

Конкурирующие эндогенные РНК как маркеры для немелкоклеточного рака легких

Бурденный А.М.^{1,2}, Логинов В.И.^{1,3}, Пронина И.В.¹, Брага Э.А.^{1,3}

¹ Научно-исследовательский институт общей патологии и патофизиологии, Москва, Россия

² Институт биохимической физики им. Н.М. Эмануэля РАН, Москва, Россия

³ Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, Россия

* burdennyu@gmail.com

Ключевые слова: аденокарцинома, плоскоклеточный рак, длинная некодирующая РНК, опухолевые маркеры

Мотивация и цель. Рак легких является наиболее часто диагностируемым злокачественным новообразованием и ведущим источником неблагоприятного исхода от рака [1]. Этот вид злокачественных новообразований, также как и другие, имеет гетерогенную природу и представлен двумя большими классами: мелкоклеточный рак легкого (МРЛ) и немелкоклеточный рак легкого (НМРЛ). На НМРЛ приходится не менее 85 % всех случаев рака легких. Этот вид рака подразделяется на аденокарциному, плоскоклеточный рак и крупноклеточный рак [2]. Отметим, что первые два класса имеют свои существенные особенности в диагностике, патогенезе, локализации, визуализации, метастазировании и лечении [3]. В связи с этим, актуальной задачей является поиск возможных молекулярных маркеров, которые могут помочь однозначно определить, как класс, так и подтип рака лёгких. Таким инструментом могут стать маркеры эпигенетической регуляции, представленные некодирующими РНК. Этот класс молекул характеризуется тем, что вовлечен в управление всех клеточных процессов, и нарушения в работе этих молекул могут стать ключевым фактором развития как рака в целом, так и НМРЛ в частности. Согласно теории конкурирующих эндогенных РНК, длинные некодирующие РНК (днРНК) конкурируют с мРНК белок-кодирующих генов за связывание с микроРНК (миРНК), тем самым влияя на уровень мРНК [4].

Целью настоящего исследования являлся поиск новых маркеров немелкоклеточного рака легкого на основе конкурентных эндогенных РНК.

Методы и алгоритмы. Данные об экспрессии генов у пациентов с НМРЛ получены из Атласа генома рака (TCGA, <https://portal.gdc.cancer.gov/>). Были отобраны дифференциально экспрессируемые (ДЭ) гены информационных (матричных) РНК (мРНК), микроРНК (миРНК) и длинных некодирующих РНК (днРНК), удовлетворяющие правилу $|\text{Log2 FoldChange}| > 3$ и $p < 0.01$. Связь днРНК с раком была изучена по базам LncRNADisease 2.0 (www.cuilab.cn/lncrnadisease) и lncATLAS (<https://lncatlas.crg.eu/>). Транскриптомный ландшафт днРНК был идентифицирован с помощью Lnc2Cancer 2.0 (www.bio-bigdata.net/lnc2cancer).

Для изучения биологических функций ДЭмРНК были использованы системы «Генная онтология» (Gene ontology, <https://geneontology.org/>) и энциклопедия генов и геномов (Киото) (KEGG, <https://www.genome.jp/kegg/>).

В настоящем исследовании была построена сеть (днРНК-миРНК-мРНК) конкурирующих эндогенных РНК, которая основана на гипотезе о том, что днРНК являются «губкой» для миРНК, тем самым осуществляя регуляцию активности

мРНК белок-кодирующих генов. Для получения данных о возможных взаимодействиях миРНК–мРНК и миРНК–днРНК в программу starBase2.0 (<http://starbase.sysu.edu.cn>) были введены данные о днРНК, миРНК, и мРНК с аномальным уровнем экспрессии (кратностью изменений >3.0 , <0.33 и $p < 0.01$). Для прогнозирования взаимодействий днРНК-миРНК использовались инструменты miRanda (<http://www.microna.gr/>) и miRCode (<http://www.mircode.org/>), а миРНК-мРНК – miRWalk 2.0 (<http://mirwalk.umm.uni-heidelberg.de/>), TargetScan (<http://www.targetscan.org/>) и miRTarBase (<https://mirtarbase.cuhk.edu.cn>). Для построения сети конкурирующих эндогенных РНК были выбраны миРНК, которые имеют негативную корреляцию с экспрессией днРНК и мРНК. Для конструирования и визуализации сети днРНК-миРНК-мРНК использовали инструментальную среду Cytoscape v3.0 (<http://www.cytoscape.org/>).

Результаты. При сравнении опухолевых тканей и прилегающих гистологически неизмененных тканей пациентов с раком легких из базы данных TCGA мы обнаружили в общей сложности 735 ДЭмРНК, 12 ДЭмиРНК и 107 ДЭднРНК. Дополнительным этапом отбора было выявление в области не далее 2000 п.н. от 5' конца гена или старта транскрипции обособленных CpG островков с использованием базы Epigenomics <https://www.ncbi.nlm.nih.gov/gene/>. В результате из вышеперечисленного осталось около 60 днРНК, 6 миРНК и около 250 мРНК белок-кодирующих генов.

