Биомедицинская химия, 2008, том 54, вып. 3, с. 278-288.

НОВОСТИ НАУКИ

ВЫСОКОСКОРОСТНОЕ СЕКВЕНИРОВАНИЕ ГЕНОМА ДЖЕЙМСА УОТСОНА

Первый полный геном, секвенированный с использованием высокоскоростных технологий будущего поколения, опубликован в апрельском журнале "Nature" [1]. Он открывает новый этап в стремительно развивающейся области науки - секвенировании генома человека.

Для секвенирования 6 миллиардов пар оснований первооткртывателя ДНК Джеймса Уотсона потребовалось только четыре месяца, небольшая группа ученых и менее 1,5 миллионов долларов. Достигнутые результаты стали доказательством того, что высокоскоростные секвенирующие машины могут расшифровывать большие и сложные геномы [2]. В данном случае использовались механизмы, разработанные 454 Life Sciences (Connecticut), - подразделение Roche Diagnostics. Они позволяют осуществлять гораздо большее количество секвенирующих реакций в одно и то же время, на той же самой поверхности, по сравнению с предыдущим поколением машин, которые производили "первые, начальные" геномы человека [3,4]. Скорость, эффективность и, стоимость "оказались в выигрыше"(см. таблицу).

Таблица. Затраты на секвенирование (по [5] с изменениями).

Геном секвенирован (год опубликования)	**Проект "Геном человека" (HGP) (2003)**	**Геном VENTER (2007)**	**Геном WATSON (2008)**
Затраченное время (от начала и до конца)	**13 лет**	**4 года**	**4,5 мес.**
Число ученых-авторов	**>2800**	**31**	**27**
Стоимость секвенирования (от начала и до конца)	**2,7 миллиарда долларов**	**100 миллионов долларов**	**<1,5 миллиона долларов**
Количество институтов-участников	**16**	**5**	**2**
Количество стран-участниц	**6**	**3**	**1**

Геном Джеймса Уотсона - не первый полный опубликованный геном; геном известного генетика и предпринимателя Craig Venter был секвенирован с использованием машин предыдущего поколения [6], стоимость секвенирования составила 100 миллионов долларов (наш журнал уже рассказывал об этой работе [7]). По мнению Jonathan Rothberg (Rothberg Institute for Childhood Diseases, Guilford, Connecticut), основателя 454 Life Sciences, ведущего автора "Watson", геном Venter стал "окончанием прошлого поколения, прошлой эпохи секвенирования". Эта работа проводилась в январе 2007 года, с "последней январской технологией"; предварительные данные были опубликованы в мае 2007. Процесс секвенирования становился все более качественным и дешевым. Rothberg называет Уотсона "первым из нас", однако в данном случае низкая стоимость секвенирования по-прежнему лишь отдаленно приближается к намеченной цели в 1000 долларов, установленной X Prize Foundation.

Не все ученые согласны с тем, что этот метод лучше. По мнению Venter, новый стандарт технологий секвенирования еще не означает новый стандарт охвата генома и независимой сборки.

Одно из опасений состоит в том, что высокоскоростные машины "разрезают" ДНК на более короткие фрагменты (в данном случае, 250 оснований) для дальнейшей расшифровки по сравнению со старым методом, используемым международным Проектом "Геном человека" (HGP) и прежней компанией Venter Celera для своих последовательностей в 2001 году. Они использовали фрагменты в 500-1000 оснований. Более короткие фрагменты превращают повторную сборку в более сложный технический процесс, и это означает то, что тяжелее исследовать области генома, которые имеют большие, повторяющиеся последовательности.

По мнению Evan Eichler, генетика из University of Washington (Seattle), 5-10% генома состоит из этих сложных областей, которые содержат гены, вызывающие то или иное заболевание, и широко варьируют среди индивидуальных организмов. Eichler полагает, что технология секвенирования коротких фрагментов может не давать необходимую информацию в этих областях. Что мы знаем об этих специфических участках в геноме Джеймса Уотсона?

Другие ученые отмечают, что авторы проекта "Watson" в основном ссылаются на последовательность HGP, которую они использовали в качестве пособия при повторной сборке фрагментов. По словам Jonathan Eisen, эволюционного биолога из University of California (Davis), который планирует использовать эти скоростные механизмы для секвенирования других разновидностей организмов, эта научная работа, безусловно, доказывает то, что, имея на руках готовый геном (геном-"ссылку"), можно с успехом осуществлять высокоскоростную процедуру секвенирования 454 Sequencing (massively-parallel pyrosequencing system). Их следующая задача будет состоять в том, чтобы продемонстрировать успешные результаты без "эталонного генома".

