

ПАНГЕНОМ ЧЕЛОВЕКА И ГЕНОМНАЯ МЕДИЦИНА



Владимир Савченко,
член-корреспондент
НАН Беларуси,
профессор



Междисциплинарный характер инноваций

Задача расшифровки генома человека потребовала международной кооперации в проведении процесса секвенирования общей последовательности нуклеотидов в молекулах геномной ДНК, включающей около 3,2 млрд оснований.

В начале XX ст. геном стал исследоваться путем сравнения внешних и прочих признаков у близнецов, родителей и детей, других родственников – целых семей. Данные наблюдения стали основой для понимания путей реализации наследственной информации, содержащейся в их геномах. Этой цели служило также изучение различных наследственных болезней, передававшихся от родителей к потомству – как сразу, так и на протяжении периода их жизнедеятельности. Постепенно накапливались клинические свидетельства о возникновении новых заболеваний в связи с появлением мутаций в геноме человека, особенно под влиянием химического или радиационного загрязнения окружающей среды.

Использование оптики микроскопа дало возможность установить связь между изменением числа или структуры 46 хромосом и определенными наследственными патологиями. Понимание зависимости между структурными изменениями хромосом или генов и состоянием здоровья человека позволило искать способы лечения генетически обусловленных заболеваний. В случае, когда мутация гена меняет активность или функцию его продукта, стало возможным избежать развития болезни путем назначения нормального химического аналога, способного участвовать в метаболизме и восполнять потерянную, но нужную функцию гена.

За последние несколько десятилетий в результате проникновения и применения физико-химических методов и информационных технологий в изучении и анализе молекулярной структуры генома человека начался нарастающий, лавинообразный процесс разработки и использования новых геномных технологий [1–2]. В результате коллективного труда на международной арене появились и были успешно задействованы для расшифровки генома человека три поколения, или фазы, технологий секвенирования последовательностей нуклеотидов в молекулах геномной ДНК. Эти инновации включали новизну, но вырастали из предыдущего опыта, ориентировались на улучшение технического устройства, автоматизацию, повышенную скорость секвенирова-

ния, миниатюрность устройств, надежность, снижение расходов.

Первое поколение упомянутых выше технологий было разработано сначала Алланом Максом и Уолтером Гилбертом в 1976–1977 гг., а также в 1977 г. Фредериком Сенгером. В 1980 г. Гилберт и Сенгер получили Нобелевскую премию за вклад в определение последовательности нуклеотидов нуклеиновых кислот. Классическая технология Maxam and Gilbert была основана на процессе гидролиза и в настоящее время уже не используется. Базой для технологии Sanger выступает процесс полимеризации и пока находит применение только в виде автоматизированной версии.

Второе поколение, известное под наименованием NGS (Next Generation Sequencing), основано на принципе «вымывай и сканируй один нуклеотид за другим» и нуждается также в дополнительной амплификации фрагментов ДНК с помощью ПЦР. Технологии теперь располагают компактными автоматизированными секвенаторами с высокой скоростью расшифровки длинных последовательностей нуклеотидов при полной расшифровке геномов (Illumina, SOLiD, Pacific Biosciences, IonTorrent).

Третье поколение технологий секвенирования получило название SMS (Single Molecule Sequencing). Они работают на отдельных молекулах ДНК и не требуют процесса амплификации изучаемой последовательности нуклеотидов, что снижает уровень ошибок. Используют разные физико-химические принципы исследования биологических структур и процессов.

Если технологии первого поколения могут еще включаться в рабочий процесс в отдельных лабораториях, то второго (NGS) сегодня доминируют на рынке – главным образом благодаря высокой скорости и значительному снижению стоимости процесса секвенирования геномной ДНК. Технологии третьего поколения продолжают изучаться и быстро совершенствоваться. За ними, судя по всему, будущее [3].

