

*Академик РАН Е.Д. Свердлов (с 1992 г. по 2011 г. – главный редактор журнала)*

## МОЛЕКУЛЯРНАЯ ГЕНЕТИКА СЕГОДНЯ. ГРОМАДНЫЕ УСПЕХИ, ТЯЖЕЛЫЕ ПРОБЛЕМЫ, БОЛЬШИЕ НАДЕЖДЫ. ОБЗОР, ПОСВЯЩЕННЫЙ 30-ЛЕТИЮ ЖУРНАЛА «МОЛЕКУЛЯРНАЯ ГЕНЕТИКА, МИКРОБИОЛОГИЯ И ВИРУСОЛОГИЯ»

Обзор посвящен 30-летию журнала «Молекулярная генетика, микробиология и вирусология» и итогам 10-летия с момента объявления о завершении секвенирования генома человека и новой постгеномной эпохи в молекулярной генетике. Обзор состоит из трех частей: 1) громадные успехи молекулярной генетики – они действительно громадны в области структурной геномики; 2) тяжелые проблемы – они невероятно сложны в области анализа жизнедеятельности и 3) большие надежды – мы должны надеяться, но наши надежды не должны оказываться пустыми фантазиями, которые никогда не осуществятся.

Журналу «Молекулярная генетика, микробиология вирусология» исполняется 30 лет. В этом же году исполняется 10 лет с момента объявления о завершении секвенирования генома человека (<http://www.hagsc.org/projects/human>), хотя соответствующая публикация появилась более года спустя [1], а черновой вариант последовательности был опубликован двумя годами раньше [2, 3].

Началась новая эпоха – постгеномная, итоги которой подвел Эрик Ландер (Eric Lander) в своем очень широком обзоре [4]. Прошли 10 лет сложной интенсивной работы, итоги которой только сейчас начинают в полной мере осмысливаться. Эпоха была наполнена технологическими революциями, особенно в области секвенирования геномов, транскриптомов и отчасти протеомов, и она характеризуется также серьезными разочарованиями относительно достижимости тех целей, которые в конечном счете ставит перед собой секвенирование – понимание функций, которые характеризуют жизнедеятельность, и их нарушений при различных патологиях. Существует громадный разрыв между структурными успехами и функциональными результатами, и это сказывается, в частности, на практических применениях результатов структурных исследований.

Я разделяю обзор на три части: громадные успехи молекулярной генетики – они действительно громадны в области структурной геномики; тяжелые проблемы – они невероятно сложны в области анализа жизнедеятельности, и большие надежды – мы должны надеяться, но наши надежды не должны оказываться пустыми фантазиями, которые никогда не осуществляются. Такие фантазии подрывают авторитет науки и, к сожалению, это произошло в области геномики, где обещания исследователей сильно превысили реальные итоги.

### Громадные успехи молекулярной генетики

Наверное, лучше всего начать этот раздел с резюме опроса, который провела редакция журнала *Nature Reviews Genetics* в октябре 2010 г., по случаю десятилетия журнала [5].

К опросу привлекли 8 заметных в нашей науке профессионалов. Их спросили, какими они видят основные достижения генетики и геномики за последние 10 лет и перспективы этих отраслей знания на будущее. Результаты были названы «Точки зрения:

десять лет генетики и геномики: Чего мы достигли и куда идем?»

Давайте познакомимся с этим спектром мнений. Вот их краткое резюме.

1. Эпигенетика вышла в лидеры в течение последнего десятилетия.

2. Одним из наиболее впечатляющих успехов было создание недорогой, высокопроизводительной методологии для изучения изменений в масштабе генома человека.

3. Это была декада улучшения понимания в исследовании генетики обычных, многофакторных заболеваний.

4. Неожиданностью последнего десятилетия было открытие абсолютно параллельного мира малых (и больших) РНК, не кодирующих белки, и их роли в регуляции экспрессии протеин-кодирующих генов

5. Вопрос, возникающий из систематического, непредвзятого (*независимого от гипотез*) объективного, высокопроизводительного анализа, заключается в том, насколько более сложной является молекулярная организация клетки относительно той, которая первоначально вырисовывалась при использовании фокусированных на отдельных генах и основанных на гипотезах (*и, следовательно, предвзятых*) редуционистских методов молекулярной биологии.

6. Достижения, полученные в исследованиях, первоначальной целью которых было «обнаружение и понимание всех генов», в какой-то степени подрывают саму концепцию термина «ген». Это особенно важно, поскольку ген часто рассматривают в качестве фундаментальной биологической единицы.

7. Мощные, далеко нацеленные догмы, возникшие на заре молекулярной биологии, такие как гипотеза Бидла и Татума (Beadle и Tatum) «один ген, один фермент, одна функция» или центральная догма Крика «ДНК создает РНК создает белок» — не могли разлететься как дым. Исследования были направлены на то, что могло бы занять их место, или, по крайней мере, дополнить эти мощные идеи для улучшения нашего понимания биологии в последней декаде и последующие годы.

8. Геномика, хотя представляет собой описательную дисциплину, революционизировала наше понимание природы эволюционных изменений неожиданным образом.

9. В течение 10 лет, следующих за объявлением о завершении проекта генома человека, произошло огромное количество революционных инноваций в области геномики, целью которых был ответ на вопрос: «является ли расшифровка генома полезной?» То есть сможет ли геномика помочь улучшить наше здоровье и нашу повседневную жизнь? Хотя все еще нет консенсуса, ученые выполнили множество фундаментальных проектов, которые показывают, что ответ, вероятно, будет «Да».

10. Наиболее впечатляющие успехи были достигнуты в области перепрограммирования клеток с помощью определенных факторов транскрипции. Способность к перепрограммированию клеток также улучшила наше понимание того, как контролируются состояние и дифференцировка клеток. Опыт перепрограммирования подтверждает, что как эмбриональные, так и дифференцированные клетки зависят от небольшого набора ключевых факторов транскрипции для установления и поддержания состояния клетки.

К этому следует добавить важнейшее событие, которое произошло уже после публикации этой статьи. Появилось 30 статей в «Nature», «Genome Research» and «Genome Biology» и, в частности, целый номер журнала «Nature» (см. вводные статьи [6-9]), где опубликованы итоги работы международного консорциума ENCODE (Encyclopedia of DNA Elements).

