

УДК 577.2.01:577.2.08

ГЕНОМИКА

© 2014 г. Г.В. Васильев

Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики
Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: genn@bionet.nsc.ru

Поступила в редакцию 11 марта 2013 г. Принята к публикации 1 февраля 2014 г.

ВВЕДЕНИЕ

Формирование геномики как общепринятого научного направления приходится на первые годы XXI в. — время завершения проекта «Геном человека». Как и собственно классическая генетика, геномика занимается изучением структуры, функций и механизмов работы генов. Однако, если генетика интересуется отдельными генами (или наборами таких генов), вовлеченными в регуляцию конкретного биологического процесса, то геномика направлена на анализ структуры генома как целого и занимается изучением всей совокупности происходящих процессов и обеспечивающих их механизмов. Поэтому само возникновение геномики как отрасли научного знания было вызвано возможностью получения огромных объемов данных о нуклеотидных последовательностях и необходимостью быстрого и качественного анализа получаемой информации.

ИСТОРИЯ РАЗВИТИЯ ГЕНОМИКИ

Строго говоря, в качестве общепринятого дня или даже года рождения новой науки — геномики невозможно выбрать какую-либо фиксированную дату. Первый полный геном был секвенирован в далеком 1977 г. (Sanger *et al.*, 1977), еще до разработки основных современных методов секвенирования ДНК, первой модификацией метода Сэнгера. Однако он составлял всего 5 375 нуклеотидов, и анализ такого объема данных не представлял сколько-нибудь заметных проблем для выполнявших данную работу

экспериментаторов. Необходимость получения данных о последовательностях генов привела к быстрому совершенствованию методов секвенирования ДНК, а это, в свою очередь, дало возможность секвенировать не только отдельные гены, но и протяженные районы генома. Первым полностью секвенированным геномом свободноживущего организма стал геном патогенной бактерии — гемофильной палочки, оказавшийся относительно небольшим даже для бактерий — 1 830 138 п.н. и содержащий 1789 генов (Fleischmann *et al.*, 1995). В последующие несколько лет количество полностью секвенированных бактериальных геномов начало стремительно расти. Однако по-настоящему ключевым событием в формировании геномики явился проект секвенирования генома человека. На выполнение этой задачи было потрачено более полутора десятилетий и более 3 млрд долларов. Международный проект стартовал в 1990 г., через некоторое время после этого корпорацией Celera был анонсирован аналогичный част-

Геном — вся совокупность наследственного материала, содержащегося в клетке и несущего биологическую информацию, необходимую для развития и функционирования всего организма.

Геномика — направление в молекулярной биологии, занимающееся исследованием структуры и функций всей совокупности генов организма или значительной их части. Для геномики характерно использование очень больших объемов данных.

ный проект, обошедшийся гораздо дешевле – в 300 млн долларов, но частично использовавший результаты международного. Первые приблизительные результаты обоих проектов были опубликованы практически одновременно в 2001 г. (International Human Genome ..., 2001; Venter *et al.*, 2001). Официально завершённым международный проект считается после опубликования первичной структуры последней из секвенированных хромосом – хромосомы I (Gregory *et al.*, 2006), и результаты его выполнения многократно окупили все затраты. Во-первых, были созданы и отработаны технологии экспериментального секвенирования полных эукариотических геномов, состоящих из миллиардов пар нуклеотидов, что позволило радикально ускорить последующие геномные проекты. Во-вторых, были разработаны совершенно новые биоинформатические алгоритмы и программы, позволяющие оперировать огромными объемами данных и проводить сборку протяженных последовательностей из относительно коротких секвенированных фрагментов. С дальнейшим развитием и удешевлением технологий секвенирования появилась возможность начать большое число новых геномных проектов, и в результате это привело к взрывному росту объема получаемых первичных данных. Хорошей иллюстрацией этого положения является экспоненциальный рост количества депонированных в базе данных GenBank полных прокариотических геномов, число которых уже превысило 40 000 (рис. 1), количество же частично секвенированных пре-

вышает его еще в несколько раз. Именно этот огромный поток экспериментальных данных и является объектом изучения науки геномики.

