

УДК 577.214.6

## ВЫБОР НАДЕЖНЫХ КОНТРОЛЬНЫХ ГЕНОВ ДЛЯ КОЛИЧЕСТВЕННОЙ ПЦР В РЕАЛЬНОМ ВРЕМЕНИ НА ОБРАЗЦАХ НЕОПУХОЛЕВОЙ ТКАНИ ЖЕЛУДКА ЧЕЛОВЕКА

© 2012 г. А. Bирerdinc<sup>1,2</sup>, R. Mehta<sup>1,2</sup>, R. Alhussain<sup>1,3</sup>, A. Afendi<sup>2,4</sup>,  
V. Chandhoke<sup>1,2</sup>, Z. Younossi<sup>1,2,4</sup>, А. Баранова<sup>1,2,5\*</sup>

<sup>1</sup>School of Systems Biology, George Mason University, Fairfax, VA, USA

<sup>2</sup>Translational Research Institute, Inova Health System, Falls Church, VA, USA

<sup>3</sup>Biology Department, College of Science, George Mason University, Fairfax, VA, USA

<sup>4</sup>Center for Liver Diseases, Inova Fairfax Hospital, Falls Church, VA, USA

<sup>5</sup>Медико-генетический научный центр Российской академии медицинских наук, Москва, 115478, Россия

Поступила в редакцию 04.04.2011 г.

Принята к печати 07.06.2011 г.

Количественная ПЦР в реальном времени – стандартный метод измерения изменений экспрессии генов относительно друг друга при изучении многофакторных заболеваний человека, в том числе ожирения и гастрита. Однако различия в количестве исходного материала, активности фермента и концентрациях ингибиторов амплификации часто приводят к ошибочной оценке уровня экспрессии гена, поэтому успех исследования зависит от правильного нормирования данных. Среди известных стратегий нормирования наиболее распространено использование в качестве внутреннего контроля стабильно экспрессирующихся генов (референсных, или контрольных генов). В случае тканей желудка человека такие гены изучены недостаточно. В представленной работе стабильность экспрессии референсных генов изучали с использованием трех популярных алгоритмов – *geNorm*, *NormFinder* и *BestKeeper*. Показано, что для определения уровней экспрессии генов методом количественной ПЦР в реальном времени в неопухолевых тканях желудка наилучшей является пара генов *HPRT1* и *GAPDH*.

**Ключевые слова:** количественная ПЦР в реальном времени, гены домашнего хозяйства клетки, внутренний контроль, желудок, ожирение.

SELECTION OF RELIABLE REFERENCE GENES FOR qRT-PCR ANALYSIS IN HUMAN NON-CANCEROUS GASTRIC TISSUE, by A. Bирerdinc<sup>1,2</sup>, R. Mehta<sup>1,2</sup>, R. Alhussain<sup>1,3</sup>, A. Afendi<sup>2,4</sup>, V. Chandhoke<sup>1,2</sup>, Z. Younossi<sup>1,2,4</sup>, A. Baranova<sup>1,2,5\*</sup> (1School of Systems Biology, George Mason University, Fairfax, VA, USA; \*e-mail: abaranov@gmu.edu; 2Translational Research Institute, Inova Health System, Falls Church, VA, USA; 3Biology Department, College of Science, George Mason University, Fairfax, VA, USA; 4Center for Liver Diseases, Inova Fairfax Hospital, Falls Church, VA, USA; 5Research Center for Medical Genetics, Russian Academy of Medical Sciences, Moscow, 115478 Russia). Real-time PCR (qRT-PCR) is the standard method for studying changes in relative gene expression in complex diseases like obesity and gastritis. However, variations in amount of starting material, enzymatic efficiency and presence of amplification inhibitors can lead to quantification errors. Hence, the need for accurate data normalization is vital. Among several known strategies for data normalization, the use of reference genes as an internal control is the most common approach. Human gastric tissue has been the least investigated for stability of reference gene expression. In this study, three popular algorithms, *GeNorm*, *NormFinder* and *BestKeeper* were used to evaluate the reference gene stability. Conclusion: *HPRT1* and *GAPDH* are the best performing pair of reference genes for qRT-PCR profiling experiments involving non-malignant gastric tissue samples.

**Keywords:** qRT-PCR, housekeeping genes, endogenous control, stomach, obesity.

До настоящего времени молекулярные события в ткани желудка человека изучали преимущественно на опухолевом материале, например на образцах рака желудка. Однако за последние годы

опубликовано много работ, убедительно указывающих на физиологическое значение изменений паттерна генной экспрессии для развития неопухолевых патологий желудка.

\* Эл. почта: abaranov@gmu.edu

Два наиболее важных заболевания человека, при которых возникают изменения в тканях желудка – гастрит и ожирение. Более того, ожирение приводит к развитию многочисленных заболеваний, таких как сахарный диабет типа 2, атеросклероз, инсулинорезистентность, неалкогольное ожирение печени (NAFLD) и синдром поликистоза яичников [1]. Недавно убедительно показали важность гормонов и рецепторов ткани желудка в регуляции чувства сытости [1–5]. Установлено, что в желудке экспрессируются многие индуцируемые гены, в том числе гены лептина [6], грелина [3], пептида YY, глюкагона, глюкагоноподобных пептидов-1 и -2 (GLP-1 и -2), оксинтомодулина [2] и других паракринных гормонов, которые влияют на уровень глюкозы в крови, экзокринную секрецию, функции адипоцитов, мышечную активность желудка и другие физиологические реакции [2]. Изучение роли ткани желудка в энергетическом гомеостазе, в патогенезе ожирения и сопутствующих заболеваний важно и потому, что за последние 5 лет резко возросло число бариатрических хирургических вмешательств. Только в США количество таких процедур достигает 220000 в год [7]. В дальнейшем число бариатрических процедур будет увеличиваться, поскольку получены доказательства их положительного влияния на общее состояние организма, в том числе при сахарном диабете типа 2 [8, 9] и сердечно-сосудистой системы [10]. В обоих случаях изменения экспрессии генов в ткани желудка, вызванные хирургическим вмешательством, требуют дальнейшего изучения.

