

ЭФФЕКТ СУПРЕССИИ P-ЭЛЕМЕНТОВ С МАРКЕРНЫМ ГЕНОМ MINI-WHITE В МЕЖГЕННЫХ РАЙОНАХ *DROSOPHILA MELANOGASTER*

© В. Н. Бабенко, В. Ф. Матвиенко

Институт цитологии и генетики СО РАН, Новосибирск;
электронный адрес: bob@bionet.nsc.ru

Рассмотрено распределение P-элементов с маркерным геном mini-white (цвет глаз) в различных по типу межгенных районах в геноме *Drosophila melanogaster*. Классификация межгенных районов проведена по количеству промоторов в нем на основе взаимной ориентации фланкирующих межгенный район генов. Выяснилось, что наибольшее количество транспозонов расположено в межгенных районах с общей промоторной областью фланкирующих генов, при этом наибольшая супрессия маркерного гена транспозонов наблюдается в районах, фланкированных генами с общей 3'-областью. Показана связь различных классов межгенных районов с различным состоянием хроматина.

Ключевые слова: *Drosophila*, взаимная ориентация генов, хроматин, время репликации.

Принятые сокращения: БД — база данных, п. н. — пара нуклеотидов, т. п. н. — тысяча пар нуклеотидов.

Области, свободные от генов, кодирующих белок, занимают порядка половины размера генома эукариот. При этом межгенные районы содержат различные элементы, начиная от транскрипционно «тихих» тандемно повторяющихся сателлитных последовательностей и заканчивая транскрипционно активной некодирующей РНК. В геноме человека с помощью ассоциативного анализа найдено множество точковых полиморфизмов, связанных с проявлением болезни или других фенотипических признаков, около половины которых лежит в межгенных областях (Hindorff et al., 2009). Таким образом, учитывая размер данной категории генома и его проявление в манифестации фенотипа, крайне актуален вопрос о его функциональной значимости.

Особенности приоритета встраивания транспозонов в геномный контекст *Drosophila melanogaster* исследуются на протяжении многих лет. Попытки найти локальный нуклеотидный контекст (Liao et al., 2000; Linheiro, Bergman, 2008, 2012) не идентифицировали однозначного консенсуса встроок. Вместе с тем анализ карты локализации транспозонов отчетливо показал корреляцию сайтов встроок с промоторными районами (Liao et al., 2000), при этом степень корреляции зависит от линии P-элементов (Babenko et al., 2010). Позже была показана большая аффинность встроок P-элементов к сайтам посадки комплекса узнавания стартов репликации (origin recognition complex, ORC; Spradling et al., 2011), окружение которых включает в себя большинство промоторных районов, но степень доступности этих сайтов во многих районах регулируется состоянием хроматина (Filion et al., 2010). Около половины проаннотированных транспозонов в *D. melanogaster* расположено в межгенных областях (Babenko et al., 2010).

В настоящей работе мы продолжили исследование анализа сайтов встроок, рассмотрев межгенные области. Выяснилось, что число встроок в межгенные области линейно зависит от числа промоторов в нем, что подчеркивает связь сайта инсерции и промотора. Примечательно, что при этом наибольшая супрессия гена-репортера принадлежит встройкам, лежащим в межгенах без промоторов.

Использовали базу данных (БД) Flybase_2012_06 (www.flybase.org) для экстракции генов для анализа и картирования межгенных районов, при этом рассматривали эухроматиновые районы хромосом X, 2 и 3. При компиляции выборки 12 492 соседних генов с перекрытием разделяли сопоставленные гены в прямой и обратной цепях во избежание введения нормировки, таким образом получили 4 класса по типу взаимной ориентации (см. таблицу). Выборка 2852 транспозонов с геном mini-white и оцененным фенотипическим параметром (цвет глаз) была взята из предыдущей нашей работы (Babenko et al., 2010). Анализ зависимости времени репликации межгенной области от числа промоторов в ней проводили по данным для линии клеток Kc157 (Schwaiger et al., 2009). За время репликации межгенной области было взято усредненное время репликации фланкирующих генов. Из рассмотрения был исключен гистонный кластер района 39D.

