

Геномика и протеомика

Лекция 6. Протеомика (бакалавры)

Составитель: проф. М.Р. Шарипова

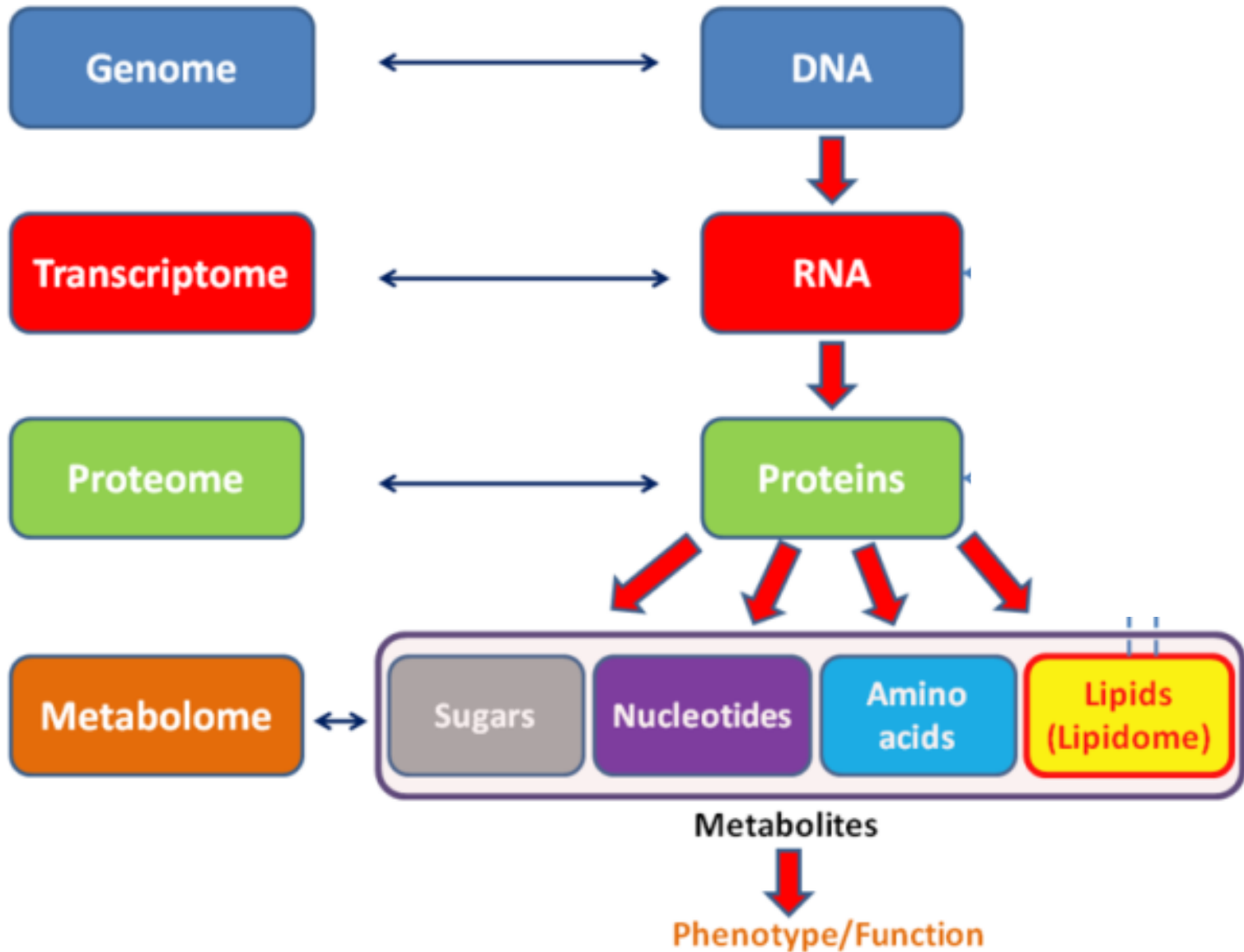
Основные определения

- **Протеом** – это белковый комплимент генома, набор белков, продуцируемых клеткой, тканью, организмом
- **Протеомика** – методы и подходы для характеристики всех белков, их содержания и взаимодействия в живых организмах
- **Цель протеомики** – определение и характеристика полного набора белков данного организма

Задачи протеомики

- **Подготовить полный каталог всех протеинов организма и расшифровать сеть их взаимодействий**
- **Систематизация знаний о белках, разработка их классификации**
- **Разработка методов и вычислительных средств для выявления механизмов функционирования белков**

Взаимосвязь омических технологий



Создание метаболических карт

Метаболические карты:

- содержат пути биосинтеза и деградациии аминокислот, нуклеотидов, углеводов и липидов
- дают полное представление о конкретных метаболических циклах, включая все промежуточные соединения
- химические структуры на карте соединены стрелками с указанием, какие ферменты (белки) участвуют в соответствующих химических превращениях

Химическое
соединение



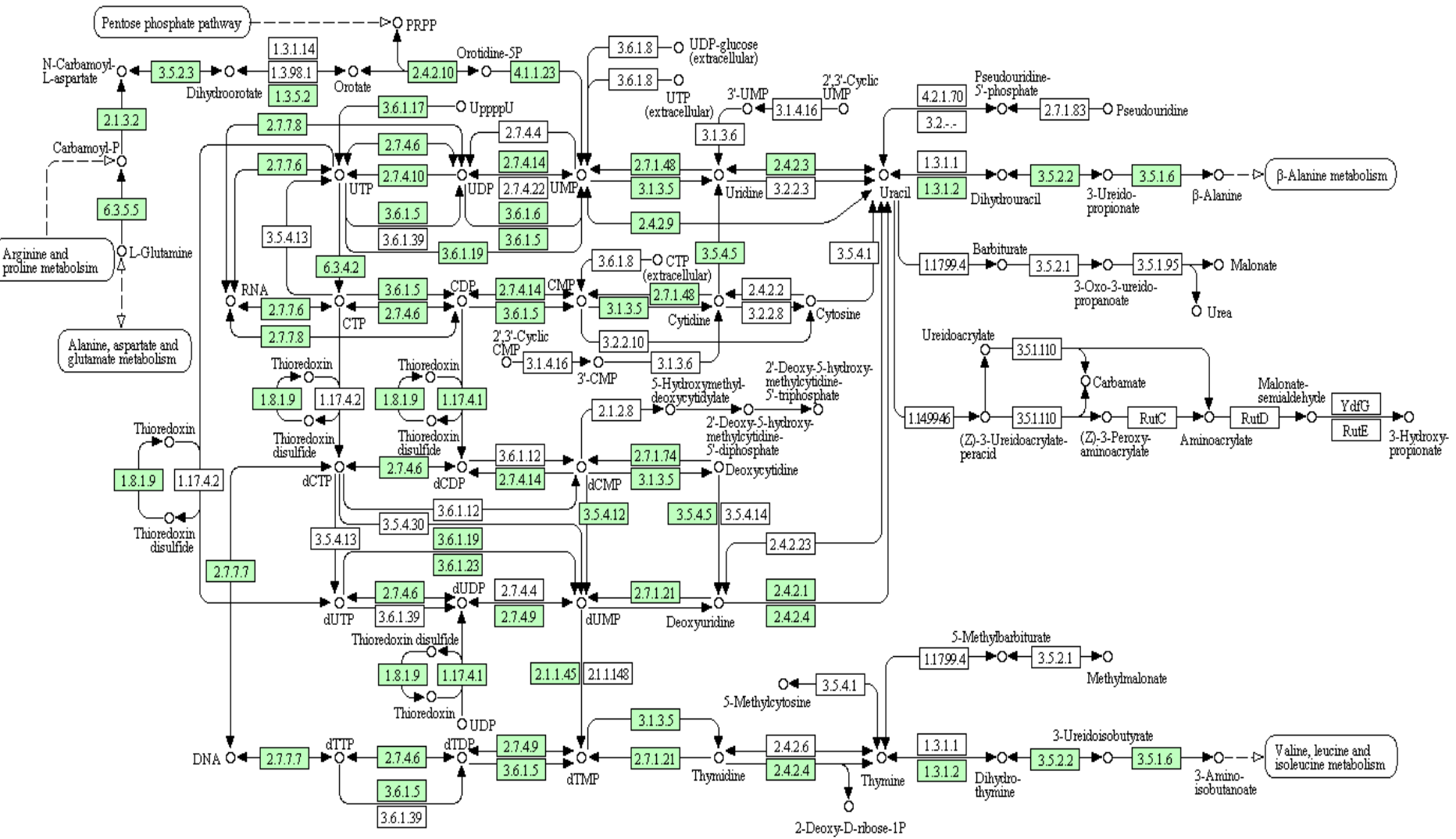
1.1.1.1.