ENCODE был организован, чтобы определить положение всех функциональных элементов в ДНК и придать им функциональный смысл. Быстро стало понятным, что структурная информация слишком скудна для того, чтобы делать функциональные выводы [10]. Исследователи смогли идентифицировать среди 3 миллиардов пар оснований генома человека примерно 1% последовательностей, кодирующих белки. Этого, конечно, недостаточно для понимания функций, и ENCODE ставил задачу найти функциональные регуляторные элементы в межгенном пространстве, точнее, в некодирующих областях генома человека. Выяснить, где и в каких клетках эти элементы активны, и как они связаны с упаковкой генома в хроматин.

Консорциум приписал потенциальную функциональную роль примерно 80% генома. При этом выявилось более чем 70 000 потенциальных промоторов и почти 400 000 потенциальных энхансеров. Это много больше, чем 20 000 генов, кодирующих белки. Показано, как эти элементы связаны с модификациями и открытостью хроматина. Однако, это пока трудно интерпретируемые результаты. Работа, несомненно, будет продолжаться, но пока она лишней раз подчеркивает громадный разрыв между структурной (очень обильной и важной) и функциональной информацией. По мнению авторов, даже только картирование этих элементов, т.е. определение их положения в геноме, сделано только наполовину, а полная характеристика всего, что геном делает, выполнена только на 10% [6]. Я думаю, что это сильно преувеличенная оценка.

То, что прошло только что перед вами – это спектр мнений людей, смотрящих на жизнь в основном с точки зрения геномики и других -омик. Авторы опроса охватили многие проблемы и сделали важные предсказания, которые я покажу в последней части обзора. Но многое осталось за скобками их анализов. Это естественно, и поэтому я продолжу цитированием (сокращенным) из интервью Е. Green [11], нынешнего директора Института исследований генома (National Human Genome Research Institute, NHGRI). Грин тоже смотрит на жизнь через то же окно генома. Интервью дано в связи с публикацией 10-летнего плана этого Института в «Nature» в феврале 2011 г. Этот план фокусируется непосредственно на клинических целях. Важно, что в отличие от далеко идущих обещаний, которые давались после завершения струк-

туры первого генома человека, на этот раз лидеры Института избегают чрезмерных обязательств. Они предсказывают, что эффект от новых технологий будет наблюдаться только к 2020 г. Чтобы достичь этого эффекта, придется создать каталоги, начиная от указателя генов, вовлеченных в возникновение редких и обычных болезней, до каталога геномов микроорганизмов, обитающих в теле человека, и до систематического перечисления мутаций в опухолях.

Итак, вот фрагмент этого интервью:

Вопрос: Что изменилось за прошедшие 8 лет, что дало возможность сформулировать новый план?

Green: Вероятно, наиболее впечатляющим являются достижения в технологиях. Восемь лет назад мы не ожидали, что цена секвенирования упадет так сильно. Понимание геномных основ болезней чрезвычайно упростилось благодаря этому. Мне кажется, что теперь мы можем действительно серьезно заняться этой проблемой.

Вопрос: Почему вы поставили 2020 г. как срок, когда геномика затронет проблемы здравоохранения? Почему это займет так много времени?

Green: Когда мы говорим с людьми, которые имеют представление об истории медицинских достижений, они обращают внимание, что реальное изменение медицинских технологий занимает большое время. Часто десятилетия. Я чувствителен к критицизму геномики теми, кто уверен, что с 2003 г. (*это год завершения секвенирования генома человека* – прим. Е. Д. Свердлов) мы не улучшили в достаточной степени здравоохранение. Поэтому этот срок отражает наше стремление быть немного более реалистичными и осторожными.

Вопрос: Вы думаете, что в 2003 г. ученые действительно верили, что их обещания будут выполнены к сегодняшнему дню, или общественность их неправильно понимала?

Green: ... Было большое количество преувеличений и энтузиазма... Может быть, были сказаны какие-то слова, что следующие 5–10 лет полностью изменят лицо медицины. Это, вероятно, было преувеличением. Знали ли мы тогда, что появится вся эта новая информация про некодирующую ДНК? Представляли ли, что выявится такая громадная сложность в структурных изменениях генома, связанных с болезнями? Чем больше мы узнаем, тем все это оказывается сложнее.

Вопрос: Где, по вашему мнению, мы будем перед 2020 г.?

Green: Я надеюсь, что к этому времени мы будем иметь громадную информацию про то, как генетические варианты играют роль в развитии болезней, и что это создаст всецело новое пространство для реального осмысления того, как предсказывать болезни, как возможно их предотвращать и, конечно, как лечить.

Вот так выглядят мнения тех, кто работает в области геномики и других -омик. Давайте попробуем посмотреть несколько более широко. Вот продвинулся за пределы генома известный генетик Aravinda Chakravarti и написал редакторскую статью в «Science» 7 октября 2011 г. по поводу предстоящего тогда конгресса генетиков человека в Монреале. Она называется: «Геномика недостаточно»[12].

### Тяжелые проблемы молекулярной генетики

«Ключевым условием использования геномики, в частности, в медицине является понимание биологических механизмов. Перенос современного знания в медицинскую практику – это важная цель для общественности, которая поддерживает медицинские исследования, и, конечно, для ученых и клиницистов. Тем не менее, несмотря на многочисленные открытия новых генов, главным препятствием является от-

существование понимания того, как эти гены влияют на фундаментальные биологические механизмы, которые нарушаются при болезни. И если мы хотим, чтобы медицина, основанная на достижениях геномики, процветала, мы должны заполнить эту зияющую дыру.

Достижения в биохимических исследованиях индуцировали большие ожидания относительно переноса этих результатов в медицинские технологии, включая индивидуализацию лечения и профилактики. Концепция индивидуальной медицины не нова в генетике. Идентификация многочисленных врожденных эффектов метаболизма и открытие ассоциированных с ними дефицитов ферментов проложили путь к их генетическому диагнозу и лечению. Тем не менее, ключевым здесь было понимание биологических механизмов.