Острый и хронический гастрит считаются наиболее важными последствиями инфицирования желудка *Helicobacter pylori*, которое встречается у 7–87% лиц с желудочно-кишечной патологией в зависимости от выборки [11]. Оставляя в стороне значительное прямое влияние этой инфекции на заболеваемость, а в случае пептической язвы и на смертность, хроническую инфекцию слизистой оболочки желудка *H. pylori* можно рассматривать как важный фактор риска рака желудка [12]. Показана также связь хеликобактерного гастрита с хронической изжогой (кислотным рефлюксом) [13], хронической крапивницей, тромбоцитопенической пурпурой [14] и другими заболеваниями. Очевидно, что *H. pylori* способствует развитию всех этих состояний, нарушая нормальное функционирование слизистой оболочки желудка и влияя на экспрессию генов в тканях желудка. Простого контакта живых бактерий *H. pylori* с клетками человека достаточно, чтобы вызвать в них экспрессию гастрина, важного регулятора секреции желудочного сока и клеточного роста [15]. В слизистой желудка при инфицировании *H. pylori* повышается секреция интерлейкинов -8, -17A и -17F [16]. Продукт гена A, ассоциированного с цитотоксином *H. pylori* (CagA), индуцирует

транскрипционную активность  $\beta$ -катенина в клеточном ядре [17].

Изучение таких распространенных патологий, как гастрит и ожирение, требует определения, часто одновременного, профиля экспрессии многих генов [18]. Высокая чувствительность и специфичность ПЦР в реальном времени, а также широкий диапазон измерений позволили этому методу стать наиболее популярным для определения профилей генной экспрессии [18, 19]. Однако количественное определение РНК с помощью количественной ПЦР в реальном времени часто затруднено из-за вариаций в количестве исходного материала, относительной эффективности полимеразы и присутствия различных ингибиторов [20]. Эти проблемы можно преодолеть при помощи точного нормирования данных [21].

Среди многих стратегий нормирования данных [22] наиболее распространено использование стабильно экспрессируемых (референсных) генов в качестве внутреннего контроля [19]. Референсными генами при изучении тканей человека часто служат гены 18S рРНК (18S), глицеральдегид-3-фосфат-дегидрогеназы (GAPDH),  $\beta$ 2-микроглобулина (B2M), РНК-полимеразы II (RP11, или POLR2A), тирозин-3-монооксигеназы/триптофан-5-монооксигеназу-активирующего белка, полипептида  $\zeta$  (YWHAZ), убиквитина C (UBC) и гипоксантин-фосфорибозилтрансферазы 1 (HPRT1) [19, 23, 24].

В качестве референсных стараются выбирать такие гены, экспрессия которых остается стабильной во всех изучаемых выборках [25]. Тем не менее, экспрессия часто используемых контрольных генов варьирует в зависимости от типа ткани и ее физиологического состояния [20, 24]. Если ген, служащий внутренним контролем, не оптимален для данной выборки и уровень его экспрессии варьирует, то результаты могут серьезно искажаться [26]. Таким образом, пригодность данного гена зависит от типа ткани или клеточной системы [24, 25, 27]. Это делает очевидной необходимость проверки приемлемости генов, используемых в качестве референсных, для каждого изучаемого типа ткани. Предложено несколько математических алгоритмов определения стабильности экспрессии референсных генов [28].

Следует отметить, что стабильность экспрессии референсных генов в ткани желудка человека изучена совершенно недостаточно. До настоящего времени проведено только одно систематическое исследование такого типа [29], в котором определен профиль экспрессии шести референсных генов в шести клеточных линиях рака желудка и в 20 парных образцах (норма/опухоль) ткани желудка. Несмотря на важность этой работы для понимания проблемы в целом, применимость ее результатов ограничена, поскольку в выборку во-

**Таблица 1.** Праймеры для амплификации шести использованных референсных генов

Ген	Номер гена	$T_{пл}$ , °C	Праймер
<i>GAPDH</i>	NM_002046.2	55	F-5'ACAGTCAGCCGCATCTTCTT R-5'GACAAGCTTCCCGTTCTCAG
<i>HPRT1</i>	NM_000194.1	55	F-5'AAGCTTGCTGGTGAAGGA R-5'AAGCAGATGGCCACAGAACT
<i>YWHAZ</i>	NM_001135702.1	55	F-5'ACTTTTGGTACATTGTGGC R-5'CCGCCAGGAAAACCAGT
<i>18S</i>	NR_003286.2	60	F-5'AGGAATTCCAGTAAGTGCG R-5'GCCTCACTAAACCATCCAA
<i>RPII</i>	NM_000937.3	60	F-5'CTTCACGGTGCTGGGCATT R-5'GTGCGGCTGCTTCCATAA
<i>UBC</i>	XM_002344708.1	60	F-5'CCTGGTGCTCCGTCTTAGAG R-5'TTTCCCAGCAAAGATCAACC

шли только образцы озлокачественных тканей или тканей, окружающих опухоль, и, возможно, подвергшихся аберрантному метилированию ДНК. Как показано ранее, нарушение метилирования ДНК часто обнаруживается в тканях, окружающих раковую опухоль желудка, и считается эпигенетической предпосылкой злокачественного перерождения [30].

В представленной работе пригодность референсных генов для оценки уровня экспрессии генов в тканях желудка определяли в образцах тканей, полученных во время бариатрических операций у больных ожирением и сопутствующими неопухолевыми заболеваниями желудка. Наиболее стабильно экспрессируемые контрольные гены выявляли с использованием программ *geNorm* [21], *NormFinder* [31] и *BestKeeper* [32]. Референсные гены, признанные наилучшими всеми тремя программами, могут быть рекомендованы для использования в качестве внутреннего контроля при определении профиля экспрессии генов в неопуховых тканях желудка человека методом количественной ПЦР в реальном времени.

### ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**Выборки.** Образцы тканей дна желудка получены от 22 человек с ожирением и диагнозом гастрит ( $n = 10$ ) и без этого диагноза ( $n = 12$ ) во время бариатрических или иных операций, погружены в жидкий азот и хранились при  $-80^{\circ}\text{C}$ . Образцы получены от лиц в возрасте  $43.8 \pm 9.9$  лет (среднее  $\pm$  SD) с индексом массы тела  $47.0 \pm 7.9$  (среднее  $\pm$  SD), из которых 73% женщины. У 35% больных диагностирован гастрит не-*H. pylori*, у 9% – гастрит с *H. pylori*. У 57% больных гастрит отсутствовал. Сахарный диабет типа 2 диагностирован у 26% больных, а неалкогольный стеатоз – у 70%. Образцы деидентифицированы в соответ-

ствии с правилами HIPAA. Исследование одобрено Независимым Экспертным Советом (IRB) Госпиталя Инова.

**Выбор референсных генов.** В качестве референсных выбраны следующие гены, ранее описанные как гены домашнего хозяйства: *18S* (18S рРНК), *GAPDH* (глицеральдегид-3-фосфат–дегидрогеназа), *HPRT1* (гипоксантин-гуанин-фосфорибозил-трансфераза I), *YWHAZ* (тирозин-3-монооксигеназа/триптофан-5-монооксигеназу-активирующий белок, полипептид  $\zeta$ ), *UBC* (убиквитин C) и *RPII*, или *POLR2A* (РНК-полимераза II).

Праймеры для генов *GAPDH* и *HPRT1* получены от Real Time Primers (“Real Time Primers”, США). Праймеры к *18S* синтезированы как описано ранее [33]. Праймеры для остальных трех генов, *RPII*, *YWHAZ* и *UBC*, подобраны с помощью программы Oligo Perfect Designer (“Invitrogen”) таким образом, чтобы перекрыть хотя бы одну интрон-экзонную границу, и синтезированы коммерчески (“Invitrogen”). Специфичность всех праймеров подтверждена с помощью NCBI Blast против всех доступных нуклеотидных последовательностей мРНК человека. Номера генов и структура праймеров приведены в табл. 1.

**Выделение РНК и обратная транскрипция.** Суммарную РНК выделяли из тканей дна желудка ( $n = 22$ ) с помощью набора Qiagen RNeasy kit (“Qiagen”) согласно протоколу производителя. Чистоту РНК определяли, измеряя соотношение поглощения раствора РНК при 260/280 нм с ожидаемыми значениями в интервале 1.8–2.00, с помощью спектрофотометра GeneQuant1300 (“GE Healthcare”, США). Целостность РНК подтверждали электрофорезом в 1%-ном агарозном геле с бромистым этидием. Обратную транскрипцию проводили с помощью набора “RT<sup>2</sup> first strand” (“SABiosciences”, США), используя 112 нг сум-

марной РНК в каждой реакции. Согласно протоколу производителя суммарную РНК обрабатывали ДНКазой для удаления геномной ДНК. В качестве затравки для обратной транскрипции использовали случайные гексамеры и олиго-dT-праймеры.

**Количественную ПЦР в реальном времени** выполняли на приборе Bio-rad CFX96 Real Time System (“BioRad Laboratories”, США) в 96-луночном формате. Смесь для ПЦР состояла из 5 мкл кДНК, что соответствует приблизительно 600 нг суммарной РНК, 0.1 мкМ праймеров от Real-TimePCRPrimers.com или 0.2 нМ праймеров Invitrogen и 1X Sso Fast Evagreen Supermix (“BioRad”) в конечном объеме 15 мкл. Для выявления возможной контаминации реагентов и/или присутствия геномной ДНК в каждом опыте добавляли контрольные лунки без матрицы и/или полимеразы. Температурный профиль реакции, повторенной в течение 50 циклов, был следующим: 1) 95°C в течение 10 мин; 2) денатурация – 10 с при 95°C; отжиг – 40 с при 55°C для Real Time-праймеров и 60°C для праймеров Invitrogen (данные амплификации обрабатывали в конце каждого шага); 3) кривая диссоциации – инкубация в течение 10 с при 95°C и 5 с при 65°C с последующим подъемом температуры до 95°C (“BioRad CFX96 Real Time System”). Специфичность продуктов определяли с использованием кривых плавления. Все образцы, составившие выборку, по крайней мере трижды амплифицировали на матрице одного и того же препарата суммарной РНК. Для дальнейшего анализа брали средние значения.

**Определение стабильности экспрессии референсных генов.** Стабильность экспрессии референсных генов в образцах оценивали с использованием трех различных статистических алгоритмов – *geNorm*, v3.4, *NormFinder* v0.953 и *BestKeeper* v1, описанных ранее [21, 31, 34]. Программу *geNorm* использовали для вычисления уровня стабильности ( $M$ ) референсных генов с помощью исходных ненормированных значений. В терминах стандартного отклонения ( $V_{jk}$ ) для каждой пары генов *geNorm* вычисляет попарную вариацию логарифмически трансформированного отношения уровней экспрессии данного гена ( $a_{ij}$ ) в каждом образце ткани ( $m$ ), при любой комбинации генов внутреннего контроля ( $j$  или  $k$ ) [21]:

$$(\forall j, k \in [1, n] \text{ и } j \neq k):$$

$$A_{jk} = \left\{ \lg_2\left(\frac{a_{1j}}{a_{1k}}\right), \lg_2\left(\frac{a_{2j}}{a_{2k}}\right), \dots, \lg_2\left(\frac{a_{mj}}{a_{mk}}\right) \right\} =$$

$$= \left\{ \lg_2\left(\frac{a_{ij}}{a_{ik}}\right) \right\}_{i=1 \rightarrow m},$$

$$V_{jk} = st.dev(A_{jk}).$$

Меру стабильности экспрессии ( $M_j$ ) вычисляли как среднюю попарную вариацию экспрессии данного гена по сравнению со всеми другими генами [21]:

$$M_j = \frac{\sum_{k=1}^n V_{jk}}{n-1}.$$

Гены с наименее стабильной экспрессией (т.е. с наибольшими значениями  $M_j$ ) последовательно удаляли при каждой итерации исключения. Эту процедуру повторяли до тех пор, пока не остались только гены с наименьшими значениями  $M_j$  и самой стабильной экспрессией. Фактор нормирования вычисляли на основании среднего геометрического для конечного оптимального набора референсных генов.

