

# №7 1998 год ПЕРВАЯ МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ ПО БИОИНФОРМАТИКЕ РЕГУЛЯЦИИ И СТРУКТУРЫ ГЕНОМА (THE FIRST INTERNATIONAL CONFERENCE ON BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE — BGRS'98)

Новосибирск – Алтай, Россия,  
24 августа – 1 сентября 1998 г.

Первая международная конференция по биоинформатике регуляции и структуры генома проходила с 24 по 28 августа 1998 г. в Новосибирском Академгородке на базе Института цитологии и генетики СО РАН. Какова цель проведения конференции с таким названием и почему она проводилась именно в Новосибирске?

В последние годы происходит стремительное накопление экспериментальных данных по различным аспектам структурно-функциональной организации молекулярно-генетических систем организмов. В качестве одного из наиболее мощных источников этой информации выступают программы крупномасштабного секвенирования (полной расшифровки) нуклеотидных последовательностей геномов про- и эукариот. В настоящее время многие сотни государственных и частных научно-исследовательских организаций в различных странах мира осуществляют исследование огромного количества геномов различных про- и эукариотических организмов. Некоторые из проектов крупномасштабного исследования геномов уже закончены или близки к завершению. Осуществлено секвенирование геномов многих сотен вирусов. Проведена полная расшифровка прокариотического генома кишечной палочки *E.coli*, а также геномов ряда других прокариотических организмов (в первую очередь – возбудителей различных заболеваний). Прочитан геном простейшего одноклеточного эукариотического организма – дрожжей. Расшифрована последовательность геномной ДНК простейшего многоклеточного организма – кольцевого червя *S.elegans*. Впечатляющие успехи достигнуты в секвенировании генома человека: расшифровано около 25-30% всей геномной ДНК. Значительный прогресс достигнут также в изучении геномов плодовой мушки *D.melanogaster*, мыши, растительного генома *Arabidopsis thaliana*. Проводятся исследования геномов различных хозяйственно полезных животных. В связи с осуществлением перечисленных выше геномных проектов накапливаются огромные объемы экспериментальных данных.

Изучение структурно-функциональной организации геномов не сводится только к расшифровке их нуклеотидных последовательностей и не является самоцелью в осуществлении геномных проектов. Программой максимум является понимание механизмов, посредством которых информация, записанная в геномной ДНК (РНК), обеспечивает самовоспроизведение и функционирование организмов про- и эукариот, их взаимодействие с окружающей средой при обмене потоками вещества, энергии и информации, реакцию на изменяющиеся условия внешней среды, адаптацию к этим условиям и т.д.

Известно, что протекание каждого из перечисленных выше процессов контролируется, как правило, большим количеством генов. Экспрессия генов основана на исключительно сложной многоэтапной последовательности молекулярно-генетических событий. Особенно сложно она организована у эукариот, включая помимо транскрипции, как первого молекулярного события, запускающего экспрессию, такие процессы, как созревание мРНК (кэпирование, сплайсинг, полиаденилирование), транспорт мРНК через ядерные поры, трансляция, формирование уникальной пространственной структуры белковых молекул, обеспечивающих выполнение структурных, ферментных, транспортных и др. функций. Для регуляции экспрессии генов огромное значение имеет структурная организация и функциональная динамика хроматина. При изучении молекулярных механизмов экспрессии генов также накапливаются огромные объемы экспериментальной информации.

