

0-787907

На правах рукописи

КРАСНОВ Георгий Сергеевич

**ПОИСК И ИДЕНТИФИКАЦИЯ
НАРУШЕНИЙ ЭКСПРЕССИИ ГЕНОВ
В РАЗЛИЧНЫХ ОПУХОЛЯХ ЧЕЛОВЕКА**

Специальность 03.01.03 – молекулярная биология

**Автореферат
диссертации на соискание учёной степени
кандидата биологических наук**

Москва 2011

Работа выполнена в лаборатории структурно-функциональной геномики (зав. лаб. доктор биологических наук Л.Ю. Фролова) Учреждения Российской академии наук Института молекулярной биологии им. В.А. Энгельгардта РАН.

Научный руководитель: кандидат химических наук
Вера Николаевна Сенченко

Официальные оппоненты: доктор биологических наук
Евгений Александрович Климов
доктор биологических наук, профессор
Александр Васильевич Карпухин

Ведущая организация Учреждение Российской академии наук
Институт биохимии им. А.Н. Баха РАН

Защита диссертации состоится «24» мая 2011 г. в 11 часов на заседании Диссертационного совета Д 002.235.01 при Учреждении Российской академии наук Институте молекулярной биологии им. В.А. Энгельгардта РАН по адресу: 119991, Москва, ул. Вавилова, 32.

С диссертацией можно ознакомиться в библиотеке Института молекулярной биологии им. В.А. Энгельгардта РАН

Автореферат разослан «21» августа 2011 г.

НАУЧНАЯ БИБЛИОТЕКА КГУ



0000677018

Учёный секретарь
Диссертационного совета
кандидат химических наук

А.М. Крицын

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. В настоящее время в развитых странах мира онкологические заболевания являются второй по распространённости причиной смертности после сердечнососудистых патологий (Jemal et al., 2010), что связано с отсутствием доступных методов диагностики, эффективных для выявления заболевания на ранних стадиях, когда хирургия и другие методы лечения позволяют добиться наилучших результатов. Это делает актуальным поиск потенциальных генов-онкомаркеров, уровень экспрессии которых значительно изменяется в опухоли по сравнению с нормальной тканью. Также в качестве онкомаркеров могут выступать альтернативные транскрипты мРНК (Cheng et al., 2005; Liu et al., 2010) или их белковые продукты, практически отсутствующие в различных нормальных тканях человека (Morre et al., 2009).

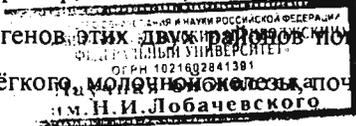
Широкая доступность большого объёма геномных, транскриптомных и эпигеномных данных делает востребованной разработку и дальнейшее совершенствование новых методов и программных продуктов, позволяющих получать сведения об онкоассоциированных изменениях уровня экспрессии и других свойствах генов на основе анализа имеющихся баз данных. В настоящее время информацию об этих изменениях наиболее полно представляют данные гибридизации на микрочипах и таговые данные – EST (экспрессируемый фрагмент последовательности, expressed sequence tag), SAGE (серийный анализ экспрессии генов, serial analysis of gene expression), RnaSeq (высокопроизводительное секвенирование).

В настоящий момент существует ряд биоинформатических подходов для выявления потенциальных генов-онкомаркеров, основанных на анализе dbEST – например, GeneHub GEPIS (Zhang et al., 2007) и EDGES (Lu et al., 2009). Однако при оценке возможных изменений экспрессии генов не учитывается распределение EST по различным клонотекам нормальной и опу-

холевой ткани, частота мутаций и наличие альтернативных или аберрантных (не встречающихся в норме) изоформ.

Для экспериментальной проверки предсказанных по данным dbEST изменений используют метод полимеразной цепной реакции в реальном времени (ПЦР-РВ), позволяющий получить точные количественные данные. Однако есть ряд проблем, связанных с этим методом, одна из которых заключается в необходимости определения эффективности реакции (Gevertz et al., 2005; Kubista et al., 2006) и более точного описания кинетики (Lalam, 2006; Rutledge and Stewart, 2008). С другой стороны, на точность результатов ПЦР-РВ влияет выбор контрольных генов, применяемых для нормирования данных. Уровень экспрессии этих генов незначительно отличается в исследуемом типе опухоли и соответствующей нормальной ткани. По данным литературы до сих пор нет единого мнения по поводу выбора приемлемых контрольных генов для рака лёгкого, почки и других видов опухолей.

Известно, что разного рода геномные нарушения – мутации, транслокации, делеции и др. приводят к развитию различных опухолей. Одним из районов, наиболее часто подверженных таким нарушениям, является короткое плечо третьей хромосомы (3p). Для генов этого района также характерно частое метилирование промоторного участка, приводящее к снижению уровня экспрессии. Известно около десяти областей (т.н. «горячих точек») короткого плеча хромосомы 3 с наиболее частыми нарушениями. Среди них можно выделить два основных района – LUCA и AP20, находящихся на участке 3p21 и подверженных нарушениям во многих видах рака (Wistuba et al., 2000; Zabarovsky et al., 2002; Ji et al., 2005; Angeloni et al., 2007; Hesson et al., 2007; Kashuba et al., 2009). Эти области содержат большое число генов-супрессоров опухолевого роста, как потенциальных (например, *ITGA9*, *APRG2*, *VILL*), так и тех, для которых экспериментально показана опухолеподавляющая активность, например, *RASSF1*, *SEMA3B* (Neufeld et al., 2007; Wang et al., 2008). Для ряда генов этих двух районов показаны снижения уровня экспрессии при раке лёгкого, молочной железы, почки и других опу-



холях (Kashuba et al., 2004; Anedchenko et al., 2008, Pronina et al., 2009). Поиск новых генов-супрессоров опухолевого роста является на сегодняшний день важной задачей не только с точки зрения практического применения – поиска мишеней для последующей разработки методов лечения с использованием таргетной и генной терапии, – но и с точки зрения фундаментальной науки – изучения механизмов онкогенеза.

Цель и задачи исследования. Цель работы – разработка и применение биоинформатического метода для поиска и идентификации нарушений экспрессии генов в девяти распространённых видах рака. Сформулированы следующие конкретные задачи исследования:

- 1) на основе анализа существующих биоинформатических подходов разработать метод поиска генов-онкомаркеров и контрольных генов; реализовать метод в виде программного продукта;
- 2) апробировать метод и применить его для поиска генов-маркеров, включая гены короткого плеча хромосомы 3, и контрольных генов;
- 3) выполнить анализ полученных результатов и сравнить с данными литературы;
- 4) создать модель для расчёта эффективности ПЦР-РВ и реализовать этот метод в виде программного продукта, позволяющего проводить математическую обработку количественных данных;
- 5) подтвердить экспериментально изменения уровня мРНК для выбранных генов-онкомаркеров и контрольных генов.

