрипции), NRF1 (митохондриальный биогенез) и NFYC (регуляция клеточного цикла) в опухолях и клеточных линиях (рис. 1А) коррелирует с их высоким пролиферативным статусом и метаболической перестройкой, необходимой для ускоренного деления раковых клеток.

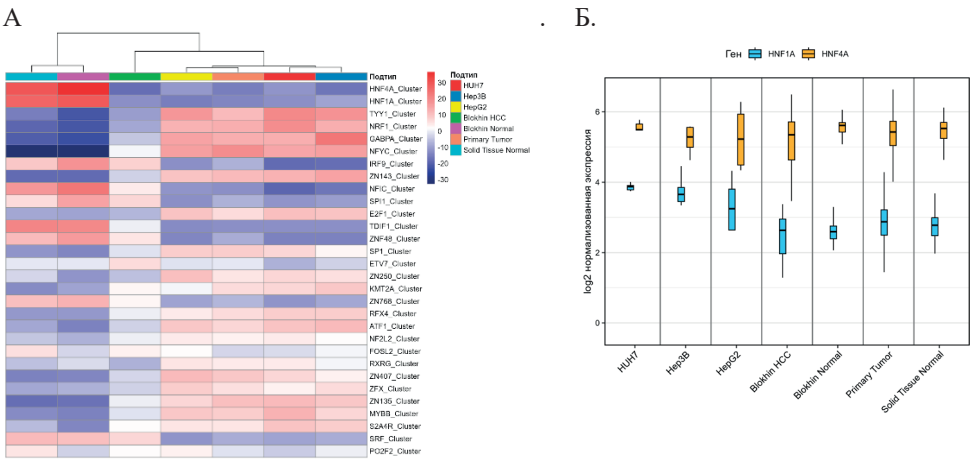


Рис. 1. (А) Активности мотивов разных классов ТФ и (Б) распределение экспрессии генов HNF1A и HNF4A для образцов здоровой печени, клеточных линий печени и ГЦК.

Мотивы связывания ТФ семейства IRF9, ответственных за взаимодействие с иммунной системой, демонстрируют высокую активность в образцах здоровой печени (рис. 1А). В клеточных линиях гены, регулируемые этими белками, не экспрессируются. В опухоли наблюдается промежуточная активность мотивов, пониженная в выборке TCGA и повышенная в выборке образцов, полученных в ФГБУ НМИЦ онкологии им. Н.Н. Блохина [5]. Это может означать отсутствие сигналов интерферонового ответа от клеток опухоли.

Ксенографты, полученные из циркулирующих в кровяном русле опухолевых клеток МКРЛ, поддерживались в пересаженном состоянии в иммунодефицитных мышах, поэтому в отличие от клеточных линий, клетки взаимодействовали между собой и формировали гетерогенные метастазирующие опухоли. Известен ряд подтипов МКРЛ, различающихся экспрессией нейроэндокринных ТФ (ASCL1, NEUROD1 и ATOH1) (рис. 2Б), которые принадлежат одному кластеру мотивов TFE2. Этот кластер находится в топе по активности ТФ, как объясняющий дисперсию транскрипции для всей выборки МКРЛ (см. рис. 2А).

