

ИНТЕГРИРОВАННАЯ РЕФЕРЕНСНАЯ ПАНЕЛЬ ГАПЛОТИПОВ *MACACA MULATTA* ДЛЯ ИМПУТАЦИИ ГЕНОТИПОВ В ПОПУЛЯЦИЯХ СМЕШАННОГО ПРОИСХОЖДЕНИЯ

Ф. С. Шарко, А. В. Прокопенко, А. О. Изотова, С. В. Тошчаков ✉

Национальный исследовательский центр «Курчатовский институт», Москва, Россия

Импутация генотипов позволяет значительно снизить затраты на получение полногеномных данных при сохранении информативности, однако ее точность напрямую зависит от наличия качественной фазированной референсной панели гаплотипов. Существующие панели для *Macaca mulatta* охватывают преимущественно индийскую линию вида и не отражают генетическую структуру лабораторных популяций смешанного происхождения. Целью ретроспективного биоинформатического исследования было создать и отвалидировать фазированную референсную панель на основе консолидации 618 общедоступных полногеномных данных *M. mulatta* обеих географических линий, обработанных по единому стандартизированному протоколу. Итоговая панель включает 33 457 491 SNP-маркер. Валидация на образцах индийского и китайского происхождения при сверхнизком покрытии (0,1x–0,9x) подтвердила высокую точность импутации, максимальную для доминирующей в панели индийской линии. Апробация на 10 независимых образцах при покрытии 0,6x показала высокий средний показатель достоверности импутации (INFO = 0,898). Созданный ресурс открывает возможности для экономически эффективного генотипирования больших когорт *M. mulatta* и служит основой для проведения полногеномных ассоциативных и популяционно-генетических исследований в отечественных и зарубежных приматологических центрах.

Ключевые слова: импутация генотипов, референсная панель, *Macaca mulatta*, полногеномное секвенирование, GLIMPSE2, SHAPEIT5, биоинформатика

Финансирование: работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (Федеральная научно-техническая программа развития генетических технологий на 2019–2030 г.), соглашение №075-15-2025-491 от 30.05.2025.

Благодарности: работа выполнена с использованием оборудования центра коллективного пользования «Комплекс моделирования и обработки данных исследовательских установок мега-класса» НИЦ «Курчатовский институт», <http://ckp.nrcki.ru/>.

Вклад авторов: С. В. Тошчаков — дизайн работы, написание рукописи; Ф. С. Шарко — биоинформатический анализ, подготовка черновика; А. В. Прокопенко, А. О. Изотова — редактирование рукописи.

Соблюдение этических стандартов: исследования одобрены этическим комитетом НИЦ «Курчатовский институт» (протокол №2-2пр от 30 июня 2025 г.).

✉ **Для корреспонденции:** Степан Владимирович Тошчаков
пл. Академика Курчатова, д. 1, г. Москва, 123098, Россия; stepan.toshchakov@gmail.com

Статья получена: 14.07.2026 **Статья принята к печати:** 14.08.2026 **Опубликована онлайн:** 21.08.2026

DOI: 10.24075/vrgmu.2026.045

Авторские права: © 2026 принадлежат авторам. **Лицензиат:** РНИМУ им. Н. И. Пирогова. Статья размещена в открытом доступе и распространяется на условиях лицензии Creative Commons Attribution (CC BY) (<https://creativecommons.org/licenses/by/4.0/>).

INTEGRATED REFERENCE PANEL OF *MACACA MULATTA* HAPLOTYPES FOR GENOTYPE IMPUTATION IN POPULATIONS OF MIXED ORIGIN

Sharko FS, Prokopenko AV, Izotova AO, Toshchakov SV ✉

National Research Centre "Kurchatov Institute", Moscow, Russia

Genotype imputation makes it possible to considerably reduce the cost of whole-genome data acquisition preserving the information value, but its accuracy depends directly on the availability of a high-quality phased reference panel of haplotypes. The existing panels for *Macaca mulatta* cover mostly the Indian species lineage and do not reflect the genetic structure of laboratory populations of mixed origin. This retrospective bioinformatics study aimed to produce and validate a phased reference panel based on the consolidation of publicly available whole-genome datasets from 618 *M. mulatta* individuals of both geographic lineages processed in accordance with the common standardized protocol. The resulting panel includes 33,457,491 SNP markers. Validation on samples of Indian and Chinese origin with ultra-low coverage (0.1x–0.9x) confirmed high imputation accuracy, which was highest for the Indian lineage dominant in the panel. Testing on 10 independent samples with the 0.6x coverage showed a high average imputation reliability score (INFO = 0.898). The created resource opens the prospects for cost-effective genotyping of large *M. mulatta* cohorts and provides the basis for whole-genome association and population genetic studies in domestic and foreign primate research centers.

Keywords: genotype imputation, reference panel, *Macaca mulatta*, whole-genome sequencing, GLIMPSE2, SHAPEIT5, bioinformatics

Funding: the study was supported by the Ministry of Science and Higher Education of the Russian Federation (Federal Scientific and Technical Program for the Development of Genetic Technologies for 2019–2030), agreement No. 075-15-2025-491 dated 30.05.2025.

Acknowledgements: the study was conducted using the equipment of the shared use center "Complex for Modeling and Processing Data from Mega-Class Research Facilities", National Research Centre "Kurchatov Institute", <http://ckp.nrcki.ru/>.

Author contribution: Toshchakov SV — research design, manuscript writing; Sharko FS — bioinformatics analysis, manuscript draft; Prokopenko AV, Izotova AO — manuscript editing.

Compliance with ethical standards: the studies were approved by the Ethics Committee of the National Research Centre "Kurchatov Institute" (protocol №2-2np dated June 30, 2025).

✉ **Correspondence should be addressed:** Stepan V. Toshchakov
Academician Kurchatov pl, 1, Moscow, 123098, Russia; stepan.toshchakov@gmail.com

Received: 14.07.2026 **Accepted:** 14.08.2026 **Published online:** 21.08.2026

DOI: 10.24075/brsmu.2026.045

Copyright: © 2026 by the authors. **Licensee:** Pirogov University. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Макак-резус (*Macaca mulatta*) относится к числу наиболее распространенных модельных организмов в биомедицинских исследованиях. В сфере инфекционной патологии этот вид служит моделью для изучения патологий человека различной этиологии, включая ВИЧ-инфекцию/СПИД, туберкулез, гепатиты, грипп [1]. Кроме того, благодаря высокому филогенетическому сходству нервной системы макака и человека, они активно используются в нейробиологии, в частности при моделировании нейродегенеративных процессов и болезни Альцгеймера [2], ряда редких наследственных заболеваний [3], а также при доклинической оценке нейроактивных лекарственных средств [4]. Широкое применение этого вида в экспериментальной биомедицине обуславливает значительный интерес к изучению генетических основ фенотипической изменчивости *M. mulatta*, в частности, различий в восприимчивости к заболеваниям и ответе на терапию [5].

