

УДК 577.21

Белок CG9890 дрозофилы, содержащий домены цинковых пальцев, колокализуется с комплексами модификации и ремоделирования хроматина на промоторах генов и участвует в регуляции транскрипции

Н. А. Фурсова, М. Ю. Мазина, Ю. В. Николенко, Н. Е. Воробьева, А. Н. Краснов*

Институт биологии гена РАН, Москва, 119334 Россия

*E-mail: krasnov@genebiology.ru

Поступила в редакцию 22.05.2020

Принята к печати 19.06.2020

DOI: 10.32607/actanaturae.11056

РЕФЕРАТ Проведено полногеномное исследование белка CG9890, содержащего домены цинковых пальцев. Показано, что белок CG9890 локализован преимущественно на промоторах активных генов. Сайты связывания CG9890 характеризуются низкой плотностью нуклеосом и колокализируются с комплексами модификации и ремоделирования хроматина SAGA и dSWI/SNF, а также с репликационным комплексом ORC. Показано, что белок CG9890 участвует в регуляции экспрессии части генов, на промоторах которых он находится, причем заметную долю этих генов составляют гены экдизонового каскада. Таким образом, белок CG9890 является новым участником транскрипционной сети клетки, локализуется на активных промоторах, взаимодействует с основными транскрипционными и репликационными комплексами и участвует в регуляции как базальной, так и индуцибельной транскрипции.

КЛЮЧЕВЫЕ СЛОВА ENY2, CG9890, дрозофила, цинковые пальцы, ChIP-Seq.

СПИСОК СОКРАЩЕНИЙ ENY2 – мультифункциональный транскрипционный фактор; C2H2 – цинковые пальцы типа C2H2; SAGA – гистон-ацетилтрансферазный комплекс; SWI/SNF – комплекс ремоделирования хроматина; AMEX – комплекс экспорта мРНК из ядра в цитоплазму; TNO – комплекс элонгации транскрипции; ORC – комплекс позиционирования точек начала репликации в геноме.

ВВЕДЕНИЕ

Ранее в нашей лаборатории был выделен и охарактеризован белок ENY2, который впоследствии оказался компонентом множества белковых комплексов, вовлеченных в регуляцию транскрипции и репликации. ENY2 входит в состав транскрипционных комплексов SAGA, AMEX, TNO и связывает различные этапы экспрессии генов – организацию транскрипционных доменов и модификацию хроматина, активацию и элонгацию транскрипции, экспорт мРНК и регуляцию пространственного расположения генов в ядре [1–7]. Кроме того, обнаружено, что ENY2 также входит в состав репликационного комплекса ORC, ответственного за позиционирование точек начала репликации [8–11].

Анализ двухгибридного взаимодействия ENY2-Su(Hw) привел к пониманию того, что Su(Hw) привлекает белок ENY2 на Su(Hw)-зависимые инсуляторы дрозофилы, что необходимо для осуществления барьерной функции [5]. Далее показали, что Su(Hw) рекрутирует гистон-ацетилтрансферазный комплекс SAGA [12] (содержащий ENY2) и комплекс ремоделирования хроматина dSWI/SNF [13–15] на Su(Hw)-зависимые инсуляторы, вызывая формирование области с низкой плотностью нуклеосом и условий для посадки репликационного комплекса ORC. Нокдаун Su(Hw) практически полностью нарушает привлечение комплексов SAGA, dSWI/SNF и ORC на Su(Hw)-зависимые инсуляторы и приводит к значительному увеличению плотности нуклео-

сом на этих регуляторных элементах [1, 2]. Показано, что Su(Hw) – это первый пример белка, ответственного за позиционирование точек начала репликации. Su(Hw) необходим для формирования 6% точек начала репликации в геноме дрозофилы, т.е. за формирование остальных 94% отвечают какие-то другие, еще не идентифицированные белки.

