

Перевод на английский язык <https://vavilov.elpub.ru/jour>

Методы высокопроизводительного репортерного анализа энхансеров

С.Е. Романов^{1, 2}, Д.А. Калашникова^{1, 2}, П.П. Лактионов^{1, 2} 

¹ Новосибирский национальный исследовательский государственный университет, лаборатория эпигенетики факультета естественных наук, Новосибирск, Россия

² Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, лаборатория геномики, Новосибирск, Россия
 laktionov@mcb.nsc.ru

Аннотация. Корректное развертывание генетических программ развития и дифференцировки опирается на тонко координированную регуляцию экспрессии специфических наборов генов. Исключительную роль в управлении этим процессом играют регуляторные элементы генома, к которым относятся промоторы, энхансеры, инсуляторы и сайленсеры. Нарушения в их работе могут приводить к развитию различных патологий, включая онкологические заболевания, пороки развития и аутоиммунные заболевания. Развитие технологий высокопроизводительного геномного анализа позволило значительно ускорить накопление информации о специфичных эпигенетических характеристиках регуляторных элементов. В совокупности с полногеномными исследованиями распределения эпигенетических меток, регуляторных белков и пространственной структуры хроматина такие данные значительно расширяют представления о принципах эпигенетической регуляции генов и позволяют осуществлять поиск потенциальных регуляторных элементов *in silico*. Вместе с тем основные экспериментальные подходы, используемые для исследования локальных характеристик хроматина, имеют ряд технических ограничений, которые снижают достоверность биоинформатической идентификации регуляторных областей генома. В связи с этим, а также с учетом вариабельности функций эпигенетических детерминант и многокомпонентной регуляции работы элементов генома определение их регуляторной роли часто требует функциональной проверки. Разработано множество методов, позволяющих провести исследование функциональной роли регуляторных элементов в масштабе генома. В настоящем обзоре кратко описаны основные экспериментальные подходы для проведения идентификации регуляторных элементов *in silico* и присущие им технические ограничения. Рассмотрены оригинальные методы высокопроизводительного репортерного анализа активности энхансеров, которые используют для валидации предсказанных регуляторных элементов и *de novo* поиска. Описанные методы анализа дают возможность оценить функциональную роль нуклеотидной последовательности регуляторного элемента, определить его точные границы, а также оценить влияние локального состояния хроматина на активность энхансеров и экспрессию генов. Применение таких методологических подходов обеспечило значительный вклад в понимание фундаментальных принципов регуляции генной экспрессии.


Ключевые слова: регуляторные элементы генома; энхансеры; высокопроизводительные методы анализа.

Для цитирования: Романов С.Е., Калашникова Д.А., Лактионов П.П. Методы высокопроизводительного репортерного анализа энхансеров. *Вавиловский журнал генетики и селекции*. 2021;25(3):344-355. DOI 10.18699/VJ21.038

Methods of massive parallel reporter assays for investigation of enhancers

S.E. Romanov^{1, 2}, D.A. Kalashnikova^{1, 2}, P.P. Laktionov^{1, 2} 

¹ Novosibirsk State University, Epigenetics Laboratory, Department of Natural Sciences, Novosibirsk, Russia

² Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences, Genomics Laboratory, Novosibirsk, Russia
 laktionov@mcb.nsc.ru

Abstract. The correct deployment of genetic programs for development and differentiation relies on finely coordinated regulation of specific gene sets. Genomic regulatory elements play an exceptional role in this process. There are few types of gene regulatory elements, including promoters, enhancers, insulators and silencers. Alterations of gene regulatory elements may cause various pathologies, including cancer, congenital disorders and autoimmune diseases. The development of high-throughput genomic assays has made it possible to significantly accelerate the accumulation of information about the characteristic epigenetic properties of regulatory elements. In combination with high-throughput studies focused on the genome-wide distribution of epigenetic marks, regulatory proteins and the spatial structure of chromatin, this significantly expands the understanding of the principles of epigenetic regulation of genes and allows potential regulatory elements to be searched for *in silico*. However, common experimental approaches used to study the local characteristics of chromatin have a number of technical limitations that may reduce the reliability of computational identification of genomic regulatory sequences. Taking into account the

variability of the functions of epigenetic determinants and complex multicomponent regulation of genomic elements activity, their functional verification is often required. A plethora of methods have been developed to study the functional role of regulatory elements on the genome scale. Common experimental approaches for *in silico* identification of regulatory elements and their inherent technical limitations will be described. The present review is focused on original high-throughput methods of enhancer activity reporter analysis that are currently used to validate predicted regulatory elements and to perform *de novo* searches. The methods described allow assessing the functional role of the nucleotide sequence of a regulatory element, to determine its exact boundaries and to assess the influence of the local state of chromatin on the activity of enhancers and gene expression. These approaches have contributed substantially to the understanding of the fundamental principles of gene regulation.

Key words: gene regulatory elements; enhancers; massive parallel assays.

For citation: Romanov S.E., Kalashnikova D.A., Laktionov P.P. Methods of massive parallel reporter assays for investigation of enhancers. *Vavilovskii Zhurnal Genetiki i Selektcii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):344-355. DOI 10.18699/VJ21.038

Введение

Корректная реализация программ развития и поддержания функций организма основана на экспрессии специфических для клеток и тканей наборов генов. Координация экспрессии генов осуществляется многоуровневой системой регуляции, включающей генетические и эпигенетические механизмы, основанные на взаимодействии геномных последовательностей, эпигенетических модификаций, регуляторных белков и специфических транскрипционных факторов. Определенные участки генома, ассоциированные с характерными эпигенетическими детерминантами, а также служащие сайтами привлечения регуляторных белков, способны модифицировать экспрессию генов. Такие регуляторные элементы генома играют ключевую роль в реализации генетических программ развития, дифференцировки, поддержания клеточного и тканевого гомеостаза (Phillips-Cremmins, Corces, 2013; Andersson et al., 2014; Kundaje et al., 2015).

Нарушения функции регуляторных геномных элементов могут приводить к развитию различных патологий, включая онкологические заболевания, пороки развития и аутоиммунные заболевания (Maurano et al., 2012; Corradin et al., 2014; Miguel-Escalada et al., 2015; Bradner et al., 2017; Chatterjee, Ahituv, 2017). В исследованиях по полногеномному поиску ассоциаций показано, что более 90 % ассоциированных с развитием заболеваний однонуклеотидных полиморфизмов располагается в некодирующих районах генома (Manolio et al., 2009; Maurano et al., 2012). Причем значительная часть таких геномных вариантов находится в областях, демонстрирующих эпигенетические характеристики энхансеров, а также затрагивает энхансеры, специфичные для клеточных линий, задействованных в патогенезе заболевания (Ernst et al., 2011; Akhtar-Zaidi et al., 2012; Trynka et al., 2013). В качестве примера можно привести генетические варианты, ассоциированные с развитием диабета второго типа, локализованные в районах потенциальных энхансеров, специфичных для клеток островков Лангерганса (Stitzel et al., 2010; Pasquali et al., 2014).

На сегодняшний день накоплено значительное количество информации о характерных эпигенетических свойствах регуляторных элементов, позволяющих идентифицировать потенциальные регуляторные участки генома *in silico* (Ernst et al., 2011). Однако валидация и функциональное описание регуляторных элементов часто требуют прямой экспериментальной проверки. Классическим

способом ее проведения являются различные вариации репортерного анализа и функционального мутагенеза. С развитием методов массового параллельного секвенирования были разработаны методики исследования, позволяющие изучать активность регуляторных элементов в масштабе генома.

В представленном обзоре будут рассмотрены существующие методологические решения в области высокопроизводительного анализа энхансеров, которые внесли значительный вклад в понимание фундаментальных принципов их работы.

Типы регуляторных элементов

Выделяют несколько типов регуляторных элементов генома: промоторы, энхансеры, инсультаторы и сайленсеры.

Промоторы располагаются в районе точки старта транскрипции и у эукариот представляют собой участок ДНК, на котором происходит сборка транскрипционного комплекса, состоящего из главных факторов транскрипции, РНК-полимеразы и прочих регуляторных белков, в том числе опосредующих взаимодействие с энхансерами (Andersson, Sandelin, 2020).

Энхансеры – это участки ДНК, содержащие сайты связывания транскрипционных факторов и кофакторов. В составе белкового комплекса они способны физически взаимодействовать с промотором гена и активировать его экспрессию (Shlyueva et al., 2014). Энхансеры могут оказывать влияние на целевые промоторы, находясь на значительном удалении и вне зависимости от взаимной пространственной ориентации (Pennacchio et al., 2013). Так, например, энхансер ZRS, доминантная мутация которого приводит к наследственным формам полидактилии, в геноме мыши находится приблизительно в 1 млн п. о. от контролируемого им гена *Sonic hedgehog* (*Shh*) (Lettice et al., 2014). В среднем энхансеры удалены от целевых генов на 20–50 тыс. п. о. в геномах позвоночных и на 4–10 тыс. п. о. – в геноме плодовой мушки (Furlong, Levine, 2018).

