

АНАЛИЗ РАКОВЫХ КЛЕТОК АДЕНОКАРЦИНОМЫ ЛЕГКИХ КУРИЛЬЩИКОВ С ИСПОЛЬЗОВАНИЕМ БИОЧИПОВ ДНК

А.В. Саечников, Н.Н. Яцков, В.В. Апанасович, В.В. Скакун

Кафедра системного анализа и компьютерного моделирования, Белорусский государственный университет

Минск, Республика Беларусь

E-mail: saetchnikov.anton@tut.by

В работе приводятся результаты исследования биофункций раковых клеток аденокарциномы легких курильщиков на основе экспериментальных данных, опубликованных в работе [3]. Представлены результаты анализа выразенных и подавленных генов. Анализ данных проводился при помощи разработанного программного обеспечения GeneExpressionAnalyser. Полученные результаты дополняют выводы, полученные в работе [3].

ВВЕДЕНИЕ

С помощью биочипов ДНК можно изучать транскрипционную активность в биологическом объекте [1]. В отличие от традиционных молекулярных биологических решений, которые предоставляют возможность исследовать единичный ген или малый набор генов, биочипы предоставляют возможность исследовать абсолютно новые и неизвестные функциональные роли генов. Мощности этих инструментов может быть использована в широком диапазоне областей применения, в том числе для обнаружения новых подтипов заболевания, создания новых диагностических инструментов и определения основных механизмов болезни или воздействия лекарства [2]. Используя микрочипы ДНК можно сравнивать экспрессию генов в больных и здоровых или обработанных лекарством клетках и определить, какие гены при каких экспериментальных условиях или в какой момент времени активируются и могут влиять на процессы, происходящие с клетке. Вышеперечисленные преимущества позволяют производить качественный анализ генов раковых клеток. Целью работы является качественный анализ функциональных изменений в раковых клетках аденокарциномы легких курильщиков на основе исследования экспериментальных данных, опубликованных в [3].

1. ЭКСПЕРИМЕНТАЛЬНЫЕ ДАННЫЕ

В работе выполнен анализ данных, полученных в работе [3]. Данные от биочипов ДНК размещены в хранилище ArrayExpress под индексом E-GEOD-43458 и представляют собой набор генов аденокарциномы легких (80 пациентов) и набор генов здоровой клетки легких (30 пациентов). Живые ткани получены из банка клеток (Хьюстон, Техас, США) онкологического центра рака легких специализированной программы исследований мастерства MD Anderson (SPORE) и классифицированы с помощью системы классификации Всемирной организации здравоохранения в 2004, как описано в [4]. Об-

разцы тканей были получены от пациентов, перенесших операцию в том же учреждении с 2003 по 2005. Подробная клинично-патологическая информация доступна для большинства пациентов, а также учтены демографические данные пациентов, история употребления сигарет (никогда не курили или курильщики, пациенты, которые выкурили, по крайней мере 100 сигарет в своей жизни). Коллекция из 80 тканей аденокарциномы легких и 30 здоровых парных тканей были быстро заморожены и сохранялись в жидком азоте для полного выделения РНК. Все злокачественные образцы содержали более 40% опухолевых клеток. Эксперимент производился на микрочипах AffymetrixGeneChipHumanGene 1.0 STArray.

II. РЕЗУЛЬТАТЫ

Экспериментальные данные обработаны с использованием программы GeneExpressionAnalyser, разработанной на кафедре системного анализа и компьютерного моделирования БГУ [5]. Проанализировано 110 наборов экспериментальных данных, в каждом наборе по 33252 гена. В наборах экспериментальных данных отсутствовали значения интенсивности ММ-проб, что обусловило выполнение RMA-нормировки. На первом этапе нормировки определены плотности интенсивности фоновой компоненты для каждого биочипа, относительно которых в дальнейшем отнормированы значения экспрессии генов. В результате нормировки преобразованы матрицы уровней экспрессии для каждого технического репликанта и сформированы матрицы значений экспрессии генов. Для данных по изучению раковых клеток выполнены следующие варианты метода анализа значимости биочипов SAM (от Significance Analysis of Microarrays): однофакторный, двухфакторный парный. Количество значимых генов в раковой клетке курящего человека, в среднем, в 13 раз больше чем у некурящего человека, что оставляет в среднем 40% от всего набора генов (таб. 1).

Таблица 1 – Количество значимых генов (Pos – выжатые гены, Neg – подавленные гены) выделенных однофакторным и двухфакторным методами SAM.