Секвенирование генома человека

Следует напомнить, что Международный проект «Геном человека» (ПГЧ) начал функционировать еще в 1998 г. и выполнялся Международным консорциумом лабораторий с целью полной расшифровки последовательностей нуклеотидов геномной ДНК. Его финансирование составляло более 3 млрд долл., а завершение работ ознаменовалось публикацией в 2001 г. в научных журналах и в Интернете

для открытого использования расшифрованных последовательностей оснований в молекулах ДНК из 23 хромосом гаплоидного генома половых клеток. Это были первые «грязные» последовательности ДНК генома человека, включавшие также возникшие пропуски и неточности. Работа продолжалась, и в 2004 г. Международный консорциум и частная компания Celera Genomics опубликовали свои, дополнительно проверенные и уточненные последовательности оснований гаплоидного генома человека. Он в качестве референтного объекта включает 3 054 815 472 пар оснований молекул ДНК половых клеток, несущих хромосому X. Диплоидные клетки тела имеют двойной набор молекул ДНК – от матери и отца, и общие последовательности нуклеотидов их ДНК превышают 6 млрд пар оснований. Было объявлено, что геном человека содержит от 20 тыс. до 25 тыс. генов, хотя в начале проекта предполагалось, что их может быть 100 тыс.

Так завершился на рубеже столетий первый этап расшифровки генома человека. Результаты секвенирования геномной ДНК человека и кодирующие последовательности генов и протеинов находятся в базе данных U.S. National Center for Biotechnology Information (США) со свободным доступом через Интернет, а также в аналогичных центрах в Европе и Азии.

На протяжении первых двух десятилетий XXI в. работы по изучению генома человека успешно продолжались и расширялись в рамках ряда новых проектов. Они создавали научную базу для практического использования новых знаний по тестированию генома для его терапии, диагностики наследственных заболеваний, создания новых лекарств направленного действия и их доставки в пораженный орган, стимулировали развитие регенеративной и персональной медицины. Среди работ, продолжающих изучение генома, следует отметить проект «1000 геномов» (2008; геномы представителей разных этнических групп), проект синтеза генома человека (GP-Write, 2016; синтез генома химическим способом), британский проект «100 тыс. геномов» (2012; выявление редких болезней, рака, инфекций), проект «100 тыс. азиатских геномов» (2016; представители 19 стран). Международный консорциум по геномике рака подготовил свое руководство для использования в клиниках для лечения пациентов. Следует отметить также национальные медицинские проекты, такие как «Эстонский геномный проект» и «Катарская геномная программа». Все эти исследования, нацеленные на изучение генетических различий между индивидами, уделяли особое внимание полиморфизму нуклеотидов.

Важным результатом выполнения различных проектов стало совершенствование геномных технологий, снижение цены и повышение эффективности технологии секвенирования ДНК примерно на два порядка. В частности, это произошло за счет использования флуоресцентных меток, их лазерного считывания и капиллярного электрофореза для разделения оснований. Таким образом был открыт технологический путь для использования алгоритмов геномной медицины в лечении пациентов. Ряд развитых стран, включая США, Францию, Соединенное Королевство и Германию, первыми предприняли шаги в направлении создания для своих граждан национальных служб геномной медицины. Как показывает практика, именно эта область становится источником инноваций для модернизации систем здравоохранения передовых современных государств.

Структура генома человека

Секвенирование молекул геномной ДНК и анализ полученных последовательностей оснований позволили вскрыть структурную организацию ядерного генома человека. Его геномная ДНК структурирована в диплоидном кариотипе человека, включающем 23 пары хромосом, причем одна пара отвечает за развитие зиготы либо по женскому пути (XX), либо по мужскому (XY). Клеточные органеллы митохондрии, участвующие в энергетическом обмене, имеют свою собственную ДНК, похожую на геном бактерии. По более полным оценкам, гаплоидный геном человека включает 3 117 275 501 пар оснований. Эти основания группируются и кодируют гены, которые после транскрипции и-РНК и трансляции закодированной последовательности аминокислот превращаются в молекулы протеинов. При этом из молекулы информационной РНК не вырезаются кодирующие интроны, а кодирующие последовательности объединяются в экзоны, которые формируют сначала полипептидную цепь аминокислот, а затем 3-мерную молекулу белка. Часть молекул геномной ДНК кодирует синтез молекул рибосомной РНК, транспортной РНК, регуляторных РНК, коротких ядерных РНК, которые не транслируются, но участвуют в жизнедеятельности клеток. Последовательности ДНК играют регуляторную и структурную роль в центромерах, теломерах, выполняют роль сигнала для начала репликации. Геном наполнен псевдогенами, «молчащими» геномами вирусов, рядом ретровирусов с полезными функциями.