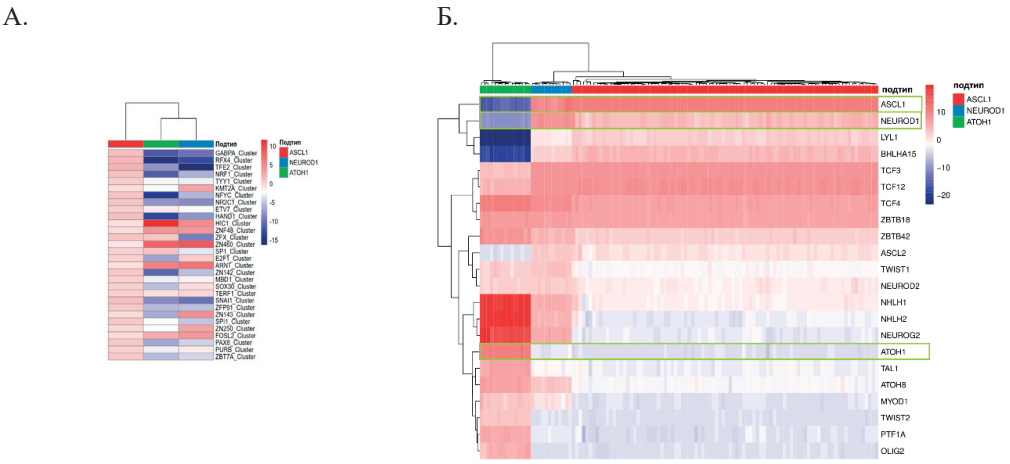


Рис. 2. Активность мотивов и экспрессия разных ТФ в ксенографтах разных подтипов МКРЛ: (А) активность мотивов разных классов ТФ; (Б) экспрессия факторов в кластере TFE2.

Выводы: Проведенный анализ показывает сложные перестройки генных сетей при раковой трансформации клеток. В то время как разница между нормальными и раковыми клетками определяется в первую очередь активностью разных семейств ТФ, распознающих непохожие мотивы, создается впечатление, что тонкие различия между разными подтипами опухолей, или различия, возникающие при развитии раковых клеток в разных условиях, т.е. непосредственно в опухоли, в иммуноизолированном ксенографте или в клеточной культуре, определяется относительной активностью разных ТФ, принадлежащих к одному и тому же семейству, и, по-видимому, конкурирующих за общие мотивы связывания.

Совместный анализ экспрессии индивидуальных ТФ и активности их мотивов открывает широкие перспективы для реконструкции генных регуляторных сетей и поиска ключевых узлов управления клеточным состоянием. В клинической практике понимание специфики взаимодействия ТФ с промоторами генов-мишеней критически важно для идентификации подтипа опухоли и поиска драйверных механизмов опухолевой прогрессии.

Благодарности: Авторы выражают благодарность И.В. Кулаковскому (Институт белка РАН), соавтору инструмента MARADONER, за консультации, продуктивное обсуждение результатов, а также за ценные замечания при подготовке текста работы.

Motif activity response analysis for identification of transcription factors driving oncogenic transformation and tumor subtype variability

Kropivnitskaya N.[1], Kaluzhskiy I.[2], Meshcheryakov G.[3], Makeev V.[4]*

[1] *Laboratory of Systems Biology and Computational Genetics, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia*
[2] *Multiomics laboratory, Life Science Center MIPT, Moscow, Russia*
[3] *Institute of Protein Research, Russian Academy of Sciences, Pushchino, Russia*
[4] *Cancer Research UK, National Biomarker Centre, Manchester, UK*
[5] *N.N. Blokhin National Medical Research Center of Oncology, Moscow, Russia*
* *vsevolod.makeev@gmail.com*

Key words: transcription factor activity; MARADONER; motif occupancy; hepatocellular carcinoma; small cell lung cancer; gene regulatory networks