Ключевым инструментом для решения таких задач служат полногеномные ассоциативные исследования (GWAS), однако их проведение на больших когортах ограничено высокой стоимостью полногеномного секвенирования [6]. В условиях ограниченного бюджета особую значимость приобретают подходы, позволяющие сократить затраты на получение геномных данных без потери их информативности. Эффективной альтернативой прямому глубокому секвенированию выступает секвенирование с низким покрытием с последующей статистической импутацией генотипов. Этот подход позволяет восстанавливать пропущенные генотипы на основе известных гаплотипов из эталонной (референсной) панели, в разы снижая стоимость анализа по сравнению со стандартным покрытием (от 20x) и делая возможным проведение популяционных исследований на выборках из тысяч образцов [7]. Ключевым фактором, определяющим эффективность импутации, является наличие репрезентативной, фазированной и генетически близкой к целевой выборке референсной панели [8].

Для *M. mulatta* создан ряд геномных ресурсов, ориентированных на каталогизацию и аннотацию известных генетических вариантов. Наиболее масштабным из них является mGAP, предоставляющий каталог аннотированных вариантов, полученных в результате секвенирования 2,5 тыс. особей [5]. Стоит также отметить базу данных RhesusBase, включающую 46,1 млн полиморфных сайтов [9] и ресурс для межвидовых сравнений внутри рода *Macaca* MACSNVdb (74,5 млн SNV) [10]. Эти ресурсы решают задачи каталогизации, аннотации и поиска известных вариантов, однако не предоставляют фазированной референсной панели гаплотипов, необходимой для восстановления полногеномных генотипов у новых, ранее не секвенированных образцов.

В современной практике фазированные референсные панели для импутации генотипов служат базовым инструментом в крупных геномных проектах, связанных с изучением человека и сельскохозяйственных животных [11, 12]. В международных репозиториях представлены обширные массивы данных полногеномного секвенирования *Macaca mulatta*. В 2025 г. была опубликована первая специализированная референсная панель для данного вида, обеспечивающая высокую точность импутации генотипов даже при уровне покрытия 0,5x [13]. Тем не менее этот ресурс ограничен преимущественно индийской линией *Macaca mulatta*. За рамками анализа остается генетическая

структура искусственно созданных закрытых популяций, генезис которых связан с гетерогенным происхождением исходного поголовья, фрагментарностью архивных данных об интродукции и многолетней панмиксией [14, 15]. Для таких популяций смешанного происхождения оптимальным решением становится использование интегрированной панели, объединяющей основные географические линии вида, что сочетает популяционную специфичность и универсальность.

Целью настоящей работы было описать разработку и валидацию фазированной референсной панели для импутации генотипов *M. mulatta*. Панель сформирована на основе результатов полногеномного секвенирования 618 образцов индийского и китайского происхождения из открытых источников, обработанных по единому стандартизированному протоколу с использованием современных алгоритмов фазирования и импутации. При оценке точности импутации по независимым тестовым образцам обеих линий получены высокие значения R^2 и конкордантности в диапазоне глубины покрытия 0,1x–0,9x при ожидаемом преимуществе доминирующей в панели индийской линии. Разработанный ресурс служит основой для проведения полногеномных ассоциативных и популяционно-генетических исследований на макаках-резус в отечественных и зарубежных приматологических центрах, генетическая структура популяций в которых может быть смешанной и априори неизвестной [16, 17].

МАТЕРИАЛЫ И МЕТОДЫ

Сбор и систематизация общедоступных данных полногеномного секвенирования

Систематический поиск данных полногеномного секвенирования *M. mulatta* осуществляли в международных репозиториях NCBI SRA [18], ENA [19] и NGDC [20]. Отбор данных производили на основании следующих критериев: подтвержденная видовая принадлежность к *M. mulatta*, глубина секвенирования не менее 20x, а также наличие верифицированных метаданных о географическом происхождении (Индия или Китай) и поле особи. В ходе первичного поиска было идентифицировано 1808 образцов. После фильтрации по величине покрытия и валидации метаданных число кандидатов составило 982. Из дальнейшего анализа исключали особи, определенные в метаданных как межвидовые гибриды (*M. mulatta* × *M. fascicularis*), а также образцы с неустановленным географическим происхождением.

На заключительном этапе формирования выборки применяли процедуру стратификации, направленную на достижение сбалансированного по полу состава внутри каждой популяционной группы и на минимизацию потенциальных систематических искажений, обусловленных неодинаковой представленностью данных из различных исследований. Итоговый массив включил 618 уникальных особей из трех проектов: PRJNA251548 (BCM, $n = 21$), PRJNA382404 (OHSU, $n = 397$) и CRA014717 (KIZ, $n = 200$). Из общего числа образцов 418 имели индийское происхождение, 200 — китайское; количество самок составило 410, самцов — 208. Сведения о географическом происхождении и поле каждого образца представлены в файле dataset-description.xlsx в открытом репозитории Zenodo: <https://doi.org/10.5281/zenodo.21296424>.

Биоинформатическая обработка и идентификация SNP

Для всех 618 образцов был разработан и применен единый стандартизированный пайплайн биоинформатической обработки, обеспечивающий консистентность данных на всех этапах. Необработанные прочтения фильтровали по качеству с помощью fastp v0.23.2 [21] с параметрами: `-w 12 -z 7 -V -g --poly_g_min_len 5 -x --poly_x_min_len 10 -5 -3 -M 30 -n 1 -e 20 -l 50 --c`. Отфильтрованные прочтения картировались на референсный геном *M. mulatta* (сборка Mmul_10, GCF_003339765.1) с использованием Bowtie2 v2.3.5.1 [22] с настройками по умолчанию. Идентификацию SNP выполняли с помощью bcftools v1.15 [23] (команды `mpileup` и `call`) с ключевыми параметрами: `--redo-BAQ --min-BQ 30 --per-sample-mF --annotate DP,AD --multiallelic-caller --variants-only`. На следующем этапе из итоговых VCF-файлов удаляли инсерции и делеции с использованием vcftools v0.1.16 (параметр `--remove-indels`), оставляя только однонуклеотидные замены (SNP) для дальнейшего анализа. Полученный VCF-файл подвергали дополнительной строгой фильтрации с использованием vcftools v0.1.16 [24] по критериям: качество генотипа (GQ) ≥ 30 , глубина прочтения (DP) 15–180, доля пропущенных генотипов на сайт $\leq 0,02$, минимальная частота минорного аллеля (MAF) $\geq 0,01$.