Ранее мы обнаружили взаимодействие ENY2 с еще одним белком, CG9890, который содержит домен цинковых пальцев C2H2-типа, как и Su(Hw) [16]. Мы предполагаем, что, как и Su(Hw), CG9890 является ДНК-связывающим белком, который привлекает ENY2-содержащие комплексы на свои сайты связывания, организуя, таким образом, регуляторные элементы генома, необходимые для функционирования клетки. Показано, что белок CG9890 локализован в ядре клетки. Биохимическими методами выявлено взаимодействие белка CG9890 с ENY2-содержащими комплексами SAGA, ORC, dSWI/SNF, TFIID и THO [16]. CG9890 взаимодействует с транскрипционными комплексами, вовлеченными в инициацию и элонгацию транскрипции, но не взаимодействует с комплексом AMEX, участвующим в экспорте мРНК из ядра в цитоплазму, что указывает на работу CG9890 в первых стадиях транскрипционного цикла.

В данной работе проведен полногеномный анализ белка CG9890 с целью идентификации и характеристики регуляторных элементов, за которые он может быть ответствен.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Антитела и клеточные линии

В работе использовали линию клеток S2 *Drosophila melanogaster*. Поликлональные антитела α-CG9890 получены из сыворотки крови кролика, иммунизированного полноразмерным белком CG9890, экспрессированным в клетках *Escherichia coli* [16].

Иммунопреципитация хроматина и полногеномное секвенирование

Иммунопреципитацию хроматина проводили согласно [1]. Для этого использовали поликлональные антитела к CG9890, полученные нами ранее [16]. ChIP-Seq-библиотеки получали с использованием набора NEBNext DNA library preparation kit (New England Biolabs). Качество библиотек проверяли с помощью Bioanalyzer. Для высокопроизводительного секвенирования брали фрагменты размером 200–500 п.н. Библиотеки были отсеквенированы на геномном секвенаторе Illumina HiSeq 2000. Полученные последовательности картировали на референсный геном дрозофилы с использованием программы Bowtie2. Для последующего анали-

за использовали только уникально картированные риды. Идентификацию координат пиков и построение полногеномного профиля (WIG-файл) для белка CG9890 проводили с использованием программы SPP (FDR < 5%) [17]. Областью пика считали геномный интервал +/- 100 п.н. от позиции вершины пика.

Биоинформатический анализ

Аннотации генов *D. melanogaster* были взяты с официального сайта FlyBase. Геном разбивали на следующие области: точки начала транскрипции (TSS), точки терминации транскрипции (TES), транскрибируемые области (участки генов за исключением TSS и TES) и межгенные области (все остальное). ChIP-Seq-пик идентифицировали как принадлежащий к одной из этих категорий при условии пересечения геномных интервалов как минимум в 10 п.н. При аннотации пиков использовали следующий приоритет геномных категорий – TSS, TES, транскрибируемые и межгенные области.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Белок CG9890 локализован преимущественно на промоторах генов

С целью определения локализации изучаемого белка в геноме мы провели иммунопреципитацию хроматина из клеток S2 с помощью поликлональных антител к белку CG9890 с последующим высокопроизводительным секвенированием (ChIP-Seq). Характерный ChIP-Seq-профиль белка CG9890 на одном из сайтов его связывания представлен на рис. 1. Идентифицировано 4709 сайтов посадки белка CG9890 в геноме дрозофилы (FDR < 5%).

Мы аннотировали полученные сайты по признаку их локализации в одном из следующих элементов генома дрозофилы: промоторы, концы генов, тела генов и межгенные области. Согласно полученным данным (рис. 2), наибольшее количество ChIP-Seq-пиков белка CG9890 (73.2%) локализуется в промоторных областях генов дрозофилы. Мы предполагаем, что, локализуясь преимущественно на промоторах генов, белок CG9890 может участвовать в функционировании регуляторных генетических элементов этого типа.