Для того чтобы читатели оценили объемы информации, получаемой только при секвенировании геномов, заметим, что для представления в печатной форме полной нуклеотидной последовательности даже такого небольшого генома, как геном *E.coli*, требуется книга объемом около 1000 страниц убогистого текста. Что же касается генома человека, то для этого потребуются около 1000 таких книг (то есть целая библиотека). В связи с исключительно высокими темпами расшифровки геномной ДНК наблюдается увеличение объема информации, накапливаемой в базах данных нуклеотидных последовательностей (EMBL и GeneBank). При этом информация, накапливаемая в базах данных, касается только одного аспекта организации геномной ДНК, а именно ее первичной структуры. Естественно, что для описания функциональных особенностей геномов, например особенностей регуляции экспрессии содержащихся в них генов, требуются неизмеримо большие объемы информации. Например, в организме человека по приближенным оценкам имеется около 100 тысяч генов, кодирующих белки. Информацию о пространственной структуре белков получают на основе рентгеноструктурного анализа, позволяющего расшифровывать пространственные координаты атомов. В настоящее время расшифрованы пространственные структуры примерно 9000 белков из различных организмов, информация о которых накапливается в базе данных PDB. Для описания пространственной структуры одного белка среднего размера (около 600 аминокислот) требуется около 200 страниц текста. Ясно поэтому, что для хранения сведений даже об этой относительно небольшой доле расшифрованных белков требуется библиотека огромного размера.

Существенно, что функционирование любого гена осуществляется в пределах некоторой геновой сети. Любая геновая сеть представляет из себя ансамбль координированно работающих генов, обеспечивающий выполнение определенной функции организма. Как правило, геновые сети включают в себя от нескольких десятков до многих сотен координированно функционирующих генов. Например, геновая сеть системы кроветворения включает в себя не менее 500-600 генов. Даже неполный список генов, входящих в геновую сеть, контролирующую прохождение клеточного цикла, включает более 100 генов и т.д.

К числу структурных компонентов геновых сетей относятся сами гены, кодируемые ими РНК и белки, пути передачи сигналов, ведущие от клеточных рецепторов к регуляторным районам генов, и т.д. Оценки показывают, что количество различных

вариантов взаимодействий между элементами даже относительно небольшой генной сети, включающей 100-200 генов, может быть исключительно велико, точно также как количество различных режимов функционирования генных сетей. Объемы экспериментальной информации о структурно-функциональной организации генных сетей, накопленные в настоящее время, весьма значительны.

С учетом сказанного становится ясна та ключевая позиция, которую занимает в современных молекулярно-биологических и молекулярно-генетических исследованиях новая наука биоинформатика. К числу ее задач относятся: создание компьютерных баз данных для хранения экспериментальных данных; разработка теоретических и компьютерных методов анализа экспериментальной информации; создание компьютерных технологий моделирования молекулярно-генетических систем и процессов, в том числе и тех, которые протекают на уровне генных сетей; моделирование структурно-функциональной организации (вторичной, третичной структуры) генетических макромолекул (ДНК, РНК и белков) на основе информации об их первичных структурах; компьютерное моделирование фундаментальных молекулярно-генетических процессов (репликации, транскрипции, сплайсинга, трансляции); компьютерное моделирование молекулярных взаимодействий между генетическими макромолекулами; исследование закономерностей молекулярной эволюции генетических макромолекул, а также молекулярно-генетических систем.

С учетом развития Интернет-технологий исключительно важное значение приобретает обеспечение сетевого доступа пользователей-молекулярных биологов к создаваемым информационным и программным ресурсам по молекулярной биологии и генетике, имеющимся в сети Интернет. Рассмотрению всех этих вопросов и была посвящена конференция BGRS-98.

