

ИССЛЕДОВАНИЯ ГЕНОМОВ К КОНЦУ 1999 ГОДА

В. Н. СОЙФЕР

Университет имени Джорджа Мэйсона, Фэйрфакс, США

GENOMIC STUDIES BY THE END OF THE 1999

V. N. SOYFER

The tempos of the studies of genomes of not only humans but of many other species are accelerated every year. The development of crucially new physical, chemical and mathematical methods, implementation of robots for the analysis of DNA nucleotide sequences as well as the development of robust computer programs distinguishes this discipline. The genomic studies became an arena of not only scientific and economic competition but also a political race at the State level.

Темпы исследования геномов не только человека, но и многих других видов живых существ ускоряются с каждым годом. Разработка принципиально новых физических, химических и математических методов, внедрение роботов в анализ последовательностей нуклеотидов ДНК, развитие самых мощных из имеющихся в мире компьютерных программ отличают эту сферу науки. Исследования генома стали ареной не только научного, но также экономического и даже политического соревнования на государственном уровне.

ЧТО ОЗНАЧАЕТ “ЗАВЕРШИТЬ СЕКВЕНИРОВАНИЕ ГЕНОМА”?

В результате исследований геномов (см. [1]) сформулированы специфические задачи, созданы методы, компьютерные программы, роботы, особый и изощренный математический аппарат. Тем самым заложены основы новой науки, названной геномикой. Только что вышел в свет первый учебник для вузов, написанный Чарлзом Кэнтором и Кассандрой Смит, так и названный “Геномика”.

При расшифровке последовательностей нуклеотидов геномов просто устроенных бактерий и вирусов гenetикам удалось с точностью до одного нуклеотида определить их последовательность в ДНК. Затем настал черед многоклеточных организмов, суммарная длина ДНК в хромосомах которых была в десятки, сотни и даже сотни тысяч раз больше. В начале декабря 1998 года было объявлено об окончании секвенирования генома круглого червя *Caenorhabditis elegans*, первого многоклеточного животного. Однако сказать однозначно, что при этом удалось определить положение *каждого* нуклеотида в ДНК этой нематоды, нельзя. Да, было доказано, что геном *C. elegans* содержит 97 млн пар оснований и несет 19 099 генов (и ни одного больше!), но тем не менее 100 или чуть больше небольших по размеру отрезков (около сотни нуклеотидов каждый) остались нерасшифрованными. К ноябрю 1999 года это число неопределенностей уменьшилось — осталось около 70 неясных точек, но они пока ускользают от исследователей. Это связано частично с тем, что в данных точках есть зоны повторения нуклеотидов. Во время наработки копий этих участков с помощью полимеразной цепной реакции (ПЦР) зоны повторяющихся нуклеотидов могут вести себя необычно: образовывать шпильки или изломы, нераспознаваемые или неправильно читаемые ДНК-полимеразами — ферментами, удваивающими (амплифицирующими) данные участки.

Другая причина неудач обусловлена тем, что иногда повторяющиеся участки просто невозможно размножить в бактериальных клетках (получить клоны этих участков), так как они обладают способностью

убивать клетки, в которых их пытаются клонировать. Но в целом остающиеся неопределенности хотя и сильно раздражают исследователей, но столь малы, что не составляют и сотой доли процента от общей длины расшифрованной ДНК и ни в одном случае не включены в участки генов, а всегда сосредоточены в межгенном пространстве. Поэтому общепризнан успех в изучении генома *C. elegans*: все до единого гена открыты, все функциональные участки (промоторы, другие рецепторные и важные в структурном отношении районы) секвенированы до последнего нуклеотида, точки генома, в которых расположены отрезки неопределенностей, известны.

Вопрос о точности изучения последовательностей ДНК стал особенно важным в отношении генома человека. В нашем геноме существует большое число повторов нуклеотидов. Кроме них в хромосомах есть теломеры, центромеры и зоны гетерохроматина, где секвенирование затруднено, так как нуклеопотеиды в них плотно сконденсированы: на сегодняшний день они попросту исключены из исследований. Участников программы это не очень беспокоит, дел и без того невпроворот.

Остается неясным, какой точности анализа надо достичь. Недавно все сходились на том, что ошибок не должно быть больше, чем одна на миллион нуклеотидов. Но добиться такой точности по всей длине генома трудно, и было заявлено, что в пределах генов частота ошибок не должна превышать 10^{-6} , а в межгенных пространствах точность может быть и в сто раз меньше. Сейчас пришли к согласию, что для рабочего варианта генома в пределах генов будет достаточно такой точности, как 10^{-4} .