Фазирование референсной панели

Для восстановления хромосомных фаз гаплотипов отфильтрованного набора данных использовали пакет SHAPEIT5 v5.1.0 [25], разработанный для крупных массивов WGS-данных и характеризующийся высокой точностью восстановления гаплотипных фаз при работе с большими выборками. Для повышения вычислительной эффективности и снижения требований к оперативной памяти применяли постадийное фазирование с разбиением генома на независимые блоки. На первом этапе с помощью модуля SHAPEIT5_phase_common обработку проводили в хромосомных блоках длиной 20 cM; для каждого блока выполняли независимое фазирование с последующей оценкой качества. С целью обеспечения целостности сборки полных хромосом и устранения артефактов, возникающих на границах блоков, осуществляли дополнительное фазирование в перекрывающихся участках на стыках (перекрытие 2 cM) с последующим объединением фаз в единую хромосомную последовательность. В результате была сформирована полностью фазированная панель, содержащая для каждой из 618 особей два гаплотипа (материнский и отцовский), что является необходимым условием для последующей импутации генотипов.

Импутация генотипов и валидация панели

Для оценки точности созданной референсной панели из ее состава были исключены два образца с известным географическим происхождением: M00016 (индийская популяция) и CRR1023808 (китайская популяция). Исходные BAM-файлы высокого покрытия для каждого из этих образцов были подвергнуты прореживанию (downsampling) до пяти уровней глубины покрытия: 0,1×, 0,3×, 0,5×, 0,7× и 0,9×, что позволяло имитировать условия секвенирования с низким покрытием.

Импутацию генотипов для смоделированных данных выполняли с использованием пайплайна GLIMPSE2 v2.0.0

[26]. Процедура включала последовательные этапы: разбиение генома на перекрывающиеся окна с помощью утилиты GLIMPSE2_chunk, подготовку референсной панели в формате GLIMPSE2 посредством GLIMPSE2_split_reference, непосредственно импутацию и фазирование через GLIMPSE2_phase, а также объединение результатов по всем окнам для восстановления полнохромосомных файлов с использованием GLIMPSE2_ligate.

Точность импутации оценивали путем сопоставления импутированных генотипов с референсными, в качестве которых использовали исходные генотипы высокой надежности тех же образцов до проведения прореживания. Для расчета применяли инструмент GLIMPSE2_concordance, вычисляющий два показателя: конкордантность (долю полностью совпавших генотипов) и R^2 (квадрат коэффициента корреляции между импутированными и референсными дозами аллелей). Данный подход позволял количественно охарактеризовать зависимость точности импутации от глубины покрытия целевого образца и степени его генетической близости к референсной панели. Референсная панель доступна в открытом репозитории Zenodo: <https://doi.org/10.5281/zenodo.21296424>.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Характеристика созданной референсной панели

В результате консолидации и унифицированной обработки данных 618 особей *M. mulatta* сформирована фазированная референсная панель гаплотипов, включающая 33 457 491 высокодостоверный SNP-маркер. Популяционная структура панели определяется составом включенных проектов. Большинство образцов имеют индийское происхождение (418 особей; 67,6%), что связано с преобладанием в выборке проекта PRJNA382404 (ONHSU, $n = 397$), образцы которого относились к индийской популяции. Китайское происхождение установлено для 200 особей (32,4%; проект CRA014717, KIZ). Благодаря такому соотношению в панели обеспечена представленность обеих основных географических линий, что расширяет область ее применения на лабораторные популяции со смешанным или неустановленным происхождением. Половой состав выборки характеризуется смещением в сторону самок: 410 самок (66,3%) и 208 самцов (33,7%), что отражает структуру исходных общедоступных WGS-данных.

Оценка точности импутации

Оценку точности импутации выполняли по двум метрикам: конкордантности (доле полностью совпавших генотипов) и R^2 (квадрату коэффициента корреляции между импутированными и референсными дозами аллелей). В отличие от конкордантности, R^2 более чувствителен к ошибкам при редких вариантах, поскольку отражает корреляционную связь между дозами аллелей. Оба показателя ожидаемо демонстрировали прямую зависимость от уровня покрытия целевого образца: с увеличением глубины точность импутации закономерно возрастала. Дополнительно оценивали точность импутации в зависимости от частоты минорного аллеля (MAF) в диапазоне от 1 до 45%.

Для образца индийского происхождения M00016 значение R^2 при увеличении покрытия с 0,1× до 0,9× возрастало с

