

УДК 577.113.5

*Г.С. Краснов, А.А. Дмитриев, А.В. Кудрявцева, Е.А. Анедченко, Н.Ю. Опарина,
В.Н. Сенченко*

Институт молекулярной биологии им. В.А. Энгельгардта РАН

Новые программы, используемые при количественной оценке копийности и уровня транскрипции генов

В настоящее время метод полимеразной цепной реакции в реальном времени (ПЦР-РВ) является основным подходом для количественной оценки уровня транскрипции генов. Однако есть ряд проблем, которые необходимо учитывать при использовании ПЦР-РВ: трудность подбора высококачественных праймеров и зондов, сложность анализа транскрипции генов со множественными изоформами, влияние ряда факторов, искажающих количественные данные, на эффективность амплификации. Нами были разработаны программы как для этапа подбора качественных наборов праймеров и зондов, так и для этапа обработки результатов ПЦР-РВ с учётом эффективности амплификации. Проведено сравнение разработанных программ с существующими аналогами.

Ключевые слова: количественная ПЦР-РВ, эффективность реакции, подбор праймеров, экспрессия генов, уровень мРНК.

Метод полимеразной цепной реакции в реальном времени (ПЦР-РВ) широко применяется во всём мире как для фундаментальных, так и для прикладных исследований. Качество полученных данных зависит от ряда факторов, таких, как: чистота матрицы ДНК, качество реактивов, выбор участка для проведения амплификации, праймеров и зондов. Кроме того, эффективность ПЦР не постоянна и зависит от многих факторов, часто неизвестных (рис. 1). В связи с этим представляется наиболее важным для получения достоверных данных подбирать качественные наборы праймеров и зондов, а также учитывать эффективность амплификации при анализе данных ПЦР-РВ. Для решения этих задач существует ряд программ, но они не полностью отвечают запросам экспериментатора. Так, при подборе праймеров затруднительно проводить предварительную селекцию наборов праймеров и зондов для того, чтобы достоверно амплифицировать фрагмент определённой изоформы мРНК изучаемого гена. При поиске таких наборов часто приходится варьировать желаемые параметры (долю нуклеотидов G и C в ПЦР-продукте, его длину, размер праймеров и др.). «Идеальные» параметры не всегда пригодны для конкретного гена, например, со сложной для проведения ПЦР нуклеотидной последовательностью. Насколько повлияют на результат ПЦР сделанные допущения в параметрах? Эти и другие вопросы поднимались ранее в литературе [1–3]. В частности, вопрос о получении специфичных праймеров к экзон-экзонным границам решается в таких программах, как ExonPrimer. Такой подход позволяет получать наборы праймеров, специфичные к мРНК изучаемого гена, но не к его геномной копии. Однако остаётся открытым вопрос, насколько подобные наборы праймеров будут специфичны к определённой из множествен-

ных изоформ. Кроме того, в присутствии геномной ДНК давать ложный сигнал может не только активная геномная копия данного гена, но и его процессированные псевдогены, содержащие те же экзон-экзонные границы, что и мРНК. Для решения этих задач нами разработана программа ПЦР-Дизайн (Primer Designer), оптимально приспособленная для анализа уровня транскрипции эукариотических генов. Программа не требует установки, работает с различными форматами нуклеотидных последовательностей, позволяет разными способами задавать или искать экзон-экзонные границы в исходной мРНК, позволяет избегать участков с известными SNP и мутациями при подборе праймеров и зондов. Выбранные праймеры можно проверять на специфичность против других изоформ мРНК гена, псевдогенов и пр. Кроме того, в Primer Designer внесён дополнительный модуль анализа качества уже известных праймеров и проб.

