

ИНФОРМАЦИОННАЯ БИОЛОГИЯ -- "КЛОНДАЙК" ДЛЯ УЧЕНЫХ

Интервью подготовила Ольга УШАКОВА, "НВС".



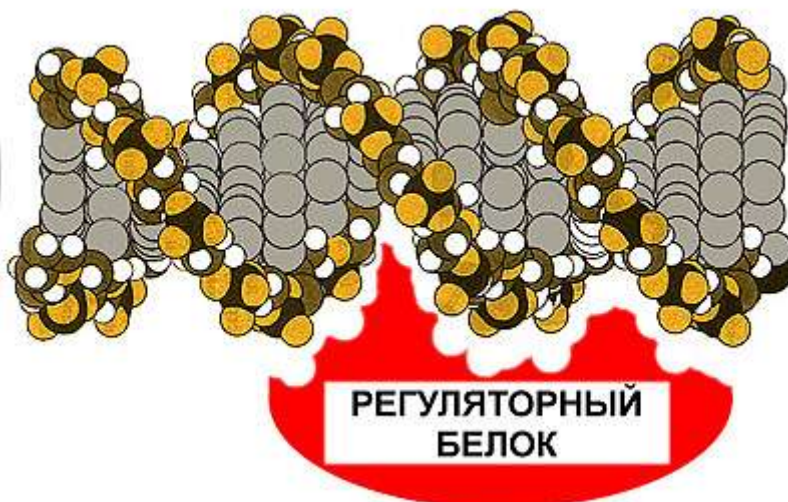
Функционирование живых организмов определяется генными сетями, управляющими молекулярными и генетическими событиями, происходящими в клетках, тканях, органах и самих организмах.

Неисповедимы пути истории развития науки. Оказывается, складывающиеся веками четкие разграничения объектов исследований ученых по различным отраслям науки на самом деле не особенно-то и существовали. Всегда физиков и математиков интересовала живая материя, а биологи и химики пытались использовать методы точных наук для изучения живых объектов. Процесс не был простым, но исторически он был обусловлен пытливостью человеческого интеллекта. В результате делались новые открытия, продвигавшие науку и человечество вперед, рождались новые отрасли наук. На пороге третьего тысячелетия мы присутствуем при рождении совершенно

оригинальных научных направлений, впитавших в себя огромный опыт развития науки нескольких веков. Об истории науки, о задачах, решаемых информационной биологией, называемой также биоинформатикой, в развитии которой участвовали ученые Сибирского отделения РАН самых разных направлений, корреспондент "НВС" беседует с заместителем директора Института цитологии и генетики СО РАН профессором Н.Колчановым.

Корр. -- Николай Александрович, сегодня к исследованиям тайн живой клетки и генома подключились представители точных наук, это проявилось достаточно внезапно, хотя готовилось довольно долго. Можно ли назвать этапы развития, чтобы понять, какой путь был пройден и какими вехами он отмечен? Ведь не сваливается же новое громадное научное направление с неба.

Двойная спираль ДНК



Н.К. -- История сама по себе длительная. Но генетика, надо сказать, изначально формировалась так, что должна была использовать определенный математический аппарат. Еще в 1866 году первый генетик, гениальный монах Мендель, чтобы обосновать закономерности, замеченные им при скрещивании линий гороха, использовал методы математической статистики. Это не случайность.

Фактически с самого начала был задан некий вектор, который показывал, что все, связанное с наследственностью, в определенном смысле может поддаваться достаточно строгому математическому анализу. Чтобы проводить такой анализ, чтобы строить математические модели процессов, необходимо большое количество экспериментальных данных. Появление в молекулярной биологии новых экспериментальных методов -- клонирование, секвенирование ДНК и др. -- привело к взрывообразному накоплению молекулярно-генетических данных.

Корр. -- И насколько этот процесс повлиял на развитие биоинформатики?

Н.К. -- Тут важно что? В 40-е и 50-е годы в биологию пришло большое число молодых активных представителей точных наук. Многие из них, возможно, были достаточно амбициозны и стремились перейти в новые области науки, где, как им казалось, они могут достигнуть новых выдающихся результатов. Среди них, как мы теперь видим, было достаточно много талантливых ученых, преимущественно физиков, с именами которых сегодня связано множество блестящих открытий, заложивших фундамент современной биологии.