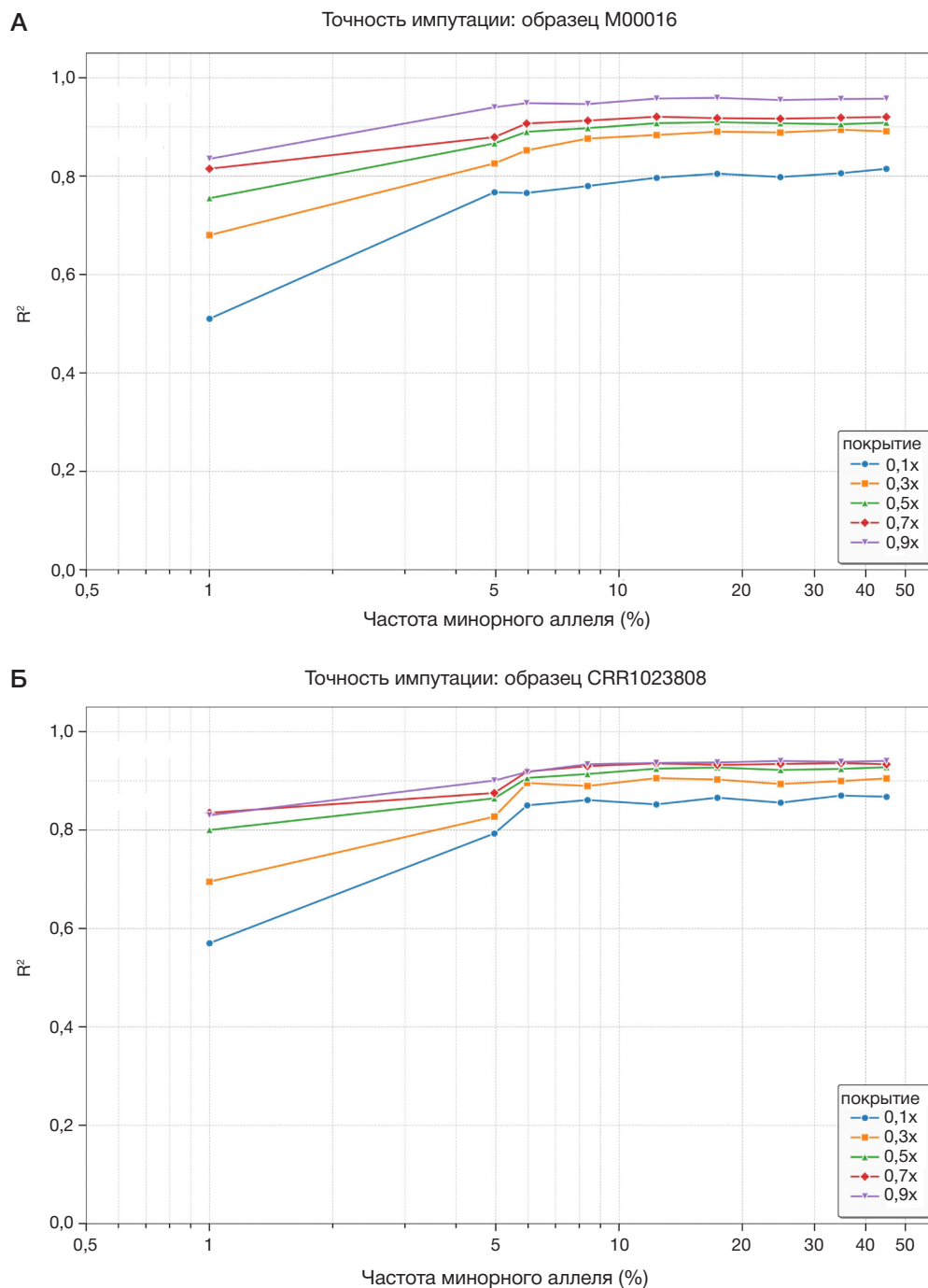


Рис. 1. Точность импутации генотипов (R^2) в зависимости от глубины покрытия и частоты минорного аллеля (MAF) для двух тестовых образцов. По оси абсцисс частота минорного аллеля (%), по оси ординат коэффициент R^2 между импутированными и референсными аллельными дозами. Каждая кривая соответствует одному из пяти уровней покрытия (0,1x; 0,3x; 0,5x; 0,7x; 0,9x). **А.** Образец индийского происхождения M00016. **Б.** Образец китайского происхождения CRR1023808

0,56 до 0,96 (рис. 1А). При стратификации по MAF установлено, что для частых вариантов ($MAF \geq 10\%$) R^2 достигал 0,95–0,96 уже при покрытии 0,7x, тогда как для редких вариантов ($MAF < 1\%$) значение R^2 составляло 0,56 при 0,1x и 0,84 при 0,9x. Конкордантность для данного образца при 0,9x варьировала от 0,61 ($MAF 1\%$) до 0,96–0,97 ($MAF \geq 10\%$).

Для образца китайской популяции CRR1023808 R^2 повышался с 0,58 при 0,1x до 0,95 при 0,9x (рис. 1Б). Для частых вариантов ($MAF \geq 10\%$) R^2 достигал 0,93–0,95 при покрытии 0,7x, для редких ($MAF 1\%$) — 0,58 при 0,1x и 0,82 при 0,9x. Конкордантность для данного образца при 0,9x составляла от 0,63 ($MAF 1\%$) до 0,95–0,97 ($MAF \geq 10\%$).

Полученные значения R^2 и конкордантности свидетельствуют о высокой эффективности созданной референсной панели для импутации генотипов у макака-резус. Для частых и среднечастотных вариантов ($MAF \geq 10\%$) значение R^2 достигает 0,93–0,96 уже при покрытии 0,7x, а конкордантность при 0,9x превышает 0,95 для обоих тестовых образцов, что подтверждает надежность восстановления генотипов для данной категории маркеров. Для редких вариантов ($MAF 1\%$) точность импутации ожидаемо ниже: R^2 составляет 0,82–0,84 при 0,9x, конкордантность — 0,86–0,88, что указывает на необходимость более глубокого покрытия (не менее 0,7x–0,9x) при проведении исследований, ориентированных

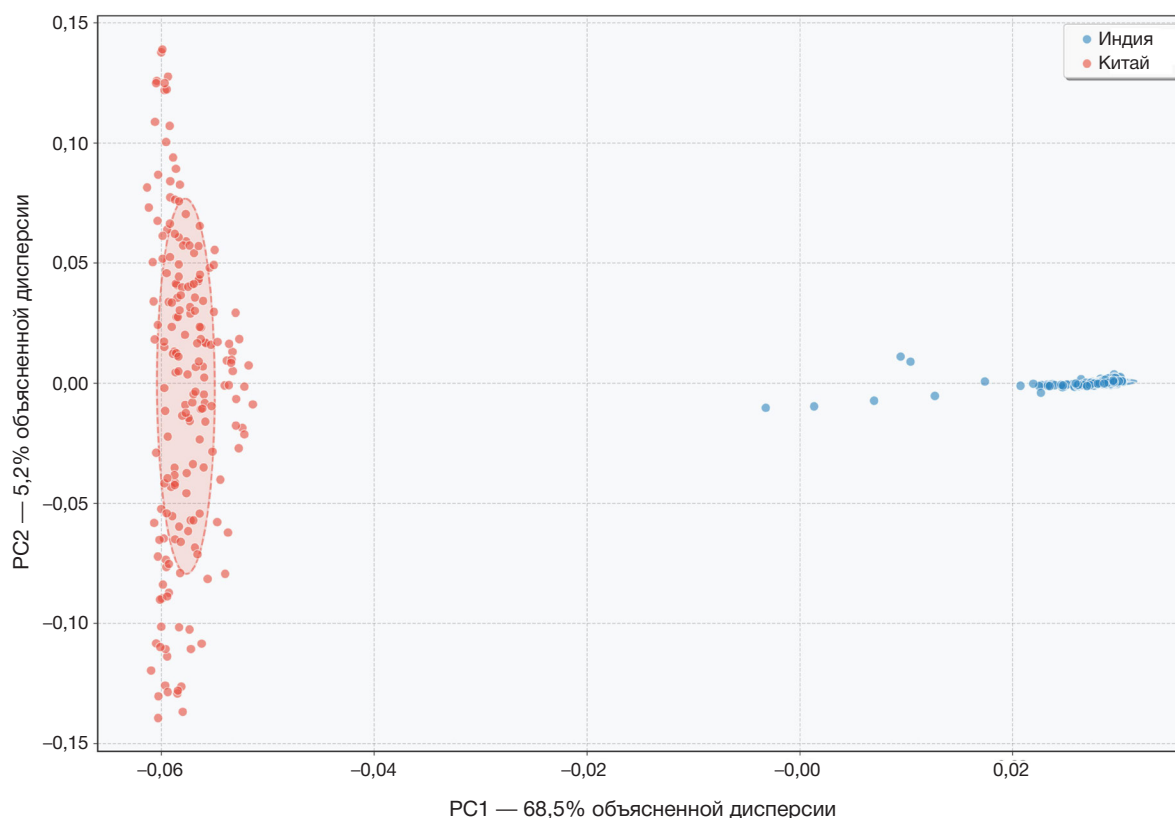


Рис. 2. Анализ главных компонент референсной панели гаплотипов *M. mulatta* ($n = 618$). Синим цветом обозначены образцы индийского происхождения ($n = 418$), красным — образцы китайского происхождения ($n = 200$)

на анализ редких генетических вариантов. Таким образом, созданная панель обеспечивает высокую точность импутации для широкого спектра вариантов и может быть использована в популяционных и ассоциативных исследованиях макак-резус.