ГЕНОМНЫЕ ПРОЕКТЫ – КЛЮЧЕВАЯ ЧАСТЬ ГЕНОМИКИ

Под общим термином «геномный проект» объединяются работы, существенно различающиеся как по используемым экспериментальным методам, так и по методам анализа полученных данных. Поэтому стоит четко разделять разные их типы.

Секвенирование *de novo*. В научной среде под «секвенированием генома» чаще всего подразумевается именно этот тип геномного проекта – определение всего набора нуклеотидных последовательностей ранее неизвестного генома. Это наиболее сложная задача, требующая большой по объему и наиболее высококвалифицированной экспериментальной работы. Простыми и относительно недорогими являются проекты секвенирования прокариотических геномов, которые, как правило, невелики по размеру (несколько млн п.н.), гаплоидны и содержат очень малое число повторяющихся последовательностей. Противоположную ситуацию представляют большие по размеру, зачастую полиплоидные геномы высших растений, содержащие большое количество слабо дивергировавших повторов (Choulet *et al.*, 2010). В состав полноценного эукариотического геномного проекта обычно входят не только собственно секвенирование ДНК, но и

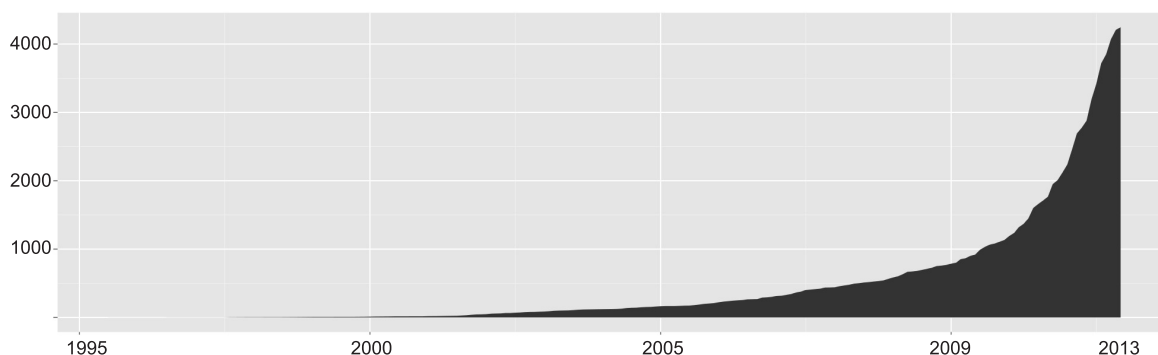


Рис. 1. Экспоненциальный рост количества секвенированных полностью либо в основном бактериальных и архейных геномов, депонированных в базе данных GenBank (по: (http://commons.wikimedia.org/wiki/File%3ABacterial_and_archeal_genome_sequences_submitted_to_Genbank.svg)).

Секвенирование ДНК – определение ее нуклеотидной последовательности.

Экзом – часть генома, представляющая экзоны, т. е. последовательности, которые транскрибируются на матричную РНК после того, как интроны удаляются в процессе сплайсинга РНК. С некоторыми поправками почти соответствует совокупности районов генов, непосредственно кодирующих белки.

привязка полученных последовательностей к конкретным хромосомам и их районам, а также подробное аннотирование обнаруженных генов. Все это даже в настоящее время требует существенных затрат как финансов, так и высококвалифицированного труда, поэтому чаще выполняется целыми консорциумами научных организаций. Упрощенные и значительно удешевленные варианты проектов секвенирования геномов *de novo* дают результаты, существенно уступающие как по полноте секвенирования, так и по надежности конечных данных и качеству аннотирования.

Ресеквенирование генома. Именно этот тип геномных проектов многие журналисты путают с предыдущим, хотя их кардинальное отличие состоит в том, что изначально исходные геномные последовательности заранее известны, и задачей является установить различия между исследуемым геномом и так называемой референсной последовательностью. Характерным примером таких проектов являются секвенирование генома русского мужчины (Чеканов и др., 2010) или проект «1000 геномов человека» (1000 Genomes Project Consortium, 2010). Важная особенность ресеквенирования – если в новом геноме есть протяженные участки ДНК, отсутствующие в референсной последовательности, то они с высокой вероятностью будут потеряны при анализе данных, даже если будут реально отсекутены.