Сегодня геномные технологии могут использоваться для рутинного сканирования в поиске генетических изменений при любой болезни. Более чем 2000 моногенных Менделевских болезней было исследовано этим путем. Предвидится нахождение изменений, которые вызывают остальные 2000 таких болезней. Но несмотря на многочисленные попытки, достижение подобного понимания обычных хронических сложных болезней остается пока невозможным. Здесь, чтобы добиться медицинской пользы, биомедицинские исследования должны двинуться дальше простых открытий генов путем картирования, секвенирования или изучения полногеномных ассоциаций (ПГА, genome-wide association study, GWAS). Фокус должен быть нацелен на понимание механизмов болезней. Мы должны знать, не только какой из вариантов ДНК какого гена ведет к болезни, но и то, как он это делает.

Уроки, извлекаемые из геномной биологии, вполне понятны. Гены и их продукты почти никогда не действуют в одиночку. Они работают в системе с другими генами и белками и в зависимости от окружения. При этом дефект в активности одного гена не обязательно повреждает всю систему. Это согласуется с наблюдениями, что большинство признаков вовлекает множество генов. Более того, вариации регуляторных систем генов происходят значительно чаще, чем изменения в структуре генов и их продуктов. Геномная биология теперь должна двигаться к клеточной биологии и физиологии (системная биология), чтобы понять, как генетические пертурбации приводят к дисрегуляции белков и их сетевых взаимосвязей (networks), а также клеток при болезнях.

Наше прогрессирующее знание генетических вариаций усложняет это понимание. Каждый индивидуум уникален по структуре своего генома. Вопрос заключается в том, являются ли биологические свойства индивидуумов и их болезни также уникальными? Или громадное разнообразие последовательностей в любой болезни сводится к меньшему набору обычных (фенотипических) индивидуальных недостатков? Фокусируясь на понимании механизмов болезни, геномная наука может дать конкретные предложения о том, когда медицинская помощь должна быть индивидуализированной, а когда универсальной. Ответы на эти вопросы будут эволюционировать по мере того, как эволюционирует наука».

Высказанные выдающимся генетиком общие мысли чрезвычайно правильно и актуальны. Нужно искать максимально общие подходы в медицине, которые не зависели бы от громадных индивидуальных различий. Иначе медицина задохнется от необходимости (утрированно говоря) производить сотни миллионов лекарств, рассчитанных на разные болезни 7 миллиардов жителей Земли. При всей их правильности, они политкорректны – автор аккуратно обходит очень острые вопросы современного состояния нашей науки, которое приводит многих к мысли о ее кризисном состоянии.

Здесь я сделаю небольшое отступление. Меня ча-

сто упрекают в том, что я слишком широко использую цитирование. Я это делаю для того, чтобы был понятен точный спектр мнений людей, формирующих лицо современной науки, а не их произвольные толкования. Свои толкования я также даю и выражаю свою позицию по поводу разных точек зрения.

Теперь я продолжу цитирование. На этот раз это будет цитирование Bruce Alberts, выдающегося биохимика, главного редактора журнала «Science», а перед этим президента Национальной Академии наук США. Он один из авторов фундаментального труда «Молекулярная биология клетки». Он очень широко смотрит на проблему перспектив развития молекулярной генетики.

Вот что он пишет 28 сентября 2012 г. в редакционной статье «The end of “small science”?» в журнале «Science» “[13].

«Эту редакционную статью меня побудил написать выпуск в этом месяце 30 статей от Encode Project Consortium. В этом продолжающемся уже 10 лет проекте принимает участие международная команда из 442 ученых, которые составили то, что называется «Энциклопедия ДНК-элементов»: исчерпывающий перечень функциональных элементов в геноме человека. Этот каталог, как полагают, стимулирует дальнейшие исследования по основам жизни, здоровья и болезни. ENCODE является примером «большой науки» – стиля исследований, который привлекает внимание и продуцирует данные с очень высокой эффективностью. Означает ли это, что весьма успешная эпоха “малой науки” биологических исследований скоро закончится? Будет ли государственное финансирование все более способствовать большим научным проектам? Я, конечно, надеюсь, что ответ будет отрицательным.

Поразительный успех Human Genome Project, который завершился в 2004 г. публикацией почти полной последовательности более 3 миллиардов нуклеотидов ДНК в хромосомах человека, стимулировал распространение других крупных “-омик”-проектов в области биологии, в том числе протеомики, транскриптомики, эпигеномики и метаболомики. Каждое из этих крупномасштабных проектов приводит к развитию новых ценных методологий, которые необходимы для того, чтобы обеспечить масштабность этого типа исследований. Но масштаб создает также группу поддержки таких проектов, которая препятствует их остановке даже при наличии признаков их убывающей результативности. В настоящее время, когда ситуация с ресурсами очень напряжена, все более критичным становится принять объективные жесткие решения о том, какие проекты имеют лучшие шансы на получение результатов, необходимых для глубокого понимания, а не просто для описания биологических систем.

Что еще предстоит понять? Как соавтор обновляемого каждые 5 лет учебника по биологии клетки, я болезненно ощутил огромный пробел, который остается в нашем понимании даже простейших клеток. Рассмотрим, например, бактерию *E. coli*, которая служила основным модельным организмом в первые годы молекулярной биологии. Очень открывает знание того, что больше чем 50 лет спустя, почти четверть из более чем 4000 белков, кодируемых её геномом, имеют функции, которые остаются неизвестными. Могут ли быть обнаружены некоторые новые функциональные классы биологических молекул общих для всех клеток, если мы сфокусируемся на исследовании таких белков?

Второй пример: типичная человеческая клетка содержит около 10 000 различных белков, организованных в сотни различных комплексов, которые функционируют как «белковые машины», активирующиеся где и когда это необходимо, для выполнения определенных функций, таких как репарация

ДНК или интеграции различных сигналов. Чтобы разобраться в биологии и получить возможные выгоды для здравоохранения от такого понимания, каждый из этих белковых комплексов должен быть детально изучен биохимиками в работах, которые обычно проводятся в небольших лабораториях.

И, наконец, мы теперь знаем, что самые интересные свойства клетки – это “возникающие” (emergent) свойства, появляющиеся в результате сложной сети взаимодействий между различными молекулами, которые включены в белковые машины. В настоящее время ученые не имеют возможности расшифровать эту сложность, и, чтобы такая возможность появилась, потребуются большая изобретательность и многие новые разработки, причем мы не можем предсказать даже, какого рода они будут. Большая часть работы должна быть сделана через научные исследования относительно простых систем, таких как бактерия кишечной палочки, небольшими коллективами с надеждой на то, что полученные результаты приведут к новым подходам и принципам, которые могут быть перенесены на более сложные клетки млекопитающих.

