

ГЕНОМ *robustus* и *sapiens*

Saccharomyces sapiens – под таким замечательным названием появилась статья в Trends in Genetics [1]. Это краткий отчет о конференции Yeast Genetics and Human Disease, состоявшейся в Балтиморе (США) в ноябре 1996 г. Участники следовали моде, заданной известным исследователем геномов Эриком Ландером, и говорили об аналогии Менделеевской периодической системы и полной геномной последовательности. Есть ли смысл в таких сравнениях? В какой-то степени безусловно: периодическая система предсказывает свойства элементов на основании состава их ядра, геномная последовательность предсказывает функции неизвестного участка ДНК на основании его последовательности. Знание свойств 63 элементов, известных во времена Менделеева, позволило построить систему, дающую возможность предсказывать свойства неизвестных элементов. Дмитрий Иванович делал такие предсказания, исходя из свойств атомов, находящихся в той же группе периодической таблицы. Функции неизвестной последовательности из одного организма могут быть предсказаны на основании свойств гомологичной известной последовательности из другого. Но на этом и заканчивается сходство. Оно скорее литературное, чем научное. В строении генов нет того, что является главным в периодической системе, – периодичности, которая в случае атомов проявляется в воспроизведении аналогичных свойств по мере возрастания заряда ядра через определенное число добавленных в него протонов.

Периодичность в свойствах атомов объясняется законами физики, диктующими определенное расположение электронов на электронных оболочках по мере возрастания числа этих электронов при переходе от атома к атому. Ничего подобного нельзя ожидать от структур геномов. Что же, свойственное геному, позволяет проводить аналогию с периодической системой элементов и что из всего этого следует для науки, которая называется Геномикой, т.е. наука о строении геномов? Что из этого следует для биологии вообще и для такого ее раздела, как эволюция, в частности?

Есть воспроизведение аналогичных функций гомологичными последовательностями. И есть некий минимальный необходимый и достаточный набор функций, которым должен обладать независимо живущий организм, чтобы иметь возможность существовать. Такой набор функций должен определяться необходимым комплектом геномных структур, который, возможно, и воспроизводится в любом живом существе. Но в каждом биологическом виде к нему добавляются

новые элементы, определяющие особенности именно этого, а не другого вида, его приспособленность к существованию в его экологической нише.

Отсюда следует простой, даже тривиальный вывод: зная последовательность генома человека и функции всех его компонентов, можно определять функции очень многих наиболее существенных структур у организмов, стоящих на более низких ступенях эволюционной лестницы. Но не наоборот. Вряд ли, к примеру, можно надеяться извлечь из дрожжевого генома информацию относительно генов, вовлеченных в программирование поведенческого комплекса человека. Тем не менее, зная структуры геномов более примитивных организмов, можно надеяться определять функции некоторых очень важных, самых жизненно необходимых генов человека. Например, генов, кодирующих ферменты метаболизма. А повреждения именно подобных генов приводят к весьма неприятным наследственным болезням.

Геномика в процессе своего развития и должна путем сравнительного анализа последовательностей геномов различных организмов определить как минимальный необходимый набор генов, так и те наслоения, которые добавлялись к нему в процессе эволюции и определяли появление новых и новых видов, в том числе и нашего с вами, *Homo sapiens*.

С этой точки зрения все организмы, в том числе и дрожжи, о которых идет речь в упомянутой статье, разумны (*sapiens*), но разумны с точки зрения эволюционно консервативной разумности строения их геномов. Разумны и мы, придя к пониманию, что геномы этих организмов многое нам скажут о нашем собственном геноме.

Возвращаясь в дрожжам, следует вслед за автором статьи [1] сказать, что 1996 год войдет в историю биологии, поскольку в этом году был секвенирован первый эукариотический геном – геном *S. cerevisiae*. Этот успех стал возможным благодаря усилиям международного коллектива, включающего более 600 ученых из разных лабораторий в разных странах мира. Последовательность 12068000 п.о., распределенных между 16 хромосомами, потенциально кодирует 5885 генов белков, большая часть которых охарактеризована как открытые рамки считывания (open reading frames, ORF), около 140 генов рибосомных РНК, 40 генов snРНК и 275 генов транспортных РНК.

Кодирующие последовательности дрожжей занимают около 70% генома и на кодирующую последовательность приходится в среднем 2 т.п.о. геномной ДНК. Дрожжи, таким образом, имеют

один из наиболее компактных геномов среди эукариот: у нематоды на кодирующую последовательность приходится 6 т.п.о., а у человека – 30. Компактность достигается за счет малого числа интронов. Только 4% дрожжевых генов, кодирующих потенциальные белки, содержат интроны. Многие дрожжевые хромосомы содержат чередующиеся GC- и AT-богатые кластеры. При этом наблюдается корреляция повышенного содержания GC с плотностью генов в данной области.

