

- lungstrends und Konsequenzen für die ärztlichen Aufgaben // München med. Wochenschr.—1986.—128, N 5.—S. 68—72.
45. Three research groups produce DNA probes for cystic fibrosis // Drug Hicence Report.—1985.—N 12.—P. 55.
  46. Atlan H. Biotechnologies, homme et nature // Futuribles.—1987.—N 108.—P. 7—25.
  47. Индукция нестабильных мутаций у *Drosophila melanogaster* микроинъекцией ДНК онкогенных вирусов в полярную плазмиду эмбрионов. Сообщ. 2. Анализ роли инъецируемой ДНК / К. Г. Газарян, С. Д. Набирочкин, Е. Н. Набирочкина и др. // Генетика.—1987.—23, № 2.—С. 214—227.
  48. Набирочкин С. Д., Набирочкина Е. Н., Газарян К. Г. Индукция нестабильных мутаций у *Drosophila melanogaster* микроинъекцией ДНК онкогенных вирусов в полярную плазмиду эмбрионов. Сообщ. 1. Зависимость от генотипа эмбриона // Там же.—С. 202—213.
  49. Emerman M., Temin H. M. Comparison of promoter suppression in avian and murine retrovirus vectors // Nucl. Acids Res.—1986.—14, N 23.—P. 9381—9396.
  50. Gélinas C., Temin H. M. Nondefective spleen necrosis virus-derived vectors define the upper size limit for packaging reticuloendotheliosis viruses // Proc. Nat. Acad. Sci. USA.—1986.—83, N 23.—P. 9211—9215.
  51. Rhode B. W., Emerman M., Temin H. M. Instability of large direct repeats in retrovirus vectors // J. Virol.—1987.—61, N 3.—P. 925—927.
  52. Plasmid mediated mutagenesis of a cellular gene in transfected eukaryotic cells / C. R. Brandt, F. M. Buonaguro, J. K. McDougall, D. A. Galloway // Nucl. Acids Res.—1987.—15, N 2.—P. 561—573.
  53. Dougherty J. P., Temin H. M. High mutation rate of a spleen necrosis virus-based retrovirus vector // Mol. and Cell. Biol.—1986.—6, N 12.—P. 4387—4395.
  54. Inactivation of a transfected gene in human fibroblasts can occur by deletion, amplification, phenotypic switching, or methylation / M. M. Gebara, C. Drevon, S. A. Harcourt et al. // Ibid.—1987.—7, N 4.—P. 1459—1464.

Ин-т молекуляр. биологии и генетики АН УССР,  
Киев

Получено 12.01.88

УДК 577.113.5+575.113

## ОПРЕДЕЛЕНИЕ ПОЛНОЙ НУКЛЕОТИДНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ ГЕНОМА ЧЕЛОВЕКА: ПРОЕКТЫ И ПЕРСПЕКТИВЫ

В. М. Кавсан

На страницах научных журналов еще продолжается дискуссия о том, нужно или не нужно расшифровывать первичную структуру всей ДНК наследственного аппарата человека, а уже ряд лабораторий США, Западной Европы и Японии получил гранты для начала работ, связанных с этой грандиозной задачей. Ее решение позволит значительно глубже понять и использовать молекулярные механизмы жизни. С одной стороны, это даст возможность распознать системы координированной регуляции экспрессии генов и локализовать гены, ответственные за длинный список наследственных болезней и, конечно, в первую очередь злокачественных новообразований. С другой стороны, определение последовательностей ДНК предоставит данные, имеющие огромный интерес сами по себе: они дадут основу для сравнения строения многих генов со сходными функциями, сопоставления организации целых семейств генов и последовательностей, ответственных за их активность, подойти к генетическому картированию индивидуальных личностей с наиболее высоким разрешением [1].

В течение последних двух лет в разных городах США прошло более десятка научных совещаний различного характера, посвященных обсуждению перспектив определения полной нуклеотидной последовательности генома человека. Одно из таких совещаний в г. Санта Фе, организованное DOE (Department of Energy), собрало чрезвычайно влиятельных ученых Соединенных Штатов в этой области естествознания. Среди них были У. Гилберт, Д. Балтимор, У. Водмер, М. Син-

гер, У. Дулитл и др., всего около 50 человек. Хотя по поводу необходимости такого проекта высказывались разные мнения, общее резюме было: "We are encouraged to proceed" («Мы настроены продолжать...»). Уже через несколько месяцев после совещания создан оргкомитет и на первый год назначен бюджет (20 млн долларов) для его работы по разработке проекта. Участие и организация совещания департаментом энергетики показательны и многообещающи: интерес крупных и состоятельных, но не имеющих прямого отношения к биологии учреждений предполагает отсутствие необходимости оттягивания средств от других биологических исследований.

