



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

## Поиск дифференциально метилированных регионов в геномах древних и современных людей

Д.Д. Бородко , С.В. Женило, Ф.С. Шарко

Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Москва, Россия  
 [daria.borodko@gmail.com](mailto:daria.borodko@gmail.com)


**Аннотация.** В настоящее время активно исследуются механизмы, регулирующие развитие различных патологий и их эволюционную динамику. Эпигенетические механизмы, такие как метилирование, играют значимую роль в эволюционных процессах, поскольку их изменения гораздо быстрее отражаются на фенотипе, чем результаты мутагенеза. В данном исследовании мы предприняли попытку разработать алгоритм для выявления дифференциально метилированных областей, связанных с метаболическим синдромом, которые изменили свое метилирование у человека при переходе от охоты и собирательства к оседлой жизни. Применение существующих методов полногеномного бисульфитного секвенирования ограничено для древних образцов из-за их низкого качества и фрагментации, и подход к получению профилей метилирования охотников-собирателей значительно отличается от подходов, используемых для современных тканей. В этой работе мы валидировали DamMet – алгоритм, реконструирующий древние метиломы. Применение DamMet к геномам неандертальца и денисовца показало средний уровень корреляции с профилями метилирования, опубликованными ранее, а также продемонстрировало занижение уровня метилирования реконструированных профилей в среднем на 15–20 %. Также мы разработали новый алгоритм на языке Python, позволяющий сравнивать метиломы в древних и современных образцах, несмотря на отсутствие профилей метилирования современных образцов костной ткани в контексте ожирения. Такой анализ подразумевает двухступенчатую обработку данных, где на первом этапе происходит идентификация тканеспецифичных областей метилирования и их фильтрация, а на втором этапе осуществляется непосредственно поиск дифференциально метилированных регионов в заданных областях, ассоциированных с интересующим исследователя заболеванием. В результате использования алгоритма на тестовых данных мы обнаружили 38 дифференциально метилированных регионов, ассоциированных с ожирением, большая часть которых принадлежала промоторным областям, и разработанный пайплайн показал достаточную эффективность в их поиске. Эти результаты подтверждают возможность восстановления профилей метилирования в древних образцах и их сравнения с современными метиломами. Также обсуждаются возможности дальнейшего развития методологии и внедрения нового шага, позволяющего изучать дифференциально метилированные позиции, связанные с эволюционными процессами.

Ключевые слова: древняя ДНК; метилирование; эпигенетика; DamMet; ДМР.

**Для цитирования:** Бородко Д.Д., Женило С.В., Шарко Ф.С. Поиск дифференциально метилированных регионов в геномах древних и современных людей. *Вавиловский журнал генетики и селекции*. 2023;27(7):820-828. DOI 10.18699/VJGB-23-95

## Search for differentially methylated regions in ancient and modern genomes

D.D. Borodko , S.V. Zhenilo, F.S. Sharko

Federal Research Center “Fundamentals of Biotechnology” of the Russian Academy of Sciences, Moscow, Russia  
 [daria.borodko@gmail.com](mailto:daria.borodko@gmail.com)

**Abstract.** Currently, active research is focused on investigating the mechanisms that regulate the development of various pathologies and their evolutionary dynamics. Epigenetic mechanisms, such as DNA methylation, play a significant role in evolutionary processes, as their changes have a faster impact on the phenotype compared to mutagenesis. In this study, we attempted to develop an algorithm for identifying differentially methylated regions associated with metabolic syndrome, which have undergone methylation changes in humans during the transition from a hunter-gatherer to a sedentary lifestyle. The application of existing whole-genome bisulfite sequencing methods is limited for ancient samples due to their low quality and fragmentation, and the approach to obtaining DNA methylation profiles differs significantly between ancient hunter-gatherer samples and modern tissues. In this study, we validated DamMet, an algorithm for reconstructing ancient methylomes. Application of DamMet to Neanderthal and Denisovan genomes showed a moderate level of correlation with previously published methylation profiles and demonstrated an underestimation of methylation levels in the reconstructed profiles by an average of 15–20 %. Additionally, we developed a new Python-based algorithm that allows for the comparison of methylomes in ancient and modern samples, despite the absence of methylation profiles in modern bone tissue within the context of obesity. This analysis involves a two-step data processing approach, where the first step involves the identification and

filtration of tissue-specific methylation regions, and the second step focuses on the direct search for differentially methylated regions in specific areas associated with the researcher's target condition. By applying this algorithm to test data, we identified 38 differentially methylated regions associated with obesity, the majority of which were located in promoter regions. The pipeline demonstrated sufficient efficiency in detecting these regions. These results confirm the feasibility of reconstructing DNA methylation profiles in ancient samples and comparing them with modern methylomes. Furthermore, possibilities for further methodological development and the implementation of a new step for studying differentially methylated positions associated with evolutionary processes are discussed.

Key words: ancient DNA; methylation; epigenetics; DamMet; DMR.

**For citation:** Borodko D.D., Zhenilo S.V., Sharko F.S. Search for differentially methylated regions in ancient and modern genomes. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):820-828. DOI 10.18699/VJGB-23-95

## Введение

В последнее время все большее внимание уделяется изучению механизмов, регулирующих развитие различных патологий и их эволюционную динамику (Briggs et al., 2009a; Niiranen et al., 2022). Особенно важную роль в этом процессе играют эпигенетические механизмы, такие как метилирование, поскольку они способны вызывать фенотипические изменения гораздо быстрее, чем обычные процессы мутагенеза (Jablonka, Raz, 2009; Feinberg, Irizarry, 2010; Zhur et al., 2021). Главной целью данного исследования было выявление дифференциально метилированных регионов (ДМР), связанных с метаболическим синдромом, которые могли бы стать потенциальными целями эпигенетической терапии метаболического синдрома.

