**Тема: Методы секвенирования 1го поколения.**

**Цель:** Ознакомиться с тероретическими основами Сэнгеровского секвенирования.

**Основные вопросы, выносимые на обсуждение семинара.**

1. «Плюс-минус» метод.
2. Метод «обрыва цепи».
3. Компоненты реакционных смесей и их функции.
4. Анализ данных Сэнгеровского секвенирования.

**Краткое содержание занятия:**

Методы расшифровки нуклеотидной последовательности нуклеиновых кислот в отечественной литературе принято называть методами **секвенирования**.

Еще в 50-е годы прошлого века были разработаны методы, позволяющие определять последовательность аминокислот в полипептидной цепи. Теоретически это несложно, поскольку все аминокислоты, встречающиеся в природных белках, имеют разные свойства.

*(Суть метода заключается в обработке исследуемого пептида определенным набором реагентов, что приводит к отщеплению одной аминокислоты с N-конца последовательности. Циклическое повторение реакции и анализ продуктов реакций дают информацию о последовательности аминокислот в пептиде)*

Поэтому, когда был расшифрован генетический код, появилась возможность восстанавливать нуклеотидную последовательность транскрибируемой ДНК по аминокислотной последовательности соответствующего белка. Однако генетический код является вырожденным. Следовательно, первичная структура ДНК, полученная на основе анализа последовательности аминокислот, не является однозначной.

К концу 60-х годов Ф.Сэнгером были разработан метод секвенирования РНК, получаемой с ДНК-матрицы при помощи РНК-полимеразы. Применив этот способ, Ш.Вейссман и У.Фирс смогли концу 1976 г. определить последовательность более половины молекулы ДНК SV40, длина которой превышает 5200 нуклеотидных пар. Следующим шагом должна была стать разработка методов прямого секвенирования ДНК.

Быстрые методы секвенирования, несомненно, приведут к более быстрому открытию одного нуклеотидного полиморфизма (SNP). Метод Сэнгера служил для определения последовательности генома с 1977 года. В настоящее время ученые сосредоточены на разработке новых методов секвенирования, которые позволят составить "генетический паспорт", и станут рутинной процедурой в исследовании человеческих генов. Такие технологии должны быть улучшены в трех направлениях: чтение длинных последовательностей, высокая пропускная способность, низкая стоимость. Если будет достигнут прогресс в этой области, то большие участки генома, а затем целые геномов людей будут полностью изучены.

**Секвенирование ДНК по Сенгеру.**

(Sanger F., Coulson A.R. A rapid method for determining sequences in DNA by primed syntesis with DNA polymerase)

|  |
| --- |
| Ферментативное секвенирование ДНК: плюс-минус метод |

Первым методом прямого ферментативного секвенирования ДНК стал метод, предложенный Ф. Сэнгером и Д. Коулсоном в 1975 г.

В качестве матрицы в реакции полимеразного копирования использовался одноцепочечный фрагмент ДНК. В качестве праймеров - синтетические олигонуклеотиды или природные субфрагменты, получаемые при гидролизе рестрицирующими эндонуклеазами, а в качестве фермента - фрагмент Кленова ДНК полимеразы I (PolI) из E.coli (*большой белковый фрагмент, образующийся при ферментативном расщеплении ДНК-полимеразы I из Escherichia coli протеазой субтилизином*).

Метод включал два этапа. Сначала в ограниченных условиях проводили полимеразную реакцию в присутствии всех четырех типов dNTP (один из них был мечен по альфа-положению фосфата), получая на выходе набор продуктов неполного копирования матричного фрагмента. Смесь очищали от несвязавшихся дезоксинуклеозидтрифосфатов и делили на восемь частей. После чего в "плюс" системе проводили четыре реакции в присутствии каждого из четырех типов нуклеотидов, а в "минус" системе - в отсутствие каждого из них. В результате, в "минус" системе терминация происходила перед dNTP данного типа, а в "плюс" системе - после него. Полученные таким образом восемь образцов разделяли с помощью электрофореза, "считывали" сигнал и определяли последовательность исходной ДНК. Этим способом была секвенирована короткая ДНК фага фХ174, состоящая из 5386 нуклеотидных пар.

**Секвенирование ДНК по Сэнгеру: метод "терминаторов"**

|  |
| --- |
| Ферментативное секвенирование ДНК: метод терминирующих аналогов трифосфатов |

В 1977 г. автор "плюс-минус" метода предложил еще один способ ферментативного секвенирования, получивший название метода терминирующих аналогов трифосфатов. Более мощный и более технологичный, этот способ, несколько модифицированный, применяется до сих пор. В основе метода тоже лежало ферментативное копирование с помощью фрагмента Кленова ДНК полимеразы I из E.coli. В качестве праймеров использовали синтетические олигонуклеотиды. Специфическую терминацию синтеза обеспечивали добавлением в реакционную смесь помимо четырех типов dNTP (один из которых был радиоактивно мечен по альфа положению фосфата) еще и одного из 2',3'-дидезоксинуклеозидтрифосфатов (ddATP, ddTTP, ddCTP или ddGTP), который способен включаться в растущую цепь ДНК, но не способен обеспечивать дальнейшее копирование из-за отсутствия 3'-ОН группы. Отношение концентраций dNTP/ddNTP авторы подбирали экспериментально, так, чтобы в итоге получить набор копий ДНК различной длины. Таким образом, для определения первичной структуры исследуемого фрагмента ДНК требовалось провести четыре реакции копирования: по одному типу терминаторов в каждой из реакций. После этого полученные продукты разгонялись в полиакриламидном геле на соседних дорожках и по расположению полос определялась последовательность нуклеотидов.

