

СрG-СВЕТОФОРЫ УЧАСТВУЮТ В АКТИВНОМ ДЕМЕТИЛИРОВАНИИ ДНК

© 2024 г. А. В. Лиознова^а, Ю. А. Медведева^{а, *}

^аФедеральный исследовательский центр “Фундаментальные основы биотехнологии”
Российской академии наук, Москва, 119071 Россия

*e-mail: ju.medvedeva@gmail.com

Поступила в редакцию 13.07.2024 г.

После доработки 04.09.2024 г.

Принята к публикации 16.09.2024 г.

Метилирование ДНК — один из важнейших механизмов, вовлеченных в эпигенетическую регуляцию экспрессии генов. Однако взаимосвязь между метилированием ДНК и экспрессией пока не до конца ясна. Есть примеры, когда изменения в уровне метилирования ДНК вызывают изменения в экспрессии генов, и наоборот — изменения экспрессии влекут за собой изменения в уровне метилирования. Ранее мы ввели понятие СрG-светофоров — отдельных СрG-сайтов, метилирование которых значимо коррелирует с экспрессией, и показали их важную роль в регуляции энхансеров. Теперь мы показали, что уровни метилирования СрG-светофоров гетерогенны в популяции клеток и предположили, что это вызвано их динамическим деметилированием. Наблюдаемое обогащение СрG-светофоров 5-гидроксиметилцитозином (5hmC) и сайтами локализации TET2 (Tet methylcytosine dioxygenase 2) подтвердило нашу гипотезу. С целью выяснить, является ли метилирование СрG-сайтов причиной или следствием экспрессии соответствующего гена, был применен метод статистической оценки направления причинной связи. В результате проведенного анализа среди СрG-сайтов идентифицированы такие, метилирование которых является причиной изменения экспрессии, и такие, для которых изменения в экспрессии вызывают изменения в метилировании. Для СрG-сайтов первого типа характерны более стабильные уровни метилирования в разных клетках и менее выраженное деметилирование в сравнении с СрG-сайтами второго типа. Также показано, что в промоторных областях доля СрG-сайтов, метилирование которых влияет на экспрессию, больше, чем в теле гена, для которого метилирование, скорее всего, следствие экспрессии. На основании этих наблюдений можно предположить, что уровни метилирования СрG-сайтов стабильны там, где они определяют экспрессию ассоциированного с этими сайтами гена, работая по принципу “выключателя”. И, наоборот, для СрG-сайтов, зависящих от экспрессии, уровень метилирования динамичен и различается между клетками в популяции, что вызвано, в первую очередь, активным деметилированием.

Ключевые слова: метилирование ДНК, СрG-сайты, экспрессия генов, эпигенетическая регуляция, причинно-следственные связи

DOI: 10.31857/S0026898424060122, **EDN:** IANHQG

ВВЕДЕНИЕ

Метилирование ДНК представляет собой ковалентную модификацию ДНК без изменения ее нуклеотидной последовательности и служит одним из важнейших механизмов эпигенетической регуляции генома. У млекопитающих чаще всего метилированы цитозины в контексте СрG-динуклеотидов, причем плотность покрытия со-

ставляет 70–80% от общего числа сайтов [1–3]. Неметилированные СрG-динуклеотиды часто обнаруживают в СрG-островках, которые присутствуют в 5'-регуляторных областях многих генов [4], а также других транскрибируемых регионах — даже вдали от генов [5]. Ранее Panchin и др. [6] предположили, что метилированные СрG-сайты остаются функциональными. И это очень услож-

Сокращения: 5hmC — 5-гидроксиметилцитозин; FDR (false discovery rate) — поправка на множественное тестирование; NPC (neural progenitor cells) — нейрональные прогениторные клетки; TSS (transcription start site) — сайт начала транскрипции; TTS (transcription termination site) — сайт терминации транскрипции; М→Э — “Метилирование→Экспрессия” (изменения уровня метилирования вызывают изменения экспрессии); ОНП — однонуклеотидный полиморфизм; Э→М — “Экспрессия→Метилирование” (изменения в экспрессии вызывают изменения уровня метилирования).

няет изучение причинно-следственных связей в механизмах регуляции.

Метилирование/деметилирование ДНК в регуляции транскрипции. Влияние метилирования ДНК на структуру хроматина имеет большое значение для развития и функционирования живого организма. Однако, взаимосвязь между метилированием ДНК и экспрессией генов неоднозначна. Ранее нами была предложена [7] и доработана [8] концепция СрG-светофоров — СрG-динуклеотидов, метилирование которых достоверно коррелирует с экспрессией расположенного рядом гена.

Чаще всего метилирование регуляторных элементов в промоторе (и энхансере [9, 10]) ведет к подавлению экспрессии генов [11–16]. Правда, описаны гены, транскрипция которых не зависит от метилирования [17], а также промоторы, метилирование только отдельных участков которых влияет на экспрессию [18, 19]. Заметим, что метилирование тела гена часто ассоциировано с активной транскрипцией. В этом случае метилирование ДНК служит не причиной, а следствием активной транскрипции. Вот один из примеров: комплекс элонгации привлекает комплекс три-метилирования гистона H3K36 (содержащий белок SETD2), а тот, в свою очередь, связывается с ДНК-метилтрансферазой DNMT3A, которая вводит метильные группы в тело гена [20].

Хотя принято считать, что метилирование соседних СрG-позиций согласовано, экспериментально выявлены случаи, когда метилирование именно отдельных СрG-позиций связано с экспрессией конкретных генов. Вот несколько примеров подобных наблюдений, опубликованных за последние годы.

→ Деметилирование отдельного СрG-динуклеотида внутри экзона-5 гена *PGLYRP3* связано с увеличением экспрессии гена. Неметилированная позиция в основном привлекает РНК-связывающие белки, в том числе факторы сплайсинга. Это значит, что деметилирование этой позиции способствует транскрипции *PGLYRP3* и/или сплайсингу пре-мРНК [21].

→ Метилирование отдельных СрG-динуклеотидов в позициях –77 и +24 регулирует экспрессию гена *TLR2* в периферических мононуклеарных клетках крови [22].

→ Метилирование одной СрG-позиции (СрG9) регулирует экспрессию гена *CHIP* при раке молочной железы [23].

→ СрG-позиция +286, находящаяся во втором экзоне гена *Il6* мышей, играет важную роль в регуляции экспрессии *Il6*. Гипометилирование СрG+286 повышает экспрессию *Il6* [24].

Активное деметилирование ДНК [25–27] — процесс, обратный метилированию, — открыт относительно недавно, видимо, по причине многостадийности [28]. Первый этап активного деметилирования — под действием белков группы ТЕТ из 5-метилцитозина (5mC) образуется 5-гидроксиметилцитозин (5hmC). Белки этой группы используют молекулярный кислород для включения гидроксильной группы в 5mC [29]. Конвертировать 5mC в 5hmC способен не только белок ТЕТ1, но и белки ТЕТ2 и ТЕТ3 [30]. Те же ТЕТ-ферменты окисляют 5hmC дальше: до 5-формилцитозина (5fC) и затем 5-карбоксилцитозина (5caC). 5fC и 5caC могут быть удалены из ДНК и заменены на неметилированный цитозин при репарации оснований [31]. Таким образом, наличие 5hmC может служить индикатором именно происходящего активного деметилирования в клетке, поскольку обратный процесс — метилирование ДНК — происходит в одну стадию, без формирования промежуточных продуктов. Однако это не дает ответа на вопрос: что первично в цепочке этих причинно-следственных связей: метилирование/деметилирование СрG-позиций или изменение уровня экспрессии гена? Ответ на этот вопрос важен и с теоретической, и с практической точек зрения — именно воздействуя на причину, можно изменять структуру хроматина и тем самым экспрессию генов, в том числе в терапевтических целях.