Анализ популяционной структуры

Для оценки генетической структуры референсной панели был проведен анализ главных компонент (PCA) на основе 33457491 SNP-маркера. Первая главная компонента (PC1, 68,5% общей дисперсии) четко разделяет образцы индийского и китайского происхождения на два обособленных кластера (рис. 2), что согласуется с ранее опубликованными данными о значительной генетической дивергенции между двумя основными географическими линиями *M. mulatta* [27]. Вторая главная компонента (PC2, 5,2% дисперсии) выявляет дополнительную структуру внутри каждой из групп: образцы индийского происхождения демонстрируют относительно компактное распределение, тогда как китайские образцы характеризуются большим разбросом вдоль PC2, что соответствует более высокой генетической гетерогенности восточноазиатской линии вследствие ее подвидового разнообразия [28].

Апробация панели на независимых образцах

Для апробации панели на независимых данных была проведена импутация генотипов 10 особей *M. mulatta* из коллекции Курчатовского комплекса медицинской приматологии НИЦ «Курчатowski институт», не входивших в состав референсной панели. Секвенирование выполняли со средним покрытием 0,6х, что соответствует схеме

эксперимента, обеспечивающей оптимальный баланс между стоимостью и достоверностью генотипирования при использовании секвенирования с низким покрытием [7, 12]. Импутацию выполняли с использованием алгоритма GLIMPSE2 на созданной референсной панели.

Качество импутации оценивали по показателю INFO/RSQ, отражающему степень статистической уверенности для каждого импутированного варианта и принимающему значения от 0 до 1. В исследованной выборке значения INFO/RSQ варьировали от 0,862 до 0,934 при среднем значении 0,898. Все идентифицированные варианты характеризовались значениями INFO > 0,8, соответствующими строгому порогу отбора для частых и среднечастотных вариантов [29, 30].

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В настоящей работе создана и валидирована фазированная референсная панель гаплотипов для импутации генотипов у *Macaca mulatta*, включающая 618 особей и 33,4 млн SNP-маркеров, обработанных по единому стандартизированному пайплайну. Среди доступных ресурсов для импутации генотипов данного вида следует выделить панель из 741 особи индийской линии [13] и масштабную когорту из 919 китайских макак с фенотипическими данными, представляющую собой ресурс, ориентированный на GWAS, но не представляющий собой единой фазированной импутационной панели [31]. Оба ресурса не охватывают обе географические линии вида в рамках единого стандартизированного пайплайна обработки. Между тем значительная часть лабораторных популяций *M. mulatta* в исследовательских центрах мира формировалась на протяжении десятилетий путем завоза животных из разных регионов и последующего

неконтролируемого скрещивания, что нередко приводило к неизвестному или смешанному происхождению животных [14, 27, 32]. Это обуславливает практическую потребность именно в интегрированном ресурсе, охватывающем обе линии, разработанном в настоящей работе.

Полученные показатели точности импутации полностью согласуются с опубликованными данными. Наибольшая точность достигнута для образцов индийского происхождения: R^2 0,95–0,96 для частых вариантов уже при покрытии 0,7х, что сопоставимо с результатами, полученными по индийской панели [13]. Для китайской линии показатели несколько ниже, что закономерно отражает ее меньшую представленность. Разработанная панель значительно превосходит ранее доступные ресурсы по плотности маркеров, обеспечивая покрытие более 33 млн SNP, что соответствует уровню современных референсных панелей для сельскохозяйственных животных и открывает возможности для высокоразрешающего популяционного и ассоциативного анализа. При покрытии уже 0,6х панель обеспечивает качество импутации, сопоставимое с коммерческими SNP-чипами, что делает подход экономически эффективным для генотипирования больших когорт.

Практическая ценность созданного ресурса определяется широким спектром биомедицинских задач, в которых *M. mulatta* выступает основной моделью: изучение инфекционных заболеваний, исследования в области нейробиологии, фармакологии и токсикологии. Полногеномное генотипирование на основе импутации открывает возможность популяционного скрининга животных на носительство вариантов, ортологических патогенных мутаций человека [33], а также контроля межиндивидуальной фармакокинетической вариабельности, обусловленной полиморфными вариантами ферментов метаболизма лекарственных веществ [34]. Этот подход доказал свою эффективность и на других модельных и сельскохозяйственных видах. Для крупного рогатого скота создана мультипородная панель на выборке из 2976 животных, включающая 61,8 млн SNP-маркеров. При использовании алгоритма GLIMPSE2, также примененного в данной работе, эта панель продемонстрировала конкордантность более 99% уже при покрытии 0,1х [35]. Для свиней описаны панели SWIM (2259 особей из 44 пород) [36] и PHARP, построенная на основе данных 1181 особи из 71 популяции [37]. Все эти панели сформированы путем объединения данных нескольких общедоступных проектов по генотипированию различных популяций, что соответствует подходу, реализованному в настоящем исследовании. Растущий интерес к данному направлению отмечается и в отечественных исследованиях: перспективы секвенирования с низким покрытием для геномной селекции крупного рогатого скота обсуждаются в обзоре российских авторов [38], а импутированные полногеномные данные уже

применяются для ассоциативного анализа хозяйственно значимых признаков в российских стадах [39]. Успешный опыт создания подобных ресурсов для других видов подтверждает, что разработанный в настоящей работе биоинформатический пайплайн воспроизводим и может быть адаптирован для создания аналогичных панелей для других видов приматов, применяемых в биомедицинских исследованиях.

Основные направления развития ресурса — расширение китайской компоненты панели на основе общедоступных данных, а также включение образцов из отечественных приматологических коллекций, что позволит повысить точность импутации для популяций со смешанным происхождением. Созданная референсная панель и разработанный биоинформатический подход закладывают инфраструктурную основу для проведения масштабных полногеномных ассоциативных и популяционно-генетических исследований на макаках-резус в отечественных лабораториях.

Ограничения исследования

К числу основных ограничений выполненной работы относятся неравномерная представленность географических линий в составе панели: количество образцов индийского происхождения (418) более чем в два раза превышает число китайских (200). Это обусловило более высокую точность импутации для индийской линии. Кроме того, панель сформирована на основе данных зарубежных репозиториях (BCM, OHSU, KIZ), в связи с чем для более точного отражения локальных популяций требуется последующее включение образцов из отечественных приматологических коллекций. Оценка точности методом искусственного прореживания выполнена на двух образцах в диапазоне покрытия 0,1–0,9х, тогда как апробация на реальных данных низкого покрытия проведена на 10 независимых образцах при среднем покрытии 0,6х. В связи с этим дальнейшая проверка на более крупной и гетерогенной выборке позволит дополнительно подтвердить воспроизводимость полученных результатов.

Выводы

Создана и валидирована фазированная референсная панель гаплотипов для импутации генотипов у *Macaca mulatta*, включающая 618 особей и 33,4 млн SNP-маркеров обеих основных географических линий вида, обработанных по единому стандартизированному пайплайну. Панель обеспечивает высокую точность импутации при сверхнизком покрытии и готова к практическому применению в популяционно-генетических и ассоциативных исследованиях. Разработанный пайплайн может служить шаблоном для создания аналогичных ресурсов для других видов приматов, используемых в биомедицинских исследованиях.