Этот массовый поток представителей точных наук отражал возникавшее в то время ощущение, что главные открытия в физике уже совершены и для реализации научного потенциала необходимо искать другие области знаний, среди которых биология представлялась наиболее таинственной и потому -- наиболее интересной. После того, как была сделана атомная бомба, после того, как были поняты принципы организации атомного ядра, многим казалось, что в этой области все уже закончено. Все это было не совсем верно, но именно на 40-50-е годы пришелся мощный поток исследователей, которые шли в молекулярную биологию.

Как ни странно, он начался с инициативы человека, достигшего в этих самых точных науках очень больших высот. Основатель квантовой физики, Э.Шредингер задумался над тем, что такое жизнь с позиции точных наук и изложил свои представления в знаменитой книге "Что такое жизнь? С точки зрения физики". Эта книга стала для многих представителей точных наук -- физиков и математиков -- своеобразной библией и стимулировала переход большого числа ученых в биологию. Один из открывателей двойной спирали ДНК Ф.Крик как раз был физиком по образованию, а его коллега Д.Уотсон -- биологом. Это был блестящий пример сотрудничества физики и биологии.

Корр. -- То есть, фактически информационная биология для определения этапов своего развития использует вехи, определившие развитие молекулярной генетики и биологии. Можно ли при этом считать ее новой наукой, или все-таки она попадает под определение синтетической науки?

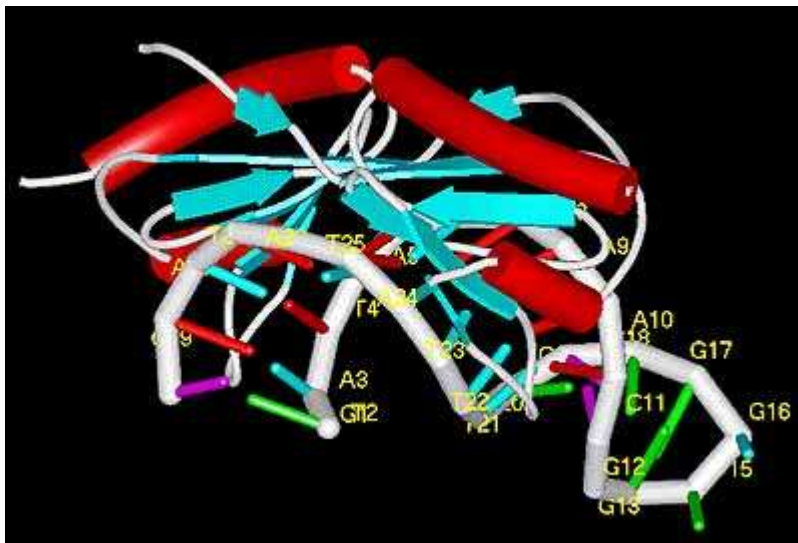
Н.К. -- На деле это все же новая наука. Еще складывающаяся. Сейчас как бы борются два названия: информационная биология и биоинформатика. Казалось бы, перестановка слов не должна сильно менять смысл, но значащей частью "биология" мы подчеркиваем, что речь идет не только о компьютерном анализе информации, полученной в результате экспериментов, но и о том, что задачей информационной биологии является понимание принципов хранения, реализации и передачи наследственной информации.

Корр. -- То есть, задачи -- биологические?

Н.К. -- Да, именно в биологических системах работают определенные законы хранения, проявления и передачи наследственной информации. Поэтому уместнее говорить именно об информационной биологии. Тем более, мировое научное сообщество в последнее время все чаще использует термин "информационная биология".

Корр. -- Мне кажется важным еще и то, что информационная биология исследует не математические или программные объекты, а биологические -- ДНК, белки, клетки, организмы и популяции.