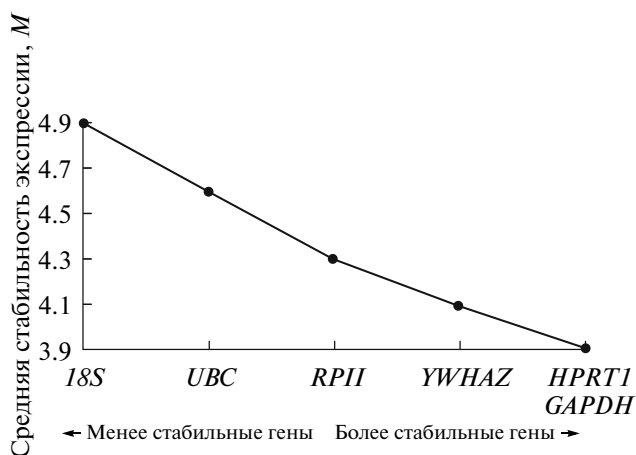
Для оценки общей изменчивости экспрессии генов-кандидатов и межгрупповой вариации в выборочных подгруппах использовали модуль *NormFinder* для программы Excel с применением следующей модели [31]:

$$y_{igj} = \alpha_{ig} + \beta_{gi} + \varepsilon_{igj}.$$

Эта модель включает в себя три следующих компонента: общий уровень экспрессии  $i$ -го гена-кандидата внутри группы  $g$  ( $a_{ig}$ ), количество образцов мРНК в выборке  $j$  ( $b_{gj}$ ) и случайную вариацию, вызванную биологическими и экспериментальными факторами ( $e_{igj}$ ). Цель состоит в нахождении двух генов с наименьшими уровнями внутри- и межгрупповой дисперсии экспрессии. Доверительный интервал для межгрупповой дисперсии определен как средняя от внутригрупповых дисперсий и представлен как величина ошибки межгрупповой дисперсии. Алгоритм программы сочетает внутригрупповую ( $s_{ig}^2$ ) и межгрупповую ( $z_{ig} - q_g, g = 1, \dots, G$ ) дисперсию и выражает уровень стабильности ( $r_{ig}$ ) каждого исследованного гена, где  $g^2$  – вариация уровней экспрессии ( $a_{ig}$ ) [31]:

$$\rho_{ig} = \frac{\hat{\gamma}^2 |d_{ig}|}{\hat{\gamma}^2 + \hat{\sigma}_{ig}^2/n_g} + \sqrt{\frac{\hat{\sigma}_{ig}^2/n_g + \frac{\hat{\gamma}^2 \hat{\sigma}_{ig}^2}{n_g}}{\hat{\gamma}^2 + \hat{\sigma}_{ig}^2/n_g}}.$$

Это выражение учитывает множественные источники вариации экспрессии и отражает общую систематическую ошибку в пересчете на один ген. Таким образом, гены, имеющие наивысший ранг (т.е. наименьший уровень стабильности, следовательно, наименьшую общую дисперсию), это наиболее стабильно экспрессируемые в данной выборке референсные гены-кандидаты. Однако поскольку уровень системной ошибки ( $r_{ig}$ ) вычислен в соответствии с нуль-гипотезой о независимости уровня экспрессии каждого гена друг от друга, проведена дальнейшая оценка внутри- и



**Рис. 1.** Стабильность экспрессии генов ( $M$ ) референсных генов-кандидатов в ткани желудка, вычисленная программой *geNorm*. Программа последовательно исключает гены с относительно высокой вариабельностью экспрессии между образцами.  $M$  представляет собой среднее от уровней стабильности остающихся генов: чем ниже величина  $M$ , тем стабильнее гены в подгруппе.

межгрупповой изменчивости. Как можно видеть из математического описания модели, увеличение числа образцов приводит к увеличению устойчивости результатов.

Алгоритм *BestKeeper* был использован для независимого подтверждения результатов, полученных с помощью алгоритмов *geNorm* и *NormFinder*. Этот алгоритм вычисляет вариацию экспрессии каждого из контрольных генов на основании точки перекреста ( $CP$ ), которая определяется как число циклов, необходимых для достижения выбранного порога флуоресценции [34]. Начальный анализ исходных значений  $CP$  включает определение стандартного отклонения ( $SD \pm CP$ ) и коэффициента вариации  $CV$  ( $\% CP$ ) каждого из референсных генов. Далее эти величины используют для ранжирования референсных генов, отличающихся наиболее стабильной экспрессией: от обнаруживающих самую низкую изменчивость до наименее стабильных с самой высокой изменчивостью. Все стабильно экспрессирующиеся референсные гены объединяются в индекс *BestKeeper*, рассчитываемый для выборки образцов с использованием среднего геометрического значений  $CP$  для каждого гена-кандидата [32]:

$$\text{индекс } BestKeeper = \sqrt[n]{CP_1 \times CP_2 \times CP_3 \times \dots \times CP_z}.$$

*BestKeeper* оценивает межгенные взаимодействия каждого референсного гена-кандидата, отражая величину вклада каждого референсного гена в общий индекс. Поскольку наличие в выборке выбивающихся значений может понизить точность анализа, *BestKeeper* определяет целостность выборки ( $InVar$ ) для различий в относительных

значениях  $CP$  ( $n$ ) и средних значений  $CP$  каждого референсного гена ( $m$ ) [32]:

$$InVar_m[\pm CP] = \frac{1}{n-1} \sum_{i=1}^n (CP_n^m - \text{среднее } CP_n)^2.$$