Химическое
соединение

- для каждой биохимической реакции приводится классификационный код соответствующего фермента
- для реакций с участием коферментов приводятся названия коферментов

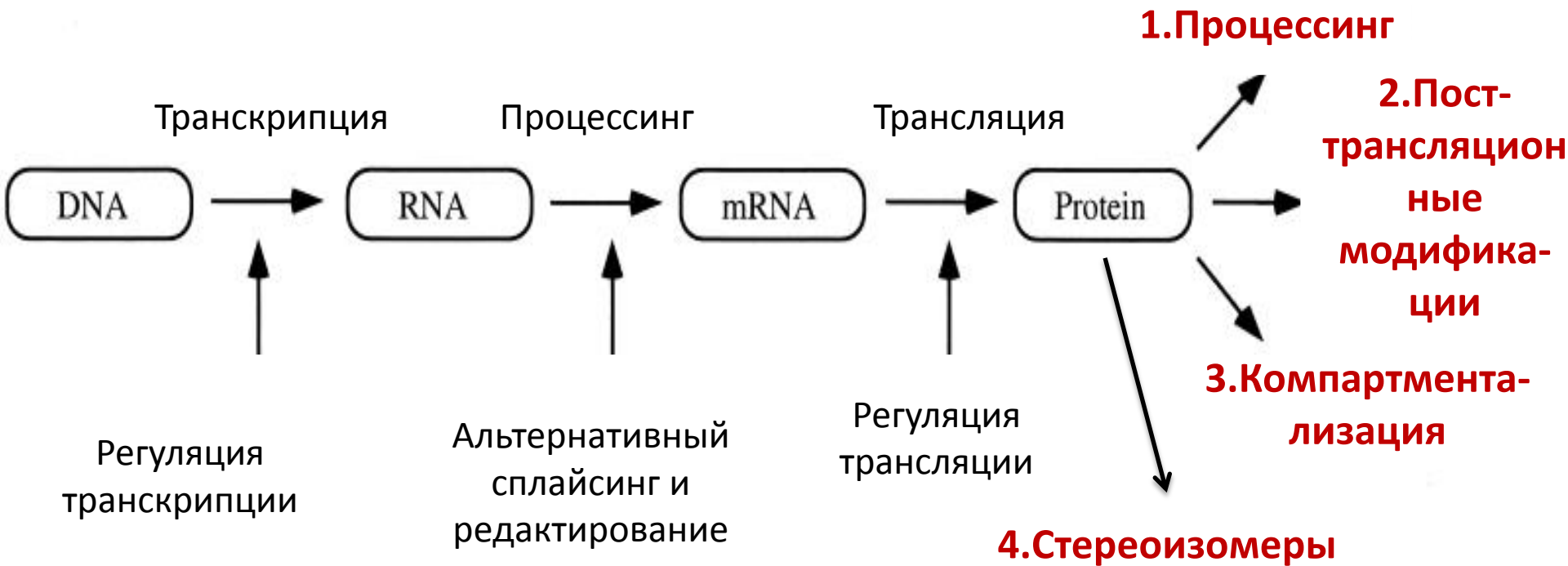
Карта метаболизма пиримидинов (сеть реакций с периодическим строением)

PYRIMIDINE METABOLISM



- Протеом – сложная и динамичная система, изменяется в процессе развития организма, а также при различных физиологических и патологических состояниях
- Если человеческий **протеом** включает до 2000000 белков на 20 000 генов, то в среднем это 100 белков на 1 ген
- Варьирует количество копий отдельных белков от 20 000 молекул на клетку (редкие) до 100 млн копий (распространенные)
- Белки, которые постоянно присутствуют в большинстве клеток – белки домашнего хозяйства (50 000 копий на клетку)

Основа динамичности протеома



Нет корреляции между наборами мРНК и белков: белки, синтезированные на рибосомах, подвергаются различным модификациям

ПОСТ-ТРАНСЛЯЦИОННЫЕ МОДИФИКАЦИИ

- ❑ присоединение различных функциональных групп (ацетил-, метил- и фосфатных групп)
 - ❑ присоединение углеводов, липидов и углеводородов (гликозилирование)
 - ❑ изменение стандартных аминокислот на нестандартные (образование цитруллина)
 - ❑ изменения в структуре белка (образование дисульфидных мостиков между цистеинами)
 - ❑ присоединение пептидов или небольших белков (убиквитинирование - добавление цепей из убиквитина в качестве сигнала для деградации этого белка)
-
- Описано более 100 типов посттрансляционных модификаций белков
 - Один белок может подвергаться разным модификациям (активно модифицируются гистоны)
 - Роль большинства модификаций пока не выяснена

Стереои́зомеры

- Белки обладают разнообразными **пространственными структурами**, которые нельзя установить по линейным последовательностям нуклеотидов или аминокислот
- Актуальной задачей остается выделение и определение структуры всех функционирующих белков
- **Определение 3D-структуры белка проводят методами рентгеноструктурного анализа и ЯМР**

Пептиды

- В динамичность протеома большой вклад вносят **пептиды** или олигопептиды, которые играют важную роль в жизнедеятельности организмов
- **Пептиды** постоянно синтезируются для регуляции физиологических процессов
- Свойства **пептидов** зависят от последовательности аминокислот, а также от её конфигурации в пространстве
- **Пептиды** характеризуются высокой скоростью обращения
- Введены термины «пептидом» и «пептидомика», которые являются частью протеома и протеомики

Пептидом – регулярная часть протеома

- Основа пептидома - пептиды, длиной до 50 аминокислотных остатков и обладающие специфическим спектром функциональной активности
- Для пептидома создан банк данных **EROP - Endogenous Regulatory OligoPeptides** – (эндогенные регуляторные олигопептиды)
- Расшифрована структура более 6000 олигопептидов, выделенных из разных организмов

- Среди пептидов чаще встречаются олигопептиды размером до 10 аминокислотных остатков
- В основном олигопептиды участвуют в регуляции нервной системы и называются нейропептидами
- С участием нервной системы осуществляются самые быстрые процессы в живом организме, поэтому пептидные регуляторы должны быть мобильными и небольшими

До сих пор не разработана строгая классификация белков и пептидов

Основной задачей биоинформатики является накопление и систематизация информации

Базы данных в протеомике

- Созданы **информационные базы данных**, в которые заносят все известные белки и пептиды
- Используют общие и специализированные **базы данных**, которые доступны в сети Интернет
- В **общих базах** содержатся сведения о всех известных белках живых организмов, т. е. о глобальном протеоме всего живого
- Общей базой является SwissProt - EMBL, она включает структуры более 200 000 белков, установленные аналитическими методами, и более 2 млн структур, которые определены на основе нуклеотидных последовательностей

PROSITE - по функционально значимым участкам белков

Prodom
по доменам белков

**RELIBASE
PROMISE**
по лиганд-рецепторным комплексам белков

PDB - по трехмерной структуре белков

**Специализированные
базы данных**

Protein Motion Database
по белковым субъединицам и доменам

Уровни функционирования генома

I. Биохимический уровень:

- Базовый
- Физиологический

II. Клеточный уровень

III. Организменный уровень

I. Биохимический уровень:

- **1) Базовые функции белков** – расшифровать все белки, закодированные в геноме и отнести их к определенным группам - ферменты, транспортные белки, структурные белки
- **2) Физиологические функции белков** –выяснить биологическую активность молекул или белковых систем (аппарат транскрипции, репликации, метаболизм и т.д.)