Белок CG9890 колокализуется с комплексами модификации и ремоделирования хроматина в областях с низкой плотностью нуклеосом

Ранее мы подтвердили взаимодействие CG9890 с белком ENY2 и выявили взаимодействие CG9890 с ENY2-содержащими белковыми комплексами SAGA, ORC, dSWI/SNF, TFIID, THO. Поэтому мы изучили геномную колокализацию белка CG9890

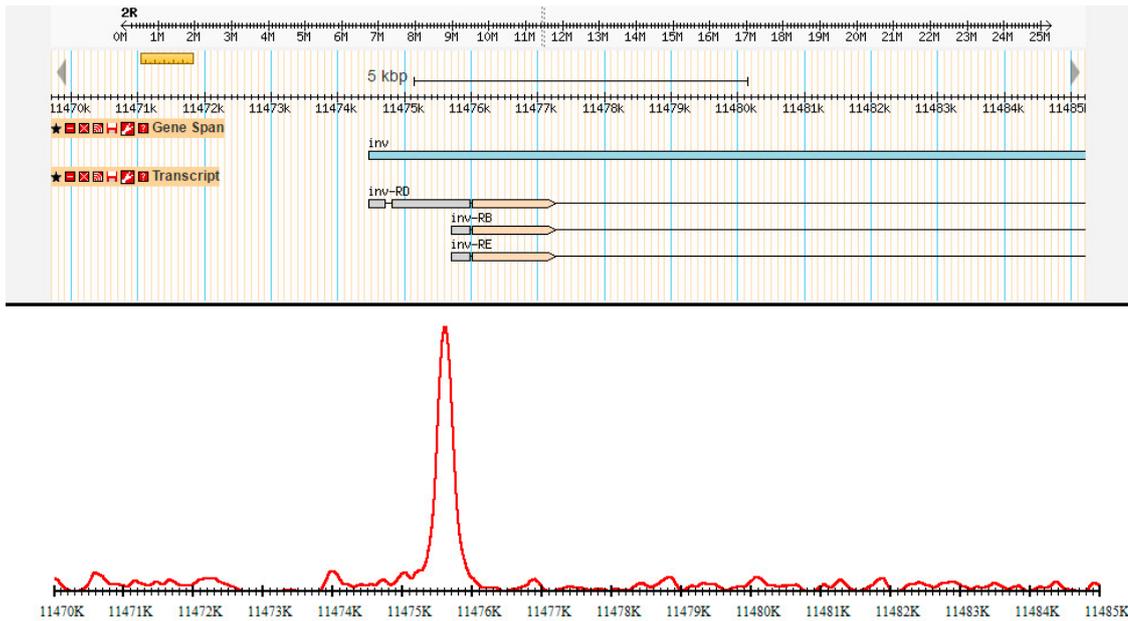


Рис. 1. Характерный ChIP-Seq-профиль белка CG9890. Представлен геномный регион, соответствующий промоторной области гена *Inv*. Верхняя часть рисунка содержит информацию об этом регионе из геномного браузера, нижняя сам профиль ChIP-Seq

с комплексами SAGA и ORC, а также с комплексом ремоделирования dSWI/SNF, который вместе с комплексом SAGA участвует в формировании нужной структуры хроматина для правильного функционирования регуляторных элементов, включая промоторы. С этой целью мы использовали программное обеспечение собственной разработки по построению усредненного профиля исследуемого фактора на заданных сайтах генома [1]. Геномные профили белков ORC2, GCN5, OSA и гистона H3 были получены ранее в нашей лаборатории. Рассчитаны усредненные профили данных белков на всех 4709 сайтах посадки белка CG9890, а также на 4709 случайных промоторах и на 4709 случайных геномных сайтах (рис. 3).