Каждый год количество фактической информации, получаемой учеными о клетках, увеличивается, и стимулированные «омик»-проектами каталоги данных расширяются с огромной скоростью. Но грандиозные проблемы, которые остаются на пути глубокого понимания химии жизни, требуют выхода за рамки подробных каталогов. Обеспечение успешного будущего для биологических наук потребует ограничений в росте числа крупных центров и «омик»-подобных проектов с тем, чтобы обеспечить большую финансовую поддержку для работы малых инновационных лабораторий, стремящихся понять удивительную сложность живых систем».

Это крайне важное философское осмысление правильности выбора направлений развития науки с тем, чтобы они были максимально информативны и полезны. Вообще говоря, могут возразить – в науке трудно предвидеть правильное направление. Но все-таки наступает момент, когда становится ясным, какое направление с высокой вероятностью неправильно. С тем, что crucialным в данный момент является функциональный анализ, согласен даже такой выдающийся лидер геномных проектов, как E. Lander [4].

Я постараюсь усилить точку зрения Alberts дальше, процитировав и прокомментировав Нобелевского лауреата Robert Weinberg [14].

Его точка зрения под названием «Гипотезы сначала» (Point: Hypotheses first) опубликована в «Nature» 1 апреля 2010 г. Я приведу ее в сокращенном варианте.

«Мало есть такого, что можно было бы продемонстрировать за все время и деньги, вложенные в геномные исследования рака, и этот подход подрывает испытанные пути исследования и организации науки.

Исследования, основанные на выдвижении гипотезы и ее проверке (hypothesis-driven research), хорошо служили нам последние полстолетия и привели к революциям в молекулярной и клеточной биологии, иммунологии, нейробиологии и генетике. Наше понимание механизмов патогенеза превзошли самые дерзкие спекуляции, которые имели место 50 лет назад.

Теперь подобные исследования под угрозой. Многие полагают, что традиционные средства не могут расшифровать грандиозную сложность, которая позволяет преуспевать отдельной клетке и сложному организму. Полагают, что недавно появившиеся технологические инновации породили альтернативный путь расшифровки. Мои студенты могут собрать определенные типы данных в 1000 и даже в 10 000 раз быстрее, чем я мог это сделать 40 лет назад.

В исследованиях рака новые технологии обещают из-

менить пути диагностики, терапии и понимания патогенеза болезни. Так что же, старые пути научной работы – тестирование гипотез стало анахронизмом? Я думаю – нет.

Эра новой биологии – геномики протеомики и метабономики – началась с секвенирования генома человека. Успех этого проекта бесспорен: десятки тысяч исследовательских программ, многие из которых сфокусированы на идентификации и характеристике определенных генов, выиграли очень сильно от создания и изучения этой базы данных.

Крупномасштабные проекты, такие как «Геном человека», выдают за будущее и центральную часть системной биологии, которая появилась недавно. Увеличивающаяся часть национального исследовательского бюджета направляется на эти цели. Но стоит ли уничтожать 20 или 30 маломасштабных проектов, основанных на проверке гипотез, ради того, чтобы освободить место исследованиям на системном уровне?

С точки зрения исследований рака (*R. Weinberg – признанный авторитет в этой области*) успехи гипотез-направляемых исследований ясны и бесспорны. Они продолжают даваться, начиная с более чем полувек назад, и продолжают давать новые концептуальные результаты неделя за неделей, месяц за месяцем. Напротив, новые пути развития биологии настолько неопробованы, что их долговременные преимущества еще трудно предвидеть....

Секвенирование полных геномов опухолей (или их экзонов) потребляет огромные ресурсы и энергии исследователей. Но дивиденды на сегодняшний день скромны: открытие нескольких новых онкогенов и генов опухолевых супрессоров... и определение степени нестабильности различных геномов опухолевых клеток. Эти массивные проекты по производству данных еще только намереваются дать понятный консенсус относительно того, как много мутаций требуется, чтобы породить человеческую опухоль, и понимание того, как развивается индивидуальная опухоль. Цена каждого концептуального понимания была очень высокой, хотя это может измениться с падением стоимости технологии. Тем временем много меньше экспериментальные исследовательские программы, которые являются проверенным источником концептуальных инноваций, борются за выживание.

Вероятно, наиболее амбициозные крупномасштабные мероприятия включают в себя сборку многих взаимодействующих сигнальных компонентов в клетке в «электромонтажные диаграммы (wiring diagrams)». Эти сложные карты, иногда называемые волосатыми шарами (hairballs), и компьютерные алгоритмы, которые моделируют сигнальные процессы, могли бы пролить свет на то, почему и как индивидуальные клетки откликаются на внешние сигналы, и предсказать их будущее поведение. Будучи эстетически приятными, они внесли малый вклад в концептуальное понимание причин того или иного поведения клетки. Некоторые полагают, что *детальное понимание отдельных компонентов сигнальных систем* является пререквизитом для предсказания поведения целых сигнальных схем.

Ставки здесь велики. Последствия сдвига финансирования, предоставляемого основными финансовыми агентствами, будут ощущаться последующими поколениями ученых. Руководство лабораториями, фокусирующимися на маломасштабных проектах, движимых гипотезами, стало непривлекательным для многих молодых ученых из-за громадных трудностей в получении денег на такие проекты. Долговременными последствиями этого будет неспособность многих биологических дисциплин привлечь ярких молодых людей, которые являются двигателями научного прогресса. Без них – мы пропали».

Нужно отметить, что в этом же журнале рядом с R. Weinberg фигурирует мнение Todd Golub [15], которое идет под заголовком «Counterpoint: Data First». Я позволю себе не цитировать его. Оно понятно из

заголовка и на данный момент является мнением тех ученых, которые коллекционируют данные и пытаются на их основании делать функциональные выводы. На мой взгляд, R. Weinberg корректно оценил успехи этого направления на сегодняшний день.

Наконец, не могу удержаться от соблазна привести высказывание еще одного Нобелевского лауреата Sydney Brenner, одного из самых ярких и оригинальных исследователей современности. Вот что он сказал, выступая в 2008 г. в Индии в Бангалоре (<http://www.genomeweb.com/blog/data-substitute-thinking>).