Н.К. -- Да -- генетические макромолекулы и живые системы. Они -- самовоспроизводятся и воспроизводят свою структуру на основе наследственной информации. Жизнедеятельность этих систем, их взаимодействие с окружающей средой на основе обмена потоками вещества, энергии и информации обеспечиваются функционированием органов, тканей и клеток, которые инструктируются информацией, закодированной в генах. Фактически, информационная биология -- это наука, которая анализирует информационное содержание геномов.



Корр. -- Николай Александрович, а насколько информационные методы исследования сами по себе представляют отражение процессов в геноме? Вы используете готовый математический аппарат других наук или создаете его под биологию?

Н.К. -- Один из самых эффективных методов анализа последовательностей нуклеотидов, из которых состоят молекулы ДНК, называется методом скрытых Марковских моделей. Своими корнями он уходит в методы анализа и распознавания языка, как такового, методы компьютерной лингвистики. Есть такое понятие, как "естественные языки", то есть, сложившиеся естественным путем, при общении людей. Методы, которые были развиты около 10 лет назад для анализа этих языков, для создания технических устройств, которые позволяют более эффективно работать с этими языками, изучать их, переводить, распознавать машинными способами -- оказались очень эффективными для нашей области. Дело в том, что в основе любого языка лежит некий способ хранения информации. И здесь существуют фундаментальные закономерности, которые можно понимать и использовать в информационной биологии. Можно привести много примеров использования методов других наук в информационной биологии.

Корр. -- Возможно, это и послужило отправной точкой бурного развития информационной биологии именно в новосибирском Академгородке?

Н.К. -- Да. В Академгородке имеются давние и устойчивые традиции проведения исследований в области теоретической генетики, математической биологии, биологической кибернетики, компьютерного моделирования биологических систем и процессов. Они были заложены еще в 60-е годы такими выдающимися учеными, как А.Ляпунов, И.Полетаев, М.Колпаков и В.Ратнер. Огромной заслугой профессора В.Ратнера является организация обучения по специальности "математическая биология" на кафедре цитологии и генетики ФЕНА НГУ. За 30 лет факультет естественных наук и физфак НГУ выпустили около 100 специалистов, составивших "костяк" группы ученых, ведущих исследования по биоинформатике в ИЦИГ СО РАН. Много матбиологов, выпускников НГУ, работает в настоящее время и в других научных организациях и университетах, как в России, так и за рубежом. При этом многие из них являются общепризнанными лидерами научных направлений в информационной биологии.

Корр. -- А насколько конкурентна информационная биология, развивающаяся в Академгородке, на мировом рынке высоких научных технологий?

Н.К. -- Конкурентна вне всяких сомнений! Об этом свидетельствуют, в частности, результаты I-й международной конференции по биоинформатике регуляции и структуры генома, проходившей на базе ИЦИГ СО РАН в августе 1998 г. у нас в Академгородке. 120 ученых приняли участие в этой конференции, в том числе 34 -- из Японии, США, Германии, Франции, Великобритании, Италии и других стран с высоким уровнем развития биоинформатики. Западными учеными двигало желание посетить конференцию, организованную на базе института, известного своими пионерными разработками в нескольких ключевых направлениях биоинформатики, в первую очередь, связанных с исследованием такой сложной проблемы, как регуляция функции генов. Сотрудники нашего института представили на конференции большое число докладов о результатах своих исследований, которые вызвали огромный интерес -- очевидно, что исследователи ИЦИГ находятся в числе несомненных лидеров в этой области.

Разработки ИЦИГ в области биоинформатики исключительно конкурентоспособны на мировом рынке высоких научных технологий. Это связано с традиционно высоким уровнем теоретических исследований российских ученых; с доступностью, благодаря Интернету, подавляющей части необходимых информационных и программных ресурсов; с нашими многолетними традициями исследований по биоинформатике, и что особенно существенно, с наличием мощного коллектива высококвалифицированных специалистов по биоинформатике, работающих в ИЦИГ в тесном контакте с математиками, физиками, химиками и биологами институтов Академгородка.

Корр. -- В нашей газете около десяти лет назад была публикация о работах профессора В.Ратнера; тогда перспективы только намечались. Но сегодня-то уже ясны цели и задачи новой науки?

