

## **БИОИНФОРМАТИКА В «ПОСТГЕНОМНУЮ ЭПОХУ»**

**В яркий, солнечный воскресный день 25 июля в новосибирском Академгородке, в Доме ученых начала работу престижная международная конференция по биоинформатике регуляции и структуры генома (BGRS'2004).**

Людмила Юдина, «НВС»

Атмосфера была праздничной, собравшаяся публика — оживленной. Гостей, прибывших в сибирский город науки со всех концов страны, а также из США, Японии, Италии, Швеции, Швейцарии, Франции, Индии, Германии и других стран сердечно приветствовали представители администрации города и области. На открытии форума выступили: директор Института цитологии и генетики СО РАН (ИЦиГ — организатор конференции) академик Владимир Шумный, председатель конференции член-корреспондент РАН Николай Колчанов и сопредседатель, профессор Билефельдского университета Ральф Хофестаedt.



**Конференцию открывает директор ИЦиГ СО РАН академик В. Шумный.**



**Дискуссия по стендовому докладу д-ра Патрицио Арриго из Института структур макромолекул (Генуя, Италия) .**

Регулярное проведение конференций BGRS каждые два года (начиная с 1998 года) стало доброй традицией СО РАН, российского и международного научного сообщества, занимающегося проблемами компьютерного исследования геномов и биоинформатики. Ныне на BGRS'2004 собралось более 200 ученых из 21 страны мира.

Биоинформатика сегодня у всех на устах. Наступление так называемой «постгеномной эпохи», времени после полного секвенирования последовательностей ДНК геномов человека и других организмов, осуществленного в рамках глобальных международных научных программ, сделало данную дисциплину особо востребованной. На сегодняшний день известны полные последовательности геномов более чем для 200 организмов, и это число продолжает стремительно расти. Сравнительный анализ последовательностей геномной ДНК родственных организмов — сравнительная геномика — позволяет лучше понять строение и функцию генов человека во многих случаях без дорогостоящих экспериментальных исследований.

Именно поэтому в ходе конференции, которая работала шесть дней, сравнительной компьютерной геномике было уделено повышенное внимание. Обсуждено и множество других проблем — компьютерная структурная и функциональная геномика; структурная компьютерная биология, компьютерная системная биология (генные сети, пути передачи сигналов в живой клетке, метаболические пути, пути экспрессии генов), компьютерная эволюционная биология (генетическое разнообразие и полиморфизм); биомолекулярные системы и процессы: анализ данных и продукция знаний; биоинформатика и образование. Последней, но не менее значимой теме — образованию — уделяется особое внимание. «Наука в Сибири» уже писала о Научно-образовательном центре «Молекулярный дизайн и экологически безопасные технологии». В НГУ на факультете естественных наук открыта кафедра информационной биологии, началась специализация студентов, многие из которых, пока в соавторстве, представляют на конференции свои первые стендовые сообщения.



**Рабочий момент конференции, в центре, у стенда — д.ф.-м.н. В. Голубятников (Институт математики СО РАН).**

На BGRS'2004 были содержательные пленарные лекции, стендовые сессии, компьютерные демонстрации, в частности по параллельным компьютерным вычислениям. В общем, все самые-самые актуальные проблемы рассматриваемой области были в центре внимания ученых и представителей промышленных кругов из биотехнических и фармацевтических компаний и фирм, пожелавших принять участие в столь крупном научном форуме.



**Сессию ИНТАС провели руководитель департамента естественных наук ИНТАС Поль Бекерс и председатель программного комитета конференции член-корреспондент РАН Н. Колчанов.**

выступил представитель Евросоюза Индриди Бенедиктсон, который посвятил собравшихся в планы Евросоюза и программу действий в области научной деятельности. Член-корреспондент Николай Колчанов рассказал о состоянии биоинформатики в странах

В один из дней Европейская организация содействия ученым из стран бывшего СССР (ИНТАС) провела свое совещание. Как заметил руководитель направления естественных наук, менеджер ИНТАС Пол Беккерс, главная задача встречи — наладить организацию работ по проектам сотрудничества Евросоюза и России, наметить график будущих встреч, определить возможные источники финансирования. Предварительно в научные организации были разосланы соответствующие анкеты, цель которых — выяснить область научных интересов исследователей. Исходя из содержания анкет строилась стратегия и тактика дальнейших действий.