«Молодые ученые не должны соблазняться системной биологией, где данные являются заменителями мышления... Существует кризис во всех науках в наши дни. Мы тонем в море данных, и все-таки мы томимся от жажды. Сегодня наука вознаграждает только тех, кто коллекционирует и распределяет данные. Нет награды тем, кто организует данные и строит на основе теории».

Я привел эти цитаты, чтобы дать почувствовать накал страстей (которые проходят мимо нас, поскольку у нас другие денежные масштабы и, соответственно, гораздо более узкий горизонт для анализа).

Многие подумают, что это просто результат борьбы поколений. R. Weinberg, B. Alberts – люди старого поколения. E. Green и другие геномные апологеты – молодое поколение. Всегда люди старого поколения, даже великие умы, воевали против новых идей и открытий, сопротивлялись новшествам: Лорд Кельвин – против расщепления атомов и электромагнитной теории света. Дэви против Дальтона, выдвинувшего закон кратности отношений. Гюйгенс и Лейбниц против теории тяготения Ньютона. Рентген запрещал употреблять в своей лаборатории слово «электрон». Кьюве, Пастер, Virchow, К. Бернар – против теории эволюции Дарвина. Изобретатель паровой машины Д. Уатт, выступая против нового варианта паровой машины, изобретенной его соотечественником Тревитиком, говорил, что это идет во вред прогрессу, и что Тревитика мало повесить за его изобретение.

Но в данном случае, на мой взгляд, дело обстоит гораздо хуже. Новые массивные подходы могут иметь принципиальный барьер, который невозможно преодолеть. Я думаю, что мы сильно недооцениваем степень этой неприятности. И я попытаюсь объяснить почему. Опять начну с цитирования.

В «Nature» 26 июля 2012 г. появились две небольшие статьи [16, 17], которые, тем не менее, выражают, возможно, в первый раз на таком форуме, как самый престижный научный журнал, тревогу не относительно того, что крупные проекты дают мало информации, поглощая огромные средства, а относительно качества получаемых при этом данных. Статья Daniel MacArthur называется «Methods: Face up to false positives». Это я бы перевел как «Встречайте ложные позитивы».

Вот какая там тревога:

«Когда исследование геномов долгожителей привело к сообщению о том, что существуют генетические варианты, сильно ассоциированные с исключительным долгожительством, это привлекло внимание средств информации и научной общественности. Это сообщение вызвало также скептицизм у генетиков. То, что индивидуальные генетические варианты могут иметь такое большое влияние на такой сложный признак, как долгожительство, было абсолютно нежиданным. Как оказалось, по крайней мере некоторые из этих результатов были неожиданными просто потому, что они были ошибочными. В отзыве этой статьи в журнале, который был опубликован годом позже, авторы признали техническую ошибку и неадекватный протокол контроля качества...»

Несколько принципов хорошо знакомы ветеранам науки: один из них – чем более неожиданным является результат, тем менее вероятно, что он правилен. Мы не можем знать, был ли этот принцип рассмотрен в каком-либо исследовании. Тем не менее, более обычно в мире, в котором неожиданные результаты могут приводить к публикациям в престижных журналах, аплодисментам и передовицам в «The New York Times», легко понять, что возможен всепоглощающий соблазн двинуться от открытия к публикации без должной проверки результатов.

В действительности, никогда не было так легко получать высокоимпактные ложнопозитивные результаты, как в геномную эру, когда массивные сложные биологические данные дешевы и легко доступны. Для ясности: большинство экспериментов в масштабах полного генома дает реальные результаты, которые было бы невозможно выявить путем исследований, направляемых гипотезой (NB! Дискуссия продолжается!). Тем не менее, охота за биологическими сюрпризами без должных мер предосторожности может легко давать богатый урожай артефактов и приводить к высокоимпактным статьям, основанным ни на чем более, кроме систематического экспериментального шума.

Статьи с ложными данными причиняют вред не только их авторам – они иницируют безнадёжные проекты... и девальвируют престиж геномных исследований. Чтобы минимизировать вред, рецензенты и редакторы должны выработать стандарты, выполнение которых требуется для того, чтобы квалифицировать находку как факт.

Далее авторы рассуждают, почему возникают ошибки. По их мнению, существует два источника: Во-первых, исключительно большой размер генома означает, что очень необычный результат может появляться случайно более часто, чем можно интуитивно ожидать. Здесь может помочь статистика.

Во-вторых, все высокопроизводительные геномные технологии дают ошибки, некоторые из которых могут быть систематическими, и на неопытный взгляд могут показаться интересным биологическим явлением. В результате исследователи могут прийти к ложным заключениям.

Авторы приводят примеры подобных ошибок. В частности, сообщалось, что редкая мутация, ведущая к потере функции в гене, названном *SIAE*, имеет большой эффект на риск аутоиммунных заболеваний. Но позже анализ более 60 000 образцов показал, что нет такой ассоциации, предполагая, что данные предыдущей публикации были получены в результате случайного события. Другой пример – публикация, сообщавшая о широко распространенной химической модификации молекул РНК. Оно было раскритиковано экспертами, которые указывали, что эта модификация является результатом типичной экспериментальной ошибки. Так оно и оказалось.

Распутывание таких противоречий после того, как результаты опубликованы, может занять годы и годы. Далее авторы предлагают меры, как уменьшить вероятность публикации ложнопозитивных результатов. Я не буду на этом останавливаться.

И в этом же журнале появился редакторский комментарий [16] под названием «Предрасположенность к ошибкам» (Error prone).

В нем говорится: «Геномика имеет потенциал для совершения революции в медицине, но становится все более ясным, что эта область вынуждена иметь дело со все возрастающим количеством болезненных проблем...»

Многие области наук о жизни в настоящее время имеют дело с массивными объемами данных, так что связанные с технологиями производства больших массивов данных артефакты не ограничиваются геномикой. Любой ученый в этой области, работающий в университете или в госпитале, может сейчас собрать образцы человеческих тканей и просеквенировать их геномы, создав огромные количества геномных данных, опыта работы с которыми у него нет. В результате снова и снова биологи не планируют эксперименты правильно и подают в печать недоработанные материалы, например, полученные с недостаточным для статистического анализа количеством образцов, а затем возмущают о случайных результатах как о биологических эффектах...»

Этот редакторский комментарий также оканчивается призывом ввести стандарты контроля качества результатов и следовать им.