Другая сложность получения качественных результатов ПЦР-РВ заключается во влиянии на них эффективности амплификации [4, 5]. Эффективность зависит от многих факторов, прежде всего — от нуклеотидной последовательности ампликона. Кроме того, она может изменяться от опыта к опыту при незначительном изменении условий реакции. Расчёт данных с учётом эффективности в каждом опыте достаточно трудоёмок. В настоящее время данные ПЦР-РВ в большинстве выпускаемых статей представляются без учёта эффективности и статистической обработки. Существуют программы, как правило, рассчитывающие эффективность реакций по кинетическим кривым, и программы, осуществляющие обработку данных с учётом уже известной заранее эффективности. Нами разработана оригинальная программа, позволяющая автоматически опреде-

лять эффективность каждой кривой, учитывать её при расчёте данных относительным количественным методом (Relative Quantification), а также производить статистическую обработку полученных результатов. Расчёт эффективности можно произвести тремя различными способами, в том числе с разработанной нами моделью, включающей пять параметров для более точного опре-

деления эффективности. Статистическая обработка производится с помощью парного *t*-теста Стьюдента или с помощью перемешивающего теста (randomization test[®]). Программа написана в виде книги Microsoft[®] Excel[®] с набором макросов, совместима с экспортируемыми файлами показаний прибора программного обеспечения (ПО) AppliedBiosystems[®].

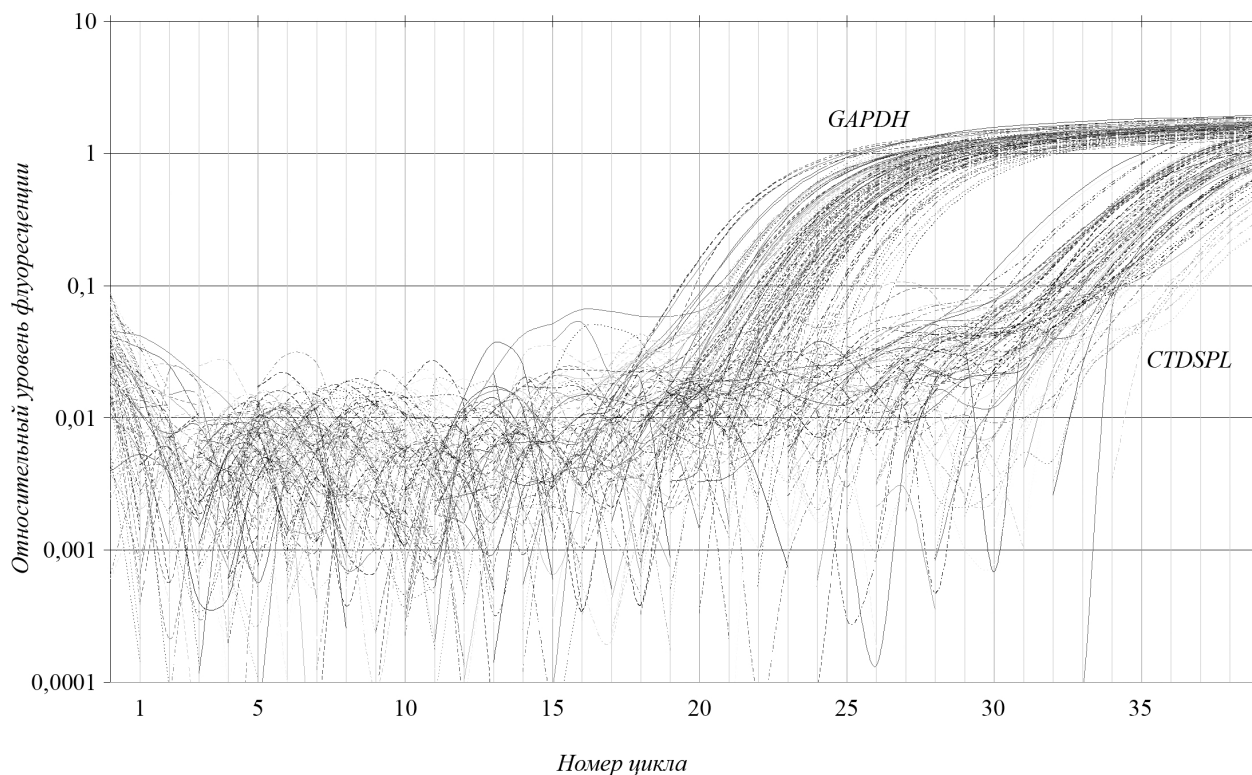


Рис. 1. Наклон групп кривых, соответствующих генам GAPDH и CTDSPL, отражает разные эффективности амплификации

Primer Designer — программа поиска оптимальных наборов праймеров и проб для количественного анализа уровня транскрипции генов. Программа Primer Designer представляет собой самостоятельное приложение, не требующее установки. При запуске программы пользователю предлагается выбор — поиск новых наборов праймеров и проб или анализ известных. На каждом этапе работы все данные и параметры можно сохранить в виде проекта и использовать его в дальнейшем. Основные реализованные в программе возможности заключаются в следующем.

