

МОДЕЛЬ РЕГУЛЯЦИИ МОНОАЛЛЕЛЬНОЙ ЭКСПРЕССИИ ГЕНОВ

© 2017 г. А.А. Миронов* **

*Факультет биоинженерии и биоинформатики Московского государственного университета имени М.В. Ломоносова, 119234, Москва, Ленинские горы, 1/73;

**Институт проблем передачи информации, 127051, Москва, Большой Каретный пер., 19

E-mail: mironov@bioinf.fbb.msu.ru

Поступила в редакцию 26.10.16 г.

Большинство высших эукариот являются диплоидами. Большинство генов экспрессируются с обеих копий хромосом, но есть целый ряд важных исключений, в частности X-хромосома у млекопитающих, гены, находящиеся под генным импринтингом, гены иммунной системы. Для некоторых случаев показана роль сенс-антисенс взаимодействие транскриптов, обеспечивающих работу только одной копии гена. В настоящей работе предложена минимальная модель, описывающая строго моноаллельную экспрессию и проведен ее анализ.

Ключевые слова: моноаллельная экспрессия, динамическая модель, генный импринтинг.

Большинство высших эукариот имеют двойной набор хромосом и являются диплоидами. Для большинства генов идет экспрессия с двух копий хромосом примерно с одинаковой интенсивностью. Однако есть ряд случаев, когда идет экспрессия только с одной из пары гомологичных хромосом. Такое явление называется моноаллельной экспрессией. Наиболее известным примером здесь является полная инактивация одной из копий X-хромосомы у самок млекопитающих. За инактивацию одной из копий X-хромосомы отвечает система Xist-Tsix [1]. Эта система содержит некоторую область на X-хромосоме, с которой синтезируется РНК Xist и ее антисмысловой партнер Tsix. Их взаимодействие отвечает за выбор одной из хромосом, которую следует инактивировать [2]. При этом выбор происходит на ранней эмбриональной стадии [3]. Другим важным примером моноаллельной экспрессии является генный импринтинг. При этом гены экспрессируются исключительно либо с отцовской либо с материнской аутосомы. У человека под генным импринтингом находятся примерно 100 генов [4]. В этом случае часто контроль экспрессии также определяется взаимодействием сенс-антисенс-РНК [5]. Строгая моноаллельная экспрессия также характерна для генов, связанных с иммунной системой, – главные комплексы гистосовместимости [6] и гены иммуноглобулинов [7], а также гены рецепторов запаха [8]. Однако есть целый ряд генов, у которых наблюдается нестрогая моноаллельность – соотношение уровня экспрессии составляет порядка 1:3 [9]. На самом деле наблюдается полный спектр

смещений экспрессии от 1:1 до 0:1 [10]. Важной особенностью всех случаев моноаллельной экспрессии является то, что при клеточном делении дочерние клетки сохраняют свой выбор [11]. Это связывают с эпигенетическим наследованием [12].

В настоящей работе построена и исследована минимальная модель, определяющая строго моноаллельную экспрессию. В основе модели лежит представление о цис- и транс-взаимодействующих элементах. Получены ограничения на параметры и проверена устойчивость к шуму.

ДИНАМИЧЕСКАЯ МОДЕЛЬ

Пусть у нас есть некий ген, представленный на двух гомологичных хромосомах. Будем считать, что активность аллелей – что бы это ни значило – равна x и y . Под активностью аллеля можно понимать уровень транскрипции mRNA или уровень продуктов процессинга транскриптов, например количество вырезанных интронов, либо уровень транскрипции «промоторной» РНК. В общем случае можно написать систему дифференциальных уравнений:

$$\begin{cases} x' = f(x, y), \\ y' = g(x, y). \end{cases} \quad (1)$$

Заметим, что x , y – это два аллеля (быть может, идентичных) на разных гомологичных хромосомах. Поэтому пока положим $g(x, y) = f(y, x)$. Будем считать для простоты, что функции – квадратичные:

$$f(x, y) = a + b_x x + b_y y + c_x x^2 + c_y y^2 + 2dx y.$$

Строгая моноаллельная экспрессия приводит к требованию наличия двух стационарных точек. Не ограничивая общности, можно потребовать, что стационарным точкам соответствуют уровни экспрессии $x = 0, y = 1$ или $x = 1, y = 0$. Это значит, что в стационарном случае мы хотим иметь либо исключительно экспрессию одного аллеля, либо другого. Именно требование наличия двух стационарных точек исключает линейные правые части. Отсюда получаем соотношения:

$$\begin{cases} x'(0, 1) = f(0, 1) = a + b_y + c_y, \\ y'(0, 1) = f(1, 0) = a + b_x + c_x. \end{cases}$$

