

Устранение экспериментальных отклонений при количественном анализе профилей метилирования ДНК посредством высокопроизводительного бисульфитного секвенирования и гибридизации на олигонуклеотидных ДНК-чипах*

Москалев Е.А.^{1,4}, Завгородний М.Г.², Майорова С.П.³, Лебедев И.Н.⁴

¹ — Институт патологии, Университет Фридриха-Александра в Эрлангене и Нюрнберге, г.Эрланген, Германия, moskalyov@mail.ru

² — Кафедра функционального анализа и операторных уравнений, Воронежский государственный университет, г.Воронеж, Россия, mzavgorodnij@yandex.ru

³ — Кафедра высшей математики и физико-математического моделирования, Воронежский государственный технический университет, г.Воронеж, Россия, spmajorova@yandex.ru

⁴ — Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский медицинский центр Российской академии наук, г.Томск, Россия, igor.lebedev@medgenetics.ru

Количественный анализ профилей метилирования ДНК является неотъемлемым аспектом исследований в области эпигенетических механизмов онкогенеза. Кроме того, детекция опухолеспецифичных эпигенетических аномалий обладает высоким диагностическим потенциалом. При анализе генов-кандидатов, как правило, используют ПЦР-амплификацию модифицированных бисульфитом генных областей. Поскольку измерение степени метилирования CpG-динуклеотидов происходит в ампликонах, ПЦР определяет точность всего анализа. Избирательная амплификация форм ДНК, различающихся характером метилирования, может затруднять интерпретацию результатов. Дополнительные экспериментальные искажения могут быть связаны с особенностями технологий детекции метилированных оснований, например специфичностью олигонуклеотидных зондов ДНК-чипов. Целью работы было исследование применимости предложенного нами ранее алгоритма для устранения экспериментальных искажений из количественных данных метилирования ДНК, которые получены методами высокопроизводительного бисульфитного секвенирования и гибридизации на олигонуклеотидных ДНК-чипах. **Материалы и методы.** Для количественной характеристики профилей метилирования ДНК вблизи точек начала транскрипции генов-кандидатов использованы методы высокопроизводительного бисульфитного секвенирования по технологии 454 и гибридизации на ДНК-чипах «Febit Biotech». Регрессионный анализ полученных величин метилирования ДНК проводили посредством кубических полиномиальных кривых. **Результаты.** Для анализа были выбраны гены, аномальное гиперметилирование которых обнаруживается в злокачественных новообразованиях: кодирующие субъединицы фермента сукцинатдегидрогеназы *SDH*, а также онкосупрессоры *CDKN2B*, *DAPK1* и *TP53*. Анализируемые регионы амплифицировали в контрольных пробах ДНК известной степени метилирования. Амплификация и последующее секвенирование ампликонов *SDHB* и *SDHD* выявили значительные отклонения от ожидаемых величин метилирования. Аналогично индексы метилирования рассмотренных CpG-динуклеотидов в генах *CDKN2B* и *DAPK1* существенно искажались при гибридизации на ДНК-чипе. С помощью уравнений полученных кубических регрессионных кривых была проведена корректировка исходных данных с практически полным устранением артефактов. **Выводы.** Применение кубической полиномиальной регрессии для корректировки экспериментальных искажений в двух разных типах количественных данных метилирования ДНК — полученных высокопроизводительным бисульфитным секвенированием и гибридизацией на олигонуклеотидных ДНК-чипах — продемонстрировало практически полное устранение артефактов, и, следовательно, универсальность метода.

Ключевые слова: метилирование ДНК, количественный анализ, ПЦР-отклонение, кубическая регрессия, высокопроизводительное бисульфитное секвенирование, олигонуклеотидные ДНК-чипы

* Авторы не имеют конфликта интересов.

Исследование поддержано научно-исследовательской стипендией Германской службы академических обменов (DAAD) и Министерства образования и науки РФ по программе «Михаил Ломоносов» для Е.А.М.

Авторы благодарят Йорга Хоайзела и Флориана Халлера за ценные обсуждения и техническую поддержку по теме исследования.

Введение

Опухолеспецифичные профили метилирования ДНК являются наиболее интенсивно исследуемым компонентом эпигенетических механизмов, которые вносят значительный вклад в образование и развитие опухолей. Особую актуальность исследование онко-ассоциированных эпигенетических изменений имеет с практической точки зрения для их детекции в качестве потенциальных онкомаркеров [1, 2]. Поскольку aberrантное метилирование отдельных локусов происходит на ранних этапах онкогенеза и является намного более частым событием, чем онкогенные мутации, а модификация остатков цитидина метильными группами химически стабильна, определение метилированных форм ДНК является перспективным направлением онкодиагностики.