Данные	Некурящие		Курящие	
	Pos	Neg	Pos	Neg
Однофакторный	281	754	12288	127
Двухфакторный парный	388	745	13566	363

Значения экспрессии выраженности генов на контрольных микрочипах (30 штук) усреднены после выполнения предварительной обработки. Затем данные от курящих и некурящих людей отнормированы относительно этих значений. Для кластеризации задействованы данные, полученные двухфакторным парным SAM-анализом. Выполнена иерархическая кластеризация данных с использованием среднего метода связывания и расстояния Махаланобиса. У данного типа кластеризации кофенетический корреляционный коэффициент численно равен 0.94. Анализ полученных результатов для коэффициента несоответствия для первых 20 уровней дендрограммы показал, что наилучший результат соответствует случаю разбиения всей совокупности значимых генов на два кластера. Исходя из методологии SAM, следует, что эти два кластера будут представлять выжатые и подавленные гены. Выделены активные функции клетки, контролируемые значимыми генами как для раковой клетки некурящего, так и раковой клетки курящего человека. Статистическая значимость функций определена с помощью точного критерия Фишера ($p < 0,01$). Выявлено, что 848 (753 биологических процессов и 95 молекулярных функций) значимых функций контролируются подавленными генами и 149 (81 биологических процессов и 68 молекулярных функций) – выжатными генами в раковой клетки некурящего человека. В раковой клетке курящего человека 229 (154 биологических процессов и 75 молекулярных функций) значимых функций контролируются подавленными генами и 425 (267 биологических процессов и 158 молекулярных функций) – выжатными генами. Определяющие функции, контролируемые выжатными генами раковой клетки курящего человека, связаны с активностью G-белка (G-protein activity) а также с деятельностью различных типов рецепторов (receptor activity). Выделить группу наиболее значимых функций, контролируемых подав-

ленными генами раковой клетки курящего человека, достаточно сложно, однако следует отметить функции, связанные с положительным хемотаксисом (positive chemotaxis). Наиболее активные функции, контролируемые выжатными генами раковой клетки некурящего человека, связаны с альфа-фукозилтрансферазной активностью (alpha fucosyl transferase activity), а контролируемые подавленными генами раковой клетки некурящего человека – с развитием кровеносных сосудов (angiogenesis). Наиболее значимы функции, контролируемые выжатными генами у раковой клетки курящего человека.

III. СРАВНЕНИЕ РЕЗУЛЬТАТОВ

В данном разделе приводится сравнение результатов данной работы и ранее опубликованной статьи [3]. Следует отметить, что авторами статьи [3] не были представлены результаты функционального анализа, поэтому рассмотрим только сравнение значимых генов. Выделены все значимые гены, опубликованные в [3], причем найдены 192 дополнительных значимых гена из общего числа исследуемых генов.

ЗАКЛЮЧЕНИЕ

В результате анализа определены группы генов и функций, имеющие положительную и отрицательные выраженности. Полученные результаты подтверждают опубликованные ранее [3]. В ходе анализа определены новые значимые гены и функции, определяющие развитие аденокарциномы легких, например, такие как ABCG2, OR5H6, TGM1 (отрицательная регуляция); CLDN14, SAA3P, VPREB3 (положительная регуляция).

1. Akondi , K. B. Emerging Trends in Genomic Approaches for Microbial bioprospecting / K. B. Akondi, V. V. Lakshmi // *Proc. OMICS* . – 2013. –Vol. 17(2). – P. 61–70.
2. Slonim, D. K. Getting Started in Gene Expression Microarray Analysis / D. K. Slonim, I. Yanai // *PLoSComput Biol.* – October 2009. – Vol. 5 (10). – P. 1–4.
3. ETS2 Mediated Tumor Suppressive Function and MET Oncogene Inhibition in Human Non-Small Cell Lung Cancer / M. Kabbout [et. al.] // *Clin Cancer Res.* – July 1, 2013. – Vol. 19. – P. 3383–3395.
4. Abnormalities of the TITF-1 lineage-specific oncogene in NSCLC: implications in lung cancer pathogenesis and prognosis / X. Tang [et al.] // *Clin Cancer Res.* – 2011. – Vol. 17. – P. 2434–2443.
5. Саечников , А. В. Анализ экспрессии генов в результате воздействия интерферона IFN-гамма на клетку с использованием программного пакета GeneExpressionAnalyser / А. В. Саечников [и др.] // *Информатика.* – 2014. – № 42 – С. 5–18.