Определение полной нуклеотидной последовательности генома человека, по мнению У. Гилберта, представляет собой «грааль генетики человека» [2]. Этот проект кажется гигантским даже в сравнении с такими, как проект физики высоких энергий или проект орбитальных станций. Если рассчитывать на использование современной технологии, он должен был бы занять 30000 человеко-лет и стоить около 2 млрд долларов (для сравнения, проект по сверхпроводимости оценивается в 3 млрд долларов).

В США пока нет единого мнения о форме разрабатываемого проекта. У. Гилберт предлагает создать специализированный Институт Генома Человека, а стратегия секвенирования должна быть в первую очередь направлена на особо интересные районы генома и участки, о которых известно, что они имеют важное значение. Однако остальные исследователи, предполагающие включиться в секвенирование человеческого генома, полагают, что будет лучше, если занимающиеся этой проблемой лаборатории будут разбросаны по стране или даже по всему миру. Это даст возможность с разных сторон улучшать и модифицировать методы секвенирования в разных лабораториях. Проект, правда, имеет и противников [3]. Мало кто отрицает нацело пользу полного секвенирования генома человека. В основном возражения сводятся к тому, что этот проект отвлечет средства от других, может быть, не менее или даже более важных биологических задач. Другие считают, что секвенирование меньших геномов, например, генома кишечной палочки или нематоды, потребует меньших затрат и в то же время даст больше информации о жизненных явлениях, чем секвенирование генома человека. Третьи думают, что полезнее определить последовательности основных, важных участков генома человека, чем расшифровывать полную последовательность всего генома. Тем не менее наиболее авторитетная группа ученых, среди которых лауреаты Нобелевской премии Д. Балтимор, У. Гилберт, Р. Далбеко, считает, что проект — это вход в новую эру биологии.

Немаловажным представляется вопрос об источнике ДНК, которую будут секвенировать. Это может быть либо ДНК одного индивидуума, либо материал из различных источников, например отдельные хромосомы от различных индивидуумов. В любом случае, конечно, будет утеря информация о вариациях человеческого генетического материала. Руководитель лаборатории в Лос-Аламосе М. Битенски предложил устроить соревнование между мужчинами и женщинами всего мира: кто сделает наибольший вклад в проект, ДНК того человека и будет секвенирована [4].

Две стратегии предлагаются для достижения цели проекта. Согласно одной из них, ДНК каждой из хромосом после их «сортировки» нарезаются рестриктазами типа *SfiI*, дающими относительно малое количество длинных фрагментов. Сравнительно небольшими усилиями путем секвенирования концов фрагментов и перекрывающихся последовательностей расположить эти фрагменты в хромосоме, после чего приступить к определению нуклеотидных последовательностей более мелких фрагментов (рис. 1). Хромосом-специфические банки относительно коротких фрагментов (5000—10000 п. п.) были сконструированы уже в середине 1986 г. В настоящее время заканчивается работа по созданию и характеристике банков, включающих фрагменты геномной ДНК человека

величиной до 50000 п. н. Разработаны, продаются и уже во многих лабораториях используются приборы для быстрого разделения хромосом в относительно больших количествах. Руководители DOE и ряда других учреждений США считают, что проект реально осуществим и его стоимость по сравнению с предполагавшейся вначале удастся в самое короткое время значительно снизить.

Другая стратегия в определении полной первичной структуры генома человека состоит в возможном сиквенировании фрагментов ДНК, дробленной случайным образом, а затем в составлении единой последовательности за счет перекрывающихся участков. В 1984 г. Е. Сэда

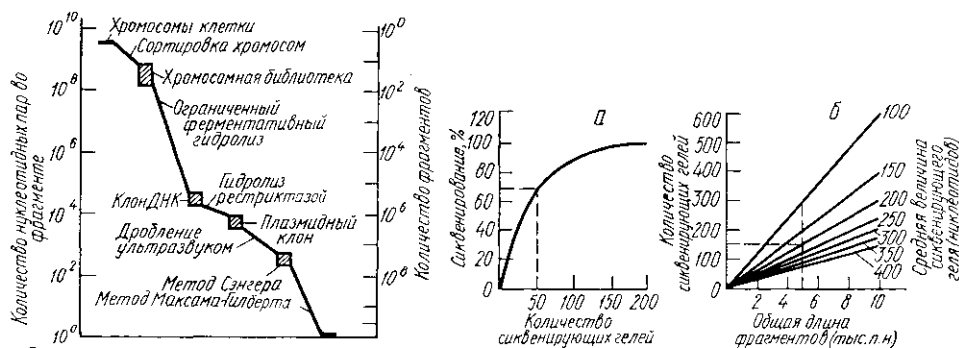


Рис. 1. Стадии фрагментирования человеческого генома [5]

Fig. 1. Fragmentation stages of human genome [5]

Рис. 2. Количество клонов и прочитанных гелей для сиквенирования ДНК [6]: а — полнота сиквенирования 5 тыс. п. н. ДНК как функция количества прочитанных гелей (предполагается, что с каждого геля считывается в среднем 250 п. н.); б — количество прочитанных гелей и полнота сиквенирования обеих цепей ДНК