К концу XX столетия геномы почти 50 видов были полностью секвенированы. Собранная информация разнообразна, порой необычна, но важна для будущего прогресса науки и промышленности. Вот один из примеров. Летом 1997 года была завершена пятилетняя работа 37 лабораторий (главным образом европейских — они расшифровали 60% генома; японских, секвенировавших 30% генома; одной корейской и двух американских лабораторий) над геномом бактерии *Bacillus subtilis* (ее ДНК содержит 4,2 Мб нуклеотидов и около 4 тыс. генов). Это был десятый по счету изученный организм, причем впервые была исследована грамположительная бактерия. К этому классу относятся такие патогены, как *Staphylococcus aureus*, вызывающий гнойные воспаления, стрептококки — источники воспаления среднего уха, пневмонии и менингитов. Помимо лучшего понимания процессов патогенности стали ясны структуры генов для многих ферментов, в том числе и промышленно важных (теперь эти гены можно искусственно

собирать из предшественников), были также секвенированы участки, в которые встроились ДНК бактериофагов, а также стало ясно, как именно эти пришельцы не только наносят вред клеткам, но иногда помогают им, придавая устойчивость к тяжелым металлам и токсинам.

ИНСТИТУТ ГЕНОМНЫХ ИССЛЕДОВАНИЙ И КОМПАНИЯ “СИЛЕРА ДЖИНОМИКС”

Пожалуй, впервые в современной науке сложилась необычная ситуация, когда в работу над исключительно дорогостоящим и важным проектом включились индивидуальные исследователи, нашедшие себе мощных спонсоров и создавшие серьезную конкуренцию учреждениям и университетам, финансируемым правительствами нескольких стран. Первоначально (в 1988 году) средства на изучение генома человека выделило Министерство энергетики США, и одним из руководителей программы “Геном человека” стал профессор Чарлз Кэнтон (написанный им совместно с Полом Шиммелом трехтомный учебник “Биофизическая химия” был переведен в СССР [2]). В 1990 году Нобелевский лауреат Джеймс Уотсон начал лоббирование конгресса США, и вскоре конгресс распорядился выделить сразу сотни миллионов долларов на изучение генома человека. Эти средства были добавлены к бюджету Министерства здравоохранения, отсюда они перетекли в ведение дирекции сети институтов, объединенных под общим названием — Национальные институты здоровья (National Institutes of Health, сокращенно NIH). В составе NIH появился новый институт — Национальный институт исследования генома человека (NHGRI, директор Фрэнсис Коллинз).

В мае 1992 года ведущий сотрудник NIH Крэйг Вентер подал заявление об уходе и объявил о создании нового, частного исследовательского учреждения — Института геномных исследований (The Institute for Genomic Research, сокращенно TIGR или ТИГР). Вентеру, к удивлению многих, удалось быстро и с большим размахом развить свой институт. Первоначальный капитал института составил ни много ни мало 70 млн долларов, собранных от частных пожертвователей. ТИГР был объявлен неприбыльным частным институтом, который не будет использовать свои результаты для обогащения или торговли. Однако одновременно была образована компания “Науки о геноме человека” (Human Genome Sciences, сокращенно HGS), которая должна была продвигать на рынок данные, получаемые сотрудниками ТИГРа. Последние обязывались бесплатно передавать компании HGS свою продукцию, а она могла распоряжаться ими как захочет.

В июне 1997 года Вентер вывел ТИГР из связки с HGS и в 1998 году организовал свою собственную

коммерческую компанию, которую назвал “Силера джиномикс” (Celera Genomics, Inc), стал ее президентом, оставшись главным научным руководителем ТИГРа, а в ТИГРе президентом стала его жена Клэйр Фрэйзер. Вентер оказался исключительно умелым руководителем. Он договорился с одной из крупных компаний по производству научного оборудования, что та предоставит в прокат ТИГРу 18–20 автоматических секвенировщиков-роботов, которые в первый же год работы позволят довести размер секвенируемых последовательностей до 60 млн оснований (одной пятой всего генома человека; такой жест был важен и для компании — лучшей рекламы своей продукции представить трудно). Позже Вентер заключил аналогичный контракт о поставке институту огромных систем усовершенствованных роботов для секвенирования протяженных кусков ДНК. Сейчас в руках Вентера сосредоточен столь мощный парк компьютеров, что его считают вторым по мощности в мире.