Термин “протеом” стали использовать по аналогии с термином “геном” для описания полного набора белков, которые данный организм способен синтезировать. Дрожжевой протеом содержит 50% белков, гомологичных белкам с известными функциями, описанным для разных организмов. На основании такого сравнительного анализа могут быть выведены закономерности, касающиеся распределения долей протеома дрожжей между различными функциональными действиями клетки (табл. 1).

Очень существенным результатом изучения структуры дрожжевого протеома является обнаружение продуктов, присутствие которых в дрожжах до этого только предполагалось или, наоборот, ставилось под сомнение. Например, гистон H1, наличие которого в дрожжах считалось сомнительным, оказался существующим. Ген гамма-тубулина, который до этого не поддавался идентификации у дрожжей, был найден в результате полного секвенирования его генома.

И тут же становится ясной важнейшая роль, которую структуры генома и протеома дрожжей могут сыграть в исследованиях функций генома и протеома человека. Уже сейчас для почти половины белков, о которых известно, что они вовлечены в возникновение генетических заболеваний у человека, найдены гомологи среди дрожжевых белков. Среди болезней, которые вызываются повреждением соответствующих генов, такие, как наследственный неполипозный рак толстой кишки, нейрофиброматоз, атаксия-телеангиэктазия и синдром Вернера.

О последнем я хотел бы сказать несколько дополнительных слов. Люди, страдающие этим наследственным заболеванием, стареют значительно раньше, чем обычные здоровые люди. В возрасте около 20 лет их волосы седеют, кожа теряет эластичность и зрение портится в результате возникновения катаракт. Хуже того: они часто заболевают раком, сердечно-сосудистыми болезнями и другими, характерными для старых людей болезнями. Большинство умирают в возрасте до 50 лет. Недавно ген *WRN*, участвующий в этом процессе старения, был клонирован большой интернациональной командой ученых [4, 5]. Оказалось, что он кодирует белок, высокомолекулярный одному из важных компонентов репликационного комплекса – геликазе, и поврежден у

Таблица 1

Функция белков	Доля белков с данной функцией от суммы белков протеома, %
Метаболизм	17
Энергетика клетки	3
Репликация, рост и деление клетки	14
Транскрипция	10
Трансляция	5
Структурные белки и др.	49
Внутриклеточный транспорт	5

людей, страдающих синдромом Вернера. Аналогичный ген *SGS* в дрожжах также кодирует геликазу. Его отсутствие приводит к тому, что дрожжевые клетки также проявляют признаки преждевременного старения. Дальнейшее исследование функций *SGS* может пролить свет на процесс нормального старения.

Этот пример показывает, что структурная гомология действительно сопровождается функциональным подобием. Многие гены человека способны замещать дрожжевые гены и выполнять их роль в дрожжах. Сейчас известен по крайней мере 71 человеческий ген, комплементирующий дрожжевые мутации [6]. Нельзя исключить, что большая часть дрожжевых белков имеет человеческие гомологи, и это облегчит идентификацию функций соответствующих продуктов у человека. В настоящее время для 31% ORF дрожжей найдены гомологи среди белков животных. Однако 60% генов дрожжей пока не имеют экспериментально установленных функций.

Наконец, определение полной структуры генома дрожжей имеет важное значение для понимания его эволюции и путей возникновения новых генов, обеспечивающих новые функции организма.

Структура генома – это лишь первый шаг на пути понимания механизмов его функционирования. В европейских и других странах созданы новые консорциумы для систематического анализа функций ORF генома дрожжей, представляющих собой возможные новые гены. Естественно, что при этом предполагается направленное делетирование определенных ORF с помощью методики, основанной на высокой эффективности и точности митотической рекомбинации у дрожжей. Планируется получить полный набор таких моногенно-делементированных дрожжевых клеток. Исследование фенотипических проявлений последствий этих делеций должно дать богатый материал для функциональных выводов.

Как уже упоминалось, сейчас из 5885 известных и предсказанных дрожжевых генов только 40% имеют известные функции. Систематическое нокаутирование генов на хромосоме VIII (нокаутировано 266 из 267 ORF) показало, что менее чем

20% этих потенциальных генов важны, и, следовательно, даже такой примитивный организм, как дрожжи, уже сильно удалился от того минимального функционального “ядра”, о котором я говорил выше. Интересно было бы провести систематическое нокаутирование “микробного минималиста” – микоплазмы [3], которая обходится 482 генами.