Начальное распределение разных категорий межгенных областей дано в таблице. Как видно, число генов с конвергентной и дивергентной ориентацией преобладает над генами с одинаковой ориентацией. Число промоторов в межгенных областях в зависимости от класса пары генов может быть равно 0, когда район фланкирован 3'-областями генов (конвергентных генов), 1, когда район содержит одну 5'- и одну 3'-область фланкирующих генов

Распределение числа ориентированных пар генов в геноме *D. melanogaster*

Класс взаимной ориентации генов	Ориентация генов по цепи (+, -)		Число пар соседних генов	Число промоторов в межгенной области с данной ориентацией генов
	левый	правый		
Коллинеарные (однаправленные) гены, прямая нить	+	+	2764	1
Ковергентные гены («хвост к хвосту»)	+	-	3473	0
Дивергентные гены («голова к голове»)	-	+	3472	2
Коллинеарные (однаправленные) гены, обратная нить	-	-	2783	1
Всего пар			12 492	

Примечание. Коллинеарные гены рассмотрены раздельно по нитям ДНК (прямая и обратная) во избежание нормировки и для контрольного дополнения.

(однаправленных), и 2, когда область содержит две 5'-области фланкирующих генов (дивергентных генов).

При анализе распределения числа встроек из предыдущей работы (Babenko et al., 2010) выяснилось, что сайты встраивания транспозонов в межгенные области линейно зависят от числа промоторов в ней (рис. 1). Примечательно, что степень достоверности корреляции крайне высока ($P < 1E-11$), несмотря на гетерогенность межгенных областей по размеру: от 2 п. н. до 345 т. п. н.

Затем было проанализировано распределение частот встроек по категориям межгенных областей и фенотипической манифестации (степень супрессии гена *mini-white*). На рис. 2 показан график, примечательной особенностью которого является класс конвергентных («хвост к хвосту») межгенных областей с наиболее значительной долей супрессированного маркера (38 % встроек от общего числа встроек этой категории) в отличие от остальных категорий (1518 % встроек). Достоверность различия вышеуказанных категорий $P < 2.1E-5$.

Мы также провели сравнение числа категоризованных межгенных областей по числу поздних и ранних районов репликаций в соответствии с работой (Schwaiger et al., 2009). Здесь также наблюдается регрессионная за-

висимость времени репликации межгенной области в зависимости от числа промоторов.

Финальным этапом анализа было исследование состояний хроматина в местах встроек по разным классам межгенных районов. Было рассмотрено 5 состояний хроматина по степени убывания его активности: «красный», «желтый», «зеленый», «синий» и «черный» (по: Filion et al., 2010). Мы исключили рассмотрение зеленого хроматина ввиду малочисленности встроек (меньше 10) и построили гистограмму относительной встречаемости такого хроматина (рис. 3). Из него видно, что наиболее активный хроматин содержат межгенные районы, фланкированные дивергентными генами (рис. 3, <-->). Наиболее гетерохроматиновые районы оказались у конвергентно ориентированных генов (рис. 3, <->). Заметим, что во всех категориях межгенных областей значительная часть встроек произошла в PolycombG (PcG)-зависимый хроматин (рис. 3, «голубой»).

Мы рассмотрели распределение и особенности супрессии транспозонов с геном-репортером в зависимости от типа межгенных областей. В соответствии с полученными результатами мы можем констатировать, что межгенные области с конвергентными генами («хвост к хвос-

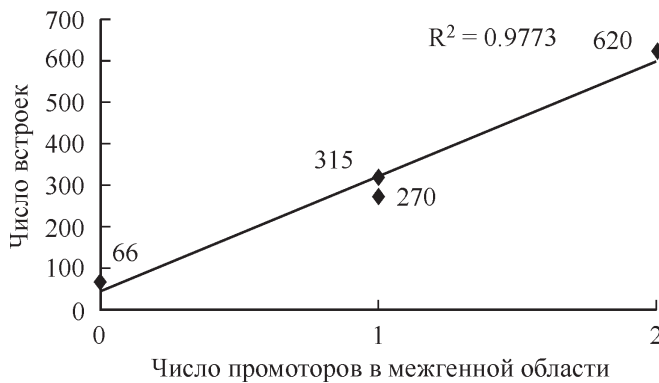


Рис. 1. Зависимость числа встроек транспозонов с известным фенотипом гена *mini-white* (Babenko et al., 2010) в межгенных областях (всего 1271 из 2852) от числа промоторов в межгенной области.

Рассмотрены 3 возможных типа межгенных областей, категоризованных по типу фланкирующих генов: гены расположены «голова к голове» (2 промотора); гены однонаправлены (1 промотор; по каждой цепи подсчет отдельный); гены расположены «хвост к хвосту» (0 промоторов). Достоверность корреляции $P < 1E-11$. Цифры у точек обозначают точное число встроек данной категории.