Сейчас ученых нередко останавливает от проведения эволюционных исследований отсутствие подходящих методов для сравнения профилей ДНК древних и современных образцов. Лабораторные протоколы, применяемые для получения этих профилей, значительно отличаются друг от друга и имеют свои особенности и ошибки. Древняя ДНК (дДНК) часто находится во фрагментированном состоянии (Sawyer et al., 2012), и с течением времени происходит естественная деградация молекулы и спонтанное дезаминирование азотистых оснований, что ограничивает доступность качественных данных (Briggs et al., 2007, 2009b). Для решения этой проблемы был предложен специальный протокол обработки образцов с применением урацил-ДНК-гликозилазы (УДГ) и эндонуклеазы, чтобы облегчить извлечение профилей метилирования и повысить их различимость (Briggs et al., 2010), а также несколько программ, позволяющих рассчитать уровни метилирования в древних образцах, последовательности которых были секвенированы с применением УДГ (Gokhman et al., 2014; Orlando et al., 2015; Hanghøj et al., 2019).

Существует два алгоритма реконструкции метилирования древних образцов, которые были разработаны для командной строки и достаточно дружелюбны для пользователя. Более ранний алгоритм epiPALEOMIX основан на исторически первом способе реконструкции метиломов, опубликованном Д. Гохманом в 2014 г., и имеет несколько модулей, в том числе модуль MethylMap, позволяющий получить уровни метилирования в регионах, заданных пользователем (Hanghøj et al., 2016). Однако в этом заключается и ограничение по его применению: пользователю необходимо знать, какие регионы он исследует. Результатом работы алгоритма является рассчитанное ко-

личество дезаминированных метилированных цитозинов в CpG контексте и покрытие; соответственно, их отношение и будет уровнем метилирования в данной позиции генома. Алгоритм DamMet более универсален, поскольку, в отличие от epiPALEOMIX, ориентирован на полногеномные исследования. Помимо этого, DamMet способен рассчитывать уровни дезаминирования в метилированных и неметилированных CpG на каждой позиции ридов, т. е. использует модель, которая наиболее точно описывает процесс дезаминирования цитозинов во фрагментах дДНК как случайный (Hanghøj et al., 2019).

В настоящее время для исследования метилирования ДНК в современных образцах применяется также метод полногеномного бисульфитного секвенирования (WGBS) (Olova et al., 2018; Suzuki et al., 2018). Для восстановления метилирования из образцов, секвенированных с помощью этой технологии, существует несколько методов (Clark et al., 1994; Bock et al., 2005), из которых самыми известными считаются Bismark, BoostMe и WGBStools. Алгоритм Bismark наиболее часто используется для препроцессинга данных WGBS, что подразумевает картирование ридов на конвертированный референсный геном, а затем подсчет количества метилированных и неметилированных цитозинов на каждой позиции генома (Krueger, Andrews, 2011). Этот метод, как и большинство методов, основанных на подсчете ридов, недостаточно эффективно преодолевает проблему малого покрытия образцов, которая часто является при использовании образцов низкого качества или проведении single cell экспериментов. Для решения этой проблемы были разработаны алгоритмы на основе машинного обучения, такие как DeepCPG и BoostMe.

DeepCPG – это алгоритм на основе нейронной сети глубокого обучения, который прогнозирует состояния метилирования малопокрытых сайтов и обнаруживает мотивы, связанные с изменением уровней метилирования и межклеточной изменчивостью (Angermueller et al., 2017). Этот инструмент применяется для улучшения качества данных single cell экспериментов. BoostMe, базирующийся на методе машинного обучения, позволяет решить проблему на этапе препроцессинга генома с помощью импутации (Zou et al., 2018). Градиентный бустинг (XGBoost), имплементированный в этом инструменте, сочетает данные нескольких образцов (более трех) для коррекции пропущенных уровней метилирования в образцах современных тканей, что позволяет использовать для восстановления метилирования образцы с низким покрытием генома. Кроме того, преимуществом BoostMe является способность

восстанавливать не только состояние данного CpG (метилирован/не метилирован), но и уровень метилирования. WGBStools – набор методов, разработанный в рамках проекта по составлению атласа метилирования современных тканей. Используется для сверхкомпактного представления картированных прочтений, статистического анализа, а также для визуализации от небольших фрагментов до целых хромосомных локусов ([https://github.com/nloyfer/wgbs\\_tools](https://github.com/nloyfer/wgbs_tools)).

Несмотря на разнообразие алгоритмов восстановления метилирования, применение WGBS технологии для образцов дДНК ограничено, так как для бисульфитной конверсии требуется большая концентрация хорошо очищенной ДНК. Также при конверсии происходит фрагментация ДНК, что еще сильнее ухудшает качество дДНК, и так значительно фрагментированной из-за деградации (Gu et al., 2011). Следовательно, алгоритмы расчета уровней метилирования, применяемые для современных образцов, не могут использоваться для реконструкции профилей метилирования древних людей. Поэтому мы сосредоточились на разработке нового алгоритма, который позволил бы сравнивать метиломы в древних и современных образцах, учитывая отсутствие доступных образцов костной ткани для проведения полногеномного бисульфитного секвенирования в контексте ожирения.

## Материалы и методы

**Подготовка данных.** Для анализа мы отобрали в базе данных GEO NCBI 11 древних геномов и 12 современных профилей метилирования, полученных с помощью методов Whole Genome Bisulfite Sequencing. При выборе древних образцов мы обратили особое внимание на их возраст, стратегию подготовки библиотек и полученное покрытие по геному: использовались только образцы, полученные

с предварительной обработкой урацил-ДНК-гликозилазой (УДГ), не моложе 3 тыс. лет до н. э. и с покрытием не менее 5x. Полные геномы древних образцов были отсеквенированы с обработкой УДГ, кроме образцов Vi33 и PES001 (Peschanitsa), которые не были обработаны УДГ перед секвенированием (табл. 1).

Выбор 12 современных образцов (Loyfer et al., 2023) определен мезодермальным происхождением тканей, из которых были подготовлены библиотеки, а также применением метода полногеномного бисульфитного секвенирования. Данные об образцах приведены в табл. 2.

**Препроцессинг ридов древней ДНК.** Древние геномы были загружены в формате bam-файлов с индексами. Согласно (Ohm et al., 2010; Gokhman et al., 2014), УДГ недостаточно эффективно работает на концах ДНК. Соответственно, для корректного анализа дДНК при помощи утилиты trimBam мы убрали по два нуклеотида с 3'- и 5'-концов последовательностей (Gansauge, Meyer, 2013; Jun et al., 2015). Для образцов Vi33 и PES001 данная процедура не проводилась, поскольку образцы не были обработаны УДГ при подготовке библиотек. Помимо этого, все риды образцов были отфильтрованы с помощью Trimmomatic (Bolger et al., 2014) по среднему качеству и длине, и в дальнейшем анализе использовались только выровненные на сборку CRCh37 (hg19) последовательности со средним качеством более 20 и длиной более 25 п. о.

