

Структурная и функциональная геномика

Структурная и функциональная геномика являются двумя тесно связанными направлениями современной молекулярной биологии, которые изучают геном на разных уровнях. Если классическая генетика часто рассматривает отдельные гены и признаки, то геномика исследует генетическую систему организма в целом. Ее развитие стало возможным благодаря технологиям секвенирования, молекулярного клонирования, ПЦР, биоинформатики и созданию крупных генетических баз данных.

Под геномом понимают совокупность генетического материала организма. В случае человека основная часть генома представлена ядерной ДНК, организованной в хромосомы, а дополнительный небольшой геном находится в митохондриях. Геном содержит не только белок-кодирующие гены, но и многочисленные некодирующие, регуляторные, повторяющиеся и структурные последовательности.

Геномика изучает строение генома, его организацию, функционирование, изменчивость и взаимодействие его компонентов. В зависимости от поставленной задачи выделяют структурную, функциональную, сравнительную, популяционную, медицинскую и другие направления геномики. В рамках данной темы особенно важно понимать различие между структурной и функциональной геномикой.

Структурная геномика отвечает прежде всего на вопрос: «Как устроен геном?» Она направлена на определение последовательности ДНК, расположения генов и других элементов, построение карт генома и изучение его физической организации.

Функциональная геномика отвечает на другой вопрос: «Как работает геном?» Она исследует активность генов, их регуляцию, взаимодействие и влияние на клеточные процессы и фенотип.

Поэтому эти направления можно кратко сопоставить следующим образом:

структурная геномика → что находится в геноме и где это расположено;
функциональная геномика → когда, где и каким образом генетическая информация используется клеткой.

Исторически структурная геномика начала активно развиваться с появлением методов секвенирования ДНК. Возможность определять последовательность нуклеотидов позволила перейти от исследования отдельных генов к анализу целых геномов. Особенно важным этапом стал международный проект Human Genome Project, в рамках которого была создана основа референсной последовательности генома человека.

Одной из основных задач структурной геномики является секвенирование генома. Исследователь получает большое количество последовательностей ДНК, после чего они подвергаются биоинформатической обработке и сборке. Современные технологии позволяют исследовать геномы различных организмов — от бактерий и вирусов до растений, животных и человека.

После получения последовательности необходимо определить, какие функциональные элементы в ней присутствуют. Этот процесс называют аннотацией генома. В ходе аннотации выявляются гены, экзоны, интроны, регуляторные области, повторяющиеся последовательности и другие элементы.

Таким образом, получение последовательности генома еще не означает полного понимания его структуры. Последовательность необходимо правильно собрать, сопоставить с известными данными и аннотировать.

Общий процесс структурно-геномного исследования можно представить следующим образом:

биологический образец → выделение ДНК → секвенирование → получение прочтений → контроль качества → сборка или выравнивание → получение геномной

последовательности → аннотация → карта генома → структурная характеристика генома.

При секвенировании *de novo* исследователь получает большое количество отдельных прочтений — *reads*. Биоинформатические алгоритмы находят перекрывающиеся участки и объединяют их в более длинные последовательности, называемые контигами — *contigs*. В дальнейшем контиги могут объединяться в более крупные структуры — *scaffolds*. Конечной целью является максимально полное восстановление организации исследуемого генома.

Современные технологии длинных прочтений значительно улучшили возможности сборки геномов. Длинные молекулы помогают преодолевать сложные повторяющиеся участки, которые трудно правильно собрать только из коротких фрагментов.

Важным результатом структурной геномики является создание геномной карты. Геномная карта описывает расположение генов, маркеров и других элементов относительно друг друга или относительно физической последовательности ДНК.

Исторически различали генетические и физические карты. Генетические карты основывались на частоте рекомбинации между маркерами. Чем чаще два локуса разделяются при рекомбинации, тем больше считается расстояние между ними. Физические карты, напротив, характеризуют реальное расположение последовательностей в молекуле ДНК.

С появлением полногеномного секвенирования стало возможным определять координаты генов непосредственно в последовательности генома. Сегодня для большинства хорошо исследованных организмов информация представляется в виде цифровых геномных карт.

В структурной геномике большое значение имеет понятие референсного генома. Это стандартная собранная последовательность, которая используется как координатная система. Референс позволяет указать, в какой хромосоме и в какой позиции расположен определенный ген или генетический вариант.

Однако референсный геном нельзя воспринимать как единственный «нормальный» геном. Между людьми существуют миллионы вариантов последовательности. Референс используется прежде всего как общая система координат, позволяющая сравнивать результаты различных исследований.

Структурная геномика также изучает различные формы генетической вариабельности. Наиболее простым примером являются одонуклеотидные варианты — *SNV*, а распространенные варианты этого типа часто рассматриваются как *SNP*. Кроме них существуют небольшие вставки и делеции — *indels*, изменения числа копий — *CNV*, а также более крупные структурные перестройки.