Научная сторона деятельности ТИГРа первоначально была встречена многими в штыки. Вентер объявил, что его группа намеревается сосредоточиться на том, чтобы выделять все без разбора синтезируемые клетками информационные РНК, готовить с помощью фермента “обратная транскриптаза” их ДНК-копии (так называемые cДНК) и секвенировать их все подряд, даже не зная, для каких целей клетки используют данные иРНК. Он упростил задачу, предложив секвенировать не полные копии генов, а короткие фрагменты иРНК, названные Expressed Sequence Tags (буквально: экспрессируемые ярлыки (этикетки) последовательностей). Такие случайным образом выбранные фрагменты — “ярлыки” можно будет идентифицировать однозначно (или почти однозначно), сопоставляя расшифрованные последовательности с уже известными. Если же подходящего гена из числа расшифрованных не найдется, то можно будет подождать некоторое время, пока нужный ген не будет идентифицирован в одном из изучаемых организмов.

Помимо интереса к изучению генома человека сотрудники ТИГРа в спешном порядке приступили к исследованиям геномов представителей эукариот: одноклеточных (дрожжей) и многоклеточных (круглого червя *Caenorhabditis elegans*, дрозофилы) и некоторых других.

Крупный вклад был сделан сотрудниками ТИГРа в современное понимание эволюции живых существ на Земле. Как было отмечено раньше [1], только благодаря геномным исследованиям стало ясно, что в ходе эволюции сначала выделились представители архей, затем эубактерий (настоящих бактерий) и позже эукариот (см. рис. 4 в статье [1]). ТИГР стал также начинателем

нескольких новых важных инициатив. Одна из них — создание искусственной ДНК, которая содержала бы минимальный набор генов, обеспечивающих независимое существование клетки. Было подсчитано, что для этого потребуется около 450 генов, и сейчас полным ходом ведется работа по их объединению в единую структуру.

Другой важный результат — совпадение нуклеотидных последовательностей ряда генов у неродственных видов. На этом основании можно предположить, что в процессе эволюции происходил перенос генов от одного вида другому. В мае 1999 года 29 сотрудников ТИГРа опубликовали результаты анализа генома термофильной неспорулирующей красной бактерии *Thermotoga maritima*, лучше всего растущей при 80°C [3]. В ее геноме, содержащем 1 860 725 оснований (для сравнения: у бактерии кишечной палочки геном содержит 4,2 Мб), найдены 1014 генов с известными функциями и 863 гена, функции которых еще предстоит открыть. Термогота первоначально привлекла ученых способностью усваивать разнообразные моно- и полисахариды: глюкозу, сахарозу, крахмал, целлюлозу, ксилан с образованием нескольких продуктов метаболизма, включая молекулярный кислород (что очень важно для расширения круга организмов, которые можно использовать в будущем для получения возобновляемых топливных источников). Показано, что на долю генов у термогоги отведено 95% генома, охарактеризованы гены, участвующие в метаболизме сахаров, переносе ионов и других компонентов в клетку и из клетки. Геномный анализ показал, что эта бактерия не несет полного набора генов для деградации целлюлозы, как термофильная бактерия *Clostridium thermocellum*, но обладает генами для усвоения глицерина, глюконатов и многочисленных сахаров, включая амилозу, мальтозу и галактозу, аминокислот аспарагина, треонина и глицина, а также что бактерия может добывать энергию не только путем ферментации сахаров, но и за счет процессов, требующих для этого наличия Fe(III).

Однако самый важный (и неожиданный!) вывод сделан именно по результатам анализа последовательностей нуклеотидов в генах. Термогота — это представитель зубактерий. Однако это первая бактерия, у которой 24% генов оказались близки по строению к генам архей. На этом основании сделан вывод, что, скорее всего, термогота возникла при захвате значительного числа архейных генов в результате так называемого параллельного переноса (одним из первых эту возможность давно обсуждал советский ученый В.М. Жданов [4]). Сотрудники ТИГРа обнаружили еще один факт, прямо указывающий на прямой перенос. Оказалось, что 81 ген термогоги, напоминающий архейные гены, собран в 15 кластеров (гроздьев), составленных из участков длиной от 4 до 20 000 оснований, причем порядок

генов в кластерах полностью повторяет порядок генов архей (таких, как *Methanococcus jannaschii* и *Archeoglobus fulgidus*), что служит сильным аргументом в пользу гипотезы переноса. Пример исследования генома бактерии термотога лучшим образом характеризует познавательную мощь геномного анализа.