Математические подходы к поиску причинно-следственных связей. Ответы на вопросы причинности (а не вопросы ассоциаций) требуют значительного расширения языка математической статистики, который разрабатывается в области под названием *causal inference* [32, 33]. Стандартный статистический анализ направлен на то, чтобы по отдельным измерениям оценить параметры распределения, которому эти измерения подчиняются. Основываясь на параметрах этих распределений, можно делать выводы об ассоциативной связи между переменными, что позволяет оценивать вероятности событий. Причинно-следственный анализ ищет ответ на вопрос, какое явление вызывает другое явление.

В настоящее время исследователи проводят поиск причинно-следственных связей между метилированием ДНК и экспрессией генов. Введем нотацию М→Э (Метилирование→Экспрессия), чтобы обозначить случаи, когда изменения в уровне метилирования являются причиной изменений в экспрессии; и Э→М (Экспрессия→Метилирование) для случаев, когда изменения в экспрессии вызывают изменения уровня метилирования. Результаты исследований пока остаются противоречивыми. Например:

→ D. Taylor с соавт. [34] выявили 213 причинно-следственных связей между экспресси-

ей (115 генов) и метилированием ДНК (190 сайтов метилирования), из которых 64% имели направление $M \rightarrow \Theta$ и 36% $\Theta \rightarrow M$. Сайты метилирования ДНК для пар $M \rightarrow \Theta$ находились ближе к сайтам начала транскрипции (transcription start site, TSS), чем для пар $\Theta \rightarrow M$;

→G. Hemani с соавт. [35] сообщали, что в среднем случаи $M \rightarrow \Theta$ встречались чаще, чем $\Theta \rightarrow M$, но при этом заметили, что результат подвержен сильным искажениям, связанным с погрешностью измерений.

→P. Нор с соавт. [36] идентифицировали 818 генов, которые влияли на паттерны метилирования ДНК в крови, то есть для них, реализуется направление $\Theta \rightarrow M$. Среди найденных генов значимо чаще встречались транскрипционные факторы;

→X. Мо с соавт. [37] выявили 5 причинно-следственных цепочек “ $M \rightarrow \Theta \rightarrow$ ревматоидный артрит” для CpG-сайтов, локализованных в CpG-островках промоторов генов;

→J. Zhang с соавт. [38] исследовали влияние переменных на выживаемость при раке шейки матки и выявили 12623 гена, для которых реализуется направление $M \rightarrow \Theta$. Интересно, что большинство этих генов содержит меньше 50 CpG-сайтов.

→K. van Eijk с соавт. [39] привели доказательства в поддержку направления $M \rightarrow \Theta$, а не наоборот. Авторы выбрали 20 CpG-сайтов, ассоциированных с экспрессией 19 генов, содержащих суммарно 147 однонуклеотидных полиморфизмов (ОНП). Для модели ОНП $\rightarrow M \rightarrow \Theta$ авторы нашли 29.9% комбинаций в ее поддержку, а для модели ОНП $\rightarrow \Theta \rightarrow M$ — 6.8%. Для локуса одного из генов выявлены причинно-следственные связи в обоих направлениях.

→A. Díez-Villanueva A. с соавт. [40] рассматривали связь между ОНП, метилированием и экспрессией. Авторы показали, что 60.1% этих комбинаций подчинялось каноническому пути $M \rightarrow \Theta$, 33.9% не имело причинно-следственной связи между метилированием и экспрессией генов, а 6% реализовали путь $\Theta \rightarrow M$;

→B. Pierce с соавт. [41], выявили более 400 пар метилирования—экспрессии (проанализировано 2913 пар) с наличием причинно-следственной связи. Из них 161 пара подчинялась модели ОНП $\rightarrow M \rightarrow \Theta$, а 125 пар — ОНП $\rightarrow \Theta \rightarrow M$; эти множества включали 119 общих пар, которые “работали” в двух направлениях. Авторы заметили, что проще выявить наличие причинно-следственной связи, чем ее направление.

В целом уже сейчас очевидно, что взаимосвязь между метилированием ДНК и экспрессией генов имеет неоднозначный характер. Ранее

нами описан механизм работы CpG-светофоров [7, 8], теперь мы детально исследовали динамику их метилирования—деметилования в процессе регуляции транскрипции и проанализировали причинно-следственные связи между процессами метилирования и экспрессии.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Выбор CpG-светофоров и контрольных позиций. В работе использовано понятие “CpG-светофор”, подробно описанное нами ранее [8] и обозначающее CpG-позиции, метилирование которых связано с экспрессией ассоциированного гена. Для 48 образцов тканей и клеток человека из проекта ENCODE (<https://www.encodeproject.org/>) проанализированы все пары “ген и CpG-динуклеотид”, где динуклеотид расположен в промежутке от -10000 п.н. в 5'-направлении от TSS и до 3'-конца гена (transcription termination site, TTS), то есть все CpG-позиции на участке $[-10000$ п.н., TTS]. При таком выборе позиций один CpG-динуклеотид может быть ассоциирован с несколькими генами, а один ген — с несколькими динуклеотидами. Для каждой пары “ген и CpG-динуклеотид” были построены два вектора: метилирования этой позиции во всех доступных образцах и экспрессии ассоциированного гена во всех доступных образцах, — после чего вычисляли коэффициент корреляции Спирмена между этими двумя векторами и вводили поправку на множественное тестирование (false discovery rate, FDR) методом Бенджамини-Хохберга. CpG-позиция называется CpG-светофором, если для нее коэффициент корреляции Спирмена между метилированием и экспрессией значим на уровне $FDR < 0.01$. В пару к каждому CpG-светофору была выбрана фоновая CpG-позиция с учетом расстояния до TSS, GC-и CpG-контента (подробный алгоритм см. в разделе “Методы” работы [8]). Процесс выбора фоновых позиций был независимо повторен 50 раз.

Анализ уровня 5-гидроксиметилцитозина. В стандартных экспериментах, использующих бисульфитную конверсию (BS), нельзя отличить 5hmC от его предшественника 5mC, так как в этих условиях обе модификации ведут себя одинаково [42] и исследователь измеряет суммарный уровень (5mC + 5hmC). С целью дифференцировать сигналы 5mC и 5hmC дополнительно проводили бисульфитную конверсию с окислением (oxBS), которая позволяет измерить уровень только 5mC (не включая 5hmC). Уровень 5hmC вычисляли как разницу (BS – oxBS).

Для анализа использовали данные Illumina 450K oxBS-array, полученные S. Field с соавт. [43] (идентификатор данных <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63179>).