Эпигенетические биомаркеры могут служить не только для детекции опухолей, но и для выявления молекулярно-генетических подгрупп, различающихся прогнозом течения заболевания и ответом на терапию (прогностические маркеры). Это в особенности актуально для новообразований с недостаточной полнотой охарактеризованными молекулярными «портретами», прогноз которых по-прежнему строится с использованием традиционных морфологических критериев, прогностическая способность которых недостаточно точна. Так, например, риск рецидива для гастроинтестинальных стромальных опухолей (ГИСО) по-прежнему оценивается на основе размера, локализации и митотической активности опухоли [3], что обуславливает необходимость в молекулярно-генетических маркерах, выявляемых доступными в клинике методами. Один из недавно обнаруженных эпигенетических маркеров — степень метилирования единичного CpG-динуклеотида, расположенного в положении -647 по отношению к гену секретируемого фосфолипидпротеина 1 *SPP1* — является независимым прогностическим признаком при ГИСО [4]. Высокий (>25%) уровень метилирования данного гена соответствует благоприятному прогнозу заболевания, в то время как низкая (<25%) степень метилирования коррелирует с низкой выживаемостью больных. Интересно, что детекция уровня метилирования гена *SPP1* обеспечивает более точную стратификацию больных ГИСО в группы риска развития рецидива, чем применяемые в настоящее время морфологические критерии. Например, несмотря на высокую митотическую активность в одном из ранее описанных образцов ГИСО (классический фактор неблагоприятного прогноза) прогрессии данной опухоли не наблюдалось на протяжении более семи лет, что коррелировало с гиперметилированием данного гена (89,9%) [4].

Использование профилей метилирования ДНК или степени метилирования единичных CpG-динуклеотидов в качестве биомаркеров было бы ценно и для дискриминации фенотипически сходных нозологических форм. Например, уровень метилирования гена *MLH1*,

кодирующего компонент системы репарации неправильно спаренных оснований, уже сейчас используется на практике для дифференциальной диагностики наследственных и спорадических форм рака прямой и толстой кишки [5]. Наследственный неполипозный рак толстой кишки, или синдром Линча, развивается в результате герминальных мутаций в одном из генов, обеспечивающих репарацию неправильно спаренных оснований, в том числе в гене *MLH1*, и сопровождается микросателлитной нестабильностью. Классический подход к выявлению наследственной формы рака прямой и толстой кишки состоит в детекции микросателлитной нестабильности, а также иммуногистохимическом исследовании компонентов системы репарации ДНК. Однако микросателлитная нестабильность характерна примерно для 20% спорадических злокачественных новообразований прямой и толстой кишки, что обуславливает необходимость дополнительных молекулярных маркеров [6]. Показана чёткая ассоциация гиперметилирования гена *MLH1*, приводящего к потере его экспрессии, с микросателлитной нестабильностью при спорадической форме рака. Определение характера метилирования данного гена в опухолевом материале, а также «жидких биопсий» уже внедрено в молекулярную клиническую диагностику.

Выявление генов, уровень метилирования которых различается в опухолях при редком заболевании триаде Карнея и при наследственном синдроме Карнея—Стратакиса позволил бы обойтись без трудозатратного анализа герминальных мутаций в генах четырёх субъединиц сукцинатдегидрогеназы методом массивного параллельного секвенирования. Данные заболевания характеризуются манифестацией первичных опухолей, а именно ГИСО и параганглиомы (характерны для обоих типов онкопатологии), а также дополняющей их при триаде Карнея пульмонарной хондромы. Для наследственного синдрома Карнея—Стратакиса свойственны герминальные инактивирующие мутации в субъединицах фермента цикла Кребса сукцинатдегидрогеназы, в то время как при триаде Карнея данных мутаций не обнаружено. Ранее мы описали возможный эпигенетический механизм инактивации экспрессии гена, кодирующего субъединицу С сукцинатдегидрогеназы *SDHC*, при триаде Карнея посредством гиперметилирования его промоторной области [7]. Важно, что данная эпигенетическая aberrация отсутствует при синдроме Карнея—Стратакиса, и уровень метилирования единичного CpG-динуклеотида из данной области позволяет дифференцировать данные заболевания.

Таким образом, описанные примеры, а также постоянно растущий список эпигенетически инактивированных генов при злокачественных новообразованиях различных локализаций свидетельствуют о том, что детекция уровня метилирования маркерных генов могла бы стать существенным дополнением к принятым способам онкодиагностики и оценки прогноза.