К структурным вариантам могут относиться крупные делеции, дупликации, инверсии и транслокации. Такие изменения иногда охватывают значительные участки генома и могут влиять сразу на несколько генов или регуляторных элементов.

Поэтому современная структурная геномика анализирует геном не только как линейную последовательность А, Т, G и С, но и как сложную систему, в которой важны количество копий участков, ориентация последовательностей и пространственная организация хромосом.

Еще одним важным направлением стало изучение пангеномов. Один линейный референс не способен полностью отражать генетическое разнообразие вида. Пангеномные подходы стремятся представить последовательности и структурные варианты, встречающиеся у множества индивидуумов или штаммов. Это особенно важно при исследовании популяционного разнообразия.

Структурная геномика тесно связана с сравнительной геномикой. Если последовательности двух организмов известны, можно сравнить расположение и структуру их генов, определить консервативные области и выявить участки, которые изменялись в ходе эволюции.

Например, если определенная последовательность сохраняется у многих эволюционно удаленных организмов, это может указывать на ее функциональную важность. Однако такое предположение требует дальнейшего функционального подтверждения.

Именно здесь структурная геномика постепенно переходит в функциональную.

Если структурная геномика отвечает на вопрос «что присутствует в геноме?», функциональная геномика исследует, «что делают эти элементы и как они взаимодействуют?»

Геном практически одинаков в большинстве клеток одного многоклеточного организма, однако клетки выполняют разные функции. Например, нейрон и клетка печени содержат в основном одну и ту же генетическую информацию, но используют различные наборы генов.

Следовательно, клеточный фенотип определяется не только наличием определенных генов, но и тем, какие гены активны, в какой степени они экспрессируются и как регулируется их активность.

Одним из основных объектов функциональной геномики является транскриптом — совокупность молекул РНК, образующихся в клетке или ткани при определенных условиях. В отличие от генома, который относительно стабилен, транскриптом является динамичным.

Если изменить условия окружающей среды, стадию развития или физиологическое состояние клетки, набор и количество РНК могут измениться. Поэтому анализ транскриптома позволяет исследовать функциональное состояние клетки.

Одним из главных методов современной функциональной геномики является RNA-seq. В этом подходе исследуют РНК, преобразуют ее в форму, пригодную для секвенирования, получают большое количество последовательностей и с помощью биоинформатики определяют, какие транскрипты представлены в образце.

Упрощенная логика RNA-seq выглядит следующим образом:

биологический образец → выделение РНК → подготовка библиотеки → секвенирование → получение reads → контроль качества → сопоставление с генами и транскриптами → количественная оценка → статистическое сравнение → функциональная интерпретация.

Особенно широко используется анализ дифференциальной экспрессии генов. Исследователь сравнивает две или несколько групп и определяет, какие гены характеризуются статистически различающимся уровнем экспрессии.

Например, можно сравнить клетки до и после определенного воздействия. Если активность группы генов возрастает после воздействия, можно предположить их участие в клеточном ответе. Однако статистическая ассоциация сама по себе не всегда доказывает причинную роль, поэтому результаты часто требуют дополнительных экспериментов.

В функциональной геномике используется также количественная ПЦР в реальном времени — qPCR. Она особенно удобна, когда необходимо исследовать ограниченное количество заранее выбранных генов. RNA-seq обеспечивает более широкий транскриптомный анализ, а qPCR часто применяется для целевых исследований и подтверждения отдельных результатов.

Функциональная геномика не ограничивается анализом РНК. Следующим уровнем является протеомика — изучение совокупности белков клетки, ткани или организма.

Это важно потому, что количество РНК не всегда прямо соответствует количеству или активности белка. После транскрипции происходят трансляция, сворачивание белков, транспорт, деградация и различные посттрансляционные модификации. Поэтому анализ только уровня мРНК не дает полного представления о состоянии клеточной системы.

Еще одним уровнем является метаболомика, изучающая низкомолекулярные продукты клеточного обмена. Изменение активности генов и белков в конечном итоге может приводить к изменению концентраций метаболитов.

Таким образом, современную функциональную геномику удобно рассматривать в контексте нескольких взаимосвязанных уровней:

геном → транскриптом → протеом → метаболом → клеточный фенотип.

Такой комплексный подход называют multi-omics, или мультиомным анализом. Его задача заключается в объединении информации нескольких уровней для более полного понимания биологических процессов.

Одной из важных задач функциональной геномики является определение функции неизвестного гена. Сам факт наличия гена в геноме еще не объясняет, какую роль он выполняет.