(Авторы провели бисульфитную конверсию, а также бисульфитную конверсию с окислением для 450 тыс. проб в клетках мозжечка человека. Авторы предложили метод анализа, при котором сигнал сглаживается, но мы приняли решение не вносить дополнительные вычислительные шумы, а воспользоваться исходными данными.) Для каждой пробы мы вычисляли разницу ($BS - \alpha BS$) и усредняли полученные величины по 4 репликам, то есть использовали значение $5hmC$, равное $avg(BS_1 - \alpha BS_1, BS_2 - \alpha BS_2, BS_3 - \alpha BS_3, BS_4 - \alpha BS_4)$. При таком подходе возможны отрицательные значения $5hmC$, возникающие вследствие неточности измерений по обоим протоколам. Интерпретировать их не представляется возможным, но это не препятствовало пониманию общей картины, так как нас интересовали различия между уровнями $5hmC$, а не абсолютные значения.

Анализ связывания TET2. Для анализа связывания белка TET2 мы воспользовались данными

ми, полученными Y. Qiao и др. [44]. Нами были выбраны полногеномные профили, полученные анализом ChIP-Seq, локализации TET2 в дифференцированных на 12 суток клетках нейронов NPC (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1668531>). Авторы определили регионы локализации белка TET2 (пики ChIP-seq) на сборке генома hg19. Нами была использована утилита liftOver [45] для того, чтобы перенести координаты регионов локализации TET2 на hg38.

Учитывая, что CrG-светофоры были выбраны на основе анализа 48 различных тканей, мы предположили, что лежащие в основе механизмы являются фундаментальными и не зависят от конкретной ткани. Тем не менее, мы выбрали дифференцирующиеся нейроны как максимально близкие (из доступных опубликованных образцов) к описанным выше данным $5hmC$ в мозжечке человека, чтобы сравнение было более адекватным.

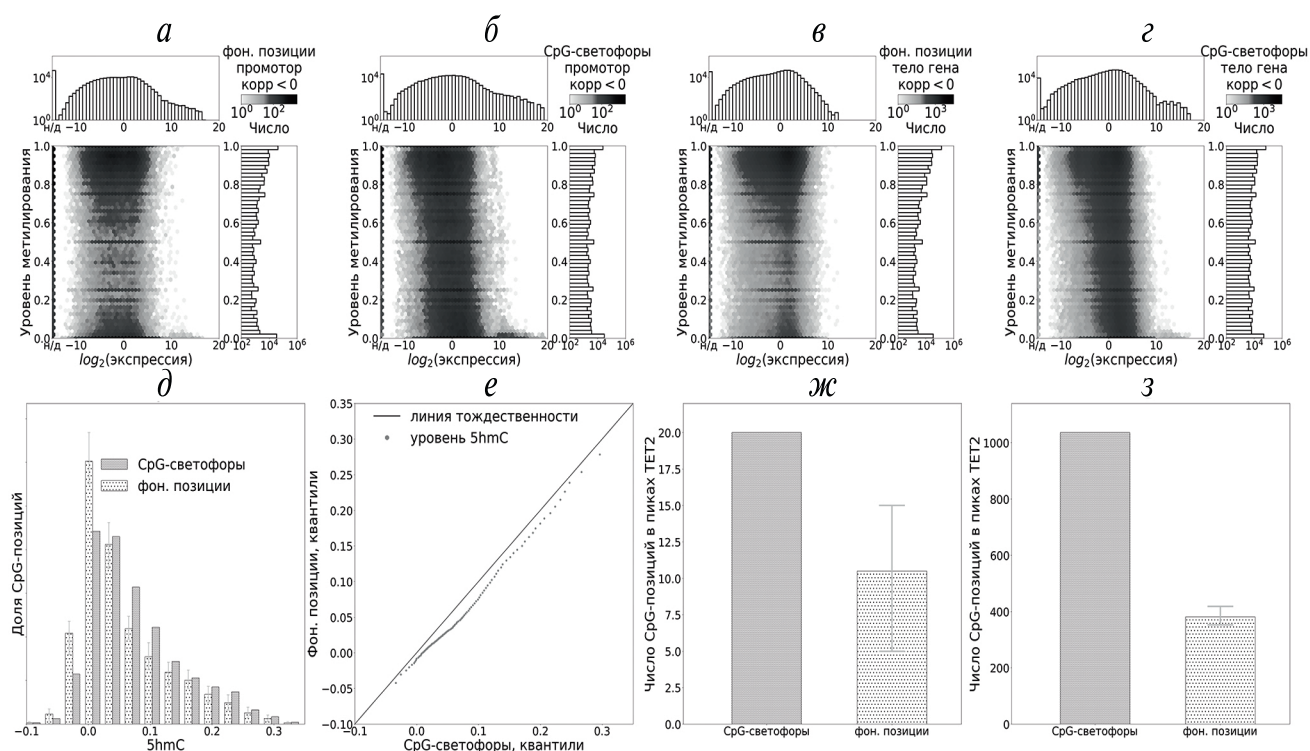


Рис. 1. Распределение метилирования и экспрессии генов, ассоциированных с CrG-светофорами и фоновыми позициями с отрицательным коэффициентом корреляции Спирмена. Тон отражает плотность точек в логарифмическом масштабе. Распределение показано для случайных фоновых CrG-позиций в промоторе (число позиций равно числу CrG-светофоров) (а); CrG-светофоров в промоторе (б); случайных фоновых CrG-позиций в теле гена (в); CrG-светофоров в теле гена [+500, TTS] (г). На панели (д) показано распределение доли CrG-светофоров и фоновых CrG-позиций в зависимости от уровня $5hmC$ в образце мозжечка; на панели (е) — квантили распределения (д). На панелях (ж) и (з) показано число CrG-позиций, локализованных с пиками сигнала TET2 в двух различных образцах соответственно: дифференцированных на 12 суток нейронах (клетки NPC) и макрофагов, полученных из моноцитов. Здесь и далее на рисунках: планки погрешностей соответствуют 5%- и 95%-ному процентилю из 50 выборок фоновых позиций; для фоновых позиций приведена медиана из 50 выборок; фон. позиции — фоновые позиции; корр — коэффициент корреляции Спирмена.

Число пиков, колокализованных с CpG-позициями в клетках NPC, невелико, поэтому для подтверждения полученных результатов мы использовали второй образец полногеномных профилей локализации TET2, полученных I. Sobo с соавт. [46]. Нами выбраны данные, полученные для макрофагов, дифференцированных из моноцитов (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206030>). Авторами опубликованы пики TET2 на сборке hg38.

Для анализа колокализации TET2 и выбранных CpG-сайтов интервалы были пересечены с помощью утилиты bedtools intersect [47].

Поиск причинно-следственных связей. Для изучения направления причинно-следственных связей в геномных данных на настоящий момент разработано несколько инструментов [48]. Для анализа направления связи мы следовали логике, описанной в книге J. Peters и др. [32], а именно логике иллюстративного примера 4.5 (рис. S1, Дополнительные материалы в электронном виде по DOI статьи и на сайте http://www.molecbio.ru/downloads/2024/6/supp_Lioznova_rus.zip). В этом примере рассмотрена выборка (X, Y), для которой построена линейная регрессия в обоих направлениях: $X \rightarrow Y$ и $Y \rightarrow X$, — после чего проанализированы остатки (ошибки) полученной регрессионной модели. Если регрессионная модель отражает истинное направление причинно-следственной связи, то распределение остатков будет независимым.