Н.К. -- Информационная биология занимает в современной биологии ключевую и исключительно важную позицию. К числу ее задач относится создание компьютерных баз данных для хранения экспериментальной информации о структуре и функции ДНК, РНК и белков, и о функционировании молекулярно-генетических систем организмов; разработка теоретических и компьютерных методов анализа геномов; создание компьютерных технологий моделирования молекулярно-генетических систем и процессов, в том числе, фундаментальных: репликации, транскрипции и т.д.; моделирование структурной организации и функции генетических макромолекул, молекулярных взаимодействий между ними; изучение закономерностей эволюции генетических макромолекул и молекулярно-генетических систем.

Таким образом, информационная биология относится к числу высоких технологий современной биологии и обеспечивает информационно-компьютерные и теоретические основы генетики и селекции, молекулярной генетики и биологии, генетической и белковой инженерии, биотехнологии, медицинской генетики, генодиагностики, генотерапии, словом, тех наук, благодаря выдающимся достижениям которых биология превратилась в одну из лидирующих наук грядущего столетия.

Корр. -- Если говорить об анализе и моделировании механизмов хранения, передачи и реализации наследственной информации в биологических системах -- это ведь внутриклеточные, эндогенные процессы?

Н.К. -- Да, именно они и моделируются. Собственно, это фундаментальная проблема информационной биологии, -- как кодируется наследственная информация в организме на молекулярном уровне, как она реализуется на уровне взаимодействия макромолекул,

каковы молекулярно-генетические механизмы, обеспечивающие хранение, реализацию и передачу наследственной информации. Задача в том, чтобы выяснить, как все это происходит: как идут эволюционные процессы, как возникают сбои при передаче наследственной информации и возникают мутации.

Корр. -- И математические методы дают возможность точно понимать эти процессы?

Н.К. -- Кое-что можно понимать совершенно точно. Например, достаточно развито на сегодня такое направление информационной биологии, как теория генных сетей. У направления есть своя история.

Парадигма генетики, сформировавшаяся в начале века, основывалась на том, что один ген кодирует один признак. Следующая формулировка была такой: один признак -- это группа взаимодействующих генов. Но первая парадигма сыграла свою фундаментальную роль, она позволила доказать, что существуют очень просто, с генетической точки зрения, контролируемые признаки.

Потом стало очевидно, что во многих случаях один признак может контролироваться группой взаимодействующих в пространстве и во времени генов: сначала включается один ген, затем другой, третий... и так складывается процесс.

А сейчас формируется представление о том, что один признак -- это продукт функционирования определенной генной сети, что функционирование любого гена осуществляется в ансамбле координированно работающих и взаимно регулируемых генов, обеспечивающих выполнение определенной функции организма. Генные сети содержат от десятков до сотен генов.

Количество различных вариантов взаимодействий между элементами любой генной сети исключительно велико, также как и количество различных режимов ее функционирования. Следовательно, в состояниях генной сети потенциально может быть закодировано огромное количество информации. Впервые представление о возможности кодирования генетической информации состояниями групп взаимодействующих генов было сформулировано в 60-годы в Академгородке Р.Чураевым.

Компьютерное моделирование генных сетей -- одна из центральных задач информационной биологии. Фактически речь идет об исследовании динамического поведения сверхбольших нелинейных открытых систем. Создание методов моделирования динамики генных сетей даст в руки исследователей мощный инструмент для предсказания признаков организмов, закодированных в их генах, позволит оценивать влияние мутаций на функциональные характеристики организмов и обеспечит возможность конструирования искусственных молекулярно-генетических систем с заданными свойствами.

Корр. -- Николай Александрович, эти представления можно назвать самыми последними, передовыми?

Н.К. -- Не совсем. В 60-70-е годы профессор В.Ратнер сформулировал представление о молекулярно-генетических системах, подразумевая под этим группы взаимодействующих генов, определяющих признаки организма.

Корр. -- Функция возникает тогда, когда система заработала. Одна, вторая, третья сеть... Сколько же их может быть?