По мнению Michael Egholm, вице-президента научно-исследовательского проекта 454, несмотря на все достижения в области секвенирования, нам известно очень мало о "правилах чтения книги жизни". Egholm представлял одну из сторон консультативного совета, разъясняющего Уотсону значение 20 мутаций в его последовательности, которые могут быть связаны с повышенным риском возникновения тех или иных заболеваний. По-существу учёные смогли рассказать очень мало. Однако, учёные уверены, это - трудное начало большого пути.

ЛИТЕРАТУРА

1. *Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., Chen Y.J., Makhijani V., Thomas Roth G., Gomes X, Tartaro K., Niazi F., Turcotte C.L., Irzyk G.P., Lupski J.R., Chinault C., Song X., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D.M., Margulies M., Weinstock G.M., Gibbs R.A., Rothberg J.M. (2008)* Nature, **452**, 872-876.
2. Olson M.V. (2008) Nature, **452**, 819-820.
3. The International Human Genome Mapping Consortium (2001) Nature, **409**, 934-941
4. *Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., Gocayne J.D., Amanatides P., Ballew R.M., Huson D.H., Russo Wortman J., Zhang Q., Kodira C.D., Zheng X.H., Chen L., Skupski M., Subramanian G., Thomas P.D., Zhang J., Miklos G.L.G., Nelson C., Broder S., Clark A.G, Nadeau J., McKusick V.A., Zinder N., Levine A.J., Roberts R.J., Simon M., Slayman C., Hunkapiller M., Bolanos R., Delcher A., Dew I., Fasulo D., Flanigan M., Florea L., Halpern A., Hannenhalli S., Kravitz S., Levy S., Mobarry C., Reinert K., Remington K., Abu-Threideh J., Beasley E., Biddick K., Bonazzi V., Brandon R., Cargill M., Chandramouliswaran I., Charlab R., Chaturvedi K., Deng Z., Di Francesco V., Dunn P., Eilbeck K., Evangelista C., Gabrielian A.E., Gan W., Ge W., Gong F., Gu Z., Guan P., Heiman T.J., Higgins M.E., Ji R.-R., Ke Z., Ketchum K.A., Lai Z., Lei Y., Li Z., Li J., Liang Y., Lin X., Lu F., Merkulov G.V., Milshina N., Moore H.M., Naik A.K., Narayan V.A., Neelam B., Nusskern D., Rusch D.B., Salzberg S., Shao W., Shue B., Sun J., Wang Z.Y., Wang A., Wang X., Wang J., Wei M.-H., Wides R., Xiao C., Yan C., Yao A., Ye J., Zhan M., Zhang W., Zhang H., Zhao Q., Zheng L., Zhong F., Zhong W., Zhu S.C., Zhao S., Gilbert D., Baumhueter S., Spier G., Carter C., Cravchik A., Woodage T., Ali F., An H., Awe A., Baldwin D., Baden H., Barnstead M., Barrow I., Beeson K., Busam D., Carver A., Center A., Cheng M.L., Curry L., Danaher S., Davenport L., Desilets R., Dietz S., Dodson K., Doup L., Ferriera S., Garg N., Gluecksmann A., Hart B., Haynes J., Haynes C., Heiner C., Hladun S., Hostin D., Houck J., Howland T., Ibegwam C., Johnson J., Kalush F., Kline L., Koduru S., Love A., Mann F., May D., McCawley S., McIntosh T., McMullen I., Moy M., Moy L., Murphy B., Nelson K., Pfannkoch C., Pratts E., Puri V., Qureshi H., Reardon M., Rodriguez R., Rogers Y.-H., Romblad D., Ruhfel B., Scott R., Sitter C., Smallwood M., Stewart E., Strong R., Suh E., Thomas R., Tint N.N., Tse S., Vech C.,Wang G., Wetter J., Williams S., Williams M., Windsor S., Winn-Deen E., Wolfe K.,Zaveri J., Zaveri K., Abril J.F., Guigo R., Campbell M.J., Sjolander K.V., Karlak B.,Kejariwal A., Mi H., Lazareva B., Hatton T., Narechania A., Diemer K., Muruganujan A., Guo N., Sato S., Bafna V., Istrail S., Lippert R., Schwartz R., Walenz B., Yooseph S., Allen D., Basu A., Baxendale J., Blick L., Caminha M., Carnes-Stine J., Caulk P., Chiang Y.-H., Coyne M., Dahlke C., Mays A.D., Dombroski M., Donnelly M., Ely D., Esparham S., Fosler C., Gire H., Glanowski S., Glasser K., Glodek A., Gorokhov M., Graham K., Gropman B., Harris M., Heil J., Henderson S., Hoover J., Jennings D., Jordan C., Jordan J., Kasha J., Kagan L., Kraft C., Levitsky A., Lewis M., Liu X., Lopez J., Ma D., Majoros W., McDaniel J., Murphy S., Newman M., Nguyen T., Nguyen N., Nodell M., Pan S., Peck J., Peterson M., Rowe W., Sanders R., Scott J., Simpson M., Smith T., Sprague A., Stockwell T., Turner R., Venter E., Wang M., Wen M., Wu D., Wu M., Xia A., Zandieh A., Zhu1 X.* (2001) Science, **291**, 1304-1351.
5. *Wadman M.* (2008) Nature, **452**, 788.
6. *Levy S., Sutton G., Ng P.C., Feuk L., Halpern A.L., Walenz B.P., Axelrod N., Huang J., Kirkness E.F., Denisov G., Lin Y., MacDonald J.R., Wing Chun Pang A., Shago M., Stockwell T.B., Tsiamouri A., Bafna V., Bansal V., Kravitz S.A., Busam D.A., Beeson K.Y., McIntosh T.C., Remington K.A., Abril J.F., Gill J., Borman J., Rogers Y.-H., Frazier M.E., Scherer S.W., Strausberg R.L., Venter J.C.* (2007) Plos Biol., **5**, e254-e286.
7. Биомедицинская химия (2007) том 53, вып.6, с.705-712.