В данной работе выбран инструмент dHSIC [49], реализующий описанную логику. Мы предполагаем линейную связь между исследуемыми переменными $\text{model} \leftarrow \text{lm}(X \sim Y)$, после чего анализируем распределение остатков `dhsic.test(model$residuals, Y)$p.value`. Такая проверка проводится для обоих потенциальных направлений причинно-следственной связи, в нашем случае для $\Theta \rightarrow M$ и $M \rightarrow \Theta$. Таким образом, в результате анализа для каждой пары “ген и CpG-динуклеотид” мы получили две величины p -value, отражающие значимость для обоих потенциальных направлений связи. С целью скорректировать p -value на полное число пар “ген и CpG-динуклеотид” мы произвели поправку FDR на множественное тестирование методом Бенджамини-Хохберга. Значимым считали p -value < 0.05 после поправки методом Бенджамини-Хохберга. Определенное направление связи между экспрессией и метилированием ДНК приписывали паре “ген и CpG-динуклеотид”, если одно из двух значений p -value было меньше заданного порога.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Гетерогенность уровня метилирования CpG-светофоров

Конкретный CpG-динуклеотид может быть либо метилирован, либо нет, то есть в отдельной клетке на уровне определенного аллеля уровень метилирования бинарен (0 или 1 в используемой нами нотации). В диплоидной клетке аллель-специфичное метилирование CpG-позиций может привести к уровню метилирования 0.5. Аллель-специфичное метилирование найдено для 10% генов человека [50] и обычно ассоциировано с генетическими полиморфизмами [50, 51] или импринтингом. Статус метилирования конкретных позиций ДНК, в отличие от генотипа, может различаться между разными клетками одного организма (даже внутри одного типа клеток). В связи с этим на практике измеряют долю метилированных CpG-позиций, а не бинарное значение метилирования в позиции [51, 52], то есть наблюдается гетерогенность уровней метилирования в данной CpG-позиции среди отдельных клеток.

Для большинства CpG-позиций, которые не относятся к CpG-светофорам (фоновые CpG-позиции), уровни метилирования близки либо к 0, либо к 1 во всех изученных типах клеток (рис. 1а, в). Это свидетельствует об однородности уровней метилирования в клеточной популяции, причем в теле гена метилирование чаще всего близко к 1. В то же время CpG-светофоры с отрицательным коэффициентом корреляции Спирмена между экспрессией и метилированием, расположенные как в промоторах, так и в теле гена, часто имеют промежуточные уровни метилирования во многих типах клеток (рис. 1б, г). Аналогичная тенденция зарегистрирована и для CpG-светофоров с положительным коэффициентом корреляции Спирмена (рис. S2, см. электронное приложение).

CpG-светофоры деметилируются динамически

В связи с тем, что уровни метилирования CpG-светофоров более гетерогенны, чем фоновых CpG-позиций, было решено проверить, является ли это наблюдение следствием более динамичного процесса деметилирования этих позиций. В качестве параметров, отражающих изменения метилирования, мы использовали уровень 5hmC и сигнал связывания белка TET2 — параметры, отражающие деметилирование этих позиций. Хотя функциональная роль 5hmC не совсем ясна, одна из наиболее популярных гипотез заключается в том, что это промежуточный продукт активного деметилирования ДНК с участием белков группы TET [53].

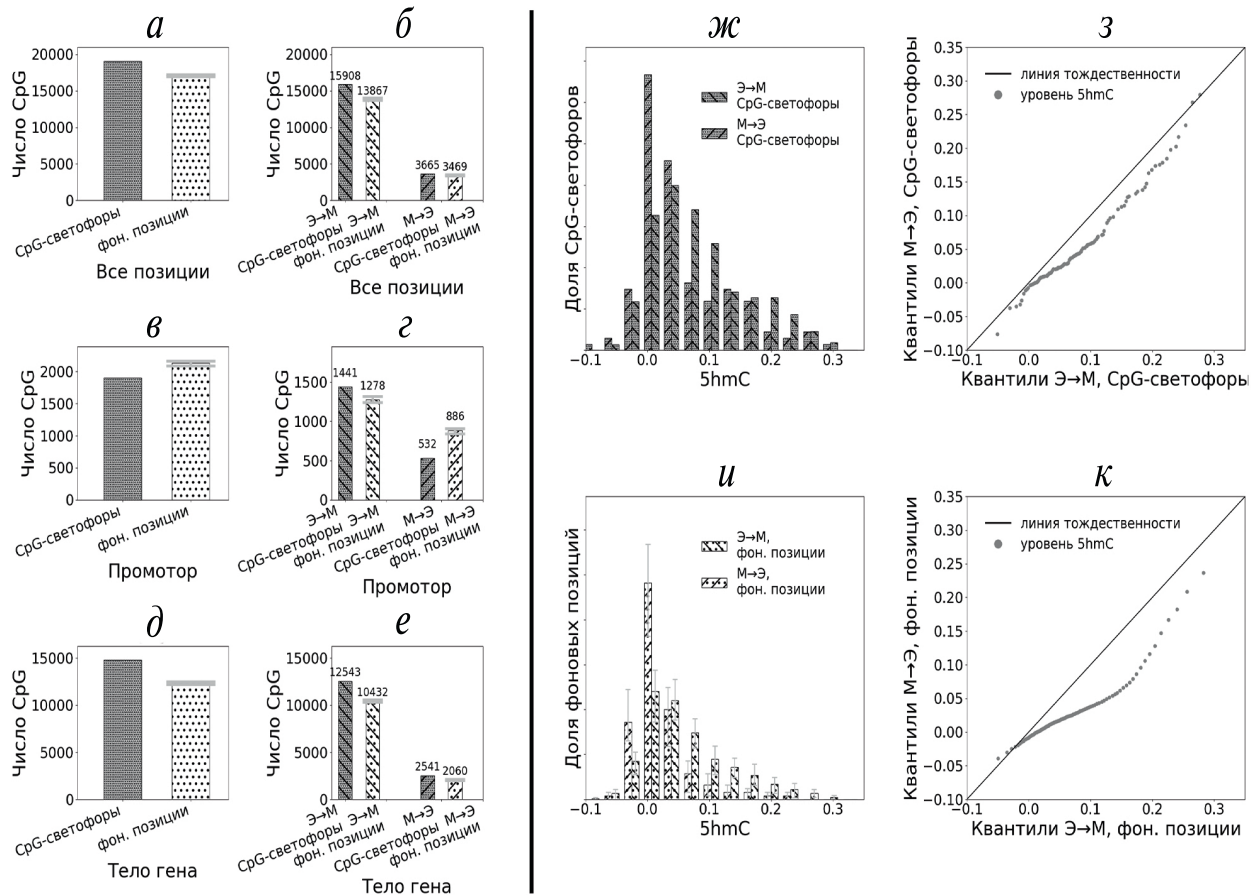


Рис. 2. Анализ причинно-следственных связей и распределения 5hmC в зависимости от направления связи. На левой панели представлены результаты количественного анализа СрG-позиций, для которых выявлено наличие причинно-следственной связи. Левая колонка показывает общее число позиций, а колонка справа — число позиций в зависимости от направления связи. На графиках показано число позиций, имеющих коэффициент причинно-следственной связи для СрG-светофоров и фоновых СрG-позиций во всем геноме (*а*) и их уточнение — число позиций в зависимости от направления связи (*б*). На видах (*в*) и (*г*) показаны позиции в промоторной области [−1000, TSS+500], на (*д*) и (*е*) — в теле гена [TSS+500, TTS]. На правой панели представлено распределение 5hmC: (*ж*) — распределение доли СрG-светофоров в зависимости от уровня 5hmC для обоих направлений причинно-следственной связи; (*з*) — квантили распределения на виде (*ж*); (*и*) — распределение, аналогичное (*ж*), только для фоновых СрG-позиций; (*к*) — квантили распределения на панели (*и*).