Степень метилирования исследуемого локуса или отдельного CpG-динуклеотида в биологической пробе является количественной величиной, которая может принимать любые значения в диапазоне от 0 до 100%. Поэтому, например, в отличие от анализа однонуклеотидных полиморфизмов для изучения характера метилирования ДНК предпочтительны методы количественного исследования. В настоящее время применяется целый ряд таких методов, включающих, к примеру, бисульфитное пиросеквенирование, высокопроизводительное бисульфитное секвенирование, анализ на олигонуклеотидных ДНК-чипах и метилспецифичную ПЦР в реальном времени. Большинство методов используют ДНК, модифицированную бисульфитом [8]. Данная химическая обработка позволяет превратить эпигенетическое различие между C и ^5mC в генетический вариант U или ^5mC вследствие избирательного дезаминирования неметилированных остатков цитозина в урацил под действием бисульфита, в то время как 5-метилцитозин в реакцию не вступает. Одним из часто используемых этапов пробоподготовки является ПЦР-амплификация исследуемых регионов, необходимая для повышения чувствительности детекции. Поскольку в действительности измерение степени CpG-метилирования происходит не в геномной ДНК, а в ампликонах, этап ПЦР-амплификации критически важен для точности всего анализа.

В ходе ПЦР-амплификации модифицированной бисульфитом ДНК нередко происходит избирательное обогащение неметилированной или, реже, метилированной формы исследуемого участка. Данный эффект, получивший название ПЦР-отклонения (англ. PCR-bias), был впервые описан в 1997 г. [9]. Механизм эффекта, по-видимому, состоит в разной эффективности амплификации метилированной и неметилированной форм ДНК, различающихся вследствие разного GC-состава температурой плавления и свободной энергией образования вторичных структур. Размер данного отклонения трудно предсказать. При этом амплификация отдельных регионов может сопровождаться очень значительным отклонением, ведущим к неверной интерпретации экспериментальных данных о степени метилирования ДНК в биологических пробах. Например, ранее мы показали значительное ПЦР-отклонение при анализе профиля метилирования гена *SFRP1*, кодирующего ингибитор онкогенного сигнального пути Wnt. Так, анализ контрольной пробы ДНК со степенью метилирования всего 12,5% (практически полностью неметилированной) приводил к ложноположительному результату о степени метилирования гена *SFRP1* 60%, что может быть ошибочно интерпретировано как высокий уровень метилирования гена [10]. И хотя ПЦР-амплификация большинства исследуемых регионов не сопровождается настолько экстремальной избирательностью по отношению к метилированным или неметилированным формам, ПЦР-отклонение является частым событием.

Предложенные ранее подходы к устранению ПЦР-отклонения сводились к тщательной оптимизации условий амплификации для каждого анализируемого региона, включая температурный режим амплификации [11], а также введение CpG-динуклеотидов в состав праймеров для ПЦР [12]. Данный локуспецифичный подход, однако, существенно увеличивает трудоёмкость анализа, не всегда обеспечивая желаемую точность результатов. Для одновременного анализа множества генов-кандидатов или проведения масштабного анализа оптимизация условий проведения амплификации тем более технически затруднительна.

Ранее мы предложили универсальный метод, позволяющий с помощью контрольных проб ДНК известной степени метилирования и регрессионного анализа не предотвращать, но выявлять и практически полностью устранять экспериментальные искажения измерений для любого региона ДНК [10]. В калибровочных пробах ДНК, получаемых смешением полностью метилированного и неметилированного образцов обработанной бисульфитом ДНК, проводили амплификацию исследуемых регионов и измеряли уровень метилирования CpG-динуклеотидов посредством пиросеквенирования. Данные отдельных CpG-динуклеотидов или средние значения степени метилирования многих CpG-динуклеотидов исследуемого участка для каждой калибровочной пробы откладывали в координатах «реальный процент метилирования» и «экспериментальный процент метилирования». По нашим данным [10], среди семейств линейных функций, квадратных и кубических парабол, гипербола регрессионная кривая на основе кубической регрессии наиболее точно описывала величины метилирования ДНК в контрольных образцах и позволяла практически полностью устранить экспериментальные искажения.

Используя метод бисульфитного пиросеквенирования, мы предположили, что алгоритм корректировки величин метилирования, полученных в эксперименте, может быть применён безотносительно к конкретному методу количественного анализа CpG-метилирования. Однако в дополнение к ПЦР-отклонению конкретный метод детекции метилированных цитозинов может вносить дополнительные искажения. Поэтому применимость предложенного алгоритма для двух других широко используемых методов количественного анализа метилирования ДНК рассмотрена в данной работе.

Цель и задачи

Цель работы состояла в исследовании применимости алгоритма устранения экспериментальных искажений данных, полученных методами высокопроизводительного бисульфитного секвенирования и гибридизации на олигонуклеотидных ДНК-чипах.

Задачами работы было измерение уровня метилирования ДНК некоторых генов-кандидатов с онко-ассо-

цированной функцией в контрольных пробах ДНК известной степени метилирования посредством указанных методов, применение кубической регрессии для корректировки полученных экспериментальных данных, и оценки применимости данного алгоритма посредством сравнения исходных и скорректированных экспериментальных данных.