Регуляторная сеть взаимодействий энхансеров с промоторами может быть достаточно сложной. Отдельные гены имеют общие энхансеры с другими генами, подвергают регуляции несколькими энхансерами и специфичными энхансерами в разных типах клеток организма. Так, экспрессия гена *Arx* контролируется четырьмя энхансерами в ткани мозга мыши (Dickel et al., 2018). Регуляция гена специфичными энхансерами наблюдается и при разви-

тии патологий, например энхансер протоонкогена *Мус* располагается в районе сайта терминации транскрипции при раке поджелудочной железы, при раке прямой кишки обнаруживается со стороны 5'-области гена, а в случае Т-клеточного острого лимфобластного лейкоза – ниже 3'-области гена (Sur, Taipale, 2016).

В исследованиях на плодовой мушке было показано, что до 30 % энхансеров способны выступать в качестве дистанционных регуляторных элементов, не затрагивая при этом экспрессию генов, расположенных между ними и целевыми генами (Ghavi-Helm et al., 2014; Kvon et al., 2014). Это означает, что должны существовать тонко настроенные регуляторные механизмы, обеспечивающие адресное взаимодействие промотора целевого гена и специфического энхансера. Выделяют несколько функционально пересекающихся концепций, описывающих механику организации промотор-энхансерных взаимодействий, основными из которых являются установление контактов посредством белковых гомо-олигомеров и формирование петель хроматина, в том числе под действием моторных белков, таких как РНК-полимераза II и когезин.

В регуляции пространственной структуры хроматина важную роль играют регуляторные элементы – инсуляторы. Взаимодействуя со специфическими белками, инсуляторы способны блокировать энхансер-промоторные взаимодействия и предотвращать распространение репрессивных меток хроматина, выступая в качестве барьерных элементов (Kellum, Schedl, 1991, 1992; Geyer, Corces, 1992; Cai, Levine, 1995). С развитием современных методов анализа пространственной структуры ядра стало очевидным, что функциональное влияние инсуляторов во многом определяется их участием в регуляции внутри- и межхромосомных контактов (Yang, Corces, 2011). Инсуляторные белки играют ключевую роль в формировании обособленных друг от друга участков генома протяженностью порядка сотен тысяч пар нуклеотидов – топологически ассоциированных доменов (ТАД) (Dixon et al., 2012). Такие участки характеризуются высокой частотой внутренних ДНК-контактов и нередко ограничены сайтами связывания инсуляторных белков, а также активно транскрибируемыми генами (Phillips-Cremins et al., 2013; Rao et al., 2014). Наряду с регуляцией пространственной структуры ядра, инсуляторы вовлечены во множество регуляторных процессов, среди которых активация и репрессия генной экспрессии, альтернативный сплайсинг, паузирование РНК-полимеразы (Shukla et al., 2011; Paredes et al., 2013; Phillips-Cremins, Corces, 2013).

Функция сайленсеров заключается в подавлении экспрессии генов, такая репрессия в основном осуществляется посредством установления репрессивных хроматиновых меток и конкуренции с активирующими белками (Li et al., 2004; Srinivasan, Atchison, 2004; Harris et al., 2005; Lanzuolo et al., 2007; Tiwari et al., 2008).

Идентификация регуляторных элементов генома

Развитие современных методов высокопроизводительного анализа значительно ускорило и упростило поиск потенциальных регуляторных элементов. Предположения о возможной регуляторной роли участка генома обычно

строятся на нескольких типах данных: (1) оценка доступности ДНК для регуляторных белков, (2) наличие характерных эпигенетических детерминант, (3) оценка генной экспрессии и (4) анализ пространственной структуры локуса.

Определенные участки активных регуляторных элементов связаны со специфическими белками, в результате чего, как правило, свободны от нуклеосом. Для идентификации локусов, свободных от нуклеосом и соответствующих потенциальным регуляторным элементам, используются методы оценки доступности ДНК для ДНКазы I (DNase-seq), микрококковой нуклеазы (MNase-seq) и транспозазы Tn5 (assay for transposase-accessible chromatin, ATAC-seq) с последующим анализом высокопроизводительным секвенированием, а также метод FAIRE-seq (Nagy et al., 2003; Gaulton et al., 2010; Song, Crawford, 2010; Buenrostro et al., 2013). Эти методы способны идентифицировать предполагаемые энхансеры, инсуляторы и сайленсеры, но не позволяют напрямую оценить, к какому функциональному классу принадлежит обнаруженный регуляторный элемент (Song et al., 2011; Murtha et al., 2014; Huang et al., 2019).

Геномное картирование специфичных факторов хроматина и гистоновых модификаций также применяется для идентификации отдельных классов регуляторных элементов. Основной метод оценки представленности таких эпигенетических детерминант в том или ином районе генома – это метод иммунопреципитации хроматина с последующим массовым параллельным секвенированием (ChIP-seq). Характерной гистоновой модификацией промоторов является гистон H3, трижды метилированный по остатку лизина в 4-й позиции (H3K4me3) (Bernstein et al., 2005). Монометилирование по этой же позиции гистона H3 ассоциировано с энхансерами, причем одновременное наличие модификации H3K27me3 свидетельствует о подготовленном к активации состоянии энхансера, в то время как наличие H3K27ac – об активном состоянии (Heintzman et al., 2007; Creighton et al., 2010; Rada-Iglesias et al., 2011; Bonn et al., 2012; Arnold et al., 2013). Для энхансеров свойственно обогащение гистоновой ацетилтрансферазой p300 (Visel et al., 2009). Картирование специфических транскрипционных факторов также используется для идентификации энхансеров. Так, например, области ДНК, обогащенные гистоновыми метками активных энхансеров, белками комплекса Mediator и мастер-регуляторами Oct4, Sox2, Nanog, Klf4 и Esrrb, представляют так называемые супер-энхансеры, контролирующие экспрессию тканеспецифичных наборов генов в эмбриональных стволовых клетках (Whyte et al., 2013). Для идентификации инсуляторов у позвоночных анализируют геномное распределение белка CTCF и кофакторов, задействованных в формировании петель, таких как Rad21 и YY (Dixon et al., 2012, 2015; Nora et al., 2017; Rao et al., 2017). Сайленсеры же, как правило, обогащены гистоновой модификацией H3K27me3, ассоциированной с действием белков-репрессоров группы Polycomb, а также модификациями H3K9me2/3, характерными для гетерохроматина (Barski et al., 2007).

Как уже упоминалось ранее, пространственная организация ядра опосредует взаимодействие целевых локусов

генома и дистальных регуляторных элементов. Пространственную структуру хроматина изучают методами фиксации ДНК-ДНК-контактов, берущими начало от метода 3C (chromosome conformation capture) (Dekker et al., 2002; Tolhuis et al., 2002). Наиболее широко используемый метод HiC позволяет определять ДНК-контакты в масштабах генома и не требует исходных предположений о конкретных контактирующих фрагментах, в отличие от более ранних методов (Gavrilov et al., 2009; Lieberman-Aiden et al., 2009). Комбинации методов анализа пространственной структуры хроматина и иммунопреципитации хроматина (ChIA-PET, HiChIP и PLAC-ChIP) дают возможность устанавливать ДНК-контакты в областях генома, специфически обогащенных конкретными хроматиновыми белками или модификациями гистонов (Fullwood, Ruan, 2009; Fang et al., 2016; Mumbach et al., 2016). Анализ ДНК-ДНК-контактов позволяет идентифицировать промотор-энхансерные взаимодействия, выделять ТАД и более крупные компартменты хроматина.

Полученные в ходе исследовательских работ данные об эпигенетических характеристиках и пространственной организации геномов модельных объектов доступны широкому кругу исследователей и депонируются в базах данных ENCODE, The Epigenome Roadmap, FANTOM и прочих (Birney et al., 2007; Bernstein et al., 2010; Andersson et al., 2014; Forrest et al., 2014; Kellis et al., 2014; Kundaje et al., 2015). Информация из таких открытых источников активно используется для предсказания и поиска потенциальных регуляторных элементов.

Однако необходимо учитывать, что методики анализа взаимодействий белок-ДНК и ДНК-ДНК способны детектировать нефункциональные взаимодействия, это может приводить к получению ложнопозитивных результатов. Локальное обогащение характерными эпигенетическими детерминантами, обнаруженное методом ChIP-seq, не обязательно свидетельствует о наличии регуляторного элемента в данном районе генома (Kvon et al., 2012). Это связано с тем, что для реализации функции регуляторного элемента часто требуется координированное связывание ряда транскрипционных факторов, и наличия только одного из них просто недостаточно (Halfon et al., 2000; Sandmann et al., 2007).