Как видно из данных, приведенных на рис. 1(*д,е*), по сравнению с фоновыми СрG-позициями большая доля СрG-светофоров обогащена 5hmC. Заметим, что здесь мы привели долю, а не общее число СрG-позиций, так как данные получены для 450 тыс. СрG-проб, выбранных по геному неслучайно, что может исказить абсолютные значения количества числа СрG-светофоров и фоновых позиций. Кроме обогащения 5hmC, СрG-светофоры гораздо чаще колокализовались с TET2, чем фоновые СрG-позиции (рис. 1*ж,з*). Это согласуется с результатами, полученными нами ранее [54].

Таким образом, полученные здесь данные служат подтверждением выдвинутой нами гипотезы о динамическом деметилировании СрG-светофоров. Модель динамического деметилирования

СрG-светофоров может объяснить наблюдаемую среди них неоднородность уровней метилирования в клеточной популяции.

Активное деметилирование происходит в позициях Э→М

На следующем этапе исследования СрG-позиции были разделены на подгруппы на основе оценки причинно-следственной связи между уровнями экспрессии и метилирования: изменения в метилировании ДНК вызывают изменения экспрессии (М→Э) или уровни метилирования зависят от интенсивности экспрессии (Э→М).

Прежде всего мы убедились, что для СрG-светофоров значимо чаще бывает определено направ-

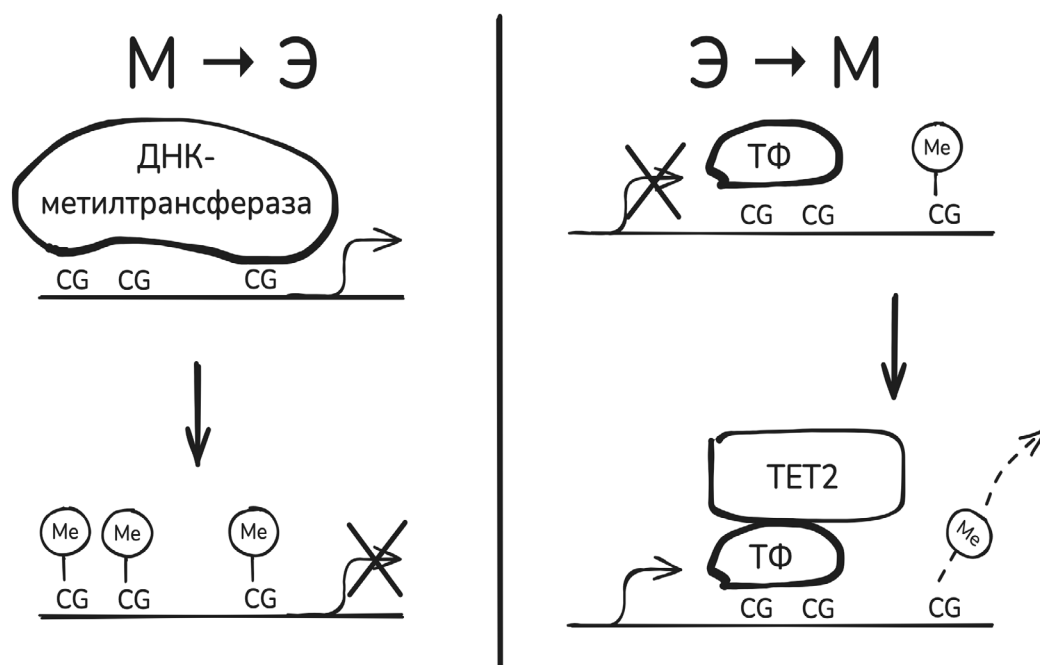


Рис. 3. Предполагаемый механизм функционирования CpG-светофоров. На левой панели представлены CpG-светофоры, в которых изменения уровня метилирования (Me) вызывают изменения экспрессии ассоциированного гена. Такие CpG-светофоры чаще встречаются в промоторах, их метилирование стабильно и работают они по принципу “вкл/выкл”. Правая панель показывает CpG-светофоры, для которых уровень метилирования зависит от экспрессии гена. Их метилирование различно между клетками и динамично, что вызвано активным деметилированием. Деметилирование происходит под действием метилцитозиндиоксигеназы TET2, которая привлекается транскрипционными факторами (ТФ).

ление причинно-следственной связи, чем для фоновых CpG-позиций (рис. 2а). При рассмотрении только промоторных областей видно, что здесь число CpG-светофоров, для которых определено направление причинно-следственной связи, меньше числа фоновых CpG-позиций (рис. 2б).

Далее мы проанализировали соотношение числа позиций, поддерживающих каждое из направлений связи. Мы заметили, что связь Э→М (уровень экспрессии определяет уровень метилирования) встречается примерно в 4 раза чаще, чем М→Э (уровень метилирования определяет уровень экспрессии) для всех позиций (рис. 2б), но в промоторной области разница меньше (рис. 2г). Эта тенденция согласуется с признанным тезисом, что метилирование промотора гена подавляет его экспрессию.

С целью лучше понять процесс деметилирования ДНК мы проанализировали уровень 5hmC в свете причинно-следственных связей между уровнями метилирования ДНК и экспрессии. Заметим, что именно CpG-позиции с направлением причинно-следственной связи Э→М обогащены 5hmC (рис. 2ж,и). Это может означать наличие положительной обратной связи между активной транскрипцией и деметилированием ДНК. Для наглядности мы привели квантиль-квантильные распределения (рис. 2з,к), которые показывают

разницу распределений при разных направлениях причинно-следственной связи. Отметим, что разница между фоновыми CpG-позициями и CpG-светофорами с направлением связи М→Э (рис. S3а,б, см. электронное приложение) более выражена, чем для всех позиций в целом (рис. 1д,е).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

На основании результатов многочисленных исследований, посвященных механизмам метилирования ДНК, ни у кого не вызывает сомнения, что это сложный процесс, регулируемый в том числе на уровне отдельных CpG-динуклеотидов. Известно, что часто соседние CpG-динуклеотиды метилируются подобным образом. В частности показано, что метилирование новосинтезированной ДНК-цепи под действием ДНК-метилтрансферазы 1 (DNMT1) происходит с учетом уровня метилирования не только той же позиции материнской цепи, но и соседних динуклеотидов [55]. Кроме того, метилирование тела активно транскрибируемого гена под действием ДНК-метилтрансферазы 3А (DNMT3А) обусловлено, как предполагают, привлечением этой метилтрансферазы белком SETD2, который регулируется комплексом элонгации [56]. Также показано,

что гетеротетрамер DNMT3A/3L содержит два активных центра и одновременно метилирует СрG-сайты на расстоянии 12 нуклеотидов [57]. При исследовании структуры фермента DNMT3A показано, что он может скоординированно метилировать СрG-сайты на расстоянии от 8 до 10 нуклеотидов [58–60].