Ресеквенирование экзона. Данный тип проектов в настоящее время наиболее часто выполняется в биомедицинских исследованиях генома человека с использованием высокопроизводительных секвенаторов второго поколения. Поскольку большая часть функционально важных районов полиморфизма находится внутри кодирующих районов гена – экзонов,

то наиболее ценная информация может быть получена в результате секвенирования совокупности всех экзонов. При этом 180 000 экзонов составляют всего около 30 млн п.н., т. е. порядка 1 % генома человека (Wang *et al.*, 2013). Соответственно, если с помощью гибридизационных методик в библиотеках для высокопроизводительного секвенирования выделить все фрагменты, содержащие последовательности экзонов, то расходы на секвенирование можно будет сократить на порядок. Как правило, при этом более 95 % кодирующих последовательностей секвенируются с глубиной покрытия, достаточной для достоверной идентификации полиморфизмов (Ng *et al.*, 2009).

Изучение транскриптома. Под транскриптомом понимается совокупность всех РНК, синтезируемых в конкретной ткани или популяции клеток, в том числе некодирующие – микроРНК, т-РНК, рибосомные РНК. Секвенирование транскриптома различных тканей на разных стадиях развития организма является обязательной частью любого полноценного проекта секвенирования генома, реже проводится в качестве самостоятельного проекта, особенно в случаях, когда уже известен геном близкородственных видов. Упрощенная версия данного метода, известная под названием RNA-Seq, наиболее часто используется в исследованиях для полномасштабного скрининга изменений экспрессии всего пула генов под воздействием различных факторов.

Секвенирование метагенома. Метагеномом называется совокупность генетического материала (ДНК), выделенного из определенного образца, содержащего разнообразные микроорганизмы, в первую очередь бактерии. Основная цель метагеномного проекта – определение видового разнообразия и относительной представленности различных микроорганизмов. Поскольку основная масса свободноживущих микробов не может быть выделена культивированием и охарактеризо-

Метагеном – генетический материал, получаемый напрямую из образцов среды без предварительного разделения отдельных организмов или видов.

вана, секвенирование метагеномов является основным методом определения видового состава во многих природных местообитаниях. Для точной идентификации видов микробов необходимы достаточно длинные фрагменты ДНК, присутствующие у всех организмов и в то же время несущие достаточно отличий. В первое время для этой цели использовались последовательности 16S рибосомальной РНК (Pace *et al.*, 1991), с накоплением больших объемов данных о полных микробных геномах стало возможным использовать другие важные гены (Segata *et al.*, 2012).

ЭКСПЕРИМЕНТАЛЬНАЯ БАЗА ГЕНОМИКИ

Осуществление проекта «Геном человека» позволило создать как экспериментальную базу геномики, так и готовый набор инструментов для теоретического анализа получаемых результатов. Это привело к взрывному росту числа начинаемых геномных проектов, что, в свою очередь, привело к быстрому совершен-

ствованию технологий секвенирования ДНК. Рассмотрим изменение расчетной стоимости ресеквенирования человеческого генома по данным National Human Genome Research Institute (рис. 2, график стоимости построен в логарифмическом масштабе). За период с 2001 г. по 2011 г. эта цена упала со 100 млн долларов до 10 тыс., т. е. в среднем снижалась в 1000 раз за год. Правда, стоит учитывать, что данная стоимость рассчитана для крупных геномных центров, фактически представляющих из себя хорошо организованное индустриальное производство. Для сравнения на этом же графике приведена кривая изменения производительности компьютеров, согласно известному закону Мура. Компьютерная отрасль традиционно считается образцом стремительно развивающейся технологии, но по сравнению с технологиями секвенирования ДНК она выглядит как черепаха по сравнению с Ахиллесом. Чем же вызвано столь стремительное падение цены?