**Реконструкция профилей метилирования древних людей.** Для восстановления метиломов древних образцов мы использовали ПО DamMet (Hanghøj et al., 2019). Пайплайн включал в себя три этапа: фильтрацию однонуклеотидных вариантов, расчет уровней дезаминирования для каждой позиции ридов и расчет уровня метилирования. Расчет SNV проводился с помощью GATK HaplotypeCaller v4.3.0.0 (Poplin et al., 2017) с последующей фильтра-

**Таблица 1.** Древние геномы, отобранные для анализа

Образец	Группа	Возраст образца, тыс. лет	Пол	Ткань	Покрытие	Профиль метилирования	Геномное окно (в CpG)	Литературный источник
Altai Neanderthal	Древний	120	Ж	Фаланга стопы	50	Gokhman et al., 2014, 2020	25	Prüfer et al., 2014
Denisovan		75	Ж	Фаланга стопы	30		25	Meyer et al., 2012
Vindija33		50	Ж	Неопознанная кость	30		50	Prüfer et al., 2017
Ust'-Ishim	ОС	45	М	Бедренная кость	42 (22 XY)	Gokhman et al., 2020	25	Fu et al., 2014
Sunghir		35	М	Бедренная кость + зубы	10.7	Эта работа	38	Sikora et al., 2017
USR1		11.5	Ж	Височная кость	17		50	Moreno-Mayar et al., 2018a
Spirit Cave		11	М	Височная кость + зубы	18		33	Moreno-Mayar et al., 2018b
Peschanitsa		11	М	Зубы	5		50	Saag et al., 2021
SF12		9	Ж	Бедренная кость	57.79		28	Günther et al., 2018
2H10 (France)		3.2	М	Зубы	13.9		33	Seguin-Orlando et al., 2021
2H11 (France)		3.2	М	Зубы	23.9		33	Seguin-Orlando et al., 2021

Примечание. Окно сглаживания – параметр усреднения уровней дезаминирования на следующем этапе анализа; ОС – охотники-собиратели.

**Таблица 2.** Современные геномы, использованные для поиска тканеспецифично метилированных регионов и ДМР

GEO accession	Пол	Возраст пациента	Орган	Ткань
GSM5652198	М	37	Толстая кишка	Фибробласты
GSM5652202	Ж	35	Сердце	
GSM5652204	М	73	Дерма	
GSM5652205	Ж	59	Скелетная мышца	Гладкие миоциты
GSM5652207	М	22	Аорта	
GSM5652209	Ж	51	Мочевой пузырь	
GSM5652210	М	24	Простата	
GSM5652211	М	57	Легочный бронх	
GSM5652212	М	83	Сердце	Кардиомиоциты
GSM2637888	–	–	Кость	–
GSM2637887	–	–	Кость	–
GSM5652218	Ж	7	Кость	Остеобласты
GSM5652177	Ж	35	Подкожная жировая ткань	Адиipoциты
GSM5652176	Ж	53	Подкожная жировая ткань	
GSM5652178	Ж	37	Подкожная жировая ткань	

цией вариантов с покрытием ниже 5 и качеством ниже 30, а также при наличии гомозиготности по альтернативному аллелю или более чем 2 альтернативных аллелей, если на этой позиции находится цитозин. Данный этап был рекомендован автором алгоритма DamMet в публикации (Hanghøj et al., 2019) и приведен в дополнительных материалах источника.

Далее проводилась реконструкция уровней метилирования с исключением найденных вариантов.

```
DamMet estDEAM -b <bam-file> -r <fasta-file> -c <chromosome> -M <expected-average-methylation> -O <out-file-prefix> -E <vcf-to-exclude> -L 25 -P 50 -q 20 -Q 20
```

Затем мы рассчитали уровень метилирования с использованием найденных уровней дезаминирования на позициях с метилированными и деметилированными цитозинами. Размер геномного окна для каждого образца указан в соответствующем столбце табл. 1 и был подобран экспериментально.

```
DamMet estF -b <bam-file> -r <fasta-file> -c <chromosome> -M <expected-average-methylation> -O <out-file-prefix> -N <genomic-window-size-in-CpGs>
```

Полученные профили метилирования были дополнительно сглажены с помощью бегущего среднего с окном сглаживания 25 CpG скриптом на языке Python.

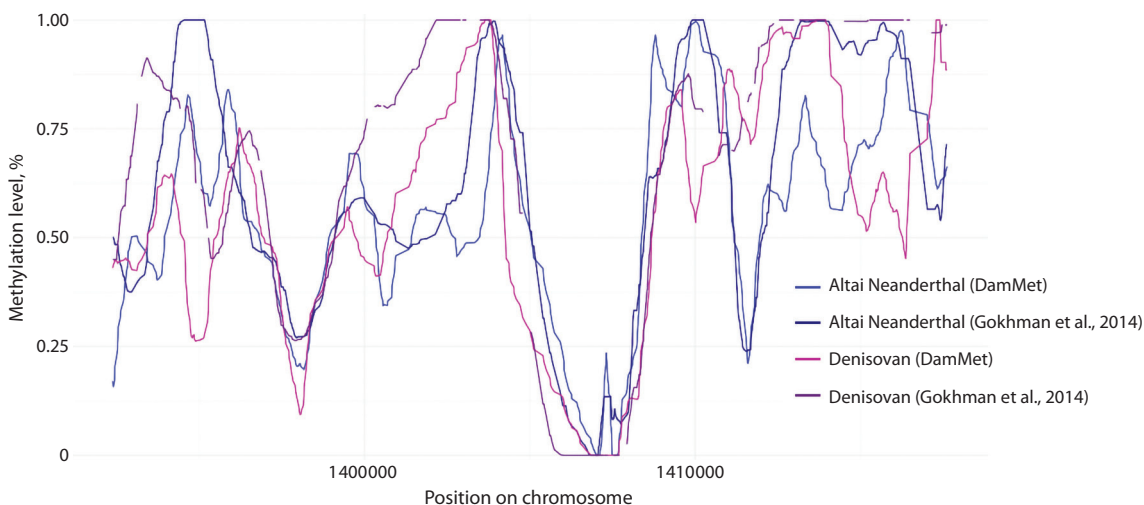
**Валидация реконструированных метиломов.** Сравнение профилей метилирования неандертальца, денисовца и усть-ишимского охотника-собирателя, полученных на предыдущем этапе, производилось с помощью языка R. Для препроцессинга данных, расчета корреляций и построения графиков мы использовали пакеты ggplot, psych, corr.test и семейство tidyverse.