1. Задание разными способами экзон-экзонных границ в мРНК. При этом возможно сравнение мРНК и геномной ДНК изучаемого гена с использованием программы Spidey [6]. В случае «сомнительности» некоторых границ (например, отсутствие канонических донорных и акцепторных сайтов) возможен поиск наборов праймеров и проб с учётом только достоверных границ. Программа позволяет формировать специфичные к мРНК наборы как праймеров и Taqman зондов, так и пары праймеров для ПЦР с использованием SYBR Green I или обычного ОТ-ПЦР.

2. Опция «ручки вариаций», позволяющая отметить на анализируемой нуклеотидной последовательности варибельные участки и не использовать их при поиске праймеров и проб.

3. При изменении пользователем параметров поиска праймеров и проб в интерактивном режиме цветом (зёленый — оптимум, красный — запрещённое значение) обозначается допустимость значения каждого параметра для качественного проведения ПЦР-РВ.

4. В отличие от аналогов, Primer Designer поддерживает различные модели расчёта термодинамических параметров олигонуклеотидов [7, 8, 9].

5. Отобранные праймеры группируются в кластеры согласно их расположению на нуклеотидной последовательности.

6. Функция анализа специфичности наборов праймеров и зондов. Эта возможность, отсутствующая у аналогов, позволяет отобрать только те наборы, которые специфичны к искомой мРНК, но не к набору нуклеотидных последовательностей, предложенных пользователем (другие изоформы, псевдогены, представители того же генного семейства).

Программа расчёта эффективности и математической обработки данных ПЦР-РВ «Анализ транскрипции генов» (АТГ, ATG, Analysis of transcription of genes) представляет собой книгу Microsoft Office Excel, включающую набор макросов. Исходными данными являются файлы, данных экспортируемые программным обеспечением Applied Biosystems.

На первом этапе вычисляется эффективность реакции амплификации и её пороговый цикл, затем пользователю предоставляется возможность обработать данные относительным количественным методом, предварительно оценив вариабельность используемых контрольных генов в группе образцов сравнения и исследуемых образцов, а также вариабельность исследуемого гена относительно контрольного в норме. Затем можно произвести статистический анализ полученных результатов. На этапе проверки пригодности контрольного гена образцы предварительно должны быть выровнены по количеству вносимой в реакцию полинуклеотидной матрицы.

Программа позволяет вычислить эффективность реакций тремя различными способами — по экспоненциальному участку (автоматически опре-

деляемые границы которого пользователь при желании может изменить), а также аппроксимацией кинетической кривой двумя моделями — моделью «сигма», определяемой уравнением

$$f = \frac{A}{1 + e^{-\frac{x-x_0}{B}}}, \quad (1)$$

и моделью, определяемой уравнением

$$f^{1+\frac{P}{C}} + f \cdot e^{\frac{B \cdot P \cdot x + W \cdot P}{C}} - P \cdot e^{\frac{B \cdot P \cdot x + W \cdot P}{C}} = 0. \quad (2)$$

Здесь f — уровень флуоресценции на цикле x , а A, B, C, W, P, x_0 — коэффициенты, определяемые из аппроксимации кинетической кривой уравнением. Эти уравнения аппроксимируют кривые, из которых вычтен уровень флуоресценции на «нулевой линии», которая определяется по методу наименьших квадратов. Модель (2) используется впервые. Способы аппроксимации кинетической кривой моделями (1) и (2) предполагают использование всей кривой, а не только её экспоненциального участка, поэтому должны быть выполнены дополнительные требования, например, кривые в фазе «плато» должны быть параллельны оси абсцисс (табл. 1, рис. 1).