Нефункциональное связывание транскрипционного фактора может быть транзитным и обусловленным общей ДНК-связывающей активностью (Hammar et al., 2012). Метод иммунопреципитации хроматина детектирует такие транзитные взаимодействия, поскольку основан на фиксации хроматина формальдегидом с формированием ковалентных сшивок между ДНК и ассоциированными белками. Некоторое время назад были предложены модификации метода, которые исключают необходимость фиксации хроматина и потенциально улучшают точность метода (Skene, Henikoff, 2017; Kaya-Okur et al., 2019). В вариации метода CUT&RUN используется микрококковая нуклеаза, слитая с белком А (Skene, Henikoff, 2017). Белок А связывается со специфическими антителами на исследуемый белок, а микрококковая нуклеаза вносит разрывы ДНК в области его связывания. Это позволяет выделить короткие геномные фрагменты, обогащенные исследуемым белком, и проанализировать их представлен-

ность высокопроизводительным секвенированием. В методе CUT&TAG вместо нуклеазы используется транспозаза Tn5, что дает возможность одновременно внедрить ДНК-адаптеры для проведения массового параллельного секвенирования, которые будут фланкировать сайт узнавания исследуемого белка (Kaya-Okur et al., 2019). Указанные методы были разработаны сравнительно недавно и пока не получили широкого применения.

Появление ложнопозитивных результатов в экспериментах ChIP-seq также может быть связано с экспериментальными вариациями, такими как режим фрагментации хроматина, глубина секвенирования и величина пороговых значений для идентификации сайтов связывания (Rye et al., 2011; Gomes et al., 2014; Jung et al., 2014). Важно также отметить, что при наличии высоко- и низкоафинных сайтов связывания белков метод ChIP-seq детектирует преимущественно все же высокоафинные (Nettling et al., 2016). Такая особенность также является ограничением метода, поскольку появляется все больше свидетельств в пользу того, что неоптимальные сайты связывания транскрипционных факторов в составе энхансеров необходимы для тонкой регуляции активности генов в ходе развития (Crocker et al., 2015, 2016; Farley et al., 2015).

Помимо технических ограничений методов анализа, важно отметить, что нередко функциональные регуляторные элементы демонстрируют наличие эпигенетических детерминант, в целом не характерных для их класса. В качестве примера можно привести экспериментально обнаруженные сайленсеры в культурах клеток K562 и HepG2B, которые, согласно данным базы ENCODE, помимо обогащения репрессивными гистоновыми модификациями H3K9me3 и H3K27me3, содержали также гистоновые метки активного хроматина H3K36me3 и H3K79me2 (Pang, Snyder, 2020). Ввиду экспериментальных ограничений методов, вариабельности функций эпигенетических детерминант и участия множества компонентов в реализации функций элементов генома определение их регуляторной роли часто требует той или иной функциональной проверки.

Методы исследования энхансеров

Основными методами оценки функциональной роли регуляторных элементов генома являются вариации репортерного анализа. Одной из первых работ, в которой была продемонстрирована функциональная роль регуляторных элементов генома, была посвящена исследованию энхансера раннего гена вируса SV40 (Banerji et al., 1981). В ней показано, что фрагмент ДНК из двух повторов протяженностью 72 п. н., расположенный в районе 5'-конца раннего гена вируса SV40, в составе генетической конструкции способен более чем в 200 раз активировать экспрессию гена β-глобина кролика в клетках HeLa (Banerji et al., 1981).

Стандартные генетические конструкции, используемые для анализа активности энхансеров, содержат репортерный ген под контролем минимального промотора, который не способен обеспечить экспрессию без дополнительного активирующего воздействия. В состав конструкции вводят геномную последовательность энхансера, причем она может располагаться как перед промотором, так и за

кодирующей последовательностью репортерного гена. Полученной конструкцией трансформируют клетки и оценивают изменение экспрессии репортерного гена относительно контрольной конструкции, в которой отсутствует потенциальный энхансер.

Одна из первых работ, направленная на функциональное тестирование энхансеров в масштабе всего генома *in vivo*, была основана на принципах классического репортерного анализа (Kvon et al., 2014). Для исследования было использовано около 8000 линий *Drosophila melanogaster*, содержащих в одном и том же участке генома трансгенную конструкцию, состоящую из последовательно расположенных: потенциального энхансера, минимального промотора и гена белка Gal4. Методом *in situ* гибридизации оценивалась экспрессия Gal4 на разных стадиях эмбриогенеза, причем для каждого потенциального регуляторного элемента было проанализировано по 400 эмбрионов, находящихся на разных стадиях развития. Было идентифицировано более трех тыс. энхансеров. Около четверти из них были расположены вблизи регулируемых генов и чуть более четверти – на удалении в 20–100 т. п. н., в среднем же они находились на расстоянии в 10 т. п. н. от генов-мишеней (Kvon et al., 2014). Ориентировочно треть обнаруженных энхансеров располагалась в межгенных областях регулируемых генов. Впоследствии также было функционально подтверждено, что энхансеры способны регулировать не только близлежащие, но и расположенные через один либо два гены (Kvon et al., 2014). Полученные данные значительно расширили представления о фундаментальных принципах функционирования энхансеров, однако выполнение подобных проектов требует колоссальных затрат времени и ресурсов.

Методы высокопроизводительного репортерного анализа энхансеров

Методы высокопроизводительного репортерного анализа пришли на смену классическим подходам и позволяют исследовать тысячи регуляторных последовательностей в одном эксперименте. Можно выделить два принципиальных подхода для проведения таких исследований (рис. 1). В рамках первого в составе репортерных конструкций под контролем фрагмента генома – потенциального энхансера и минимального промотора – располагается репортерный ген, содержащий перед сигналом полиаденилирования вырожденную короткую нуклеотидную последовательность – ДНК-баркод (см. рис. 1, а). В случае активации экспрессии репортерного гена, ДНК-баркод будет содержаться на 3'-конце его транскрипта. Высокопроизводительным секвенированием полученного пула конструкций определяют уникальные ДНК-баркоды, соответствующие каждому из исследуемых фрагментов генома (рис. 2). После проведения трансгенеза с использованием таких конструкций выполняется анализ представленности ДНК-баркодов методом высокопроизводительного анализа транскриптома (RNA-seq). Уровень экспрессии того или иного ДНК-баркода позволяет сделать вывод об активирующей способности соответствующего ему конкретного регуляторного элемента. Данный подход лежит в основе методов количественной оценки активности фрагментов

генома, объединенных термином MPRA (massive parallel reporter assay), вариации которых будут рассмотрены в представленном обзоре (Kwasnieski et al., 2012, 2014; Melnikov et al., 2012; Kheradpour et al., 2013; Maricque et al., 2017).

Второй принципиальный подход позволяет оценить качественную способность фрагмента генома проявлять свойства энхансеров. На первом этапе подготавливается пул генетических конструкций, последовательно содержащих исследуемый фрагмент генома, минимальный промотор и репортерный ген, кодирующий флуоресцентный белок либо люциферазу. На следующем этапе проводится трансгенез полученным пулом конструкций и при помощи проточной цитометрии отбираются клетки, экспрессирующие флуоресцентный белок. Активация экспрессии репортерного гена означает, что фрагмент генома, содержащийся в составе конструкции, способен демонстрировать свойства энхансера. Из полученных клеток выделяют ДНК, амплифицируют фрагменты конструкций, соответствующих исследуемым фрагментам генома, и осуществляют массовое параллельное секвенирование для идентификации конкретных фрагментов генома, обнаруживающих свойства энхансеров. К таким методам можно отнести FIREWACH и SIF-seq (Dickel et al., 2014; Murtha et al., 2014) (см. рис. 1, б, в).

В настоящее время используют и комбинации двух подходов, описанных выше, когда на первом этапе методами проточной цитометрии отбирают клетки, несущие конструкции, содержащие потенциальные энхансеры. Затем количественно оценивают активирующую способность конкретных фрагментов генома посредством анализа представленности ДНК-баркодов методом RNA-seq (Maricque et al., 2018).

Методы MPRA-анализа успешно применяются для исследования активирующих свойств нуклеотидной последовательности энхансеров, функционального влияния мотивов связывания регуляторных белков, а также для поиска и валидации энхансеров. С помощью такой методологии изучали влияние одиночных мутаций в составе трех энхансеров, активных в клетках печени, – ALDOB, ECR11 и LTV1 (Patwardhan et al., 2012). Для этого была синтезирована ДНК-библиотека, содержащая более 100000 мутированных вариантов исследуемых энхансеров. Такие фрагменты были клонированы в конструкции, содержащие минимальный промотор, ген люциферазы и транскрибируемые ДНК-баркоды. Полученные ДНК-библиотеки вводили в печень мышей, и спустя сутки транскриптом клеток печени был проанализирован методом RNA-seq (Kim, Ahituv, 2013).