Литература

- Gardner MB, Luciw PA. Macaque models of human infectious disease. *ILAR J*. 2008; 49: 220–55. Available from: <https://doi.org/10.1093/ilar.49.2.220>.
- Joho D, Sato K, Kumita W, Saido TC, Sasaki E, Sasaguri H. Advances in non-human primate models for Alzheimer's disease research. *Neurosci Res*. 2025; 221: 104984. Available from: <https://doi.org/10.1016/j.neures.2025.104984>.
- Vallender EJ, Hotchkiss CE, Lewis AD, Rogers J, Stern JA, Peterson SM, et al. Nonhuman primate genetic models for the study of rare diseases. *Orphanet J Rare Dis*. 2023; 18: 20. Available from: <https://doi.org/10.1186/s13023-023-02619-3>.
- Warren WC, Harris RA, Haukness M, Fiddes IT, Murali SC, Fernandes J, et al. Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. *Science*. N.Y. 2020; 370: eabc6617. Available from: <https://doi.org/10.1126/science.abc6617>.
- Bimber BN, Yan MY, Peterson SM, Ferguson B. mGAP: the macaque genotype and phenotype resource, a framework for accessing and interpreting macaque variant data, and identifying new models of human disease. *BMC Genomics*. 2019; 20: 176. Available from: <https://doi.org/10.1186/s12864-019-5559-7>.
- Bimber BN, Raboin MJ, Letaw J, Nevenon KA, Spindel JE, McCouch SR, et al. Whole-genome characterization in pedigreed non-human primates using genotyping-by-sequencing (GBS) and imputation. *BMC Genomics*. 2016; 17: 676. Available from: <https://doi.org/10.1186/s12864-016-2966-x>.
- Rubinacci S, Ribeiro DM, Hofmeister RJ, Delaneau O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat Genet*. 2021; 53: 120–6. Available from: <https://doi.org/10.1038/s41588-020-00756-0>.
- Mitt M, Kals M, Pärn K, Gabriel SB, Lander ES, Palotie A, et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet*. 2017; 25: 869–76. Available from: <https://doi.org/10.1038/ejhg.2017.51>.
- Zhong X, Peng J, Shen QS, Chen J-Y, Gao H, Luan X, et al. RhesusBase PopGateway: Genome-Wide Population Genetics Atlas in Rhesus Macaque. *Mol Biol Evol*. 2016; 33: 1370–5. Available from: <https://doi.org/10.1093/molbev/msw025>.
- Du L, Guo T, Liu Q, Li J, Zhang X, Xing J, et al. MACSNVdb: a high-quality SNV database for interspecies genetic divergence investigation among macaques. *Database (Oxford)*. 2020; 2020: baaa027. Available from: <https://doi.org/10.1093/database/baaa027>.
- Quick C, Anugu P, Musani S, Weiss ST, Burchard EG, White MJ, et al. Sequencing and imputation in GWAS: Cost-effective strategies to increase power and genomic coverage across diverse populations. *Genet Epidemiol*. 2020; 44: 537–49. Available from: <https://doi.org/10.1002/gepi.22326>.
- Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, et al. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat Genet*. 2012; 44: 631–5. Available from: <https://doi.org/10.1038/ng.2283>.
- Watowich MM, Chiou KL, Graves B, Montague MJ, Brent LJJ, Higham JP, et al. Best practices for genotype imputation from low-coverage sequencing data in natural populations. *Mol Ecol Resour*. 2025; 25: e13854. Available from: <https://doi.org/10.1111/1755-0998.13854>.
- Kanthaswamy S, Trask JS, Ross CT, Kou A, Houghton P, Smith DG, et al. A large-scale SNP-based genomic admixture analysis of the captive rhesus macaque colony at the California National Primate Research Center. *Am J Primatol*. 2012; 74: 747–57. Available from: <https://doi.org/10.1002/ajp.22025>.
- Kanthaswamy S, Satkoski J, Kou A, Malladi V, Glenn Smith D. Detecting signatures of inter-regional and inter-specific hybridization among the Chinese rhesus macaque specific pathogen-free (SPF) population using single nucleotide polymorphism (SNP) markers. *J Med Primatol*. 2010; 39: 252–65. Available from: <https://doi.org/10.1111/j.1600-0684.2010.00430.x>.
- Wang J, Wang M, Moshiri A, Harris RA, Raveendran M, Nguyen T, et al. Genetic diversity of 1,845 rhesus macaques improves genetic variation interpretation and identifies disease models. *Nat Commun*. 2024; 15: 5658. Available from: <https://doi.org/10.1038/s41467-024-49922-6>.
- Rogers J. Genomic resources for rhesus macaques (*Macaca mulatta*). *Mamm Genome*. 2022; 33: 91–9. Available from: <https://doi.org/10.1007/s00335-021-09922-z>.
- Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res*. 2022; 50: D387–90. Available from: <https://doi.org/10.1093/nar/gkab1053>.
- David Y, Alisha A, Awais A, Rajkumar D, Dipayan G, Muhammad H, et al. The European Nucleotide Archive in 2025. *Nucleic Acids Res*. 2026; 54: D120–7. Available from: <https://doi.org/10.1093/nar/gkae1295>.
- CNCB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. *Nucleic Acids Res*. 2025; 53: D30–44. Available from: <https://doi.org/10.1093/nar/gkaf978>.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018; 34: 1884–90. Available from: <https://doi.org/10.1093/bioinformatics/bty560>.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9: 357–9. Available from: <https://doi.org/10.1038/nmeth.1923>.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021; 10: giab008. Available from: <https://doi.org/10.1093/gigascience/giab008>.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011; 27: 2156–8. Available from: <https://doi.org/10.1093/bioinformatics/btr330>.
- Hofmeister RJ, Ribeiro DM, Rubinacci S, Delaneau O. Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nat Genet*. Nature Publishing Group. 2023; 55: 1243–9. Available from: <https://doi.org/10.1038/s41588-023-01415-w>.
- Rubinacci S, Hofmeister RJ, Sousa da Mota B, Delaneau O. Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes. *Nat Genet*. 2023; 55: 1088–90. Available from: <https://doi.org/10.1038/s41588-023-01438-3>.
- Smith DG. Genetic characterization of Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). *Comp Med*. 2005; 55: 227–30.
- Liu Z, Tan X, Orozco-terWengel P, Zhou X, Zhang L, Tian S, et al. Population genomics of wild Chinese rhesus macaques reveals a dynamic demographic history and local adaptation, with implications for biomedical research. *Gigascience*. 2018; 7: giy106. Available from: <https://doi.org/10.1093/gigascience/giy106>.
- Stahl K, Papiol S, Budde M, Heilbronner M, Oraki Kohshour M, Falkai P, et al. Aggregating single nucleotide polymorphisms improves filtering for false-positive associations postimputation. *G3 (Bethesda)*. Bethesda, Md. 2025; 15: jkaf043. Available from: <https://doi.org/10.1093/g3journal/jkaf043>.
- Fu J, Zhang Q, Wang J, Wang M, Zhang B, Zhu W, и др. Cross-ancestry genome-wide association studies of brain imaging phenotypes. *Nat Genet*. 2024; 56: 1110–20. Available from: <https://doi.org/10.1038/s41588-024-01766-y>.
- Zhang B-L, Chen Y, Zhang Y, Qiao Y, Wu Y, Zhang Y, et al. Forward and reverse genomic screens enhance the understanding of phenotypic variation in a large Chinese rhesus macaque cohort. *Nat Commun*. Nature Publishing Group. 2025; 16: 8703. Available from: <https://doi.org/10.1038/s41467-025-63747-x>.
- Lyke MM, Bagwell A, Newman D, Galindo S, Church T, Christensen C, et al. Mergers and migrations: drivers of population genetic structure in a captive rhesus macaque colony. *BMC Genomics*. 2026; Available from: <https://doi.org/10.1186/s12864-026-13090-2>.
- Wang J, Wang M, Moshiri A, Harris RA, Raveendran M, Nguyen T, et al. Genetic diversity of 1,845 rhesus macaques improves genetic variation interpretation and identifies disease models. *Nat Commun*. 2024; 15: 5658. Available from: <https://doi.org/10.1038/s41467-024-49922-6>.
- Uno Y, Uehara S, Yamazaki H. Genetic polymorphisms of drug-metabolizing cytochrome P450 enzymes in cynomolgus and rhesus monkeys and common marmosets in preclinical studies