Напротив, активное деметилирование, видимо, представляет более локальный процесс. Показано, что связывание белков группы ТЕТ с ДНК происходит специфически и определяется ее последовательностью (сиквенсспецифично) [61]. В работах Suzuki с соавт. [62, 63] показано, что белок ТЕТ2, который не имеет ДНК-связывающего домена, привлекается рядом транскрипционных факторов, которые сиквенсспецифично связывают ДНК. Следовательно, этот механизм реализуется скорее локально, и это подтверждено нами ранее [54].

Можно предположить, что метилирование цитозинов в районе СрG-светофоров может накапливаться вследствие отсутствия связывания транскрипционных факторов, что приводит к метилированию СрG-светофоров, т.е. их метилирование в этом случае не причина, а следствие “пассивная” метка отсутствующей экспрессии гена в результате инактивации его регуляторного элемента. Эта гипотеза подтверждается результатами нескольких исследований [64, 65], в том числе, полученными нами теперь.

Деметилирование регуляторного района также может быть следствием связывания определенных факторов транскрипции, которые привлекают ТЕТ-белки. В этом случае выглядит логичным, что деметилирование происходит не в самом сайте связывания, а в его ближайших окрестностях, с которыми деметилаза может входить в прямой контакт.

Мы предполагаем, что СрG-позиции, в которых уровень метилирования определяет уровень экспрессии, — это районы, где метилирование ДНК работает как своего рода переключатель, который устанавливается одновременно и далее поддерживается в таком состоянии при делении клеток. Таким образом, уровень метилирования оказывается близким к 0 или 1 во всех клетках этого типа, а районы, где уровень экспрессии определяет уровень метилирования, гетерогенны и, видимо, возникают в значительной степени как результат активного деметилирования с использованием ТЕТ-белков, связывание которых происходит динамически в разных клетках (рис. 3).

Таким образом, на основании экспериментально полученных результатов можно сделать вывод, что метилирование промоторов генов

не всегда вызывает подавление экспрессии [66]. Следовательно, для решения медицинских и других прикладных задач воздействие на уровень метилирования не всегда правомерно. Метилирование СрG-позиций может быть не причиной, а лишь следствием изменения экспрессии. Разработанная нами модель позволит в будущем более точно выбирать СрG-мишени для воздействия на экспрессию ассоциированных с ними генов.

Работа частично поддержана грантом РНФ-23-14-00371.

Настоящая статья не содержит каких-либо исследований с использованием людей или животных в качестве объектов.

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Ziller M.J., Gu H., Müller F., Donaghey J., Tsai L.T.Y., Kohlbacher O., De Jager P.L., Rosen E.D., Bennett D.A., Bernstein B.E., Gnirke A., Meissner A. (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature*. **500**, 477–481.
2. Schroeder D.I., Blair J.D., Lott P., Yu H.O.K., Hong D., Crary F., Ashwood P., Walker C., Korf I., Robinson W.P., LaSalle J.M. (2013) The human placenta methylome. *Proc. Natl. Acad. Sci. USA*. **110**, 6037–6042.
3. Loyfer N., Magenheimer J., Peretz A., Cann G., Bredno J., Klochender A., Fox-Fisher I., Shabi-Porat S., Hecht M., Pelet T., Moss J., Drawshy Z., Amini H., Moradi P., Nagaraju S., Bauman D., Shveiky D., Porat S., Dior U., Rivkin G., Or O., Hirshoren N., Carmon E., Pikarsky A., Khalaileh A., Zamir G., Grinbaum R., Abu Gazala M., Mizrahi I., Shussman N., Korach A., Wald O., Izhar U., Erez E., Yutkin V., Samet Y., Rotnemer Golinkin D., Spalding K.L., Druid H., Arner P., Shapiro A.M.J., Grompe M., Aravanis A., Venn O., Jamshidi A., Shemer R., Dor Y., Glaser B., Kaplan T. (2023) A DNA methylation atlas of normal human cell types. *Nature*. **613**, 355–364.
4. Hernando-Herraez I., Garcia-Perez R., Sharp A.J., Marques-Bonet T. (2015) DNA methylation: insights into human evolution. *PLoS Genet*. **11**, e1005661.
5. Medvedeva Y.A., Fridman M.V., Oparina N.J., Malko D.B., Ermakova E.O., Kulakovskiy I.V., Heinzl A., Makeev V.J. (2010) Intergenic, gene terminal, and intragenic CpG islands in the human genome. *BMC Genomics*. **11**, 1–16.
6. Panchin A.Y., Makeev V.J., Medvedeva Y.A. (2016) Preservation of methylated CpG dinucleotides in human CpG islands. *Biol. Direct*. **11**, 1–15.
7. Pardo L.M., Rizzu P., Francescatti M., Vitezic M., Leday G.G.R., Sanchez J.S., Khamis A., Takahashi H., van de Berg W.D.J., Medvedeva Y.A., van de Wiel M.A., Daub C.O., Carninci P., Heutink P. (2013) Regional