NIH ПРЕТВОРЯЕТ В ЖИЗНЬ НОВЫЙ ПЛАН

В феврале 2008 года консультанты Национальных Институтов Здоровья США, (NIH, Bethesda, Maryland) представили практически готовый план по "усовершенствованию" перегруженной системы экспертного анализа и экспертных оценок NIH. Они предлагают NIH провести полную "ревизию", своего рода "реконструкцию", которая позволит ускорить проведение экспертного анализа и откроет перспективные возможности для привлечения новых идей и концепций. Один из вариантов - упростить процесс пересмотра заявок на предоставление грантов. Предложенные нововведения обнадёжили исследователей, хотя некоторые из них считают, что NIH следует сперва проверить "план реконструкции" на практике.

Летом 2007 года директор NIH Elias Zerhouni уже задавался вопросом о том, как помочь NIH справиться с такими "перегрузками" и облегчить положение рецензентов. Учреждение получает рекордное количество заявок - около 80000 ожидается в 2008 году - в тот самый момент, когда бюджет NIH переживает не лучшие времена. Zerhouni сформировал два консультативных комитета, один внутренний в NIH и другой внешний, и поставил перед ними задачу: как финансировать "лучшее в науке" и при этом свести к минимуму "административные сложности". Многие предложения, принятые этими группами, представлены в докладе, опубликованном в журнале "Science" (от 14 декабря 2007, стр. 1708).

Основная рекомендация объединенной группы состоит в том, чтобы избегать рутинного пересмотра поступающих предложений и их повторного рассмотрения более 2 раз. Эти заявки с "внесенными поправками и изменениями" ставят в очередь перед новыми заявками, и разумнее дать шанс именно таким заявителям. Талант зачастую уступает место настойчивости, и по мнению Zerhouni, необходимо изменить существующее положение вещей.

Консультанты считают, что некоторые заявки должны быть сразу помечены как "не рекомендованные для повторного представления" во время первого просмотра. По словам Keith Yamamoto, сопредседателя внешней группы экспертов (University of California, San-Francisco), такие "быстрые отказы" могли бы в целом сэкономить 20 % времени. Предложения, которые преодолевают этот первый барьер, но не оцениваются как лучшие, могут столкнуться с более жестким отбором. Группа экспертов, скорее предпочла бы отказаться от категории "исправленных" заявок и рассматривать все предложения как "новые". Раздел, посвященный опровержениям экспертных оценок, упразднен; вместо этого, автор гранта просто включает их в новую заявку.

Кроме того, группа консультантов рекомендует редактирование критериев экспертного анализа и проводимых процедур. NIH должен сократить форму заявки, составляющую 25 страниц, и уделить больше внимания самим инновациям, упразднив при этом методы и предварительные данные. Отделы по изучению предложений должны оценивать все заявки, даже отклоненные, по пяти основным критериям (например, производит впечатление или нет?), чтобы заявители могли как-то ориентироваться. Консультанты также предлагают другой способ избежать двусмысленности: помимо оценок и баллов, каждую заявку должны поставить на определенное место (с первого до последнего). Для более качественной работы число рецензентов для каждого предложения должно быть удвоено с двух до четырех и более.