Fig. 2. The number of clones and gel readings necessary for DNA sequencing [6]: а — sequence completion of 5 kb DNA as a function of the number of gel readings (it is assumed that on the average 250 base pairs are read every reading); б — the number of gel readings and sequence completion of both DNA chains

в Нац. ин-те генетики разработал метод «дробовика» (shotgun-method), позволяющий анализировать 7—8 тыс. п. н. в день [6]. Этот метод в последнее время получил значительное развитие и заключается в дроблении ДНК ультразвуком на короткие фрагменты, которые затем клонируются случайным образом в составе бактериофага *M13*, после чего используется улучшенная техника Сэнгера [7]. Далее результаты анализа обрабатываются компьютером для составления последовательности фрагментов (рис. 2). При такой стратегии расчеты показывают, что человеческий геном вручную (без автоматических машин) может быть сиквенирован за 1 год десятью тысячами исследователей, если принять величину ДНК около 1 м или 2,9 млрд п. н.

Конечно, такой стратегический подход к подготовке материала, как метод «дробовика» с последующим использованием высокой степени автоматизации сиквенса и мощных компьютеров для обработки полученных результатов, весьма перспективен. Однако сейчас пока еще большинство методов быстрого сиквенирования длинных фрагментов ДНК включает прогрессивное укорачивание фрагментов с последующим субклонированием в составе специальных векторов для сиквенирования (так называемых «сиквенирующих векторов»). Прогрессивное укорачивание ДНК осуществлялось обычно с помощью ДНКазы I, нуклеазы *Bal31* или экзонуклеазы III с последующей обработкой нуклеазой  $S_1$ . Эти методики занимают довольно много времени, а скорости гидролиза *Bal31* или экзонуклеазой III далеко не постоянны, что создает дополнительные трудности. Ахмед [8], а потом Пенг и Ву [9] предложили метод транспозон-опосредованных делеций. Они сконструировали плазмиду, соединяющую участок инициации репликации *ori* бактериофага *M13* и участок транспозона *Tn9* (рис. 3). Длинный фрагмент ДНК,

клонированный в составе такого вектора, будет прогрессивно укорачиваться с одного конца путем транспозон-зависимых делеций *in vivo*. После этого плазмиды, включающие различные по размерам фрагменты ДНК, могут быть превращены в одноцепочечную ДНК в том же хозяине, после чего последовательность можно определить с помощью техники Сэнгера. Метод не требует дополнительного субклонирования.

Очистка многочисленных фрагментов ДНК после клонирования в составе бактериофагов или плазмид представляет длительный и трудоемкий процесс. Автоматический экстрактор нуклеиновых кислот фирмы

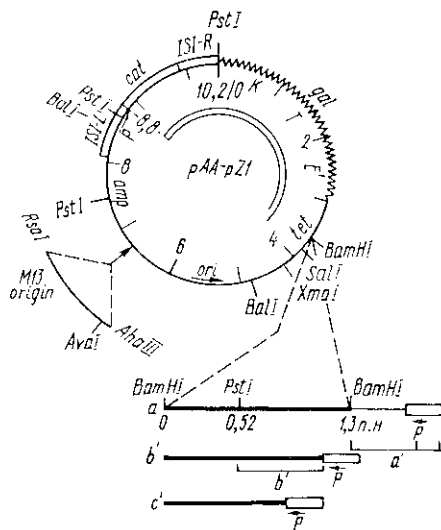


Рис. 3. Плаزمида pAA-PZ1-H3 до и после транспозон-опосредованной делеции [9]  
Fig. 3. pAA-PZ1-H3 plasmid before and after the transposon-mediated deletion [9]

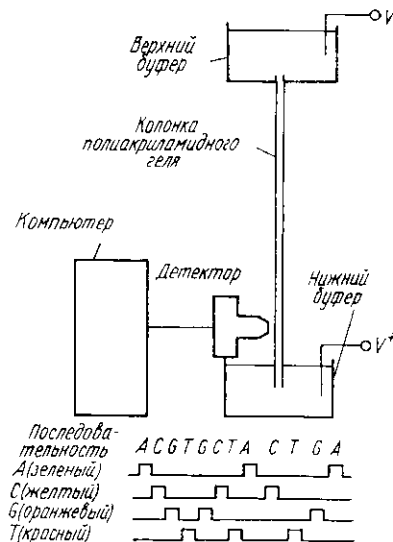


Рис. 4. Упрощенная схема автоматического сиквенатора ДНК [12]  
Fig. 4. Simplified scheme of the automatic DNA sequencer [12]

“Applied Biosystems” (США) осуществляет очистку ДНК или РНК из вирусов, клеток, биологических жидкостей, тканей. Использование автоматизации в этом процессе стандартизует процедуру, уменьшает возможность загрязнения нуклеазами, дает длинные молекулы ДНК, позволяет обрабатывать одновременно в едином модуле до восьми проб. Границы объема обрабатываемого биологического материала довольно широки — от 5 мкл до 1 мл. В автомате используются экстракция и колоночная хроматография, полностью заменившая центрифугирование. Вся процедура занимает около 3 ч [10].