Материалы и методы

Контрольные пробы ДНК известной степени метилирования — 0, 25, 50, 75 и 100% — получали смешиванием полностью метилированной и неметилированной ДНК (EpiTect PCR control DNA, «Qiagen»), обработанной бисульфитом. Участки, расположенные вблизи точек начала транскрипции генов *SDHB* и *SDHD*, которые кодируют субъединицы фермента сукцинатдегидрогеназы, амплифицировали методом ПЦР в 50 мкл реакционной смеси, содержащей фермент Taq ДНК-полимеразу с «горячим стартом» (HotStarTaq, «Qiagen»), буфер в 10x разведении, 1,5 мМ MgCl₂, 200 мкМ каждого dNTP, 0,3 мкМ каждого праймера, 0,52 Ед. фермента. Программа амплификации включала этап активации ДНК-полимеразы при 95°C в течение 15 мин, 1 мин денатурации при 95°C, 2 мин отжига праймеров при 60°C и элонгацию в течение 3 мин. При данном температурном режиме проводили 20 циклов амплификации, при этом в каждом последующем цикле температура отжига снижалась на 0,5°C. Дальнейшие 40 циклов амплификации проводили, денатурируя ДНК при 95°C в течение 1 мин и последующем отжиге праймеров при фиксированной температуре 50°C в течение 2 мин и элонгации в течение 2 мин. Условия проведения реакции и программа амплификации гена *SDHC* отличались температурой отжига праймеров, составлявшей в первой и второй частях программы 61°C и 51°C соответственно. Амплификацию ге-

нов *CDKN2B*, *DAPK1* и *TP53* проводили аналогично ранее опубликованной схеме [13]. Нуклеотидные последовательности праймеров представлены в таблице.

Ампликоны генов *SDHB*, *SDHC* и *SDHD* очищали с использованием набора Agencourt AMPure XP («Beckman Coulter», США). Их концентрацию измеряли флуориметрическим методом с использованием набора Quant-iT PicoGreen dsDNA Assay Kit («Life Technologies», США) на анализаторе Synergy 2 Multi-Mode Microplate Reader («Biotek», США). Высокопроизводительное секвенирование проводили по технологии 454, («Roche», Германия) по описанному ранее протоколу [6]. Полученные нуклеотидные последовательности сравнивали с конвертированными бисульфитом последовательностями генов в программе Amplicon Variant Analyser Software v. 2.5 («Roche»). Процент метилирования CpG-динуклеотидов в пробах вычисляли с помощью программы BISMA [13].

Синтез олигонуклеотидных ДНК-чипов *in situ* проводили на экспериментальной установке Geniom One («Febit Biotech AG», Германия) в соответствии с ранее опубликованным протоколом [15]. Пробоподготовку меченной биотином ДНК, гибридизацию, отмывание, детекцию и анализ флуоресцентных сигналов осуществляли, как описано ранее [16].

Результаты и обсуждение

Для проверки предположения об универсальности предложенного ранее метода корректировки величин метилирования ДНК, получаемых в эксперименте, и возможности его применения не только для данных пиросеквенирования, но и для других методов количественного анализа мы рассмотрели экспериментальные данные высокопроизводительного бисульфитного секвенирования по технологии 454, а также результаты гиб-

Таблица

Нуклеотидные последовательности праймеров, по [6, 12]

Ген	Нуклеотидная последовательность праймеров (5'-3'), F: прямой; R: обратный	Длина ампликона, п.н.
SDHB-F	TAGTTTGTTAAGATGATGAAATTT	316
SDHB-R	CTTTCAAAAATAAACTAAACTTAAATA	
SDHC-F	GAAAAATAATTAGTAAATTAGTTAGGTAG	414
SDHC-R	ACTAAATCACCTCAACAACAAC	
SDHD-F	TATTAAGGAAGGTGAAATTTTTTTT	317
SDHD-R	TCCTAAAACTCAAATCATCCAC	
CDKN2B-F	GTTTTTTGGTTTAGWTGAAAAGGGAATTTTTTGT	575
CDKN2B-R	AACCCATAAACCCCAACTACCTAAATC	
DAPK1-F	GGAATATAGTTAGGGAGTGAGTG	152
DAPK1-R	CTACTCCAAATATAAAAAACAACC	
TP53-F	TTGATGAGAAGAAAGGATTTAGTTGA	484
TP53-R	AAAAACTTACCCAATCCAAAAAAC	