Один из классических подходов заключается в нарушении функции гена и последующем наблюдении за изменением фенотипа. Если изменение определенного гена приводит к изменению конкретного клеточного процесса, это дает информацию о возможной функции данного гена.

Другой подход заключается в повышении или восстановлении активности гена и анализе возникающих изменений. Особенно убедительными являются исследования, в которых функциональная гипотеза подтверждается несколькими независимыми методами.

Современные технологии CRISPR значительно расширили возможности функциональной геномики. С помощью таких систем можно направленно изменять определенные гены или регулировать их активность, а затем исследовать последствия.

Особенно интересны CRISPR-скрининги, позволяющие параллельно исследовать функции большого количества генов. Вместо изучения одного гена можно анализировать сотни или тысячи генетических мишеней и определять, какие из них влияют на исследуемый фенотип.

Функциональная геномика также изучает регуляторные элементы генома. Не вся функционально значимая ДНК кодирует белки. Большое значение имеют промоторы, энхансеры и другие последовательности, регулирующие активность генов.

Чтобы понять работу генома, необходимо определить не только расположение гена, но и механизмы, которые контролируют его включение и выключение.

Здесь большое значение имеет эпигеномика. Она изучает изменения состояния хроматина и химические модификации ДНК и связанных с ней белков, которые влияют на активность генов без обязательного изменения нуклеотидной последовательности.

Одним из наиболее известных эпигенетических механизмов является метилирование ДНК. Также большое значение имеют модификации гистонов и организация хроматина.

Поэтому современное представление о работе генома можно выразить следующим образом:

последовательность ДНК + регуляторные элементы + состояние хроматина + транскрипционные факторы → характер экспрессии генов → клеточная функция.

Для изучения взаимодействия белков с определенными областями генома используются специальные методы. Например, технологии семейства ChIP-seq позволяют исследовать распределение определенных белков или модификаций хроматина по геному.

Методы исследования доступности хроматина помогают определить области генома, находящиеся в более открытом состоянии и потенциально доступные для регуляторных белков.

Еще более сложным уровнем является изучение трехмерной организации генома. ДНК внутри ядра не располагается как вытянутая линейная нить. Она формирует сложную пространственную структуру, благодаря которой удаленные в линейной последовательности участки могут физически сближаться.

Это имеет функциональное значение. Например, регуляторный элемент может находиться на значительном линейном расстоянии от гена, но благодаря пространственной организации хроматина взаимодействовать с его регуляторной областью.

Следовательно, современная функциональная геномика изучает геном не только как последовательность, но и как динамическую трехмерную систему.

Большой шаг вперед связан с развитием одноклеточной геномики — single-cell technologies. Традиционные методы часто анализируют смесь миллионов клеток. Полученный результат представляет собой усредненный сигнал.

Однако ткань может содержать множество различных типов и состояний клеток. Если анализировать их вместе, важные различия могут быть скрыты.

Технологии single-cell RNA-seq позволяют исследовать транскриптом отдельных клеток. В результате можно выявлять клеточные субпопуляции и изучать, как экспрессия генов различается между отдельными клетками.

Следующим направлением стала пространственная транскриптомика. Она позволяет исследовать экспрессию генов с сохранением информации о расположении клеток или областей внутри ткани.

Таким образом, исследователь получает не только информацию о том, какие гены активны, но и где именно в ткани наблюдается эта активность.

Современная функциональная геномика тесно связана с биоинформатикой. После RNA-seq или других высокопроизводительных экспериментов исследователь получает огромный набор данных. Его необходимо обработать, нормализовать, статистически сравнить и биологически интерпретировать.

Например, после анализа экспрессии могут быть обнаружены сотни генов, активность которых различается между группами. Простого списка генов недостаточно. Необходимо понять, какие биологические процессы объединяют эти гены.

Для этого используется функциональная аннотация, Gene Ontology, анализ метаболических и сигнальных путей и другие биоинформатические подходы.

Логика функционального анализа может выглядеть следующим образом:

список измененных генов → функциональная аннотация → группировка по биологическим процессам → анализ сигнальных и метаболических путей → формирование биологической гипотезы → экспериментальная проверка.

Особенно важно понимать, что структурная и функциональная геномика не существуют независимо друг от друга. Чтобы исследовать функцию гена, сначала необходимо знать его структуру, последовательность и положение в геноме. С другой

стороны, знание последовательности без функционального анализа дает лишь ограниченное понимание биологии организма.

Предположим, в геноме обнаружен определенный вариант. Структурная геномика позволяет установить его координату и определить, в каком гене или регуляторном участке он расположен. Функциональная геномика позволяет исследовать, связан ли этот вариант с изменением экспрессии, структуры транскрипта или других клеточных процессов.