Количество генов в секвенированных геномах разных организмов с известной или предполагаемой функцией

Организм	Общее число генов	Гены с известной функцией	% генов, функция которых известна или предполагается
Человек (3000 Мб)	20 000	8000	40
Нематода (97Мб)	19 099	7000	40
Дрожжи (12 Мб)	6 034	3086	63
<i>E. coli</i> (4.6 Мб)	4 288	2656	62
<i>B. subtilis</i> (4.2Мб)	4 000	2320	58
<i>Haemophilus influenzae</i> (1.8Мб)	1 740	1015	58
<i>Helicobacter pylori</i> (1.7Мб)	1 590	907	57
<i>Mycoplasma genitalium</i> (0.6Мб)	470	324	69

I. Биохимический уровень:

На основании геномного анализа видно, что биохимический уровень функционирования генома сложен и неоднозначен:

- 1) В каждом секвенированном геноме есть гены с неизвестной функцией
- 2) Объем информации об известных функциях уменьшается с увеличением сложности генома
- 3) Функцию определяют на основании гомологии с другими известными белками

Это правило не всегда работает: *nif*-ген у бактерий участвует в фиксации азота, у дрожжей и других организмов выполняет другие функции

Гомология структур ≠ аналогия функции

II. Клеточный уровень функционирования:

- Единицей функционирования на этом уровне является клетка
- Отражает **внутри- и внеклеточные взаимодействия** между белками или белковыми системами:
 - Транспорт и секреция
 - Движение
 - Регуляторные пути
 - Цитоскелет
 - Метаболические циклы
 - Защита и хранение
- Взаимодействия между белками нельзя напрямую прочитать в геноме
- Они формируются аналитическим путем, включая модельные эксперименты *in vivo* и *in vitro*
- Проводят **кластерный анализ** – распределение белков по группам, выполняющим общую функцию

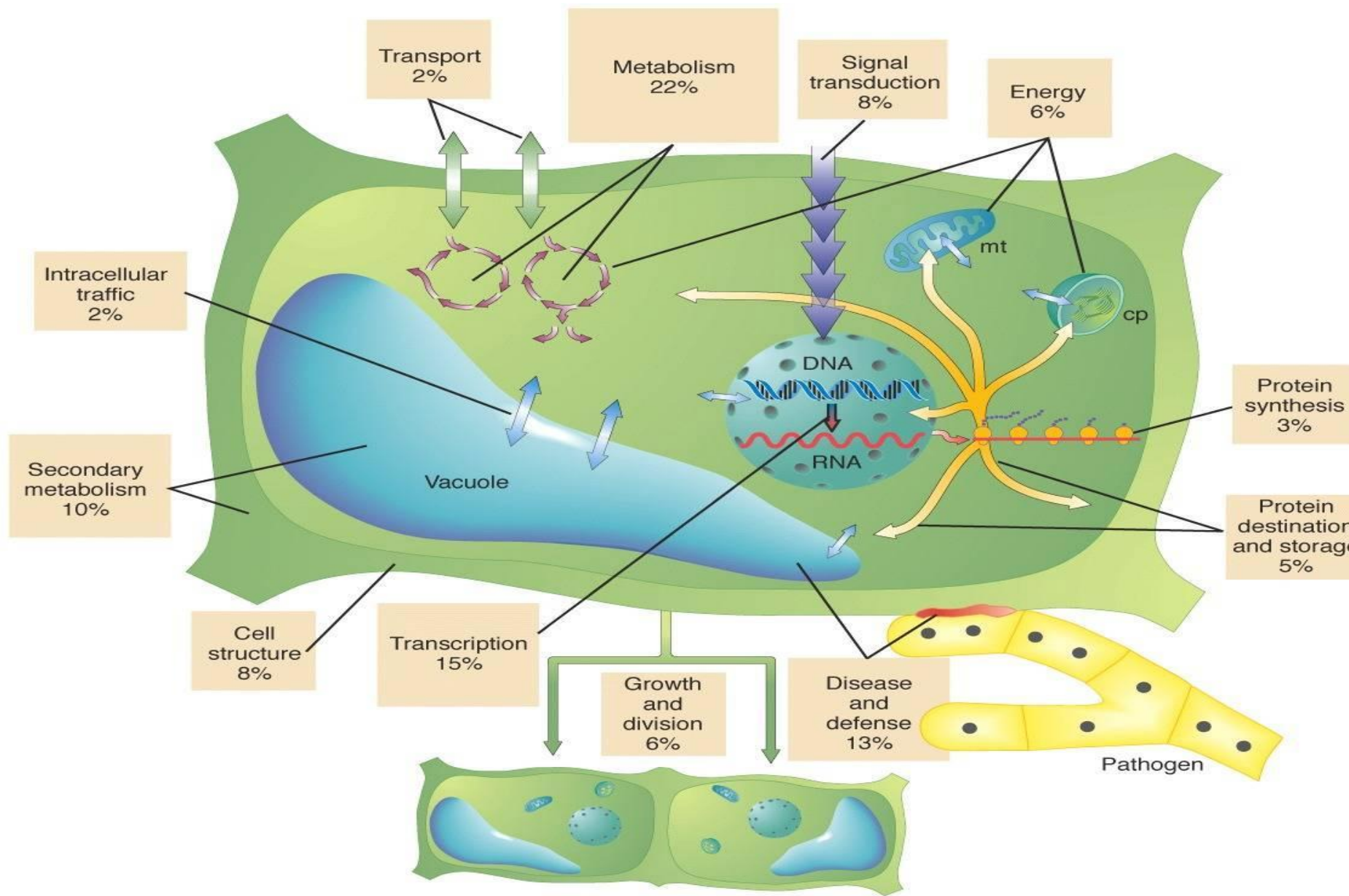
Кластерный анализ генома человека

Диаграмма протеома клетки человека

Кластеры (по функции)	Пропорции, %
ДНК-репликация, транскрипция и трансляция	22
Метаболизм	17
Клеточное деление	12
Защита	12
Сигнальные белки и регуляция	12
Структура	8
Неизвестные функции	17

Диаграмма отражает протеом клетки человека на основании полной геномной последовательности

Кластерный анализ генома арабидопсиса



III. Уровень функционирования органов и целого организма

- Это более сложно организованный уровень функционирования, чем клеточный
- Этот уровень пока остается недостижимым до понимания
- Пример – организация нервной системы:
 - Нематода (80 К-каналов, 90 лиганд-зависимых рецепторов, 1000 G-протеин-зависимых рецепторов – в передаче импульса участвуют 300 нейронов)
 - Пчела – 1 000 000 нейронов
 - Человек – 100 млрд нейронов