- for humans. *Biochem Pharmacol.* 2018; 153: 184–95. Available from: <https://doi.org/10.1016/j.bcp.2017.12.015>.
35. Zhang Z, Wang A, Hu H, Wang L, Gong M, Yang Q, et al. The efficient phasing and imputation pipeline of low-coverage whole genome sequencing data using a high-quality and publicly available reference panel in cattle. *Animal Research and One Health.* 2023; 1: 4–16. Available from: <https://doi.org/10.1002/aro2.8>.
 36. Ding R, Savegnago R, Liu J, Long N, Tan C, Cai G, et al. The SWine IMputation (SWIM) haplotype reference panel enables nucleotide resolution genetic mapping in pigs. *Commun Biol.* 2023; 6: 577. Available from: <https://doi.org/10.1038/s42003-023-04933-9>.
 37. Wang Z, Zhang Z, Chen Z, Sun J, Cao C, Wu F, et al. PHARP: a pig haplotype reference panel for genotype imputation. *Sci Rep.* 2022; 12: 12645. Available from: <https://doi.org/10.1038/s41598-022-15851-x>.
 38. Kostyunina O, Koldichev N, Nemkovskiy G, Traspov A, Ermilov A, Bakoev F, et al. Low-Coverage Whole-Genome Sequencing (lcWGS) in Cattle: Analysis of Potential and Prospects for Application. *Animals (Basel).* 2025; 15: 3538. Available from: <https://doi.org/10.3390/ani15243538>.
 39. Genome-wide association study of carcass traits in an Angus beef cattle Russian herd using imputed Whole-Genome Sequence data. The Thirteenth International Multiconference. Institute of Cytology and Genetics, the Siberian Branch of the Russian Academy of Sciences; 2022 [дата обращения: 4 август 2026 г.]. Available from: <https://doi.org/10.18699/SBB-2022-397>.
- ## References
1. Gardner MB, Luciw PA. Macaque models of human infectious disease. *ILAR J.* 2008; 49: 220–55. Available from: <https://doi.org/10.1093/ilar.49.2.220>.
 2. Joho D, Sato K, Kumita W, Saido TC, Sasaki E, Sasaguri H. Advances in non-human primate models for Alzheimer's disease research. *Neurosci Res.* 2025; 221: 104984. Available from: <https://doi.org/10.1016/j.neures.2025.104984>.
 3. Vallender EJ, Hotchkiss CE, Lewis AD, Rogers J, Stern JA, Peterson SM, et al. Nonhuman primate genetic models for the study of rare diseases. *Orphanet J Rare Dis.* 2023; 18: 20. Available from: <https://doi.org/10.1186/s13023-023-02619-3>.
 4. Warren WC, Harris RA, Haukness M, Fiddes IT, Murali SC, Fernandes J, et al. Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. *Science, N.Y.* 2020; 370: eabc6617. Available from: <https://doi.org/10.1126/science.abc6617>.
 5. Bimber BN, Yan MY, Peterson SM, Ferguson B. mGAP: the macaque genotype and phenotype resource, a framework for accessing and interpreting macaque variant data, and identifying new models of human disease. *BMC Genomics.* 2019; 20: 176. Available from: <https://doi.org/10.1186/s12864-019-5559-7>.
 6. Bimber BN, Raboin MJ, Letaw J, Nevenon KA, Spindel JE, McCouch SR, et al. Whole-genome characterization in pedigreed non-human primates using genotyping-by-sequencing (GBS) and imputation. *BMC Genomics.* 2016; 17: 676. Available from: <https://doi.org/10.1186/s12864-016-2966-x>.
 7. Rubinacci S, Ribeiro DM, Hofmeister RJ, Delaneau O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat Genet.* 2021; 53: 120–6. Available from: <https://doi.org/10.1038/s41588-020-00756-0>.
 8. Mitt M, Kals M, Pärn K, Gabriel SB, Lander ES, Palotie A, et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet.* 2017; 25: 869–76. Available from: <https://doi.org/10.1038/ejhg.2017.51>.
 9. Zhong X, Peng J, Shen QS, Chen J-Y, Gao H, Luan X, et al. RhesusBase PopGateway: Genome-Wide Population Genetics Atlas in Rhesus Macaque. *Mol Biol Evol.* 2016; 33: 1370–5. Available from: <https://doi.org/10.1093/molbev/msw025>.
 10. Du L, Guo T, Liu Q, Li J, Zhang X, Xing J, et al. MACSNDVdb: a high-quality SNV database for interspecies genetic divergence investigation among macaques. *Database (Oxford).* 2020; 2020: baaa027. Available from: <https://doi.org/10.1093/database/baaa027>.
 11. Quick C, Anugu P, Musani S, Weiss ST, Burchard EG, White MJ, et al. Sequencing and imputation in GWAS: Cost-effective strategies to increase power and genomic coverage across diverse populations. *Genet Epidemiol.* 2020; 44: 537–49. Available from: <https://doi.org/10.1002/gepi.22326>.
 12. Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, et al. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat Genet.* 2012; 44: 631–5. Available from: <https://doi.org/10.1038/ng.2283>.
 13. Watowich MM, Chiou KL, Graves B, Montague MJ, Brent LJJ, Higham JP, et al. Best practices for genotype imputation from low-coverage sequencing data in natural populations. *Mol Ecol Resour.* 2025; 25: e13854. Available from: <https://doi.org/10.1111/1755-0998.13854>.
 14. Kanthaswamy S, Trask JS, Ross CT, Kou A, Houghton P, Smith DG, et al. A large-scale SNP-based genomic admixture analysis of the captive rhesus macaque colony at the California National Primate Research Center. *Am J Primatol.* 2012; 74: 747–57. Available from: <https://doi.org/10.1002/ajp.22025>.
 15. Kanthaswamy S, Satkoski J, Kou A, Malladi V, Glenn Smith D. Detecting signatures of inter-regional and inter-specific hybridization among the Chinese rhesus macaque specific pathogen-free (SPF) population using single nucleotide polymorphic (SNP) markers. *J Med Primatol.* 2010; 39: 252–65. Available from: <https://doi.org/10.1111/j.1600-0684.2010.00430.x>.
 16. Wang J, Wang M, Moshiri A, Harris RA, Raveendran M, Nguyen T, et al. Genetic diversity of 1,845 rhesus macaques improves genetic variation interpretation and identifies disease models. *Nat Commun.* 2024; 15: 5658. Available from: <https://doi.org/10.1038/s41467-024-49922-6>.
 17. Rogers J. Genomic resources for rhesus macaques (*Macaca mulatta*). *Mamm Genome.* 2022; 33: 91–9. Available from: <https://doi.org/10.1007/s00335-021-09922-z>.
 18. Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res.* 2022; 50: D387–90. Available from: <https://doi.org/10.1093/nar/gkab1053>.
 19. David Y, Alisha A, Awais A, Rajkumar D, Dipayan G, Muhammad H, et al. The European Nucleotide Archive in 2025. *Nucleic Acids Res.* 2026; 54: D120–7. Available from: <https://doi.org/10.1093/nar/gkaf1295>.
 20. CNCB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. *Nucleic Acids Res.* 2025; 53: D30–44. Available from: <https://doi.org/10.1093/nar/gkae978>.
 21. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018; 34: 1884–90. Available from: <https://doi.org/10.1093/bioinformatics/bty560>.
 22. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012; 9: 357–9. Available from: <https://doi.org/10.1038/nmeth.1923>.
 23. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021; 10: giab008. Available from: <https://doi.org/10.1093/gigascience/giab008>.
 24. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011; 27: 2156–8. Available from: <https://doi.org/10.1093/bioinformatics/btr330>.
 25. Hofmeister RJ, Ribeiro DM, Rubinacci S, Delaneau O. Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nat Genet.* 2023; 55: 1243–9. Available from: <https://doi.org/10.1038/s41588-023-01415-w>.
 26. Rubinacci S, Hofmeister RJ, Sousa da Mota B, Delaneau O. Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes. *Nat Genet.* 2023; 55: 1088–90. Available from: <https://doi.org/10.1038/s41588-023-01438-3>.
 27. Smith DG. Genetic characterization of Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). *Comp Med.* 2005; 55: 227–30.
 28. Liu Z, Tan X, Orozco-terWengel P, Zhou X, Zhang L, Tian S,