В силу симметрии по x, y соотношения для точки $(1, 0)$ будут такими же. Мы хотим, чтобы стационарные точки были устойчивыми и чтобы вблизи них не было колебаний, выходящих в отрицательную область. Это значит, что стационарные точки должны быть устойчивыми узлами. Для выполнения этих условий есть некоторые условия. Вычислим соответствующие собственные значения в точке $(0, 1)$:

$$\begin{vmatrix} \frac{\partial f(x, y)}{\partial x} - \lambda & \frac{\partial f(x, y)}{\partial y} \\ \frac{\partial f(y, x)}{\partial x} & \frac{\partial f(x, y)}{\partial y} - \lambda \end{vmatrix}_{(0,1)} = 0.$$

Вычисляя соответствующие производные, получаем характеристическое уравнение:

$$\lambda^2 - 2\lambda(b_x + c_x + d) - (b_y + 2d)(b_y + 2c_y) + (b_x + 2c_x)(b_x + 2d) = 0.$$

Корни уравнения:

$$\begin{aligned} \lambda_{1,2} &= (b_x + c_x + d) \pm \sqrt{\Delta}, \\ \Delta &= (b_x + c_x + d)^2 + (b_y + 2d)(b_y + 2c_y) - \\ &\quad - (b_x + 2c_x)(b_x + 2d). \end{aligned}$$

Мы хотим, чтобы стационарная точка $(0, 1)$ была устойчивым узлом. Для этого необходимо, чтобы оба корня уравнения были действительными отрицательными. Получаем ограничения на коэффициенты:

$$\Delta = (b_x + c_x + d)^2 + (b_y + 2d)(b_y + 2c_y) - (b_x + 2c_x)(b_x + 2d) \geq 0, \quad (2)$$

$$(b_y + 2d)(b_y + 2c_y) - (b_x + 2c_x)(b_x + 2d) \leq 0, \quad (3)$$

$$b_x + c_x + d \leq 0. \quad (4)$$

Неравенство (2) необходимо выполнить, чтобы корни были действительными, неравенство (3) – чтобы корни были одного знака, неравенство (4) – чтобы корни были отрицательными.

Состояние аллеля чувствует цис-продукты и транс-продукты. Цис-продукты – то, что как бы привязано к той же хромосоме, откуда идет транскрипция, например, мРНК, висающая на полимеразе – цис-продукт. Быть может, что-то привязывает РНК-продукт к хромосоме, с которой идет транскрипция – тогда это тоже цис-продукт. Как только продукт покидает хромосому – он становится транс-продуктом, поскольку может влиять на экспрессию как своего родного аллеля, так и на экспрессию гомологичного аллеля на другой хромосоме. Все, что теряет связь со своей хромосомой – это транс-продукт. Поэтому для аллеля x цис-продуктом будет экспрессия x , в то время как транс-продуктом будет смесь продуктов x и y :

$$z = \alpha x + \beta y, \quad y = \frac{z - \alpha x}{\beta}.$$

Поскольку диффузия происходит не мгновенно, концентрация вклада цис-продукта в транс-продукт будет выше: $\alpha \geq \beta$, впрочем, это не важно. Преобразуем правую часть уравнений к новым переменным x, z :

$$a + B_x x + B_z z + C_x x^2 + C_z z^2 + 2Dxz,$$

где

$$\begin{aligned} B_x &= \frac{\beta b_x - \alpha b_y}{\beta}, \quad B_z = \frac{b_y}{\beta}, \quad C_x = \frac{\beta^2 c_x + \alpha^2 c_y - 2\alpha\beta d}{\beta^2}, \\ C_z &= \frac{c_y}{\beta^2}, \quad D = \frac{\beta d - \alpha c_y}{\beta^2}. \end{aligned}$$

ИНТЕРПРЕТАЦИЯ ПАРАМЕТРОВ

Весь процесс можно представить, например, как сенс-антисенс-транскрипцию. Цис-взаимодействие – это взаимодействие транскриптов, так или иначе связанных (физически) с самим аллелем, например, взаимодействие транскрипта, связанного с РНК-полимеразой. Транс-взаимодействие – это взаимодействие транскрипта с транскриптами, плавающими в нуклеоплазме. Будем считать, что x – актуальный аллель (тот, на который мы в настоящее время смотрим). Тогда параметры можно интерпретировать следующим образом:

a – фоновая активность. Должна быть неотрицательна;

B_x – линейная самоактивация или самоподавление. Линейность означает непосредственное влияние транскрипта на активность. Более детальная интерпретация будет дана ниже;