ридизации на олигонуклеотидных ДНК-чипах. Выбор генов-кандидатов для исследования был обусловлен их онко-ассоциированной функцией и, следовательно, потенциальным практическим применением получаемых данных метилирования. Для секвенирования были выбраны гены, кодирующие субъединицы фермента сукцинатдегидрогеназы *SDHB*, *SDHC* и *SDHD*, аномальное гиперметилирование которых в опухолях Карни было показано нами ранее [7]. В качестве контрольных проб были использованы образцы ДНК 0, 25, 50, 75 и 100% метилирования. Секвенирование амплифицированных участков данных генов, расположенных в непосредственной близости от точек начала транскрипции позволило охватить 15, 13 и 25 последовательных CpG-динуклеотидов исследуемой области генов *SDHB*, *SDHC* и *SDHD* соответственно. Вычисленные на основе данных секвенирования проценты метилирования CpG-динуклеотидов в гене *SDHB* существенно отличались от ожидаемых (рис. 1А). Несколько меньшее отклонение экспериментальных величин было показано при анализе характера метилирования гена *SDHD*, и практически отсутствовало для гена *SDHC* (рис. 1Б, В). Кубические полиномиальные регрессионные кривые описывали калибровочные данные с высокой точностью с коэффициентами детерминации R^2 , близкими к 1. Использование уравнений полученных регрессионных кривых позволило скорректировать экспериментальные данные, практически полностью устранив экспериментальные искажения.

Аналогичные результаты, демонстрирующие применимость предложенного метода корректировки, были получены при анализе величин метилирования ДНК, измеренных методом гибридизации на олигонуклеотидных ДНК-чипах. Гибридизационный анализ на ДНК-чипах сопряжён с возникновением большего числа экспериментальных искажений (Рис. 2А). В отличие от количественных данных, полученных методами секвенирования, для которых основной причиной искажений является, по-видимому, обогащение неметилированных или метилированных форм ДНК в ходе амплификации, большой вклад в точность анализа на ДНК-чипах вносит специфичность функционирования олигонуклеотидных зондов, т.е. их способность дискриминировать метилированный и неметилированный статус исследуемых CpG-динуклеотидов. Дополнительным фактором, снижающим точность анализа, может быть недостаточная специфичность гибридизации ДНК-мишени. Традиционный подход к решению данной проблемы состоит в тщательном отборе олигонуклеотидных зондов и использовании лишь тех олигонуклеотидов, которые способны высокоточно дискриминировать калибровочные образцы ДНК известной степени метилирования, например, 0, 25, 50, 75 и 100% метилирования [16]. Недостатком данного подхода является относительно малое число зондов, которые преодолевают данный фильтр. Поскольку каждый олигонуклеотидный

зонд предназначен для измерения степени метилирования одного или нескольких расположенных рядом CpG-динуклеотидов, потеря существенной доли зондов при их отборе приводит к значительному снижению разрешения анализа профилей метилирования ДНК. Возможность не исключать значительную часть зондов, но корректировать получаемые данные, могла бы обеспечить большее разрешение и, таким образом, большую информативность анализа.

Мы рассмотрели несколько типичных олигонуклеотидных зондов ДНК-чипа, гибридизация ДНК-пробы с которыми сопряжена со значительными экспериментальными искажениями. Примеры зондов для анализа индекса метилирования CpG-динуклеотидов в онко-ассоциированных генах *CDKN2B* и *DAPK1* представлены на рис. 2Б, В. В каждом случае использование алгоритма корректировки позволяло практически полностью устранять экспериментальные искажения.

Выводы

Таким образом, применение предложенного нами ранее метода корректировки экспериментальных искажений для двух разных типов количественных данных, полученных высокопроизводительным бисульфитным секвенированием и гибридизацией на олигонуклеотидных ДНК-чипах, продемонстрировало практически полное устранение артефактов, и следовательно, универсальность метода.

Список литературы

1. Пономарева А.А., Рыкова Е.Ю., Чердынцева Н.В. и др. Анализ уровень метилирования гена *RARβ2* в циркулирующих ДНК крови у больных раком легкого как потенциального прогностического маркера. Вопросы онкологии. 2011;57(3):302-307.
2. Брызгунова О.Е., Лактионов П.П. Формирование пула циркулирующих ДНК крови: источники, особенности строения и циркуляции. Биомедицинская химия. 2015;61(4):409-426.
3. Miettinen M., Lasota J. Gastrointestinal stromal tumors: pathology and prognosis at different sites. *Semin Diagn Pathol*. 2006;23(2):70-83.
4. Haller F., Zhang J.D., Moskalev E.A. et al. Combined DNA methylation and gene expression profiling in gastrointestinal stromal tumors (GISTs) reveals hypomethylation of *SPP1* as an independent prognostic factor. *International Journal of Cancer*. 2015;136(5):1013-1023.
5. Gausachs M., Pilar M., Corral J. et al. *MLH1* promoter hypermethylation in the analytical algorithm of Lynch syndrome & a cost-effectiveness study. *Eur J Hum Genet*. 2012; 20(7):762-768.
6. Bettstetter M., Dechant S., Ruemmele P. et al. Distinction of hereditary nonpolyposis colorectal cancer and sporadic microsatellite-unstable colorectal cancer through quantification of *MLH1* methylation by real-time PCR. *Clin Cancer Res*. 2007;13:3221-3228.