Научная новизна и практическая ценность работы. Разработанный биоинформатический метод поиска онкоассоциированных генов и реализующее его приложение CompEST включает преимущества имеющихся на настоящий момент других методов и лишён их недостатков; в то же время этот подход дополнен рядом других возможностей – учётом мутаций, нарушений сплайсинга, степени равномерности распределения EST по клонотекам. Приложение CompEST, реализующее данный метод, имеет простой пользовательский интерфейс, что делает программу доступной и удобной для лю-

бого, даже неподготовленного пользователя. Дальнейшее применение этот подход может найти при анализе другого типа транскриптомных таговых данных, полученных методом высокопроизводительного секвенирования (RnaSeq).

Отобранные при помощи разработанного метода гены с потенциально изменённым в опухоли уровнем экспрессии или часто встречающимися нарушениями сплайсинга в дальнейшем могут быть использованы в составе панелей генов при разработке диагностических и, возможно, прогностических наборов.

Впервые предложен и успешно применён биоинформатический метод поиска новых контрольных генов. Контрольный ген *RPNI*, отобранный в ходе исследования, может быть использован для нормирования данных по оценке изменений экспрессии генов для рака лёгкого и почки. Успешно показана применимость предложенной в диссертационной работе и реализованной в программе «АТГ» («Анализ транскрипции генов») кинетической модели ПЦР-РВ, а также преимущество этой модели по сравнению с другими распространёнными методами расчёта эффективности реакций.

Апробация работы. Диссертационная работа представлена и обсуждена на семинаре Лаборатории структурной и функциональной геномики ИМБ РАН им. В.А. Энгельгардта. Результаты работы представлены: на 4-м съезде Российского общества медицинских генетиков, Ростов-на-Дону, 14—20 мая 2010 года; на 6-й Российской конференции по фундаментальной онкологии, Санкт-Петербург, 16 апреля 2010 года; на 5-м Московском международном конгрессе «Биотехнология: состояние и перспективы развития», Москва, 16—20 марта 2009 года; на 4-й международной конференции по проблемам вычислительной биологии, Москва, 20—23 июля 2009 года; на 9-й международной конференции молодых онкологов «Проблемы экспериментальной и клинической онкологии», Киев, 23—24 апреля 2008 года; на Итоговых конференциях по результатам выполнения мероприятий за 2007—2009 гг. в рамках приоритетного направления «Живые системы» ФЦП «Исследования

и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007—2012 годы», Москва; на 49-й научной конференции МФТИ, 21 ноября 2006 г., Москва; на 4-м съезде Российского общества биохимиков и молекулярных биологов, Новосибирск, 11—15 мая 2008 года.

Публикации. По материалам диссертационной работы опубликовано 5 статей в научных журналах и зарегистрирована программа для ЭВМ.

Структура и объём диссертации. Диссертация изложена на 113 страницах, содержит 19 рисунков и 7 таблиц и состоит из введения, обзора литературы, результатов и обсуждения, выводов и списка литературы (283 наименования).

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Метод анализа базы данных dbEST и программа CompEST

Предложен биоинформатический метод поиска генов, характеризующихся изменениями уровня мРНК или ее нуклеотидной последовательности в опухоли по сравнению с нормой. Разработанный подход основан на сравнительном анализе коротких последовательностей кДНК – EST, и включает: нормирование данных EST по контрольным генам; учёт неравномерности распределения EST по клонотекам; учёт частоты нарушений сплайсинга, мутаций, повторяющихся элементов и различий этих частот между нормой и опухолью; возможность отбора потенциально белок-кодирующих aberrантных транскриптов. Метод реализован в программе CompEST, которая имеет простой пользовательский интерфейс, что делает ее удобной в работе и доступной для любого, даже неподготовленного пользователя. В программе предусмотрена возможность автоматического обновления файлов данных – нуклеотидных последовательностей геномной ДНК, расположения генов, выравнивания EST и др. с серверов UC Santa Cruz и NCBI.

Нормирование данных с использованием контрольных генов

Одна из основных предпосылок, заложенных в основу всех таговых методов, заключается в том, что число EST для определённого гена коррелирует с уровнем его мРНК. Однако для корректного сравнения числа клонов различных генов в клонотеках нормальной ткани и опухоли и проведения оценки возможных изменений эти значения должны быть нормированы. Одна из особенностей разработанного метода, отсутствующая в ряде других аналогов (Baranova et al., 2001; Zhang et al., 2007; Lu et al., 2009), заключается в применении сразу нескольких контрольных генов для нормирования данных EST.

Предложенный в работе метод предполагает проведение нормирования с учётом того, что чем большее число клонов EST соответствует данному гену, тем сильнее выражена корреляция с действительным уровнем мРНК. Введём обозначения:

$N_{i,j}$ – нормированное число EST для целевого гена i в клонотеке j ;

$E_{i,j}$ – исходное число EST для целевого (не контрольного) гена i в клонотеке j ;

$N_{i,j}$ – нормированное число EST для целевого гена i в клонотеке j ;

$C_{k,j}$ – число EST для контрольного гена k в клонотеке j ;

K – общее число контрольных генов;

P_j – общее число EST (представительность) в клонотеке j ;

J – общее число клонотек.

Метод предполагает расчёт нормированных количеств клонов EST $N_{i,j}$ по формуле:

$$N_{i,j} = \frac{E_{i,j}}{\prod_{k=1}^K (C_{k,j} + a\beta_{k,j})^{w_{k,j}}}$$

Здесь: a – постоянный коэффициент, значение которого в CompEST по умолчанию равно 0.5 (оптимальное значение установлено эмпирическим путём);

$W_{k,j}$ – вес контрольного гена k , отражающий степень статистической значимости данных EST;

$\beta_{k,j}$ – дополнительное слагаемое, включённое с целью устранения случаев обращения знаменателя в нуль при отсутствии в клонотеке j клонов EST контрольного гена k ($\beta_{k,j}$ равно ожидаемому числу клонов EST этого гена в соответствии с данными остальных клонотек):

$$\beta_{k,j} = \frac{P_j}{J-1} \sum_{l=1, l \neq j}^J \frac{C_{k,l}}{P_l}$$

Отбор потенциальных контрольных генов

При исследовании видов рака, для которых отсутствуют данные об оптимальных контрольных генах, можно произвести поиск наиболее подходящих кандидатов с помощью CompEST. Для учёта различных характеристик гена, таких как стабильность числа EST в различных нормальных и опухолевых клонотеках, частоты мутаций, повторяющихся элементов и нарушений сплайсинга, предложена оценочная функция S , значение которой отражает потенциальную степень пригодности определённого гена в качестве контрольного. В общем виде значение S равно:

$$S = \frac{F}{Q \cdot \prod D_i}$$

где множитель F , экспрессионный фактор, отвечает за общий уровень и стабильность экспрессии гена во всех клонотеках (нормальной ткани и опухоли), корректирующий множитель Q отвечает за частоту мутаций, нарушений сплайсинга и повторяющихся элементов в клонах EST, соответствующих данному гену, множители D_i – за степень отличия этих характеристик гена между нормой и опухолью.

