

Введение в ДНК-технологии: история развития, основные понятия и современные направления

ДНК-технологии являются одним из важнейших направлений современной молекулярной биологии и биотехнологии. Они объединяют совокупность методов, позволяющих выделять ДНК из биологических образцов, определять ее количество и качество, амплифицировать отдельные участки, разделять фрагменты, устанавливать последовательность нуклеотидов, выявлять генетические варианты, клонировать гены, исследовать геномы и в определенных случаях направленно изменять генетический материал. Современные ДНК-технологии находятся на пересечении генетики, молекулярной биологии, биохимии, микробиологии, биоинформатики и биотехнологии и широко используются в медицине, фармацевтике, сельском хозяйстве, экологии, криминалистике и фундаментальных исследованиях.

Чтобы понять современные ДНК-технологии, необходимо обратиться к истории изучения наследственного материала. Одним из первых важнейших открытий стало исследование швейцарского ученого Фридриха Мишера. В 1869 году при изучении клеточных ядер он обнаружил неизвестное вещество, богатое фосфором, которое назвал «нуклеином». Впоследствии было установлено, что это вещество связано с тем, что сегодня мы называем нуклеиновыми кислотами. Однако в XIX веке ученые еще не знали, что именно ДНК выполняет функцию хранения наследственной информации.

Важным этапом стал эксперимент Фредерика Гриффита, проведенный в 1928 году с бактериями *Streptococcus pneumoniae*. Гриффит обнаружил явление трансформации: определенные свойства одного бактериального штамма могли передаваться другому. Он предположил существование некоторого «трансформирующего фактора», однако его химическая природа оставалась неизвестной. В 1944 году Освальд Эвери, Колин Маклауд и Маклин Маккарти получили данные, показывающие, что этим трансформирующим фактором является ДНК. В 1952 году Альфред Херши и Марта Чейз в экспериментах с бактериофагами дополнительно подтвердили генетическую роль ДНК.

Следующий фундаментальный этап произошел в 1953 году, когда Джеймс Уотсон и Фрэнсис Крик предложили модель двойной спирали ДНК. При разработке модели существенное значение имели рентгенодифракционные исследования, связанные с работами Розалинд Франклин и Мориса Уилкинса, а также уже известные закономерности состава азотистых оснований. Модель двойной спирали позволила объяснить, каким образом молекула ДНК может хранить и копировать генетическую информацию.

Дезоксирибонуклеиновая кислота представляет собой полимер, состоящий из нуклеотидов. Каждый нуклеотид включает три основных компонента: остаток фосфорной кислоты, сахар дезоксирибозу и одно из четырех азотистых оснований — аденин, тимин, гуанин или цитозин. Аденин и гуанин относятся к пуриновым основаниям, а тимин и цитозин — к пиримидиновым.

Две полинуклеотидные цепи ДНК соединены между собой благодаря комплементарному взаимодействию оснований. Аденин преимущественно образует пару с тимином, а гуанин — с цитозином. Поэтому последовательность одной цепи позволяет определить последовательность комплементарной цепи. Например, если одна цепь имеет последовательность

5'-ATGCCGTA-3',

то комплементарная ей цепь будет

3'-TACGGCAT-5'.