C_x – квадратичная самоактивация. Квадратичность означает, что на активность влияет продукт взаимодействия цис-продукта с самим собой. Более детальная интерпретация будет дана ниже;

B_z – линейная активация или подавление аллеля x плавающими в растворе продуктами аллелей x, y . Это взаимодействие не отличает рассматриваемый ген от любых других – здесь нет адресации именно к рассматриваемому гену. Поэтому этот параметр можно положить $B_z = 0$;

C_z – квадратичная активация или подавление аллеля x взаимодействующими продуктами, плавающими в растворе. По-видимому, особого смысла в таком взаимодействии нет (см. рассуждение про предыдущий параметр), поэтому можно положить $C_z = 0$;

D – взаимодействие транс-продуктов (плавающих продуктов x, y) с цис-продуктом x . Это очень важное взаимодействие, которое измеряет «дозу активного гена», и если доза гена большая, то идет его подавление;

α, β – относительные доли продуктов аллелей x, y в том, что плавает в растворе вблизи гена x . Без потери общности можно положить $\alpha = \beta = 1$.

Таким образом, получаем набор ограничений. Из положения стационарных точек и из $C_z = 0$ и $B_z = 0$ получаем:

$$B_z = 0 \Rightarrow b_y = 0; \quad a + b_y + c_y = 0 \Rightarrow a = 0$$

$$C_z = 0 \Rightarrow c_y = 0; \quad a + b_x + c_x = 0 \Rightarrow c_x = -b_x.$$

С учетом этих равенств ограничения (2)–(4) примут вид:

$d^2 + b_x(b_x + 2d) \geq 0$ – чтобы корни были действительными,

$b_x(b_x + 2d) \leq 0$ – чтобы корни были одного знака,

$d \leq 0$ – чтобы корни были отрицательными.

Первое неравенство выполнено всегда. Действительно,

$$d^2 + b_x(b_x + 2d) = d^2 + b_x^2 + 2db_x = (d + b_x)^2 \geq 0.$$

Из двух последних неравенств следует, что $b_x \geq 0, b_x + 2d \leq 0$. Окончательно имеем:

$$d \leq 0, \tag{5}$$

$$0 \leq b_x \leq -2d. \tag{6}$$

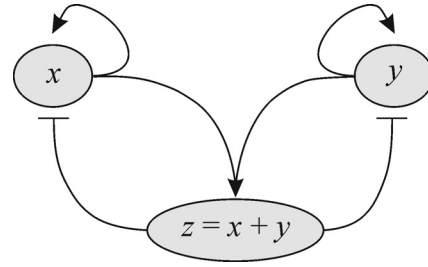


Рис. 1. Схема взаимодействий.

На интерпретации коэффициентов B_x и C_x можно остановиться подробнее. Забудем пока про взаимодействие и второй аллель. Тогда (при условии $a = 0$) уравнение активности аллеля имеет вид:

$$x' = B_x x + C_x x^2.$$

Поскольку $B_x \geq 0$ и $C_x \leq 0$, то простым преобразованием переменной $x = -B_x/C_x \tilde{x}$ приходим к логистическому уравнению:

$$\tilde{x}' = B_x \tilde{x}(1 - \tilde{x}).$$

Смысл этого уравнения достаточно прост. Активация происходит по «сайтам активации». Чем больше активность, тем активнее происходит активация (первый множитель в правой части), но количество «сайтов активации» сокращается, о чем говорит второй множитель. Решение этого уравнения:

$$\tilde{x} = \frac{e^{B_x t}}{A + e^{B_x t}},$$

где A – произвольная константа интегрирования. В исходных переменных решение имеет вид:

$$x = -\frac{C_x e^{B_x t}}{A + B_x e^{B_x t}}, \quad x = \frac{C_x e^{B_x t}}{A + B_x e^{B_x t}}.$$

Еще раз отметим, что константа B_x положительна, а константа C_x – отрицательна.

Схема взаимодействия, которая соответствует модели, представлена на рис. 1. На рис. 2а представлено поле для некоторого допустимого набора параметров. Рис. 2б представляет решение системы уравнений для начальных значений (0,01; 0,011). Видно, что на старте сначала оба аллеля имеют заметную активность, а потом один из них уходит в 0, а другой выходит на значение 1, что и требуется.