Так как белок CG9890 локализован преимущественно на промоторах генов, то на сайтах связывания CG9890 можно будет найти обогащение любых промоторных факторов, включая комплексы SAGA, dSWI/SNF и ORC, относительно среднегеномного уровня. Однако, как можно видеть из рис. 3, белки GCN5 (комплекс SAGA), OSA (комплекс dSWI/SNF) и ORC2 (комплекс ORC) обогащены на сайтах связывания белка CG9890 не только по сравнению со среднегеномным уровнем, но и по сравнению со случайными промоторами. Полученный результат свидетельствует, что данная колокализация обусловлена не просто случайным совпадением на промоторах генов, а тем, что именно сайт связывания белка CG9890 способствует локализации комплексов SAGA, dSWI/SNF и ORC. Кроме того, как следует из рис. 3, сайты связывания белка CG9890 характеризуются более низкой нуклеосомной плотностью (представленностью гистона H3), чем в среднем по ге-

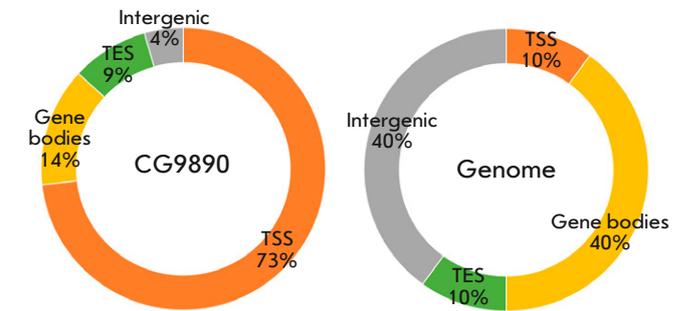


Рис. 2. Распределение сайтов связывания белка CG9890 относительно аннотированных элементов генома дрозофилы (левая картинка). Для сравнения показана относительная представленность всех аннотированных элементов в геноме (правая картинка). TSS – промоторная область, TES – конец гена, Gene bodies – область гена между TSS и TES, Intergenic – межгенные области

ному и на промоторах, что указывает на активное состояние данных регуляторных элементов.

С помощью второго подхода мы рассчитали количество сайтов белка CG9890, перекрывающихся с сайтами белков GCN5, OSA и ORC2. В качестве контрольного фактора был выбран хорошо известный промоторный белок BEAF-32 [18]. Координаты ChIP-Seq-пики для белка BEAF-32 получены из NCBI GEO (GSE35648). Пики двух белков считали пересекающимися при условии пересечения их геномных интервалов как минимум в 10 п.н. Полученные данные представлены на рис. 4.