Заключение

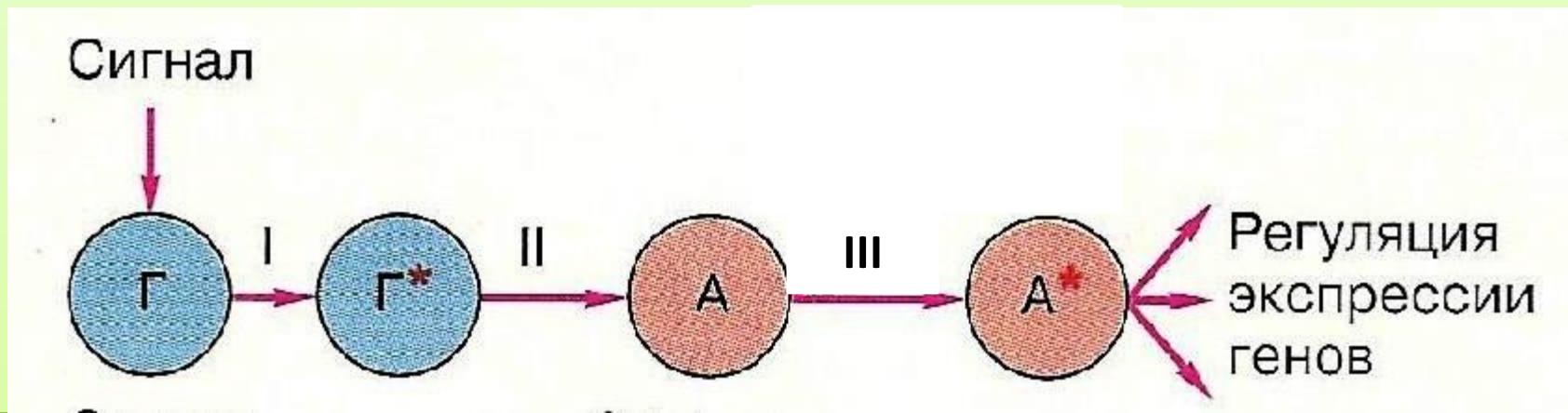
- Знание последовательности ДНК не несет информации о роли кодируемых белков в функционировании клетки
- На каждом уровне функционирования протеом динамичен
- Для понимания функционирования клетки задача анализа белков становится приоритетной
- **Основная цель** – использование данных о геномах для предсказания функций белков и выявления взаимосвязей между белками и другими молекулами

Типы взаимодействия генных продуктов

Продукты генов функционируют не обособленно, а во взаимосвязи с продуктами других генов:

1) Интегрирующие системы взаимодействия –

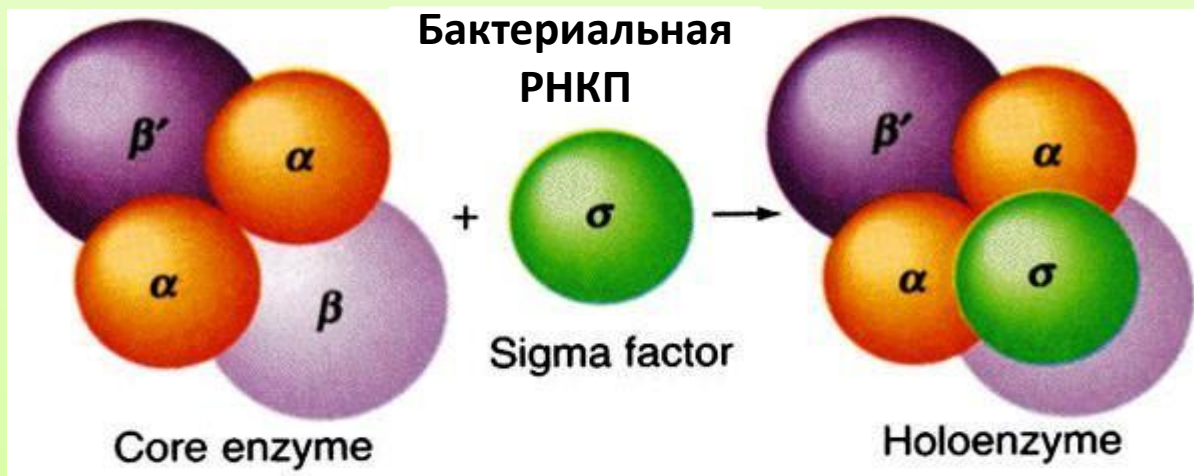
- такое взаимодействие нескольких белков приводит к одному результату
- продукты этих генов не функционируют по отдельности, а только в интеграции
- пример – сигнальные системы регуляторных путей



Типы взаимодействия генных продуктов

2) Субъединичное взаимодействие –

- это образование функционирующих мультиподъединичных комплексов
- от состава комплекса зависят его функции
- пример – РНК-полимеразы, проявление их активности определяется набором субъединиц
- **субъединичная комбинаторика значительно повышает функциональный потенциал протеома**



Типы взаимодействия генных продуктов

3) Множественные продукты одного гена –

- образование множества изоформ на матрице одного гена на основе механизмов альтернативного сплайсинга, редактирования, процессинга, модификации и др.
- Образуются белки, которые могут выполнять сходные функции и заменять друг друга, либо вступать в антагонистические отношения



Гены-
матрешки



Типы взаимодействия генных продуктов

4) Увеличение функциональной нагрузки –

- С увеличением сложности организма происходит увеличение функциональной нагрузки на единицу экспрессии
- Это явление называют **функциональной** **плейотропией** (один и тот же белок выполняет множество функций)
- Пример: цитокины, способны взаимодействовать с разными белками и при этом выполнять разную функцию
- **Основа функциональной плейотропии – усложнение способов регуляции экспрессии**

Типы взаимодействия генных продуктов

5) **Функциональная избыточность** –

- Это выполнение сходных функций белков, которые кодируют гены-паралоги
- Сложные геномы содержат семейства генов, которым соответствуют семейства белков. Формируется избыточность молекул, выполняющих сходную функцию
- Пример: опухолевые супрессоры (p53, p73, p40, p51), у мыши с нокаутом по p73 опухоль не развивается.
- **Избыточность приводит к тому, что генная мутация малозаметна на уровне фенотипа, обеспечивает страховку от вредных мутаций и является механизмом стабильности генома**

Типы взаимодействия генных продуктов

6) Эпистаз и эпигенез –

- Вклад в протеом вносят эпистатические и эпигенетические эффекты
- **Эпистаз** – взаимодействие генов, при котором один неаллельный ген интерферирует на уровне фенотипа с другим неаллельным геном
- **Эпигенез** – нет изменений на уровне генотипа, но есть на уровне фенотипа, изменения в хроматине, не затрагивающие ДНК, влияют на экспрессию.
- Пример: серповидно-клеточная анемия вызывает смерть в раннем возрасте, но человек с такой мутацией может оставаться здоровым, т.к. на ее проявление могут влиять другие генетические локусы

Заключение

- Большинство признаков являются сложными
- Для их формирования необходимо участие продуктов многих генов
- Поиск генов, вовлеченных в сложные признаки, является более трудоемким, чем картирование гена, и нуждается в разработке новых аналитических подходов

Методы протеомики

1) Характеристика протеома путем ORF-анализа

- Выявление ORF путем поиска старт-кодона и стоп-кодона
- Для эукариот предсказание происходит путем поиска экзонов
- Анализ ORF проводят путем поиска гомологии к известным генам по базам данных
- Последовательности отдельных кластеров или доменов, их локализация и ориентация, тоже являются важной информацией для идентификации

2) Нокауты и Нокдауны

- **Нокдаун** - методика понижения экспрессии гена
- **Нокаут** – инактивация гена, разрушают ген и анализируют фенотип

Эффекты нокаутов

Организм	Нокаут с изменением фенотипом	Нокаут без изменения фенотипа
Дрозофила	20	50
Мышь	20	70

- **Нокауты и нокдауны** редко дают однозначную информацию о функции гена, поскольку продукты экспрессии находятся в цепи сложных взаимодействий
- При нокаутах фенотип либо не меняется, либо дает множественные изменения
- Нокаут гена в разных линиях животных может приводить к разным эффектам
- Это значит, что выводы о функции генов нельзя однозначно переносить на другие организмы
- **Чем сложнее организм, тем неоднозначней нокаут**

Методы протеомики

3) Двумерный электрофорез (2D-электрофорез)

разделяет белки по 2-м характеристикам белковой молекулы:

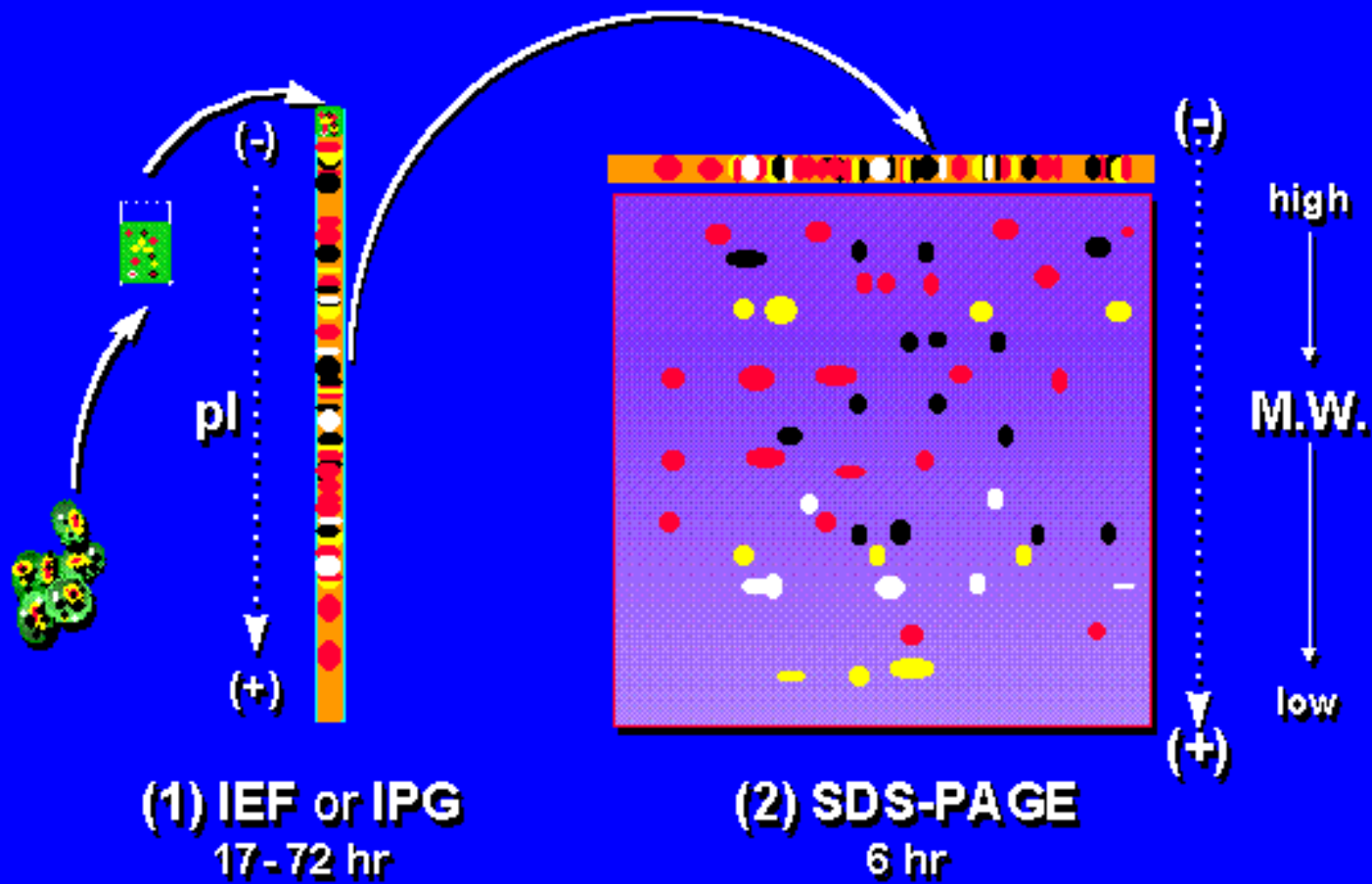
1 - молекулярной массе (M_r)

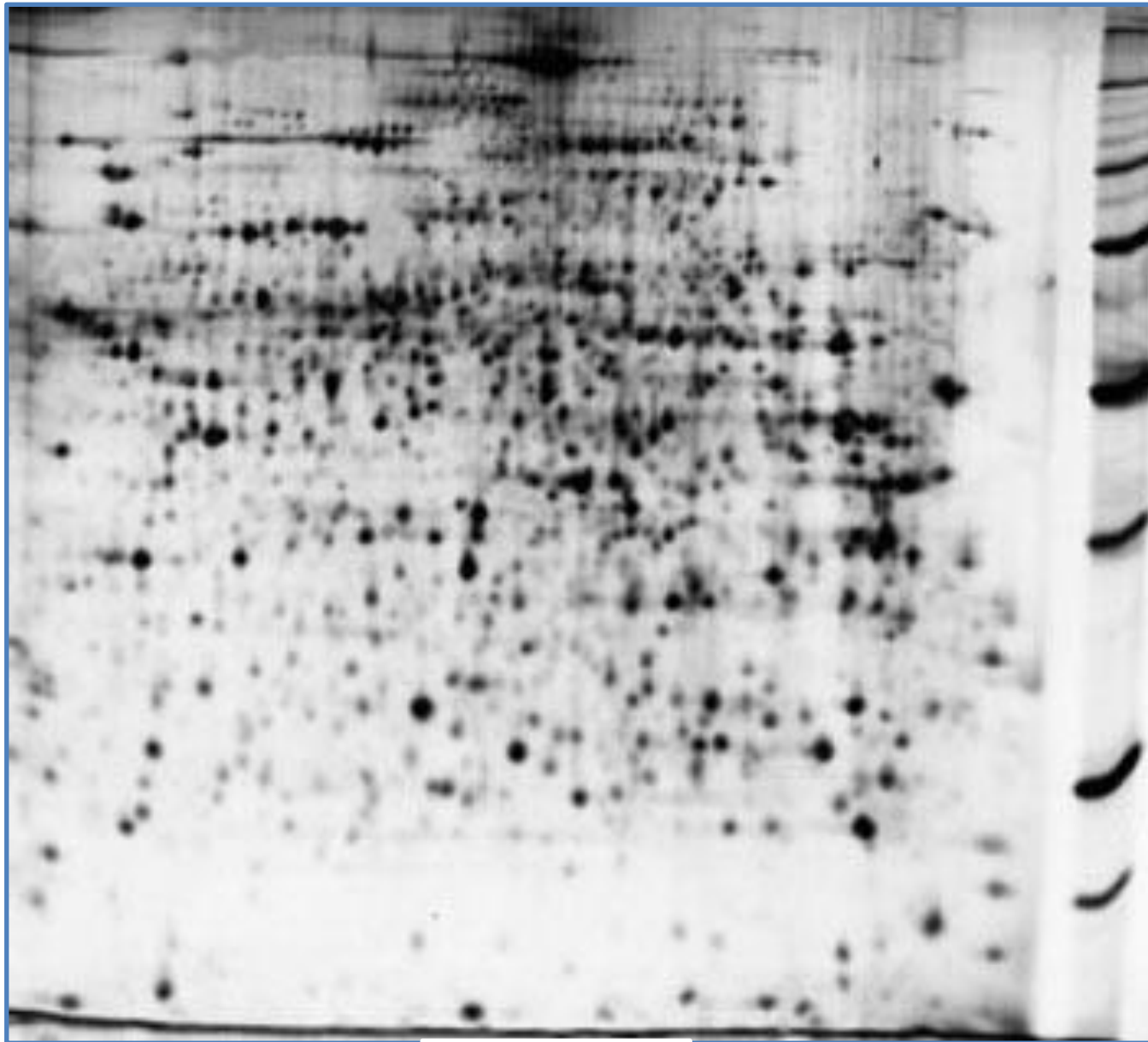
2 - изоэлектрической точке (pI)

- Сначала белки разделяют в первом направлении по заряду согласно их изоэлектрической точке (pI) путем изоэлектрического фокусирования (ИЭФ), а затем во втором направлении – по молекулярной массе (M_r) с помощью электрофореза в ПААГ
- 2D-электрофорез позволяет одновременно разделять и выявлять **тысячи белков** в одном образце (обычный электрофорез меньше 100)