Систематическое исследование влияния изменения структуры генов на их функции, пример которого дает нокаутирование ORF у дрожжей, открывает новую стадию познания геномов, которая может быть названа функциональной геномикой, в отличие от предыдущего этапа, осуществлявшего чисто структурный анализ. Полагаю, что функциональная геномика не будет ограничиваться исследованиями только ORF. Геном содержит массу функциональных элементов, которые важны для регуляции его экспрессии и репликации. Все они должны идентифицироваться, и это открывает совершенно новую страницу в изучении геномов – интегральную функциональную геномику, которая должна ставить своей целью познание всего ансамбля регуляторных процессов, осуществляющихся в организме под контролем генома.

Эта грандиозная задача немислима без развития новых технологий исследования как в секвенировании, так и в особенности в исследованиях функций.

Что касается секвенирования, то авторы статьи [2] дают следующие выкладки относительно зависимости стратегии полногеномного секвенирования от длины секвенируемого генома.

При длине генома менее 2×10^6 п.о. может быть эффективно и экономично использована стратегия секвенирования методом “шот-ган”, т.е. клонирование непосредственно рестриктных фрагментов генома в подходящий вектор, секвенирование множества случайно выбираемых клонированных последовательностей и сборка полной структуры на основании перекрывающихся последовательностей секвенированных фрагментов. Так были секвенированы геномы микоплазм и нескольких бактерий, включая *H. influenzae*.

Для геномов размером между 2×10^6 и 6×10^6 п.о. секвенирование становится значительно более сложным и дорогим. Здесь важное значение приобретает предварительное создание библиотек перекрывающихся последовательностей, использование дистанционной PCR-амплификации (long-range PCR) или прямого PCR-секвенирования.

Секвенирование геномов, превышающих по размеру 6×10^6 п.о., обычно требует создания библиотек клонов в космидах или других векторах с высокой емкостью и трудоемкой работы по заполнению пробелов в покрытии генома такими клонами. Планы по секвенированию больших геномов почти всегда недооценивают затраты на эти стадии.

Что касается секвенирования генома дрожжей, то, по мнению авторов работы [2], ключевым мо-

ментом для их успеха было существование готовых геномных библиотек, практически не содержащих пробелов в клонированной геномной ДНК.

Здесь необходимо дать комментарий. В последнее время часто высказывается мнение, что следует резко сократить затраты на разработку новых методов секвенирования и основные средства, выделяемые на проект “Геном человека”, сосредоточить на секвенировании уже известным способом, который дает возможность осуществить в разумные сроки сиквенс генома человека. Эта в целом очень рациональная точка зрения тем не менее, на мой взгляд, не должна восприниматься как сигнал к полному прекращению разработок новых технологий. Скорее всего, работающий сегодня подход даст возможность определить существенную часть последовательности генома человека, но оставит также и значительное число “дырок”, заполнение которых будет невозможным без новых методов секвенирования. Выкладки, приведенные выше, кажутся весьма весомым доводом в пользу этого мнения. Нужны методы систематического последовательного секвенирования длинных фрагментов ДНК, которые дали бы возможность избежать потерь фрагментов генома при субклонировании громадного количества коротких фрагментов, неизбежных при современных методах.

Среди задач функциональной геномики одной из важнейших является идентификация спектров транскрипции генов в различных клетках организма. Традиционно для этой цели развивались такие методы, как норзерн-блот-гибридизация или RT-PCR (амплификация мРНК после ее обратной транскрипции). Оба эти метода предполагают анализ мРНК с помощью имеющихся гибридационных проб, специфичных для данной отдельной мРНК, или пар праймеров, опять-таки специфичных для данной отдельной мРНК. Такие виды анализа присутствия по отдельности каждой из исследуемых мРНК в данном типе клеток абсолютно неприемлемы по причине их низкой скорости. Необходим интегральный анализ присутствия всех возможных мРНК в клетке.

Современные возможности химических методов исследования нуклеиновых кислот открывают такую перспективу. Речь идет об использовании ДНК-“чипов” – микропластинок, содержащих на поверхности десятки тысяч идентификационных нуклеотидных последовательностей, характерных для отдельных индивидуальных мРНК. Гибридизация такого типа с мечеными (радиоактивно или флуоресцентно) кДНК или мРНК, выделенными из данного типа клеток, позволяет сразу определять весь набор мРНК, присутствующих в этих клетках [7]. Если эта технология окажется работающей, то она вытеснит современные методы анализа тканеспецифичности экспрессии генома.