**Анализ данных Сэнгеровского секвенирования.**

|  |
| --- |
| Ферментативное секвенирование ДНК: плюс-минус метод |

Первым методом прямого ферментативного секвенирования ДНК стал метод, предложенный Ф. Сэнгером и Д. Коулсоном в 1975 г.

В качестве матрицы в реакции полимеразного копирования использовался одноцепочечный фрагмент ДНК. В качестве праймеров - синтетические олигонуклеотиды или природные субфрагменты, получаемые при гидролизе рестрицирующими эндонуклеазами, а в качестве фермента - фрагмент Кленова ДНК полимеразы I (PolI) из E.coli (*большой белковый фрагмент, образующийся при ферментативном расщеплении ДНК-полимеразы I из Escherichia coli протеазой субтилизином*).

Метод включал два этапа. Сначала в ограниченных условиях проводили полимеразную реакцию в присутствии всех четырех типов dNTP (один из них был мечен по альфа-положению фосфата), получая на выходе набор продуктов неполного копирования матричного фрагмента. Смесь очищали от несвязавшихся дезоксинуклеозидтрифосфатов и делили на восемь частей. После чего в "плюс" системе проводили четыре реакции в присутствии каждого из четырех типов нуклеотидов, а в "минус" системе - в отсутствие каждого из них. В результате, в "минус" системе терминация происходила перед dNTP данного типа, а в "плюс" системе - после него. Полученные таким образом восемь образцов разделяли с помощью электрофореза, "считывали" сигнал и определяли последовательность исходной ДНК. Этим способом была секвенирована короткая ДНК фага фХ174, состоящая из 5386 нуклеотидных пар.

**Секвенирование ДНК по Сэнгеру: метод "терминаторов"**

|  |
| --- |
| Ферментативное секвенирование ДНК: метод терминирующих аналогов трифосфатов |

В 1977 г. автор "плюс-минус" метода предложил еще один способ ферментативного секвенирования, получивший название метода терминирующих аналогов трифосфатов. Более мощный и более технологичный, этот способ, несколько модифицированный, применяется до сих пор. В основе метода тоже лежало ферментативное копирование с помощью фрагмента Кленова ДНК полимеразы I из E.coli. В качестве праймеров использовали синтетические олигонуклеотиды. Специфическую терминацию синтеза обеспечивали добавлением в реакционную смесь помимо четырех типов dNTP (один из которых был радиоактивно мечен по альфа положению фосфата) еще и одного из 2',3'-дидезоксинуклеозидтрифосфатов (ddATP, ddTTP, ddCTP или ddGTP), который способен включаться в растущую цепь ДНК, но не способен обеспечивать дальнейшее копирование из-за отсутствия 3'-ОН группы. Отношение концентраций dNTP/ddNTP авторы подбирали экспериментально, так, чтобы в итоге получить набор копий ДНК различной длины. Таким образом, для определения первичной структуры исследуемого фрагмента ДНК требовалось провести четыре реакции копирования: по одному типу терминаторов в каждой из реакций. После этого полученные продукты разгонялись в полиакриламидном геле на соседних дорожках и по расположению полос определялась последовательность нуклеотидов.

**Анализ данных массового параллельного секвенирования.**

Сборка генома — процесс объединения большого количества коротких фрагментов ДНК (ридов) в одну или несколько длинных последовательностей (контигов и скаффолдов) в целях восстановления последовательностей ДНК хромосом, из которых возникли эти фрагменты в процессе секвенирования.

Сборка генома является очень сложной вычислительной задачей, в частности, осложнённой тем, что геномы часто содержат большое количество одинаковых повторяющихся последовательностей (так называемые геномные повторы). Эти повторы могут быть длиной в несколько тысяч нуклеотидов, а также встречаться в тысяче различных мест в геноме. Особенно богаты повторами большие геномы растений и животных, в том числе геном человека.

Существует два подхода для сборки геномов — основанный на перекрытии overlap-layout-consensus (применяется для длинных фрагментов), а также основанный на графах де Брюйна (применяется для коротких фрагментов).

Overlap-Layout-Consensus

При секвенировании методом дробовика все ДНК организма сначала разрезают на миллионы маленьких фрагментов до 1000 нуклеотидов в длину. Затем алгоритмы сборки генома рассматривают полученные фрагменты одновременно, находя их перекрытия (overlap), объединяя их по перекрытиям (layout) и исправляя ошибки в объединённой строке (consensus). Данные шаги могут повторятся несколько раз в процессе сборки.

Данный подход был наиболее распространён для сборки геномов до появления секвенирования следующего поколения.

Графы де Брёйна

С развитием технологий секвенирования следующего поколения, получение фрагментов стало на порядок дешевле, но размер фрагментов стал меньше (до 150 нуклеотидов), а количество ошибок при чтении фрагментов увеличилось (до 3%). При сборке таких данных получили распространение методы [3], основанные на графах де Брёйна.