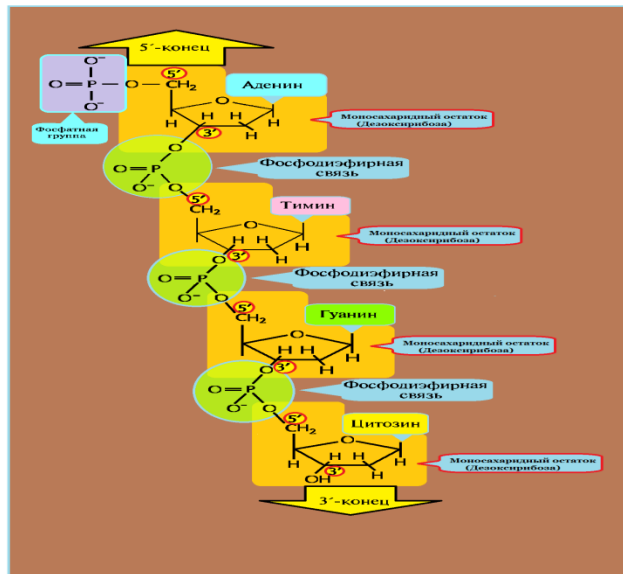


Рисунок 1 : ДНК — строение одной цепочки нуклеотидов

Цепи ДНК имеют противоположную ориентацию и называются антипараллельными. Одна ориентирована в направлении 5'→3', другая — 3'→5'. Понимание направления цепей особенно важно при изучении репликации ДНК, проектировании праймеров, проведении полимеразной цепной реакции и анализе результатов секвенирования.

Одним из центральных понятий молекулярной генетики является ген. Ген можно определить как участок ДНК, функционально связанный с образованием определенного продукта — РНК или, через промежуточную РНК, белка. При этом ген не следует рассматривать только как участок, непосредственно кодирующий белок. В геноме присутствует большое количество некодирующих и регуляторных последовательностей, играющих важную роль в контроле активности генов.

Совокупность генетического материала организма называется геномом. ДНК эукариот организована в хромосомы и связана с белками. Определенное положение генетической последовательности на хромосоме называется локусом. Различные варианты последовательности одного локуса могут называться аллелями.

Для понимания функционирования генетической информации используется концепция центральной догмы молекулярной биологии, которую в упрощенном виде можно представить следующим образом:

ДНК → РНК → белок.



Рисунок 2 : Центральная догма молекулярной биологии

При репликации происходит копирование ДНК. При транскрипции информация ДНК используется для синтеза РНК. При трансляции последовательность информационной РНК считывается рибосомой и используется для синтеза полипептидной цепи. Однако современное понимание передачи и реализации генетической информации значительно сложнее этой упрощенной схемы. В клетке существуют различные типы некодирующих РНК, альтернативный сплайсинг, многочисленные механизмы регуляции экспрессии генов, эпигенетические процессы и обратная транскрипция.

Большое значение для развития ДНК-технологий имело открытие ферментов, способных работать с нуклеиновыми кислотами. Особую роль сыграли рестрикционные эндонуклеазы, или рестриктазы. Эти ферменты способны распознавать определенные последовательности ДНК и осуществлять разрезание молекулы в определенных участках. Другой важный фермент — ДНК-лигаза — позволяет соединять фрагменты ДНК.

Использование рестриктаз и лигаз стало основой технологии рекомбинантной ДНК. В 1970-х годах исследователи получили возможность соединять фрагменты ДНК различного происхождения. Например, интересующий ген можно встроить в бактериальную плазмиду. Плазмиды представляют собой небольшие молекулы ДНК, способные реплицироваться в бактериальных клетках. Модифицированная плазида может использоваться как вектор для переноса и размножения определенного фрагмента ДНК.

Общую схему молекулярного клонирования можно представить следующим образом:

интересующий фрагмент ДНК → получение или амплификация фрагмента → подготовка вектора → соединение фрагмента с вектором → получение рекомбинантной молекулы → введение конструкции в клетки → отбор клеток → размножение → анализ полученных клонов.

Технология рекомбинантной ДНК стала основой современной биотехнологической промышленности. Благодаря ей стало возможно получать различные рекомбинантные белки с использованием клеточных систем.

Следующей революцией в молекулярной биологии стало появление полимеразной цепной реакции — PCR, или ПЦР. Развитие PCR в 1980-х годах связано прежде всего с работами Кэри Муллиса. Главная идея PCR заключается в избирательном многократном копировании определенного участка ДНК *in vitro*.

Для классической PCR необходимы матричная ДНК, два праймера, термостабильная ДНК-полимераза, дезоксирибонуклеозидтрифосфаты — dNTP, соответствующий буфер и ионы, необходимые для работы фермента. Реакция проводится в приборе, называемом термоциклером.