В 2021 г. был секвенирован геном женщины, который не содержал мужской Y-хромосомы; анализ подтвердил наличие в нем 63 494 последовательностей нуклеотидов, характерных для разных генов, в том числе 19 969 – кодирующих синтез протеинов. Геномная ДНК включала также регуляторные последовательности, гены для синтеза различных клеточных РНК, последовательности генов с неизвестной функцией, а также множества интронов, повторов LINE и SINE. Лишь 1,5% геномной ДНК занимают последовательности для генов, кодирующих синтез протеинов, из которых в основном и состоит организм человека.

По мере секвенирования новых геномов человека накапливались последовательности геномной ДНК, которые расширяли объем известного геномного биоразнообразия человека. Возникло предложение упорядочить эту информацию путем создания пангенома человека с целью его дальнейшего практического использования при реализации новых проектов.

Референтный пангеном человека

Пангеном вбирает в себя все разнообразие генов, зафиксированное в индивидуальных геномах, и поэтому характеризует как обязательные для жизнедеятельности вида функции, так и специфические, в том числе редкие и уникальные. Как система он включает ядро из обязательных генов, которые представлены и функционируют в каждом геноме изучаемого множества, оболочку из вспомогательных генов у двух и более членов множества и облако из единичных и редких генов, специфичных для множества. Компактное графическое изображение пангенома дает полное представление о разнообразии входящих в него генов и их взаимодействии с самыми разными условиями окружающей среды. Темп эволюции пангенома зависит от изменения частоты дупликации генов, возникновения новых и динамики потери менее эффективных, а также взаимодействия видового генома с множеством мобильных элементов, циркулирующих в популяции или биоценозе.

Осуществляется с помощью диаграммы Венна, которая напоминает распустившийся цветок с венчиком лепестков. В центре располагаются группы функциональных генов, обеспечивающих базовые экзистенциальные функции: размножения, процессов жизнедеятельности, адаптации к среде обитания и выживания. Различия в ядрах пангеномов у

штаммов или видов бактерий могут достигать 30%, поскольку они способны нести мобильные плазмиды ДНК с генами не только от своего, но и других видов. Такие пангеномы принято называть открытыми, в отличие от закрытых (у линий или семей отдельных видов эукариот) с минимальным количеством мобильных генов. Таким образом, формирование пангеномов позволяет наглядно представлять структуру генетического разнообразия, что важно при оценке эволюционных перспектив редких или охраняемых видов.

В оболочке пангенома сгруппированы гены, связанные с приспособлением к конкретным экологическим нишам. Но такие гены могут быть представлены не во всех геномах изучаемого множества. Это может быть результатом адаптации к локальным условиям фотосинтеза или симбиоза с другими видами в биоценозе. Оболочка и ядро пангенома отражает степень его вариабельности, в то время как во внешнем облаке сосредоточены только уникальные или специфичные гены, характерные лишь для отдельных редких геномов изучаемого системного пангенома. Различают пангеномы, сформированные на аннотациях отдельных генов и их семей, а также основанные на нуклеотидных последовательностях геномных ДНК.

При использовании первого метода необходимо сначала выявить в результате анализа аннотаций их ортологические ряды для построения графов. Ортологические группы кластеризуют с учетом порога вхождения гена в группу. На основе полученных данных строится структура пангенома. Если необходимо, проводится одно или несколько выравниваний последовательностей ядра пангенома, реконструкция филогенетических связей и повторные визуализации. Существует список различных компьютерных программ для проверки и выбора лучших вариантов анализа собранных данных и визуализации результатов.

Второй метод построения пангенома для различных по численности общностей организмов основывается на полученных последовательностях нуклеотидов их геномной ДНК. При этом делаются многократные выравнивания этих последовательностей с использованием графов. Это связано с тем, что у видов эукариот доминирует вариабельность из разных единичных нуклеотидов, и с их точной идентификацией возникают трудности в результате частых ошибок. Для второго метода также имеются специальные компьютерные программы и онлайн-сервисы, облегчающие анализ последовательностей, исправление выявленных ошибок и построение лучшего пангенома.

Геномные технологии позволили многократно секвенировать геном человека (как мужчин, так и женщин) из разных этнических групп, разнообразных географических и экологических зон и сейчас, когда назрела необходимость создания планетарного пангенома человека, выходят из состояния инструмента и становятся основой создания медицинского сервиса обследования и диагностики для предупреждения и лечения наследственных заболеваний. В 2022 г. была опубликована статья о необходимости подготовки пангенома человека, а в 2023 г. появился хорошо документированный коллективный отчет о проделанной работе по проектированию первого референтного пангенома [4, 5].