В результате анализа установлено, что большинство одиночных мутаций оказывали слабый эффект на активность исследуемых энхансеров. Кроме того, показано, что нарушающие функцию энхансера мутации затрагивают предсказанные биоинформатическими методами сайты связывания активных в клетках печени транскрипционных факторов HNF4 и HNF1 (Kel et al., 2003). Важно отметить, что эксперимент также продемонстрировал серьезные расхождения в теории и практике. Так, в составе энхансера ECR11 мутации, вызывающие функциональные нарушения, были сконцентрированы в районе, не содержа-

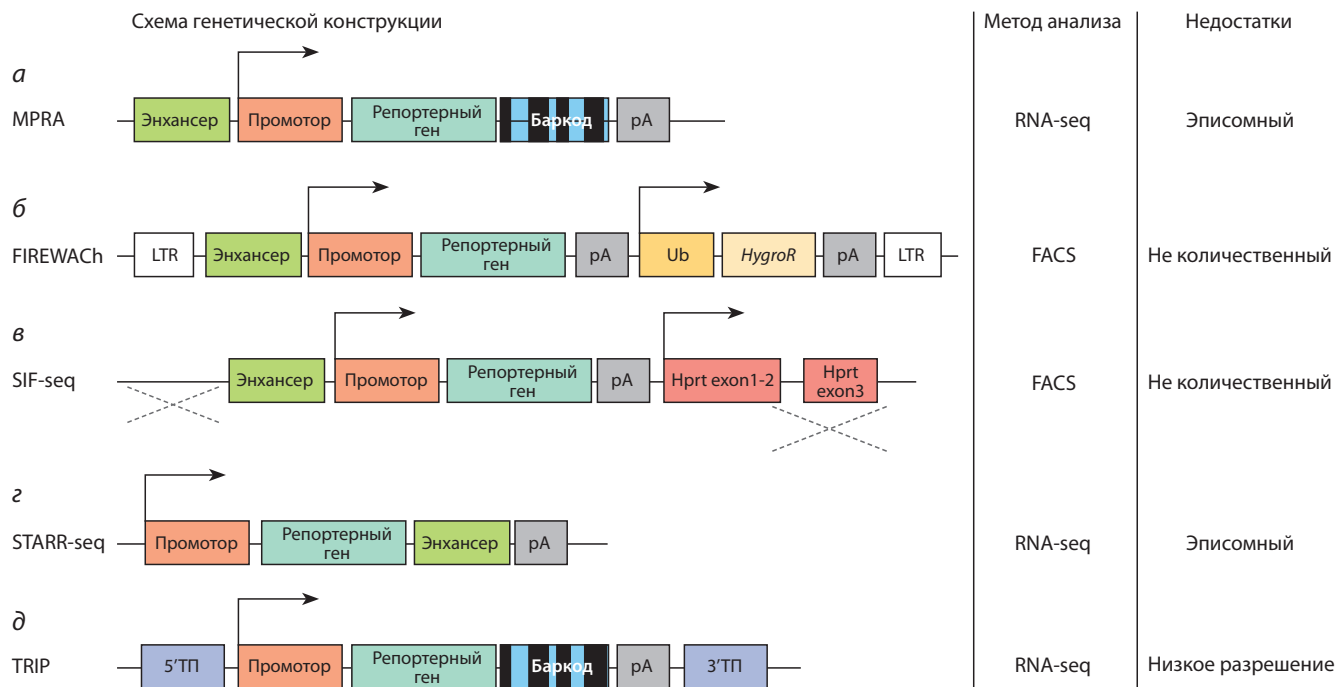


Рис. 1. Схема генетических конструкций, используемых для проведения MPRA-анализа (а–г) и метода TRIP.

pA – сигнал полиадеилирования; LTR – длинный концевой повтор; Ub – промотор убиквитина; *HygroR* – ген резистентности к гигромицину; Hprt exon – экзон гена *Hprt*.

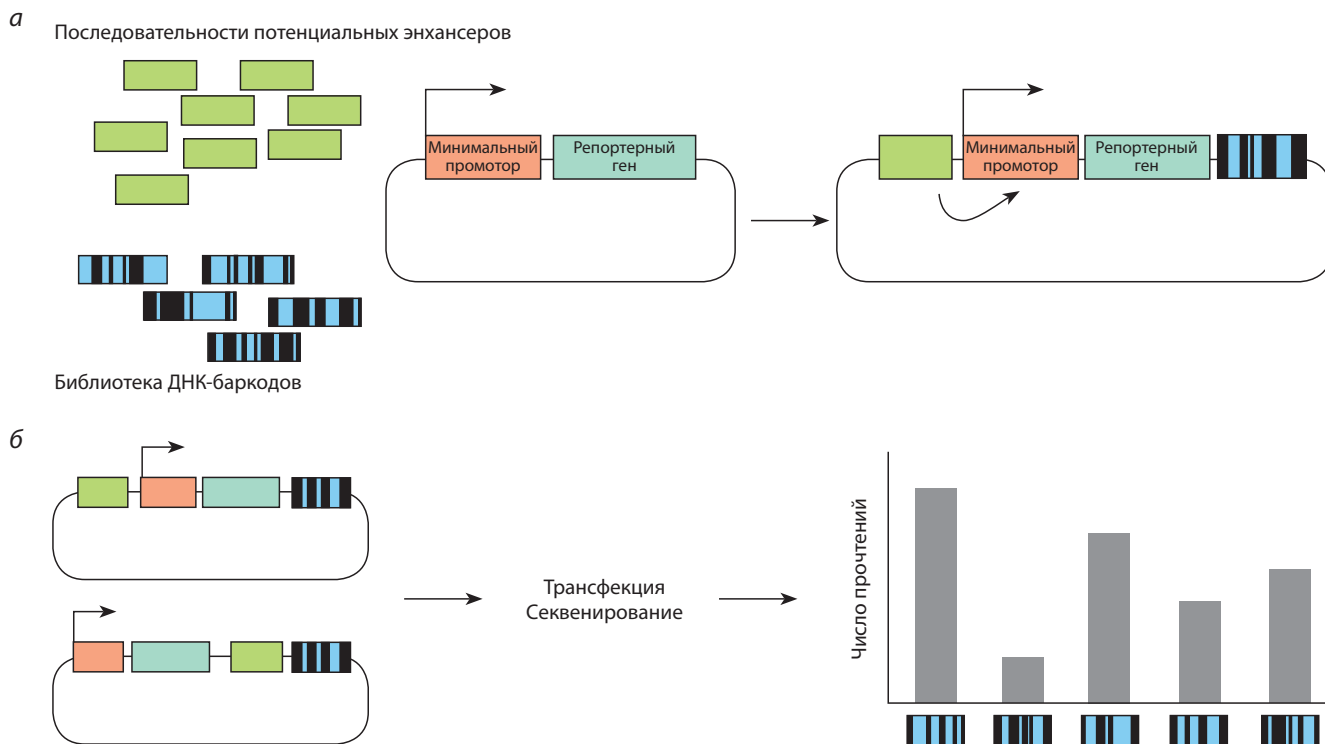


Рис. 2. Принцип работы методов MPRA-анализа.

а – на первом этапе MPRA-анализа готовят пул потенциальных регуляторных последовательностей. Для получения таких последовательностей используют технологии синтеза либо обогащение методами иммунопреципитации хроматина и пр. Затем создают пул генетических конструкций, содержащих минимальный промотор и репортерный ген под контролем исследуемых регуляторных элементов. Каждому регуляторному элементу в составе таких конструкций соответствует уникальный ДНК-баркод, располагающийся на конце кодирующей последовательности репортерного гена; **б** – после проведения трансфекции клеток полученными конструкциями, в случае если исследуемый регуляторный элемент проявляет свойства энхансера, происходит активация экспрессии репортерного гена. Методом RNA-seq оценивается уровень экспрессии уникальных ДНК-баркодов в транскриптоме клетки. Нормализация на представленность ДНК-баркодов в исходном пуле генетических конструкций и установление соответствия уникальных ДНК-баркодов конкретным ДНК-фрагментам позволяют сделать вывод о наличии регуляторной функции последних.

щем предсказанных сайтов связывания транскрипционных факторов, мутации же в области, имеющей большинство таких предсказанных сайтов, не приводили к изменению активности энхансера. С одной стороны, это наглядно демонстрирует, что MPRA-анализ применим для уточнения границ энхансеров, а с другой, подчеркивает важность экспериментальной проверки предиктивных данных.

Методологию MPRA также используют для *de novo* поиска и валидации предсказанных энхансеров. Элегантный подход поиска энхансеров был реализован в методе STARR-seq (Arnold et al., 2013) (см. рис. 1, з). Авторы использовали способность энхансеров активировать экспрессию вне зависимости от положения относительно гена и разработали репортерные конструкции, в составе которых за минимальным промотором следовала открытая рамка считывания, в которую клонировали исследуемый фрагмент генома. После трансформации клеток такими конструкциями, в случае если участок генома проявляет энхансерную функцию, это будет приводить к его транскрипции. Уровень же представленности этого фрагмента в транскриптоме клетки позволяет сделать вывод о выраженности активирующей функции. Такой подход полностью исключает необходимость применения ДНК-баркодов, поскольку в их роли выступают сами исследуемые фрагменты.

Для проведения эксперимента была подготовлена библиотека плазмид, содержащих миллионы случайных фрагментов генома плодовой мушки. После трансфекции культуры клеток S2 был проведен высокопроизводительный анализ профиля транскрипции методом RNA-seq. Были идентифицированы тысячи геномных фрагментов, демонстрирующих свойства энхансеров. Наиболее активные из них находились вблизи генов домашнего хозяйства и транскрипционных факторов, контролирующих развитие. Показано также, что около трети фрагментов, проявляющих ярко выраженные активирующие свойства, в геноме клеток S2 располагались в областях закрытого хроматина, в котором отсутствовала модификация активных энхансеров H3K27ac. Таким образом, маловероятно, что в геноме исследуемых клеток эти фрагменты способны выполнять роль энхансеров, и эта находка, скорее, подчеркивает некоторые ограничения эпизодических методов MPRA, которые будут рассмотрены ниже.