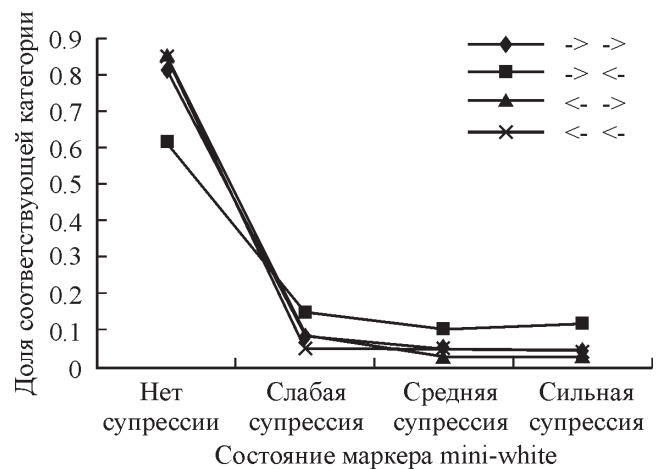


Рис. 2. Распределение степени супрессии гена *mini-white* от типа межгенной области.

→ и ← обозначают направление фланкирующих генов 5' → 3' и 3' → 5' соответственно. Достоверность превышения класса встроек с выраженной супрессией в конвергентных межгенных областях (красный цвет) $P < 2.4E-5$, точный критерий Фишера для таблиц 2 × 2.

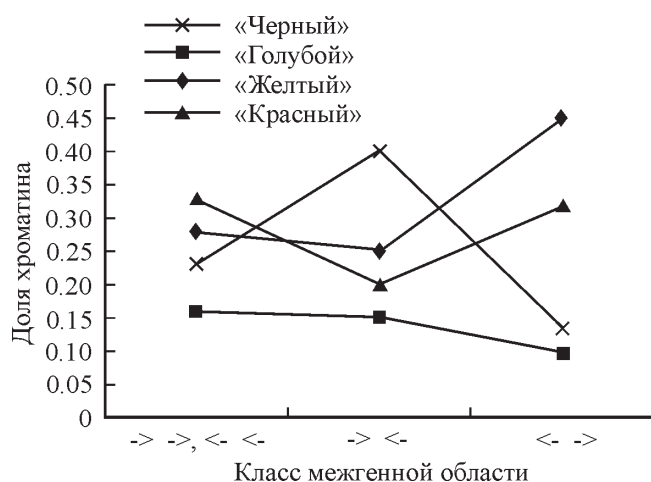


Рис. 3. Относительная доля состояния хроматина, в которой локализованы P-элементы в соответствии с номенклатурой хроматина (Filion et al., 2010).

«Красный» и «желтый» — открытый хроматин, «голубой» — PcG-хроматин, «черный» — репрессированный эухроматин. → и ← обозначают направление фланкирующих генов 5' → 3' и 3' → 5' соответственно.



Рис. 4. Зависимость времени репликации межгенной области от числа промоторов в ней по данным для линии клеток Kc157 (Schwaiger et al., 2009).

За время репликации межгенной области взято усредненное время репликации фланкирующих генов. Из рассмотрения был исключен гистонный кластер района 30D. Цифры у точек — число межгенных областей, по которым оценивали параметр времени репликации.

ту») более склонны к поздней репликации (рис. 4), гетерохроматинизации и другим признакам интеркалярного гетерохроматина (Belyakin et al., 2010), нежели дивергентные и коллинеарные межгенные области. Это объясняет более выраженную супрессию относительно немногочисленных встроек P-элементов в конвергентные межгенные области и существенное подавление экспрессии репортера (Babenko et al., 2010).

Как показано в предыдущей работе (Babenko et al., 2010), наиболее приемлемый для встройки P-элементов — это открытый хроматин (рис. 3, «красный», «желтый»; Fillion et al., 2010), в меньшей степени — PcG-хроматин («голубой»). «Черный» хроматин не является характерным для встроек (Babenko et al., 2010), его значительное присутствие на рис. 3 объясняется его распространенностью в эухроматиновом геноме *D. melanogaster* (около 50 % от размера генома; Filion et al., 2010). Таким образом, присутствие «черного» хроматина на рис. 3 в соответствующей группе коррелирует с дефици-

том относительной доли встроек. Следствием большой доли «черного» хроматина на рис. 3 в конвергентных межгенных областях (средней категории на рис. 3) является наименьшее число встроек в эту межгенную категорию. При этом показано, что именно «черный» хроматин имеет наибольший потенциал супрессии путем подавления транскрипции рассматриваемых P-элементов (Filion et al., 2010), что наблюдается и в нашем случае.

В свою очередь дивергентные межгенные области имеют выраженные признаки открытого хроматина (рис. 3) и как следствие наибольшее насыщение встройками P-элементов (рис. 1) при минимальной частоте подавления транскрипции (рис. 2).