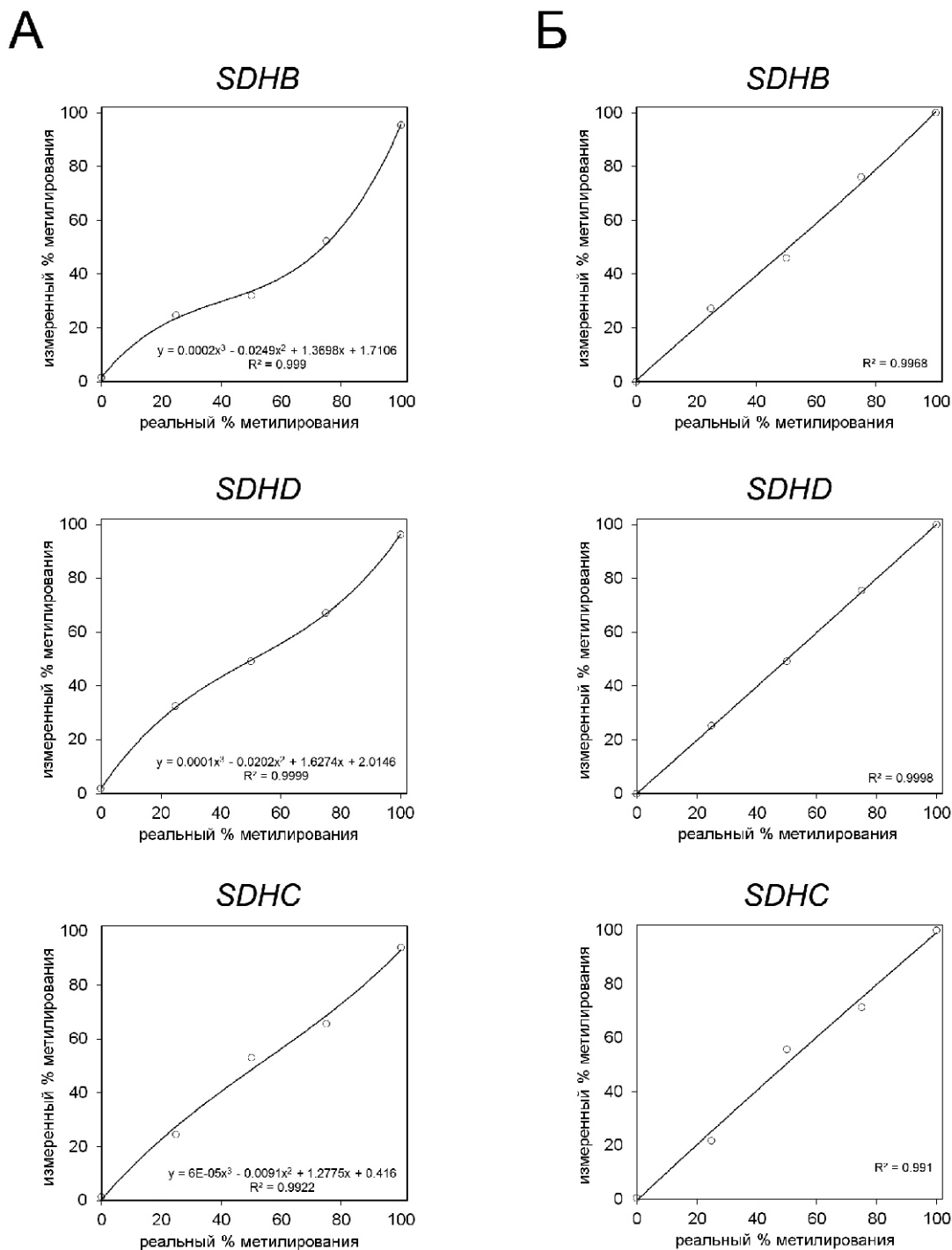


Рис. 1. Репрезентативные примеры экспериментальных отклонений, полученные при анализе степени метилирования ДНК методом высокопроизводительного бисульфитного секвенирования.

(А) Для визуализации экспериментальных отклонений степень метилирования исследованных регионов в контрольных пробах («измеренный % метилирования», ось ординат) представлен как функция реального процента метилирования в геномной ДНК (ось абсцисс). Каждая величина метилирования является средним уровнем метилирования 15, 13 и 25 последовательных CpG-динуклеотидов исследуемой области генов *SDHB*, *SDHC* и *SDHD*. Показаны примеры сильного и умеренного искажения для генов *SDHB* и *SDHD*, соответственно, а также практически полное отсутствие отклонений для гена *SDHC*. На графиках представлены кубические полиномиальные регрессионные кривые, описывающие экспериментальные величины.

(Б) Результаты корректировки экспериментальных величин из панели (А) посредством кубической регрессии. Линейный характер функции $y(x)$ демонстрирует эффективное устранение искажений из экспериментальных величин.