Существенная разница в исследованиях академических групп (университетских лабораторий и институтов, работающих за счет грантов, предоставляемых правительствами) и коммерческих групп пока еще мало заметна, хотя в одном отношении различия уже начали проявляться: ТИГР, например, передает гласности не все свои результаты, да и те, что публикует, может обнародовать только через несколько месяцев после их получения, в то время как участники официально спонсируемого проекта “Геном человека” в соответствии с установленной договоренностью обязаны передавать полученные результаты о секвенировании участков ДНК длиной не менее тысячи оснований в открытый для всех международный банк в течение первых 24 часов после их получения. Академические группы не претендуют на патентование данных, а “придерживание” данных ТИГРом связано с тем, что это время нужно для заполнения заявок на патентование генов последовательностей. Несколько десятков тысяч заявок уже готово. Результаты такой деятельности могут коренным образом изменить будущую ситуацию на финансовых рынках мира: патенты в будущем могут принести невиданные дивиденды, возможно триллионы долларов. Ведь к концу XX века самыми процветающими компаниями в мире стали фармацевтические, медицинские и сельскохозяйственные, для которых данные о генах жизненно необходимы.

ВЛОЖЕНИЯ В ПРОЕКТ “ГЕНОМ ЧЕЛОВЕКА” РАЗЛИЧНЫХ ПРАВИТЕЛЬСТВ

Ожидание гигантских прибылей от будущего внедрения результатов изучения геномов хорошо поняли не только в США. В ведущих странах Запада началась настоящая гонка в отношении вклада средств в исследования геномов. 3 мая 1999 года британский “Велком траст” (формально правительство Великобритании финансирует британскую часть проекта “Геном человека” через этот частный благотворительный фонд) добавил дополнительно 100 млн фунтов стерлингов (примерно 167 млн долларов) нескольким английским лабораториям, занимающимся исследованиями генома человека, из них 77 млн долларов было выделено на 1999 год Сэнгеровскому центру в Кембридже. Этим шагом британский фонд постарался стимулировать своих соотечественников. Вскоре в Колд Спринг Харборской лаборатории под Нью-Йорком (где почетным президентом

работает Джеймс Уотсон) за закрытыми дверями состоялось заседание всех сторон, участвующих в международном проекте “Геном человека”, после чего руководство проекта объявило, что “рабочий вариант” человеческого генома будет готов не к 2003, а в 2000 году¹.

Для того чтобы объяснить публике, как можно столь вольно манипулировать, казалось бы, строго продуманными научными планами, был использован следующий аргумент. Как уже было сказано, можно по-разному подходить к критериям точности секвенирования геномов. При первоначальном объявлении сроков завершения проекта в 2003 году предполагалось, что точность исследования генома составит 99,99%. Потом сроки подвинули, основываясь на том, что для биологов и медиков хватит и 90%-ной точности, зато отрапортовать о завершении генома можно будет к концу 2000 года.

Правда, американскому правительству, чтобы не отстать в гонке, пришлось пойти на серьезные дополнительные траты. Уже 15 марта дирекция NHGRI сообщила, что получила дополнительно 81,5 млн долларов на программу генома человека и что эти деньги будут немедленно распределены между тремя американскими центрами: Уайтхедским центром геномных исследований Массачусетского технологического института, который получит 33 млн долларов, медицинским факультетом Вашингтонского университета, расположенного в Сент-Луисе в штате Миссури, которому выделили 35 млн долларов, и Бэйлоровским медицинским центром в городе Хьюстон (штат Техас), которому досталось 13,5 млн долларов. Согласно американским прогнозам, этим шагом будет обеспечено положение, при котором три центра смогут завершить к весне 2000 года две трети работ по составлению первичной карты генома человека, а британский Сэнгеровский центр успеет закончить около 33% всей работы.

¹ 2 декабря 1999 года журнал “Nature” обнародовал данные, касающиеся крупного прорыва в исследовании генома человека: в основном усилиями английских ученых при активном участии других европейских, японских и американских лабораторий был завершен полный анализ одной из хромосом человека (правда, одной из самых маленьких) — хромосомы 22. Важным преимуществом работы, проводившейся под руководством Иэна Дэнэма (Ian Dunham), было то, что отрезки ДНК клонировали в искусственных бактериальных хромосомах и после секвенирования сравнивали друг с другом, наращивая отсекуемые участки по результатам перекрытия концов последовательно секвенированных отрезков хромосом. Дэнэм считает, что такой подход привлекательнее метода, предложенного Вентером, предпочитающим секвенирование случайных кусков в надежде, что позже мощные компьютеры помогут собрать их в единое целое. Такая “беспорядочная стрельба” по последовательностям — это ахиллесова пята Вентера, считает Дэнэм и повторяют многие другие специалисты.