На нынешней сессии ИНТАС, заседания которой проходили в плотном режиме в течение всего дня и при широком составе участников,

бывшего Советского Союза, провел обзор программ. Был обсужден график двусторонних встреч.

Следует отметить, что ИНТАС с особым вниманием относится к Новосибирску — Сибирское отделение РАН имеет многолетний опыт организации междисциплинарных исследований в области математики и информатики, естественных наук и практику международного сотрудничества в этих отраслях знаний.

### Говорят участники конференции



Профессор Юрий Богданов, Институт общей генетики им. Н. И. Вавилова РАН

— Юрий Федорович, в день открытия конференции вы выступили с докладом, который, наверняка, никого не оставил равнодушным. Посвятите читателей в область ваших научных интересов.

— До недавнего времени я заведовал сугубо экспериментальной лабораторией. Сейчас мы стали активно использовать комплекс методов — то, что ныне называется биоинформатикой, для предсказания некоторых интересующих нас явлений, в частности, — результатов взаимодействия белков с ДНК. Нас интересовали построения специфических структур, которые возникают в клетке только в тот момент, когда она переходит от деления путем митоза к делению путем мейоза.

Митозом делятся все клетки тела, способные к размножению. А мейозом — только генеративные, которые дают начало половым клеткам. При этом происходят существенные перестройки внутри хромосом и в клетке. В дело вступают специфические гены, специфические белки, которые раньше молчали. Как говорится — всю жизнь молчат, а при созревании половых клеток начинают работать. И возникает целый ряд вопросов, на которые требуется дать ответ.

Много исследований ведется на простом модельном объекте — дрожжах. Открыто 360 специфических генов, работающих только при переходе клетки от деления путем митоза к делению путем мейоза. У дрозофилы таких генов открыто всего 125. Это не значит, что их именно столько, просто, как я думаю, еще предстоит открыть остальные, недостающие до 360.



Рабочий момент конференции.

Нам захотелось узнать, а можно ли предсказывать существование некоторых генов. Например, найти у дрозофилы такие гены, которые еще не открыты, но должны быть. С помощью биоинформационных методов это впервые удалось.

На прошлой конференции в 2002 году я доложил результаты такой работы. Был предсказан неизвестный белковый продукт. То есть о белке, который кодируется

определенным геном, ничего не знали. Его удалось предсказать по аналогии с белками дрожжей. За основу взяли структуру аналогичного белка дрожжей. Дело в том, что сам процесс мейоза, деления генеративных клеток, консервативен. Схема выработалась миллионы лет назад и осталась неизменной у дрожжей, насекомых, позвоночных животных.

Мы предположили, что белок у дрозофилы должен иметь такую же схему строения. Что такое — схема строения? Блочная организация, доменная организация, белок ведь состоит из блоков.

К тому времени весь геном дрозофилы только что секвенировали в США, и было установлено, что он содержит 13600 «открытых рамок считывания», т.е. потенциальных генов. С одной из «рамок» могла бы считываться информационная РНК для искомого нами белка. Как ее найти?

Среди 125 известных генов мейоза у дрозофилы мы, зная микроскопическую картину мейоза, предвидели лишь один ген — кандидат на эту роль. Дело в том, что мутация гена  $c(3)G$  лишала хромосомы способности соединяться попарно с помощью застежки «молнии» и вступать в рекомбинацию. Мы предположили, что у мутантов дефектен белок, формирующий субмикроскопические зубцы застежки. Сам искомый белок должен быть аналогичен (по вторичной структуре и конформации) белку Zip1 дрожжей (от слова zipper — застежка-молния).