*BestKeeper* также определяет параметр целостности выборки для каждого референсного гена с помощью следующего анализа [32]:

$$InVar_m[\pm x\_fold] = E_m^{InVar[\pm CP]}.$$

В программе *BestKeeper* приемлемыми признаются только выборки со скорректированной по выходу внутренней дисперсией в пределах трехкратного повышения или понижения экспрессии. Следовательно, выпадающие значения элиминируются, что увеличивает надежность индекса *BestKeeper*.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В нашей работе определен профиль экспрессии шести референсных генов, наиболее часто используемых при изучении тканей человека методом количественной ПЦР в реальном времени [33, 34], в том числе генов *18S*, *ACTB* ( $\beta$ -актин), *GAPDH*, *B2M*, *HPRT1*, *YWHAZ*, *UBC* и *RPII*, или *POLR2A*. Стабильность экспрессии этих генов анализировали в 22 образцах тканей дна желудка человека с помощью трех популярных алгоритмов: *geNorm* v3.4 [21], *NormFinder* [31] и *BestKeeper* [32]. Стабильность экспрессии каждого гена в каждом образце ткани оценивали с использованием среднего из трех измеренных значений  $Ct$ . Входные данные для алгоритмов *geNorm* и *NormFinder* из исходных значений  $Ct$  переводили в относительные количества РНК при помощи метода сравнения  $Ct$ , тогда как в алгоритме *BestKeeper* в качестве входных данных использовали исходные значения  $Ct$  [28].

### Алгоритм *geNorm*

Программа *geNorm*, которая ранжирует референсные гены-кандидаты в соответствии с относительным уровнем их экспрессии, разработана Vandesompele et al [21]. Эта программа позволяет ранжировать гены по стабильности их экспрессии — от наименее стабильного к наиболее стабильному. С помощью этой программы получено следующее ранжирование: *18S*  $\rightarrow$  *UBC*  $\rightarrow$  *RPII*  $\rightarrow$  *YWHAZ*  $\rightarrow$  *GAPDH* и *HPRT1*. Значения  $M$ , определенные для этих генов в тканях желудка, составили 4.90, 4.60, 4.30, 4.09 и 3.9 соответственно (рис. 1). С помощью последовательной элиминации наименее стабильных генов на основе их значений  $M$  гены *HPRT1* и *GAPDH* посчитали самыми стабильными референсными генами. Основываясь на попарной вариации, вычисленной для

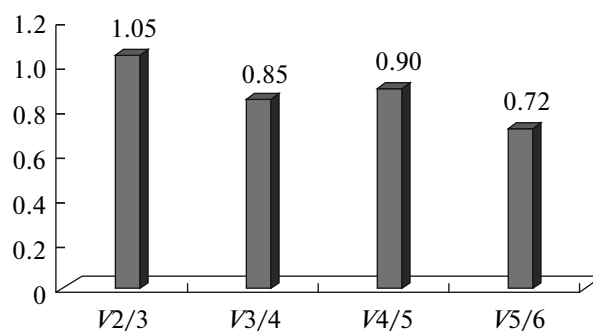
двух последовательных факторов нормирования ( $NF_n$  и  $NF_{n+1}$ ), *geNorm* определил, что при изучении ткани желудка для надежного контроля достаточно двух референсных генов, используемых одновременно (рис. 2).

**Алгоритм NormFinder**

В отличие от *geNorm* программа *NormFinder* оценивает стабильность экспрессии на основании математической модели [31]. Эта программа располагает гены-кандидаты по их минимальной меж- и внутригрупповой вариации. Другими словами, в дополнение к стабильности генов внутри выборок эта программа оценивает стабильность генов в разных выборках и подвыборках. В нашей работе выборка образцов ( $n = 22$ ) была разделена на две – с гастритом ( $n = 10$ ) и без гастрита ( $n = 12$ ). При этом *GAPDH* и *HPRT1* со значениями стабильности 0.356 считали наиболее высокоранговыми на основе комбинации наименьшей меж- и внутригрупповой дисперсии (табл. 2). Дальнейшее экспертное изучение генов с наименьшими значениями меж- и внутригрупповой вариации показало, что стабильность гена *18S*, имевшего второй снизу уровень межгрупповой вариации после *GAPDH*, характеризуется наиболее высоким уровнем внутригрупповой вариации. Следовательно, стабильность гена *18S* невелика (0.868). С другой стороны, ген *HPRT1* показал больший, чем у *18S*, уровень межгрупповой вариации, но самый низкий уровень внутригрупповой вариации. Следовательно, алгоритмом *NormFinder* ген *HPRT1* был отобран как стабильный. Таким образом, результаты, полученные с помощью алгоритма *NormFinder*, хорошо согласуются с оценкой программы *geNorm*.

**Алгоритм BestKeeper**

Алгоритм *BestKeeper* был выбран в качестве независимого метода ранжирования, позволяюще-



**Рис. 2.** Определение оптимального количества референсных генов для нормирования с помощью анализа попарной вариации программой *geNorm*. Высота столбцов выражает уровень изменения фактора нормирования после включения дополнительного референсного гена. Большая вариабельность отражает большой эффект добавленного гена и означает, что данный ген должен быть включен в вычисление фактора нормирования. Авторы программы *geNorm* предлагают рассматривать значение  $V > 0.15$  как пороговое для включения в анализ дополнительного референсного гена.

го осуществлять не прямое сравнение результатов, полученных с помощью программ *geNorm* и *NormFinder* [32]. В отличие от *geNorm* и *NormFinder* в качестве входных данных *BestKeeper* использует исходные значения  $St$  для каждого гена. Согласно этому алгоритму, стабильные референсные гены выбирали с помощью двух главных критериев: низкой дисперсии (SD) и высокой корреляции (значения  $r$  и  $p$ ) в образцах [32].