Как следует из вышесказанного, генная ориентация не всегда определяет свойства хроматина. Например, распределение парной генной ориентации в 52 недореплицированных районах политенных хромосом (Belyakin et al., 2010) практически совпадает со средним по геному, т. е. случайно. Распределение ориентированных пар в выборке междисков также оказалось случайным (из личной беседы с Е. С. Беляевой). Примером неслучайного направления генов является выборка генов мужского тестиса (Shevelyov et al., 2009). В выборке 1652 генов этого класса присутствует достоверный избыток коллинеарно направленных генов.

В целом результаты работы подтверждают предыдущие выводы (Babenko et al., 2010), но при этом проливают свет на соответствие типов межгенных областей различным состояниям хроматина, что в целом приводит к выводу о неслучайной природе взаимной ориентации соседних генов.

Авторы благодарны Т. И. Аксенович, И. Ф. Жимулеву и Е. С. Беляевой за критические замечания. Часть вычислений производили на кластере НКС-160 в Сибирском суперкомпьютерном центре СО РАН в Институте вычислительной математики и математической геофизики (www.sccc.ru).

Список литературы

- Babenko V. N., Makunin I. V., Brusentsova I. V., Belyaeva E. S., Maksimov D. A., Belyakin S. N., Maroy P., Vasil'eva L. A., Zhimulev I. F. 2010. Paucity and preferential suppression of transgenes in late replication domains of the *D. melanogaster* genome. BMC Genomics. 11 : 318.
- Belyakin S. N., Babenko V. N., Maksimov D. A., Shloma V. V., Kvon E. Z., Belyaeva E. S., Zhimulev I. F. 2010. Gene density profile reveals the marking of late replicated domains in the *Drosophila melanogaster* genome. Chromosoma. 119 : 589—600.
- Filion G. J., van Bemmel J. G., Braunschweig U., Talhout W., Kind J., Ward L. D., Brugman W., de Castro I. J., Kerkhoven R. M., Bussemaker H. J., van Steensel B. 2010. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. Cell. 143 : 212—224.
- Hindorff L. A., Sethupathy P., Junkins H. A., Ramos E. M., Mehta J. P., Collins F. S., Manolio T. A. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Nat. Acad. Sci. USA. 106 : 9362—9367.
- Liao G. C., Rehm E. J., Rubin G. M. 2000. Insertion site preferences of the P transposable element in *Drosophila melanogaster*. Proc. Nat. Acad. Sci. USA 97 : 3347—3351.
- Linheiro R. S., Bergman C. M. 2008. Testing the palindromic target site model for DNA transposon insertion using the *Drosophila melanogaster* P-element. Nucl. Acids Res. 36 : 6199—6208.

Linheiro R. S., Bergman C. M. 2012. Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. PLoS One. 7(2) : e30008.

Schwaiger M., Stadler M. B., Bell O., Kohler H., Oakeley E. J., Schubeler D. 2009. Chromatin state marks cell-type- and gender-specific replication of the *Drosophila genome*. Genes Develop. 23 : 589—601.

Shevelyov Y. Y., Lavrov S. A., Mikhaylova L. M., Nurminsky I. D., Kulathinal R. J., Egorova K. S., Rozovsky Y. M., Nurminsky

D. I. 2009. The B-type lamin is required for somatic repression of testis-specific gene clusters. Proc. Nat. Acad. Sci. USA. 106 : 3282—3287.

Spradling A. C., Bellen H. J., Hoskins R. A. 2011. *Drosophila* P elements preferentially transpose to replication origins. Proc. Nat. Acad. Sci. USA. 108 : 15 948—15 953.

Поступила 26 XI 2012

P ELEMENTS COMPRISING MINI-*WHITE* MARKER GENE SUPPRESSION FEATURES
IN INTERGENIC REGIONS OF *DROSOPHILA MELANOGASTER* GENOME

V. N. Babenko, V. F. Matvienko

Institute of Cytology and Genetics SB RAS, Novosibirsk;
e-mail: bob@bionet.nsc.ru

12 492 intergenic regions were stratified into four classes according to protein coding genes mutual orientation flanking an intergenic region. It was revealed that transposon insertion sites number linearly correlates with number of promoters (none, one, or two) in an intergenic region. The vast majority of transposons reside in intergenic regions with two promoters. Remarkably, the suppression manifestation, on the contrary, was most pronounced in «promoterless» intergenic regions. Discussion of the phenomenon is based on the chromatin state analysis of various intergenic regions.

Key words: *Drosophila melanogaster*, mutual gene orientation, chromatin, replication timing.