Несомненно, такие меры нужны. Но, на мой взгляд, проблема ошибок сложнее. В такой конкурентной области, как геномика, ученые неизбежно ищут все более быстрые и эффективные пути генерации и анализа данных. В относительно короткое время с 2005 г. начали развиваться технологии секвенса следующей генерации (next generation sequencing, NGS). Они, как полагают, фундаментально изменили скорость геномных исследований и открыли новые пути исследования геномов и транскриптомов. Например, в июле 2012 г. фирма «Ion Torrent» в Калифорнии объявила, что она может секвенировать 100 геномов менее чем за 30 дней и по цене меньше чем 1000 долларов за геном [16]...

Тем не менее, это драматическое увеличение в скорости секвенирования происходит за счет значительного увеличения частоты ошибок. В настоящее время ошибки при непосредственном прочтении данного фрагмента составляют примерно 0,05–2% [18–21] в зависимости от разных факторов. Существует множество алгоритмов для биоинформационного понижения этой высокой ошибочности (см., например, обзор [21]). Благодаря высокой скорости анализа данная последовательность может быть прочитана много раз и, таким образом, ошибки прочтения снижены. Так, сообщают о секвенировании трех геномов человека, при котором средняя частота прочтения одной последовательности составляла 45–87 раз [18]. При этом точность определения последовательности в конечном варианте, по их данным, составляла одну ошибку на 100 000 пар оснований. Это вполне сопоставимо с точностью, достигаемой при использовании классических методов. Однако, когда речь идет о таком громадном геноме, как геном человека, – 3 миллиарда пар оснований, то такая частота ошибок означает, что следует ожидать 30 000 ошибок просто от неверных прочтений. При этом не нужно забывать, что существуют еще множественные источники ошибок, перечислять которые не входит в задачи данного обзора, но которые читатель может найти в недавних обзорах [19, 20, 22–24]. Существуют вариации генома, особенно трудно поддающиеся анализу новыми методами и вызывающие ошибки: повторяющиеся элементы, небольшие делеции и вставки, так называемая вариабельность по числу копий (copy number variants,

CNVs) и структурные перестройки генома [25–27]. Совокупность этих ошибок превращает, как это отмечают авторы обзора [23], поиски геномных изменений, приводящих к функциональным изменениям, в поиски иголки в стоге сена. Легко представить себе, как, по мере накопления геномной информации, накапливаются ошибки, и в конечном счете доходят до уровня невозможности их проверки, делая проблему трудноразрешимой или неразрешимой (для полноты картины см. также [28]). Замечу также, что хотя современные секвенирующие аппараты могут прочитывать ~250 миллиардов пар оснований в неделю по сравнению с ~25000 в 1990 г. и ~5 миллионами в 2000 г., их недостаток – прочитываемые последовательности, которые много короче, чем ~700 оснований, рутинно прочитываемых методом Сэнгера, делают проблематичной корректную сборку последовательности генома. Поэтому в большинстве случаев для сборки в качестве матрицы используют существующие геномные последовательности [4]. Понятно, что в исследуемом геноме могут быть варианты, несовместимые с прежними последовательностями.

Вот примерно так выглядит ситуация с молекулярной генетикой сегодня: громадные объемы структурной информации, получаемые со скоростью, превосходящей способность компьютеров эти данные обрабатывать (есть прогнозы, что в ближайшее время анализ данных секвенирования, а не само секвенирование, будет главным камнем преткновения у геномных проектов [29]), и сильное отставание в функциональном понимании, которое, по большому счету, продвинулось не очень сильно.

Поэтому я закончу этот неприятный раздел призывом S. Brenner [30]: «Самое время остановиться и спросить самих себя, что мы ожидаем найти в конце этого громадной головоломки -омиков (This is a good time to pause and ask ourselves what we expect to find at the end of this immense omic brainbow)».

### **Большие планы и надежды молекулярной генетики**

Вот как их видит E. Lander [4] в области исследований эволюционной и медицинской генетики человека. Я даю мои комментарии курсивом.

1. Предвидится громадный прогресс в клонировании генов ~1800 Менделевских заболеваний, которые до сих пор не идентифицированы, а также в выявлении неописанных новых Менделевских заболеваний. Это может быть осуществлено семейным анализом с использованием высокоточного полногеномного секвенирования. Поскольку рецессивные Менделевские заболевания часто связаны с полным отсутствием белка, кодируемого поврежденным геном, их трудно лечить иным путем, кроме как заместительной белковой терапией. Геномика выработала в некоторых случаях новые подходы. Но критическим шагом будет разработка генно-терапевтических подходов.

2. В случае сложных болезней, для которых в последние годы было идентифицировано более 1000 локусов, вовлеченных в болезненные фенотипы, в следующую декаду будут исследованы десятки тысяч больных с использованием секвенирования и других методов ПГА (GWAS). При этом будут развиваться

методы функционального анализа, выявляться новые гены и сигнальные системы. Кроме того, будут создаваться модельные животные, мимикрирующие физиологию человека в исследуемых заболеваниях.

*На основании сегодняшних данных трудно предвидеть успех этого направления, если не будут найдены принципиально новые пути анализа сложных систем. Сегодня они анализируются теми же методами, как и простые моногенные признаки. Однако при этом не учитываются «возникающие» в результате взаимодействий новые непредсказуемые свойства. Недавно это выявилось в эффекте «прячущейся» наследуемости (missing heritability). Я не могу анализировать этот достаточно специфический феномен в небольшом обзоре, но отсылаю заинтересованных читателей к обзорам [31, 32]. Он как раз заключается в том, что представления об аддитивном эффекте генов, формирующих сложный признак, не реализуются на практике при анализе вкладов в данный признак этих генов по отдельности.*