В названии конференции отражен тот факт, что геномы рассматриваются с двух точек зрения: структурной и регуляторной. О важной роли структурного аспекта при компьютерном анализе информации, получаемой в рамках геномных проектов, указывалось выше. Что же касается регуляторных механизмов функционирования геномов, то интерес к их изучению в последние годы растет стремительно. Ведущие ученые, работающие в области экспериментального и компьютерного исследования геномов, приходят к выводу, что в самом ближайшем будущем произойдет радикальное смещение научной активности от прямого секвенирования геномной ДНК (что представляет собой, вообще говоря, уже чисто технологическую задачу) к исследованию функционирования геномов и, в первую очередь, изучению молекулярно-генетических особенностей их регуляции. О важности разворачивания теоретических и компьютерных исследований в области регуляторной геномики свидетельствует, в частности, и такой факт, хорошо известный специалистам, что структурные части генов, кодирующих белки, занимают в геномной ДНК человека и других высших эукариот примерно 3-5%, в то время как остальная часть геномов занята некодирующей ДНК, причем во многих случаях с очень выраженным регуляторным потенциалом. Известно также, что регуляторные районы генов высших эукариот имеют зачастую огромные размеры, до 20-30 тыс. пар оснований, что может быть на порядок больше, чем размер кодирующих частей генов. Кроме того, структурно-функциональная организация регуляторных районов генов высших эукариот характеризуется высокой сложностью, наличием большого количества иерархических уровней регуляции и огромным разнообразием регуляторных элементов. Дополнительную сложность регуляторным механизмам придает тот факт, что они реализуются в составе генных сетей.

В целом результаты сравнительного анализа механизмов регуляции функции генов подтверждают вывод знаменитого отечественного эволюциониста академика И.И.Шмальгаузена о том, что ключевые события прогрессивной эволюции организмов, направленные на повышение надежности и устойчивости их воспроизведения и повышения степени автономизации от повреждающих условий внешней среды, заключаются в усложнении именно генетических регуляторных механизмов. Этот вывод, сделанный более 50 лет назад, в настоящее время блестяще подтверждаются современной молекулярной биологией и генетикой.

Биоинформатика как наука, поворачивающаяся лицом к проблемам регуляции функции геномов, должна решить большое количество актуальных задач, о которых много говорилось на конференции. Прежде всего, весьма мала доля компьютерных баз данных, содержащих информацию о регуляции функции геномов (среди общего количества имеющихся баз данных по молекулярной биологии и генетике). Это указывает на необходимость форсированной разработки "регуляторных баз данных" научными организациями мирового биологического сообщества. О том, насколько велико здесь отставание, свидетельствуют оценки, согласно которым лишь незначительная доля всей экспериментальной информации по регуляции экспрессии генов, доступной в научной литературе, собрана в компьютерных базах данных.

Весьма важной задачей является разработка универсальных способов представления экспериментальной информации по регуляции экспрессии генов в компьютерных базах данных. В настоящее время не существует ни единых стандартов по описанию данных, ни типичных моделей их представления. Решению этой задачи было посвящено большое количество докладов на конференции BGRS'98 (секция "Базы данных по регуляторным геномным последовательностям и регуляторным белкам").

Хотя в проблеме распознавания кодирующих частей генов в геномах про- и эукариот в настоящее время достигнут существенный прогресс, ее нельзя считать до конца решенной. Особенно большие сложности возникают при применении имеющихся методов к анализу сверхдлинных участков секвенированной геномной ДНК размером до сотен тысяч пар оснований. С учетом этого работа одной из секций конференции BGRS'98 "Распознавание генов и их анализ; общие проблемы компьютерного изучения геномов" была посвящена распознаванию кодирующих частей генов. На этой же секции рассматривались общие проблемы теоретического и компьютерного анализа геномов.

Исключительно актуальной для биоинформатики регуляции геномов является задача распознавания функциональных сайтов и регуляторных элементов. В настоящее время она еще далека от решения. Созданные к настоящему времени компьютерные методы еще далеки от способности точно распознавать регуляторные элементы и функциональные сайты во вновь секвенированной геномной ДНК. Рассмотрению подходов к решению этой задачи была посвящена секция "Регуляторные геномные последовательности: компьютерный анализ и распознавание, молекулярные механизмы функционирования".