В обоих использующихся в настоящее время методах сиквенирования фрагментов ДНК — энзиматическом Сэнгера [7] и химическом Максама—Гилберта [11] — обычно проводят четыре отдельные реакции. В энзиматическом методе эти реакции дают фрагменты ДНК, заканчивающиеся аденозином (А), цитозином (С), гуанозином (G) либо тимидином (Т). В химическом методе обычно получают фрагменты, заканчивающиеся G, G+A, C+T или C. В обоих случаях продукты реакций разделяются электрофоретически на соседствующих дорожках полиакриламидного геля с высоким разрешением. Получающийся автограф всего геля используется для определения относительных размеров фрагментов ДНК, образующихся в каждой из четырех реакций. Нуклеотидная последовательность ДНК определяется прямо из этой информации.

Хотя обе техники по критериям настоящего времени считаются высокоэффективными (на сегодняшний день определена последовательность более  $10^6$  п. н.), они, тем не менее, требуют значительной затраты

труда квалифицированных сотрудников и для их осуществления необходимо применение дорогостоящих радиоактивных изотопов.

Использование вместо радиоактивной флюоресцентной метки помимо прочих преимуществ способствует также автоматизации процесса. Л. Худ и соавт. из Калифорн. технол. ин-та поставили для себя целью решение двух задач. Во-первых, получать, сохранять и анализировать информацию о нуклеотидной последовательности прямо во время электрофореза (рис. 4). Это сделает ненужным перенос информации с геля на рентгеновскую пленку (радиоавтография), устранив тенденциозный ручной анализ и необходимость изготовления нескольких перекрывающихся гелей. Вместо этого единственный гель может «прогоняться» до тех пор, пока не будет достигнуто нужное разрешение. Во-вторых, избежать применения радиоактивных изотопов — опасных, дорогих и нестабильных [12].

Для решения обеих задач вместо изотопов стали использовать флюорофоры. Флюоресценция дает достаточную чувствительность для детекции в течение реального времени малых количеств ДНК, присутствующих на «сиквенирующем» геле (около 10 фмолей в полосе). Для каждого основания при этом используется свой флюорофор (т.е. для четырех оснований используются четыре различных флюорофора) в каждой специфической реакции. После этого продукты всех четырех реакций соединяются и вместе подвергаются электрофорезу. Образующиеся в каждой реакции фрагменты ДНК определяются в нижней части геля и идентифицируются по цвету. Эти данные постоянно собираются специальными датчиками во время электрофореза и сохраняются, а затем анализируются компьютером для выдачи готовой нуклеотидной последовательности. Принцип вначале был применен к стандартной технике сиквенирования ло Сэпгеру, а для затравки полимеразных реакций использованы синтезированные химическим путем флюоресцирующие олигодезоксирибонуклеотиды. Один из способов получения таких затравок заключается в синтезе производного тимидипа, содержащего фосфорамидитное начало у 3'-углерода и защищенную амминогруппу у 5'-углерода. При использовании этой молекулы в конечном цикле присоединения олигонуклеотидного синтеза фосфорамидитным методом продукт после снятия защиты и отщепления от твердой фазы представляет собой олигонуклеотид, содержащий единственную алифатическую аминогруппу на 5'-конце. Этот материал может быть теперь легко конъюгирован с различными коммерческими аминореактивными флюоресцирующими красителями.

Выбор флюоресцентной метки имеет большое значение для автоматического сиквенатора: 1) абсорбционный и эмиссионный максимумы должны быть в видимом районе спектра, как можно дальше от длинноволновой области, чтобы свести к минимуму рассеяние и флюоресцирующей фазой; 2) эмиссионные максимумы флюорофоров должны значительно отличаться друг от друга, чтобы краски были хорошо различимы; 3) краски должны быть сильно флюоресцирующими для достаточной чувствительности; 4) краски не должны заметно мешать гибридизации олигонуклеотидного праймера, так как при этом заметно уменьшается эффективность синтеза в соответствующих реакциях; 5) краски не должны значительно паразитировать и различаться в электрофоретической подвижности соответствующих ДНК-фрагментов; на самом деле это все же происходит и компьютеру приходится вводить соответствующие поправки. При нанесении на каждую дорожку геля по одной краске разрешение достигает более 400 оснований в течение около 16 ч, если же нанести все четыре краски, — разрешение пока только около 200 оснований, но система быстро модифицируется и улучшается. В конце 1986 г. машина при автоматизации всех реакций давала 1 % ошибок, сиквенируя одну цепь и 0.01 %, сиквенируя обе цепи, что соответствует лучшим стандартам на сегодня (1 ошибка на 10000 оснований) [13].