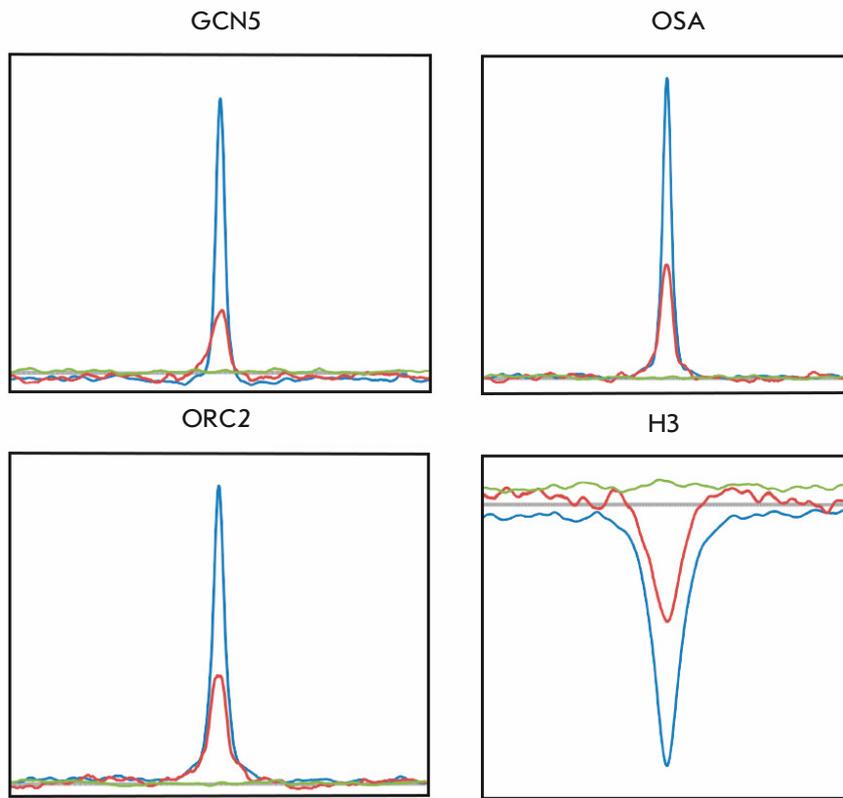


Рис. 3. Усредненный геномный профиль (log₂ обогащения) белков GCN5 (комплекс SAGA), OSA (комплекс dSWI/SNF), ORC2 (комплекс ORC), а также гистона H3 в области от -5 до +5 т.п.н относительно указанных ниже сайтов. Синим цветом отмечен усредненный профиль указанного белка на 4709 сайтах CG9890 в геноме, красным на таком же количестве случайно выбранных промоторов, зеленым на таком же количестве случайных участков генома

— Сайты CG9890
— промоторы
— случайные сайты

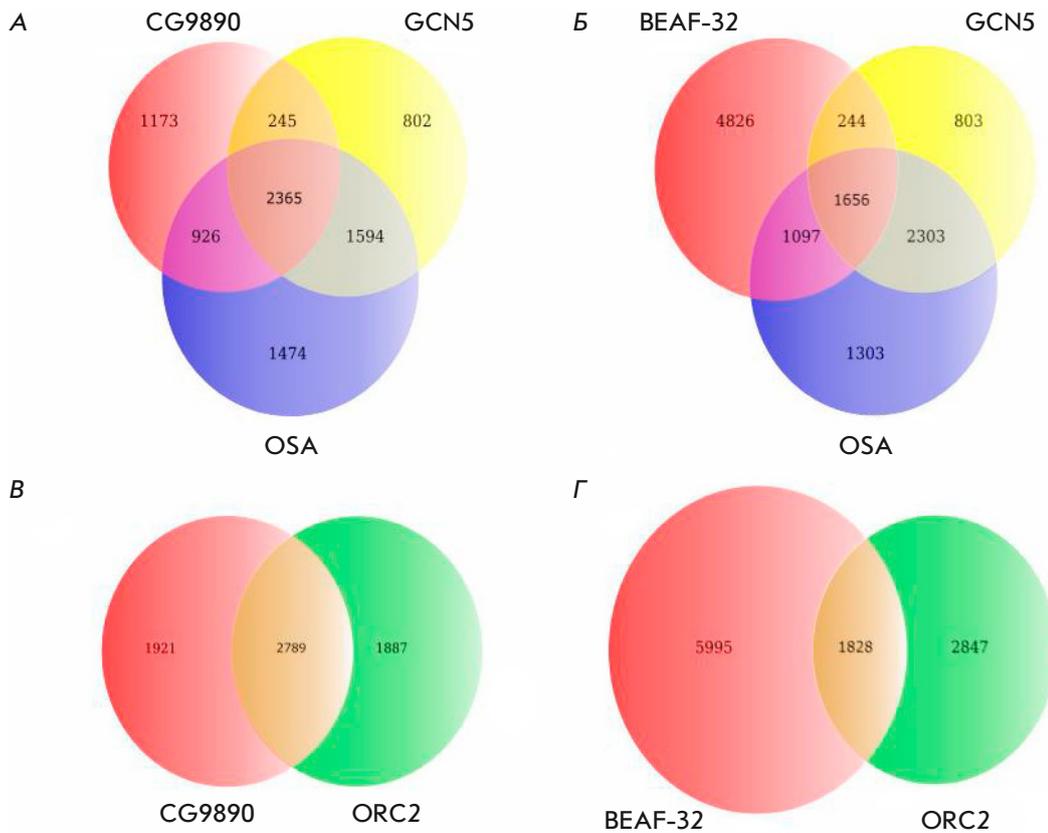


Рис. 4. Диаграммы Эйлера–Венна, показывающие перекрывание сайтов связывания белков CG9890 и BEAF-32 с сайтами связывания GCN5 (комплекс SAGA), OSA (комплекс dSWI/SNF) и ORC2 (комплекс ORC).
 А – CG9890, GCN5, OSA;
 Б – BEAF-32, GCN5, OSA;
 В – CG9890 и ORC2;
 Г – BEAF-32 и ORC2