- differences in gene expression and promoter usage in aged human brains. *Neurobiol. Aging*. **34**, 1825–1836.
8. Lioznova A.V., Khamis A.M., Artemov A.V., Besedina E., Ramensky V., Bajic V.B., Kulakovskiy I.V., Medvedeva Y.A. (2019) CpG traffic lights are markers of regulatory regions in human genome. *BMC Genomics*. **20**, 102.
 9. Pan X., Li X., Sun J., Xiong Z., Hu H., Ning S., Zhi H. (2022) Enhancer methylation dynamics drive core transcriptional regulatory circuitry in pan-cancer. *Oncogene*. **41**, 3474–3484.
 10. Benetatos L., Vartholomatos G. (2018) Enhancer DNA methylation in acute myeloid leukemia and myelodysplastic syndromes. *Cell. Mol. Life Sci.* **75**, 1999–2009.
 11. Stein R., Razin A., Cedar H. (1982) In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. *Proc. Natl. Acad. Sci. USA*. **79**, 3418–3422.
 12. Anastasiadi D., Esteve-Codina A., Piferrer F. (2018) Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin*. **11**, 1–17.
 13. Kang J.G., Park J.S., Ko J.H., Kim Y.S. (2019) Regulation of gene expression by altered promoter methylation using a CRISPR/Cas9-mediated epigenetic editing system. *Sci. Rep.* **9**, 1–12.
 14. Antequera F., Bird A. (1993) CpG islands. *EXS*. **64**, 169–185.
 15. Deaton A.M., Bird A. (2011) CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022.
 16. Meehan R., Lewis J., Cross S., Nan X., Jeppesen P., Bird A. (1992) Transcriptional repression by methylation of CpG. *J. Cell Sci. Suppl.* **16**, 9–14.
 17. Bird A.P. (1984) DNA methylation versus gene expression. *J. Embryol. Exp. Morphol.* **83**(Suppl), 31–40.
 18. Eckhardt F., Lewin J., Cortese R., Rakyan V.K., Attwood J., Burger M., Burton J., Cox T.V., Davies R., Down T.A., Haefliger C., Horton R., Howe K., Jackson D.K., Kunde J., Koenig C., Liddle J., Niblett D., Otto T., Pettett R., Seemann S., Thompson C., West T., Rogers J., Olek A., Berlin K., Beck S. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38**, 1378–1385.
 19. Weber M., Hellmann I., Stadler M.B., Ramos L., Pääbo S., Rebhan M., Schübeler D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466.
 20. Jones P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492.
 21. Sobiak B., Leśniak W. (2019) The effect of single CpG demethylation on the pattern of DNA-protein binding. *Int. J. Mol. Sci.* **20**(4), 914.
 22. Bordagaray M.J., Fernández A., Astorga J., Garrido M., Hernández P., Chaparro A., Lira M.J., Gebicke-Haerter P., Hernández M. (2022) CpG single-site methylation regulates TLR2 expression in proinflammatory PBMCs from apical periodontitis individuals. *Front. Immunol.* **13**, 861665.
 23. Gohno T., Hanamura T., Kurosumi M., Takei H., Yamaguchi Y., Hayashi S.I. (2022) One DNA methylation regulates gene expression of human breast cancer and predicts recurrence. *Anticancer Res.* **42**, 759–766.
 24. Lian B.S.X., Kawasaki T., Kano N., Ori D., Ikegawa M., Isotani A., Kawai T. (2022) Regulation of expression by single CpG methylation in downstream of transcription initiation site. *iScience*. **25**, 104118.
 25. Wu X., Zhang Y. (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**, 517–534.
 26. Wu S.C., Zhang Y. (2010) Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**, 607–620.
 27. Prasad R., Yen T.J., Bellacosa A. (2021) Active DNA demethylation — the epigenetic gatekeeper of development, immunity, and cancer. *Adv. Genet.* **2**, e10033.
 28. Moore L.D., Le T., Fan G. (2013) DNA methylation and its basic function. *Neuropsychopharmacology*. **38**, 23–38.
 29. Tahiliani M., Koh K.P., Shen Y., Pastor W.A., Bandukwala H., Brudno Y., Agarwal S., Iyer L.M., Liu D.R., Aravind L., Rao A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. **324**, 930–935.
 30. Ito S., D'Alessio A.C., Taranova O.V., Hong K., Sowers L.C., Zhang Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*. **466**, 1129–1133.
 31. Shi D.Q., Ali I., Tang J., Yang W.C. (2017) New insights into 5hmC DNA modification: generation, distribution and function. *Front. Genet.* **8**, 100.
 32. Peters J., Janzing D., Scholkopf B. (2017) *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press.
 33. Pearl J. (2010) An introduction to causal inference. *Int. J. Biostat.* **6**, Article 7. doi: 10.2202/1557–4679.1203
 34. Taylor D.L., Jackson A.U., Narisu N., Hemani G., Erdos M.R., Chines P.S., Swift A., Idol J., Didion J.P., Welch R.P., Kinnunen L., Saramies J., Lakka T.A., Laakso M., Tuomilehto J., Parker S.C.J., Koistinen H.A., Davey Smith G., Boehnke M., Scott L.J., Birney E., Collins F.S. (2019) Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *Proc. Natl. Acad. Sci. USA*. **116**, 10883–10888.
 35. Hemani G., Tilling K., Davey Smith G. (2017) Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081.
 36. Hop P.J., Luijk R., Daxinger L., van IJterson M., Dekkers K.F., Jansen R., BIOS Consortium, van Meurs J.B.J., 't Hoen P.A.C., Ikram M.A., van Greevenbroek M.M.J., Boomsma D.I., Slagboom P.E., Veldink J.H., van Zwet E.W., Heijmans B.T. (2020) Genome-wide identification of genes regulating DNA

- methylation using genetic anchors for causal inference. *Genome Biol.* **21**, 220.
37. Mo X.B., Zhang Y.H., Lei S.F. (2021) Integrative analysis identifies potential causal methylation-mRNA regulation chains for rheumatoid arthritis. *Mol. Immunol.* **131**, 89–96.
 38. Zhang J., Lu H., Zhang S., Wang T., Zhao H., Guan F., Zeng P. (2021) Leveraging methylation alterations to discover potential causal genes associated with the survival risk of cervical cancer in TCGA through a two-stage inference approach. *Front. Genet.* **12**, 667877.
 39. van Eijk K.R., de Jong S., Boks M.P.M., Langeveld T., Colas F., Veldink J.H., de Kovel C.G.F., Janson E., Strengman E., Langfelder P., Kahn R.S., van den Berg L.H., Horvath S., Ophoff R.A. (2012) Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics.* **13**, 636.
 40. Díez-Villanueva A., Jordà M., Carreras-Torres R., Alonso H., Cordero D., Guinó E., Sanjuan X., Santos C., Salazar R., Sanz-Pamplona R., Moreno V. (2021) Identifying causal models between genetically regulated methylation patterns and gene expression in healthy colon tissue. *Clin. Epigenetics.* **13**, 162.
 41. Pierce B.L., Tong L., Argos M., Demanelis K., Jasmine F., Rakibuz-Zaman M., Sarwar G., Islam M.T., Shahriar H., Islam T., Rahman M., Yunus M., Kibriya M.G., Chen L.S., Ahsan H. (2018) Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. *Nat. Commun.* **9**, 804.
 42. Jin S.G., Kadam S., Pfeifer G.P. (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* **38**, e125.
 43. Field S.F., Beraldi D., Bachman M., Stewart S.K., Beck S., Balasubramanian S. (2015) Accurate measurement of 5-methylcytosine and 5-hydroxymethylcytosine in human cerebellum DNA by oxidative bisulfite on an array (OxBS-array). *PLoS One.* **10**, e0118202.
 44. Qiao Y., Wang X., Wang R., Li Y., Yu F., Yang X., Song L., Xu G., Chin Y.E., Jing N. (2015) AF9 promotes hESC neural differentiation through recruiting TET2 to neurodevelopmental gene loci for methylcytosine hydroxylation. *Cell Discovery.* **1**, 1–19.
 45. Hinrichs A.S., Karolchik D., Baertsch R., Barber G.P., Bejerano G., Clawson H., Diekhans M., Furey T.S., Harte R.A., Hsu F., Hillman-Jackson J., Kuhn R.M., Pedersen J.S., Pohl A., Raney B.J., Rosenbloom K.R., Siepel A., Smith K.E., Sugnet C.W., Sultan-Qurraie A., Thomas D.J., Trumbower H., Weber R.J., Weirauch M., Zweig A.S., Haussler D., Kent W.J. (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598.
 46. Cobo I., Tanaka T.N., Mangalhara K.C., Lana A., Yeang C., Han C., Schlachetzki J., Challcombe J., Fixsen B.R., Sakai M., Li R.Z., Fields H., Mokry M., Tsai R.G., Bejar R., Prange K., de Winther M., Shadel G.S., Glass C.K. (2022) DNA methyltransferase 3 alpha and TET methylcytosine dioxygenase 2 restrain mitochondrial DNA-mediated interferon signaling in macrophages. *Immunity.* **55**, 1386–1401.e10.
 47. Quinlan A.R., Hall I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* **26**, 841–842.
 48. Statnikov A., Henaff M., Lytkin N.I., Aliferis C.F. (2012) New methods for separating causes from effects in genomics data. *BMC Genomics.* **13**(Suppl. 8), S22.
 49. Pfister N., Bühlmann P., Schölkopf B., Peters J. (2018) Kernel-based tests for joint independence. *J. R. Stat. Soc. Series B Stat. Methodol.* **80**, 5–31.
 50. Zhang Y., Rohde C., Reinhardt R., Voelcker-Rehage C., Jeltsch A. (2009) Non-imprinted allele-specific DNA methylation on human autosomes. *Genome Biol.* **10**, R138.
 51. Shoemaker R., Deng J., Wang W., Zhang K. (2010) Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res.* **20**, 883–889.
 52. Rahmani E., Schweiger R., Rhead B., Criswell L.A., Barcellos L.F., Eskin E., Rosset S., Sankararaman S., Halperin E. (2019) Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. *Nat. Commun.* **10**, 3417.
 53. Hackett J.A., Sengupta R., Zyllicz J.J., Murakami K., Lee C., Down T.A., Surani M.A. (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science.* **339**, 448–452.
 54. Romanova E.I., Zubritskiy A.V., Lioznova A.V., Ogunleye A.J., Golotin V.A., Guts A.A., Lennartsson A., Demidov O.N., Medvedeva Y.A. (2022) RUNX1/CEBPA mutation in acute myeloid leukemia promotes hypermethylation and indicates for demethylation therapy. *Int. J. Mol. Sci.* **23**(19), 11413.
 55. Haerter J.O., Lövkqvist C., Dodd I.B., Sneppen K. (2014) Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res.* **42**, 2235–2244.
 56. Weinberg D.N., Papillon-Cavanagh S., Chen H., Yue Y., Chen X., Rajagopalan K.N., Horth C., McGuire J.T., Xu X., Nikbakht H., Lemiesz A.E., Marchione D.M., Marunde M.R., Meiners M.J., Cheek M.A., Keogh M.C., Bareke E., Djedid A., Harutyunyan A.S., Jabado N., Garcia B.A., Li H., Allis C.D., Majewski J., Lu C. (2019) The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature.* **573**, 281–286.
 57. Emperle M., Bangalore D.M., Adam S., Kunert S., Heil H.S., Heinze K.G., Bashtrykov P., Tessmer I., Jeltsch A. (2021) Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase. *Nucleic Acids Res.* **49**, 8294–8308.
 58. Jurkowska R.Z., Anspach N., Urbanke C., Jia D., Reinhardt R., Nellen W., Cheng X., Jeltsch A. (2008) Formation of nucleoprotein filaments by mammalian