Гены с максимальными значениями S могут рассматриваться как наиболее перспективные контрольные гены-кандидаты.

Идентификация генов с потенциально изменённым уровнем экспрессии

Для поиска генов с изменённой экспрессией в опухоли по сравнению с нормой, разработанный метод также предполагает использование оценочной функции W , аналогично S , отражающей ряд характеристик гена. Эта функция представляет собой произведение двух величин – W_F , отражающей оценку изменения экспрессии гена, и W_Q , отражающей оценку различий частот нарушений сплайсинга, мутаций и повторяющихся элементов в клонотеках опухоли по сравнению с нормальной тканью. В свою очередь, W_F – отношение двух экспрессионных факторов: F_N для нормальной ткани и F_T для опухоли, а множитель W_Q – отношение корректирующих множителей Q_N и Q_T :

$$W = W_F W_Q = F_T / F_N \cdot Q_N / Q_T$$

Значения F_N , F_T , Q_N , Q_T вычисляются аналогично значениям компонентов F и Q оценочной функции S , однако отдельно для клонотек нормы и опухоли.

В качестве потенциально инактивированных отбирали гены с минимальными значениями оценочной функции W – те гены, для которых присутствуют клоны EST, равномерно распределённые по клонотеками нормальных тканей, и практически отсутствуют EST (или характеризуются высокой частотой нарушений сплайсинга) в клонотеках опухоли. Аналогично, в качестве генов с повышенным уровнем экспрессии в опухоли отбирали гены с максимальными значениями W .

Идентификация часто встречающихся нарушений сплайсинга

Третья из основных возможностей, предоставляемых программным продуктом CompEST, – поиск встречающихся нарушений сплайсинга. При различных заболеваниях человека, в том числе онкологических, достаточно часто помимо количественных изменений содержания мРНК в клетке, происходят изменения непосредственно и в самом транскрипте (Cheng et al., 2005; Liu et al., 2010). Это могут быть как мутации единичных нуклеотидов,

так и нарушения сплайсинга (в основном, сохранение интрона и пропуск экзона) (Brinkman 2004; Greenman et al., 2007). Наряду с этими достаточно частыми явлениями обнаружены мРНК, для которых старт транскрипции смещён относительно нормального положения на десятки и даже сотни нуклеотидов (Smith et al., 2009; Takahashi and Nagai, 2009; Liu, 2010). Присутствие или отсутствие таких транскриптов в ткани или кровотоке пациента может служить основой для создания эффективных диагностических наборов и систем.

Аберрантные транскрипты, выявленные только в опухолевых клетках и отсутствующие в нормальных тканях, могут включать: транскрипты, не кодирующие белок; ранее не описанные изоформы (в т.ч. активированные при канцерогенезе эмбриональные изоформы); продукты опухолеассоциированных нарушений сплайсинга, способные кодировать белок (Morre et al., 2009). Для этих транскриптов характерно сохранение рамки считывания в результате изменений сплайсинга, а также отсутствие новых стоп-кодонов (как в добавленных участках, так и на их границах). Такие перестройки могут не влиять на синтез белка по матрице данной мРНК, что и стало критерием отбора возможных аберрантных белковых продуктов. Однако при этом активность белка может быть существенно снижена, например, для изоформы 2 каталитической субъединицы теломеразы hTERT, которая характеризуется делецией в кодирующей области (Mavrogianou et al., 2007; Lincz et al., 2008). Существование таких аберрантных белковых продуктов может быть основой для создания наборов протеомной диагностики.

Таким образом, разработанный биоинформатический метод поиска онкоассоциированных генов, реализованный в приложении CompEST, основан на построении оценочной функции, включающей в себя несколько характеристик каждого гена. Метод позволяет идентифицировать гены с потенциально изменённым уровнем экспрессии в опухоли и часто встречающимися нарушениями сплайсинга, а также контрольные гены, для которых характерны минимальные изменения в опухоли.

Выбор генов с потенциально изменённым уровнем мРНК в различных видах опухолей

С помощью разработанного метода идентифицированы гены с потенциально изменённым уровнем экспрессии при раке толстой кишки, лёгкого, почки, предстательной железы, печени, молочной железы, поджелудочной железы, кожи и головного мозга. Для анализа использованы EST клонотек первичных опухолей и нормальных тканей. При поиске исключены EST, содержащие более 5% несоответствий и 15% повторяющихся элементов.

Идентифицировано 24 гена с пониженным и 12 генов с повышенным уровнем экспрессии в опухолях толстой кишки (Таблица 1). В нашей лаборатории методами ОТ-ПЦР и ПЦР-РВ подтверждены частые (до 80%) и значительные снижения уровня мРНК генов *DES*, *SLC26A3*, *ZG16*. Анализ биомедицинских публикаций и патентный поиск показали, что ген *ZG16* – новый перспективный маркер рака толстой кишки. В ряде работ также обнаружено снижение уровня мРНК генов *MEP1A* (Rosmann et al., 2002) и *GPRC5A* (Acquafreda et al., 2009) при раке толстой кишки, что согласуется с результатами проведённого биоинформатического анализа. Однако для двух генов – *AGR2* и *EIF4A1* экспериментальная проверка, выполненная в нашей лаборатории, не выявила значительных изменений, предсказанных при помощи CompEST. Эти гены в таблице не представлены.

При помощи CompEST также отобрано около 100 перспективных генов с предположительными изменениями уровня экспрессии для 8-ми других видов рака (Таблица 1). Для большинства этих генов показана связь с онкогенезом, а для многих – предсказанные изменения подтверждены экспериментально согласно данным литературы. Так, например, ген *NOLC1* относится к потенциальным онкогенам, поскольку его белковый продукт способен увеличивать скорость прогрессии различных опухолей (Hwang et al., 2009). Повышение уровня экспрессии гена *POSTN* характерно при раке поджелудочной железы; этот ген перспективен для дальнейшего клиничес-

Таблица 1. Потенциальные гены-онкомаркеры для различных видов рака.