**Поиск тканеспецифичных областей метилирования.** Для поиска регионов, метилирование которых примерно одинаково во всех тканях мезодермального происхож-

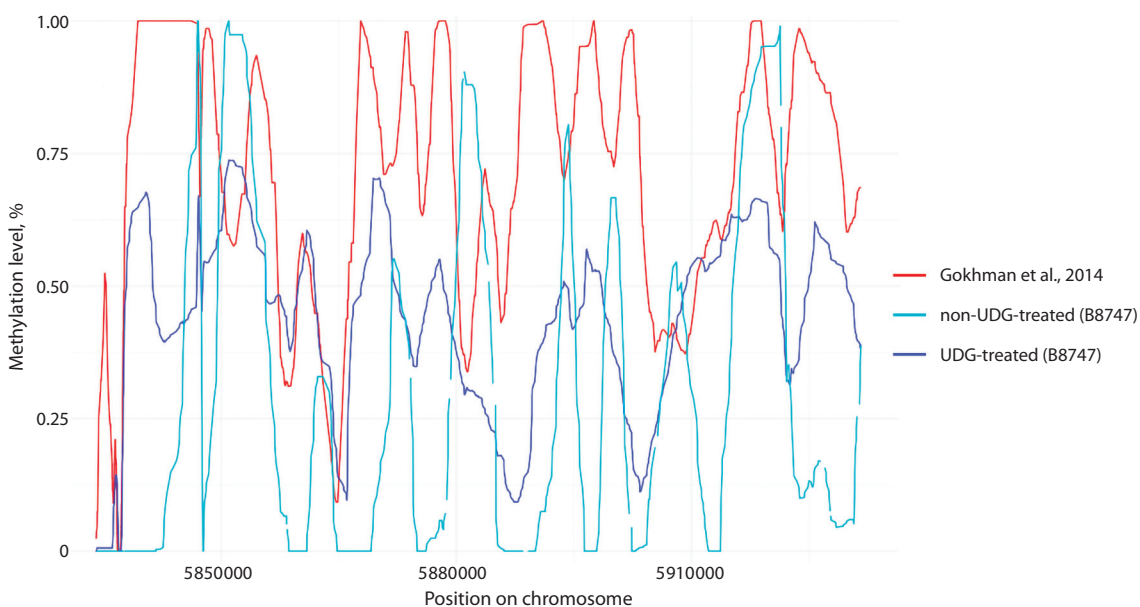
дения, был разработан скрипт на языке Python. На вход подаются значения метилирования, рассчитанные алгоритмом Bismark (Krueger, Andrews, 2011) после картирования образцов, упоминаемых выше. Сравнение значений метилирования проводилось попозиционно с помощью ANOVA для выявления различий в трех группах тканей (фибробласты, миоциты, остеобласты) и исключения из древних костных и современных адипоцитарных профилей метилирования позиций, в которых было обнаружено дифференциальное метилирование ( $p < 0.05$ ).

**Поиск ДМР.** Подготовленные на предыдущих этапах профили метилирования охотников-собирателей (OC) и современных людей сравнивались методом ANOVA, как и при поиске тканеспецифичного метилирования. В первой итерации образцы разделялись на три группы: кости охотников-собирателей, адипоциты здоровых людей и адипоциты пациентов с ожирением. При наличии уровня значимости  $p < 0.05$  CpG-сайт отбирался для последующего анализа с помощью *post hoc* теста Тьюки. CpG-сайт считался дифференциально метилированным, если изменение метилирования было значимо ( $p < 0.05$ ) при сравнении костей OC с адипоцитами тучных людей и незначимо при сравнении OC с контролем. Во второй итерации мы изменили деление по группам: все образцы были костными, группы представляли собой образцы разного возраста (анатомически древние люди, охотники-собиратели и современные люди), для облегчения нагрузки на вычислительные ресурсы сравнения проводили только в регионах, ассоциированных с ожирением. Для сборки полученных дифференциально метилированных сайтов в регионы мы воспользовались ПО combined-pvalues (<https://github.com/brentp/combined-pvalues>), в основе которого лежит метод поправок на множественное тестирование Штуффера–Липтака (Pedersen et al., 2012). Статус изменения метилирования определялся сравнением средних значений метилирования региона в группах.





**Рис. 1.** Профили метилирования денисовца и неандертальца, восстановленные DamMet и опубликованные Д. Гохманом. В поле зрения – деметилированный CpG-островок chr1:1406845–1407821.



**Рис. 2.** Сравнение уровней метилирования на участке хромосомы 2 образца Vi33 в присутствии и в отсутствие УДГ обработки библиотек с ранее опубликованными профилями Д. Гохмана.