На момент завершения проекта «Геном человека» геномное секвенирование осуществлялось почти исключительно по методу Сэнгера

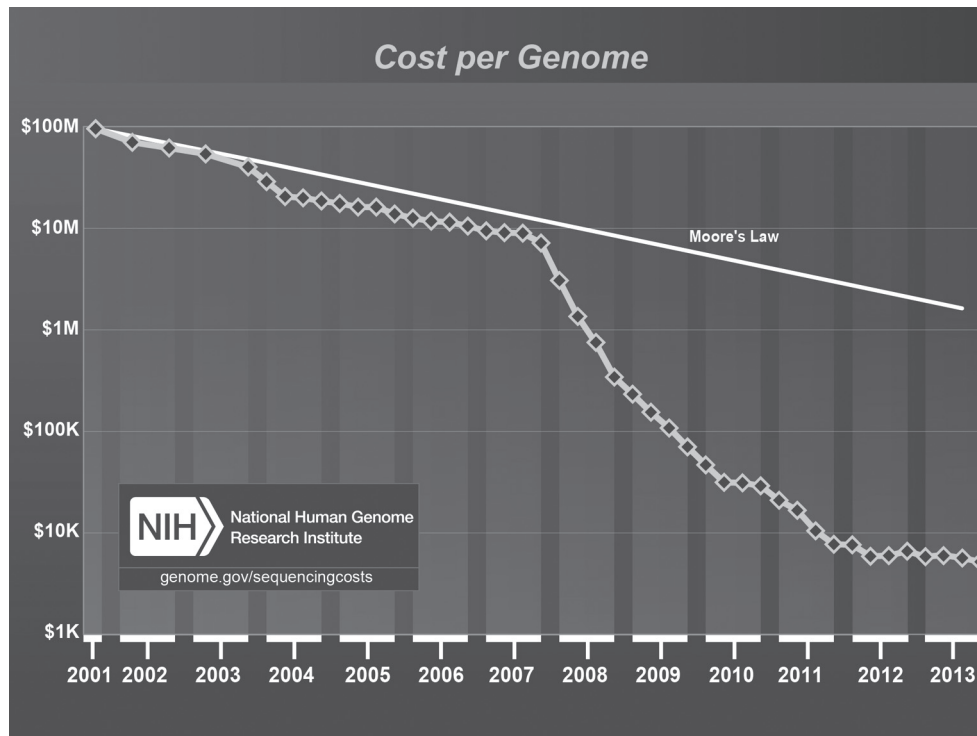


Рис. 2. Радикальное снижение расчетной цены ресеквенирования человеческого генома показано ломаной линией в логарифмическом масштабе. Для сравнения белой линией приведена кривая роста производительности компьютеров, согласно закону Мура (genome.gov/sequencingcosts).

Секвенирование по Сэнгеру – метод секвенирования ДНК, основанный на случайном обрыве синтезируемой ферментом второй цепи ДНК при включении меченого дидезоксинуклеотида. На сегодняшний день высоко автоматизировано и проводится с использованием капиллярных секвенаторов.

с использованием капиллярных секвенаторов – т.е. технологии первого поколения. Секвенирование по Сэнгеру дает последовательности очень высокого качества и достаточно длинные (500–800 п.н.) для эффективной сборки в протяженные контиги, однако требует очень высоких затрат времени, денег и высококвалифицированного труда. Стоимость использования данной технологии снижалась, но довольно медленно. Первый скачок удешевления произошел в 2004 г. с появлением первого секвенатора второго поколения (NGS) – пиросеквенатора 454 фирмы «Roche» – представителя новой технологии массового параллельного секвенирования. Именно с использованием подобных приборов было осуществлено секвенирование первого индивидуального генома Дж. Уотсона (Wheeler *et al.*, 2008) всего за 1 млн долларов. Ключевой особенностью технологий второго поколения является получение молекулярных колоний на твердофазном носителе с помощью ПЦР с одной молекулы и дальнейшее параллельное одновременное секвенирование сотен миллионов и миллиардов этих кластеров в ходе синтеза новых цепей. Дальнейшее радикальное падение цены в 2008–2011 гг. было вызвано появлением и совершенствованием двух других технологий – SOLiD, основанной на секвенировании путем последовательного лигирования флуоресцентно меченых олигонуклеотидов; и Illumina, основанной на использовании обратимой терминации синтеза флуоресцентно мечеными нуклеотидами. Появившийся последним прибор «Proton Torrent», основанный на принципе полупроводникового секвенирования, уже не привел к новому существенному падению цены. Современные представители всех этих четырех технологий второго поколения показаны на рис. 3.