Каждый цикл PCR включает три основных стадии. Первая стадия — денатурация, при которой при повышенной температуре двухцепочечная ДНК разделяется на отдельные цепи. Вторая стадия — отжиг, при котором температура снижается и праймеры связываются с комплементарными участками матрицы. Третья стадия — элонгация, когда ДНК-полимераза удлиняет праймер и синтезирует новую цепь ДНК в направлении 5'→3'.

Цикл повторяется много раз. В идеализированной модели при стопроцентной эффективности количество целевого продукта может увеличиваться приблизительно в два раза за цикл. Теоретически это можно выразить формулой:

$$N = N_0 \times 2^n,$$

где N_0 — исходное число копий, а n — количество циклов. Например, 2^{30} составляет приблизительно 1,07 миллиарда. Однако реальная PCR не сохраняет стопроцентную эффективность на протяжении всех циклов, поэтому это значение является теоретическим.

Существуют различные модификации PCR. Классическая PCR используется для амплификации выбранных участков ДНК. Multiplex PCR позволяет исследовать несколько мишеней в одной реакции. Reverse-transcription PCR начинается с РНК, которая сначала превращается в комплементарную ДНК — кДНК. Real-time PCR, или qPCR, позволяет регистрировать накопление продукта в ходе реакции с помощью флуоресцентного сигнала. Сочетание обратной транскрипции и количественной PCR часто обозначается как RT-qPCR.

После проведения классической PCR полученные фрагменты часто анализируют методом гель-электрофореза. ДНК благодаря фосфатным группам имеет отрицательный заряд и в электрическом поле перемещается через матрицу геля. Более короткие фрагменты обычно перемещаются через гель быстрее, чем более длинные. Для определения приблизительного размера используют DNA ladder — набор фрагментов ДНК известной длины. Размер фрагментов ДНК обычно выражают в парах оснований — bp.

Однако наличие PCR-продукта еще не сообщает исследователю полную последовательность нуклеотидов. Для определения последовательности ДНК используется секвенирование.

Одним из наиболее известных методов является секвенирование по Сэнгеру, разработанное Фредериком Сэнгером и его коллегами в 1970-х годах. Метод основан на синтезе ДНК с использованием специальных терминаторов цепи. Современные автоматизированные системы позволяют регистрировать флуоресцентные сигналы и представлять результат в форме электроферограммы. Последовательность определяется по порядку зарегистрированных оснований. Метод Сэнгера до сих пор широко используется для анализа отдельных генов, проверки PCR-продуктов и подтверждения определенных генетических вариантов.

Развитие методов секвенирования подготовило условия для одного из крупнейших международных научных проектов — Human Genome Project. Проект «Геном человека» официально начался в 1990 году. В 2003 году было объявлено о завершении основной стадии проекта. Он позволил создать фундаментальную основу для изучения последовательности человеческого генома и значительно ускорил развитие геномики и биоинформатики.

Следующий технологический скачок связан с появлением секвенирования нового поколения — Next-Generation Sequencing, или NGS. Главное отличие NGS от классического подхода Сэнгера заключается в возможности параллельно анализировать огромное количество фрагментов нуклеиновых кислот.

Общая схема NGS включает получение биологического материала, выделение ДНК или РНК, оценку качества материала, подготовку библиотеки, секвенирование, получение большого количества коротких или, в зависимости от платформы, более длинных последовательностей, последующий биоинформатический анализ и биологическую интерпретацию.

С использованием NGS можно выполнять полногеномное секвенирование — Whole Genome Sequencing, WGS. При этом анализируется практически весь доступный геном. Whole Exome Sequencing, WES, направлено преимущественно на исследование белок-кодирующих экзонов. Targeted sequencing используется для анализа заранее выбранной группы генов или геномных участков. RNA-seq позволяет исследовать

транскриптом — совокупность молекул РНК в клетке или ткани при определенных условиях.

Одним из важных объектов современных генетических исследований являются генетические варианты. Наиболее распространенным типом вариабельности являются однонуклеотидные варианты. Термин SNP — Single Nucleotide Polymorphism — используется для однонуклеотидных полиморфизмов, распространенных в популяции.