Н.К. -- По приблизительным оценкам, в геноме человека имеется около 100 тысяч генов. Если предположить, что каждый ген будет взаимодействовать с каждым из остальных, то мы получим около десяти миллиардов взаимодействий, причем, одного какого-то типа. Реально, такой глобальной сети, видимо, нет, но, в принципе, генные сети могут иметь огромные размеры. Например, генная сеть, контролирующая процесс кроветворения, содержит не менее 500--600 генов и тысячи других молекулярных компонентов.

Корр. -- Н-да, и это только одна из многих, многих функций... А существуют ли генные сети, контролирующие процессы развития?

Н.К. -- Мы как раз ведем исследования в этой области. Вообще, многие лаборатории сейчас занимаются проблемами генных сетей, контролирующих процессы развития бактерий, растений, животных и человека. Исследуется также гомеостаз -- постоянство внутренней среды организма. Например, генные сети липидного метаболизма (обмена веществ) контролируют обмен липидов, то есть жиров. При возникновении дефектов в этой генной сети обмен липидов нарушается.

Корр. -- Значит, все полные люди -- это те, у кого нарушается состояние генной сети липидного метаболизма, то есть, попросту, обмена жиров, усвоения, окисления, выведения?

Н.К. -- В определенном смысле -- да. Либо дефект достался по наследству, либо является приобретенным. В любом случае возникает особый вариант состояния генной сети. А нарушение генной сети липидного метаболизма приводит, например, к нарушению работы сердечно-сосудистой системы...

Корр. -- А вот, скажем, возрастные проблемы, старение тканей?

Н.К. -- Интенсивное функционирование генной сети, которая базируется на функциях соматических клеток, может привести к определенным изменениям в них. В этих клетках наследственная память может подвергаться повреждениям под воздействием каких-то факторов. Как следствие, могут возникнуть определенные дефекты в генных сетях. Если бы речь шла о том, что один ген -- один признак... Но ведь в сетях взаимодействуют тысячи генов, и вероятность возникновения дефектов тоже увеличивается в тысячи раз. Проблема старения -- это проблема функционирования и надежности любых сложных систем. Чем сложнее система, тем больше вероятность того, что какой-то элемент этой системы выйдет из строя. И если этот элемент обеспечивает какой-то лимитирующий процесс, то система будет работать со сбоями или откажет. Если же таковой не является лимитирующим, то система будет вполне нормально действовать.

Корр. -- А все-таки, возможно, что решение каких-то крупных проблем геронтологии находится где-то в рамках развития проблематики информационной биологии?

Н.К. -- Думаю, что для некоторых -- да.

Корр. -- Николай Александрович, в любом случае, важны, как я понимаю, прежде всего, огромный массив данных, накапливаемый информационной биологией, и методы, которые позволяют выстроить из него систему?

Н.К. -- Именно систему. И определить, как зависят выходные параметры этой системы -- это собственно является признаком на молекулярном, клеточном, организменном уровне - от ее внутренних параметров. В действительности, проблема генных сетей -- это проблема описания биологических процессов на самых различных уровнях, от молекулярного до организменного.

Мы сейчас в наших исследованиях подходим к ситуации, когда накопленные экспериментальные данные позволяют строить достаточно реалистичные модели.

Корр. -- А объектами для ваших моделей являются какие-то белки, ДНК, клетка, ткани?

Н.К. -- Не только. У нас в лаборатории -- еще и организм. Дело в том, что, если при функционировании геномной сети гены, входящие в нее, активны в клетках разных тканей, то взаимодействие между этими генами относится к уровню организма.

Сейчас становится реалистичным описание конкретных геномных сетей в организме человека. В частности, в нашей лаборатории в сотрудничестве с другими организациями ведутся работы по описанию геномных сетей, которые контролируют противовирусный ответ, систему липидного обмена, систему кроветворения и многие другие.

Я думаю, что моделирование геномных сетей, которые контролируют важные с медицинской точки зрения признаки, это дело ближайших лет.

Корр. -- А как ваши модели выглядят? Они -- в пробирках, на экране компьютера, в формулах?

Н.К. -- Тридцать лет назад одним из наиболее распространенных методов представления моделей были системы дифференциальных уравнений. Существенное отличие нынешней информационной биологии от ее корней, в том, что она использует, как сложные компьютерные модели, так и информацию, накопленную в базах данных. На экране дисплея модель может быть представлена в виде графического образа, отражающего особенности изучаемого процесса. Зачастую, эти картины не только информативны, но и красочны.