В результате функционального анализа было показано, что мРНК с повышенным уровнем экспрессии в основном участвуют в митотическом клеточном цикле и делении клеток, а мРНК с пониженной экспрессией в основном участвуют в передаче сигнала в клетке, клеточной адгезии и иммунном ответе. Мы предположили, что гены участвующие в клеточном цикле, сигнальных путях p53, PI3K-Akt, Ca²⁺ могут быть вовлечены в прогрессию НМРЛ.

Всего в сеть конкурентных эндогенных РНК были вовлечены 6 днРНК (ADAMTS9-AS2, HAND2-AS1, HOTAIRM1, MEG3, SNHG12, NCK1-AS1) 4 миРНК (miR-124-3p, miR-203a-3p, miR-125b-5p, miR-129-5p) и 26 мРНК. В опухолевых тканях уровень NCK1-AS1 был повышен, тогда как уровень ADAMTS9-AS2, HAND2-AS1, HOTAIRM1, MEG3, SNHG12 значительно снижался в сравнении с таковым прилежащих гистологически неизмененных тканей. Количество узлов и линий в сети составило 36 и 60 соответственно. Из 6 днРНК количество миРНК, связанных с ADAMTS9-AS2, HAND2-AS1, MEG3 было больше, чем у других днРНК, что указывает на их важную роль прогрессии НМРЛ. Из них 3 днРНК, 2 микроРНК и 15 мРНК были тесно связаны с общей выживаемостью.

Это исследование имеет некоторые ограничения. Во-первых, мы не проверяли эти новые биомаркеры с использованием дополнительных выборок данных. Во-вторых, мы не проводили молекулярно-биологические эксперименты для дальнейшей проверки функций и механизмов регуляции идентифицированных эндогенных РНК у пациентов с НМРЛ. Необходимы дальнейшие исследования с использованием клинических образцов и функциональных экспериментов на клеточных линиях для подтверждения наших выводов.

Выводы. Таким образом, мы построили сеть днРНК-миРНК-мРНК у пациентов с НМРЛ и определили 15 мРНК, 2 микроРНК и 3 днРНК в качестве факторов прогноза общей выживаемости в вышеупомянутой популяции. Настоящее исследование позволяет глубже оценить роли каждого элемента сети эндогенных

РНК, у пациентов с НМРЛ, а некоторые из генов этой сети могут быть использованы в качестве новых потенциальных терапевтических мишеней.

Финансирование: Исследование поддержано государственным заданием FGUF-2023-0001.

Competing endogenous RNAs as non-small cell lung cancer markers

Burdenny A.M.^{1,2*}, Loginov V.I.^{1,3}, Pronina I.V.¹, Braga E.A.^{1,3}

¹ Institute of General Pathology and Pathophysiology, Moscow, Russia

² Emanuel Institute for Biochemical Physics, Moscow, Russia

³ Bochkov Research Centre for Medical Genetics, Moscow, Russia

* burdenny@gmail.com

Key words: lung adenocarcinoma, squamous cell lung carcinoma, long non-coding RNA, oncomarkers

Motivation and Aim. Lung cancer is the most frequently diagnosed malignant neoplasm and the leading source of cancer adverse outcome [1]. This type of malignant neoplasm, as well as others, has a heterogeneous nature and is represented by two large classes: small cell lung cancer (MRL) and non-small cell lung cancer (NSCLC). NSCLC accounts for at least 85% of all lung cancer cases. This type of cancer is divided into adenocarcinoma, squamous cell carcinoma and large cell carcinoma [2]. It should be noted that the first two classes have their own significant features in diagnosis, pathogenesis, localization, visualization, metastasis and treatment [3]. In this regard, an urgent task is to search for possible molecular markers that can help uniquely identify both the class and subtype of lung cancer. Markers of epigenetic regulation, represented by non-coding RNAs, can become such a tool. This class of molecules is characterized by the fact that it is involved in the regulation of all cellular processes, and dysregulation in the work of these molecules can become a key factor in the development of both cancer in general and NSCLC in particular. According to the theory of competing endogenous RNAs, long non-coding RNAs (lncRNAs) compete with the mRNAs of protein-coding genes for binding to microRNAs (miRNAs), thereby affecting the mRNA level [4].

The aim of this study was to search for new markers of non-small cell lung cancer based on competitive endogenous RNAs.

Methods and Algorithms. Gene expression dataset of patients with NSCLC were obtained from the Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>). Differentially expressed (DE) genes of informational (matrix) RNAs (mRNAs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) that satisfying the rule $|\text{Log2FoldChange}| > 3$ and $p < 0.01$ were selected. The association of lncRNAs with cancer was studied using LncRNADisease 2.0 (www.cuilab.cn/lncrnadisease) and lncATLAS (<https://lncatlas.crg.eu/>) databases. The transcriptomic landscape of lncRNAs was identified using Lnc2Cancer 2.0 (www.bio-bigdata.net/lnc2cancer).