Например, в определенной позиции последовательности у части людей может находиться гуанин G, а у других — аденин A. В таком случае возможны генотипы GG, GA и AA. GG и AA являются гомозиготными генотипами, а GA — гетерозиготным.

SNP могут находиться в экзонах, интронах, промоторных и других регуляторных участках, а также в межгенных областях. Некоторые варианты практически не влияют на фенотип, тогда как другие могут изменять структуру белка, уровень экспрессии гена или быть статистически связаны с определенными биологическими признаками.

Процесс определения генетических вариантов конкретного организма называется генотипированием. Для этого могут применяться PCR-RFLP, аллель-специфическая PCR, TaqMan-анализ, секвенирование по Сэнгеру, SNP-микрочипы и NGS.

Помимо SNP существуют другие типы генетических маркеров и вариантов. К ним относятся короткие tandemные повторы STR, вариабельные tandemные повторы VNTR, вставки и делеции InDel, а также изменения числа копий участков генома — CNV.

Одним из наиболее значительных достижений молекулярной биологии XXI века стало развитие технологий направленного редактирования генома. Наиболее известной системой является CRISPR-Cas. В природе CRISPR-системы являются компонентом адаптивной защиты бактерий и архей от чужеродного генетического материала. Впоследствии элементы этих систем были адаптированы для биотехнологических задач.

В широко известной системе CRISPR-Cas9 направляющая РНК помогает комплексу распознавать комплементарный участок ДНК, расположенный рядом с соответствующей PAM-последовательностью. Белок Cas9 может осуществить разрыв ДНК. После этого клеточные системы репарации восстанавливают поврежденный участок. При репарации могут возникать изменения последовательности либо, при соответствующем экспериментальном дизайне, может использоваться матрица для направленной репарации.

Два часто обсуждаемых механизма репарации — NHEJ, нехомологичное соединение концов, и HDR, гомологически направленная репарация. На основе CRISPR были разработаны и более специализированные технологии. Base editing позволяет в определенных системах изменять отдельные основания без классического двуцепочечного разрыва ДНК. Prime editing расширяет возможности внесения заданных изменений. CRISPRi используется для подавления активности определенных генов, а CRISPRa — для повышения их экспрессии.

ДНК-технологии сегодня имеют огромное значение для медицины. Они используются при диагностике наследственных заболеваний, молекулярной классификации опухолей, обнаружении генетических вариантов, фармакогенетических исследованиях и разработке некоторых подходов генной терапии.

Особое направление представляет персонализированная, или прецизионная, медицина. Ее идея заключается в использовании совокупности клинических, молекулярных и других данных для более индивидуального подхода к диагностике и лечению. Фармакогенетика изучает, каким образом генетические варианты могут быть

связаны с различиями в эффективности и безопасности лекарственных препаратов у разных людей.

Большое значение ДНК-технологии приобрели в онкологии. Опухолевые клетки накапливают генетические и эпигенетические изменения. Использование PCR, qPCR, секвенирования, NGS и других методов позволяет выявлять некоторые из этих изменений. Отдельным направлением является анализ циркулирующей опухолевой ДНК — ctDNA. Фрагменты опухолевой ДНК могут присутствовать в плазме крови. Анализ таких молекул входит в более широкое направление, называемое жидкостной биопсией.

ДНК-технологии широко используются и в сельском хозяйстве. Генетические маркеры позволяют проводить marker-assisted selection — маркер-ассоциированную селекцию. В этом случае отбор растений или животных может основываться не только на наблюдаемом фенотипе, но и на наличии определенных генетических маркеров. Современное сельское хозяйство также использует геномную селекцию, молекулярную диагностику патогенов, трансгенные технологии и редактирование генома.

Еще одной областью применения является криминалистическая генетика. Для идентификации биологических образцов часто исследуют STR-маркеры. Типичная схема включает получение образца, выделение и количественную оценку ДНК, амплификацию набора STR-локусов, разделение продуктов и получение генетического профиля. При этом совпадение генетических профилей требует корректной статистической интерпретации.