Т а б л и ц а 1

Расчёт эффективности ПЦР-РВ для генов GAPDH и CTDSPL тремя различными способами

Ген	Эффективность ПЦР-РВ		
	по экспоненциальному участку	аппроксимация уравнением	
		сигма (1)	разработанная модель (2)
GAPDH	99 ± 10	103 ± 9	97 ± 7
CTDSPL	79 ± 12	87 ± 15	83 ± 11

После определения эффективности для каждой кривой программа позволяет исключить некачественные кривые, а затем предлагает выбрать способ усреднения. С целью увеличения точности эффективность может быть усреднена в рамках одного гена, а также в рамках одного повтора — если измерения проводятся на той же матрице в одинаковых условиях. Стоит отметить, что при вычислении эффективности с использованием обеих моделей значение порогового цикла может быть определено непосредственно из уравнения, аппроксимирующего кинетическую кривую.

Перед проведением дальнейших расчётов в значения C_t вносится поправка, преобразующая их к «эффективным» пороговым циклам C_t^{eff} , которые реализовались бы при эффективности, равной 100%.

Дополнительные возможности. Программа позволяет оценить вариабельность уровня транскрипции контрольных генов, для чего определяются C_t^{eff} для всех образцов.

Далее можно провести статистическую обработку полученных данных. Обработку можно проводить как с помощью парного t -теста Стьюдена

та, так и с помощью перемешивающего теста (randomization test[®]).

Проблема подбора качественных праймеров и зондов для ПЦР-РВ и её решение с помощью различных современных программ. Наиболее качественные современные программы для поиска праймеров для ПЦР-РВ реализованы в виде веб-приложений: PROBEmer, Integrated DNA Technologies SciTools: PrimerQuest и другие. Задача специфической амплификации мРНК или кДНК в присутствии геномной ДНК, решаемая методом поиска праймеров на экзон-экзонных границах, реализуется в таких программах, как EasyExonPrimer, ExPrimer, PerlPrimer [10]. Однако все эти программы не предусматривают контроль специфичности полученных наборов по сравнению с другими изоформами гена и пр., а предлагают проверять каждый олигонуклеотид с помощью BLAST против баз данных нуклеотидных последовательностей. Реализованный в Primer Designer контроль специфичности хоть и не отменяет необходимости проверять праймеры, используя BLAST, однако позволяет резко сократить количество потенциально подходящих олигонуклеотидов. Кроме того, группировка наборов

праймеров и проб по кластерам полезна при подборе двух и более участков для определённой мРНК, в то время как аналогичные программы перечисляют все варианты праймеров как независимые, что делает трудоемким отбор двух независимых ампликонов. Так же стоит отметить как нововведения вышеперечисленные «ручку вариаций» и интерактивный режим определения допустимости изменений параметров. Таким образом, Primer Designer — это комплексное средство, оптимально отвечающее требованиям экспериментатора и удачно сочетающее классические методы и новые опции, отсутствующие в программах-аналогах.

С помощью Primer Designer были получены и экспериментально проверены праймеры к генам RASSF1 (изоформы A, C), HYAL1, HYAL2, VILL, ITGA9, CTDSPL, TIMP3, DAPK1, VHL, SEMA3B, SEMA3F, APRG1, NPRL2, FTL, FTH и др.

Обработка данных ПЦР-РВ: сравнение программ. На сегодняшний день существует большое количество программ, предназначенных для обработки данных ПЦР-РВ (qBase, DART-PCR, BestKeeper, GenEx, GeNorm, qCalculator, qGene, LinReg PCR). Однако большинство из них разделяется на программы, рассчитывающие эффективность, и программы, непосредственно обрабатывающие результаты. Разработанная нами программа представляет комплексное решение и включает:

- 1) расчёт эффективности тремя различными способами;
- 2) математическую обработку данных ПЦР-РВ относительным количественным методом, предусматривающим двойное сравнение данных для парных образцов: исследуемый ген — контрольный ген, исследуемые образцы — образцы сравнения;
- 3) статистическую оценку измеряемых изменений Ct , то есть позволяет оценить вариабельность контрольных генов в обеих группах образцов;
- 4) пригодность выбранных образцов сравнения;
- 5) использование нескольких образцов сравнения;
- 6) использование нескольких контрольных генов;
- 7) статистическую обработку полученных результатов.