Интересная модификация метода STARR-seq была использована в работе по исследованию энхансеров в эмбриональных стволовых клетках (ЭСК) человека (Barakat et al., 2018). В оригинальной работе (Arnold et al., 2013) ДНК-библиотеки были созданы путем ультразвуковой дезинтеграции геномной ДНК *D. melanogaster* и массового клонирования полученных фрагментов. Однако такой подход плохо применим для более крупных геномов, поскольку достижение достаточной представленности регуляторных элементов в итоговых ДНК-библиотеках – крайне затруднительная задача. Действительно, согласно расчетным данным, применение оригинального метода STARR-seq для исследования регуляторных элементов генома мыши потребует создания более 200 млн уникальных конструкций (Murtha et al., 2014). Экспериментальная проверка показала, что использование плазмидной библиотеки, содержащей 1.3 млн уникальных фрагмен-

тов генома человека, позволило идентифицировать лишь шесть энхансеров (Murtha et al., 2014).

Для преодоления этого ограничения был предложен метод ChIP-STARR-seq, в рамках которого на первом этапе методом иммунопреципитации хроматина проводили выделение фрагментов генома, обогащенных транскрипционными факторами OCT4, NANOG, а также гистоновыми модификациями H3K4me1 и H3K27ac (Barakat et al., 2018). Полученные ДНК-фрагменты затем были клонированы в состав ДНК-библиотек, аналогичных используемым в оригинальном методе. Было установлено, что только часть фрагментов генома, демонстрирующих обогащение указанными факторами, проявляла энхансерную активность. Лишь около 25 % фрагментов, связанных OCT4, демонстрировали свойства энхансеров. Для фрагментов, обогащенных NANOG и гистоновыми модификациями H3K4me1 и H3K27ac, эта доля составила 15, 9 и 10 % соответственно. Показано, что ни отдельные факторы, ни их комбинации не способны однозначно предсказать энхансеры. Кроме того, выявлена группа энхансеров, ассоциированных с регуляцией общих клеточных процессов, которые ранее не были обнаружены в эмбриональных стволовых клетках. Оказалось, что такие энхансеры демонстрируют достаточно слабое обогащение ТФ OCT4 и NANOG, а также гистоновой модификацией H3K4me1, и, вероятно, по этой причине раньше их не детектировали в поисковых работах на основе метода иммунопреципитации хроматина.

Депонированная в базах данных информация о геномном распределении гистоновых модификаций и регуляторных белков, а также доступности хроматина дает возможность предполагать, какие участки генома способны проявлять свойства энхансеров. С применением MPRA-методологии исследована активность регуляторных элементов клеток K562 и эмбриональных стволовых клеток человека E1, идентифицированных на основе анализа структуры хроматина и аннотированных в базе ENCODE (Kwasniewski et al., 2014). Оказалось, что лишь около четверти из них оказывали влияние на экспрессию генов, что подчеркивает важность подобной экспериментальной проверки (Kwasniewski et al., 2014). Вместе с тем подобный эффект может быть обусловлен экспериментальными ограничениями описанных методов MPRA. Например, тем фактом, что описанные выше подходы являются эпизодическими – репортерные конструкции не встраиваются в геном, а значит, активность энхансеров оценивается вне хроматинового контекста. Значительные отличия в активности энхансеров в составе плазмид и при интеграции в геном были подтверждены и экспериментально (Inoue et al., 2017).

Такое экспериментальное расхождение выглядит логичным, ведь наблюдение о влиянии структуры хроматина на регуляцию генов было продемонстрировано в классических генетических экспериментах еще в начале прошлого века (Muller, 1930). Применение оригинального высокопроизводительного метода репортерного анализа (метод TRIP), пересекающегося с MPRA-подходами, позволило детализировать локальные эффекты хроматина на экспрессию генов в эмбриональных стволовых клетках мыши (Akhtar et al., 2013) (рис. 1, д). В этом исследова-

нии с использованием системы геномной интеграции на основе транспозазы PiggyBac в геном клеток были случайным образом встроены репортерные конструкции, содержащие на 3'-конце репортерного гена уникальные ДНК-баркоды.

На следующем этапе такие встройки были прокартированы, и каждый ДНК-баркод ассоциирован с конкретным геномным локусом. Суммарно было получено более 17 тыс. таких встроек. Далее при помощи метода RNA-seq были проанализированы представленности ДНК-баркодов, что позволило оценить транскрипционную активность каждой встройки и влияние на нее локальной структуры хроматина. Репортерные конструкции, интегрированные в районы компактизованного хроматина и области доменов, ассоциированных с ядерной ламиной, ожидаемо демонстрировали сниженный уровень экспрессии. Репортерные же конструкции, располагающиеся в пределах 200 т. п. н. от активных генов, также транскрибировались более активно. Интересно, что приблизительно в таких же границах наблюдалась повышенная встречаемость энхансеров. Энхансеры оказывали активирующее влияние на экспрессию репортерных конструкций на удалении вплоть до 20 т. п. н. Важно отметить, что в данном случае речь идет о линейном расстоянии и не учитывается пространственная структура хроматина. Предполагается, что формирование протяженных активно транскрибирующихся участков обусловлено воздействием нескольких энхансеров. Это дополнительно подчеркивает необходимость исследования регуляторных элементов в условиях, приближенных к нативным.

Влияние хроматина на функцию регуляторных элементов в той или иной степени учитывается в методах MPRA-анализа, подразумевающих геномную интеграцию репортерной конструкции (Dickel et al., 2014; Murtha et al., 2014; Maricque et al., 2017, 2018). Методы FIREWACH и SIF-seq были использованы для идентификации энхансеров в ЭСК мыши, однако не позволяли оценивать активность регуляторных элементов количественно (см. общее описание подходов выше) (Dickel et al., 2014; Murtha et al., 2014).

Метод FIREWACH основан на геномной интеграции репортерных конструкций при помощи лентивирусной трансдукции (см. рис. 1, б). Такой метод геномной интеграции, хотя и демонстрирует некоторые закономерности, все же обеспечивает встройки конструкции в случайные области генома (Yang et al., 2008). Следовательно, адекватное сравнение активности различных регуляторных элементов затруднительно, поскольку с высокой долей вероятности репортерные конструкции будут встраиваться в различные участки генома с непредсказуемым влиянием локального хроматинового окружения.

Метод SIF-seq лишен этого недостатка, поскольку интеграция репортерных конструкций осуществляется в один и тот же участок генома, расположенный в районе гена *Hprt* (Dickel et al., 2014) (см. рис. 1, в). Однако в этом случае все тестируемые регуляторные элементы потенциально оказываются в одинаковом хроматиновом окружении. Поскольку корректная работа энхансера определяется специфическим набором хроматиновых

факторов, высока вероятность того, что он станет нефункционален, будучи перенесенным в нетождественное хроматиновое окружение.

Рассмотренные выше подходы не позволяли ответить на один из основополагающих вопросов в области понимания принципов работы энхансеров, а именно, в какой степени их активность определяется последовательностью ДНК, а в какой – свойствами окружающего хроматина? Систематическое исследование этого вопроса проведено в работе, в которой было изучено влияние различного хроматинового окружения на сравнительную активность энхансеров (Maricque et al., 2018). В этом исследовании были подготовлены 15 линий клеток K562, содержащих единичные встройки репортерных конструкций, располагающиеся в различном хроматиновом окружении и содержащие сайты распознавания Cre-рекомбиназы (*loxP*), позволяющие направленно встраивать трансгены. Важно отметить, что в составе таких встроек за пределами фланкированного *loxP*-сайтами фрагмента находились ДНК-баркод и сигнал полиаденилирования, причем каждой линии соответствовал единственный уникальный ДНК-баркод.

Описанные линии были объединены в общий пул, с использованием которого проводили Cre-опосредованную направленную интеграцию конструкций, содержащих репортерный ген, оканчивающийся транскрибируемым вырожденным ДНК-баркодом, под контролем минимального промотора и исследуемого фрагмента генома. В качестве таких фрагментов были 300 искусственно синтезированных регуляторных элементов, которые ранее были исследованы методом эписомного MPRA-анализа и ранжированы по уровню активности (Kwasnieski et al., 2014). Для каждого геномного фрагмента предварительно были установлены соответствующие им уникальные ДНК-баркоды. При успешной интеграции происходила замена исходной *loxP*-кассеты на репортерную конструкцию, содержащую предполагаемые энхансеры. Причем в случае активации репортерного гена в его составе будут транскрибироваться два ДНК-баркода, один из которых позволяет установить исследуемый регуляторный элемент, а второй – определить, в клетки какой линии произошла встройка, и, следовательно, выяснить, в каком хроматиновом окружении находится данная репортерная конструкция. Анализ представленности комбинаций ДНК-баркодов в транскриптоме клеток дал возможность оценить уровень активности исследуемых регуляторных элементов в различном хроматиновом окружении. Было установлено, что хроматиновое окружение ожидаемо оказывает серьезный эффект на активность цис-элементов. Однако в одном и том же хроматиновом окружении сравнительная активность различных регуляторных элементов сохраняется. Продемонстрировано также, что активность промотора влияет на экспрессию репортерных конструкций, но вместе с тем не меняет сравнительную активность регуляторных элементов. Полученные результаты свидетельствуют в пользу модели, согласно которой нуклеотидная последовательность энхансера определяет его общую активность, которая уже модулируется структурой хроматинового окружения.