Локализация опухоли	Гены с предположительно повышенным уровнем экспрессии	Гены с предположительно сниженным уровнем экспрессии
толстая кишка	<u>CALU</u> , <u>GSTP1</u> , <u>IMPDH2</u> , <u>KARS</u> , <u>MGAT1</u> , <u>MYO1D</u> , <u>PKM2</u> , <u>POLDIP2</u> , <u>POLR3H</u>	<u>ATPIF1</u> , <u>CES2</u> , <u>DES</u> , <u>GPRC5A</u> , <u>FUT6</u> , <u>MEP1A</u> , <u>SLC26A3</u> , <u>PGM1</u> , <u>ZG16</u> и др.
молочная железа	<u>IPO4</u> , <u>GANAB</u> , <u>ZNF207</u> , <u>NONO</u> , <u>NOLC1</u>	<u>USP24</u> , <u>KRT5</u> , <u>GJA4</u> , <u>PPP1R16B</u>
предстательная железа	<u>SLC25A3</u> , <u>NME1-NME2</u> , <u>SMG7</u> , <u>MRFAP1</u> , <u>LOC100288366</u> , <u>ODC1</u>	<u>MSMB</u> , <u>SEMG1</u> , <u>SEMG2</u> , <u>HSPA1A</u> , <u>NDE1</u> , <u>ALDH1A1</u>
почка	<u>LOC100133042</u> , <u>ANXA2</u> , <u>EEF1D</u> , <u>TXNDC5</u>	<u>CPM</u> , <u>UMOD</u> , <u>BHMT</u> , <u>SLC12A1</u> , <u>ACSF2</u> , <u>SLC22A8</u> , <u>CDH16</u> , <u>PRSS8</u>
печень	<u>PKM2</u> , <u>SNRPB</u> , <u>CCT7</u> , <u>CKAP4</u> , <u>CDC20</u>	<u>HPS3</u> , <u>CLU</u> , <u>TM4SF4</u> , <u>SDS</u>
поджелудочная железа	<u>BGN</u> , <u>ANXA2</u> , <u>LUM</u> , <u>POSTN</u> , <u>PGM3</u> , <u>LAMB1</u>	<u>SST</u> , <u>PPY</u> , <u>VWA1</u> , <u>CDKN1C</u> , <u>PRKAG2</u> , <u>BRWD1</u>
головной мозг (глиобластома)	<u>IDI</u> , <u>HAX1</u> , <u>ITGAV</u> , <u>GPR62</u> , <u>SMARCC2</u>	<u>STXBPI</u> , <u>PFKP</u> , <u>NNAT</u> , <u>FAM107A</u> , <u>PLK2</u>
лёгкое	<u>SSU72</u> , <u>KLF2</u> , <u>LPCAT1</u>	<u>GBP3</u> , <u>SEPT7</u> , <u>TM9SF3</u> , <u>COPB1</u> , <u>FTL</u> , <u>FTH</u> , <u>LI-MAL</u> , <u>KTNI</u> , <u>NCOR1</u>
кожа (меланома)	<u>CD63</u> , <u>SILV</u> , <u>SNRPB</u> , <u>PRAME</u> , <u>MAGED1</u> , <u>MAGED2</u> , <u>CCT8</u>	<u>TSPAN5</u> , <u>DMKN</u> , <u>APL2</u> , <u>CAVI</u>

Примечания: 1) Функции генов описаны в диссертации. 2) Жирным выделены гены, для которых ранее показана связь с онкогенезом. 3) Сплошной линией подчеркнуты гены, для которых получены экспериментальные подтверждения, пунктиром – гены, для которых предсказанные изменения наблюдали для других видов рака; при этом не учитывали публикации с данными крупномасштабных методов (микрочипы и др.). 4) В пунктирную рамку обведены гены, для которых предсказанные изменения не соответствуют изменениям, описанным в литературе для других видов рака.

кого применения в диагностике (Tilman et al., 2007). Показано понижение уровня экспрессии гена *CDKN1C* при раке молочной железы, обусловленное метилированием промоторной области (Sato et al., 2005). Особый интерес представляют гены, для которых предсказанные изменения уровня экспрессии обнаружены для других видов рака. Так, например, для *TXNDC5*, отобранного при помощи CompEST в качестве гена с повышенным уровнем экспрессии при раке почки, выявлено повышение уровня экспрессии при раке толстой кишки, шейки матки, желудка и лёгкого (Wang et al., 2007; Zhang et al., 2010). Однако для некоторых генов результаты выполненного в работе анализа не соответствуют экспериментальным данным для других видов рака. Например, для гена *NCOR1* с потенциально сниженным уровнем экспрессии при раке лёгкого ранее показано повышение уровня при раке предстательной железы (Ting et al., 2007).

Исследование функциональных свойств выявленных генов свидетельствует о том, что значительная часть из них вовлечена в важные клеточные процессы, чаще всего связанные с онкогенезом, такие как регуляция клеточной пролиферации (*CAVI*, *CLU*, *CDKN1C*, *LAMB1*, *MAGED1*, *NME1-NME2* и др.), организация внеклеточного матрикса (*APLP2*, *ANXA2*, *LUM*, *POSTN*, *VWAI*), ангиогенез (*ANXA2*, *BGN*, *CAVI*, *GJA4*, *ID1*, *ITGAV*) и ряд других. Применение биоинформатических подходов при поиске опухолевых маркеров позволяет учесть характер экспрессии гена в различных тканях. В частности, из анализа исключены гены, высокий уровень экспрессии которых характерен для крови, сосудистых и мышечных тканей: плазминоген *PLG*, преальбумин *TTR*, сывороточные амилоиды *SAA1* и *SAA2*, миозин гладких мышц *MYH11*. Их выбор в качестве вероятных генов-онкомаркеров мог быть обусловлен высоким содержанием этих тканей в образцах, из которых получены клонотекки. Кроме того, исключение из анализа генов с высоким уровнем экспрессии в клетках крови – неотъемлемый этап отбора сывороточных онкомаркеров, поскольку

высокое содержание белка в сыворотке крови пациента существенно снижает чувствительность анализа и становится препятствием для эффективного применения потенциальных диагностических маркеров.