На рис. 2 можно также заметить, что в последние 2 года падение стоимости секвенирования

приостановилось, хотя еще недавно постоянно были слышны прогнозы о неизбежном достижении в ближайшие годы порога стоимости ресеквенирования индивидуального генома в 1 000 долларов. Тому, что эти прогнозы вряд ли сбудутся, есть несколько причин. Во-первых, это экономическая целесообразность. Уже в настоящий момент все большую долю цены занимает стоимость пробоподготовки, которую трудно снижать дальше. Например, полный экзом человека составляет только около 1 % генома (Wang *et al.*, 2013), а стоимость его секвенирования в китайском Институте геномных исследований (BGI) равна 25 % стоимости секвенирования полного генома. Во-вторых, снижение стоимости неизбежно ведет к снижению реального качества получаемых данных, что многократно увеличивает затраты труда и времени на адекватный анализ полученных последовательностей. Не случайно крайне дорогой и весьма трудоемкий метод Сэнгера до сих пор активно применяется в тех геномных проектах, где необходимы высокая достоверность и полнота получаемых данных (Lindblad-Toh *et al.*, 2011).

Кроме того, основные надежды на следующий этап падения цен связываются с разработкой третьего поколения секвенаторов – «одномолекулярных» секвенаторов. Их клю-

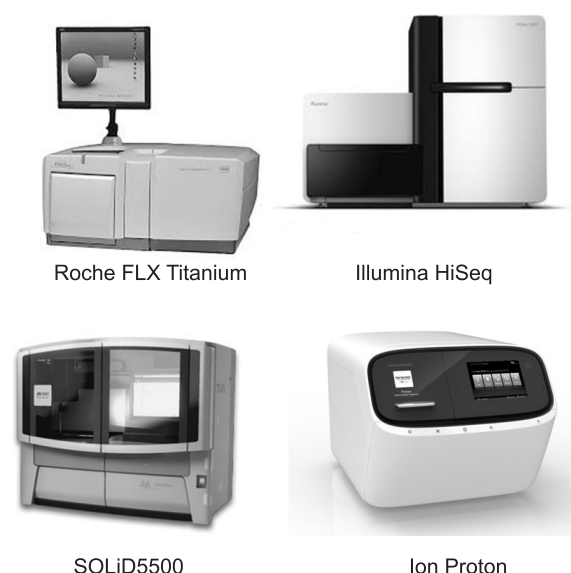


Рис. 3. Приборы-представители четырех основных технологий высокопроизводительного секвенирования ДНК второго поколения (NGS-системы).

чевой особенностью является отсутствие ПЦР в процессе пробоподготовки, при этом секвенируется каждая единичная молекула ДНК, а не кластеры идентичных копий. Это является важным достоинством, поскольку именно наличие ПЦР во время пробоподготовки вызывает большую часть проблем, характерных для секвенаторов второго поколения. За последние годы было анонсировано несколько десятков фирм, готовых предложить работающую модель одномолекулярного секвенатора в самое ближайшее время. Однако эти обещания так и не претворились в реальность. Единственным представителем реально работающих в настоящее время одномолекулярных секвенаторов является прибор «Pacific Biosciences» (PacBio). Его особенностью является возможность непрерывного чтения молекул ДНК длиной до нескольких тысяч нуклеотидов при крайне невысокой его точности – более 10 % ошибок (Quail *et al.*, 2012), что делает этот прибор весьма полезным для узкого круга задач, но малоприменимым для большинства приложений. Для использования в геномных проектах любая технология должна обладать тремя условиями: высокая производительность, низкая цена и высокая точность прочтения. К сожалению, ни один одномолекулярный секвенатор до сих пор не обладает одновременно всеми тремя достоинствами и поэтому безнадежно проигрывает приборам второго поколения.