Заклучение

Методы MPRA дают возможность детально исследовать регуляторный потенциал фрагментов генома и служат удобным инструментом для исследования влияния вариаций нуклеотидной последовательности на их функцию. Однако необходимо отметить ограничения методов, которые должны учитываться при интерпретации полученных результатов. Общий недостаток всех вариаций методов MPRA – это необходимость использования минимального промотора, который не способен активировать экспрессию репортерного гена в отсутствие энхансера, поскольку наличие базальной активности может существенно исказить результаты. Вместе с тем выбранный промотор может в значительной мере воздействовать на активность конкретного энхансера (Zabidi et al., 2015; Maricque et al., 2018). Идеальным экспериментом представляется анализ активности энхансеров в комбинации с различными промоторами, однако выполнение таких работ представляется крайне сложным и трудоемким.

Синтез фрагментов ДНК, используемых в качестве исследуемых регуляторных элементов, накладывает ограничения на общую длину такого фрагмента. Как правило, длина исследуемых фрагментов ограничена приблизительно 200 п. н., что часто усложняет анализ вклада остальных участков энхансеров, выпадающих за эти пределы (Kwasniewski et al., 2014). Основанные на эписомных конструкциях методы MPRA не учитывают возможного влияния хроматинового окружения на регуляторный элемент, поэтому могут быть использованы для исследования непосредственной активирующей способности последовательности ДНК. Методы MPRA, включающие геномную интеграцию репортерных конструкций, позволяют в некоторой степени преодолеть это ограничение. Однако случайная или сайт-специфичная интеграция все же не может обеспечить исследование активности регуляторного элемента в нативном хроматиновом окружении. Это является серьезным ограничением методов MPRA, поскольку функция регуляторного геномного элемента зависит от структуры окружающего хроматина и пространственной организации локуса.

Современные методы высокопроизводительного направленного мутагенеза CRISPR/Cas9, а также методы направленной модуляции экспрессии, использующие инактивированную форму эндонуклеазы Cas9 (dCas9), слитой с активаторными либо репрессорными белками, позволяют исследовать регуляторные элементы в нативном геномном окружении (Chavez et al., 2015; Sanjana et al., 2016; Canver et al., 2017; Li et al., 2020). При очевидных преимуществах, такие методы также не лишены потенциальных недостатков. Например, точечные замены, возникающие в результате направленного мутагенеза, могут оказаться недостаточными для нарушения функции энхансера. Кроме того, направленный мутагенез сопряжен с ошибками в распознавании целевых районов генома, потенциально усложняющими интерпретацию полученных результатов.

Важно отметить, что белок-репрессор KRAB, широко используемый для направленной инактивации энхансеров, может инициировать формирование участков гетерохроматина протяженностью 1–2 т. п. н. (Gasparini et al.,

2019). Эта особенность может снижать разрешающую способность метода и усложнить определение конкретных функциональных фрагментов энхансера, а также усилить нежелательные побочные эффекты в случае наличия ошибочных сайтов узнавания dCas9. Помимо возможных технических сложностей, в случае успешного нарушения функции энхансера фенотипические проявления могут достаточно быстро восстанавливаться за счет предположительного существования дублирующих энхансеров (Diao et al., 2016).

Технологии MPRA и высокопроизводительные методы на основе системы CRISPR/Cas9, скорее, являются взаимодополняющими для детального описания регуляторных функций исследуемых фрагментов генома. Вместе с обширными массивами накопленных данных о структуре и пространственной организации хроматина в различных клетках и тканях использование таких методов дает возможность значительно продвинуться в области понимания механизмов точной регуляции экспрессии генов в ходе развития и при различных патологиях. Все это позволяет надеяться, что в скором времени современная геномика сможет перейти от подробного функционального описания регуляторных элементов к созданию количественных биологических моделей регуляции генной экспрессии.

Список литературы / References

- Akhtar-Zaidi B., Cowper-Sal-lari R., Corradin O., Saiakhova A., Bartels C.F., Balasubramanian D., Myeroff L., Lutterbaugh J., Jarrar A., Kalady M.F., Willis J., Moore J.H., Tesar P.J., Laframboise T., Markowitz S., Lupien M., Scacheri P.C. Epigenomic enhancer profiling defines a signature of colon cancer. *Science*. 2012;336:736-739.
- Akhtar W., de Jong J., Pindyurin A.V., Pagie L., Meuleman W., de Ridder J., Berns A., Wessels L.F., van Lohuizen M., van Steensel B. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell*. 2013;154:914-927.
- Andersson R., Gebhard C., Miguel-Escalada I., Hoof I., Bornholdt J., Boyd M., Chen Y., ..., Suzuki H., Hayashizaki Y., Muller F., Forrest A.R.R., Carninci P., Rehli M., Sandelin A. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507:455-461.
- Andersson R., Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* 2020;21:71-87.
- Arnold C.D., Gerlach D., Stelzer C., Boryn L.M., Rath M., Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013;339:1074-1077.
- Banerji J., Rusconi S., Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*. 1981;27:299-308.
- Barakat T.S., Halbritter F., Zhang M., Rendeiro A.F., Perenthaler E., Bock C., Chambers I. Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell*. 2018;23:276-288 e278.
- Barski A., Cuddapah S., Cui K., Roh T.Y., Schones D.E., Wang Z., Wei G., Chepelev I., Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129:823-837.
- Bernstein B.E., Kamal M., Lindblad-Toh K., Bekiranov S., Bailey D.K., Huebert D.J., McMahon S., Karlsson E.K., Kulbokas E.J. 3rd, Gingeras T.R., Schreiber S.L., Lander E.S. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 2005;120:169-181.
- Bernstein B.E., Stamatoyannopoulos J.A., Costello J.F., Ren B., Milosavljevic A., Meissner A., Kellis M., Marra M.A., Beaudet A.L., Ecker J.R., Farnham P.J., Hirst M., Lander E.S., Mikkelsen T.S., Thomson J.A. The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* 2010;28:1045-1048.