Как можно видеть из *рис. 4*, около 60% сайтов белка CG9890 перекрываются с сайтами белка ORC2, субъединицы комплекса ORC, что, в свою очередь, составляет около 60% сайтов белка ORC2. Уровень перекрывания сайтов ORC2 с сайтами BEAF-32, другого фактора, локализирующегося на промоторах, существенно ниже, несмотря на то, что сайтов посадки белка BEAF-32 в геноме гораздо больше. Из *рис. 4* видно, что белок CG9890 колокализуется с белками GCN5 и OSA на половине своих сайтов посадки в геноме, что существенно выше аналогичного значения для контрольного белка BEAF-32.

Белок CG9890 участвует в регуляции экспрессии генов

Ранее нами было показано, что белок CG9890 взаимодействует с белком ENY2, который координирует множество этапов регуляции экспрессии генов. Выявлено взаимодействие белка CG9890 с ENY2-содержащими комплексами SAGA, ORC, dSWI/SNF, TFIID и TPOC [16], т.е. комплексами, вовлеченными в инициацию и элонгацию транскрипции. Учитывая, что CG9890 найден преимущественно на промоторах генов, мы решили исследовать, к каким изменениям в экспрессии CG9890-ассоциированных генов приведет снижение внутриклеточного уровня белка CG9890 методом РНК-интерференции. В результате оптимизации условий РНК-интерференции удалось добиться эффективного уменьшения экспрессии из-

учаемого белка в клетках более чем в 5 раз по количеству мРНК и фактически полное истощение количества белка (ниже предела детекции методом вестерн-блоттинга).

Используя метод ОТ-кПЦР, мы проанализировали изменение уровня мРНК 21 CG9890-ассоциированного гена в клетках после РНК-интерференции по сравнению с контрольными образцами. Результаты этого эксперимента представлены на *рис. 5*. После нокдауна белка CG9890 количество мРНК семи из этих генов снизилось как минимум на 20%, а количество РНК трех генов возросло как минимум на 20%. Таким образом, белок CG9890 действительно участвует в регуляции экспрессии по крайней мере части генов, на промоторах которых он локализуется.

Среди 10 генов, экспрессия которых изменилась статистически значимо при РНК-интерференции белка CG9890, пять являются генами экдизонового каскада. Их транскрипция значительно активируется в процессе ответа на экдизон. Это дает возможность воспользоваться удобной модельной системой индукции клеток экдизоном для более детального изучения функционирования белка CG9890 в регуляции экспрессии этих генов. Преимущество этой системы состоит в том, что с ее помощью можно изучать динамические процессы индуцибельной регуляции экспрессии генов, а не только поддержание базальной транскрипции [19].