ДНК-технологии также существенно изменили экологические исследования. Одним из перспективных подходов является environmental DNA, или eDNA. Организмы постоянно оставляют генетический материал в окружающей среде. Фрагменты ДНК можно обнаруживать в воде, почве, донных отложениях и других образцах. Анализ eDNA позволяет проводить мониторинг биоразнообразия, обнаруживать редкие или инвазивные виды и исследовать экосистемы без необходимости непосредственного обнаружения каждого организма.

С этим направлением тесно связана метагеномика. В классической микробиологии исследование микроорганизма часто связано с его культивированием. Однако многие микроорганизмы трудно выращивать в стандартных лабораторных условиях. Метагеномный подход позволяет выделить совокупную ДНК непосредственно из образца и исследовать генетический материал всего микробного сообщества. Метагеномика широко применяется при изучении микробиома человека, почвы, воды и других экологических систем.

Развитие NGS сделало биоинформатику обязательной составляющей современных ДНК-технологий. Секвенаторы способны генерировать огромные объемы данных, которые невозможно полноценно анализировать вручную. Поэтому после лабораторной части начинается компьютерная обработка.

Типичный процесс анализа данных секвенирования можно представить следующим образом:

получение исходных последовательностей → контроль качества → фильтрация → выравнивание на референсный геном или сборка последовательностей → поиск генетических вариантов → аннотация → статистический анализ → биологическая интерпретация.

Одним из классических биоинформатических инструментов является BLAST — Basic Local Alignment Search Tool. Он позволяет сравнивать исследуемую последовательность с последовательностями, содержащимися в базах данных. С помощью BLAST можно определить, на какие известные последовательности похож

исследуемый фрагмент, оценить степень сходства и получить информацию о потенциальной принадлежности последовательности к определенному гену или организму.

Современная геномика все активнее взаимодействует с искусственным интеллектом и машинным обучением. Эти методы используются для анализа больших массивов генетической информации, классификации биологических объектов, поиска сложных закономерностей, анализа экспрессии генов и интеграции геномных данных с другими типами информации. Поэтому современный специалист в области ДНК-технологий должен обладать не только лабораторными навыками, но и базовым пониманием статистики, программирования и биоинформатики.

Важно понимать, что современное ДНК-исследование представляет собой не отдельную лабораторную реакцию, а последовательный научный процесс. Сначала формулируется исследовательский вопрос. Затем выбирается биологический материал, определяется дизайн исследования, проводится выделение ДНК, оцениваются ее количество и качество, выбирается подходящий молекулярный метод, выполняется эксперимент, после чего полученные результаты подвергаются статистическому и биоинформатическому анализу. Завершающими этапами являются биологическая интерпретация, проверка воспроизводимости и формулирование научных выводов.

Выбор метода всегда должен зависеть от поставленной задачи. Если необходимо определить наличие определенного участка ДНК, можно использовать PCR. Если требуется установить генотип известного SNP, можно применить TaqMan-анализ, PCR-RFLP или другой валидированный метод генотипирования. Если необходимо определить последовательность отдельного PCR-продукта, удобно использовать секвенирование по Сэнгеру. Для одновременного исследования большого числа генов могут применяться таргетные NGS-панели. Для исследования белок-кодирующих областей используется WES, а для более широкого исследования генома — WGS. Если необходимо исследовать экспрессию большого количества генов, применяется RNA-seq.

Следовательно, один из главных принципов современной молекулярной биологии заключается в том, что сначала определяется научный вопрос, а затем выбирается технология, способная дать необходимую информацию.

При этом возможности современных ДНК-технологий не следует переоценивать. Каждый метод имеет ограничения. Качество результата зависит от качества биологического материала, правильности выделения ДНК, отсутствия контаминации, специфичности праймеров и реагентов, качества оборудования, дизайна исследования, статистических методов и корректности интерпретации.

Особенно важно различать статистическую ассоциацию и причинно-следственную связь. Если определенный SNP чаще встречается в одной исследуемой группе, это само по себе еще не доказывает, что именно этот вариант непосредственно вызывает наблюдаемый фенотип. Необходимы независимые выборки, функциональные исследования и учет возможных смешивающих факторов.