Electrophoresis 2D PAGE

Two Dimensional Electrophoresis



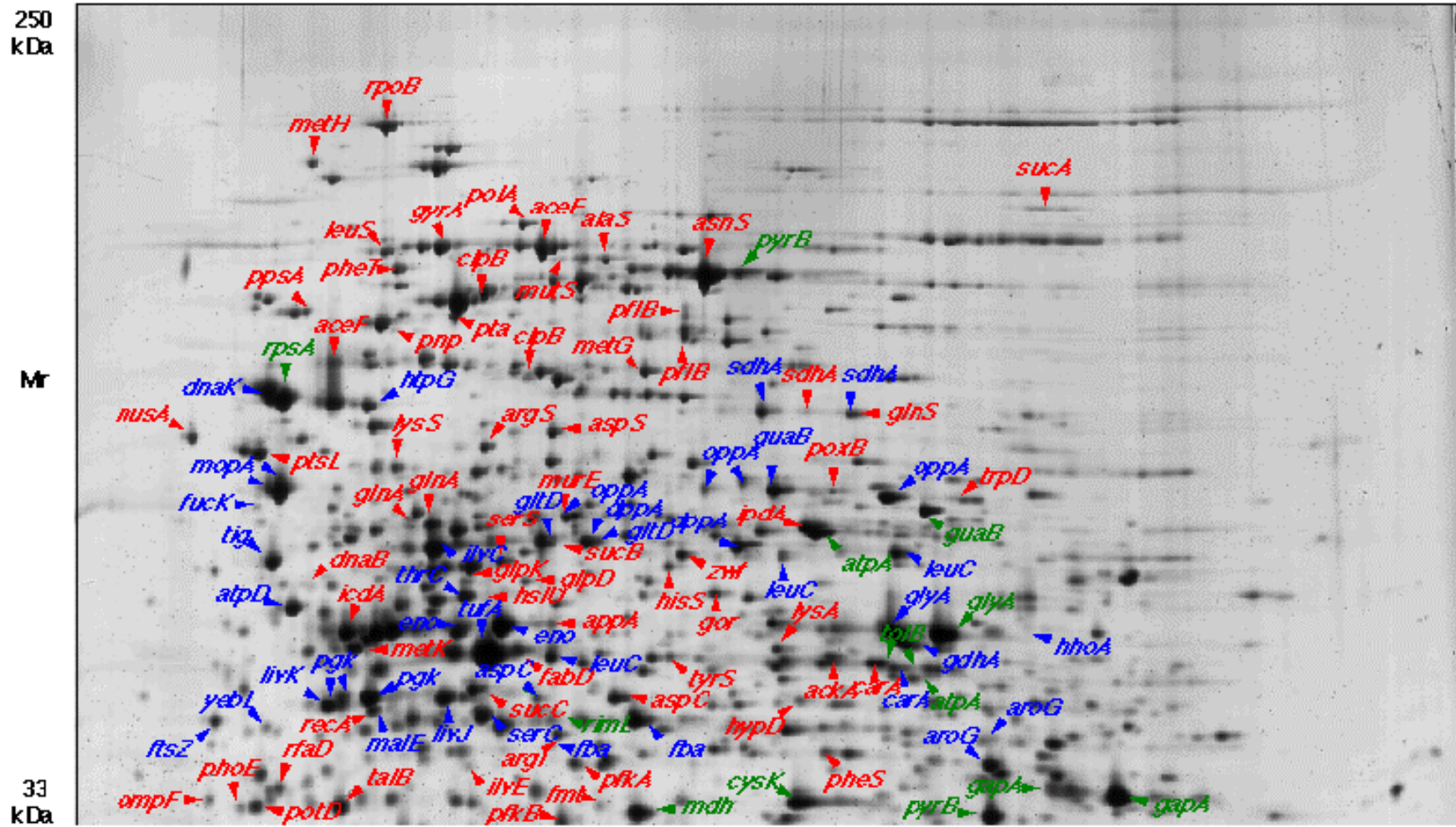


**Протеомная
карта
сложных
белковых
смесей**

pI

Протеомная карта белковых смесей

стрелками обозначены белки,
идентифицированные с помощью масс-спектрометрии



Фрагмент протеомной карты *E. coli*

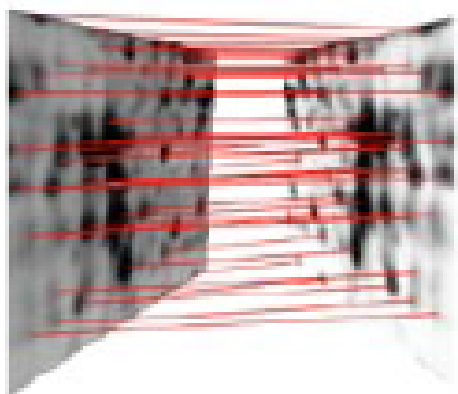
Сравнительный протеомный анализ

Сравнивают протеомные карты на различных этапах развития и/или в различных условиях

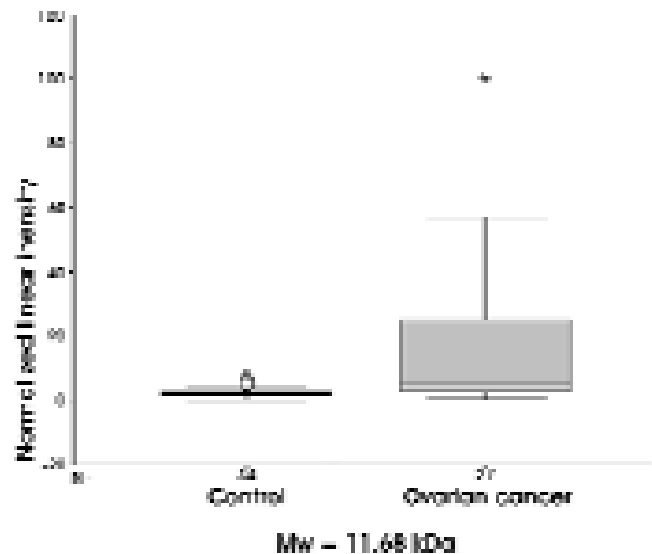
Для этого осуществляют следующее:

- **1.** Выравнивание - поиск одинаково расположенных белковых зон
- **2.** Сопоставление белковых зон
- **3.** Общий статистический анализ - оценка достоверности изменений
- **4.** Количественные характеристики пятен:
 - координаты пятна
 - площадь пятна
 - оптическая плотность пятна
 - объем пятна