Как указывалось выше, важнейшее значение для понимания механизмов регуляции геномов имеет понимание механизмов функционирования генных сетей. Для успешного решения этой задачи требуется создание баз данных по генным сетям, а также методов компьютерного моделирования их функции. Последняя задача является исключительно сложной с учетом того, что в

функционировании генных сетей принимают участие многие сотни генов, мРНК, белков. Фактически здесь мы сталкиваемся с необходимостью моделирования сверхбольших динамических нелинейных открытых систем. Стандартных подходов к решению этой задачи в настоящее время не существует. Именно поэтому секция “Генные сети: базы данных, компьютерный анализ, моделирование метаболических путей” являлась одной из наиболее интересных на конференции BGRS’98.

Сравнительный анализ геномов является одним из наиболее распространенных и эффективных методов изучения их структурно-функциональной организации. Именно поэтому в настоящее время создано огромное разнообразие подходов к сравнительному анализу геномной ДНК. Однако с учетом исключительной важности этого направления и огромных объемов информации, которая требует сравнительного анализа, необходимо создание более эффективных и высокопроизводительных методов сравнительного анализа геномов. Успешное решение проблемы осложняется также тем, что требуется осуществлять сравнительный анализ нуклеотидных последовательностей сверхбольшой длины (до десятков миллионов пар оснований). С учетом этого в рамках BGRS’98 работала секция “Сравнительное и эволюционное изучение геномов”. Внимание к этой секции было обусловлено также и тем, что эволюционная идея является одним из “краеугольных камней” современной биологии. В связи с этим естественен интерес исследователей не только к особенностям структурно-функциональной организации геномов, но также к механизмам, темпам и направленности их эволюции.

Белки являются одним из трех классов генетических макромолекул, относясь наряду с ДНК и РНК к числу базовых типов строительных элементов, на основе которых функционируют молекулярно-генетические системы организмов. В ходе секвенирования геномной ДНК выявляется огромное количество генов, кодирующих белки. В связи с этим к числу ключевых проблем биоинформатики относится успешное решение проблемы предсказания пространственной структуры белков по их аминокислотным последовательностям. Рассмотрению этой и других проблем, связанных со структурной организацией белков, была посвящена секция “Анализ структуры белков”.

Следует отметить, что биоинформатика по своей природе является интегративной наукой, возникшей на стыке экспериментальных наук (молекулярной биологии, молекулярной генетики, генетической и белковой инженерии, биотехнологии и т.д.), и теоретических наук (теоретическая биология, информатика, теория баз данных, математические методы моделирования сложных систем, теория анализа данных и принятия решений, химическая кинетика, термодинамика, молекулярная физика и т.д.). Вот почему было естественным включение в программу конференции такой секции, как “Биоинформатика и экспериментальная молекулярная биология”. Задача этой секции состояла в построении своеобразного моста между системами взглядов, принятыми в экспериментальных и теоретических науках.

Наконец, в рамках конференции работала секция “Биоинформатика и образование”. На этой секции рассматривались проблемы создания информационно-компьютерных систем, предназначенных для обучения. С учетом очень быстрого накопления экспериментальных данных по структурно-функциональной организации геномов и огромных объемов этой информации задача знакомства с этой информацией в ходе учебного процесса является нетривиальной. Создание специализированных учебных курсов, основанных на профессиональных разработках, созданных в области биоинформатики, может помочь решению этой задачи.

По каждому из этих направлений были сделаны пленарные и симпозиальные доклады, были проведены стендовые сессии, компьютерные демонстрации и круглые столы. В целом на конференции было сделано более чем 50 докладов. Их дополнили 80 стендовых сообщений, в которых детально описывались специальные базы данных (компьютерные программы и модели).

Выбор Новосибирска как места проведения конференции обусловлен тем, что в Новосибирском Академгородке имеются давние и устойчивые традиции в области теоретической генетики, математической биологии, биологической кибернетики, компьютерного моделирования биологических систем и процессов. Они были заложены еще в 60- 70-е годы в результате работ таких выдающихся ученых, как А.А.Ляпунов, М.Г.Колпаков, И.А.Полетаев и В.А.Патнер.