Поиск генов хромосомы 3 с изменённым уровнем экспрессии при раке лёгкого и почки

Для идентификации онкоассоциированных генов, локализованных на хромосоме 3 человека, использован разработанный метод анализа базы данных dbEST, реализованный в приложении CompEST, дополненный анализом баз данных OncoPrint и SAGE. Для рака почки отобран ряд генов с предположительно изменённым уровнем экспрессии. Среди них:

- 4 гена с предположительно повышенным уровнем экспрессии: *IMPDH2* (опухолеассоциированный антиген рака толстой кишки, 3p21.2), *MCM2* (прогностический маркер рака почки, 3q21), *SOX2* (маркер степени дифференциации клеток рака желудка, 3q26.3-q27), *KALRN* (3q21.2);
- 6 генов с пониженным уровнем экспрессии: *CLDN16* (3q28), *SEMA3G* (обладает противоопухолевой активностью, 3p21.1), *MST1* (известный ген-супрессор, 3p21), *KNG1* (3q27), *ALS2CL* (3p21.31).

Для рака лёгкого идентифицированы:

- 3 гена с предположительно повышенным уровнем экспрессии: *ZBED2* (другие гены этого семейства характеризуются повышенным уровнем экспрессии при различных видах рака, 3q13.2), *AGTR1* (терапевтическая мишень для рака молочной железы, 3q21-q25), *AMT* (3p21.2-p21.1);
- 7 генов с пониженным уровнем экспрессии: *TNFSF10* (фактор некроза опухолей с высокой селективной противоопухолевой активностью, 3q26), *LIMD1* (известный супрессор опухолевого роста, 3p21.3), *FRMD4B* (3p14.1), *LAMP3* (отмечено понижение уровня экспрессии при раке молочной железы, 3q26.3-q27), *ITPR1* (участвует в апоптозе, 3p26-p25), *VGLL3* (потенциальный ген-супрессор опухолевого роста,

3p12.1), *WWTR1* (белок участвует в апоптозе, однако для этого гена отмечено *повышение* уровня экспрессии при раке молочной железы, 3q23-q24).

Таким образом, предложенный метод позволяет провести масштабный поиск генов, вовлечённых в канцерогенез, и выявить новые, не известные ранее опухолевые маркеры.

Новая модель расчёта эффективности ПЦР в реальном времени

Практически все изменения экспрессии генов, предсказанные при помощи методов, основанных на EST, SAGE и микрочипах, требуют экспериментальной проверки точными количественными методами. Для этой цели чаще всего применяют ПЦР в реальном времени. Одна из основных трудностей, связанных с этим методом, заключается в корректном вычислении и учёте эффективности реакции (E). Эффективность ПЦР-РВ представляет собой величину, отражающую возрастание содержания ампликона за один цикл реакции. На практике значение E далеко не всегда близко к 100%, и часто такое допущение, хотя и упрощает расчёты, но приводит к значительному искажению результатов. Методы расчёта эффективности в настоящий момент продолжают совершенствоваться. В работе предложена новая модель, описывающая кинетику ПЦР в реальном времени, применимая для расчёта E путём аппроксимации кинетической кривой уравнением:

$$\frac{dR}{dc} = E \frac{R(F_0 - R)(P_0 - \varepsilon R)}{(F_0 - R)(P_0 - \varepsilon R) + aR(P_0 - \varepsilon R) + bR(F_0 - R)e^{\xi c}}$$

где R – концентрация ампликона на цикле c ; F_0 – исходная концентрация праймеров; P_0 – полимеразы; a, b, ε, ξ – коэффициенты; E – искомая эффективность реакции. Вывод уравнения представлен в диссертационной работе. Аппроксимация кинетических кривых этим уравнением для определения эффективности и дальнейшая обработка данных ПЦР-РВ реализована в программе «АТГ» (Анализ Транскрипции Генов).

***RPNI* – новый контрольный ген, применимый
для нормирования данных при оценке изменений
экспрессии генов в опухолях лёгкого и почки**

ПЦР в реальном времени – основной метод количественной оценки экспрессии генов, который также широко используют для уточнения данных, получаемых при помощи микрочипов. Основные требования, обеспечивающие достоверные и воспроизводимые результаты, полученные этим методом, включают стандартизацию всех этапов исследования, а также нормирование данных, которое позволяет сравнивать уровни экспрессии генов в различных образцах мРНК/кДНК с неизвестной концентрацией. Для нормирования можно использовать количество клеток, массу или объем образца/раствора, концентрацию суммарной или рибосомальной РНК. Однако чаще всего для этого используют содержание мРНК так называемых контрольных генов со стабильным уровнем экспрессии в исследуемых образцах, например, *GAPDH*, *ACTB*, *HPRT1*, *TBP* и др.

Задача данного исследования – идентификация, по крайней мере, одного нового контрольного гена, пригодного для оценки изменений уровня мРНК в таких распространённых типах опухолей, как немелкоклеточный рак лёгкого (НМРЛ) и светлоклеточный почечноклеточный рак (СПР). Определены основные принципы отбора контрольных генов:

1. экспрессия гена должна быть стабильна (данные EST, микрочипов и др.);
2. для гена не должны быть характерны частые нарушения сплайсинга;
3. ген должен относиться к группе «генов домашнего хозяйства» (housekeeping genes);
4. ген должен быть локализован в районе, для которого не характерны частые геномные aberrации при изучаемых типах рака (НМРЛ и СПР);
5. ген не должен содержать частые однонуклеотидные полиморфизмы, особенно в транскрибируемой области.

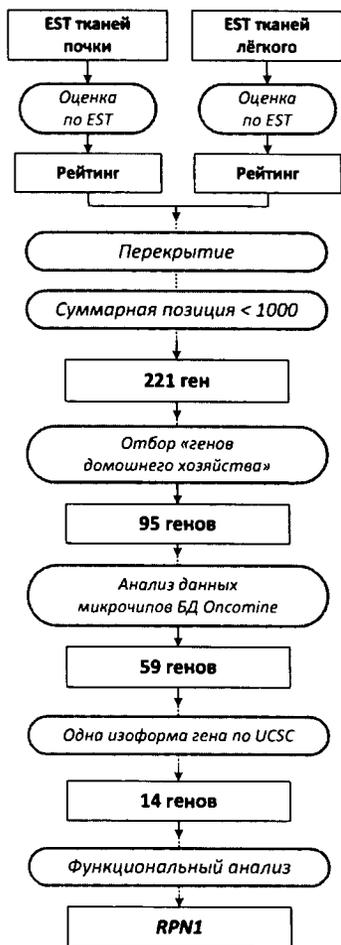


Рис. 1. Схема отбора потенциальных контрольных генов. Составлен рейтинг наиболее перспективных генов-кандидатов. Проведена поэтапная фильтрация генов согласно экспрессионным данным в нормальных тканях, анализу транскриптомной базы данных Oncomine, данным альтернативного сплайсинга (UCSC) и анализу известных функций белков.