Объединенная группа консультантов должна помочь NIH расходовать средства более эффективно. Принимая во внимание, что многократные гранты предоставляются лишь небольшому проценту исследователей, NIH должен гарантировать оптимальное использование своих ресурсов, требуя от исследователей, по крайней мере, 20 % участия в каждом гранте. Это может ограничить большинство исследователей получением трех или четырех грантов.

По словам Zerhouni, главная цель - обеспечить максимальную поддержку молодым исследователям. Консультанты предлагают NIH отдельно рассматривать заявки, подающиеся впервые, привлекая при этом к работе рецензентов широкого профиля, скорее, чем узких специалистов. Для поощрения научных проектов, связанных с повышенным риском, группа предлагает NIH посвятить, по крайней мере, 1% выделяемых исследовательских грантов, таким механизмам, как, например, Pioneer Award, в основе которого лежит прежде всего достижение самого ученого, а не научно-исследовательский проект. По мнению Yamamoto, NIH мог бы выделять 300-400 грантов в год для таких "смельчаков", в пять раз больше, чем сейчас.

ИНСТИТУТ ГОВАРДА ХЬЮЗА ВЫДЕЛИЛ 300 МИЛЛИОНОВ ДОЛЛАРОВ МОЛОДЫМ ИССЛЕДОВАТЕЛЯМ

Обеспокоенная тем, что молодые исследователи продолжают "топтаться на месте", одна из крупнейших в мире благотворительных организаций, работающих в сфере биомедицинских исследований США, на этой неделе обнародовала план по "спасению" некоторых из них. Медицинский Институт Говарда Хьюза (Howard Hughes Medical Institute, HHMI, Chevy Chase, Maryland), выделяет 300 миллионов долларов на 6 лет для поддержки талантливых исследователей, которым приходится бороться за получение первого независимого федерального гранта. Hughes намеревается профинансировать около 70 человек в этом году.

Президент HHMI Thomas Cech рассматривает новую программу как "аварийную" помощь молодым ученым, которым на протяжении 5 лет распределяли непропорционально низкобюджетные средства в Национальных Институтах Здоровья (NIH, Bethesda, Maryland). Поскольку доля успешных исследовательских грантов NIH ничтожно мала, Cech и другие руководители обеспокоены тем, что многие молодые исследователи откажутся от фундаментальных исследований.

HHMI планирует помочь ученым с лабораторной практикой и поддерживать их на протяжении 2-6 лет в качестве доцентов - именно тех, кто рано добивается карьерных достижений, но в настоящее время имеет только 18% вероятности получить свой первый грант NIH R01. По словам Cech, в ведущих научных заведениях молодые талантливые ученые загнаны и подавлены. Молодые исследователи, получившие назначение сроком на 6 лет, в среднем получат приблизительно 700000 долларов в год - что равно двум грантам NIH R01 - и разделят предоставленные средства на исследования, заработную плату и службы института.

"Аварийная помощь" Howard Hughes - это "капля в море" по сравнению с тем вызовом, который получил NIH. Впервые в истории NIH намеревается предоставлять премии R01 в среднем 1500 молодым исследователям ежегодно, в 2006 году эта цифра составляла 1353 человека. Готовятся и другие проекты NIH: в этом году Pathway to Independence ("Путь к независимости") предоставит около 170 премий на обучение и исследования, и New Innovators ("Молодые новаторы") предоставят 24 гранта на 5 лет молодым ученым для важнейших исследований и разработок.

HHMI планирует профинансировать по крайней мере еще один проект вплоть до 2011 года, чтобы помочь молодым людям преодолеть этот сложный период. Chech надеется, что не придется проводить подобную программу больше, чем дважды.

НОВЫЕ УСПЕХИ В ПРЕДСКАЗАНИИ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ БЕЛКОВ

150,000 пользователей персональных компьютеров стали участниками научного эксперимента - ученые предсказали структуру белка, используя только его аминокислотную последовательность. По мнению экспертов, проект стремительно прогрессирует, что вселяет надежду на достижение ощутимых результатов.