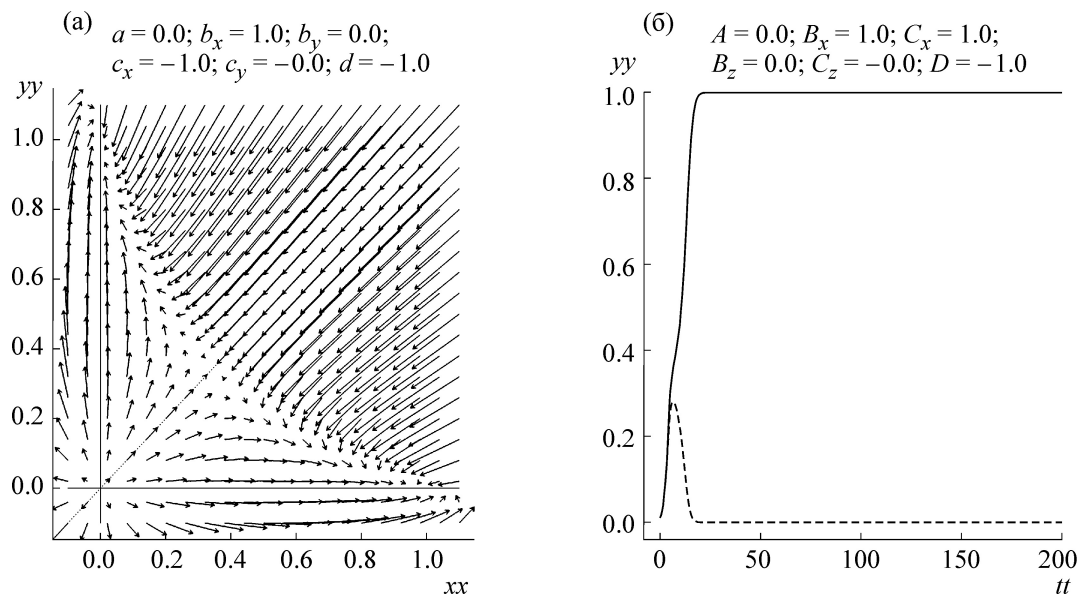


Рис. 2. (а) – Поле производных для некоторого набора параметров – по осям активности аллелей. Стрелки показывают направление изменений активности (вектор x', y'); (б) – пример поведения системы – решение системы уравнений.

УЧЕТ ШУМА ЭКСПРЕССИИ

На самом деле активность аллелей есть флуктуирующая случайная величина. Поэтому мы добавим в уравнения (1) шум. Ясно, что шум должен быть пуассоновским, отражающим флуктуации:

$$\begin{cases} x' = f(x_r, y_r), \\ x' = f(y_r, x_r), \\ x_r = \frac{\xi(\lambda x)}{\lambda}, \\ y_r = \frac{\xi(\lambda y)}{\lambda}. \end{cases} \quad (7)$$

где $\xi(a)$ – случайная величина, распределенная согласно распределению Пуассона с параметром a . Параметр λ отражает количество молекул при активности, равной единице. На рис. 3а показано решение уравнения (7) при тех же параметрах и $\lambda = 100$. Теперь посмотрим, как такая система будет реагировать на деление клеток. Пусть каждые 25 единиц времени происходит деление клеток, т.е. количество продуктов x, y уменьшается в два раза. Рис. 3б показывает поведение такой системы. Видно, что эта модель поддерживает специфичность аллелей при делении клеток. При этом достаточно, чтобы продукты аллелей были в среде.

ОБСУЖДЕНИЕ

Построена простейшая модель, объясняющая явление строгой моноаллельной экспрессии генов. Явление строгой моноаллельной экспрессии встречается в ряде специальных случаев – на второй копии X-хромосомы млекопитающих, при генном импринтинге, в генах иммунной системы и в генах рецепторов запаха. Для ряда случаев было показано экспериментально, что имеет место двунаправленная экспрессия некодирующих РНК с области контроля локуса. По-видимому, взаимодействие этих взаимно-комплементарных РНК отвечает за выбор хромосомы, с которой происходит экспрессия. В дальнейшем вступают эпигенетические механизмы поддержания активного или пассивного поведения соответствующих локусов. Предложенная модель устанавливает некоторые ограничения на систему контроля локуса. В частности, транс-взаимодействие должно ингибировать активность локуса, в то время как цис-взаимодействие должно активировать активность. При этом активация происходит по логистической модели «сайтов активации» и самоактивация не должна быть слишком сильной. Модель предсказывает существование белков, которые обеспечивают связь цис-продуктов с локусом и чувствуют наличие транс-продуктов. Однако существует довольно большое множество генов, которые находятся не под строгой моноаллельной экспрессией, а поддерживают некоторое соотношение уровней экспрессии аллелей. По-