Среди перечисленных программ следует отметить qBase — комплексное программное решение, помогающее на всех этапах работы с ПЦР-РВ, от планирования опыта до обработки данных и графического представления результатов, однако не позволяющее рассчитывать эффективность для каждой реакции. Однако эта программа не ориентирована на парное сравнение образцов. Программа также включает ряд статистических тестов. DART-PCR (Data Analysis for Real-Time PCR)

представляет собой комплексное решение, позволяющее как обрабатывать данные относительным количественным методом, так и оценивать эффективность методом построения графика амплификации, исключая выпадающие кривые с использованием алгоритма ANOVA. REST — классическая программа для обработки данных ПЦР-РВ относительным количественным методом [11]. Используемая в программе модель объединяет в одном расчёте нормирование с количественной оценкой экспрессии гена. Оценка вариабельности уровня транскрипции контрольных генов реализована в других программах, таких, как BestKeeper и GeNorm.

Разработанные программы представляют собой оригинальные авторские продукты, отличающиеся от аналогов рядом инноваций. Обе программы зарегистрированы в Роспатенте: ПЦР-Дизайн (Primer Designer) — свидетельство №2008612865, 2008 г., АТГ (ATG) — свидетельство №2008612585, 2008 г., и доступны для некоммерческого использования при обращении к авторам (gskrasnov@mail.ru, versen@eimb.ru, oparina@gmail.com).

Литература

1. *Ellinghaus P., Badehorn D., Blumer R., Becker K., Seedorf U.* Increased efficiency of arbitrarily primed PCR by prolonged ramp times // *Biotechniques*. — 1999. — V. 26, N. 4. — P. 626–630.
2. *Sandhu K.S., Acharya K.K.* ExPrimer: to design primers from exon-exon junctions // *Bioinformatics*. — 2005. — V. 21, N. 9. — P. 2091–2092.
3. *Shulzhenko N., Smirnova A.S., Morgun A., Gerbase-DeLima M.* Specificity of alternative splice form detection using RT-PCR with a primer spanning the exon junction // *Biotechniques*. — 2003. — V. 34, N. 6. — P. 1244–1249.
4. *Marino J.H., Cook P., Miller K.S.* Accurate and statistically verified quantification of relative mRNA abundances using SYBR Green I and real-time RT-PCR // *J. Immunol. Methods*. — 2003. — V. 283, N. 1–2. — P. 291–306.
5. *Yuan J.S., Reed A., Chen F., Stewart C.N. Jr.* Statistical analysis of real-time PCR data // *BMC Bioinformatics*. — 2006. — V. 22, N. 7. — P. 85.
6. *Wheeler S.J., Church D.M., Ostell J.M.* Spidey: a tool for mRNA-to-genomic alignments // *Genome Res*. — 2001. — V. 11, N. 11. — P. 1952–1957.
7. *Breslauer K.J., Frank R., Blöcker H., Marky L.A.* Predicting DNA duplex stability from the base sequence // *Proc. Natl. Acad. Sci.* — 1986. — V. 83. — P. 3746–3750.
8. *Sugimoto N., Nakano S., Yoneyama M., Honda K.* Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes // *Nucleic Acids Res.* — 1996. — V. 24, N. 22. — P. 4501–4505.

9. *SantaLucia J. Jr, Allawi H.T., Seneviratne P.A.* Improved nearest-neighbor parameters for predicting DNA duplex stability // *Biochemistry*. — 1996. — V. 35, N. 11. — P. 3555–3562.

10. *Marshall O.J.* PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR // *Bioinformatics*. — 2004. — V. 20, N. 15. — P. 2471–2472.

11. *Pfaffl M.W., Horgan G.W., Dempfle L.* Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR // *Nucleic Acids Res.* — 2002. — V. 30, N. 9. — P. 36.

Поступила в редакцию 10.02.2009.