Определение формы белка обычно осуществляют при помощи рентгеноструктурного анализа белковых кристаллов, и химики, изучающие природу белков, долгое время скептически относились к попыткам заменить этот практический метод моделированием или теорией. По мнению Michael Levitt, компьютерного биолога из Stanford University, само понятие моделирование на данном этапе развития белковой химии носит несколько устрашающий характер. В журнале "Nature" опубликована статья, в которой David Baker, биохимик из University of Washington (Seattle) и его коллеги сообщают о результатах исследований, которые могут развеять такие скептические настроения (B. Qian et al. Nature doi:10.1038/nature06249; 2007).

Форму белка, а следовательно и его активность, определяет точный фолдинг аминокислотной последовательности. По мнению Eleanor Dodson, структурного биолога из University of York (UK), если представить, что структура белка напоминает извивающуюся змею, тогда можно свободно экспериментировать на любом участке аминокислотной последовательности. Форма, полученная в результате эксперимента, зависит от молекулярных взаимодействий каждого аминокислотного остатка с "соседями", с окружающими водными молекулами и другими удаленными остатками, которые располагаются ближе в результате фолдинга. Моделирование в данном случае представляет собой огромную проблему.

Методика Baker объединяет уже известную информацию о структурах белка с огромными компьютерными возможностями Berkeley Open Infrastructure for Network Computing. Это программное обеспечение, разработанное в University of California (Berkeley), позволяет компьютерным пользователям участвовать в научных проектах (самый популярный проект - поиск внеземных цивилизаций в форме SETI@home); 150000 добровольцев загрузили копию программы лаборатории David Baker Rosetta@home program.

Rosetta разбивает аминокислотную последовательность белка на короткие фрагменты, которые могут быть противопоставлены идентичным фрагментам белков с известными структурами. Эти формы предлагают множество способов, как "связать" исследуемый белок воедино, и программа отбирает лишь те варианты, которые минимизируют свободную энергию структуры (показатель её стабильности). Неоднократно используя программу на тысячах компьютеров, исследователи экспериментально приблизились к более точной модели белка.

Когда ученые ввели последовательность T0283, бактериального белка, состоящего из 112 аминокислот, сеть предложила несколько миллионов структур после миллиона часов вычислений. Эти миллионы уменьшили до пяти дальнейшим повторным компьютерным анализом, и впервые с высокой точностью удалось предсказать одну из структур белка, коррелирующую с определенной структурой кристалла.

Хотя точность структуры не совсем отвечала кристаллическим моделям с высоким разрешением, этого оказалось вполне достаточно для исследователей, чтобы упростить процесс получения структур кристаллов с помощью рентгеновских лучей в будущем. Ученые должны производить модели на основе кристаллов, "обогащенных" маркерами тяжелых металлов, или же имеющими другие показатели будущей структуры, например, форму родственного белка

для того, чтобы превращать рентгеновские модели в структуры. Структура Rosetta для T0283 оказалась вполне подходящей. Эта программа способна найти такие "ориентиры" для белков, у которых недостаточно "полезных родственников" и чьи структуры по-прежнему не разгаданы.

По мнению Rhiju Das, соавтора David Baker, работающего над проектом Rosetta@home project, необходимо совершенствоваться дальше. Каждый персональный компьютер работает изолированно. Если бы программу можно было переписать и запускать одновременно на многих процессорах одного суперкомпьютера, Rosetta могла бы стать значительно мощнее.

Более точное предсказание структуры белка позволит в будущем производить уникальные белки. Baker использует Rosetta для отбора последовательностей, которые соответствуют необходимым структурам. Его лаборатория совместно с биохимиком Bill Schief (University of Washington) в настоящее время работают над перепроектированием белка gp120 вируса иммунодефицита человека для производства вакцины, стимулирующей иммунную систему различными способами, на основе естественного вируса. Измененный белок должен выявить антитела, которые атакуют вирус более эффективно, чем антитела, созданные после инфицирования.

Время, когда разработчики белков могли посчитать кристаллизацию ненужным процессом, давно в прошлом. По мнению David Baker, объединение двух методов конструирования (сбор экпериментальных данных и моделирование) открывает для биологов новые перспективы: создание большего количества белков в короткие сроки.

ПОСТГЕНОМНАЯ ВИЗУАЛЬНАЯ МОДЕЛЬ

Наиболее характерная особенность "постгеномной эры" в биологии заключается в накоплении огромного количества генотипных и фенотипных данных, которые должны быть систематизированы, проанализированы, визуализированы и интерпретированы. Такой ряд возможностей стал основным для современной биоинформатики. На сегодняшний день наиболее популярным графическим представлением для визуализации стала "кластерная тепловая карта", которая компактизирует большое количество информации в небольшое пространство для получения когерентных моделей данных. Однако, несмотря на ее популярность, возникает вопрос: оптимальны ли такие карты для визуального интегрирования информации и генерирования свежих гипотез? Ответить на этот вопрос можно, изучив достоинства и недостатки визуализации тепловых карт.