Использование флюоресцентной метки в технике Максама—Гил-

берта позволит использовать дешевые и стабильные реагенты, что даст возможность применить их для автоматизации процесса. Техника Максама—Гилберта включает пять последовательных этапов: специфические модификации оснований, остановка реакций, серии пересаживаний этанолом и центрифугирования, расщепление пиперидином и повторные выпаривания при уменьшенном давлении. Эта техника имеет целый ряд недостатков: сама методика довольно длительна, всегда есть возможность потери фрагмента ДНК при осаждении этанолом и центрифугировании; гидразин может копреципитировать с ДНК при осаждении этанолом и мешать специфичности реакции. Автоматизация процесса могла бы в какой-то мере устранить отмеченные недостатки.

Первый шаг к автоматизации техники сиквенирования ДНК по Максаму—Гилберту был сделан Чувпило и Кравченко [14], предложившими проводить реакции на твердой подложке, что устраняло потери и резко сократило время обработки. Недостаток метода — механическая непрочность предложенной ими подложки (бумага DE 81) — был устранен модификацией Розенталя и соавт. [15, 16] с использованием новой, основанной на целлюлозе, подложки с анионообменными свойствами. Однако такой подход хорош для сиквенирования только коротких олигонуклеотидов и не подходит для длинных фрагментов ДНК. Кроме того, пиперидин должен все же удаляться повторной лиофилизацией, что отнимает относительно много времени. Поэтому следующим шагом в автоматизации техники Максама—Гилберта было использование обращенно-фазовой хроматографии. Здесь радиоактивно меченный обычным способом фрагмент ДНК наносит в равных количествах на четыре колонки  $C_{18}$  Sep-рак. ДНК иммобилизуется смолой не изученным до конца способом и обрабатывается в течение 2 мин диметилсульфатом в колонке G, 4 мин — 50%-ной муравьиной кислотой в колонке A, 3 мин — 60%-ным гидразином в колонке T и 4 мин — 60%-ным гидразином + 1,5 M хлористым натрием в колонке C. Реакции идут на смоле с той же скоростью, что и в растворе, их останавливают в нужное время пропусканием через колонки раствора, согласно протоколу Максама—Гилберта, после чего колонки промывают 5%-ным ацетонитрилом в триэтиламмоний бикарбонате, удаляющим гидразин, муравьиную кислоту, диметилсульфат и др., а ДНК элюируют 10%-ным пиперидином. Элюаты нагревают до 90 °С, добавляют уксусную кислоту для образования пиперидинацтата, пропускают через свежие четыре колонки, новые элюаты выпаривают и запускают на «сиквенирующий» гель. При этом способе нет угрозы потери ДНК при осаждении этанолом, снижения специфичности в результате копреципитации этанолом гидразина, не нужны многочисленные центрифугирования, удаление пиперидина легко осуществляется в вакууме. Вся процедура занимает около 1,5 ч.

Сиквенирующая машина на основе техники Максама—Гилберта уже продается японской фирмой "Sciko". Эта машина обрабатывает одновременно 12 образцов и выдает за день последовательность до 6000 оснований. Работа этой машины основана на стандартной технике Максама—Гилберта с использованием радиоактивного фосфора для метки сиквенируемого фрагмента и радиоавтографии гелей. Лимитирующий фактор — приготовление гелей — устраняется выпуском готовых полиакриламидных блоков 20×40 см и 0,2 мм толщиной, уже покрытых ацетатной пленкой и годных для экспозиции с рентгеновской пленкой, так называемый "Gensog gel" фирмы "Fuji Photo Film Co." (Япония). Стоимость одной пленки 16 долларов. Для увеличения разрешения такие пленки предполагается делать также с контролируемым градиентом плотности.

Естественно, что работа с гелями и рентгеновскими пленками производится роботами. Для быстрой и правильной интерпретации радиоавтографов ряд фирм продает автоматические сканнеры. Так, фирма "Bio Rad" (США) предлагает Gene-Master DNA Workstation для введения, сохранения и анализа последовательностей нуклеиновых кислот. Для введения в компьютер нуклеотидной последовательности

используется тактильный карандаш, позволяющий считывать ее прямо с радиоавтографа в три раза быстрее, чем при обычном чтении. Пропуски избегаются специальным приспособлением, и голос повторяет каждое основание, которое вводится, чтобы проверять последовательность без отрыва от радиоавтографа. Скanners, выпускающиеся фирмами "Hitachi Software" и "Seiko" (Япония) для учета полос на радиоавтографе, имеют либо встроенные камеры, либо контактные сенсорные датчики. Правильное узнавание истинного сигнала на геле, необходимая коррекция деформированных полос и аккуратное удаление артефактных полос производятся с помощью специальной программы компьютера. С автографов среднего качества и разрешением до 250 оснований на каждом можно считывать нуклеотидную последовательность до 60 тыс. оснований в день с ошибкой менее 1%. Точность может быть увеличена в 10 раз при использовании улучшенного оптического процессора и компьютера более высокой квалификации. Проверка последовательности по двум комплементарным цепям ДНК снизит величину ошибки до 0,001%.