Многоэтапный процесс отбора контрольных генов-кандидатов, включающий анализ dbEST, Oncomine, базы данных хромосомных перестроек Mitelman, позволил идентифицировать ген *RPN1* (схема отбора представлена на рис. 1). В качестве начального и основного этапа отбора послужил анализ применимости генов в качестве контрольных на основе построения оценочной функции. При построении этой функции наибольший вес имели компоненты, ответственные за учёт доли клонотек, в которых присутствуют соответствующие данному гену EST, и учёт различий нормированного числа EST в норме и опухоли. Наименьший вес имели компоненты, отвечающие за учёт частоты несоответствий и повторов в EST. Средний вес имели компоненты, ответственные за учёт стандартного отклонения количеств EST между различными клонотеками нормы или опухоли, и учёт изменений частоты несоответствий и нарушений сплайсинга в клонотеках опухоли по сравнению с клонотеками нормальной ткани. При анализе базы данных Oncomine обязательным условием отбора генов-кандидатов являлось присутствие не более одного случая значительного повышения или понижения уровня мРНК этих генов во всех исследованных образцах рака лёгкого и почки.

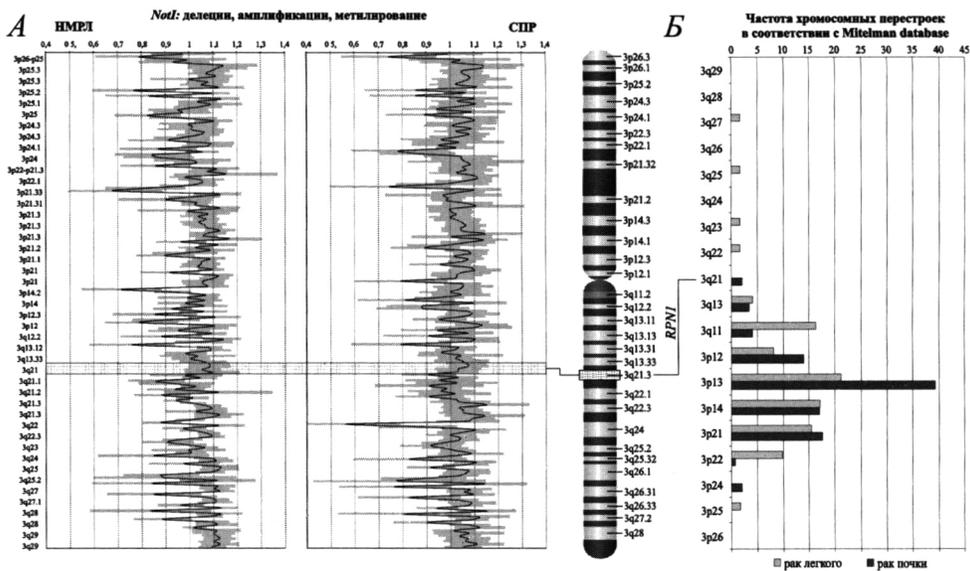


Рис. 2. Частота геномных и эпигеномных нарушений в районах хромосомы 3 при раке легкого и почки. А. Результаты гибридизации ДНК на *NotI*-микрочипах для образцов немелкоклеточного рака лёгкого (НМРЛ) и светлоклеточного почечно-клеточного рака (СПР). Серым цветом показаны стандартные отклонения уровня флуоресценции для различных образцов). Значения в пределах 0,8—1,2 свидетельствуют об отсутствии делеций, амплификаций или метилирования участков в исследуемом районе. Б. Относительная частота хромосомных перестроек при раке легкого и раке почки согласно цитогенетическим данным (Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer). Расположение гена *RPN1* отмечено на хромосоме 3.

При канцерогенезе распространены различные геномные нарушения – мутации, делеции, амплификации, а также гиперметилирование промоторных областей генов-супрессоров опухолевого роста, происходящие уже на ранних стадиях развития опухоли. Поэтому дополнительным свидетельством приемлемости применения гена *RPN1* в качестве контрольного может стать геномная и эпигеномная стабильность соответствующего хромосомного локуса. В соответствии с базой данных Mitelman (<http://cgap.nci.nih.gov/Chromosomes/Mitelman>) ген *RPN1* находится в одном из наиболее стабильных при раке легкого и почки районов (рис. 2).

Ген *RPNI* кодирует мембранный белок рибофорин 1, часть N-олигосахарил-трансферазного комплекса, связывающего обогащённые остатками маннозы олигосахариды с аспарагиновыми остатками в синтезируемых полипептидных цепях. Показано, что рибофорин 1 необходим для N-гликозилирования и доставки субстрата в активный центр олигосахарилтрансферазы (Wilson and High, 2007). Эти процессы играют ключевую роль в большинстве тканей, активно синтезирующих белок, в том числе и опухолях.

Экспериментальная проверка применимости гена RPNI в качестве контрольного

Уровень мРНК генов *RPNI*, *GUSB*, *GAPDH* измерен методом ПЦР-РВ для 28-и образцов НМПЛ, 21-го образца СПР и соответствующих образцов прилегающих гистологически нормальных тканей. Начальную концентрацию образцов мРНК выравняли при помощи спектрофотометра NanoDrop ND-1000 (NanoDrop Technologies Inc.). Методом ПЦР-РВ оценили вариабельность уровня мРНК гена *RPNI* по сравнению с *GAPDH* для НМПЛ и *GUSB* для СПР. Значения эффективных пороговых циклов C_T^{eff} , отражающие начальную концентрацию транскрипта с учетом эффективности реакций, представлены на рис. 3. Средние значения эффективных пороговых циклов для гена *RPNI* достаточно близки между собой, 50% этих значений находится в достаточно узком диапазоне, как для НМПЛ, так и СПР.