СОВРЕМЕННОЕ СОСТОЯНИЕ И ПЕРСПЕКТИВЫ РАЗВИТИЯ

Современная геномика в первую очередь ориентирована на анализ данных больших ге-

номных проектов. Более простым их типом является секвенирование бактериальных геномов, поскольку бактериальный геном относительно невелик и, как правило, содержит малое число повторяющихся последовательностей (Gregory, 2005). На начало 2014 г. в базе данных GOLD зарегистрировано более 32 000 проектов по секвенированию индивидуальных геномов бактерий. Из них полноценно законченными являются только 2 711, основная масса – более 60 % – находится на различных этапах выполнения, а 9 591 – т. е. почти треть, брошена на промежуточной стадии (permanent draft) (рис. 4). Еще более наглядным является положение с секвенированием геномов эукариот. Из общего количества в 4 834 анонсированных проектов доведены до полноценного завершения только 153, еще 162 остановлены на стадии черновика (permanent draft), а подавляющая масса – более 90 % – застыла на разных стадиях выполнения. Отсюда ясно, что в современных условиях очень легко начать новый геномный проект, относительно несложно получить значительный объем первичных данных нуклеотидных последовательностей этого генома, но крайне сложно завершить проект с получением полноценной, хорошо собранной и аннотированной геномной последовательности.

Основной причиной такого положения дел является то, что скорость развития методов теоретического анализа получаемых последовательностей ДНК существенно уступает скорости совершенствования технологий секвенирования. На настоящий момент проведение качественного анализа полученных данных занимает в несколько раз больше времени, чем собственно экспериментальная работа. Пол-

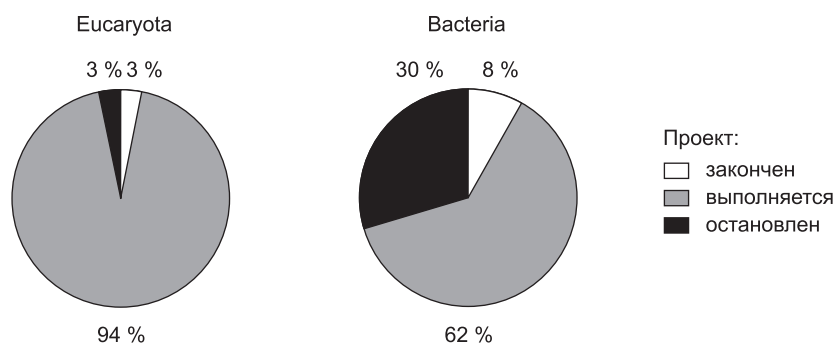


Рис. 4. Статус выполнения геномных проектов (слева – эукариотические, справа – прокариотные) на начало 2014 г., согласно базе данных GOLD.

ноценная же сборка генома, его аннотация и тем более контроль их правильности требуют еще больших трудозатрат. Результатом такого положения дел стала заметная «инфляция» самого понятия «геном». Даже выполненный с максимально возможным в реальности качеством проект «Геном человека» дал в результате секвенирования только 92,3 % от всей геномной последовательности (Schmutz *et al.*, 2004). Неизвестными остались не только центромерные и теломерные районы хромосом, протяженные зоны высокоповторяющихся последовательностей, но и большие кластеры мультигенных семейств, например генов иммуноглобулинов. Проект «1000 геномов» ограничился ресеквенированием в среднем 85 % референсной последовательности (1000 Genomes Project Consortium, 2010), что составляет менее 80 % реального генома. А в настоящий момент «секвенированным геномом» могут назвать набор достаточно коротких контигов (Veenstra, 2010; Cuomo *et al.*, 2012) после проведения первичного анализа без установления их хромосомной локализации и взаимного расположения. Несомненно, даже такие неполные данные чрезвычайно полезны при проведении сравнительных эволюционных и молекулярно-биологических исследований, позволяют получать большое количество ценнейших научных данных, но при условии адекватного отношения к их полноте и достоверности. К сожалению, стоит ожидать, что доля высококачественно выполненных геномных проектов по отношению к их общему числу в дальнейшем будет только снижаться, что потребует пропорционального улучшения качества их теоретического анализа.