На этом гонка отнюдь не затихла. Как сообщил журнал “Science” со ссылкой на газету “Ля Монд” от 14 мая 1999 года, французское правительство решило в этот момент “впрыснуть” дополнительно 330 млн долларов на ближайшие три года в бюджет расположенного рядом с Парижем исследовательского центра генома в Иври. Этим шагом французское правительство хотело бы устранить свое отставание по сравнению с США и Великобританией и обеспечить себе возможность запатентовать достаточное количество разработок в этой области. Только тогда можно надеяться, что в будущем, когда результаты изучения генома человека будут внедрены в медицинскую и промышленную практику, приток денег французским компаниям будет большим.

В июне 1999 года Германия, которая до этого выделяла явно недостаточно средств на исследования генома человека (всего 23 млн долларов в год начиная с 1996 года), изменила свой подход: на ближайшие пять лет было отпущено 550 млн долларов. В ноябре—декабре 1999 года стало ясно, что ученым удалось убедить правительство увеличить ежегодные траты на исследования генома человека до 280 млн долларов.

13 июля 1999 года об увеличении выделяемых средств на работу по секвенированию генома человека объявило правительство Японии. Вклад Японии в проект “Геном человека” составлял до этого небольшую величину (японские ученые к тому времени изучили не более 8% генома, средств было выделено недостаточно, хотя в последние три года ежегодные траты достигали 560 млрд иен и составили четверть средств, расходуемых в США). Теперь правительство Японии решило вложить в ближайшие пять лет 2 трлн иен (17 млрд долларов, или около 0,2% валового национального продукта Японии), что позволит японским ученым расшифровать до трети генома человека к 2001 году. Этот огромный по размерам план стал частью общих усилий по широкомасштабному развитию биотехнологии в стране. Для этого японское правительство решило расширить в ближайшие десять лет японский биотехнологический рынок в 25 раз, доведя масштаб ежегодных сделок на нем до 25 трлн иен (213 млрд долларов) и создав условия для возникновения около 1000 биотехнологических частных фирм к 2010 году.

То, что участвовавшая в начале создания международного проекта “Геном человека” Россия фактически приостановила свой вклад в него, можно рассматривать однозначно: Россия обрекает себя в этом отношении на скатывание на уровень второстепенных государств, обреченных на экономическую зависимость в будущем от тех, кто вложил средства в эту научную область.

ФУНКЦИОНАЛЬНАЯ ГЕНОМИКА

Как упоминалось раньше [1], не без изрядной доли удивления генетики установили, что, как правило, лишь небольшая часть генома эукариотов занята генами (например, у человека гены занимают лишь около 3% от всей ДНК), в то время как функциональная роль большей части ДНК остается нераскрытой, возможно, все остальное играет какую-то пока не очень ясную структурную роль. Остается спорным и вопрос о числе генов у человека. Еще недавно ученые располагали данными лишь о 10 000 генов, в то время как на самом деле их может оказаться от 50 000 до 120 000 (согласия на этот счет достичь пока не удалось: некоторые, как Вентер, считают, что у человека не больше чем 50 000—60 000 генов, другие, как Лерой Худ, утверждают, что генов должно быть не меньше 120 000; изучение генома человека позволит разрешить это противоречие).

Как же выявить все гены? Со времени Томаса Ханта Моргана, исследовавшего дрозофилу в 20—30-е годы XX века, за изменениями генов следят по трем главным показателям: видимым мутациям, летальным и стерильности. Под видимыми мутациями подразумевают не только случаи, когда какая-то часть тела меняет свои очертания, но и когда изменения затрагивают биохимические реакции (биохимические мутации) или структуру хромосом (так называемые хромосомные aberrации) и вообще возникают любые отличимые особенности в строении и поведении организмов.

Для выявления генов стали применять два главных метода: в каждый из участков ДНК, несущих обязательные для генов участки промоторов, операторов, старт- и стоп-кодонов с их характерными последовательностями нуклеотидов, искусственно вносили точечные мутации, после чего следили за тем, какие функции оказываются измененными или нокаутированными, либо внедряли короткие чужеродные участки ДНК — транспозоны (см. статью В.А. Гвоздева в СОЖ [5]). Во многих случаях нокаутирование генов помогало убедиться в том, что в данной точке располагается ген, ибо после этого наступал летальный исход, или терялась способность давать потомство (наступала стерильность), или возникали видимые изменения.