Примерная локализация гена  $c(3)G$  в хромосоме 3 у дрозофилы была известна, и мы компьютерными методами начали искать в базе данных об этом районе, (составляющем 700 тыс. пар нуклеотидов), такую рамку считывания, которая могла бы кодировать похожий белок. Мы понимали, что гомологии белка (т.е. сходства последовательности аминокислот) у дрозофилы и дрожжей быть не может, но белок должен быть таких же размеров, с такой же организацией из трех доменов и способностью центрального домена формировать альфа-спираль длиной около 40 нанометров. Эти условия диктовались сходством электронно-микроскопической картины попарно соединенных хромосом в мейозе у дрожжей и у дрозофилы.

Просмотрели открытые рамки считывания почти для 80 генов в районе поиска. Пришлось использовать программы, позволяющие прогнозировать вторичную структуру виртуального белка, его физико-химические свойства и распределение электростатических зарядов в молекулах. И когда уже потеряли всякую надежду, Татьяна Гришаева нашла такую рамку считывания на границе предсказанной (зарубежными генетиками) зоны локализации гена  $c(3)G$  на микроскопической карте хромосом. Это был ген CG17604 по геномной карте компании «Селера».

Мы заключили, что этот виртуальный ген должен быть давно известным геном  $c(3)G$  и кодировать белок, аналогичный белку Zip1 дрожжей. Послали сообщение в Drosophila Information Service, и на той же неделе получили электронное письмо из США от проф. Скотта Хоули, в котором он сообщал, что экспериментально доказал, что ген  $c(3)G$  кодирует белок, формирующий застежку-молнию между хромосомами в мейозе у дрозофилы. Результаты наших работ совпали, но экспериментальная работа группы



Д-р Акинори Сарай  
(Япония) .

Хоули заняла около семи лет, а наша компьютерная работа силами трех человек — лишь около трех месяцев. Статьи вышли из печати одновременно.

Эта история была рассказана мною на конференции BGRS в 2002, а сейчас мы с Сергеем Дадашевым и Татьяной Гришаевой доложили новую работу, в которой доказываем, что хромосомная ДНК крепится к зубцам застежки молнии в мейозе у человека с помощью повторяющихся последовательностей нуклеотидов из семейства Alu. Она также целиком сделана средствами биоинформатики, но при этом, как и в истории с дрозофилой, мы опираемся на тридцатилетний опыт экспериментальной работы в области цитологии и генетики мейоза и на знание его универсальных законов, на которых, кстати, построена вся классическая генетика. Биоинформатика не заменяет классических наук и мозг ученого.



Профессор Михаил Гельфанд, Институт проблем передачи информации РАН

— Михаил Сергеевич, ваши сообщения на научных форумах вызывают неизменный интерес. Какой проблеме посвятили выступление на нынешней конференции?

— Я выступил с докладом «Эволюция регуляторных систем». Речь шла о бактериях. Идея довольно простая. Сейчас уже известно столько бактериальных геномов, что путем сравнения разных геномов друг с другом можно рассматривать целые метаболические или регуляторные системы.

Метаболическая система — это набор генов, которые участвуют в одном и том же процессе, в разных его стадиях. Например, вы кормите бактерию каким-то сахаром, а она его утилизирует — у нее есть система питания этим сахаром. Бактерия может выполнять необходимые ей функции: нужна аминокислота — есть метаболический путь для производства этой аминокислоты.

У регуляторной системы — иные задачи. Если гены работают вместе, то они должны включаться и выключаться одновременно: то есть, если вы кормите бактерию сахаром, например, фруктозой, то включается система для утилизации именно этого сахара. А если ей не хватает метионина, аминокислоты, то включается система для производства метионина, потому что без него жить нельзя. Или — белок-транспортёр, который берет метионин из внешней среды. Геномов сейчас так много, что можно следить за судьбой метаболических и регуляторных систем, т.е., повторяю, можно сравнивать уже не отдельные гены, а целые системы.