Какое место в современной геномике занимают российские научные институты? Наблюдается довольно интересная картина. Первый российский проект полного секвенирования бактериального генома выполнен в Институте физико-химической биологии Министерства здравоохранения России (Lazarev *et al.*, 2011); первый в России проект ресеквенирования генома человека сделан на базе Курчатовского института (Чеканов и др., 2010); секвенирование метагенома уникальных бактерий подледного антарктического озера Восток осуществляется в Петербургском институте ядерной физики. Приборной основой современной геномики

являются высокопроизводительные секвенаторы второго поколения. В России число реально работающих таких приборов (всех 4 платформ) составляет менее двух десятков. Это примерно аналогично их количеству в такой африканской стране, как ЮАР, и принципиально уступает их количеству не только в США (более тысячи) или в Европе, но и в Китае, где один филиал Пекинского института геномных исследований (BGI) в г. Шэньчжэнь имеет их более двух сотен. Приведенные примеры позволяют заключить, что биологические институты РАН по каким-то причинам оказались вдали от передовой линии стремительного развития современной геномики.

Изменения, внесенные развитием современных технологий секвенирования ДНК в современную молекулярную и эволюционную биологию, во многом по своей революционности сравнимы с изменениями, произошедшими после открытия того, что именно ДНК является носителем генетической информации. Как правило, прочтение последовательности геномной ДНК называют «расшифровкой генома». Однако это утверждение верно лишь для прокариотических геномов, которые содержат небольшую долю некодирующих последовательностей, а регуляторные системы, управляющие экспрессией генов, сравнительно просты. Лучшим доказательством истинности этого является успешное создание группой ученых под руководством Крейга Вентера полностью синтетических геномов микроорганизмов, никогда не существовавших в природе (Gibson *et al.*, 2010). После введения искусственного генома в безгеномную клетку эти бактерии смогли жить и размножаться, и это доказывает, что все биохимические и регуляторные процессы нового синтетического организма были идентифицированы правильно. Совершенно иная ситуация с геномом эукариот. Его можно определить как «переписанный», но не расшифрованный, подобно древней рукописи на неизвестном языке. Кодированные белки и РНК-районы, функции которых сравнительно нетрудно установить, составляют лишь несколько процентов от генома, и ранее считалось, что до 95 % нашего генома составляет «мусорная ДНК», «junk DNA» (Ohno, 1972). Однако исследования последнего времени в рамках проекта «ENCODE» пока-

зали, что до 80 % некодирующей ДНК может иметь регуляторные функции (Ecker *et al.*, 2012; Pennisi, 2012). Это означает, что большинство регуляторных элементов, обеспечивающих дифференцировку клеток, поддержание этого статуса или взаимодействие между различными клетками организма, равно как и изменения в работе генов в ответ на внешние воздействия, работают только в небольших временных интервалах. Они могут оказывать существенное влияние только в определенные короткие интервалы времени или в ограниченной популяции клеток и большую часть времени в большинстве клеток не проявляться вообще. Поэтому прямое экспериментальное определение роли эпизодически включающихся регуляторных элементов сопряжено с большими трудностями, дорого и весьма трудоемко. Например, в работе Аттаназио с соавт. (Attanasio *et al.*, 2013) было показано, что тонкая морфология морды мышей регулируется в ходе эмбрионального развития многими энхансерами, в том числе расположенными на расстоянии более 10 т.п.н, а не кодирующими частями генов. Кроме того, время активной работы этих энхансеров в конкретных группах клеток относительно непродолжительно. Реальная расшифровка эукариотического генома должна включать в том числе и описание тонкостей работы регуляторной части генома, и эта гигантская по объему работа представляет собой наиболее важное поле деятельности науки геномики.