С момента их появления более 10 лет назад (I) кластерные тепловые карты появились более чем в 4000 биологических или биомедицинских публикациях. Они использовались для двухмерного изображения моделей во всех типах молекулярных данных, включая экспрессию мРНК и микроРНК экспрессию белка, номер копии ДНК, метилирование ДНК, концентрацию метаболитов и активность лекарств [1-8]. Они оказались полезными для данных микрочипов [2] и иногда применялись для "интегромного" слияния [1, 9, 10] различных типов молекулярной информации.

В случае с данными генной экспрессии цвет, обозначающий точку в сетке тепловой карты, указывает, сколько определенной РНК или белка экспрессируется в данном образце. Уровень высокой генной экспрессии обычно обозначается красным, а низкой - зеленым и синим. Когерентные модели (фрагменты) цвета генерируются иерархическим кластерингом на горизонтальных и вертикальных осях, чтобы свести подобное с подобным. Кластерные отношения обозначаются древоподобными структурами, смежными с тепловой картой и фрагменты цвета могут обозначать функциональные отношения между генами и образцами. Иногда кроме кластеринга по одной или обеим осям откладывается, например, время серийных измерений. Когерентные модели цвета не могут существовать без некоторой основы для функциональной последовательности по обеим осям.

Кластерная тепловая карта, привлекательная на первый взгляд, имеет свои недостатки и определенный потенциал для неверной интерпретации или неправильного использования. Наиболее заметным среди них является понимание данных только первого порядка; сложные модели нелинейных отношений среди малого количества образцов вряд ли обнаруживаются. Для выявления таких взаимоотношений [11] был разработан компьютерный вариант, основанный на "бикластеринге". Вторая проблема состоит в том, что в иерархическом кластеринге (иерархическая система классификации белков) каждое разветвление кластерного древа может "раскачиваться" в любом направлении в каждой ветке древа; поэтому должно быть какое-то объективное правило (до некоторой степени, правда, спорное): в каком направлении каждая ветка будет фактически "раскачиваться". Есть также соблазн выбрать маленький поднабор переменных (например, гены в микрочиповом исследовании) и представить их в кластерной тепловой карте. Это обычная практика при обнаружении новых биомаркеров и сигнатур генной экспрессии для различения подтипов заболеваний таких, например, как рак [12]. Однако, если выбирать сигнатуру, состоящую только из нескольких десятков генов из 10000-ного набора, даже тогда рандомизированные данные могут составить кластерные тепловые карты, которые оказываются ложными при выявлении различий двух подклассов.

Однако даже без учета этих недостатков создание кластерных тепловых карт - удивительно утонченный процесс, который требует использования большого количества вариантов, чем и продиктован тип и значение возникающей модели. Необходимые решения включают: (i) алгоритм предварительной обработки (например, тип второстепенного вычитания, нормирование и фильтрация данных), который, возможно, минимизирует шум в системе при хранении значимого сигнала: (ii) алгоритм кластеринга (например, усредненное связывание, полное связывание или центроидное связывание), который определяет, как будут группироваться данные; (iii) метрическое расстояние (например, евклидово расстояние или корреляция), которое определяет то, что подразумевается под схожестью генов или образцов друг к другу; и (iv) цветная схема (линейная, логарифмическая, квантильная, двухцветная или трехцветная), которая определяет, какие образцы в данных будут выделены визуально. Следует принять решение об использовании относительных или абсолютных данных. Один набор данных может быть вычтен из другого для создания кластерной карты "разницы" (например, данные генной экспрессии до и после воздействия лекарства на клетки). Дело в том, что многие различные тепловые карты, каждая со своим собственным визуальным значением, могут быть созданы в результате одного и того же эксперимента. Если не указаны детали параметрических вариантов, в лучшем случае анализ будет неполным, а в худшем - неверно интерпретирован.