Фирма "Genofit" (Франция) предлагает прибор "Genoscope", позволяющий автоматически считывать разделяющиеся радиоактивные полосы по мере их разрешения у дна «сиквенирующего» геля. При этом можно использовать технику Сэнгера или Максама—Гилберта. Автоматический контроль компьютером, отсутствие необходимости обрабатывать автордиограммы с геля и т. д. в значительной мере сокращают время эксперимента.

Фирмой "Seiko" (Япония) вместе с серийным выпуском сиквенирующей машины, основанной на стандартной технике Максама—Гилберта, разработана модель машины для сиквенирования техникой Сэнгера. Производительность такой машины — до 30000 оснований в день и уже сейчас имеются идеи по ее улучшению. Одновременно и параллельно японскими учеными развиваются техника сиквенирования с использованием специфических флуоресцентных меток, характерных для каждого основания, и автоматизация этого процесса.

Задача стоит поистине грандиозная. Еще сегодня обычная техника сиквенирования позволяла каждый год добавлять в геномный банк компьютера последовательности до  $10^6$  оснований. Но только человеческий геном состоит из  $3 \cdot 10^9$  пар оснований, распределенных между 23 парами хромосом! Тем не менее история технологии показывает, что на смену повторяющимся простым действиям всегда приходила сложная индивидуальная задача для машины. Сиквенирование ДНК — первый кандидат на автоматизацию, что даст возможность освободить квалифицированных исследователей от повторяющихся механических и часто очень скучных и утомительных манипуляций, повысить объективность оценки результатов, обеспечить более быстрое накопление данных и удешевить стоимость анализа. Если сейчас опытный исследователь определяет последовательность  $10^3$  оснований в день ( $10^5$  оснований в год) с примерной стоимостью 1 доллар за основание, то машина должна довести это количество до  $10^6$  оснований в день стоимостью примерно 0,1 доллар за основание.

Для обработки огромного количества поступающей информации потребуются компьютеры Сгау-класса [18]. Конечно, понадобятся большие усилия для развития техники манипуляций с такой массой данных. Какой-то опыт оперирования нуклеотидными последовательностями уже имеется в американском «National DNA database», известном как «Gen Bank», однако в скором будущем появится примерно в  $10^3$  больше информации, чем имеется в настоящее время. Сейчас «Gen Bank» тратит 30 центов за одно основание при операциях по сохранению 6 млн оснований. В будущем, даже если стоимость снизится на порядок, затраты на оперирование накопленными данными составят более  $10^5$  долларов в течение 10 лет [13].

Не последнюю роль играет написание последовательностей огромной длины, читабельное как для исследователей, так и для машины.

Хайаши и Мунаката предлагают записывать четыре основания ДНК в виде четырех музыкальных нот [19]. Лейз и Финдлей [20] предпочитают графическое изображение, где четыре основания распределены в линейном формате: значения для G, A, T и C будут +2, +1, -1 и -2. В примере, показанном на рис. 5, 100 нуклеотидов из кодирующей последовательности мажорного гена комплекса гистосовместимости мыши изображены предлагаемым способом. Отмечены важные последовательности ТАТА-боксы (здесь ТАТТАТ) и АТG-кодон. Такое напи-

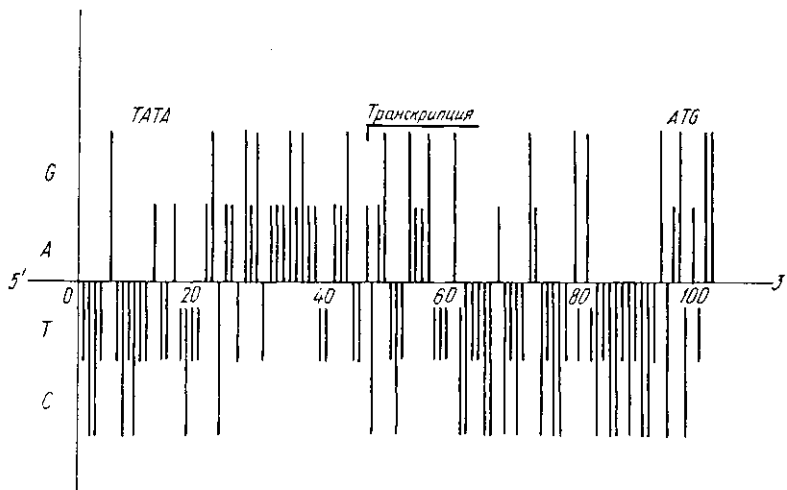


Рис. 5. Последовательность 100 н. мажорного гена комплекса гистосовместимости мыши E [20]. Верхние и нижние штрихи — А, G и C, T соответственно; короткие и длинные штрихи соответствуют GC- и AT-богатым участкам последовательности  
Fig. 5. Sequence of hundred nucleotides of the rat E major histocompatibility gene [20]

сание имеет несколько интересных преимуществ. Во-первых, легко различаются участки, богатые пуринами и пиримидинами. Во-вторых, сайты рестрикции здесь выявляются как вращательно-симметричные участки, легко выхватываемые глазом. В-третьих, весьма мало нужно времени для определения того, что пара CG встречается довольно редко. В-четвертых, чтобы превратить одну последовательность в последовательность комплементарной цепи достаточно только перевернуть ее вдоль горизонтальной оси. Такой узкий формат легко читается машиной и позволяет соответствующим образом подготовленному читателю перенести последовательность с печатной страницы прямо в компьютер.