Таблица 2

Организм	Размер генома, 10 ⁶ п. о.	Предположительное количество генов	Идентифицировано генов	
			число	%
<i>S. cerevisiae</i>	12.1	6034	3086	63
<i>E. coli</i>	4.6	4288	2656	62
<i>B. subtilis</i>	4.2	4000	2320	58
<i>Synechocystis</i> sp	3.6	3168	1402	48
<i>Archaeoglobus fulgidus</i>	2.2	2471	1193	44
<i>H. influenzae</i>	1.8	1740	1015	58
<i>Methanobacterium thermoautotrophicum</i>	1.8	1855	816	44
<i>Helicobacter pylori</i>	1.7	1590	907	57
<i>Methanococcus jannaschii</i>	1.7	1692	776	46
<i>Borrelia burgdorferi</i>	1.3	863	499	58
<i>Mycoplasma pneumoniae</i>	0.8	677	333	49
<i>M. genitalium</i>	0.6	470	324	69

Нужно заметить, что идея использования иммобилизованных на твердой поверхности олигонуклеотидов для секвенирования нуклеиновых кислот методом гибридизации секвенируемой ДНК с этой поверхностью была предложена А.Д. Мирзабековым с соавторами. Развитие этой идеи для анализа содержания мРНК в клетках представляется очень плодотворным.

Понятно, что экспоненциально накапливающаяся информация требует новых подходов к ее хранению и обработке. Это невозможно без развития компьютерных, легко доступных баз данных. Но это отдельная история.

Заканчивая, следует заметить, что сейчас секвенируется очень большое количество геномов [8–10]. Только что появилась информация о сиквенсе генома классика молекулярной генетики — *E. coli*, который состоит из 4638858 п.о. и содержит 4286 ORF. Это сделано в США. В Японии секвенирован другой штамм *E. coli*. Фирма Human Genome Science секвенировала третий, патогенный штамм. От 6 до 12 бактериальных патогенов сейчас в работе. Этот список включает в себя возбудителей различных оппортунистических инфекций, менингита, туберкулеза, пневмонии, гонорей, сифилиса, холеры, тифа и некоторых других болезней. Начато исследование малярийного паразита *Plasmodium falciparum*, длина генома которого около 30 млн. п.о. В прошлом году секвенирован геном *Staphylococcus aureus*. Изучаются геномы также и непатогенных бактерий, в частности археобактерий. Некоторые из них способны продуцировать очень полезные вещества. Например, один из секвенируемых в настоящее время микробов, *Clostridium acetobutylicum*, синтезирует ацетон и бутанол. Успешно изучается геном нематоды, обсуждается секвенирование генома арабидопсиса и множества других модельных организмов.

Сводка организмов, полный геном которых уже секвенирован, приведена в табл. 2 [10].

Здесь есть пространство для научно-экономических соображений. Например, авторы статьи [2] указывают, что с накопленным с *S. cerevisiae* опытом достаточно несложно секвенировать геномы других дрожжей, в частности имеющих промышленное или медицинское значение. В то же время они полагают, что ввиду усилий, которые требует геном человека, возможно, нецелесообразно распылять силы и внимание и секвенировать слишком много модельных геномов. Вместо этого более полезно секвенировать геномы наиболее важных паразитов, которые поражают миллионы людей, особенно в развивающихся странах. Есть в этой точке зрения своя правда. Ресурсы наше безграничны, и мы должны все больше думать, как их расходовать наиболее рационально с точки зрения увеличения шансов на выживание наших потомков в условиях стремительного роста человеческой популяции и надвигающегося экологического кризиса.

СПИСОК ЛИТЕРАТУРЫ

1. Richards W. *Saccharomyces sapiens* // Trends Genetics. 1997. V. 13. P. 49–50.
2. Goffeau A. et al. Life with 6000 Genes // Science. 1996. V. 274. P. 546–567.
3. Goffeau A. Life With 482 Genes // Science. 1995. V. 270. P. 445–446.
4. Chang-En Yu et al. Positional Cloning of the Werner's Syndrome Gene // Science. 1996. V. 272. P. 258–262.
5. Aging Gene Discovered // Science. 1996. V. 272. P. 193–194.
6. Botstein D., Chervitz S.A., Cherry J.M. Yeast as a Model Organism // Science. 1997. V. 277. № 5330. P. 1259–1260.
7. To Affinity... and Beyond! (Editorial) // Nature Genetics. 1997. V. 14. № 4. P. 367–370.
8. Focus Moves from Transcriptional Mapping to Gene-Functions Studies // Human Genome News. 1996. V. 8. № 2. P. 4–6.
9. Fox J.L. Microbial Genomics: Milestones Mount Exponentially // Nature Biotechnol. 1997. V. 15. P. 211–212.
10. Pennisi E. Laboratory Workhorse Decoded // Science. 1997. V. 277. P. 1432–1434.

Е.Д. Сverdlov, ИБХ РАН, Москва