Постгеномные наборы данных становятся все больше и больше. Десять лет назад микрочипы производили тысячи чисел; а сейчас они часто производят миллионы. Программное, материальное обеспечение, математические алгоритмы и человеческий разум загружены до предела потоком данных. Человеческие и вычислительные ресурсы научно-исследовательских учреждений часто не

соответствуют поставленной задаче. Наша потребность в графических представлениях, разъясняющих модели данных и способствующих возникновению новых гипотез, будет только возрастать в следующие несколько лет, особенно при идентификации биомаркеров, полезных для "персонифицированного" лечения таких заболеваний, как рак. Благодаря способности человеческого глаза распознавать образцы, связанные с современным анализом данных, изотерическая научная информация получит всестороннюю оценку [13]. Последние десять лет в постгеномной биологии в этих целях использовалась универсальная кластерная тепловая карта, пусть и не всегда успешно.

ЛИТЕРАТУРА

1. *Weinstein J.N.* (1997) Science, **275**, 343.
2. *Eisen M.B., Spellman P.T., Brown P.O., Botstein D.* (1998) Proc. Natl. Acad. Sci. USA, **95**, 14863.
3. *Ross D.T.* (2000) Nat. Genet., **24**, 227.
4. *Scherf U.* (2000) Nat. Genet., **24**, 236.
5. *Szakacs G.* (2004) Cancer Cell, **6**, 129.
6. *Zeeberg B.R.* (2005) BMC Bioinformatics, **6**, 168.
7. *Nishizuka S.* (2003) Proc. Natl. Acad. Sci. USA, **100**, 14229.
8. *Brauer M.J.* (2006) Proc. Natl. Acad. Sci. USA, **103**, 19302.
9. *Myers T.G.* (1997) Electrophoresis, **18**, 467.
10. *Weinstein J.N., Pommier Y.* (2003) C.R.Biol., **326**, 909.
11. *Kluger Y., Basri R., Chang J.T., Gerstein M.* (2003) Genome Res., **13**, 703.
12. *Golub T.R.* (1999) Science, **286**, 531.
13. *Nesbit J., Bradford M.* (2007) Science, **317**, 1857.

РЕКОНСТРУКЦИЯ ГЕНОМОВ

Главный герой поэмы "Наследственность" английского классика Томаса Гарди олицетворяет идею о том, что вся жизнь представляет собой "спуск по лестнице" из одного поколения в следующее. С научной точки зрения, репликация и воспроизведение генетического материала, ДНК или РНК, - важнейший механизм, при помощи которого каждое поколение передает "инструкции", определяющие наследственные черты и особенности потомков. Gibson и соавторы [1] попытались избежать "природных ограничений" путем объединения научных данных и сырьевых химических препаратов для конструирования полного набора генетического материала или генома, кодирующего живую бактерию. Первое конструирование генома, кодирующего самовоспроизводящийся организм, открывает новые важные перспективы в области генетики и биотехнологий, подчеркивает необходимость совершенствования технологий секвенирования ДНК и укрепляет значимость продолжающихся общественных споров относительно создания более простых и легких способов проектирования организмов.

Авторы использовали многоступенчатый процесс для конструирования генома *Mycoplasma genitalium*. Сначала из компьютерной базы данных получили информацию, определяющую последовательность ДНК генома в 582970 пар оснований, которую нужно синтезировать, и разделили ее на короткие фрагменты ДНК длиной до 7000 пар оснований. Затем коммерческие поставщики ДНК

сконструировали эти фрагменты. Сырьевые материалы, полученные из сахарного тростника, использовали для синтеза специфических олигонуклеотидов, коротких фрагментов ДНК в несколько сотен пар оснований [2]. Затем поставщики объединили подгруппы олигонуклеотидов для производства необходимых генетических "кассет" (фрагментов ДНК) [3]. Gibson и соавторы использовали иерархическую схему для сборки, проверки и восстановления более длинных фрагментов ДНК, в конечном счете воссоздавая геном в полную длину.

Вся жизнь кодируется определенным генетическим материалом, поэтому современные и будущие достижения в области синтезирования ДНК и технологий конструирования генома чрезвычайно важны. Например, Национальные институты здоровья США ежегодно тратят 1,5 миллиарда долларов на поддержку "ручной" работы с ДНК [4], которая требует неимоверных усилий от биологов и инженеров.

Новая улучшенная технология позволяет воспроизводить любые молекулы ДНК быстро, надежно и экономично, что расширяет масштабы технических исследований [5] и может стать основной целью скоординированных общественных научных проектов. К сожалению, на сегодняшний день таких проектов пока нет.