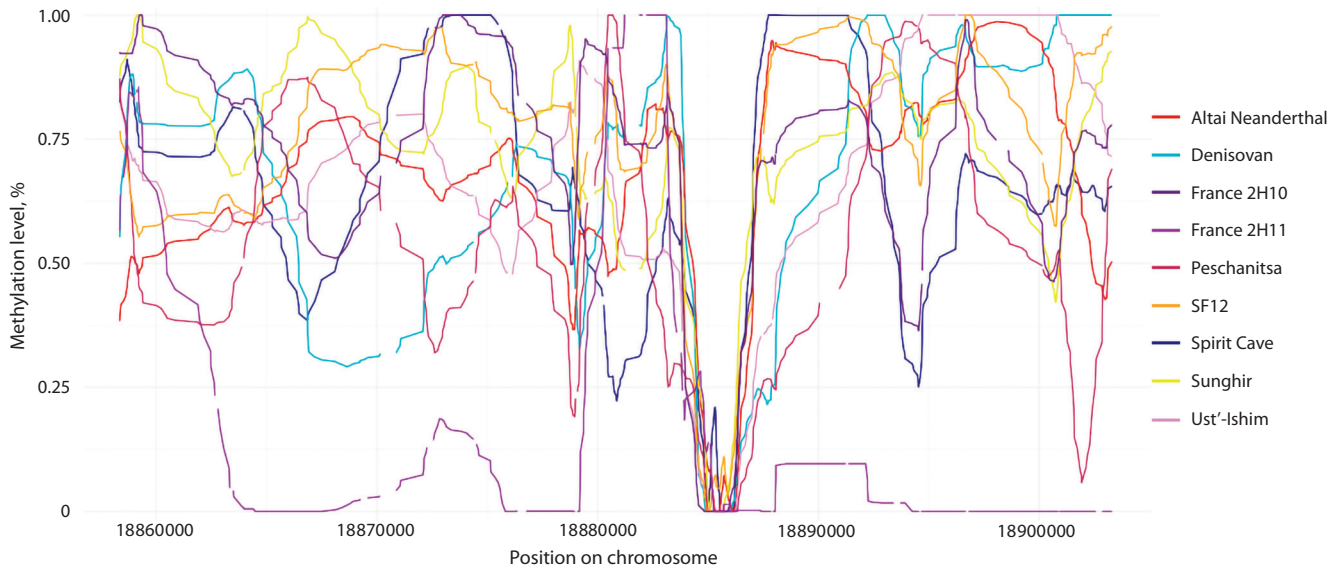
Уровни метилирования всех образцов были сглажены с помощью бегущего среднего с окном 25 CpG.

## Результаты

В данном исследовании мы реконструировали 11 профилей метилирования древних людей с помощью инструмента DamMet. В первую очередь необходимо было разработать пайплайн, который позволял бы восстанавливать метиломы с высокой точностью; для этого мы подали на вход в пайплайн геномы неандертальца и денисовца, прошедшие обработку УДГ. Профили для этих организмов уже были опубликованы ранее (Gokhman et al., 2014, 2020), что позволяет нам провести валидацию пайплайна. Мы обнаружили, что рассчитанное нами метилирование в среднем на 15–20 % ниже, чем ранее опубликованное, при этом в целом профили метилирования схожи (рис. 1).

Коэффициент корреляции профилей метилирования в обоих случаях более 85 %:  $r_{\text{Denisovan}} = 0.87$ ,  $r_{\text{Neanderthal}} = 0.9$  ( $p < 0.05$ ).

Поскольку у нас было несколько образцов, которые не прошли обработку урацил-ДНК-гликозилазой на этапе подготовки библиотек, мы также хотели удостовериться, что DamMet способен реконструировать профили метилирования и в отсутствие этого этапа пробоподготовки. Для этого мы выбрали образец Vi33, для которого в открытом доступе были опубликованы последовательности, как прошедшие УДГ обработку, так и не прошедшие ее. Параметры пайплайна не изменялись для этих расчетов, т. е. условия были одинаковы при реконструкции мети-

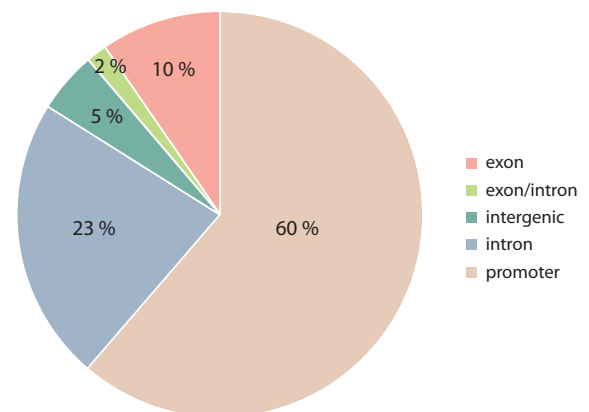


**Рис. 3.** Профили метилирования охотников-собирателей, реконструированные с помощью DamMet. Область протяженного деметилирования принадлежит CpG-островку chr21:18884807–18886111 (GRCh37 hg19).

ломов обеих библиотек. Мы обнаружили среднюю корреляцию профиля метилирования, полученного в присутствии УДГ, с рассчитанным Д. Гохманом ( $r = 0.57$ ); паттерны метилирования представлены на рис. 2. Однако метилом, полученный без обработки УДГ, имеет слабую корреляцию с опубликованным ( $r = 0.14$ ), и паттерны метилирования совпадают в основном в деметилированных CpG-островках независимо от того, проводим мы последующее сглаживание бегущим средним или нет.

Далее мы обработали с помощью нашего пайплайна восемь геномов охотников-собирателей, для которых ранее реконструкция профилей метилирования не проводилась (см. табл. 1). Полученные профили в целом имели похожую на другие древние метиломы форму, в том числе проявляли полное деметилирование некоторых CpG-островков (рис. 3), схожее с профилем ранее реконструированного усть-ишимского охотника-собирателя (Gokhman et al., 2020). Несмотря на то что образец PES001 не был обработан УДГ при подготовке библиотек, полученный нами профиль метилирования имел тенденции, схожие с прочими профилями охотников-собирателей, поэтому не был исключен из дальнейшего анализа.

Как утверждают авторы метода, реконструированные с помощью DamMet профили метилирования могут быть использованы для прямого сравнения с современными данными, однако метилирование может различаться в клетках разного происхождения, и, следовательно, прямому сравнению подлежат только профили метилирования, полученные из одних тканей. Насколько нам известно, секвенирование костных тканей в контексте ожирения не проводилось, поэтому для конечного сравнения мы выбрали образцы адипоцитов подкожной клетчатки и висцерального жира, которые имеют схожие паттерны метилирования. Однако эти паттерны могут серьезно отличаться от тех, что демонстрируют кости и другие ткани мезодермального происхождения, поэтому мы разрабо-



**Рис. 4.** Процентное соотношение ДМР в различных геномных областях.

тали скрипт на языке Python, который осуществляет поиск дифференциально метилированных позиций в мезодермальных тканях и исключает их из дальнейшего анализа. В основу лег метод дисперсионного анализа в трех группах с последующей проверкой попарными сравнениями и поправкой на множественные сравнения. Образцы мезодермального происхождения были разделены на группы в соответствии с типом ткани: фибробласты, мышечные клетки и остеобласты. Всего исследовали 26.5 миллиона CpG-позиций, из них около 206 тысяч имели дифференциальное метилирование хотя бы в одной из групп, тогда как более 26 миллионов не проявляли значимых различий.