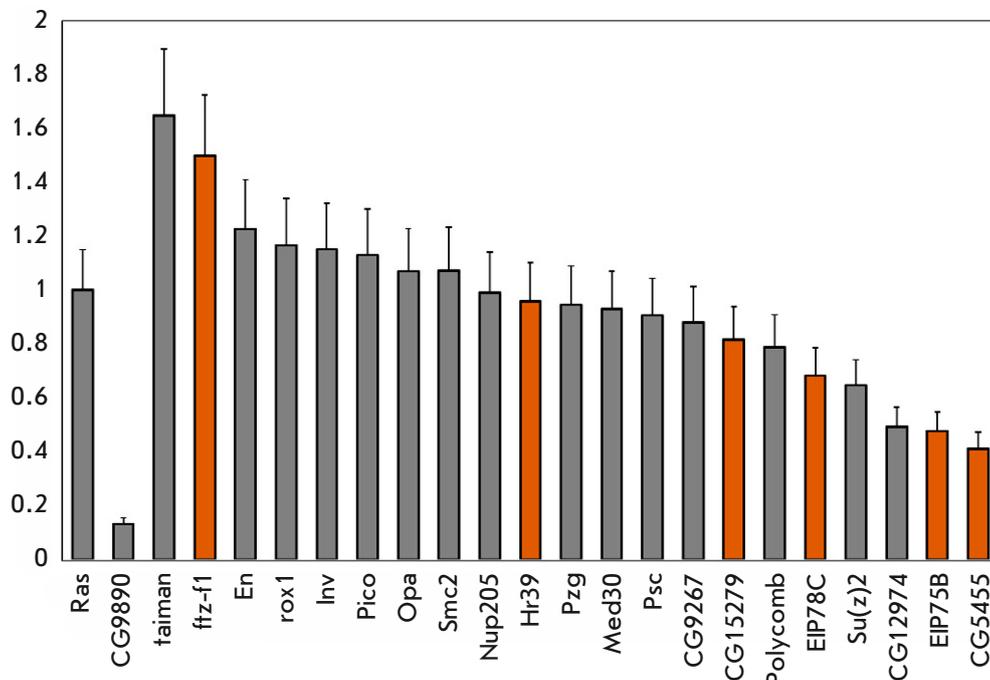


Рис. 5. Изменение уровней экспрессии CG9890-ассоциированных генов после РНК-интерференции белка CG9890. По оси ординат показано изменение уровня мРНК генов после РНК-интерференции относительно исходного уровня. За единицу принят исходный уровень. Для нормирования использовали содержание мРНК гена *Ras* в соответствующих образцах. Линии погрешностей соответствуют стандартной ошибке среднего. Оранжевые столбики соответствуют генам экдизонового каскада

ЗАКЛЮЧЕНИЕ

В ходе предыдущих исследований мы обнаружили, что инсуляторный белок Su(Hw), содержащий домен цинковых пальцев, взаимодействует с ENY2 и привлекает ENY2-содержащие комплексы на Su(Hw)-зависимые инсуляторы дрозофилы, участвуя одновременно в регуляции транскрипции и в позиционировании точек начала репликации. Также мы обнаружили взаимодействие ENY2 с еще одним белком CG9890, который содержит домен цинковых пальцев, как и Su(Hw). Биохимическими методами показано взаимодействие белка CG9890 с ENY2-содержащими комплексами SAGA, ORC, dSWI/SNF, TFIID и TPOC. Мы предполагаем, что, как и Su(Hw), белок CG9890 является ДНК-связывающим белком, который привлекает ENY2-содержащие комплексы на свои сайты связывания, организуя таким образом регуляторные элементы генома, необходимые для функционирования клетки. В данной работе мы идентифицировали сайты связывания CG9890 в геноме и показали, что они локализованы преимущ-

ественно на промоторах генов. Установлена полногеномная корреляция сайтов связывания CG9890 и ENY2-содержащих комплексов SAGA, ORC и dSWI/SNF. Сайты связывания белка CG9890 характеризуются более низкой нуклеосомной плотностью (представленностью гистона H3), чем в среднем по геному и в среднем на промоторах, что указывает на активное состояние данных регуляторных элементов. Белок CG9890 участвует в регуляции экспрессии части генов, на промоторах которых он находится, причем заметную долю этих генов составляют гены экдизонового каскада. Таким образом, белок CG9890 – это новый участник транскрипционной сети клетки, который локализуется на активных промоторах, взаимодействует с основными транскрипционными и репликационными комплексами и участвует в регуляции как базальной, так и индуцибельной транскрипции. ●

Работа выполнена при поддержке Российского научного фонда, грант № 20-14-00269.