Анализ численной дисперсии (SD ( $\pm St$ ) и коэффициента вариации  $CV$  ( $\%St$ )) для всех референсных генов-кандидатов во всех 22 образцах позволил провести ранжирование относительной стабильности генной экспрессии. Из шести потенциальных референсных генов наибольшую вариабельность экспрессии проявили *18S* ( $CV = 22.73$ ) и *UBC* ( $CV = 8.42$ ), поэтому они были исключены. На следующем этапе с использованием коэффи-

**Таблица 2.** Сравнение списков генов с наивысшим рангом, полученных тремя различными алгоритмами

Ген	<i>BestKeeper</i> , коэффициент корреляции $R$	<i>geNorm</i> , стабильность экспрессии, $M$	<i>NormFinder</i> , уровень стабильности, $p_{ig}$
<i>GAPDH</i>	0.045	3.90	0.513
<i>HPRT1</i>	0.112	3.90	0.490
<i>RPII</i>		4.30	0.666
<i>YWHAZ</i>		4.09	0.543
<i>UBC</i>		4.60	0.738
<i>18S</i>		4.90	0.868

**Таблица 3.** Результаты анализа корреляции с помощью *BestKeeper*

Ген	Коэффициент корреляции, $r$	Уровень значимости, $p$
<i>YWHAZ</i>	0.785	<0.001
<i>RPII</i>	0.557	<0.001
<i>HPRT1</i>	0.112	<0.001
<i>GAPDH</i>	0.045	<0.001

циента корреляции Пирсона ( $r$ ) и уровня значимости ( $p$ ) оценили межгенные взаимодействия всех возможных пар референсных генов. Из оставшихся четырех генов высокую корреляцию ( $0.506 < r < 0.807$ ) с индексом *BestKeeper* с высоким уровнем значимости ( $p < 0.02$ ) обнаружили уровни экспрессии *HPRT1* и *GAPDH* (табл. 2). Индекс *BestKeeper* используется для описания связи между каждым входящим в него референсным геном-кандидатом и индексом, основанном на коэффициенте корреляции Пирсона ( $r$ ), коэффициенте детерминации ( $r^2$ ), и уровнем значимости ( $p$ ).

Таким образом, результаты, полученные с помощью всех трех алгоритмов, оказались сходными. В тканях желудка больных ожирением наиболее стабильными референсными генами были *HPRT1* и *GAPDH*, а наименее стабильным – ген *18S* (табл. 3).

### ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Разница в уровнях экспрессии генов, измеряемая методом количественной ПЦР в реальном времени, представляет собой комбинацию истинных биологических различий между образцами и различий, обусловленных несовершенством экспериментального метода, например, связанных с разным количеством исходного материала и присутствием ингибиторов, влияющих на эффективность ферментов [28]. Для учета влияния этих небиологических факторов на результаты опыта данные, полученные с помощью количественной ПЦР, необходимо нормировать по внешнему или внутреннему контролю. В настоящее время наиболее предпочтительным считается нормирование с использованием референсных генов [28]. К недостаткам этого подхода относится потребность в контрольном гене, стабильно экспрессирующемся в каждом экспериментальном условии. Сам по себе такой подход превращает проблему в заколдованный круг, поскольку не существует никакого внешнего стандарта, позволяющего оценить стабильность экспрессии референсных генов [28].

Чтобы найти выход из этой ситуации разработаны три алгоритма – *geNorm*, *NormFinder* и *BestKeeper* [28], в каждом из которых используется свой метод нахождения идеального референсного гена или комбинации таких генов [28].

Первый из алгоритмов, применяемый в программе *geNorm* [21], количественно оценивает стабильность экспрессии референсных генов с использованием двух параметров: средней стабильности экспрессии ( $M$ ) и попарной вариации ( $V$ ). При низком значении  $M$  экспрессия более стабильная и пригодность этого гена как референсного возрастает. В этом алгоритме величина  $M$  определяется из попарных сравнений одного гена-кандидата со всеми остальными без учета уровня геновой экспрессии в каждом образце. Когда два референсных гена обнаруживают идентичное отношение экспрессии, делается вывод об относительной стабильности экспрессии. Таким образом, анализ с помощью программы *geNorm* не зависит от вариаций в количестве исходного материала между выборками и не требует нормального распределения данных. Результаты применения алгоритма *geNorm* в минимальной степени подвержены влиянию интенсивности экспрессии генов-кандидатов, что составляет еще одно преимущество данной программы. К тому же, поскольку этот подход основан на множественных попарных сравнениях, уменьшается необходимость в использовании выборки большого размера [34, 35]. Однако на результаты анализа с помощью *geNorm* влияет корегуляция генов-кандидатов; чтобы преодолеть это ограничение следует использовать референсные гены с разной физиологической функцией.

Анализ шести референсных генов с помощью алгоритма *geNorm* выявил *GAPDH* и *HPRT1* как наилучшую комбинацию референсных генов для неопухолевых образцов тканей желудка (рис. 1), в то время как *18S* и *UBC* описаны как наименее стабильные. С целью определения идеального количества референсных генов вычисляли попарную вариацию ( $V$ ) двух последовательных факторов нормирования ( $NF_n$  и  $NF_{n+1}$ ). Показано, что добавление большего числа референсных генов не ведет к значимому увеличению попарной вариации ( $V_3 = 1.05$ ), а использования двух референсных генов достаточно для достоверного нормирования экспрессии генов в данной ткани (рис. 2).

Алгоритм *NormFinder* работает с корегулируемыми референсными генами, преодолевая, таким образом, ограничения, присущие программе *geNorm*. Программа *NormFinder* использует подход, основанный на математической модели, ранжирующий гены по уровню внутри- и межгрупповой дисперсии. Действительно, *NormFinder* учитывает гетерогенность выборок, но подчерки-

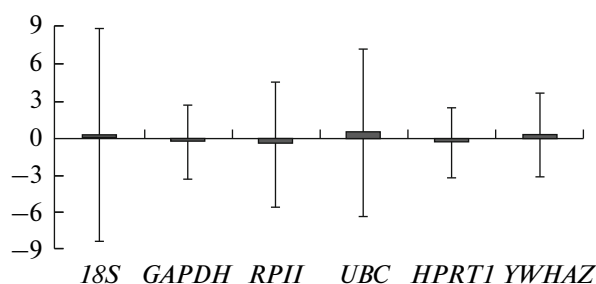


важает важность общей интенсивности экспрессии каждого гена, преувеличивая вклад генов со сходными уровнями общей экспрессии (в терминах Ct). Следовательно, надежность этого метода и размер выборки оказываются взаимосвязанными.