Мы провели поиск ДМР для современных образцов костных тканей, но при этом искали изменения только в 642 регионах, для которых в литературе было показано дифференциальное метилирование в контексте ожирения. В данном случае мы проводили попозиционный ANOVA-анализ для групп древних людей, охотников-собирателей

и современных людей (костная ткань) с предварительной фильтрацией нетканеспецифичных CpG-сайтов. Было идентифицировано 38 ДМР, в которых пересечение хотя бы с одним из обозначенных 642 регионов составило 60 % не менее 20 CpG-сайтов. Как видно на рис. 4, около 60 % найденные ДМР лежат в промоторных областях генов, 35 % принадлежат ДМР в теле гена, и лишь 5 % ДМР находятся в межгенных областях. 94 % этих ДМР гиперметилированы, что способствует подавлению экспрессии генов, ассоциированных с ожирением.

#### Дополнительные данные и программный код

Профили метилирования древних людей и скрипты на языке Python, использованные для анализа в данной работе, доступны в репозитории GitHub: <https://github.com/bor-d/ancDMR>

#### Заключение

В настоящее время существует несколько методов восстановления профилей метилирования древних организмов, из которых в основном используются epiPALEOMIX (Hanghøj et al., 2016) и DamMet (Hanghøj et al., 2019). Оба эти метода, согласно исследованиям, демонстрируют значительную точность, но ограничены качеством древних образцов. Для нашего анализа мы остановились именно на втором методе, так как авторы при разработке учитывали возможность сравнения полученных значений метилирования с профилями, для которых использовались другие технологии секвенирования, что сделало метод универсальным. Однако при валидации пайплайна мы заметили, что значения метилирования, полученные с помощью DamMet, довольно сильно отклоняются от опубликованных ранее (Gokhman et al., 2014, 2020). Разработчики DamMet пишут, что их инструмент имеет свойство занижать значения метилирования по сравнению с профилями epiPALEOMIX, который не учитывает SNV, ошибки секвенирования и дезаминирование неметилированных цитозинов, что мы и наблюдали при реконструкции профилей неандертальца и денисовца. В то же время полученная нами корреляция восстановленных с помощью DamMet значений метилирования с ранее опубликованными имеет положительные значения, как и при сравнении, проведенном авторами DamMet. Соответственно, инструмент эффективно работает и для исследования ранее не восстановленных профилей, которые впоследствии действительно могут использоваться для дальнейшего сравнения с современными метиломами.

В качестве демонстрации разработанного нами пайплайна мы осуществили поиск ДМР в геномах охотников-собирателей и современных людей в контексте ожирения, выявив 38 регионов, среди которых около двух третей находится в промоторных областях, а значит, изменение метилирования промоторов данных генов может быть ассоциировано с их экспрессией. Шаги, на которые мы разделили пайплайн, разрешают проблемы, с которыми может столкнуться исследователь, обнаруживший дефицит опубликованных профилей метилирования нужных тканей в контексте исследуемого им состояния, поскольку они позволяют снизить вероятность получения ложноположительных ДМР, обусловленных тканеспецифич-

ностью. Для применения данного пайплайна при поиске ДМР, ассоциированных с другим заболеванием, исследователю потребуется провести анализ соответствующей литературы и выбрать регионы, метилирование которых связано с интересующим его заболеванием. Однако, не смотря на исключение тканеспецифичных регионов и фильтрацию по регионам, ассоциированным с болезнью, в процессе поиска ДМР все еще будут принимать участие CpG-сайты, метилирование которых было изменено в ходе эволюционного перехода от древних людей (*Homo sapiens neandertaliensis*) к людям современного типа (*Homo sapiens sapiens*).

#### Список литературы / References

- Angermueller C., Lee H.J., Reik W., Stegle O. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome Biol.* 2017;18(1):67. DOI 10.1186/s13059-017-1189-z
- Bock C., Reither S., Mikeska T., Paulsen M., Walter J., Lengauer T. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics.* 2005;21(21):4067-4068. DOI 10.1093/bioinformatics/bti652
- Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120. DOI 10.1093/bioinformatics/btu170
- Briggs A.W., Stenzel U., Johnson P.L.F., Green R.E., Kelso J., Prüfer K., Meyer M., Krause J., Ronan M.T., Lachmann M., Pääbo S. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA.* 2007;104(37):14616-14621. DOI 10.1073/pnas.0704665104
- Briggs A.W., Good J.M., Green R.E., Krause J., Maricic T., Stenzel U., Lalueza-Fox C., Rudan P., Brajković D., Kučan Ž., Gušić I., Schmitz R., Doronichev V.B., Golovanova L.V., de la Rasilla M., Fortea J., Rosas A., Pääbo S. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science.* 2009a;325(5938):318-321. DOI 10.1126/science.1174462
- Briggs A.W., Good J.M., Green R.E., Krause J., Maricic T., Stenzel U., Pääbo S. Primer extension capture: targeted sequence retrieval from heavily degraded DNA sources. *J. Vis. Exp.* 2009b;31:1573. DOI 10.3791/1573
- Briggs A.W., Stenzel U., Meyer M., Krause J., Kircher M., Pääbo S. Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* 2010;38(6):e87. DOI 10.1093/nar/gkp1163
- Clark S.J., Harrison J., Paul C.L., Frommer M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 1994;22(15):2990-2997. DOI 10.1093/nar/22.15.2990
- Feinberg A.P., Irizarry R.A. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc. Natl. Acad. Sci. USA.* 2010;107(Suppl.1):1757-1764. DOI 10.1073/pnas.0906183107
- Fu Q., Li H., Moorjani P., Jay F., Slepchenko S.M., Bondarev A.A., Johnson P.L.F., Aximu-Petri A., Prüfer K., de Filippo C., Meyer M., Zwyns N., Salazar-García D.C., Kuzmin Y.V., Keates S.G., Kosintsev P.A., Razhev D.I., Richards M.P., Peristov N.V., Lachmann M., Douka K., Higham T.F.G., Slatkin M., Hublin J.J., Reich D., Kelso J., Viola T.B., Pääbo S. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature.* 2014;514(7523):445-449. DOI 10.1038/nature13810
- Gansauge M.-T., Meyer M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 2013;8(4):737-748. DOI 10.1038/nprot.2013.038
- Gokhman D., Lavi E., Prüfer K., Fraga M.F., Riancho J.A., Kelso J., Pääbo S., Meshorer E., Carmel L. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science.* 2014;344(6183):523-527. DOI 10.1126/science.1250368