- Birney E., Stamatoyannopoulos J.A., Dutta A., Guigo R., Gingeras T.R., Margulies E.H., Weng Z., ..., Lander E.S., Koriabine M., Nefedov M., Osoegawa K., Yoshinaga Y., Zhu B., de Jong P.J. Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. *Nature*. 2007;447:799-816.
- Bonn S., Zinzen R.P., Girardot C., Gustafson E.H., Perez-Gonzalez A., Delhomme N., Ghavi-Helm Y., Wilczynski B., Riddell A., Furlong E.E. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* 2012;44:148-156.
- Bradner J.E., Hnisz D., Young R.A. Transcriptional addiction in cancer. *Cell*. 2017;168:629-643.
- Buenrostro J.D., Giresi P.G., Zaba L.C., Chang H.Y., Greenleaf W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*. 2013;10:1213-1218.
- Cai H., Levine M. Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. *Nature*. 1995;376:533-536.
- Canver M.C., Bauer D.E., Orkin S.H. Functional interrogation of non-coding DNA through CRISPR genome editing. *Methods*. 2017; (121-122):118-129.
- Chatterjee S., Ahituv N. Gene regulatory elements, major drivers of human disease. *Annu. Rev. Genomics Hum. Genet.* 2017;18:45-63.
- Chavez A., Scheiman J., Vora S., Pruitt B.W., Tuttle M., Eswar P.R.I., Lin S., Kiani S., Guzman C.D., Wiegand D.J., Ter-Ovanesyan D., Braff J.L., Davidsohn N., Housden B.E., Perrimon N., Weiss R., Aach J., Collins J.J., Church G.M. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods*. 2015;12:326-328.
- Corradin O., Saiakhova A., Akhtar-Zaidi B., Myeroff L., Willis J., Cowper-Sallari R., Lupien M., Markowitz S., Sacheri P.C. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res*. 2014;24:1-13.
- Creyghton M.P., Cheng A.W., Welstead G.G., Kooistra T., Carey B.W., Steine E.J., Hanna J., Lodato M.A., Frampton G.M., Sharp P.A., Boyer L.A., Young R.A., Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA*. 2010;107:21931-21936.
- Crocker J., Abe N., Rinaldi L., McGregor A.P., Frankel N., Wang S., Alsawadi A., Valenti P., Plaza S., Payne F., Mann R.S., Stern D.L. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell*. 2015;160:191-203.
- Crocker J., Noon E.P., Stern D.L. The soft touch: low-affinity transcription factor binding sites in development and evolution. *Curr. Top. Dev. Biol.* 2016;117:455-469.
- Dekker J., Rippe K., Dekker M., Kleckner N. Capturing chromosome conformation. *Science*. 2002;295:1306-1311.
- Diao Y., Li B., Meng Z., Jung I., Lee A.Y., Dixon J., Maliskova L., Guan K.L., Shen Y., Ren B. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res*. 2016;26:397-405.
- Dickel D.E., Ypsilanti A.R., Pla R., Zhu Y., Barozzi I., Mannion B.J., Khin Y.S., Fukuda-Yuzawa Y., Plajzer-Frick I., Pickle C.S., Lee E.A., Harrington A.N., Pham Q.T., Garvin T.H., Kato M., Osterwalder M., Akiyama J.A., Afzal V., Rubenstein J.L.R., Pennacchio L.A., Visel A. Ultraconserved enhancers are required for normal development. *Cell*. 2018;172:491-499.e415.
- Dickel D.E., Zhu Y., Nord A.S., Wylie J.N., Akiyama J.A., Afzal V., Plajzer-Frick I., Kirkpatrick A., Gottgens B., Bruneau B.G., Visel A., Pennacchio L.A. Function-based identification of mammalian enhancers using site-specific integration. *Nat. Meth.* 2014;11:566-571.
- Dixon J.R., Jung I., Selvaraj S., Shen Y., Antosiewicz-Bourget J.E., Lee A.Y., Ye Z., Kim A., Rajagopal N., Xie W., Diao Y., Liang J., Zhao H., Lobanenkov V.V., Ecker J.R., Thomson J.A., Ren B. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518:331-336.
- Dixon J.R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., Hu M., Liu J.S., Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376-380.
- Ernst J., Kheradpour P., Mikkelsen T.S., Shores N., Ward L.D., Epstein C.B., Zhang X., Wang L., Issner R., Coyne M., Ku M., Durham T., Kellis M., Bernstein B.E. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473:43-49.
- Fang R., Yu M., Li G., Chee S., Liu T., Schmitt A.D., Ren B. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res*. 2016;26:1345-1348.
- Farley E.K., Olson K.M., Zhang W., Brandt A.J., Rokhsar D.S., Levine M.S. Suboptimization of developmental enhancers. *Science*. 2015;350:325-328.
- Forrest A.R., Kawaji H., Rehli M., Baillie J.K., de Hoon M.J., Haberland V., Lassmann T., ..., Bajic V.B., Taylor M.S., Makeev V.J., Sandelin A., Hume D.A., Carninci P., Hayashizaki Y. A promoter-level mammalian expression atlas. *Nature*. 2014;507:462-470.
- Fullwood M.J., Ruan Y. ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* 2009;107:30-39.
- Furlong E.E.M., Levine M. Developmental enhancers and chromosome topology. *Science*. 2018;361:1341-1345.
- Gasparini M., Hill A.J., McFaline-Figueroa J.L., Martin B., Kim S., Zhang M.D., Jackson D., Leith A., Schreiber J., Noble W.S., Trapnell C., Ahituv N., Shendure J. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell*. 2019;176:1516.
- Gaulton K.J., Nammo T., Pasquali L., Simon J.M., Giresi P.G., Fogarty M.P., Panhuis T.M., Mieczkowski P., Secchi A., Bosco D., Berney T., Montanya E., Mohlke K.L., Lieb J.D., Ferrer J. A map of open chromatin in human pancreatic islets. *Nat. Genet.* 2010;42:255-259.
- Gavrilov A., Eivazova E., Priozhkova I., Lipinski M., Razin S., Vasetzky Y. Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods Mol. Biol.* 2009;567:171-188.
- Geyer P.K., Corces V.G. DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. *Genes Dev*. 1992;6:1865-1873.
- Ghavi-Helm Y., Klein F.A., Pakozdi T., Ciglar L., Noordermeer D., Huber W., Furlong E.E. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*. 2014; 512:96-100.
- Gomes A.L., Abeel T., Peterson M., Azizi E., Lyubetskaya A., Carvalho L., Galagan J. Decoding ChIP-seq with a double-binding signal refines binding peaks to single-nucleotides and predicts cooperative interaction. *Genome Res*. 2014;24:1686-1697.
- Halfon M.S., Carmena A., Gisselbrecht S., Sackerson C.M., Jimenez F., Baylies M.K., Michelson A.M. Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell*. 2000;103:63-74.
- Hammar P., Leroy P., Mahmutovic A., Marklund E.G., Berg O.G., Elf J. The lac repressor displays facilitated diffusion in living cells. *Science*. 2012;336:1595-1598.
- Harris M.B., Mostecky J., Rothman P.B. Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. *J. Biol. Chem.* 2005;280:13114-13121.
- Heintzman N.D., Stuart R.K., Hon G., Fu Y., Ching C.W., Hawkins R.D., Barrera L.O., Van Calcar S., Qu C., Ching K.A., Wang W., Weng Z., Green R.D., Crawford G.E., Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 2007;39:311-318.
- Huang D., Petrykowska H.M., Miller B.F., Elnitski L., Ovcharenko I. Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome Res*. 2019;29:657-667.
- Inoue F., Kircher M., Martin B., Cooper G.M., Witten D.M., McManus M.T., Ahituv N., Shendure J. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res*. 2017;27:38-52.