В случае алгоритма *NormFinder* идеальный референсный ген (гены) имеет наиболее близкую к нулю внутри- и межгрупповую дисперсию. Когда гены ранжируются только по уровню их стабильности, комбинация *GAPDH* и *HPRT1* оказывается наилучшей для использования в качестве внутреннего контроля (табл. 2). Интересно, что в нашей работе ген *18S* характеризовался наименьшей межгрупповой изменчивостью после *GAPDH*; *HPRT1* следовал за ним (рис. 3) и был более стабилен. Этот вывод принимает во внимание низкую внутригрупповую изменчивость *HPRT1*, в то время как внутригрупповая изменчивость *18S* была наивысшей (рис. 3). Таким образом, *18S* оценен как наименее стабильный и, следовательно, удален из дальнейшего рассмотрения. Заметим, что оба алгоритма выделяют ген *18S* как один из наименее стабильных. Этот результат не был неожиданным, поскольку ранее уже приводились аргументы против использования генов рРНК в качестве референсных, в том числе из-за затруднений при вычитании весьма высокого уровня экспрессии *18S* из базового уровня экспрессии [21].

Важно отметить, что одна и та же пара генов, *HPRT1* и *GAPDH*, признана наиболее стабильной как программой *geNorm*, так и *NormFinder*. Для независимого сравнения результатов, полученных с помощью *geNorm* и *NormFinder*, данные количественной ПЦР проанализировали с помощью еще одного алгоритма – *BestKeeper* [32]. Этот алгоритм определяет стабильность экспрессии референсных генов на основании как общей стабильности экспрессии, так и собственного индекса, вычисляемого программой *BestKeeper* (значения *r* и *p*). На основании низкого коэффициента вариации *CV* и высокого коэффициента корреляции (*r*) с индексом *BestKeeper*, *GAPDH* и *HPRT1* посчитали наиболее стабильными (табл. 2). Как и ранее, *18S* посчитали наименее стабильным и исключили его из дальнейшего рассмотрения.

В нашей работе результаты, полученные тремя алгоритмами, соответствовали друг другу. Недавно на смешанной выборке из раковых и прилегающих к опухоли нормальных тканей референсными признаны существенно различные наборы генов [29]. Так, с помощью *geNorm* пару генов *HPRT1* и *RPL29* оценили как наиболее стабильную, в то время как *18S* и *GAPDH* были наименее стабильными. С другой стороны, программа *NormFinder* самой стабильной посчитала комбинацию генов *B2M* и *RPL29*. Разница в стабильности экспрессии *GAPDH*, отмеченная нами и в работе [36], может определяться опухолеспецифичным



**Рис. 3.** Определение наиболее стабильного референсного гена с помощью программы *NormFinder*. Две рассмотренных группы представляли собой образцы тканей больных ожирением с гастритом/без гастрита ( $n = 22$ ). Прямоугольники – межгрупповая дисперсия, линии ошибок представляют среднюю внутригрупповую дисперсию. Идеальный референсный ген имеет наиболее близкую к нулю межгрупповую дисперсию и наименьшую из возможных линию ошибок.

увеличением экспрессии *GAPDH*. Экспрессия *GAPDH*, как показано ранее для других опухолевых тканей, может возрастать до таких значений, что этот ген предлагали использовать в качестве биомаркера определенных типов рака [37]. С другой стороны, подтверждена надежность *GAPDH* как референсного гена для неопухолевых клеток и тканей человека [38], что позволяет считать различия в результатах изучения двух выборок тканей желудка не артефактом, а отражением природы этих выборок.

Однако насколько обнаруженные различия в стабильности экспрессии генов определяются набором использованных генов или же особенностями опухолевых и неопухолевых образцов тканей желудка, с полной определенностью сказать невозможно. В представленной работе использованы неопухолевые ткани желудка, поэтому сделанные нами выводы о стабильности экспрессии контрольных генов относятся только к образцам этого типа. Важно, что три различных алгоритма признали наиболее стабильно экспрессирующимся один и тот же набор референсных генов. Это позволяет нам рекомендовать пару *HPRT1/GAPDH* как наилучшую комбинацию референсных генов для изучения неопухолевых заболеваний желудка методом количественной ПЦР в реальном времени. Не исключено, что в случае опухолевых и неопухолевых тканей желудка необходимо использовать разные наборы референсных генов.

Мы рекомендуем использовать пару генов *GAPDH* и *HPRT1* в качестве стабильных референсных генов, наиболее подходящих для изучения экспрессии генов в неопухолевых тканях желудка человека. Использование этих генов как референсной пары поможет в дальнейшем увеличить надежность количественной ПЦР в реальном времени в данной модельной системе.