- Gokhman D., Nissim-Rafinia M., Agranat-Tamir L., Housman G., Garcia-Pérez R., Lizano E., Cheronet O., Mallick S., Nieves-Colón M.A., Li H., Alpaslan-Roodenberg S., Novak M., Gu H., Osinski J.M., Ferrando-Bernal M., Gelabert P., Lipende I., Mjunga D., Kondova I., Bontrop R., Kullmer O., Weber G., Shahar T., Dvir-Ginzberg M., Faerman M., Quillen E.E., Meissner A., Lahav Y., Kandel L., Liebergall M., Prada M.E., Vidal J.M., Gronostajski R.M., Stone A.C., Yakir B., Lalueza-Fox C., Pinhasi R., Reich D., Marques-Bonet T., Meshorer E., Carmel L. Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat. Commun.* 2020; 11(1):1189. DOI 10.1038/s41467-020-15020-6
- Gu H., Smith Z.D., Bock C., Boyle P., Gnirke A., Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protoc.* 2011;6(4): 468-481. DOI 10.1038/nprot.2010.190
- Günther T., Malmström H., Svensson E.M., Omrak A., Sánchez-Quinto F., Kılınc G.M., Krzewińska M., Eriksson G., Fraser M., Edlund H., Munters A.R., Coutinho A., Simões L.G., Vicente M., Sjölander A., Sellevold B.J., Jørgensen R., Claes P., Shriver M.D., Valdiosera C., Netea M.G., Apel J., Lidén K., Skar B., Storå J., Götherström A., Jakobsson M. Population genomics of Mesolithic Scandinavia: investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.* 2018;16(1):e2003703. DOI 10.1371/journal.pbio.2003703
- Hanghøj K., Seguin-Orlando A., Schubert M., Madsen T., Pedersen J.S., Willerslev E., Orlando L. Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX. *Mol. Biol. Evol.* 2016;33(12):3284-3298. DOI 10.1093/molbev/msw184
- Hanghøj K., Renaud G., Albrechtsen A., Orlando L. DamMet: ancient methylome mapping accounting for errors, true variants, and post-mortem DNA damage. *GigaScience.* 2019;8(4):giz025. DOI 10.1093/gigascience/giz025
- Jablonka E., Raz G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* 2009;84(2):131-176. DOI 10.1086/598822
- Jun G., Wing M.K., Abecasis G.R., Kang H.M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* 2015;25(6): 918-925. DOI 10.1101/gr.176552.114
- Krueger F., Andrews S. RBismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011;27(11): 1571-1572. DOI 10.1093/bioinformatics/btr167
- Loyfer N., Magenheimer J., Peretz A., Cann G., Bredno J., Klochendler 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., Khalailah A., Zamir G., Grinbaum R., Gazala M.A., Mizrahi I., Shussman N., Korach A., Wald O., Izhar U., Erez E., Yutkin V., Samet Y., Golinkin D.R., 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. A DNA methylation atlas of normal human cell types. *Nature.* 2023;613(7943):355-364. DOI 10.1038/s41586-022-05580-6
- Meyer M., Kircher M., Gansauge M.-T., Li H., Racimo F., Mallick S., Schraiber J.G., Jay F., Prüfer K., de Filippo C., Sudmant P.H., Alkan C., Fu Q., Do R., Rohland N., Tandon A., Siebauer M., Green R.E., Bryc K., Briggs A.W., Stenzel U., Dabney J., Shendure J., Kitzman J., Hammer M.F., Shunkov M.V., Derevianko A.P., Patterson N., Andrés A.M., Eichler E.E., Slatkin M., Reich D., Kelso J., Pääbo S. A high-coverage genome sequence from an archaic Denisovan individual. *Science.* 2012;338(6104):222-226. DOI 10.1126/science.1224344
- Moreno-Mayar J., Potter B., Vinner L., Steinrücken M., Rasmussen S., Terhorst J., Kamm J., Albrechtsen A., Malaspina A., Sikora M., Rether J., Irish J., Malhi R., Orlando L., Song Y., Nielsen R., Meltzer D., Willerslev E. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature.* 2018a; 553(7687):203-207. DOI 10.1038/nature25173
- Moreno-Mayar J.V., Vinner L., Damgaard P.B., de la Fuente C., Chan J., Spence J.P., Allentoft M.E., Vimala T., Racimo F., Pinotti T., Rasmussen S., Margaryan A., Orbeogo M.I., Mylopotamitaki D., Wooller M., Bataille C., Becerra-Valdivia L., Chivall D., Comeskey D., Deviese T., Grayson D.K., George L., Harry H., Alexandersen V., Primeau C., Erlandson J., Rodrigues-Carvalho C., Reis S., Bastos M.Q.R., Cybulski J., Vullo C., Morello F., Vilar M., Wells S., Gregersen K., Hansen K.L., Lynnerup N., Mirazón Lahr M., Kjær K., Strauss A., Alfonso-Durruty M., Salas A., Schroeder H., Higham T., Malhi R.S., Rasic J.T., Souza L., Santos F.R., Malaspina A.-S., Sikora M., Nielsen R., Song Y.S., Meltzer D.J., Willerslev E. Early human dispersals within the Americas. *Science.* 2018b;362(6419). DOI 10.1126/science.aav2621
- Niiranen L., Leciej D., Edlund H., Bernhardsson C., Fraser M., Sánchez Quinto F., Herzog K.H., Jakobsson M., Walkowiak J., Thalman O. Epigenetic modifications in modern and ancient genomes. *Genes.* 2022;13(2):178. DOI 10.3390/genes13020178
- Ohm J.E., Mali P., Van Neste L., Berman D.M., Liang L., Pandiyan K., Briggs K.J., Zhang W., Argani P., Simons B., Yu W., Matsui W., Van Criekinge W., Rassool F.V., Zambidis E., Schuebel K.E., Cope L., Yen J., Mohammad H.P., Cheng L., Baylin S.B. Cancer-related epigenome changes associated with reprogramming to induced pluripotent stem cells. *Cancer Res.* 2010;70(19):7662-7673. DOI 10.1158/0008-5472.CAN-10-1361
- Olova N., Krueger F., Andrews S., Oxley D., Berrens R.V., Branco M.R., Reik W. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol.* 2018;19(1):33. DOI 10.1186/s13059-018-1408-2
- Orlando L., Gilbert M.T.P., Willerslev E. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* 2015;16(7):395-408. DOI 10.1038/nrg3935
- Pedersen B.S., Schwartz D.A., Yang I.V., Kechris K.J. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics.* 2012;28(22):2986-2988. DOI 10.1093/bioinformatics/bts545
- Poplin R., Ruano-Rubio V., DePristo M.A., Fennell T.J., Carneiro M.O., Van der Auwera G.A., Kling D.E., Gauthier L.D., Levy-Moonshine A., Roazen D., Shakir K., Thibault J., Chandran S., Whelan C., Lek M., Gabriel S., Daly M.J., Neale B., MacArthur D.G., Banks E. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv.* 2017. DOI 10.1101/201178
- Prüfer K., Racimo F., Patterson N., Jay F., Sankararaman S., Sawyer S., Heinze A., Renaud G., Sudmant P.H., de Filippo C., Li H., Mallick S., Dannemann M., Fu Q., Kircher M., Kuhlwillm M., Lachmann M., Meyer M., Ongyerth M., Siebauer M., Theunert C., Tandon A., Moorjani P., Pickrell J., Mullikin J.C., Vohr S.H., Green R.E., Hellmann I., Blanche H., Cann H., Kitzman J.O., Shendure J., Eichler E.E., Lein E.S., Bakken T.E., Golovanova L.V., Doronichev V.B., Shunkov M.V., Derevianko A.P., Viola B., Slatkin M., Reich D., Kelso J., Pääbo S. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature.* 2014;505(7481): 43-49. DOI 10.1038/nature12886
- Prüfer K., de Filippo C., Grote S., Mafessoni F., Korlević P., Hajdinjak M., Vernot B., Skov L., Hsieh P., Peyrégne S., Reher D., Hopfe C., Nagel S., Maricic T., Fu Q., Theunert C., Rogers R., Skoglund P., Chintalapati M., Dannemann B., Nelson B.J., Key F.M., Rudan P., Kučan Ž., Gušić I., Golovanova L.V., Doronichev V.B., Patterson N., Reich D., Eichler E.E., Slatkin M., Schierup M.H., Andrés A.M., Kelso J., Meyer M., Pääbo S. A high-coverage Neanderthal genome from Vindija Cave in Croatia. *Science.* 2017;358(6363):655-658. DOI 10.1126/science.aao1887
- Saag L., Vasilyev S.V., Varul L., Kosorukova N.V., Gerasimov D.V., Oshibkina S.V., Griffith S.J., Solnik A., Saag L., D'Atanasio E., Metspalu E., Reidla M., Rootsi S., Kivisild T., Scheib C.L., Tambets K., Kriiska A., Metspalu M. Genetic ancestry changes in Stone to Bronze Age transition in the East European plain. *Sci. Adv.* 2021;7:eabd6535. DOI 10.1126/sciadv.abd6535