- Jung Y.L., Luquette L.J., Ho J.W., Ferrari F., Tolstorukov M., Minoda A., Issner R., Epstein C.B., Karpen G.H., Kuroda M.I., Park P.J. Impact of sequencing depth in ChIP-seq experiments. *Nucleic Acids Res.* 2014;42:e74.
- Kaya-Okur H.S., Wu S.J., Codomo C.A., Pledger E.S., Bryson T.D., Henikoff J.G., Ahmad K., Henikoff S. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* 2019;10:1930.
- Kel A.E., Gossling E., Reuter I., Cheremushkin E., Kel-Margoulis O.V., Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* 2003;31:3576-3579.
- Kellis M., Wold B., Snyder M.P., Bernstein B.E., Kundaje A., Marinov G.K., Ward L.D., Birney E., Crawford G.E., Dekker J., Dunham I., Elnitski L.L., Farnham P.J., Feingold E.A., Gerstein M., Giddings M.C., Gilbert D.M., Gingeras T.R., Green E.D., Guigo R., Hubbard T., Kent J., Lieb J.D., Myers R.M., Pazin M.J., Ren B., Stamatoyannopoulos J.A., Weng Z., White K.P., Hardison R.C. Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. USA.* 2014;111:6131-6138.
- Kellum R., Schedl P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell.* 1991;64:941-950.
- Kellum R., Schedl P. A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol. Cell. Biol.* 1992;12:2424-2431.
- Kheradpour P., Ernst J., Melnikov A., Rogov P., Wang L., Zhang X., Alston J., Mikkelsen T.S., Kellis M. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.* 2013;23:800-811.
- Kim M.J., Ahituv N. The hydrodynamic tail vein assay as a tool for the study of liver promoters and enhancers. *Methods Mol. Biol.* 2013;1015:279-289.
- Kundaje A., Meuleman W., Ernst J., Bilenky M., Yen A., Heravi-Moussavi A., Kheradpour P., ..., Hirst M., Meissner A., Milosavljevic A., Ren B., Stamatoyannopoulos J.A., Wang T., Kellis M. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518:317-330.
- Kvon E.Z., Kazmar T., Stampfel G., Yanez-Cuna J.O., Pagani M., Schernhuber K., Dickson B.J., Stark A. Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature.* 2014;512:91-95.
- Kvon E.Z., Stampfel G., Yanez-Cuna J.O., Dickson B.J., Stark A. HOT regions function as patterned developmental enhancers and have a distinct *cis*-regulatory signature. *Genes Dev.* 2012;26:908-913.
- Kwasnieski J.C., Fiore C., Chaudhari H.G., Cohen B.A. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res.* 2014;24:1595-1602.
- Kwasnieski J.C., Mogno I., Myers C.A., Corbo J.C., Cohen B.A. Complex effects of nucleotide variants in a mammalian *cis*-regulatory element. *Proc. Natl. Acad. Sci. USA.* 2012;109:19498-19503.
- Lanzuolo C., Roue V., Dekker J., Bantignies F., Orlando V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* 2007;9:1167-1174.
- Lettice L.A., Williamson I., Devenney P.S., Kilanowski F., Dorin J., Hill R.E. Development of five digits is controlled by a bipartite long-range *cis*-regulator. *Development.* 2014;141:1715-1725.
- Li K., Liu Y., Cao H., Zhang Y., Gu Z., Liu X., Yu A., Kaphle P., Dickerson K.E., Ni M., Xu J. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat. Commun.* 2020;11:485.
- Li L., He S., Sun J.M., Davie J.R. Gene regulation by Sp1 and Sp3. *Biochem. Cell Biol.* 2004;82:460-471.
- Lieberman-Aiden E., van Berkum N.L., Williams L., Imakaev M., Ragozy T., Telling A., Amit I., Lajoie B.R., Sabo P.J., Dorschner M.O., Sandstrom R., Bernstein B., Bender M.A., Groudine M., Gnirke A., Stamatoyannopoulos J., Mirny L.A., Lander E.S., Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009;326:289-293.
- Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorf L.A., Hunter D.J., McCarthy M.I., ..., Clark A.G., Eichler E.E., Gibson G., Haines J.L., Mackay T.F., McCarroll S.A., Visscher P.M. Finding the missing heritability of complex diseases. *Nature.* 2009;461:747-753.
- Maricque B.B., Chaudhari H.G., Cohen B.A. A massively parallel reporter assay dissects the influence of chromatin structure on *cis*-regulatory activity. *Nat. Biotechnol.* 2018. DOI 10.1038/nbt.4285.
- Maricque B.B., Dougherty J.D., Cohen B.A. A genome-integrated massively parallel reporter assay reveals DNA sequence determinants of *cis*-regulatory activity in neural cells. *Nucleic Acids Res.* 2017;45:e16.
- Maurano M.T., Humbert R., Rynes E., Thurman R.E., Haugen E., Wang H., Reynolds A.P., ..., Ziegler S., Cotsapas C., Sotoodehnia N., Glass I., Sunyaev S.R., Kaul R., Stamatoyannopoulos J.A. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337:1190-1195.
- Melnikov A., Murugan A., Zhang X., Tesileanu T., Wang L., Rogov P., Feizi S., Gnirke A., Callan C.G. Jr., Kinney J.B., Kellis M., Lander E.S., Mikkelsen T.S. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 2012;30:271-277.
- Miguel-Escalada I., Pasquali L., Ferrer J. Transcriptional enhancers: functional insights and role in human disease. *Curr. Opin. Genet. Dev.* 2015;33:71-76.
- Muller H.J. Types of visible variations induced by X-rays in *Drosophila*. *J. Genet.* 1930;299-334.
- Mumbach M.R., Rubin A.J., Flynn R.A., Dai C., Khavari P.A., Greenleaf W.J., Chang H.Y. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods.* 2016;13:919-922.
- Murtha M., Tokcaer-Keskin Z., Tang Z., Strino F., Chen X., Wang Y., Xi X., Basilico C., Brown S., Bonneau R., Kluger Y., Dailey L. FIREWACH: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat. Methods.* 2014;11:559-565.
- Nagy P.L., Cleary M.L., Brown P.O., Lieb J.D. Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc. Natl. Acad. Sci. USA.* 2003;100:6364-6369.
- Nettling M., Treutler H., Cerquides J., Grosse I. Detecting and correcting the binding-affinity bias in ChIP-seq data using inter-species information. *BMC Genomics.* 2016;17:347.
- Nora E.P., Goloborodko A., Valton A.L., Gibcus J.H., Uebersohn A., Abdennur N., Dekker J., Mirny L.A., Bruneau B.G. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell.* 2017;169:930-944.e922.
- Pang B., Snyder M.P. Systematic identification of silencers in human cells. *Nat. Genet.* 2020;52:254-263.
- Paredes S.H., Melgar M.F., Sethupathy P. Promoter-proximal CCCTC-factor binding is associated with an increase in the transcriptional pausing index. *Bioinformatics.* 2013;29:1485-1487.
- Pasquali L., Gaulton K.J., Rodriguez-Segui S.A., Mularoni L., Miguel-Escalada I., Akerman I., Tena J.J., ..., Berney T., Gloyn A.L., Ravassard P., Skarmeta J.L.G., Muller F., McCarthy M.I., Ferrer J. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* 2014;46:136-143.
- Patwardhan R.P., Hiatt J.B., Witten D.M., Kim M.J., Smith R.P., May D., Lee C., Andrie J.M., Lee S.I., Cooper G.M., Ahituv N., Pennacchio L.A., Shendure J. Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat. Biotechnol.* 2012;30:265-270.
- Pennacchio L.A., Bickmore W., Dean A., Nobrega M.A., Bejerano G. Enhancers: five essential questions. *Nat. Rev. Genet.* 2013;14:288-295.
- Phillips-Cremins J.E., Corces V.G. Chromatin insulators: linking genome organization to cellular function. *Mol. Cell.* 2013;50:461-474.

- Phillips-Cremins J.E., Sauria M.E., Sanyal A., Gerasimova T.I., Lajoie B.R., Bell J.S., Ong C.T., Hookway T.A., Guo C., Sun Y., Bland M.J., Wagstaff W., Dalton S., McDevitt T.C., Sen R., Dekker J., Taylor J., Corces V.G. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*. 2013;153:1281-1295.
- Rada-Iglesias A., Bajpai R., Swigut T., Brugmann S.A., Flynn R.A., Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 2011;470:279-283.
- Rao S.S., Huntley M.H., Durand N.C., Stamenova E.K., Bochkov I.D., Robinson J.T., Sanborn A.L., Machol I., Omer A.D., Lander E.S., Aiden E.L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665-1680.
- Rao S.S.P., Huang S.C., Glenn St Hilaire B., Engreitz J.M., Perez E.M., Kieffer-Kwon K.R., Sanborn A.L., Johnstone S.E., Bascom G.D., Bochkov I.D., Huang X., Shamim M.S., Shin J., Turner D., Ye Z., Omer A.D., Robinson J.T., Schlick T., Bernstein B.E., Casellas R., Lander E.S., Aiden E.L. Cohesin loss eliminates all loop domains. *Cell*. 2017;171:305-320 e324.
- Rye M.B., Saetrom P., Drablos F. A manually curated ChIP-seq benchmark demonstrates room for improvement in current peak-finder programs. *Nucleic Acids Res*. 2011;39:e25. DOI 10.1093/nar/gkq1187.
- Sandmann T., Girardot C., Brehme M., Tongprasit W., Stolc V., Furlong E.E. A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev*. 2007;21:436-449.
- Sanjana N.E., Wright J., Zheng K., Shalem O., Fontanillas P., Joung J., Cheng C., Regev A., Zhang F. High-resolution interrogation of functional elements in the noncoding genome. *Science*. 2016;353:1545-1549.
- Shlyueva D., Stampfel G., Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet*. 2014;15:272-286.
- Shukla S., Kavak E., Gregory M., Imashimizu M., Shutinoski B., Kashlev M., Oberdoerffer P., Sandberg R., Oberdoerffer S. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*. 2011;479:74-79.
- Skene P.J., Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*. 2017;6:e21856. DOI 10.7554/eLife.21856.
- Song L., Crawford G.E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc*. 2010;2010(2):pdbprot 5384.
- Song L., Zhang Z., Grasfeder L.L., Boyle A.P., Giresi P.G., Lee B.K., Sheffield N.C., ..., Winter D., Clarke N.D., Birney E., Iyer V.R., Crawford G.E., Lieb J.D., Furey T.S. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res*. 2011;21:1757-1767.
- Srinivasan L., Atchison M.L. YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev*. 2004;18:2596-2601.
- Stitzel M.L., Sethupathy P., Pearson D.S., Chines P.S., Song L., Erdos M.R., Welch R., Parker S.C., Boyle A.P., Scott L.J., Margulies E.H., Boehnke M., Furey T.S., Crawford G.E., Collins F.S. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab*. 2010;12:443-455.
- Sur I., Taipale J. The role of enhancers in cancer. *Nat. Rev. Cancer*. 2016;16:483-493.
- Tiwari V.K., McGarvey K.M., Licchesi J.D., Ohm J.E., Herman J.G., Schubeler D., Baylin S.B. PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol*. 2008;6:2911-2927.
- Tolhuis B., Palstra R.J., Splinter E., Grosveld F., de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell*. 2002;10:1453-1465.
- Trynka G., Sandor C., Han B., Xu H., Stranger B.E., Liu X.S., Raychaudhuri S. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet*. 2013;45:124-130.
- Visel A., Blow M.J., Li Z., Zhang T., Akiyama J.A., Holt A., Plajzer-Frick I., Shoukry M., Wright C., Chen F., Afzal V., Ren B., Rubin E.M., Pennacchio L.A. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009;457:854-858.
- Whyte W.A., Orlando D.A., Hnisz D., Abraham B.J., Lin C.Y., Kagey M.H., Rahl P.B., Lee T.I., Young R.A. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153:307-319.
- Yang J., Corces V.G. Chromatin insulators: a role in nuclear organization and gene expression. *Adv. Cancer Res*. 2011;110:43-76.
- Yang S.H., Cheng P.H., Sullivan R.T., Thomas J.W., Chan A.W. Lentiviral integration preferences in transgenic mice. *Genesis*. 2008;46:711-718.
- Zabidi M.A., Arnold C.D., Schernhuber K., Pagani M., Rath M., Frank O., Stark A. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*. 2015;518:556-559.

ORCID ID

S.E. Romanov orcid.org/0000-0002-5989-5756
D.A. Kalashnikova orcid.org/0000-0002-0209-4616
P.P. Laktionov orcid.org/0000-0003-2174-6496

Благодарности. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-34-70141.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 23.11.2020. После доработки 28.03.2021. Принята к публикации 28.03.2021.