ЛИТЕРАТУРА

- Чеканов Н.Н., Булыгина Е.С., Белецкий А.В. Характеристика генома русского мужчины на основе анализа однонуклеотидных полиморфизмов и реконструкции последовательностей ДНК // *Acta Naturae*. 2010. Т. 2. С. 140–145.
- 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing // *Nature*. 2010. V. 467. P. 1061–1073.
- Attanasio C., Nord A.S., Zhu Y. *et al.* Fine tuning of craniofacial morphology by distant-acting enhancers // *Science*. 2013. V. 342. P. 1241006.
- Choulet F., Wicker T., Rustenholz C. *et al.* Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces // *Plant Cell*. 2010. V. 22. P. 1686–1701.
- Cuomo C.A., Desjardins C.A., Bakowski M.A. *et al.* Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth // *Genome Res*. 2012. V. 22. P. 2478–2488.
- Ecker J.R., Bickmore V.A., Barroso I. *et al.* Genomics: ENCODE explained // *Nature*. 2012. V. 489. P. 52–55.
- Fleischmann R.D., Adams M.D., White O. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. // *Science*. 1995. V. 269. P. 496–512.
- Gibson D., Glass J., Lartigue C. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome // *Science*. 2010. V. 329. P. 52–56.
- Gregory S.G., Barlow K.F., McLay K.E. *et al.* The DNA sequence and biological annotation of human chromosome 1 // *Nature*. 2006. V. 441. P. 315–321.
- Gregory T.R. Synergy between sequence and size in large-scale genomics // *Nat. Rev. Genet.* 2005. V. 6. P. 699–708.
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome // *Nature*. 2001. V. 409. P. 860–921.
- Lazarev V.N., Levitskii S.A., Basovskii Y.I. *et al.* Complete genome and proteome of *Acholeplasma laidlawii* // *J. Bacteriol.* 2011. V. 193. P. 4943–4953.
- Lindblad-Toh K., Garber M., Zuk O. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals // *Nature*. 2011. V. 478. P. 476–482.
- Ng S., Turner E.H., Robertson P.D. *et al.* Targeted capture and massively parallel sequencing of twelve human exomes // *Nature*. 2009. V. 461. P. 272–276.
- Ohno S. So much «junk» DNA in our genome // *Brookhaven Symp. Biol.* 1972. V. 23. P. 366–370.
- Pace N.R., DeLong E.F., Pace N.R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing // *J. Bacteriol.* 1991. V. 173. P. 4371–4378.
- Pennisi E. Genomics. ENCODE project writes eulogy for junk DNA // *Science*. 2012. V. 337. P. 1159–1161.
- Quail M., Smith M.E., Coupland P. *et al.* A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers // *BMC Genomics*. 2012. V. 13. P. 341.
- Sanger F., Air G.M., Barrell B.G. *et al.* Nucleotide sequence of bacteriophage phi X174 DNA // *Nature*. 1977. V. 265. P. 687–695.
- Segata N., Waldron L., Ballarini A. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes // *Nature Meth.* 2012. V. 9. P. 811–814.
- Schmutz J., Wheeler J., Grimwood J. *et al.* Quality assessment of the human genome sequence // *Nature*. 2004. V. 429. P. 365–368.
- Veenstra J.A. Neurohormones and neuropeptides encoded by the genome of *Lottia gigantea*, with reference to other mollusks and insects // *Gen. Comp. Endocrinol.* 2010. V. 167. P. 86–103.
- Venter J.C., Adams M.D., Myers E.W. *et al.* The sequence of the human genome // *Science*. 2001. V. 291. P. 1304–1351.
- Wang Z., Liu X., Yang B.Z., Gelernter J. The role and challenges of exome sequencing in studies of human diseases // *Front. Genet.* 2013. V. 4. P. 160.
- Wheeler D.A., Srinivasan M., Egholm M. *et al.* The complete genome of an individual by massively parallel DNA sequencing // *Nature*. 2008. V. 452. P. 872–876.