3. В области онкологии задачей является снижение смертности от рака. В следующую декаду следует разработать детальный каталог всех генов, которые могут служить существенными мишенями терапевтических средств во всех типах человеческого рака, а также разработать все раковые клеточные линии и необходимые животные модели. Получить данные по мутантным генам, в некоторых случаях идентифицировать мишени воздействия, клеточные сигнальные системы и взаимодействия, которые могут использоваться для терапевтического вмешательства, для создания модельных животных, позволять скрининг терапевтических средств на раковых клетках с известными молекулярными механизмами, делающими эти клетки раковыми. Будут идентифицироваться также все пути, посредством которых опухоли развивают лекарственную устойчивость. Эффективная терапия рака должна быть комбинированной, чтобы воздействовать на множественные мишени, чтобы уменьшить шансы возникновения устойчивости. Такие проекты, как Атлас ракового генома (The Cancer Genome Atlas) и международный Консорциум ракового генома (The International Cancer Genome Consortium), к сегодняшнему дню очень расширились. По мере того как геномика будет внедряться в клиническую практику, следует выработать механизмы, с помощью которых все больные раком смогут добровольно предоставлять свои геномные и клинические данные для использования в открытых исследовательских проектах. Это позволит сильно ускорить их прогресс. *См. предыдущий раздел и мнение R. Weinberg по этому поводу.*

4. Важнейшей задачей является использование геномной информации для реконструкции с максимальной детальностью наиболее существенных событий в эволюции человека. Это включает в себя анализ структуры древней популяции человечества в Африке; потоков генов с архаичными гомининами как перед, так и после их миграции из Африки, и влияние позитивной и негативной селекции в недавнее и древнее время. В следующую декаду будут собираться геномные базы данных для всех современных популяций человека, ископаемых остатков гоминин и человекообразных обезьян. Будут развиваться соответствующие технологии анализа, в том числе компьютерные. *Здесь я должен сказать, что эта задача, в отли-*

*чие от предыдущих, мне представляется наиболее реальной, поскольку она в основном может основываться только на структурной информации, и даже значительное число ошибок в структуре может оказаться несущественным для построения эволюционных взаимосвязей.*

5. Исследование модулярной клеточной биологии является одной из основных задач. Как раньше казалось нереалистичным составление полного каталога клеточных компонентов, так и сегодня таким же нереалистичным кажется составление исчерпывающей схемы их взаимосвязей. Однако сейчас требуется начать систематический анализ этого уровня организации. Система клеточных взаимодействий не бесконечно сложна. Она организована вокруг ограниченного репертуара клеточных модулей, повторное использование которых является одним из основных принципов эволюции. Эти модули включают комплексы белков, метаболические пути, цис-регуляторные системы взаимодействий и сигнальные сети. Составление полного каталога этих модулей связано со многими проблемами и потребует развития новых подходов. *Безусловно, такую задачу нужно ставить, и в процессе ее решения будут развиваться соответствующие подходы. Однако конечный ее успех – дело достаточно отдаленного будущего; вряд ли можно ожидать его в следующую декаду.*

6. Мы должны научиться программировать клетки с такой же легкостью, с какой мы программируем компьютеры. Основы уже были заложены – было сделано открытие, что взрослая соматическая клетка может быть перепрограммирована в плюрипотентную клетку, подобную стволовой, которая затем может быть превращена в дифференцированную клетку другого типа, чем исходная. *Здесь можно с уверенностью сказать, что точного соответствия геномов исходной клетки и перепрограммированной не может быть. Перепрограммированная клетка обязательно будет содержать какое-то число мутаций, возникающих стохастически и с непредсказуемым эффектом на фенотип новой клетки. Здесь работает фундаментальный запрет биологии на существование двух одинаковых сложных биологических систем [33].*

7. Наконец, фундаментальные науки должны использовать геномную информацию для развития терапевтических наук, используя неортодоксальные подходы. Для этого нужно растить новую генерацию молодых ученых. Ключевые задачи здесь включают: создание больших библиотек низкомолекулярных соединений, которые могут быть использованы для скрининга биологически активных кандидатов в лекарства; интенсивные методы фенотипического анализа, используя геномные подходы для быстрого выявления модуляторов клеточных процессов, и систематические методы нахождения белковых мишеней и механизмов действия выявленных потенциальных лекарств. Революция в медицине потребует многих декад, чтобы полностью свершиться. Геномика только еще начала проникать в биомедицинские исследования. Мы не должны обещать быстрых успехов. В то же время мы не должны терять уверенность в конечном успехе геномной медицины, которая будет преобразовать

здравоохранение для наших детей и детей наших детей.

*Это очень аккуратный подход, полностью отличающийся от фанфар, сопровождавших расшифровку генома человека. Тогда обещали чуть ли не немедленную революцию в здравоохранении, как вы могли прочесть в интервью E. Green.*

На этом E. Lander заканчивает свои задачи и прогнозы. Он естественно ограничивается геномом человека.

Вот планы и прогнозы из цитированной в первой части статьи [5]:

1. Одной из основных проблем, которая касается многих областей, станет интеграция массива ряда данных, генерируемых новыми поколениями методов секвенирования (НПС) (next generation of sequencing, NGS)... Другой проблемой является выход за пределы статического взгляда, которые дают данные определения структур геномов.

2. Несмотря на достижения в геномных технологиях в течение последней декады, существует только небольшое число успешных идентификаций функциональных генетических вариантов, которые играют роль в сложных признаках, таких как рост, диабет, сердечно-сосудистые заболевания или аутизм. Многие из локусов, идентифицированные исследованиями ПГА (GWAS), представляют только долю фенотипических вариантов и риска заболевания. В большинстве случаев функциональные биологические варианты еще не идентифицированы. Таким образом, ключевой задачей в будущем станет обнаружение «наследуемости», пропущенной при исследованиях ПГА.

3. Основной задачей станет установить корреляцию изменений экспрессии гена и его функции с вариациями генома. Небольшие различия в экспрессии и функции некоторых генов могут приводить к специфическим клеточным фенотипам, которые оказывают очень важное плейотропное действие на развитие, а также на восприимчивость и устойчивость к заболеваниям.

4. В целом мы не должны ожидать улучшения здоровья за счет полученных на основе генетических и геномных данных знаний механизмов заболевания быстрее, чем через 20 лет. Это путь от выявленного молекулярного механизма до одобренного лекарственного средства.

5. Сложность транскриптомов, протеомов и интерактомов выявила в последнюю декаду их связь с классическими проблемами. Сто лет спустя после Wilhelm Johannsen (1857–1927, датский ботаник, физиолог и генетик, который ввел термины «генотип и фенотип») взаимоотношения между генотипом и фенотипом остаются далекими от полного понимания. Следующая декада принесет новые попытки связать генотипы и фенотипы... Тесно связанные биологические компоненты образуют сложные, динамичные и логические сети (networks) или системы, которые проявляют возникающие в процессе их образования (emergent) свойства. Эти свойства непредсказуемы.