Motivation and Aim: Transcriptional regulation, mediated by binding of transcription factors (TFs) to regulatory regions such as promoters and enhancers, is a fundamental process determining the cellular phenotype and identity. DNA sites binding TFs contain “motifs”, specific nucleotide patterns recognized by TF’s DNA-binding domains. TFs from the same structural family typically recognize similar motifs, and these binding motifs are now characterized for the majority of human TFs [1]. Transcriptomic profiling (e.g., RNA-Seq) provides data on gene expression levels. Since human promoter locations are well-mapped, Motif Activity Response Analysis (MARA) can be employed to determine which DNA motifs and corresponding TFs drive transcriptional changes across specific sample sets based on motif presence in the promoters of differentially expressed genes [2].
Methods and Algorithms: Analysis was performed using the MARADONER tool [3], which implements a Bayesian multiple linear regression model. Input data consisted of TMM-normalized and batch-corrected expression profiles, alongside a motif-to-promoter occupancy matrix calculated using the sumoccupancy method for promoter regions (–500/+2000 bp). This approach accounts for clusters of binding sites, enhancing model robustness against technical noise. Post-processing involved evaluating the significance of identified activities using ANOVA with False Discovery Rate (FDR) correction. The analysis identified key regulators associated with both tumor stages and subtypes, with results integrated with gene expression data.
Results: This study examines two applications of MARADONER: (1) a cohort of individual hepatocellular carcinoma (HCC) tumors, corresponding cell lines, and healthy liver samples; and (2) small cell lung cancer (SCLC) xenografts [4]. Both cancers exhibit high biological heterogeneity and aggressive clinical progression.
In healthy liver samples, high activity was observed for motif clusters recognized by key hepatic differentiation regulators, such as HNF1A and HNF4A (Fig. 1A). Conversely, these clusters showed decreased activity in tumor samples and cell lines. Notably, while HNF1A and HNF4A motif activities were comparable in “Solid Normal Tissue,” the expression levels of their corresponding genes differed significantly (Fig. 1B). The lack of direct correlation between TF expression and motif activity suggests that the functional state of a regulator is determined not only by concentration but also by epigenetic context, post-translational modifications, or cofactor availability. Opposite dynamics were observed for motifs associated with cell proliferation (TYY1, NRF1, and NFYC), where increased activity in tumors correlates with high proliferative status.
Binding motifs of the IRF9 transcription factor family, which are responsible for interacting with the immune system, demonstrate high activity in healthy liver samples (Fig. 1A). In cell lines, the genes regulated by these proteins are not expressed. In tumors, intermediate motif activity is observed – decreased in the TCGA cohort and increased in the cohort of samples obtained from the N.N. Blokhin National Medical Research Center of Oncology [5]. This may indicate an absence of interferon response signals from the tumor cells.

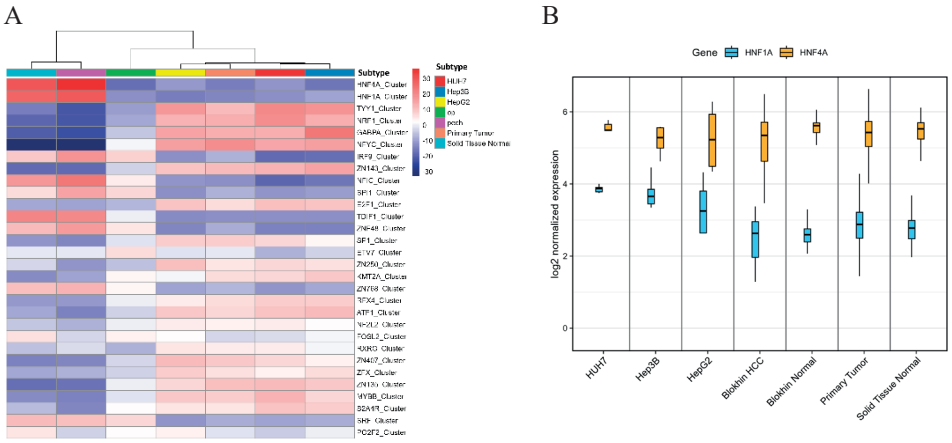


Fig. 1. Comparison of transcriptional profiles across liver samples: (A) Motif activities of various TF classes; (B) Gene expression distribution of HNF1A and HNF4A in healthy liver tissue, liver cell lines, and HCC samples.

Xenografts derived from circulating SCLC tumor cells were maintained in immunodeficient mice. In xenographt cells interact with each other and form heterogeneous metastatic tumors. A few SCLC subtypes are known to differ in the expression of neuroendocrine transcription factors (TFs) such as ASCL1, NEUROD1, and ATOH1 (Fig. 2B). These factors recognise similar motifs and belong to the same TFE2 motif cluster. This cluster is among the top in TF activity, serving as a key factor explaining the variance in transcription across the entire SCLC sample set (Fig. 2A).