Считается, что самые ранние открытия кодируемых генетически функций зависят от анализа взаимосвязи между естественными или беспорядочно генерированными мутациями и фенотипами [6]. За последние 30 лет в процессе создания [7] и разработки [8] технологии секвенирования ДНК появился дополнительный метод обнаружения генетических функций. Сравнивая информацию о нуклеотидных последовательностях ДНК различных организмов, исследователи могут теперь идентифицировать последовательности, которые оставались неизменными на протяжении миллионов лет эволюции [9].

Для подтверждения и исчерпывающей идентификации всех функций, закодированных в природной последовательности ДНК, необходимы две дополнительные технологии. Определенные последовательности ДНК, влияющие на фенотипы, должны быть целенаправленно изменены и предполагаемый эффект должен быть подтвержден. "Посторонние" (несущественные) последовательности ДНК должны быть удалены, разрушены или модифицированы как ненужные. До настоящего времени применение этих дополнительных подходов было ограничено короткими последовательностями ДНК [10] или хорошо изученными организмами. Разрабатывая технологии конструирования генома, Gibson и соавторы надеются более тщательно исследовать, можно ли разрушить индивидуальные гены или их комбинации. Возможность претворять в жизнь множественные изменения в последовательностях природной ДНК [13], а также создавать и тестировать синтетические системы [14] откроет исследователям новый путь в мир "генетического конструирования жизни".

"Синтетический" геном в 582970 пар оснований, произведенный Gibson красноречиво показывает, что теперь можно конструировать геномы для всех известных вирусов человека, включая опасные патогены (например, оспы), на основе доступной информации, методах и материалах о секвенировании ДНК. Пока процесс конструирования генома, также как производство инфекционного агента с учетом заново синтезированного, но инертного генома, требует привлечения высоко квалифицированных специалистов и значительных ресурсов. Предприняты международные усилия для установления и координирования безопасности среди конкурирующих поставщиков ДНК [15]. Новые задачи и цели - сосредоточиться на развитии профессиональных сообществ и усовершенствовать стандарты практики среди биоконструкторов.

ЛИТЕРАТУРА

1. *Gibson D.G, Benders G.A., Andrews-Pfannkoch C., Denisova E.A., Baden-Tillson H., Zaveri J., Stockwell T. B., Brownley A., Thomas D.W., Algire M.A., Merryman C., Young L., Noskov V. N., Glass J.I., Venter J.C., Hutchison C.A., Smith H.O.* (2008) Science, **319**, 1215.
2. *Sanghvi Y.* (2007) A Roadmap to the Assembly of Synthetic DNA from Raw Materials, http://hdl.handle.net/1721.1/39657.
3. *Kodumal S.J., Patel K.G., Reid R., Menzella H.G., Welch M., Santi D.V.* (2004) Proc. Natl. Acad. Sci. USA, **101**,15573.
4. *Bugl H., Danner J.P., Molinari R.J., Mulligan J., Roth D.A., Wagner R., Budowle B., Scripp R.M., L. Smith J.A., Steele S.J., Church G., Endy D.* (2006) A Practical Perspective on DNA Synthesis and Biological Security, http://dspace.mit.edu/bitstream/1721.1/40280/l/PPDS.pdf.
5. *Baker D.* (2006) Sci. Am., **294**, 44.
6. *Studier F.W., Hausmann R.* (1969).Virology, **39**, 587.
7. *Sanger F., Nicklen S., Coulson A.R.* (1977) Proc. Natl. Acad. Sci. USA, **74**, 5463.
8. *Carlson R.* (2003) Biosecur. Bioterror., **1**, 203.
9. *Bejerano G, Pheasant M., Makunin I., Stephen S., Kent W.J., Mattick J.S., Haussler D.* (2004) Science, **304**, 1321.
10. *Schneider T.D., Stormo G.D.* (1989) Nucleic Acids Res., **17**, 659.
11. *Murray A.W., Szostak J.W.* (1983) Nature, **306**, 189.
12. *Hutchison C.A., Peterson S.N., Gill S.R., Cline R.T., White O., Fraser C.M., Smith H.O., Venter J.C.* (1999) Science 286, 2165.
13. *Chan L.Y., Kosuri S., Endy D.* (2005) Mol. Syst. Biol., **1**, 2005.0018
14. *Elowitz M.B., Leibler S.* (2000) Nature, **403**, 335.
15. *Bugl H., Danner J.P., Molinari R.J., Mulligan J.T., Park H.O., Reichert B., Roth D.A., Wagner R., Budowle B., Scripp R.M., L. Smith J.A., Steele S.J., Church G., Endy D.* (2007) Nat. Biotechnol., **25**, 627.

По материалам журнала "Nature" и "Science" при участии Рыженковой О.Н.