- Sawyer S., Krause J., Guschanski K., Savolainen V., Pääbo S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One*. 2012;7(3):e34131. DOI 10.1371/journal.pone.0034131
- Seguin-Orlando A., Donat R., Der Sarkissian C., Southon J., Thèves C., Manen C., Tchérémissinoff Y., Crubézy E., Shapiro B., Deleuze J., Dalén L., Guilaine J., Orlando L. Heterogeneous hunter-gatherer and steppe-related ancestries in Late Neolithic and Bell Beaker genomes from present-day France. *Curr. Biol*. 2021;31(5):1072-1083. DOI 10.1016/j.cub.2020.12.015
- Sikora M., Seguin-Orlando A., Sousa V.C., Albrechtsen A., Korneliusen T., Ko A., Rasmussen S., Dupanloup I., Nigst P.R., Bosch M.D., Renaud G., Allentoft M.E., Margaryan A., Vasilyev S.V., Veselovskaya E.V., Borutskaya S.B., Deviese T., Comeskey D., Higham T., Manica A., Foley R., Meltzer D.J., Nielsen R., Excoffier L., Lahr M.M., Orlando L., Willerslev E. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science*. 2017;358(6363):659-662. DOI 10.1126/science.aao1807
- Suzuki M., Liao W., Wos F., Johnston A.D., DeGrazia J., Ishii J., Bloom T., Zody M.C., Germer S., Grealis J.M. Whole-genome bisulfite sequencing with improved accuracy and cost. *Genome Res*. 2018;28(9):1364-1371. DOI 10.1101/gr.232587.117
- Zhur K.V., Trifonov V.A., Prokhortchouk E.B. Progress and prospects in epigenetic studies of ancient DNA. *Biochemistry (Mosc.)*. 2021; 86(12-13):1563-1571. DOI 10.1134/S0006297921120051
- Zou L.S., Erdos M.R., Taylor D.L., Chines P.S., Varshney A., Parker S.C.J., Collins F.S., Didion J.P. BoostMe accurately predicts DNA methylation values in whole-genome bisulfite sequencing of multiple human tissues. *BMC Genomics*. 2018;19(1):390. DOI 10.1186/s12864-018-4766-y

---

#### ORCID ID

D.D. Borodko [orcid.org/0000-0003-3596-5470](https://orcid.org/0000-0003-3596-5470)

S.V. Zhenilo [orcid.org/0000-0003-0874-1594](https://orcid.org/0000-0003-0874-1594)

F.S. Sharko [orcid.org/0000-0002-1189-5597](https://orcid.org/0000-0002-1189-5597)

**Благодарности.** Работа выполнена при финансовой поддержке проекта Минобрнауки России, системный номер № 075-10-2020-116 (номер гранта 13.1902.21.0023).

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 15.07.2023. После доработки 04.10.2023. Принята к публикации 05.10.2023.