— Вы участвовали во всех конференциях по биоинформатике регуляции и структуры генома?

— Во все четырех.

— И как, на ваш взгляд, идет развитие темы?

— Как и полагается — накапливается информация, растет число фактов и открытий, о чем можем судить по доложенным материалам. Относительно нынешней конференции могу сказать одно — здесь собрались очень сильные специалисты, было много содержательных, глубоких докладов.

Что касается моей работы, то за последние два года удалось сделать некоторое количество нетривиальных биологических утверждений. Возьмем, к примеру, эволюцию. Ее очень интересно изучать, но проверить все утверждения трудно — кто знает, что происходило раньше! Мы можем строить более или менее убедительные модели и обсуждать, насколько они хороши.

С другой стороны, мы можем исследовать целые системы и делать совершенно конкретные проверяемые предсказания. Более того — со временем все увереннее предсказывать функции отдельных генов. Раньше мы могли сказать примерно так — этот белок похож на уже известный, наверное, он делает то же самое. А сейчас, поскольку мы смотрим на гены в функциональном контексте, то выдаем обширную информацию.

Возьмем такой пример. В какой-то бактерии найден метаболический путь почти целиком, а фермента для промежуточной стадии мы не знаем. Но можем модифицировать пустоты в наших представлениях о бактериях и заполнить их путем применения более тонких методов, которые и пытаемся развивать.

Это уже совершенно конкретное биологическое утверждение. Мы не говорим — эволюция шла таким-то способом, а утверждаем — неизвестно, какой бы белок в данной бактерии выполнял ту или иную функцию, но у нас есть хороший кандидат, помогающий ответить на вопрос.

Работы в данном направлении ведем довольно давно, но именно последние годы отмечены неожиданными и красивыми нетривиальными предсказаниями. Даже удалось опубликовать результаты раньше экспериментаторов.



Профессор Евгений Коротков, Московский инженерно-физический институт.

— Евгений Вадимович, чем для вас интересна конференция?

— На ней много ученых из России, тех, кто продолжает работать в российских академических учреждениях. И — молодежи нашей, отечественной, и тех, кто временно (надеюсь) трудится за пределами страны.

Внимание к биоинформатике сейчас повышенное. Она находится на стыке наук. Многие специалисты в своей работе используют те и другие методы.

— Можно ли вас отнести к их числу?

— Я работаю в двух учреждениях. Профессор Московского инженерно-физического института и сотрудник кафедры кибернетики, где у меня группа студентов, овладевающих биоинформатикой — готовим специалистов с 1995 года. И еще есть научная группа в Центре биоинженерии РАН.

— С каким материалом выступили здесь?

— Мы разработали новый метод по поиску периодичности в символьных последовательностях, и он дает неожиданные результаты, которые мы не смогли бы получить, используя классические методы. Удалось показать, что белковые семейства или

белки, которые выполняют одну и ту же биологическую функцию, обладают однотипной скрытой периодичностью. Можно считать, что все они имеют как бы скрытую волну, которая наложена на эту аминокислотную последовательность.

Наши результаты имеют и практическое применение. С одной стороны, зная, что там есть такая же волна, мы можем абсолютно точно идентифицировать биологическую функцию аминокислотной последовательности, ничего не зная о ней больше. То есть, исходя из имеющихся фактов, предсказывать биологическую функцию аминокислотной последовательности.

С другой стороны, волны в последовательности имеют отношение и к теории солитонов. Еще Энрико Ферми начал решать эти задачи. Затем работы продолжил австралийский ученый Крускал. Им удалось доказать, что если имеются определенные периоды в одномерном кристалле с нелинейными связями, то в этом случае возможно образование таких комбинаций, которые будут очень долго сохраняться в этой структуре.