Анализ изображений электрофореграмм



Выравнивание,
сопоставление
изображений

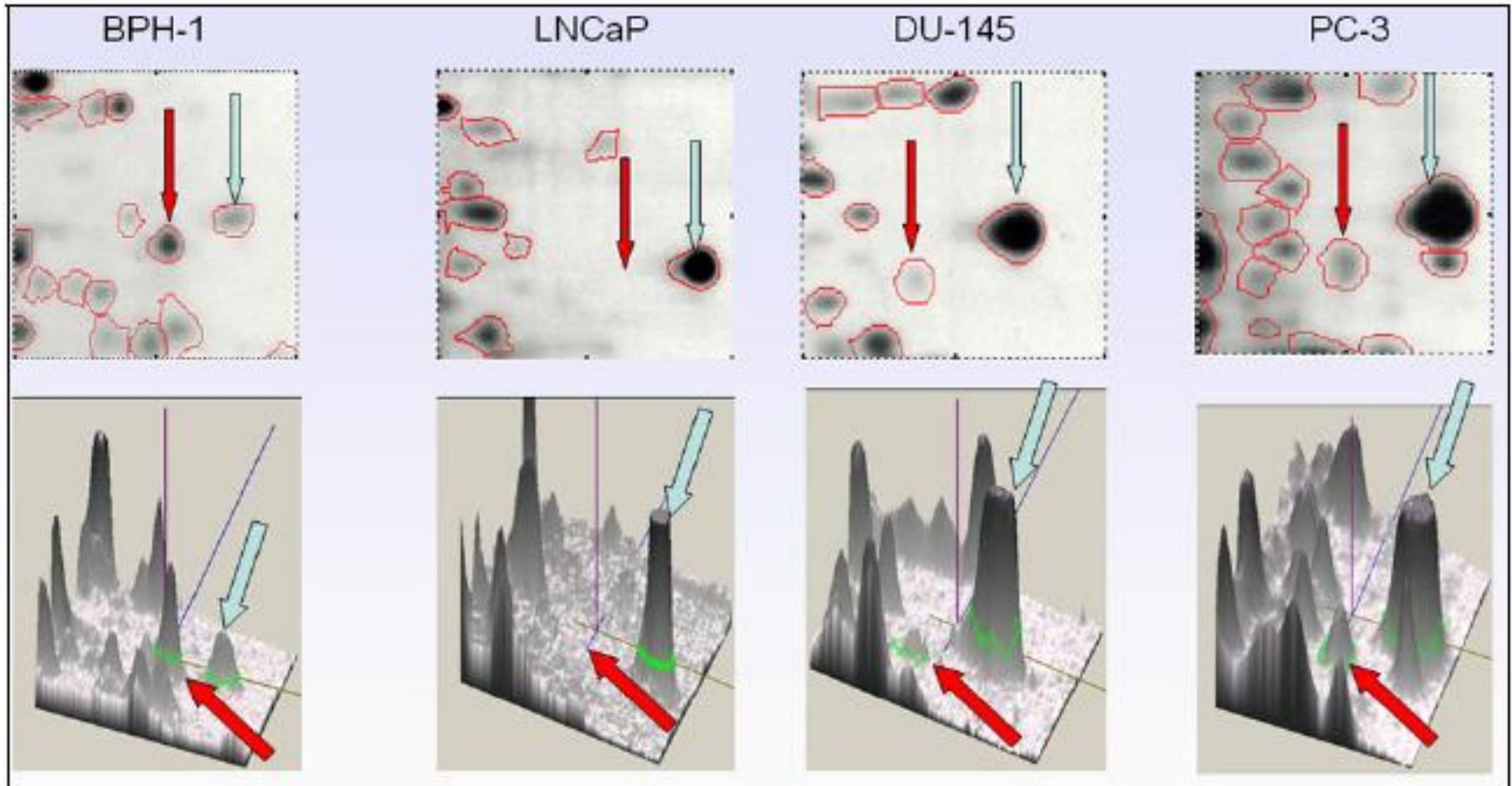


Статистический
анализ различий
экспрессии белка

Сравнительный протеомный анализ программой ImageMaster 2D Platinum

построение трехмерной модели (3D) пятен -

пятно представлено в виде пика, который отражает его форму и объем



Эукариотический трансляционный фактор 1
элонгации альфа 1 (EEF1A1) →

Ингибитор сериновых протеаз, серпин Н1,
(синонимы - белок теплового шока 47,
SERPINH1), →

2D карта клеточных белков двух вариантов

Окраска DIGE Cy3/Cy5:

Электрофорез 2-х образцов
окрашенных белков штамма
Mycoplasma gallisepticum проведен
на одном геле

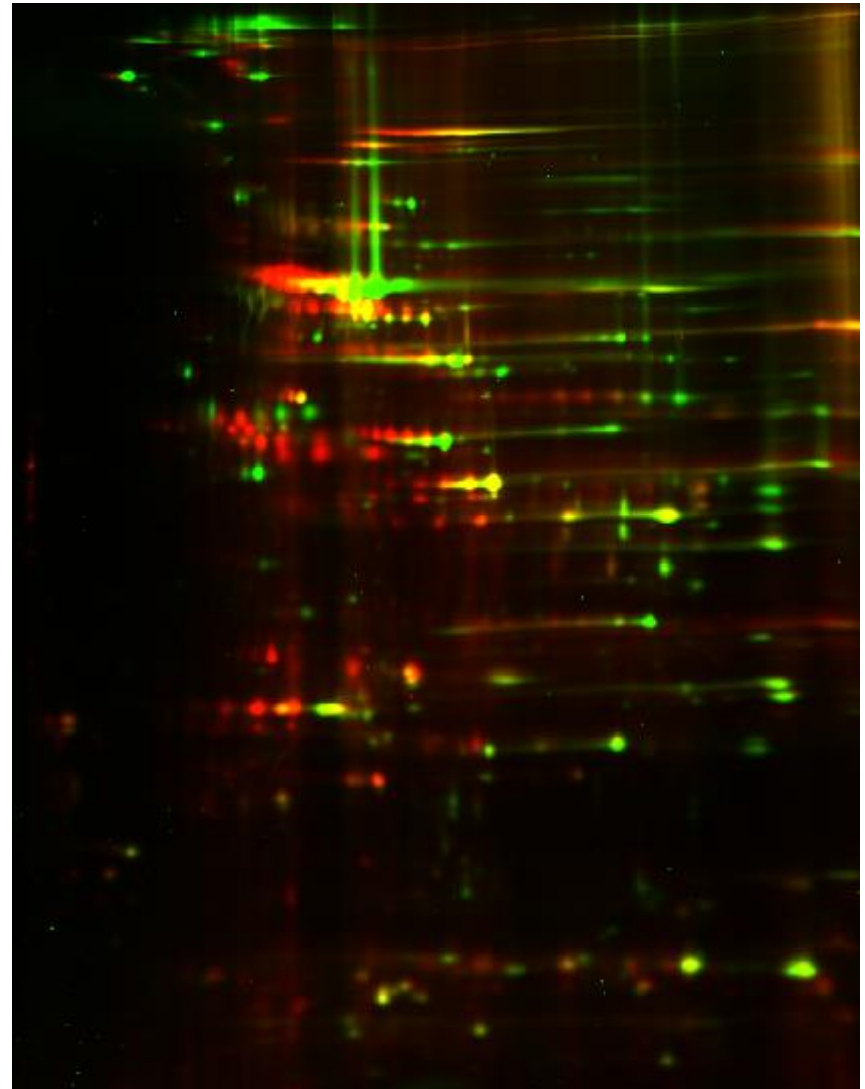
- Зеленым (Cy3) – клеточные белки из контрольных клеток
- Красным (Cy5) – клеточные белки из обработанных клеток

Достоинства:

- чувствителен
- удобен для оценки экспрессии белка

Недостатки:

- недолговечность флуоресценции (сутки)
- непригоден для вырезания пятен



Двумерный электрофорез позволяет

- Разделять белковые смеси на составляющие белковые компоненты
- Сравнивать белковые профили в образцах
- Изучать реакцию клеток на физиологический или патологический фактор (добавка к культуре лекарственного средства)
- Все обнаруженные изменения могут быть оценены на качественном и количественном уровне

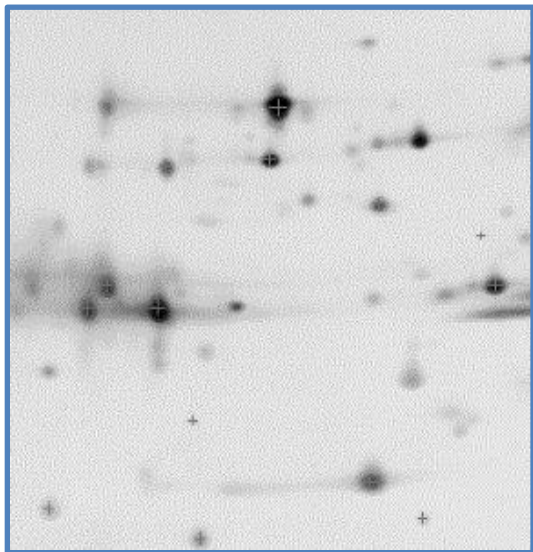
Методы протеомики

- **4) Масс-спектрометрия** - метод, позволяющий установить количественный и качественный состав белков в исследуемом образце (очищенный белок или клеточный лизат)
- В основе метода - законы движения заряженных частиц в магнитном или электрическом поле
- Измеряют отношение массы заряженных частиц (ионов) к их заряду -- m/z
- **Масс-спектр** - это рассортировка заряженных частиц по их массам, а точнее по отношениям массы к заряду