Последовательности нуклеотидов геномной ДНК человека, находящиеся в открытом доступе, свободно используются два последних десятилетия в различных лабораториях при выполнении новых международных и национальных исследовательских проектов. Однако теперь пришло время позаботиться о создании инновационного обобщенного пангенома человека. Это обычно делается для различных видов прокариот и эукариот в рамках изучения и охраны природного биоразнообразия. Поэтому для этой цели был образован Консорциум для создания референтного пангенома человека. Была поставлена цель дополнительно выверить имеющиеся длинные и короткие последовательности нуклеотидов геномной ДНК из ранее секвенированных мужских и женских геномов с построением графов и включением туда новых последовательностей теломер и центромер для повышения качества и полноты референтного пангенома человека. Он должен содержать более полное и корректное представление глобального геномного разнообразия, точно отражать ассоциации между генами и вызываемыми ими болезнями в популяциях человека, расширять масштабы изучения геномной изменчивости многочисленных повторов и полиморфных участков генома. Это необходимо как для проведения дальнейших медико-биологических исследований, так и организации точной геномной медицины для пациентов.

Первый референтный геном человека важен для инициации проектов и организации обмена данными, выполнения функциональной и структурной аннотации отдельных генов и их семей, осуществления эпигеномного анализа, для оценки и сравнения полученных результатов при создании новых или завершении начатых исследовательских проектов. В настоящее время в качестве референтного генома используется мозаика последовательностей

нуклеотидов GRCh38, полученная ранее всего от двух десятков индивидов. Однако она не может рассматриваться в качестве глобальной реферативной выборки для всего нашего вида, так как включает также многие короткие последовательности длиной 50–300 пар оснований, что затрудняет текущее изучение участков генома, содержащих высокое число повторов, границы которых становятся размытыми при стыковке коротких ансамблей нуклеотидов. Кроме того, многие последовательности нуклеотидов для отдельных хромосом содержат уже известное количество ошибок, пропусков и неполных данных. Это естественно для первой референтной последовательности нуклеотидов GRCh38, полученной 20 лет назад.

Структура референтного пангенома

При его проектировании и формировании следовало устранить все выявленные недостатки. Прежде всего авторы пангенома расширили в 2 раза охват исходной геномной базы – до 47 секвенированных женских и мужских геномов из разных географических зон планеты. Это дает возможность более полно судить о геномном и генетическом разнообразии нашего вида. Консорциум «Теломера к теломере» (T2T) успешно завершил расшифровку последовательностей нуклеотидов теломерных участков гаплоидного генома, и эту информацию включили в структуру пангенома человека. Использование данных T2T-CHM13 способствовало дополнительному открытию полиморфизма одиночных нуклеотидов (SNP) объемом 3,7 млн единиц и уточнению действительного числа копий вариантов (CNV) в сохраненных данных проекта «1000 геномов». Дополнены данные о частоте полиморфизма структурных вариантов в популяциях и мелких потерь и вставок нуклеотидов как источников геномного разнообразия.

Авторы осуществили полную проверку, верификацию и создали аккуратный ансамбль последовательностей нуклеотидов 47 геномов человека, отобранных для демонстрации существующего геномного разнообразия всем желающим. Это означает, что новые проекты по дальнейшему изучению нашего генома, как и использованию пангенома в целях диагностики и разработки протоколов терапии в рамках геномной медицины, могут свободно осуществляться. Сами геномные технологии становятся на путь развития от научного инструмента

для анализа к активному протоколу терапии с целью дальнейшего совершенствования существующей ныне системы здравоохранения.

Труд по составлению аннотаций генов и отдельных гаплотипов для множества из 47 разных геномов авторы выполнили с пониманием ценности такой информации для пользователей. При конструировании пангенома была использована визуализация путем построения графа с узлами в точках пересечения ребер, составленных на основе имеющихся последовательностей нуклеотидов. Каждый такой узел может иметь только две ориентации – вперед или назад. Между двумя соседними узлами возникает четыре крайние точки, отражающие их взаимную ориентацию. Последовательности гаплотипа можно представить в виде шагов прогулки по графу. Модель генома – в виде сочетания многочисленных вершин и множества соединяющих их ребер на основе последовательностей, которые были использованы для построения графа. Последовательности каждого гаплотипа находят свое место недалеко от его ближайшей вершины.