- an DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic Acids Res.* **36**, 6656–6663.
59. Jia D., Jurkowska R.Z., Zhang X., Jeltsch A., Cheng X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*. **449**, 248–251.
 60. Glass J.L., Fazzari M.J., Ferguson-Smith A.C., Greally J.M. (2009) CG dinucleotide periodicities recognized by the Dnmt3a-Dnmt3L complex are distinctive at retroelements and imprinted domains. *Mamm. Genome*. **20**, 633–643.
 61. Ginno P.A., Gaidatzis D., Feldmann A., Hoerner L., Imanci D., Burger L., Zilbermann F., Peters A.H.F.M., Edenhofer F., Smallwood S.A., Krebs A.R., Schübeler D. (2020) A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat. Commun.* **11**, 1–16.
 62. Suzuki T., Maeda S., Furuhashi E., Shimizu Y., Nishimura H., Kishima M., Suzuki H. (2017) A screening system to identify transcription factors that induce binding site-directed DNA demethylation. *Epigenetics Chromatin*. **10**, 60.
 63. Suzuki T., Shimizu Y., Furuhashi E., Maeda S., Kishima M., Nishimura H., Enomoto S., Hayashizaki Y., Suzuki H. (2017) RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. *Blood Adv.* **1**, 1699–1711.
 64. Thurman R.E., Rynes E., Humbert R., Vierstra J., Maurano M.T., Haugen E., Sheffield N.C., Stergachis A.B., Wang H., Vernot B., Garg K., John S., Sandstrom R., Bates D., Boatman L., Canfield T.K., Diegel M., Dunn D., Ebersol A.K., Frum T., Giste E., Johnson A.K., Johnson E.M., Kutayavin T., Lajoie B., Lee B.K., Lee K., London D., Lotakis D., Neph S., Neri F., Nguyen E.D., Qu H., Reynolds A.P., Roach V., Safi A., Sanchez M.E., Sanyal A., Shaffer A., Simon J.M., Song L., Vong S., Weaver M., Yan Y., Zhang Z., Zhang Z., Lenhard B., Tewari M., Dorschner M.O., Hansen R.S., Navas P.A., Stamatoyannopoulos G., Iyer V.R., Lieb J.D., Sunyaev S.R., Akey J.M., Sabo P.J., Kaul R., Furey T.S., Dekker J., Crawford G.E., Stamatoyannopoulos J.A. (2012) The accessible chromatin landscape of the human genome. *Nature*. **489**, 75–82.
 65. Wang H., Maurano M.T., Qu H., Varley K.E., Gertz J., Pauli F., Lee K., Canfield T., Weaver M., Sandstrom R., Thurman R.E., Kaul R., Myers R.M., Stamatoyannopoulos J.A. (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688.
 66. de Mendoza A., Nguyen T.V., Ford E., Poppe D., Buckberry S., Pflueger J., Grimmer M.R., Stolzenburg S., Bogdanovic O., Oshlack A., Farnham P.J., Blancafort P., Lister R. (2022) Large-scale manipulation of promoter DNA methylation reveals context-specific transcriptional responses and stability. *Genome Biol.* **23**, 1–31.

CpG Traffic Lights are Involved in Active DNA Demethylation

A. V. Lioznova¹, Yu. A. Medvedeva^{1, *}

¹Federal Research Centre “Fundamentals of Biotechnology”,
Russian Academy of Sciences, Moscow, 119071 Russia

*e-mail: ju.medvedeva@gmail.com

DNA methylation is one of the most important mechanisms involved in the epigenetic regulation of gene expression. However, the relationship between DNA methylation and expression is still not fully understood. There are examples where changes in DNA methylation level cause changes in gene expression, and vice versa – changes in expression entail changes in the methylation level. Earlier, we introduced the concept of CpG traffic lights, individual CpG sites whose methylation significantly correlates with expression, and showed their important role in enhancer regulation. Here, we showed that the methylation levels of CpG traffic lights are heterogeneous in the cell population and suggested that this is due to their dynamic demethylation. The observed enrichment of CpG traffic lights with 5-hydroxymethylcytosine (5hmC) and TET2 (Tet methylcytosine dioxygenase 2) localization sites has now confirmed our hypothesis. In order to find out whether the methylation of CpG sites is a cause or a consequence of the expression of the corresponding gene, we applied the method of causal inference. As a result, among the CpG sites, we distinguished those for which methylation was the cause of expression changes and those for which expression changes caused methylation changes. CpG sites of the first type were characterized by more stable methylation levels in different cells and less pronounced demethylation compared to CpG sites of the second type. It was also shown that the proportion of CpG sites whose methylation affected expression was greater in promoter regions than in the gene body, for which methylation was likely to be a consequence of expression. Based on these observations, we can assume that the methylation levels of CpG sites, which determine the expression of the gene associated with them, are stable and work on the principle of a “switch”. Conversely, for expression-dependent CpG sites, methylation levels are dynamic and vary between cells in the population, primarily due to active demethylation.

Keywords: DNA methylation, CpG sites, gene expression, epigenetic regulation, causal inference