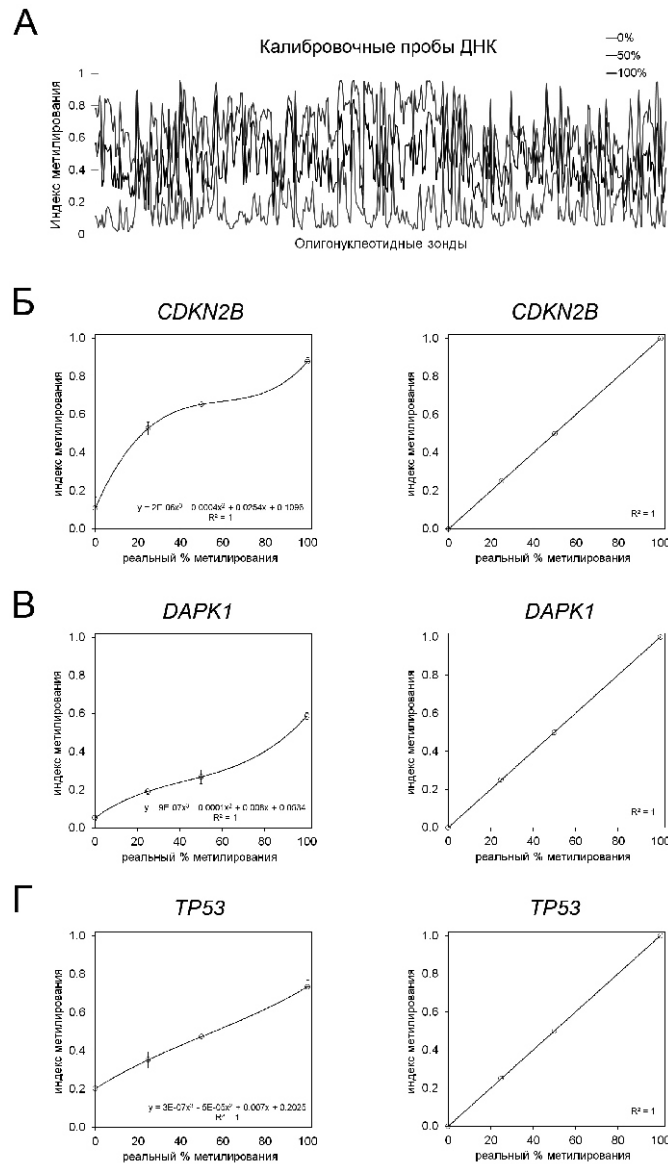


Рис. 2. Типичные примеры искажений, полученные при гибридационном анализе метилирования ДНК на олигонуклеотидных ДНК-чипах. (А) Способность олигонуклеотидных зондов чипа дискриминировать контрольные образцы ДНК разной степени метилирования. Кривые описывают средние значения вычисленного индекса метилирования (ось ординат) для 421 анализируемых CpG-динуклеотидов (ось абсцисс). Степень метилирования каждого CpG-динуклеотида определяется парой зондов, один из которых полностью комплементарен метилированному (М-зонд) или неметилированному аллелю (U-зонд). Анализируемый CG-динуклеотид соответствовал центральному нуклеотиду зонда. Индекс метилирования вычислен как отношение флуоресцентного сигнала от М-зонда к сумме интенсивностей сигналов от пары зондов — индекс метилирования, $I_M/(I_M + I_U)$, где I_M и I_U — интенсивности флуоресцентных сигналов от М- и U-зондов соответственно. Каждая пара зондов представлена на ДНК-чипе в шести повторностях. Как видно из диаграммы, способность олигонуклеотидных зондов дискриминировать контрольные пробы ДНК различается. (Б), (В), (Г) Репрезентативные примеры пар олигонуклеотидных зондов. Степень метилирования исследованных CpG-динуклеотидов в контрольных пробах («индекс метилирования», ось ординат) представлен как функция реального процента метилирования в геномной ДНК (ось абсцисс). На графиках представлены кубические полиномиальные регрессионные кривые, описывающие экспериментальные величины. Полосы погрешностей соответствуют стандартным отклонениям от шести повторностей. (Б) Пара зондов для анализа трёх CpG-динуклеотидов гена *CDKN2B* способна достаточно точно детектировать калибровочные пробы ДНК с крайними значениями процента метилирования, однако зависимость $y(x)$ сильно отличается от линейной. (В) Пара зондов для измерения индекса метилирования двух CpG-динуклеотидов в гене *DAPK1* обеспечивает близкую к линейной зависимость $y(x)$, но существенно занижает измеряемую величину метилирования для калибровочной пробы 100% метилирования. (Г) Пара зондов для измерения индекса метилирования двух CpG-динуклеотидов в гене *TP53* демонстрирует практически линейную зависимость $y(x)$, однако имеет место существенное искажение измеряемых величин метилирования для обеих крайних калибровочных проб ДНК 0 и 100% метилирования.