Логика развития науки привела к тому, что в настоящее время большинство направлений, продолжающих эти исследования, сконцентрированы в Институте цитологии и генетики СО РАН, где имеется несколько лабораторий, работающих в области биоинформатики, с общим количеством ученых, превышающим 50 человек. Несмотря на общеизвестные трудности, которые переживает российская наука, ИЦиГ СО РАН сумел не только сохранить традиции и подходы к решению проблем теоретической и компьютерной биологии, заложенные “отцами-основателями”, но умел их развить в соответствии с новейшими тенденциями молекулярной биологии и генетики. Именно признание весомого вклада ученых ИЦиГ СО РАН в современную биоинформатику обусловило исключительно высокую представительность конференции BGRS’98.

В организационный/программный комитет конференции входили известные ученые из 10 стран, работающие в области биоинформатики и экспериментальной молекулярной биологии и генетики, в том числе:

Nikolay Kolchanov, Institute of Cytology and Genetics, Novosibirsk, Russia (председатель оргкомитета конференции);

Victor Solovyev, The Sanger Centre, Cambridge, UK (сопредседатель);

Lev Kisselev, Engelhardt Institute of Molecular Biology, Moscow, Russia;

Vladimir Tumanian, Engelhardt Institute of Molecular Biology, Moscow, Russia;

Martin Bishop, Human Genome Mapping Project Resource Centre (UK);

George Michaels, George Mason University, USA (координатор со стороны США);

Philip Bourne, SDSC, San-Diego, USA;

Soren Brunak, Center for Biological Sequence Analysis, Denmark;

Philipp Bucher, Swiss Institute for Experimental Cancer Research, Switzerland;

Jean-Michel Claverie, Structural & Genetic Information (CNRS), France;

Julio Collado-Vides, National University of Mexico, Mexico;

Luciano Milanesi, ITBA, Milan, Italy;

Christian Overton, Center for Bioinformatics University of Pennsylvania USA;

Michael Zhang Cold Spring Harbor Laboratory, Cold Spring Harbor, USA;

Bruce Roe, University of Oklahoma, USA;

Masaru Tomita, Bioinformatics Laboratory of Keio University, Japan;

Edgar Wingender, GBF, Braunschweig, Germany.

На конференции было зарегистрировано 120 ученых. Наряду с российскими учеными в ней участвовали исследователи из Японии (8 человек), США (6 человек), Германии (5 человек), Израиля (3 человека), Франции (3 человека), Великобритании (2 человека), Италии (2 человека) и Южной Африки, Сингапура, Китая, Швейцарии, Швеции (по одному человеку).

К числу отличительных особенностей конференции BGRS'98 относится высокий уровень компьютерных демонстраций. Имевшееся техническое обеспечение позволило всем участникам конференции провести Интернет-демонстрации своих разработок в реальном времени.

Полностью программу конференции, ее труды и другую информацию о конференции можно найти на WWW-сайте конференции <http://bgrs.bionet.nsc.ru>.

Сообщения участников конференции объемом до 4-6 страниц опубликованы в Трудах конференции (два тома, общий объем 450 страниц).

Статьи, отобранные оргкомитетом (около 30), в настоящее время находятся в редподготовке для публикации в специальном номере международного журнала "BIOINFORMATICS".

В постановлении конференции отмечается, что ее тематика является очень актуальной. Принято решение о проведении второй международной конференции по биоинформатике и структуре генома. Она будет проведена либо в 1999 году в Великобритании (Кембридж) на базе Сенджеровского центра по изучению генома человека, либо в 2000 году, вновь на базе ИЦиГ СО РАН. Отмечается также, что регулярное проведение этой конференции особенно актуально для исследований, проводимых в рамках геномных проектов, в том числе Российского национального проекта "Геном человека".

*Н.А.Колчанов*, д.б.н., проф.,  
председатель оргкомитета конференции BGRS'98,  
ИЦиГ СО РАН, Новосибирск