СПИСОК ЛИТЕРАТУРЫ

- Vorobyeva N.E., Mazina M.U., Golovnin A.K., Kopytova D.V., Gurskiy D.Y., Nabirochkina E.N., Georgieva S.G., Georgiev P.G., Krasnov A.N. // Nucl. Acids Res. 2013. V. 41. № 11. P. 5717–5730.
- Mazina M., Vorob'eva N.E., Krasnov A.N. // Tsitologiya. 2013. V. 55. № 4. P. 218–224.
- Vorobyeva N.E., Nikolenko J.V., Krasnov A.N., Kuzmina J.L., Panov V.V., Nabirochkina E.N., Georgieva S.G., Shidlovskii Y.V. // Cell Cycle. 2011. V. 10. № 11. P. 1821–1827.
- Kopytova D.V., Krasnov A.N., Orlova A.V., Gurskiy D.Y., Nabirochkina E.N., Georgieva S.G., Shidlovskii Y.V. // Cell Cycle. 2010. V. 9. № 3. P. 479–481.
- Kurshakova M., Maksimenko O., Golovnin A., Pulina M., Georgieva S., Georgiev P., Krasnov A. // Mol. Cell. 2007. V. 27. № 2. P. 332–338.
- Krasnov A.N., Kurshakova M.M., Ramensky V.E., Mardanov P.V., Nabirochkina E.N., Georgieva S.G. // Nucl. Acids Res. 2005. V. 33. № 20. P. 6654–6661.
- Gurskiy D., Orlova A., Vorobyeva N., Nabirochkina E., Krasnov A., Shidlovskii Y., Georgieva S., Kopytova D. // Nucl. Acids Res. 2012. V. 40. № 21. P. 10689–10700.
- Kopytova D., Popova V., Kurshakova M., Shidlovskii Y., Nabirochkina E., Brechalov A., Georgiev G., Georgieva S. // Nucl. Acids Res. 2016. V. 44. № 10. P. 4920–4933.
- Eaton M.L., Prinz J.A., MacAlpine H.K., Tretyakov G., Kharchenko P.V., MacAlpine D.M. // Genome Res. 2011. V. 21. № 2. P. 164–174.
- Masai H., Matsumoto S., You Z., Yoshizawa-Sugata N., Oda M. // Annu. Rev. Biochem. 2010. V. 79. P. 89–130.
- MacAlpine H.K., Gordan R., Powell S.K., Hartemink A.J., MacAlpine D.M. // Genome Res. 2010. V. 20. № 2. P. 201–211.
- Baptista T., Grunberg S., Minoungou N., Koster M.J.E., Timmers H.T.M., Hahn S., Devys D., Tora L. // Mol. Cell. 2017. V. 68. № 1. P. 130–143.
- Shi J., Zheng M., Ye Y., Li M., Chen X., Hu X., Sun J., Zhang X., Jiang C. // Nucl. Acids Res. 2014. V. 42. № 15. P. 9730–9739.
- Harikrishnan K.N., Chow M.Z., Baker E.K., Pal S., Bassal S., Brasacchio D., Wang L., Craig J.M., Jones P.L., Sif S., et al. // Nat. Genet. 2005. V. 37. № 3. P. 254–264.
- Elfring L.K., Daniel C., Papoulas O., Deuring R., Sarte M., Moseley S., Beek S.J., Waldrip W.R., Daubresse G., DePace A., et al. // Genetics. 1998. V. 148. № 1. P. 251–265.
- Fursova N.A., Nikolenko J.V., Soshnikova N.V., Mazina M.Y., Vorobyova N.E., Krasnov A.N. // Acta Naturae. 2018. V. 10. № 4. P. 110–114.
- Kharchenko P.V., Tolstorukov M.Y., Park P.J. // Nat. Biotechnol. 2008. V. 26. № 12. P. 1351–1359.
- Zhao K., Hart C.M., Laemmli U.K. // Cell. 1995. V. 81. № 6. P. 879–889.
- Mazina M.Y., Nikolenko J.V., Fursova N.A., Nedil'ko P.N., Krasnov A.N., Vorobyeva N.E. // Cell Cycle. 2015. V. 14. № 22. P. 3593–3601.