- et al. Population genomics of wild Chinese rhesus macaques reveals a dynamic demographic history and local adaptation, with implications for biomedical research. *Gigascience*. 2018; 7: giy106. Available from: <https://doi.org/10.1093/gigascience/giy106>.
29. Stahl K, Papiol S, Budde M, Heilbronner M, Oraki Kohshour M, Falkai P, et al. Aggregating single nucleotide polymorphisms improves filtering for false-positive associations postimputation. *G3* (Bethesda). Bethesda, Md. 2025; 15: jkaf043. Available from: <https://doi.org/10.1093/g3journal/jkaf043>.
 30. Fu J, Zhang Q, Wang J, Wang M, Zhang B, Zhu W, и др. Cross-ancestry genome-wide association studies of brain imaging phenotypes. *Nat Genet*. 2024; 56: 1110–20. Available from: <https://doi.org/10.1038/s41588-024-01766-y>.
 31. Zhang B-L, Chen Y, Zhang Y, Qiao Y, Wu Y, Zhang Y, et al. Forward and reverse genomic screens enhance the understanding of phenotypic variation in a large Chinese rhesus macaque cohort. *Nat Commun*. Nature Publishing Group. 2025; 16: 8703. Available from: <https://doi.org/10.1038/s41467-025-63747-x>.
 32. Lyke MM, Bagwell A, Newman D, Galindo S, Church T, Christensen C, et al. Mergers and migrations: drivers of population genetic structure in a captive rhesus macaque colony. *BMC Genomics*. 2026; Available from: <https://doi.org/10.1186/s12864-026-13090-2>.
 33. Wang J, Wang M, Moshiri A, Harris RA, Raveendran M, Nguyen T, et al. Genetic diversity of 1,845 rhesus macaques improves genetic variation interpretation and identifies disease models. *Nat Commun*. 2024; 15: 5658. Available from: <https://doi.org/10.1038/s41467-024-49922-6>.
 34. Uno Y, Uehara S, Yamazaki H. Genetic polymorphisms of drug-metabolizing cytochrome P450 enzymes in cynomolgus and rhesus monkeys and common marmosets in preclinical studies for humans. *Biochem Pharmacol*. 2018; 153: 184–95. Available from: <https://doi.org/10.1016/j.bcp.2017.12.015>.
 35. Zhang Z, Wang A, Hu H, Wang L, Gong M, Yang Q, et al. The efficient phasing and imputation pipeline of low-coverage whole genome sequencing data using a high-quality and publicly available reference panel in cattle. *Animal Research and One Health*. 2023; 1: 4–16. Available from: <https://doi.org/10.1002/aro2.8>.
 36. Ding R, Savegnago R, Liu J, Long N, Tan C, Cai G, et al. The SWine IMputation (SWIM) haplotype reference panel enables nucleotide resolution genetic mapping in pigs. *Commun Biol*. 2023; 6: 577. Available from: <https://doi.org/10.1038/s42003-023-04933-9>.
 37. Wang Z, Zhang Z, Chen Z, Sun J, Cao C, Wu F, et al. PHARP: a pig haplotype reference panel for genotype imputation. *Sci Rep*. 2022; 12: 12645. Available from: <https://doi.org/10.1038/s41598-022-15851-x>.
 38. Kostyunina O, Koldichev N, Nemkovskiy G, Traspov A, Ermilov A, Bakoev F, et al. Low-Coverage Whole-Genome Sequencing (lcWGS) in Cattle: Analysis of Potential and Prospects for Application. *Animals* (Basel). 2025; 15: 3538. Available from: <https://doi.org/10.3390/ani15243538>.
 39. Genome-wide association study of carcass traits in an Angus beef cattle Russian herd using imputed Whole-Genome Sequence data. The Thirteenth International Multiconference. Institute of Cytology and Genetics, the Siberian Branch of the Russian Academy of Sciences; 2022 [дата обращения: 4 август 2026 г.]. Available from: <https://doi.org/10.18699/SBB-2022-397>.