Gene ontology (<https://geneontology.org/>) and the Kyoto Encyclopedia of Genes and Genomes (KEGG, <https://www.genome.jp/kegg/>) systems were used to study the biological functions of lncRNAs.

In present study, a network (lncRNA-miRNA-mRNA) of competing endogenous RNAs was constructed, which is based on the hypothesis of lncRNAs were to be the "sponge" for miRNAs, thereby regulating the activity of mRNA protein-coding genes. To obtain data of possible miRNA-mRNA and miRNA-lncRNA interactions, the data of abnormal lncRNAs, miRNAs, and mRNAs expression levels (multiplicity of changes >3.0 , <0.33 , and $p<0.01$) were entered in the starBase2.0 program (<http://starbase.sysu.edu.cn>). Database miRanda (<http://www.microrna.gr/>) and miRCode (<http://www.mircode.org/>) tools were used to predict lncRNA-miRNA interactions, and miRWalk 2.0 (<http://mirwalk.umm.uni-heidelberg.de/>), TargetScan (<http://www.targetscan.org/>) and miRTarBase (<https://mirtarbase.cuhk.edu.cn>) databases were used for miRNA-mRNA interactions. To build a network of competing endogenous RNAs, miRNAs that have a negative correlation with the expression of lncRNA and mRNA were selected. Cytoscape v3.0 instrumental environment was used to construct and visualize the lncRNA-miRNA-mRNA network (<http://www.cytoscape.org/>).

Results. When comparing tumor tissues and adjacent histologically unchanged tissues of lung cancer patients from the TCGA database, we found a total of 735 DE mRNAs, 12 DE miRNAs and 107 DE lncRNAs. An additional stage of selection was the identification of isolated CpG islands in the region no farther than 2000 bp from the 5' end or the start of transcription of the gene using the Epigenomics database <https://www.ncbi.nlm.nih.gov/gene/>. As a result, about 60 lncRNAs, 6 miRNAs and about 250 mRNAs of protein-coding genes remained from the above mentioned.

As a result of functional analysis, it was shown that mRNAs with an increased level of expression are mainly involved in the mitotic cell cycle and cell division, and mRNAs with reduced expression are mainly involved in signal transmission in the cell, cell adhesion and immune response. We hypothesized that genes involved in the cell cycle, signaling pathways p53, PI3K-Akt, Ca²⁺ may be involved in the progression of NSCLC. A total of 6 lncRNAs (*ADAMTS9-AS2*, *HAND2-AS1*, *HOTAIRM1*, *MEG3*, *SNHG12*, *NCK1-AS1*), 4 miRNAs (miR-124-3p, miR-203a-3p, miR-125b-5p, miR-129-5p) and 26 mRNAs were involved in the network of competitive endogenous RNAs. In tumor tissues, the level of *NCK1-AS1* was increased, whereas the level of *ADAMTS9-AS2*, *HAND2-AS1*, *HOTAIRM1*, *MEG3*, *SNHG12* significantly decreased in comparison with that of adjacent histologically unchanged tissues. The number of nodes and lines in the network was 36 and 60, respectively. From the 6 lncRNAs, the number of miRNAs associated with *ADAMTS9-AS2*, *HAND2-AS1*, and *MEG3* was higher than that for other lncRNAs, indicating their important role in NSCLC progression. Of these set, 3 lncRNAs, 2 microRNAs, and 15 mRNAs were closely associated with overall survival. This study has some limitations. Firstly, we did not test these new biomarkers using additional data samples. Secondly, we did not conduct molecular biological experiments to further verify the functions and mechanisms of regulation of identified endogenous RNAs in patients with NSCLC. Further studies using clinical samples and functional experiments on cell lines are needed to confirm our findings.

Conclusion. Thus, we constructed a lncRNA-miRNA-mRNA network in patients with NSCLC and identified 15 mRNAs, 2 microRNAs and 3 lncRNAs as predictors of overall survival in the above-mentioned population. This study allows us to more deeply evaluate the roles of each element of the endogenous RNA network in patients with NSCLC, and some of the genes in this network may be used as new potential therapeutic targets.

Funding: The study is supported by the FGFU-2023-0001 state assignment.

Список литературы/References

- 1 Bray F., Ferlay J., Soerjomataram I., Siegel R.L., Torre L.A., Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424
- 2 Inamura K. Lung cancer: understanding its molecular pathology and the 2015 WHO classification. *Front Oncol.* 2017;7:193
- 3 Chen J.W., Dhahbi J. Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. *Sci Rep.* 2021;11(1):13323. doi 10.1038/s41598-021-92725-8
- 4 Yu Y., Ren K. Five long non-coding RNAs establish a prognostic nomogram and construct a competing endogenous RNA network in the progression of non-small cell lung cancer. *BMC Cancer.* 2021;21(1):457. doi 10.1186/s12885-021-08207-7