ЗАВЕРШЕНИЕ ИЗУЧЕНИЯ ГЕНОМА ДРОЗОФИЛЫ

Крэйг Вентер еще в конце 1998 года заявил, что его ТИГР и “Силера джиномикс” ускорят темпы своей работы и обгонят лаборатории, использующие преимущественно правительственные субсидии, а в октябре 1999 года он объявил о своих планах приблизить срок полного изучения генома человека. По его словам, теперь для этого появилась реальная возможность, так

как в сентябре 1999 года была завершена техническая часть работы по секвенированию кусков ДНК, покрывающих полностью геном дрозофилы (содержащий 1,8 млрд оснований; напомним, что весь геном человека сосредоточен примерно в 3,2–3,5 млрд оснований ДНК), и все освободившиеся машины-роботы быстро переключили для анализа человеческой ДНК. Компьютерные данные об участках генома дрозофилы были переданы Вентером в лабораторию Калифорнийского университета в Беркли (туда же сейчас поступили данные европейских ученых, изучавших геном дрозофилы). В Беркли создана система алгоритмов по выявлению перекрываемости частей генома, которая позволит завершить обработку всех данных в рекордно короткий срок, после чего не позднее начала 2000 года полный геном дрозофилы будет опубликован.

У дрозофилы методы точечного мутагенеза для выявления генов применить не удалось, зато были широко использованы мобильные элементы — транспозоны (так называемые Р-элементы). Благодаря этому удалось дать грубую предварительную оценку числа генов у дрозофилы — от 8600 до почти 17 000. Завершающаяся в Беркли работа даст точную цифру. Некоторые из исследователей высказывают предположение, что число генов у дрозофилы может оказаться примерно на 10% больше найденного до сих пор. Но даже если это так на самом деле, общее число генов дрозофилы будет меньше числа генов, найденных у круглого червя *C. elegans*. Как могло случиться, что значительно большие по размеру и более сложно устроенные дрозофилы обходятся меньшим числом генов, чем просто устроенные микроскопические круглые черви, сказать никто пока не может.

К моменту окончания работы с дрозофилой стало возможным сравнить последовательности сходных по функциям генов в разных видах. При этом выявлена интересная закономерность: гены, кодирующие белки с существенными для жизни функциями, сходны. Складывается впечатление, что гены существенные сохраняют консерватизм наследственной структуры, а гены, определяющие несущественные функции, более изменчивы.

Другой важный результат получен при сравнении числа генов, соответствующих полосам (дискам), видимым в гигантских (политенных) хромосомах дрозофилы. Ученик Моргана Кэлвин Бриджес первым высказал в 1925 году предположение, что каждый диск соответствует одному гену. Такое соответствие было возведено в правило. Теперь стало ясно, что на самом деле такого правила не существует. В одном из участков с 69 дисками было найдено 218 генов, отвечающих за синтез белков, и 11 генов — за синтез транспортных РНК. Раньше молекулярно-генетическими методами к

тому же выводу о несоответствии дисков генам пришел российский ученый, Соросовский профессор И.Ф. Жимулев [6]).

УЛУЧШЕНИЕ ТОЧНОСТИ СЕКВЕНИРОВАНИЯ И ДРУГИХ ПОКАЗАТЕЛЕЙ РАБОТЫ

Три основные американские лаборатории, вовлеченные в анализ генома человека, сравнили в начале 1999 года свои результаты и установили, что на каждые 10 000 изученных оснований ошибочным оказалось не более одного и только в лучших случаях одна ошибка встречалась на 100 000 оснований.

Стоимость затрат на секвенирование упала. Если в 1996 году ученые затрачивали на одно основание до 2 долларов, то к началу 1998 года затраты снизились до 50 центов за нуклеотид, в середине 1999 года — до 41 цента и к сентябрю 1999 года — до 20–30 центов на основание. Уайтхедовский центр Массачусетского института сумел увеличить в 1998 году продуктивность секвенирования в 12 раз по сравнению с предыдущим годом без дополнительных затрат.

Громадной важности шаг был сделан в Вашингтонском университете в Сент-Луисе. Долгое время узким местом оставался компьютерный анализ перекрываемости огромного числа генных последовательностей. Эта чисто математическая задача требовала непомерно большой памяти компьютеров, принципиально новых шагов в математическом оперировании огромными массивами данных либо какого-то нового лабораторного принципа. Разработав процедуру внедрения кусков человеческой ДНК длиной в 400 000–500 000 оснований (400–500 Кб) в так называемые искусственные бактериальные хромосомы и одноразового их секвенирования в составе таких хромосом, ученые преодолели и эту трудность. К июлю 1999 года компьютерные технологии сравнения между собой нуклеотидов в участках ДНК длиной 400–500 Кб были отработаны и этим облегчено выявление зон их взаимного перекрытия. Теперь процедура наращивания размера отсеквенированных частей гигантских молекул индивидуальных хромосом человека была упрощена.