Клеточные пути, в которых участвует белковый продукт гена *RPNI*, не связаны с процессами, в которые вовлечены гены *GAPDH* (катализ обмена углеводов) и *GUSB* (гидролиз гликозамингликанов), что позволяет исключить возможность их ко-регуляции и её влияние на результаты. Комбинация генов *RPNI* и *GAPDH* пригодна для нормирования данных при исследовании рака лёгкого, а *RPNI* и *GUSB* – рака почки. Ранее в нашей лаборатории успешно выполнен анализ экспрессии ряда генов-супрессоров опухолевого роста с использованием двух пар контрольных генов, включающих *RPNI*, в

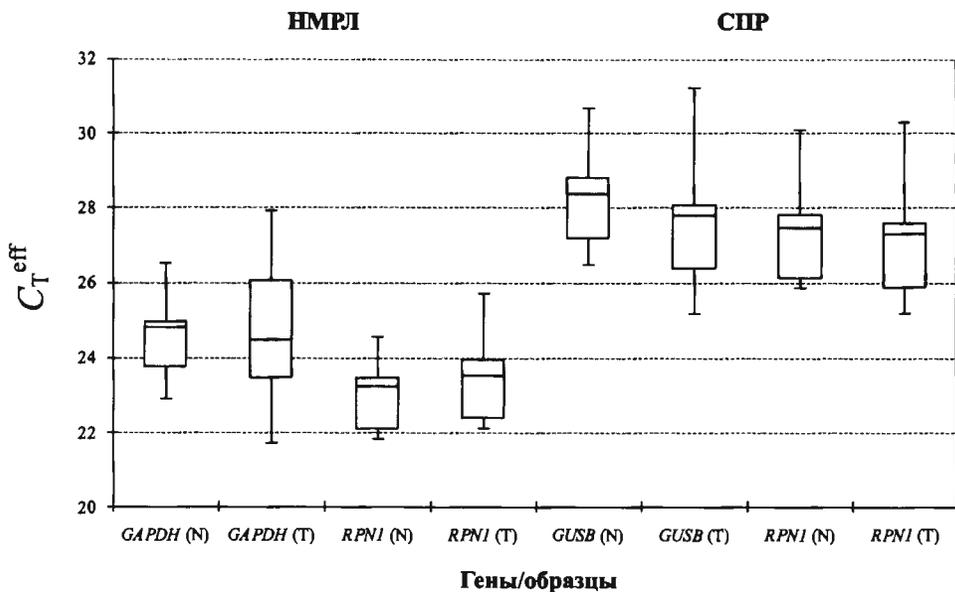


Рис. 3. Эффективные значения порогового цикла C_T^{eff} для контрольных генов *GAPDH* и *GUSB* и потенциального контрольного гена *RPN1* в 28-и образцах НМРЛ и 21-ом образце СПР. Прямоугольником обозначена область, в которую попадает 50% всех значений (между 25-м и 75-м перцентильями). Горизонтальная линия в прямоугольниках – медиана. Ограничивающие отрезки – максимальные и минимальные значения C_T^{eff} .

опухолях лёгкого и почки (Анедченко и др., 2008; Кудрявцева и др., 2009; Senchenko et al., 2010).

Методом сравнительной гибридизации на *NotI*-микрочипах показано отсутствие делеций, амплификаций и/или метилирования геномного локуса *RPN1* во всех 56-и образцах НМРЛ, и только в одном из 42-х образцов СПР обнаружено снижение сигнала (Павлова и др., 2009). Таким образом, совокупность полученных результатов – биоинформатический анализ (рис. 1, 2Б), экспериментальная оценка частоты геномных и эпигеномных нарушений с помощью гибридизации на *NotI*-микрочипах (рис. 2А) и оценка вариабельности уровня мРНК методом ПЦР-РВ (рис. 3) – позволяет предложить *RPN1* в качестве нового контрольного гена для нормирования транскриптомных данных при исследовании НМРЛ и СПР.

Согласно неопубликованным данным Забаровского и соавт., отсутствие делеций и/или метилирования гена *RPNI* характерно и для других видов рака – рака молочной железы, шейки матки, простаты, яичников, толстой кишки, что указывает на возможность его использования в качестве универсального контрольного гена для эпителиальных опухолей. Однако для оценки возможности применения гена *RPNI* при исследовании других видов рака необходимы дополнительные исследования.

Сравнительный анализ экспрессии генов *FTL* и *FTH*, кодирующих субъединицы ферритина, при раке лёгкого

Анализ базы данных dbEST с помощью CompEST выявил потенциальное понижение уровня экспрессии генов лёгкой и тяжёлой цепей ферритина для рака лёгкого. Ферритин – основной переносчик негемового железа в организме, представляющий собой глобулярный белковый комплекс, состоящий из 24 субъединиц (12 тяжёлых субъединиц, *FTH* и 12 лёгких – *FTL*) и выполняющий роль основного внутриклеточного депо железа. Многие опухолевые заболевания человека сопровождаются увеличением содержания в сыворотке крови этого белка, однако, возможные изменения экспрессии этих генов на уровне транскрипции в опухолях изучены недостаточно (Milman and Pedersen, 2002; Kashyap et al., 2005), в том числе при немелкоклеточном раке лёгкого (Fracchia et al., 1999; Zaleska et al., 2005). В связи с этим в ряде работ выдвинуто предположение, что при различных видах рака повышенный уровень сывороточного ферритина может быть связан с его синтезом непосредственно клетками самой опухоли. Однако данных об изменении содержания ферритина в опухолевых клетках на настоящий момент нет – экспрессия генов *FTL* и *FTH*, как на уровне транскрипции, так и синтеза белка непосредственно в опухолевых тканях лёгкого остается малоизученной.

В связи с этим в нашей лаборатории проведено исследование изменения уровня мРНК этих генов в образцах плоскоклеточного рака лёгкого (ПКРЛ) методом ПЦР в реальном времени. Показано существенное (более 2

раз) и частое снижение уровня мРНК генов *FTL* и *FTH* (рис. 4) непосредственно в первичной опухоли уже на первой, а также на последующих стадиях заболевания – в среднем в 11 и 9 раз в 83% и 73% случаев, соответственно.

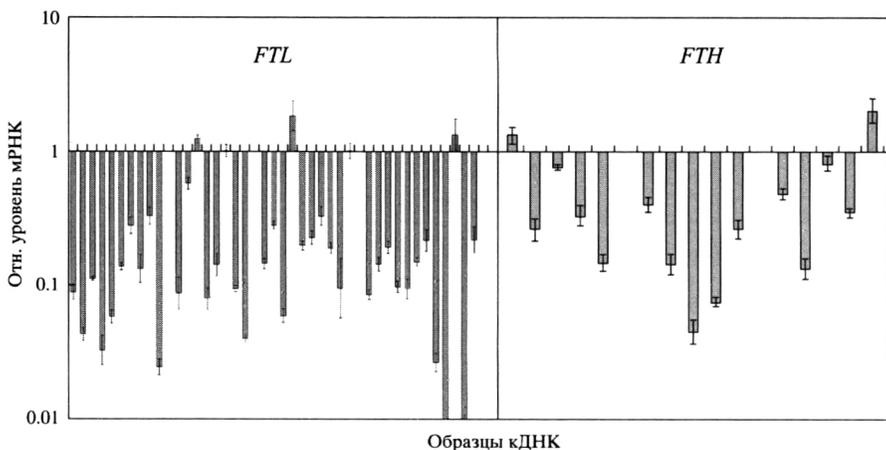


Рис. 4. Относительный уровень мРНК генов *FTL* и *FTH* в образцах ПКРЛ.