Conclusion: The analysis reveals complex rewiring of gene regulatory networks during malignant transformation. While the distinction between normal and cancerous cells is primarily driven by the activity of different transcription factor families recognizing distinct motifs, finer differences appear to be governed by a different mechanism. Specifically, the subtle variations between tumor subtypes – or changes arising from different growth environments (e.g., the primary tumor, immunoisolated xenografts, or cell cultures) – are likely to be determined by the relative expression of different TFs recognising similar motifs. These factors likely compete for the shared binding sites.

The integrated analysis of individual TF expression and their corresponding motif activities offers broad opportunities for reconstructing gene regulatory networks and identifying key nodes controlling cellular states. In clinical practice, understanding the specificity of TF interactions with target gene promoters is critical for identifying tumor subtypes and uncovering the driver mechanisms of tumor progression.

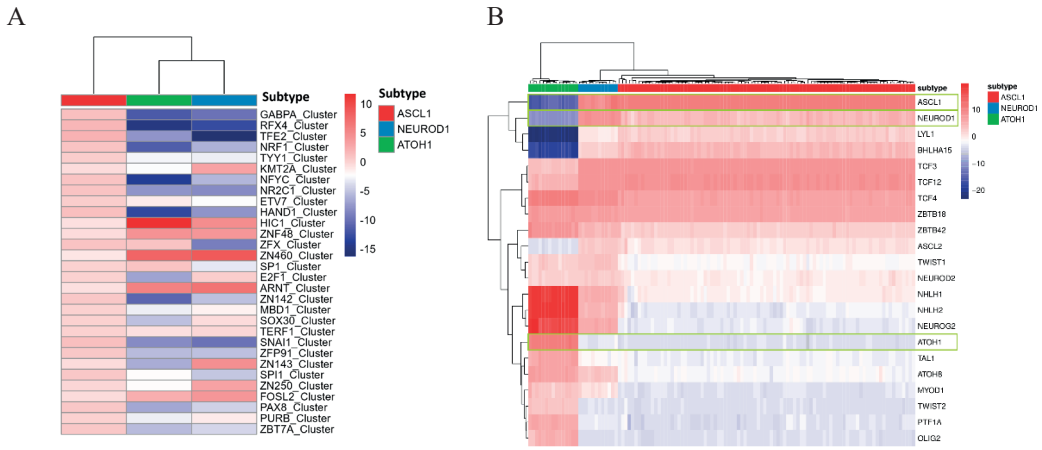


Fig. 2. TF motif activity and gene expression in xenografts of different SCLC subtypes: (A) Motif activities of various TF classes; (B) Expression levels of TFs within the TFE2 cluster.

Acknowledgements: The authors are grateful to Kulakovskiy I. (Institute of Protein Research, RAS), co-developer of the MARADONER tool, for consultations, productive discussion of the results, and for helpful comments on the manuscript.

Список литературы/References

1. Vorontsov IE, Eliseeva IA, Zinkevich A, Nikonov A, et al., HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors, *Nucleic Acids Research*, Volume 52, Issue D1, 5 January 2024, Pages D154–D163
2. FANTOM Consortium; Suzuki H, Forrest AR, van Nimwegen E, et al., The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet*. 2009 May;41(5):553-62. doi: 10.1038/ng.375.
3. Meshcheryakov G., and Buyan, AI., MARADONER: Motif Activity Response Analysis Done Right, 2026, https://arxiv.org/abs/2602.03343.
4. Orenstein Y, Linhart C, Shamir R (2012) Assessment of Algorithms for Inferring Positional Weight Matrix Motifs of Transcription Factor Binding Sites Using Protein Binding Microarray Data. *PLoS ONE* 7(9): e46145. https://doi.org/10.1371/journal.pone.0046145.
5. Simpson, K.L., Stoney, R., Frese, K.K. et al. A biobank of small cell lung cancer CDX models elucidates inter- and intratumoral phenotypic heterogeneity. *Nat Cancer 1*, 437–451 (2020). https://doi.org/10.1038/s43018-020-0046-2