Еще одной важнейшей проблемой является биоэтика. Генетические данные человека являются чувствительной информацией. При проведении исследований необходимо учитывать информированное согласие участников, конфиденциальность, защиту персональных данных, правила хранения биологических образцов и генетической информации.

Особенно серьезные этические вопросы возникают в области редактирования генома человека. Необходимо различать исследования и терапевтические подходы, направленные на соматические клетки, и изменения зародышевой линии, которые

потенциально могут передаваться последующим поколениям. Последний случай вызывает особенно сложные научные, этические и правовые дискуссии.

Современная молекулярная биология постепенно переходит от исследования отдельных генов к комплексному анализу биологических систем. Геномика изучает геном, транскриптомика — совокупность РНК, протеомика — белки, метаболомика — метаболиты. Объединение нескольких уровней называется multi-omics.

Общую концепцию можно представить следующим образом:

ДНК → РНК → белки → метаболиты → клеточные процессы → фенотип.

Комплексный анализ этих уровней позволяет лучше понимать механизмы функционирования организма и развитие патологических процессов.

К современным направлениям ДНК-технологий также относится long-read sequencing — секвенирование длинных молекул ДНК. Более длинные прочтения позволяют лучше исследовать сложные участки генома, структурные варианты и повторяющиеся последовательности.

Еще одним важным направлением является single-cell sequencing — анализ отдельных клеток. Традиционное исследование ткани часто дает усредненный результат для миллионов клеток. Однако ткань состоит из различных клеточных популяций. Анализ отдельных клеток позволяет исследовать эту гетерогенность.

Следующим этапом является spatial omics, включая пространственную транскриптомику. Эти технологии позволяют исследовать молекулярные характеристики клеток с учетом их расположения непосредственно в ткани. Это особенно важно для изучения опухолей, развития органов и взаимодействия различных клеточных популяций.

Таким образом, развитие ДНК-технологий можно представить как последовательный исторический процесс. В 1869 году было открыто вещество, которое впоследствии связали с нуклеиновыми кислотами. В первой половине XX века была установлена генетическая роль ДНК. В 1953 году была предложена модель двойной спирали. В 1970-х годах получили развитие технологии рекомбинантной ДНК и секвенирование. В 1980-х годах появилась PCR. Проект «Геном человека» значительно ускорил развитие геномики. В XXI веке широкое распространение получили NGS, высокопроизводительная биоинформатика и технологии редактирования генома. Сегодня активно развиваются long-read sequencing, single-cell genomics, spatial omics, multi-omics и применение искусственного интеллекта.

В результате современную ДНК-технологию можно представить как целую систему:

биологический образец → выделение ДНК → контроль количества и качества → PCR или другой метод подготовки → генотипирование или секвенирование → получение молекулярных данных → биоинформатический анализ → статистическая обработка → биологическая интерпретация → подтверждение результата → научный или практический вывод.

Для будущего специалиста по биотехнологии особенно важно понимать, что владение ДНК-технологиями не означает простое умение выполнить PCR или запустить секвенатор. Необходимо понимать биологическую основу метода, правильно сформулировать исследовательский вопрос, выбрать подходящую технологию, использовать необходимые положительные и отрицательные контроли, оценить ограничения метода и корректно интерпретировать полученные результаты.

Таким образом, ДНК-технологии прошли путь от фундаментального открытия молекулы наследственности до высокопроизводительного анализа целых геномов и направленного редактирования отдельных генетических локусов. Они стали одной из основных технологических платформ современной биологии и биотехнологии и

продолжают быстро развиваться. Знание принципов PCR, электрофореза, генотипирования, секвенирования, NGS, биоинформатики и геномного редактирования является фундаментальной основой подготовки современного биотехнолога.

Главный вывод данной лекции можно сформулировать следующим образом: современная ДНК-технология — это не один лабораторный метод, а комплексный процесс, объединяющий получение биологического материала, молекулярный эксперимент, анализ данных и биологическую интерпретацию. При этом выбор конкретной технологии всегда определяется научным вопросом, который исследователь хочет решить.