Корр. -- Результаты каких экспериментов используются для моделирования?

Н.К. -- Видите ли, мы сами экспериментами не занимаемся. Мы собираем в базах данных информацию из публикаций. И далее можем использовать ее для моделирования. Создание баз данных -- дело сложное и трудоемкое, требующее высокой квалификации экспертов-биологов.

Сейчас в нашем институте информационной биологией занимаются несколько лабораторий: молекулярно-генетических систем -- это как раз лаборатория профессора В.Ратнера, лаборатории теоретической генетики, молекулярной эволюции и лаборатория рекомбинационного и сегрегационного анализа.

В наших лабораториях можно сегодня описывать и моделировать практически все уровни организации и эволюции живых систем.

Корр. -- Николай Александрович, вы назвали в числе других лабораторию молекулярной эволюции. Эволюционная идея вписывается в объект исследований?

Н.К. -- Эволюционная идея всегда была одним из "краеугольных камней" биологии. В настоящее время, в эпоху массовых геномных исследований, теория эволюции имеет возможность получить ответ на многие из волновавших ее ранее вопросов, касающихся механизмов эволюции геномов, темпов возникновения генетической изменчивости, направленности эволюции геномов. Накапливается все больше доказательств того, что ключевые события прогрессивной эволюции организмов, направленные на повышение надежности и устойчивости их воспроизведения, на адаптацию к принципиально новым средовым условиям и так далее, обеспечиваются изменением и усложнением именно

генетических регуляторных систем организмов. Исследования по генетике поведения, проводившиеся в свое время под руководством академика Д.Беляева, дают огромный материал для анализа. Это невероятно ценные экспериментальные данные.

Сегодня очевидно, что теория молекулярной эволюции превратилась в науку, имеющую очень важные практические приложения при изучении геномов, и может рассматриваться как одно из наиболее развитых и перспективных направлений информационной биологии.

Корр. -- Вы относите информационную биологию к высоким технологиям, значит, она должна заинтересовать тех, кто формирует рыночные отношения.

Н.К. -- Это своеобразный "Клондайк", как в смысле богатства идей для самой биологии и других наук, так и в смысле возможных инвестиций. О возможностях взаимодействия информационной биологии с другими науками я говорил выше. Информационная биология является одной из наиболее выгодных областей коммерческих инвестиций. Сферами ее приложения являются биотехнология, генетическая и белковая инженерия, создание трансгенных форм животных и растений, фармакология, генодиагностика, генотерапия...

Объемы инвестиций, выделяемых в развитых странах на проведение исследований и научно-конструкторских разработок в этих направлениях, составляют десятки миллиардов долларов. Так как информационная биология является одним из "становых хребтов" высоких биологических технологий, инвестиции, вкладываемые в ее развитие, окупаются очень быстро. Наиболее эффективный канал возвращения вложенных средств состоит в исключительно большой прибыли, получаемой при продаже баз данных и пакетов прикладных программ.

Например, покупка крупной фирмой лицензии на использование широко известной базы данных Swiss-Prot, содержащей информацию о первичных структурах белков, может стоить свыше ста тысяч долларов. Стоимость баз данных, содержащих информацию о секвенированных участках геномной ДНК человека или, скажем, каких-то полезных растений или животных, может составлять от сотен тысяч до нескольких миллионов долларов. В настоящее время в развитых странах имеется несколько тысяч фирм, работающих в области высоких биологических технологий. Так что, информационная биология имеет громадный рынок для коммерческой реализации своих разработок.

На рисунках:

-- участок двойной спирали ДНК -- момент (условно) взаимодействия с белком;

-- цветная схема кристаллической структуры комплекса дрожжевого ТАТА-бок-связывающего белка (ТВР) с ДНК (Y.Kim, J.H.Geiger, S.Hahn, P.B.Sigler, Nature, 365, pp.512-520, 1993). Схема взята из базы данных "Активность", где документировано свыше 400 экспериментов по характеристикам комплексов ДНК с регуляторными белками).