В белках мы видим похожую картину. И уже появилась идея, что белки представляют из себя такие же тепловые насосы, которые закачивают часть тепловой энергии в резонансные частоты, как раз и определяющиеся скрытой периодичностью.

В принципе такая гипотеза может объяснить механизм работы ферментов. Сейчас настал такой момент, когда одновременно с нашими теоретическими исследованиями следует провести синтез искусственных белков и показать, что возможно рассчитывать резонансные частоты и синтезировать новые ферменты с абсолютно новыми биологическими функциями.

Например, все знают о ДДТ, которым буквально засорен земной шар. Препарат очень медленно разлагается, и в природе нет фермента, который бы его расщеплял. А если рассчитать резонансные частоты, то можно сделать искусственный белок, который стал бы расщеплять ДДТ.

Конечно, многое из того, о чем говорят ученые, поначалу кажется фантастикой. Но известно также множество случаев, когда самые несбыточные на первый взгляд идеи становились реальностью.

Или можно создать ферменты небелковой природы для нефтехимии. Ведь крекинг-процесс требует высоких температур и затрат энергии. И если удастся получить фермент небелковой природы, расщепляющий углеводороды, технология станет много проще. Нужные цепочки можно сделать не только из аминокислот, важно понимать, как работает сам фермент. Если идея теплового резонансного насоса будет реализована, вся нефтехимия сильно изменится, будут не нужны огромные заводы, и можно будет получать требуемые продукты (бензин, керосин и т.д.) буквально при комнатной температуре.

Пока это все было из области теории, но пришло время, когда мы можем создавать искусственные ферменты. Только исследования требуют существенного финансирования — в этом вся проблема.

\* \* \*

29 июля. Конференция близится к своему завершению. Но еще продолжаются заседания, двусторонние встречи. Заняты все компьютеры, установленные в холле рядом с залом

заседаний. Слишком велико желание общаться. Организатор серии конференций BGRS, председатель ее Программного комитета член-корреспондент Николай Колчанов завершил очередные переговоры, а его уже приглашают на следующее мероприятие. Попросила Николая Александровича уделить мне несколько минут, чтобы оценить итоги конференции.

— Сегодня предпоследний день работы. Реализована основная часть программы, включая большую часть научных сессий по проблемам компьютерной геномики и структурной компьютерной биологии, эволюционной проблематике, методам анализа и статистической обработки данных для биоинформатики. Остается провести только одну сессию — по системной компьютерной биологии, которая является сейчас ключевым направлением биоинформатики.

Хочу отметить высокую активность участников.

Завязалось большое количество контактов и не только с российскими учеными из разных городов, но и с иностранными.

Это очень продуктивная конференция с точки зрения обсуждения самых горячих точек биоинформатики, налаживания нужных и важных контактов, механизмов сотрудничества.

Прошло серьезное совещание, посвященное проблемам взаимодействия российской биоинформатики с европейским сообществом. Обсуждение было конструктивным. Прозвучало много конкретных предложений, которые затем, после конференции, будут обсуждаться европейскими коллегами и администраторами программ. И это убедительный пример действия обратной связи, когда оценивается ситуация и создаются новые механизмы.



**Экспрессивно, «на пальцах», дает пояснения слушателям д-р Рамачандран Сринивасан (Институт геномики и системной биологии, Дели, Индия) .**

Содержательными были небольшие параллельные рабочие совещания — их проведено несколько. Например, одно, высоко оцененное всеми, — по проблемам компьютерного анализа изображений. Изображения, получаемые с помощью микроскопов или другой техники — один из важнейших способов исследования биологических объектов. Договорились в следующий раз провести рабочее совещание по этой тематике.

\* \* \*

Крупнейшая конференция, BGRS'2004, завершила работу на оптимистической ноте, очертив круг проблем, на которых будут сосредоточены исследования по биоинформатике регуляции и структуры генома. Результаты, думается, будут обнародованы на BGRS'2006.

Фото В. Новикова.