Помимо трех названных выше центров, финансируемых в основном американским правительством через сеть NIH, в США функционирует также еще один крупный центр — Объединенный институт геномных исследований Соединенных Штатов, расположенный в Волнот Крике (Калифорния), получающий средства от Министерства энергетики США. К июню 1999 года этот институт подписал договор с одной из компаний по производству автоматических секвенаторов, позволивший институту взять в аренду

на несколько лет 24 суперкомпьютера, настроенных на задачу декодирования вновь секвенируемой ДНК (так называемых MegaBACE DNA sequencers), что немедленно утроило скорость работы (сотрудники института анализируют 5-, 16- и 19-ю хромосомы человека), и скорость секвенирования была доведена до абсолютно немислимых еще год назад величин: 14 млн оснований в день! Руководители Министерства энергетики заявили корреспонденту журнала “Nature”, что они планируют предоставить институту дополнительные средства, чтобы удвоить или даже утроить скорость работы в самое ближайшее время.

Разработаны и другие совершенно новые методы секвенирования. Один из них базируется на возможности избирательно присоединять тяжелые атомы металлов (нерадиоактивные изотопы) к определенным нуклеотидам с последующим масс-спектрометрическим сканированием молекул ДНК, пропускаемых через тончайший (нанометровый) микрокапилляр. Устройство читает нуклеотидную последовательность практически безошибочно. При этом не нужно дорогостоящих и отнимающих уйму труда операций по химическому секвенированию, использованию наборов рестрикционных ферментов и прочих ухищрений. Метод начала использовать компания “Секвенум”, зарегистрированная в городе Сан-Диего (Калифорния) и руководимая Чарлзом Кэнтором.

Другой подход основан на присоединении флуоресцентных меток к ДНК, разрезании ДНК одним или несколькими рестрикционными ферментами на достаточно протяженные куски и оптическом анализе кусков. Так как флуоресцентные метки, сорбирующиеся на индивидуальных нуклеотидах, создают для каждого участка ДНК светящуюся картинку, характерную только для него, можно сравнивать ее с имеющимися в памяти компьютеров картинками. Для этого сотрудники компании “Силера джиномикс” создали прибор оптического “обстрела” протяженных ДНК (система Visionade) и математический алгоритм Gentig. Если после оптического просмотра остаются сомнения в точности нуклеотидных последовательностей в каких-то коротких участках, только эти участки и надлежит секвенировать химически. Оптическая “стрельба” по нарезанному участкам ДНК позволила достичь небывалой скорости в секвенировании: в свое время изучение генома кишечной палочки потребовало работы нескольких сот человек в течение 12 месяцев, в то время как система Visionade помогла расшифровать этот же геном в несколько минут.

АНАЛИЗ ГЕНОМОВ РАЗНЫХ ОРГАНИЗМОВ

В последнее время важные для медицины и сельского хозяйства сведения о геномах получены в разных странах. Так, британские исследователи из Сэнгеровского центра и Института молекулярной медицины Оксфордского университета полностью раскодировали две из четырнадцати хромосом основного патогена, вызывающего смертные случаи при заболевании малярией, *Plasmodium falciparum*. Секвенированы геномы большого числа микроорганизмов, вызывающих болезни человека.

Завершается работа над геномом растения *Arabidopsis thaliana*, представителя семейства крестоцветных (полностью данные станут доступными во второй половине 2000 года, но уже сейчас очевидно, как распределены от 20 000 до 25 000 генов по своим функциям; рис. 1). В проекте участвовали ученые из США, Европы и Японии. Это растение стало таким же важным объектом генетических исследований, как в свое время дрозофила. Отличие арабидопсиса от многих других высших растений заключается в том, что в его относительно компактной ДНК длиной 130 Мб (у арабидопсиса пять хромосом) содержится очень мало повторяющихся последовательностей. Это особенно удивительно, так как в близких к арабидопсису разновидностях капусты (кочанной, брокколи, кольраби, брюссельской и др.), рапса, сурепицы, редиски на долю повторов приходится иногда до трех четвертей генома.



*Белковые реакции включают в себя реакции, осуществляемые продуктами генов, вовлеченных в синтез, распад, модификацию, хранение и внутриклеточное перемещение различных продуктов метаболизма.