## СПИСОК ЛИТЕРАТУРЫ

1. Yamada T., Oka Y., Katagiri H. 2008. Inter-organ metabolic communication involved in energy homeostasis: Potential therapeutic targets for obesity and metabolic syndrome. *Pharmacol. Therapeutics*. **117**(1), 188–198.
2. Murphy K.G., Bloom S.R. 2006. Gut hormones and the regulation of energy homeostasis. *Nature*. **444**(7121), 854–859.
3. Badman M.K., Flier J.S. 2005. The gut and energy balance: visceral allies in the Obesity Wars. *Science*. **307**(5717), 1909–1914.
4. Gale S.M., Castracane V.D., Mantzoros C.S. 2004. Energy homeostasis, obesity and eating disorders: recent advances in endocrinology. *J. Nutrition*. **134**(2), 295–298.
5. Ravussin E., Bogardus C. 2000. Energy balance and weight regulation: genetics versus environment. *Br. J. Nutrition*. **83**, S17–S20.
6. Bado A., Levasseur S., Attoub S., et al. 1998. The stomach is a source of leptin. *Nature*. **394**(6695), 790–793.
7. Buchwald H., Oien D. 2009. Metabolic/bariatric surgery Worldwide. 2008. *Obesity Surgery*. **19**(12), 1605–1611.
8. Gill R.S., Karmali S., Sharma A.M. 2011. Treating type 2 diabetes mellitus with sleeve gastrectomy in obese patients. *Obesity (Silver Spring)*. **19**(4), 701–702.
9. Nandagopal R., Brown R.J., Rother K.I. 2011. Resolution of type 2 diabetes following bariatric surgery: implications for adults and adolescents. *Diabetes Technol. Therapeutics*. **12**(8), 671–677.
10. Hofsø D., Nordstrand N., Johnson L.K., et al. 2010. Obesity-related cardiovascular risk factors after weight loss: a clinical trial comparing gastric bypass surgery and intensive lifestyle intervention. *Eur. J. Endocrinol.* **163**(5), 735–745.
11. Ford A.C., Axon A.T.R. 2010. Epidemiology of helicobacter pylori infection and public health implications. *Helicobacter*. **15**, 1–6.
12. Chao C., Hellmich M.R. 2010. Gastrin, inflammation, and carcinogenesis. *Curr. Opin. Endocrinol. Diabetes Obesity*. **17**(1), 33–39.
13. Yaghoobi M., Farrokhyar F., Yuan Y., Hunt R.H. 2010. Is there an increased risk of GERD after *Helicobacter pylori* eradication? a meta-analysis. *Am. J. Gastroenterol.* **105**(5), 1007–1013.
14. Hernando-Harder A.C., Booken N., Goerdts S., et al. 2009. *Helicobacter pylori* infection and dermatologic diseases. *Eur. J. Dermatol.* **19**(5), 431–444.
15. Tucker T.P., Gray B.M., Eaton K.A., Merchant J.L. 2010. *Helicobacter pylori* induction of the gastrin promoter through GC-rich DNA elements. *Helicobacter*. **15**(5), 438–448.
16. Kimang'a A., Revathi G., Kariuki S., et al. 2010. IL-17A and IL-17F Gene expression is strongly induced in the mucosa of *H. pylori* – infected subjects from Kenya and Germany. *Scandinavian J. Immunol.* **72**(6), 522–528.
17. Pelz C., Steininger S., Weiss C., et al. 2011. A novel inhibitory domain of *Helicobacter pylori* protein CagA reduces CagA effects on host cell biology. *J. Biol. Chem.* **286**(11), 8999–9008.
18. Logan J., Edwards K., Saunders N. 2009. *Real-time PCR: current technology and applications*. Horizon Scientific Press.
19. de Jonge H.J.M., Fehrmann R.S.N., de Bont E.S.J.M., et al. 2007. Evidence based selection of housekeeping genes. *PLoS ONE*. **2**(9), e898.
20. Bustin S. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* **25**(2), 169–193.
21. Vandesompele J., de Preter K., Pattyn F., et al. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**(7), research0034.1–research0034.11.
22. Huggett J., Dheda K., Bustin S., Zumla A. 2005. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.* **6**(4), 279–284.
23. Catalán V., Gómez-Ambrosi J., Rotellar F., et al. 2007. Validation of endogenous control genes in human adipose tissue: relevance to obesity and obesity-associated type 2 diabetes mellitus. *Horm. Metab. Res.* **39**(07), 495–500.
24. Thellin O., Zorzi W., Lakaye B., et al. 1999. Housekeeping genes as internal standards: use and limits. *J. Biotechnol.* **75**(2–3), 291–295.
25. Suzuki T., Higgins P.J., Crawford D.R. 2000. Control selection for RNA quantitation. *Biotechniques*. **29**(2), 332–337.
26. Dheda K., Huggett J., Chang J., et al. 2005. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal. Biochem.* **344**(1), 141–143.
27. Schmittgen T.D., Zakrajsek B.A. 2000. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *J. Biochem. Biophys. Methods*. **46**(1–2), 69–81.
28. Vandesompele J., Kubista M., Pfaffl M.W. 2009. In: *Real-time PCR: current technology and applications*. Eds Logan J., Edwards K., Saunders N. Horizon Sci. Press.
29. Rho H., Lee B., Choi E., et al. 2010. Identification of valid reference genes for gene expression studies of human stomach cancer by reverse transcription-qPCR. *BMC Cancer*. **10**(1), 240.
30. Ushijima T. 2007. Epigenetic field for cancerization. *J. Biochem. Mol. Biol.* **40**(2), 142–150.
31. Andersen C.L., Jensen J.L., Ørntoft T.F. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**(15), 5245–5250.
32. Pfaffl M.W., Tichopad A., Prgomet C., Neuvians T.P. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**(6), 509–515.

33. Baranova A., Hammarsund M., Ivanov D., et al. 2003. Distinct organization of the candidate tumor suppressor gene *RFP2* in human and mouse: multiple mRNA isoforms in both species- and human-specific antisense transcript RFP2OS. *Gene*. **321**, 103–112.
34. Rasmussen R. 2001. In: *Rapid cycle real-time PCR: methods and applications*. Eds Meuer S., Wittwer C., Nakagawara K. Heidelberg: Springer-Verlag Press, 21–34.
35. Mehta R., Birerdinc A., Hossain N., et al. 2010. Validation of endogenous reference genes for qRT-PCR analysis of human visceral adipose samples. *BMC Mol. Biol.* **11**(1), 39.
36. Cicinnati V., Shen Q., Sotiropoulos G., et al. 2008. Validation of putative reference genes for gene expression studies in human hepatocellular carcinoma using real-time quantitative RT-PCR. *BMC Cancer*. **8**(1), 350.
37. Tokunaga K., Nakamura Y., Sakata K., et al. Enhanced expression of a glyceraldehyde-3-phosphate dehydrogenase gene in human lung cancers. *Cancer Res.* **47**(21), 5616–5619.
38. Zainuddin A., Chua K., Rahim N., Makpol S. 2010. Effect of experimental treatment on GAPDH mRNA expression as a housekeeping gene in human diploid fibroblasts. *BMC Mol. Biol.* **11**(1), 59.