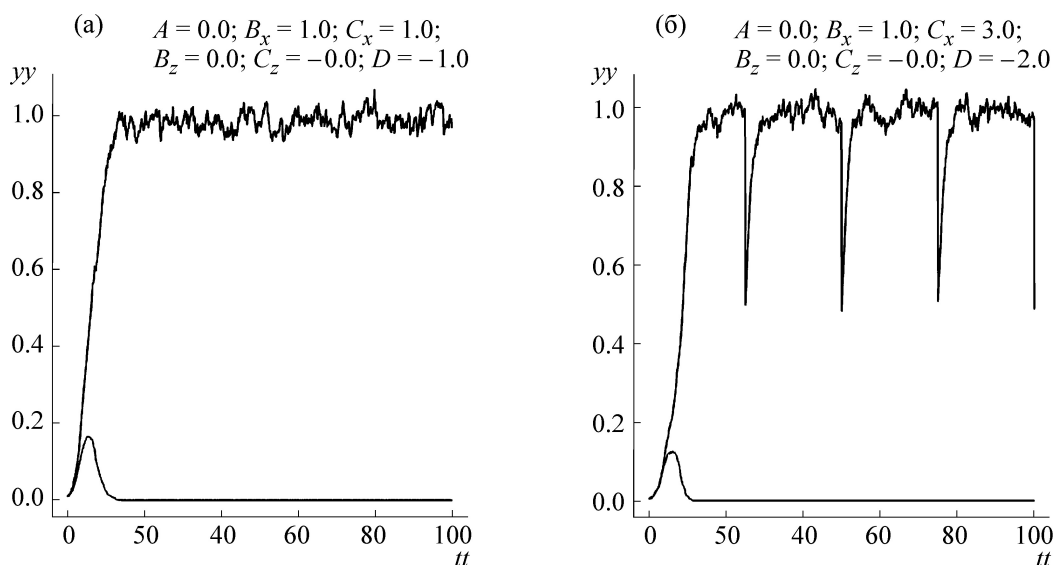


Рис. 3. (а) – Решение системы уравнений (7) при тех же параметрах, что и в предыдущем случае; (б) – решение уравнений при условии, что каждые 25 единиц времени происходит деление.

видимому, в этом случае работают другие механизмы, а предложенная модель не претендует на описание этих случаев.

Работа выполнена при финансовой поддержке Российского научного фонда, грант № 14-14-00088.

СПИСОК ЛИТЕРАТУРЫ

1. S. Augui, E. P. Nora, and E. Heard, *Nat. Rev. Genet.* **12** (6), 429 (2011).
2. S. K. Kota, *BioEssays* **31** (12), 1280 (2009).
3. M. Melé and J. L. Rinn, *Mol. Cell.* **62** (5), 657 (2016).
4. D. P. Barlow, M. S. Bartolomei, *Cold Spring Harb. Perspect. Biol.* **6** (2), pii: a018382 (2014).
5. J. J. Quinn and H. Y. Chang, *Nat. Rev. Genet.* **2016** (1), 47 (2016).
6. T. Metsalu, T. Viltrop, A. Tüürats, et al., *Epigenetics* **9** (10), 1397 (2014).
7. C. Vettermann and M. S. Schlissel, *Immunol. Rev.* **237** (1), 22 (2010).
8. X. J. Tian, H. Zhang, J. Sannerud, and J. Xing, *Proc. Natl. Acad. Sci. USA* **113** (21), E2889 (2016).
9. A. Chess, *Nat. Rev. Genet.* **13** (6), 421 (2012).
10. B. Reinius and R. Sandberg, *Nat. Rev. Genet.* **16** (11), 653 (2015).
11. D. J. Burgess, *Nat. Rev. Genet.* **15** (3), 143 (2014).
12. A. Nag, S. Vigneau, V. Savova, et al., *G3: Genes/Genomes/Genetics* **5** (8), 1713 (2015).

A Model for Regulation of Monoallelic Gene Expression

A.A. Mironov* **

*Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Leninskiye Gory 1/73, Moscow, 119234 Russia

**Institute for Information Transmission Problems, Russian Academy of Sciences, Bol'shoy Karetnyi per. 19, Moscow, 127051 Russia

The majority of higher eukaryotes are diploids. Most genes are expressed from both chromosomal copies, but there are several important exceptions, in particular, the imprinted form of mammalian X-chromosome, genes of the immune system. In some cases, the role of interaction between sense and antisense transcripts, providing the work of only one copy of the gene, was shown. In this paper, a minimal model in which the cells exhibit strict monoallelic expression is proposed and analyzed.

Key words: monoallelic expression, dynamic model, genetic imprinting