Таким образом, уже сегодня имеется и интенсивно разрабатывается техника, позволяющая определить полную последовательность генома человека в реальные сроки с относительно умеренными затратами средств. Соединенные Штаты, Япония, страны Западной Европы по-разному намечают подходы к решению этой задачи, которые, однако, не исключают друг друга и способны к объединению. Один из проектов, предлагаемых Ф. Блатнером из Висконсинского ун-та (США), предполагает образование станций, разбросанных по всей стране. Каждая станция будет обслуживаться двумя техническими работниками и четырьмя автоматическими сиквенирующими машинами. Каждая машина будет иметь 40 линий. Каждая линия способна сиквенировать 750 остатков за смену. Время работы — 2 смены по 8 ч 5 дней в неделю. Следовательно, 750 остатков · 40 линий · 4 машины · 2 смены · 5 дней = 1,2 млн п. н. в неделю или 60 млн п. н. в год. Количество нуклеотидных последовательностей, которое необходимо сиквенировать с учетом перекрывающихся фрагментов, нужно увеличить по меньшей мере в 10 раз. Значит, для определения первичной структуры всего человеческого генома нужно сиквенировать  $35 \cdot 10^9$  п. н., т. е. 580 станций закончат работу за 1 год. Стоимость каждой станции оценивается в



0,5 млн долларов, вся стоимость проекта будет значительно меньше предполагавшейся вначале — менее 300 млн долларов [21].

В Японии, напротив, склоняются к образованию единого большого комплекса для определения последовательностей ДНК (фабрика сиквенса), который может быть международным центром для обслуживания лабораторий и индивидуальных исследователей всего мира. Сердцем такого комплекса должна стать высокопроизводительная сиквенирующая супермашина, подобная высокопродуктивным промышленным линиям, которые не только позволяют развить большую скорость процесса и дать экономию труда и затрат, но также обеспечивают легкость контроля и воспроизводимость, увеличивающие достоверность результатов. При Совете науки и технологии Японии создан Комитет по разработке автоматизации сиквенирования ДНК, работающий в содружестве с рядом промышленных компаний. В настоящее время все элементы разработанной линии сиквенирования уже существуют с интерфейсами между ними. Конечно, система включает устройства-роботы для отбора блюшек, экстракции и очистки ДНК из последних. За основу работы сиквенирующей машины в такой линии выбрана техника Сэнгера, хотя химический сиквенс по Максаму — Гилберту также предполагается использовать в случаях, когда применение техники Сэнгера неадекватно. Вся система, однако, роботизирована не полностью, так что опытный оператор играет в ней основную роль для получения точных результатов [22].

Предлагаются и уже вводятся в жизнь и другие проекты внедрения в последовательность человеческого генома. Так, С. Бреннер предлагает характеризовать клетки из разных дифференцированных типов тканей по генам, активным в клетках этих тканей, используя технику копирования молекул информационных РНК с помощью обратной транскрипции. Одна из задач заключается в идентификации целых классов генов, близких по строению и/или функции, которые затем могут быть размещены в человеческом геноме с помощью молекулярной гибридизации. Для этой цели С. Бреннер (директор Отделения молекулярной генетики в Кембридже) набирает команду из 6—12 человек, которая будет работать в Бабрахэме (близ Кембриджа) на базе Института исследований физиологии и генетики животных. Финансирование работы будет осуществлено отчасти полученной им премией (Louis Jeantet Foundation Prize) — 300 тыс. фунтов стерлингов; остальные деньги (до 200—250 тыс. фунтов стерлингов в год) будут предоставлены Советом мед. исследований и из других источников [23].

Конечно, сейчас трудно предсказать, какой из обсуждаемых подходов будет наиболее выигрышным и кого в гандикапе по определению полной нуклеотидной последовательности человеческого генома ждет триумфальный успех. Одно лишь несомненно: определение первичной структуры геномной ДНК будет важнейшей частью активности молекулярных биологов будущего и даже, возможно, последнего десятилетия нынешнего столетия. И сейчас, хотим мы этого или нет, наступают эра массированного чтения последовательностей оснований. Наиболее развитые капиталистические страны Запада и Востока интенсивно ведут организационные и технические работы, направленные на ускорение этого процесса. Отставание в нем, по выражению известного японского исследователя А. Вада [5], можно сравнить с отставанием в строительстве больших ускорителей частиц, гигантских телескопов, долгосрочных программ исследования космоса.