Рис. 1. Функциональная классификация генов, предполагаемая для 1,9 Мб участка генома *Arabidopsis*

Одним из неожиданных итогов геномики, существенных для будущего сельскохозяйственного производства, стало развитие нового направления, название которого в прямом переводе на русский звучит несколько неуклюже — питательная геномика. Известно, что многие сельскохозяйственные культуры несут недостаточное количество незаменимых для человека аминокислот (тех, которые не синтезируются в теле человека), микроэлементов, металлов, витаминов или, напротив, содержат вещества, в больших количествах вредные или даже опасные для человека. В последние десятилетия интерес врачей и диетологов к потреблению так называемой здоровой пищи, которая содержала бы сбалансированное количество всех нужных человеку ингредиентов пищи, небывало вырос. Практически на каждом пищевом продукте на Западе проставлены цифры, говорящие о том, какую долю от ежедневно рекомендованных норм потребления того или иного соединения приносит данный продукт. Сейчас установлено, во сколько раз необходимо увеличить потребление того или иного витамина и микроэлемента, чтобы понизить во столько-то раз риск раковых, сердечно-сосудистых, респираторных, обменных и иных заболеваний. На фоне этих успехов стало ясно, что изучение геномов растений, их метаболизма (целиком зависящего от определенных генов), разработка биотехнологических операций по переносу генов позволяют надеяться, что в ближайшее время, в считанные годы, ученые научатся получать растения с заранее выбранными свойствами в отношении их питательной ценности.

Помимо нескольких сот микроэлементов, металлов, минералов, витаминов, биологически активных соединений растения синтезируют от 80 000 до 100 000 других веществ, относящихся к нескольким классам соединений (из них более 700 каротиноидов, более 200 фитостероидов, более 4000 фенольных соединений, глюкозидов и пр.). Часть из них не поступает в пищу из-за сложившихся веками национальных и религиозных пристрастий к потреблению тех или иных растений или из-за биологических особенностей растений. Теперь в ожидании скорого окончания исследований генома риса (возможно, работа будет завершена в течение двух-трех лет), пшеницы, кукурузы (исследования начаты), дополненных данными об арабидопсисе и ряде других видов, ученые-диетологи получают возможность сформулировать лучше запросы к биотехнологам, а те смогут направленно вводить в растения желательные гены даже от растений, непригодных к пищевому использованию. Уже сегодня отработаны схемы клонирования в бактериях генов растений, управляющих синтезом многих витаминов, каротиноидов, жиров, а также белков, вовлеченных в перенос металлов в клетки растений. В этом отношении задержки

со стороны биотехнологов не предвидится (см. статью Ю.Ю. Глебы [7]). А вот тот факт, что у большого числа гомологичных генов, уже исследованных в разных растениях, установлено сходство в нуклеотидной последовательности (гомология наследственных записей была предсказана в начале 20-х годов XX века советским биологом Н.И. Вавиловым), позволяет надеяться, что для внесения нужных изменений в структуру синтезируемых растениями белков не потребуется громоздких и дорогостоящих манипуляций. Наличие таких ортологических последовательностей (Вавилов ввел другой термин — “гомологические ряды наследственной изменчивости”, но теперь генетики используют для сходных по строению генов термин “ортологические гены”) в большинстве исследованных видов растений свидетельствует о реалистичности этих надежд.

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

1. Сойфер В.Н. Международный проект “Геном человека” // Соросовский Образовательный Журнал. 1988. № 12. С. 4–11.
2. Кэнтон Ч., Шиммель П. Биофизическая химия. М.: Мир, 1984.
3. Nelson R. et al. Evidence for lateral gene transfer between *Archaea* and *Bacteria* from genome sequence of *Thermotoga maritima* // Nature. 1999. Vol. 399. P. 323–329.
4. Жданов В.М., Тихоненко Т.И. Вирусы и генетический обмен в биосфере // Современные проблемы биологии / Под ред. В.Н. Сойфера. М.: Знание, 1974. С. 165–190.
5. Гвоздев В.А. Подвижные ДНК эукариот // Соросовский Образовательный Журнал. 1998. № 8. С. 8–16.
6. Жимулев И.Ф. Современные представления об организации и функционировании полигенных хромосом // Там же. 1997. № 11. С. 2–7.
7. Глеба Ю.Ю. Биотехнология растений // Там же. 1998. № 6. С. 3–8.

Рецензент статьи Ю.А. Владимиров

Добавление при корректуре. 14 марта 2000 года президент США Уильям Клинтон и премьер-министр Великобритании Энтони Блэйр обратились с призывом к ученым, работающим над программой “Геном человека”, принять все меры к тому, чтобы максимально уменьшить коммерциализацию получаемых данных относительно генома человека. Это заявление сделано с учетом двух главных обстоятельств: (1) того, что не позднее осени этого года рабочий план будет полностью составлен, и (2) существенной разницы в обнаружении получаемых данных научными организациями, работающими на средства правительств США и Англии, и коммерческими организациями.

* * *

Сведения об авторе – см. с. 9.