Визуализация системы пангенома заключается в выявлении кластеризации вершин и ребер в пространстве графа, что позволяет сделать анализ его структуры. Мы сразу начинаем понимать, каким сложным и многокластерным образованием является спроектированный и затем построенный пангеном. Следует обратить особое внимание на самые крупные вершины и самые многократные связи, поскольку именно на них держатся все остальные составляющие пангенома.

Авторский коллектив проектирования и осуществления грандиозного проекта пангенома проделал масштабную работу по сбору, верификации и анализу коротких и длинных последовательностей нуклеотидов, их тщательному анализу и использованию для построения трех разных моделей графов для их сравнения и более полной визуализации всей накопленной геномной информации. Разработчики пангенома выбрали и реализовали три разные граф-конструкции: Minigraph, MC и PGGB.

Minigraph использует minimap2 algorithm³⁴ для быстрого переноса выбранных известных последовательностей нуклеотидов GRCh38 и CHM13 в пространство создаваемого графа. Все обнаруженные новые структурные варианты размером ≥ 50 пар оснований постепенно вносятся дополнительно. Но Minigraph не создает своих последовательностей и не дополняет имеющихся линейных структур из обоих источников. Это означает, что референтные последовательности нуклеотидов сохраняются в исход-



Ансамбль последовательностей нуклеотидов
47 геномов человека

Источник: <https://edition.cnn.com/2023/05/10/health/human-pangenome-science-wellness/index.html>

ном состоянии и находятся в единой локации создаваемого графа пангенома.

МС-модель имеет цель сформировать пангеном и учесть при этом наряду с двумя исходными источниками также все мелкие структурные варианты последовательностей вплоть до уровня SNP. Но, с другой стороны, из структуры МС-графа было решено удалить длинные последовательности ДНК-сателлита, который не связан ни с какой конкретной хромосомой и поэтому представляет собой фактически нечто вроде информационного шума. Удаление последовательностей размером ≥ 100 kb, не являющихся референтными, упрощает структуру МС-графа без потери его качества, поскольку два исходные и вновь позднее добавленные структурно и функционально значимые элементы геномного разнообразия гаплотипов сохранены как референтные. С осторожностью контролировались и были также удалены и не вносились в МС-граф сомнительные и ошибочные последовательности, обнаруженные в ансамблях в точках их соприкосновения, а также в гаплотипах, когда проводилась декомпозиция отдельных хромосом. Исправленные версии графов на базе секвенсов GRCh38 и CHM13 были сделаны, когда выяснилось, что при соприкосновениях коротких последовательностей нуклеотидов могут возникать длинные делеции, что неприемлемо для референтного пангенома.

Модель построения графа PGGB отличается от описанных выше Minigraph и МС прежде всего тем, что она предусматривает использование всех последовательностей нуклеотидов, содержащихся в исходных секвенсах GRCh38 и CHM13, на основе симметричного сравнения всех изучаемых геномов – каждого с каждым. Это значит, что все без исключения ансамбли нуклеотидов играют роль

референтных. При наличии огромного количества сомнительных парных последовательностей, горячих точек структурных вариаций, включая центромеры и единичные повторы, для которых нужно строить субграфы, возникает тенденция снижения уровня изменчивости в последовательностях повторов. Кроме того, без фильтрации быстро эволюционирующих последовательностей сателлитной ДНК или сомнительных участков быстро растет объем коротких и единичных вариантов изменчивости, что увеличивает степень и сложность геномного разнообразия. В этом также есть и положительный аспект, поскольку в пангеноме более полно отражается фактическая структурная и функциональная геномная вариация среди людей. Кроме того, упрощается выявление и анализ с единой позиции всех структурных и функциональных аспектов изменчивости с учетом аннотаций новых ансамблей и их интеграции в общую картину пангеномной вариативности. Представлено аллельное разнообразие генов и его сложная и временами причудливая архитектура. Достоинством PGGB-модели пангенома и сконструированного на его основе графа являются возможности анализа множества скопированных и сохраненных данных о структурных вариациях (SVs), представляющих практический медицинский интерес для диагностики и выбора метода лечения. Сам процесс конструирования графа включал три фазы: а) выравнивание исходных последовательностей нуклеотидов; б) трансформация выровненных последовательностей в граф без потерь; в) нормализация графа с помощью алгоритма для упрощения сложного комплекса возникших мотивов среди повторов.