DETERMINATION OF THE COMPLETE NUCLEOTIDE SEQUENCE  
OF THE HUMAN GENOME: PROJECTS AND PROSPECTS

V. M. Kavsun

Institute of Molecular Biology and Genetics,  
Academy of Sciences of the Ukrainian SSR, Kiev

Summary

The lively discussion on human genome sequencing is continuing on the pages of science and popular journals. Despite this fact, the most developed west countries began the organization and technical works directed to the bulk reading of the DNA sequences. The cost of the suggested project became lower in the course of this review writing. Therefore, the work on this project is assumed to be completed within the real terms.

1. *Dulbecco R.* A turning point in cancer research: sequencing the human genome // *Science*.— 1986.— **231**, N 4742.— P. 1055—1057.
2. *Lewin R.* Proposal to sequence the human genome stirs debate // *Ibid.*—**232**, N 4758.— P. 1598—1600.
3. *Walsh J. B., Marks J.* Sequencing the human genome // *Nature*.— 1986.— **322**, N 6080.— P. 590.
4. *Palca J.* Department of energy on the map // *Ibid.*—**321**, N 6068.— P. 371.
5. *Wada A., Soeda E.* Strategy for building an automatic and high speed DNA-sequencing system // *Proc. 4th Congr. of the Fed. of Asian and Oceanian Biochemists*.— Singapore, 1986.— P. 1—16.
6. *Soeda E.* A «shotgun» DNA sequencing technology which can analyze genes at a rate of 8000 per day // *Technocrat*.— 1984.— **17**, N 7.— P. 39—42.
7. *Sanger F., Nicklen S., Coulson A. R.* DNA sequencing with chain-terminating inhibitors // *Proc. Nat. Acad. Sci. USA*.— 1977.—**74**, N 2.— P. 5463—5467.
8. *Ahmed A. A.* A rapid procedure for DNA sequencing using transposon-promoted deletions in *Escherichia coli* // *Gene*.— 1985.—**39**, N 2—3.— P. 305—310.
9. *Peng Z.-G., Wu R.* A simple and rapid nucleotide sequencing strategy and its application in analyzing a rice histone 3 gene // *Ibid.*— 1986.—**45**, N 3.— P. 247—252.
10. *Automated nucleic acid extractor* // *Process Biochem.*— 1986.—**21**, N 4.— P. 13.
11. *Maxam A. M., Gilbert W.* Sequencing end-labelled DNA with base-specific chemical cleavages // *Meth. Enzymol.*— 1980.— **65**, pt 1.— P. 499—560.
12. *Fluorescence detection in automated DNA sequence analysis* / L. M. Smith, I. Z. Sanders, R. I. Kaiser et al. // *Nature*.— 1986.— **321**, N 6071.— P. 674.
13. *Lewin R.* DNA sequencing goes automatic // *Science*.— 1986.—**233**, N 4759.— P. 24.
14. *Chuapilo S. A., Kravchenko V. V.* A simple and rapid method for sequencing DNA // *FEBS Lett.*— 1985.— **179**, N 1.— P. 34—37.
15. *Solid-phase methods for sequencing of nucleic acids. 1. Simultaneous sequencing of different oligodeoxyribonucleotides using a new, mechanically stable anion-exchange paper* / A. Rosenthal, S. Schwerther, V. Hahn, T. Hunger // *Nucl. Acids Res.*— 1985.— **13**, N 4.— P. 1173—1184.
16. *Gross B., Rosenthal A.* Simultaneous sequencing of RNA by solid-phase chemical degradation using DE 81 anion-exchange paper // *Gene Anal. Tech.*— 1987.— **4**, N 3.— P. 57—61.
17. *Jagadeeswaran P., Kaul R. K.* Use of reverse-phase chromatography in the Maxam-Gilbert method of DNA sequencing a step toward automation // *Ibid.*— 1986.—**3**, N 5.— P. 79—85.
18. *Parkinson D.* More parallel than others // *Nature*.— 1985.—**316**, N 6031.— P. 765—766.
19. *Hayashi K., Munakata N.* Basically musical // *Ibid.*— 1984.— **310**, N 5973.— P. 95.
20. *Lathe R., Findlay R.* Machine-readable DNA sequences // *Ibid.*— **311**, N 5987.— P. 610.
21. *Palca J.* The numbers game // *Ibid.*— 1986.— **321**, N 6068.— P. 371.
22. *Wada A.* Automated high-speed DNA sequencing // *Ibid.*— 1987.— **325**, N 6107.— P. 771—772.
23. *Maddox I.* Brenner homes on the human genome // *Ibid.*— **326**, N 6109.— P. 119.

Ин-т молекуляр. биологии и генетики АН УССР, Киев

Получено 25.02.88