Таким образом, полный анализ может выглядеть следующим образом:

геномная последовательность → обнаружение варианта → определение его положения → анализ регуляторного или кодирующего эффекта → исследование экспрессии → анализ белка и клеточной функции → оценка фенотипического значения.

Одной из важнейших областей применения геномики является медицина. Структурная геномика позволяет обнаруживать генетические варианты, а функциональная помогает понять их возможное биологическое действие.

Однако обнаружение генетического варианта еще не означает доказательства его роли в заболевании. Для интерпретации необходимы популяционные данные, функциональные исследования, статистические доказательства и другие источники информации.

В онкологии структурная геномика используется для обнаружения соматических мутаций и структурных перестроек опухолевого генома. Функциональная геномика позволяет исследовать, какие гены и сигнальные пути изменяют свою активность в опухолевых клетках.

В микробиологии структурная геномика позволяет секвенировать геномы микроорганизмов, определять набор генов и сравнивать штаммы. Функциональная геномика помогает выяснять, какие гены активируются при различных условиях среды и какие из них участвуют в адаптации.

В сельскохозяйственной биотехнологии геномика используется для изучения генетического разнообразия растений и животных, поиска маркеров и исследования генов, связанных с хозяйственно важными признаками.

В экологических исследованиях геномные подходы позволяют анализировать генетическое разнообразие природных популяций и микробных сообществ.

Большое значение имеет также сравнительная геномика. Сравнивая геномы различных организмов, можно находить гомологичные гены, консервативные последовательности и специфические генетические элементы.

Если два гена у разных организмов имеют сходную последовательность, это может указывать на общее эволюционное происхождение. Однако степень сходства последовательности не всегда напрямую определяет идентичность функции, поэтому необходим функциональный анализ.

Современная геномика постепенно переходит от изучения отдельных генов к анализу генных сетей. Гены редко действуют полностью независимо. Продукт одного гена может регулировать другой ген, белки могут взаимодействовать между собой, а метаболические пути образуют сложные сети.

Поэтому функциональная геномика стремится понять не только функцию отдельного гена, но и его место в общей клеточной системе.

В этом контексте важным понятием является системная биология. Она рассматривает клетку как совокупность взаимосвязанных молекулярных сетей и использует данные геномики, транскриптомики, протеомики и метаболомики для построения целостной модели.

Развитие искусственного интеллекта и машинного обучения также влияет на геномику. Такие методы используются для анализа больших наборов генетических данных, выявления закономерностей и интеграции нескольких типов информации. Однако вычислительный прогноз должен рассматриваться как часть исследования и требует соответствующей проверки.

Для студентов полезно четко различать основные уровни анализа:

ДНК отвечает на вопрос: какая генетическая информация присутствует?

РНК показывает: какие гены транскрибируются?

Белки показывают: какие функциональные молекулы образуются?

Метаболиты отражают: какие биохимические процессы происходят?

Фенотип показывает: каким становится конечный биологический результат?

Следовательно, современное исследование может объединять несколько уровней:

геном → эпигеном → транскриптом → протеом → метаболом → фенотип.

Такой подход значительно информативнее анализа одного показателя, но одновременно создает большие объемы данных и требует сложной статистической и биоинформатической обработки.

Если представить структурную и функциональную геномику как единый исследовательский процесс, получится следующая последовательность:

биологический образец → выделение нуклеиновых кислот → секвенирование → сборка и аннотация генома → определение генов и вариантов → анализ экспрессии → исследование регуляции → анализ белков и метаболических путей → интеграция данных → функциональная интерпретация → экспериментальное подтверждение → биологический вывод.

Структурная геномика дает своеобразную карту генома, а функциональная геномика объясняет, как элементы этой карты работают.

Это различие можно запомнить очень просто:

Структурная геномика: «Что? Где? В какой последовательности?»

Функциональная геномика: «Когда? Где активно? Как регулируется? Что делает?»

При этом граница между ними условна. Современные исследования часто одновременно используют полногеномное секвенирование, RNA-seq, эпигеномные методы, протеомику и биоинформатический анализ.

В результате формируется системное представление о геноме не как о статичном наборе генов, а как о динамической регулируемой системе.

Главная идея структурной геномики заключается в установлении последовательности, расположения и организации элементов генома. Главная идея функциональной геномики — в определении активности, регуляции, взаимодействия и биологической роли этих элементов.

Таким образом, структурная и функциональная геномика дополняют друг друга. Первая позволяет прочесть и картировать генетическую информацию, а вторая помогает понять, каким образом эта информация реализуется в клетке и формирует биологические свойства организма.

Для будущего специалиста по биотехнологии важно понимать общую логику:

структура генома → генетическая вариабельность → регуляция генов → экспрессия → белки → метаболические процессы → клеточная функция → фенотип.

Именно переход от описания последовательности ДНК к пониманию функции генома является одной из центральных задач современной геномики.