Отметим, что Консорциум по конструированию пангенома человека завершил начатую работу

и представил это справочное издание в журнал Nature для всеобщего доступа. Первый пангеном человека основан на диплоидных последовательностях нуклеотидов из геномов 47 предварительно отобранных индивидов обоих полов на различных континентах и прошел необходимую верификацию на уровне 99%. Выравнивания этих последовательностей и их информационный анализ осуществлялись по компьютерным программам и алгоритмам. Исходные последовательности геномов точны как на уровне хромосом, так и оснований, включая единичные. К исходным последовательностям нуклеотидов и гаплотипов авторы добавили выявленные новые аллели в структурно сложных локусах. К эталонным последовательностям GRCh38 добавили 119 млн пар оснований из эухроматических полиморфных участков и от 1115 дупликаций генов. Структурная вариация геномов затронула еще около 90 млн пар оснований. Анализ структуры пангенома позволил уменьшить число скрытых ошибок их обнаружения в коротких последовательностях и значительно увеличить количество структурных вариантов по сравнению с исходными данными о последовательностях нуклеотидов в геномах. Необходимо отметить роль информационных технологий в контроле процесса формирования графов, визуализации полученных результатов и определении их надежности. Таким образом, первый пангеном человека начинает выполнять свою роль координатора усилий по подготовке новых проектов и надежного справочника по структуре нашего генома. Можно утверждать, что появился научный фундамент для дальнейшего развития геномной медицины.

Сервис «Геномная медицина»

Попробуем осмыслить, каким образом мы можем использовать научные достижения геномики для лечения наследственных заболеваний, консультирования молодых людей, создающих семьи, или пожилых, боящихся развития старческих немощей. Ответ будет простой: в сотрудничестве клинической медицины и геномики с ее технологиями при поддержке населения можно создать новый сервис – «Геномная медицина».

Первоочередная задача – формирование профессиональной структуры, которая займется планированием и организацией новой, доступной для населения службы, в функции которой будет входить анализ геномов, консультирование по его результатам, определение диагноза и выбор стра-

тегии лечения, если оно потребуется. Общее руководство этой структурой логично будет возложить на национальный орган здравоохранения. Законодателям в соответствии с Конституцией предстоит определить порядок формирования новой службы, ее права, обязанности и ответственность за результаты работы. Финансирование могут осуществлять государство, бизнес и сами пациенты путем взаимной кооперации. Что же нам нужно будет строить и создавать, как стимулировать использование геномных технологий в рутинной клинической практике? Как будем налаживать социальный контакт и отношения между пациентами, членами их семей и персоналом клиник для гарантирования интересов пациентов? Ответы на эти вопросы нам предстоит искать в ближайшем будущем.

В таком компактном государстве, как Республика Беларусь, лидером нового сервиса должна стать столица как административный и научный центр при тесной кооперации с региональными центрами. Их в стране, таким образом, суммарно семь, и в практическом отношении это оптимальное количество для успешной реализации задач, стоящих перед национальной службой геномной медицины. Ключевыми ее рабочими структурами станут Центр секвенирования генома, Центр анализа генома и диагностики, Центр выбора и реализации стратегий лечения, Центр обучения и повышения квалификации. Залог успешной деятельности нового сервиса – желание медицинских работников, ученых и пациентов сотрудничать и оказывать взаимную помощь в выполнении задач службы по организации и предоставлении качественной медицинской помощи. Это важно как для наших граждан, так и для иностранцев, которые смогут воспользоваться им, как только сервис «Геномная медицина» начнет успешно функционировать. Особенно полезной он будет для детей и лиц, страдающих онкологическими заболеваниями. ■

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Савченко В.К. [и др.]. Человек. Биотехнологии. Биоэтика / под общ. ред. члена-корр. В.К. Савченко. – СПб., 2021.
2. Савченко В.К. Геном человека: эволюция, технологии, этика. – Минск, 2022.
3. Куэйл М.А. [и др.]. Рассказ о трех платформах секвенирования следующего поколения: сравнение секвенсоров Ion Torrent, Pacific Biosciences и Illumina MiSeq // BMC Genomics. 2012. Vol. 13. P. 341.
4. Wang T., Antonacci-Fulton L., Howe K. [et al.]. The Human Pangenome Project: a global resource to map genomic diversity // Nature. 2022. Vol. 604. P. 437–446.
5. Liao W.W., Asri M., Ebler J. [et al.]. A draft human pangenome reference // Nature. 2023. Vol. 617. P. 312–324.