6. В области медицины с технологией, которую мы имеем сейчас, мы способны расшифровать почти все известные моногенные заболевания и, фокусируясь на развитии более совершенных ин-

тегративных анализов и инструментов, начнем накапливать более полное понимание сложных заболеваний, включая рак.

7. Мы увидим некоторые весьма впечатляющие и важные достижения в нашем понимании процессов развития и механизмов болезней, что поставит перед нами новые проблемы.

8. Следует ожидать существенных успехов в биологии развития, в которой изучение регуляторных механизмов, контролирующих состояния клетки, будет продолжать давать новые и неожиданные открытия. Например, недавнее новое понимание процессов подавления активности генов, которое критично для определения и поддержания состояния клетки, предполагает, что в эти процессы часто вовлечены молекулы некодирующей РНК (нк-РНК). Многие еще остается узнать о генах нк-РНК человека и функциях различных классов нк-РНК в клетках разного типа. Поскольку некоторые регуляторные нк-РНК являются коротко живущими, их идентификация и определение их функций представляет проблему.

9. Перед биомедицинским сообществом стоят по крайней мере две основные проблемы. Первая – начать процесс, направленный на понимание того, как ~1500 факторов транскрипции человека контролируют сотни состояний клетки, которые имеются у человека. Вторая – узнать, как модифицировать активность факторов транскрипции, используя малые молекулы, полученные химически.

Вот такие мнения относительно перспектив развития исследований человека в следующие лет десять.

Но максимального эффекта геномной революции следует ожидать в тех областях, где мы имеем дело с организмами, имеющими относительно простой геном. Такой как у вирусов (десятки-сотни тысяч нуклеотидных звеньев), бактерий (несколько миллионов пар оснований) или дрожжей с несколькими десятками миллионов пар оснований с минимальным количеством повторяющихся элементов. Это касается различного рода патогенов. В этой области полногеномное секвенирование действительно обещает преобразить сферу терапии и диагностики [34], создания эффективных вакцин, исследование различных симбиотических связей, эволюцию патогенов внутри хозяев, структуры микробных популяций, в том числе микробиомов человека [35]. Можно твердо надеяться, что мы станем свидетелями революции в диагностике, терапии и профилактике инфекционных болезней.

### **Заключение. Читателям журнала**

Я пытался дать по возможности более широкую картину событий, которые разворачивались в последние годы в области молекулярной генетики, в основном человека, поскольку это ближе к моим собственным исследованиям. Картина, безусловно, неполная, но я надеюсь, что она дает представление о динамике нашей науки. Хотелось бы также надеяться, что читатели нашего журнала будут иметь возможность не только наблюдать за происходящими событиями, но и активно в них участвовать. Страницы нашего журнала открыты для публикации ваших достижений, идей и размышлений.

## ЛИТЕРАТУРА

1. Consortium. *Nature* 2004;431:931-45.
2. *Lander E.S.* et al. *Nature* 2001;409:860-921.
3. *Venter J.C.* et al. *Science* 2001; 291: 1304-51.
4. *Lander E.S.* *Nature* 2011; 470: 187-97.
5. *Heard E., Tishkoff S., Todd J.A., Vidal M., Wagner G.P., Wang J.* et al. *Nat Rev Genet* 2010; 11: 723-33.
6. *Maher B.* *Nature* 2012; 489: 46-8.
7. *Ecker J.R., Bickmore W.A., Barroso I., Pritchard J.K., Gilad Y., Segal E.* *Nature* 2012; 489: 52-5.
8. *Neph S.* et al. *Nature* 2012; 489: 83-90.
9. *Skipper M., Dhand R., Campbell P.* *Nature* 2012; 489: 45.
10. *Birney E.* et al. *Nature* 2007; 447: 799-816.
11. *Green E.* *Science* 2011; 331: 660.
12. *Chakravarti A.* *Science* 2011; 334: 15.
13. *Alberts B.* *Science* 2012; 337: 1583.
14. *Weinberg R.* *Nature* 2010; 464: 678.
15. *Golub T.* *Nature* 2010; 464: 679.
16. Editorial. *Nature* 2012; 487: 406.
17. *Macarthur D.* *Nature* 2012; 487: 427-8.
18. *Drmanac R.* et al. *Science* 2010; 327: 78-81.
19. *Su Z., Ning B., Fang H., Hong H., Perkins R., Tong W., Shi L.* *Expert Rev Mol Diagn* 2011;11: 333-43.
20. *Kinde I., Wu J., Papadopoulos N., Kinzler K.W., Vogelstein B.* *Proc Natl Acad Sci USA* 2011;108: 9530-5.
21. *Yang X., Chockalingam S.P., Aluru S.* *Brief Bioinform* 2012. Apr 6. [Epub ahead of print]
22. *Ku C.S., Cooper D.N., Polychronakos C., Naidoo N., Wu M., Soong R.* *Ann Neurol* 2012; 71: 5-14.
23. *Fernald G.H., Capriotti E., Daneshjou R., Karczewski K.J., Altman R.B.* *Bioinformatics* 2011; 27: 1741-8.
24. *Auffray C., Caulfield T., Houry M.J., Lupski J.R., Schwab M., Veenstra T.* *Genome Med* 2011; 3: 6.
25. *Alkan C., Coe B.P., Eichler E.E.* *Nat Rev Genet* 2011;12: 363-76.
26. *Meyerson M., Gabriel S., Getz G.* *Nat Rev Genet* 2010;11: 685-96.
27. *Treangen T.J., Salzberg S.L.* *Nat Rev Genet* 2012;13: 36-46.
28. *Brunham L.R., Hayden M.R.* *Science* 2012; 336: 1112-3.
29. *Pennisi E.* *Science* 2011; 331: 666-8.
30. *Brenner S., Sejnowski T.J.* *Science* 2011; 334: 567.
31. *Manolio T.A.* et al. *Nature* 2009; 461: 747-53.
32. *Marian A.J.* *Curr Opin Cardiol* 2012; 27: 197-201.
33. *Sverdlov E.D.* *Biochemistry (Mosc)* 2009; 74: 939-44.
34. *Koser C.U.* et al. *PLoS Pathog* 2012; 8: e1002824.
35. *Wilson D.J.* *PLoS Pathog* 2012; 8: e1002874.