7. Haller F., Moskalev E.A., Faucz F.R. et al. Aberrant DNA hypermethylation of SDHC: a novel mechanism of tumor development in Carney Triad. *Endocrine Related Cancer*. 2014;21(4):567-577.

8. Frommer M., McDonald L.E., Millar D.S. et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA*. 1992;89(5):1827-1831.

9. Warnecke P.M., Storzaker C., Melki J.R. et al. Detection and measurement of PCR bias in quantitative methylation analysis of bisulfite-treated DNA. *Nucleic Acids Research*. 1997;25(21):4422-4426.

10. Moskalev E.A., Zavgorodnij M.G., Majorova S.P. et al. Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression. *Nucleic Acids Research*. 2011;39(11):e77.

11. Shen L., Guo Y., Chen X. et al. Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. *Bio-Techniques*. 2007;42(1):48-58.

12. Wojdacz T.K., Hansen L.L., Dobrovic A. A new approach to primer design for the control of PCR bias in methylation studies. *BMC Res. Notes*. 2008;1:54.

13. Москалев Е.А. Аномальные картины метилирования геномной ДНК при хроническом В-клеточном лимфолейкозе: дис. канд. биол. наук. Воронежский государственный университет, Воронеж, 2007.

14. Rohde C., Zhang Y., Reinhardt R. et al. BISMA — fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics*. 2010;11:230.

15. Guimil R., Beier M., Scheffler M. et al. Geniom technology — the benchtop array facility. *Nucleosides Nucleotides Nucleic Acids*. 2003;22(5-8):1721-1723.

16. Mund C., Beier V., Bewerunge P. et al. Array-based analysis of genomic DNA methylation patterns of the tumour suppressor gene p16INK4A promoter in colon carcinoma cell lines. *Nucleic Acids Res*. 2005;33(8):e73.

Correction of measurement biases in DNA methylation studies by means of next generation sequencing and hybridisation on oligonucleotide microarrays

Moskalev E.A.^{1,4}, Zavgorodnij M.G.², Majorova S.P.³, Lebedev I.N.⁴

¹ — Institute of Pathology, Friedrich-Alexander-University of Erlangen-Nuremberg, Erlangen, Germany, moskalyov@mail.ru

² — Chair of Functional Analysis and Operational Equations, Voronezh State University, University, Voronezh, Russia, mzavgorodnij@yandex.ru

³ — Chair of Higher Mathematics and Physical and Mathematical Modeling, Voronezh State Technical University, Voronezh, Russia, spmajorova@yandex.ru

⁴ — Research Institute of Medical Genetics, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, Russia, igor.lebedev@medgenetics.ru

Quantitative analysis of DNA methylation patterns is an intrinsic part of epigenetic cancer research. Besides, detection of tumour specific epigenetic aberrations holds promise for cancer diagnostics. PCR amplification of loci of interest from bisulfite treated DNA is a common sample preparation step. Given the fact amplicons and not the genomic DNA is used for quantification of the methylation percentages, the PCR amplification step is limiting factor for the accuracy of the entire analysis. Preferential amplification of either methylated or unmethylated alleles may hamper the correct interpretation of results. Additional biases may be introduced by the quantification technologies on their own, for example the specificity of oligonucleotide probes of a microarray. Aim. The aim of present work was to evaluate the applicability of the cubic polynomial regression for correction of measurement biases in quantitative DNA methylation data that were obtained by the next generation sequencing and hybridisation analysis on the oligonucleotide microarrays. Materials and Methods. Next generation bisulfite 454 sequencing and hybridisation on the *in situ* synthesised microarrays (Febit Biotech) was employed to quantify DNA methylation in vicinity of the transcriptional start sites. Cubic polynomial regression was used to analyse the calibration data. Results. Candidate genes were selected that exhibit aberrant DNA methylation patterns in cancer. Specifically, these included genes coding for the subunits of succinate dehydrogenase (*SDH*) as well as tumour suppressor genes *CDKN2B*, *DAPK1* and *TP53*. The interrogated DNA loci were amplified from control DNA samples of defined methylation percentages. Substantial deviations of apparent methylation percentages from theoretically expected values were detected by the sequencing of *SDHB* and *SDHD* amplicons. Likewise, the methylation indices of the *CDKN2B* and *DAPK1* CpG sites that were analysed were significantly biased in the microarray data. Correction of biased data was performed by using the equations of respective regression curves resulting in almost complete removal of the artefacts. Conclusions. Successful application of cubic polynomial regression was achieved for correcting measurement biases in two different types of quantitative DNA methylation data — namely next generation bisulfite sequencing and hybridisation on the oligonucleotide microarrays — demonstrating the universal applicability of this correction process.

Key words: DNA methylation, quantitative analysis, PCR-bias, cubic polynomial regression, next-generation sequencing, oligonucleotide microarrays