Однако далеко не всегда изменения содержания мРНК приводят к соответствующим изменениям количества белка. Одним из основных факторов, определяющих соотношение между содержанием мРНК и белка в клетке, является микроРНК. Повышение содержания в клетке определённой микроРНК негативно сказывается на трансляции мРНК-мишени. При помощи веб-сервисов miRanda (<http://www.microrna.org/>) и TargetScan (<http://www.targetscan.org/>) определены наиболее вероятные микроРНК, отвечающие за регуляцию трансляции мРНК генов *FTL* и *FTH*. В соответствии с полученными данными, экспрессия гена *FTL* регулируется тремя микроРНК, содержание двух из которых – hsa-miR-22 и hsa-miR-24 – повышено при раке лёгкого (Cheng et al., 2005; Gao et al., 2010; Liu et al., 2010). Кроме того, для hsa-miR-22 отмечено онкогенное воздействие на клетки бронхов (Liu et al., 2010). Регуляция трансляции гена *FTH* осуществляется более чем десятью различными микроРНК, для одной из которых, hsa-miR-186, выяв-

лено существенное увеличение содержания при аденокарциноме лёгкого (Tang et al., 2010; Zhang et al., 2010).

В то же время, анализ протеомных данных для различных видов рака (базы данных PRIDE, Peptidome, Tranche и др.) показал лишь небольшое число случаев обнаружения белков FTL и FTH в опухолевых образцах. Таким образом, вероятнее всего, подавление экспрессии генов *FTL* и *FTH* происходит не только на уровне транскрипции, но и на уровне трансляции. Это говорит о том, что повышение содержания ферритина в сыворотке крови, по-видимому, обусловлено повышенным синтезом FTL и FTH в других тканях организма, либо ускоренным распадом клеток и высвобождением цитоплазматического ферритина.

ВЫВОДЫ

1. Разработан биоинформатический метод поиска генов-онкомаркеров, основанный на анализе базы данных dbEST, позволяющий идентифицировать гены с потенциально изменённым в опухоли уровнем экспрессии и часто встречающимися нарушениями сплайсинга. Подход реализован в Windows-приложении CompEST.
2. С помощью CompEST идентифицировано более 100 потенциальных генов-маркеров различных видов рака – толстой кишки, молочной, предстательной и поджелудочной желёз, почки, печени, лёгкого, глиобластомы и меланомы (в том числе 19 генов-маркеров, локализованных на коротком плече хромосомы 3).
3. Разработан новый метод расчёта эффективности ПЦР в реальном времени, использующий аппроксимацию кинетической кривой уравнением, и реализованный в программе «АТГ» (Анализ Транскрипции Генов).
4. Впервые идентифицирован контрольный ген *RPN1*, пригодный для нормирования количественных данных при исследовании рака лёгкого и почки.
5. Выборочная экспериментальная оценка экспрессии отобранных при помощи CompEST генов-маркеров толстой кишки (*ZG16*, *DES*, *SLC26A3*) подтвердила частое и значительное снижение уровня их мРНК. Для гена *ZG16* результаты получены впервые.
6. Предсказанное при помощи CompEST снижение уровня мРНК генов лёгкой и тяжёлой цепей ферритина (*FTL* и *FTH*) при раке лёгкого подтверждено методом ПЦР в реальном времени.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Г. С. Краснов, Н. Ю. Опарина, А. А. Дмитриев, А. В. Кудрявцева, Е. А. Анедченко, Т. Т. Кондратьева, Е. Р. Забаровский, В. Н. Сенченко. 2011. Новый контрольный ген *RPNI* для нормирования количественных данных при раке лёгкого и почки. *Молекулярная биология*. **45**, 238-248.
2. V.N. Senchenko, E.A. Anedchenko, T.T. Kondratieva, G.S. Krasnov, A.A. Dmitriev, V.I. Zabarovska, T.V. Pavlova, V.I. Kashuba, M.I. Lerman, E.R. Zabarovsky. 2010. Simultaneous down-regulation of tumor suppressor genes *RBSP3/CTDSPL*, *NPRL2/G21* and *RASSF1A* in primary non-small cell lung cancer. *BMC Cancer*. **10**, 75.
3. Г.С. Краснов, А.А. Дмитриев, А.В. Кудрявцева, Е.А. Анедченко, Н.Ю. Опарина, В.Н. Сенченко. 2010. Новые программы, используемые при количественной оценке копийности и уровня транскрипции генов. *Труды МФТИ*. **2**, 23-27.
4. А.В. Кудрявцева, Е.А. Анедченко, Н.Ю. Опарина, Г.С. Краснов, К.Н. Кашкин, А.А. Дмитриев, И.Б. Зборовская, Т.Т. Кондратьева, Т.В. Виноградова, М.В. Зиновьева, Е.П. Копанцев, В.Н. Сенченко. 2009. Экспрессия генов *FTL* и *FTH*, кодирующих субъединицы ферритина при раке лёгкого и почки. *Молекулярная биология*. **43**, 1044-1054.
5. Е. А. Анедченко, А. А. Дмитриев, Г. С. Краснов, Т. Т. Кондратьева, Е. П. Копанцев, Т. В. Виноградова, М. В. Зиновьева, И. Б. Зборовская, Б. Е. Полоцкий, О. В. Сахарова, В. И. Кашуба, Е. Р. Забаровский, В. Н. Сенченко. 2008. Подавление экспрессии генов *RBSP3/CTDSPL*, *NPRL2/G21*, *RASSF1A*, *ITGA9*, *HYAL1* и *HYAL2* при немелкоклеточном раке лёгкого. *Молекулярная биология*. **42**, 965-976.
6. Г.С. Краснов, Н.Ю. Опарина. Программа «Идентификация изменений транскриптома» («CompEST»). Номер регистрационного свидетельства 2008613673.

Заказ № 138-а/04/11 Подписано в печать 15.04.2011 Тираж 100 экз. Усл. п.л. 1,2



ООО "Цифровичок", тел. (495) 649-83-30
www.cfr.ru ; e-mail: info@cfr.ru

18-