Масс-спектрометрия

- Для проведения анализа пятно вырезают из 2D-ПААГ-ЭФ-геля и элюируют
- Полученный белок гидролизуют на фрагменты путем расщепления трипсином
- Смесь образовавшихся фрагментов подвергают масс-спектрометрии
- Получают спектральные данные для каждого фрагмента смеси по соотношению - m/z
- Поиск в базах данных измеренных масс пептидов
- Сборка последовательности белка

Стратегия анализа: 2DЭф + MALDI-MS



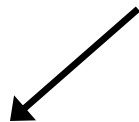
2D электрофорез белков



Компьютерный анализ изображений гелей



Вырезание и трипсинолиз белков в
фрагментах геля



Получение
MALDI масс-спектра суммарного
гидролизата



Поиск в базе данных
измеренных масс пептидов
(пептидный фингерпринт)

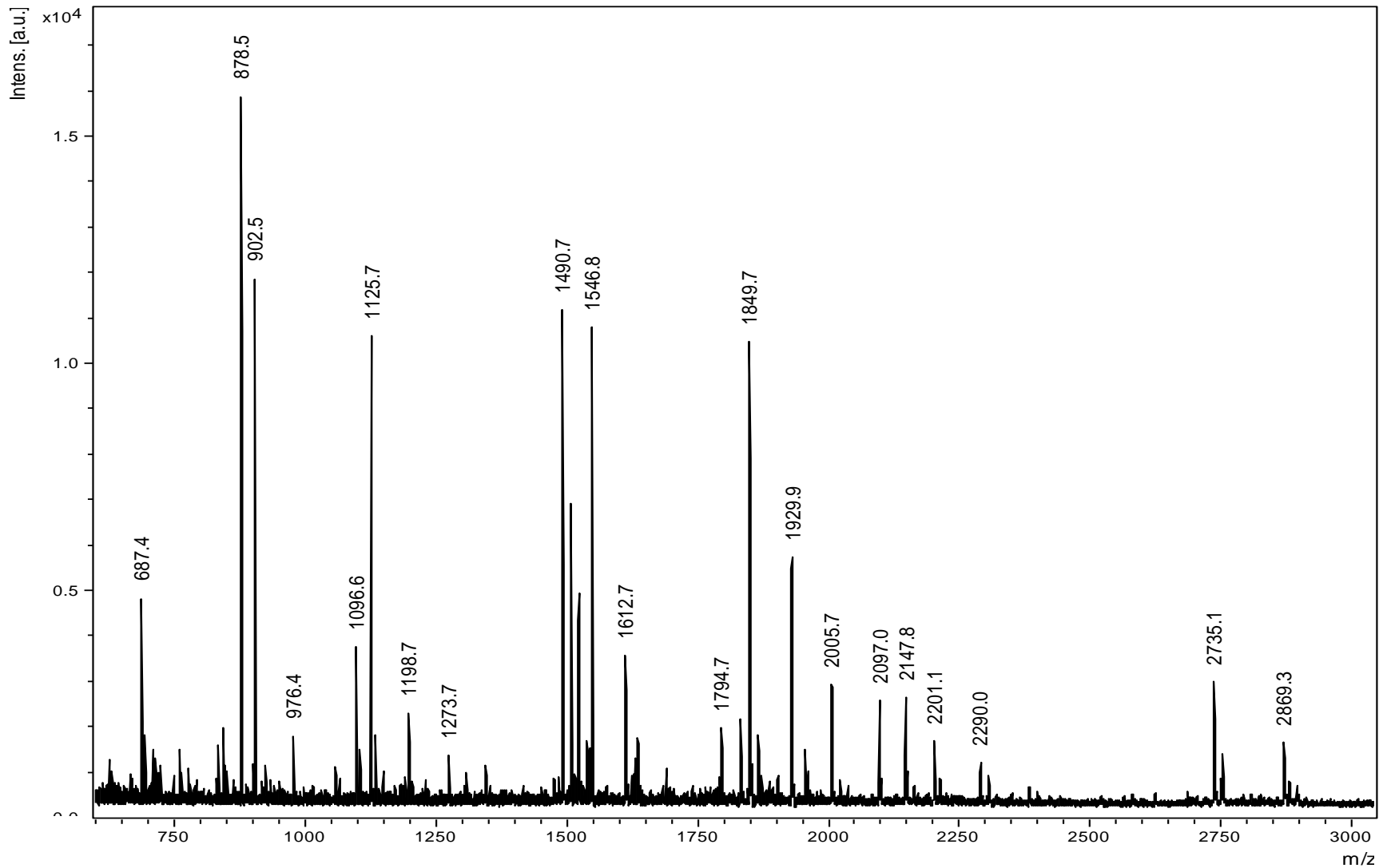


Post Source Decay-MS, TOF-TOF
фрагментация отдельных пептидов



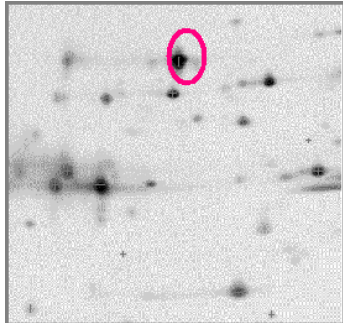
Поиск в базе данных
измеренных масс фрагментов

Пример MALDI масс-спектра: триптический гидролизат фрагмента белка М1 вируса гриппа

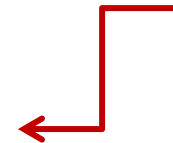
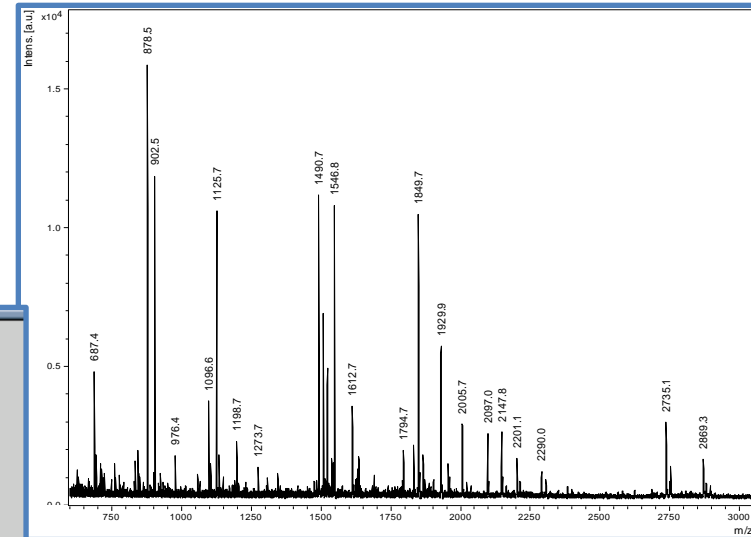
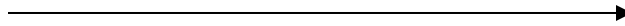


**MLLTQVQTYVLSIIPSGPLKAEIAQRLEDVFAGKNTDLEVLMEWLKTRPILSPLTKGILGFVFTLTPSERGLQ
RRRFVQNALNGNDPNNMDKAVKLYRKLKREITFHGAKEISLSYSAGALASCMGLIYNRMGAVTTEVAFGLV
CATCEQIADSQHRSHRQMVT TTNPLIRHENRMVLASTTAKAMEQMAGSSEQAAEAMEVASQARQMVQAMR
TIGTHPSSAGLKNLLENLQAYQKRMGVQMQRFK**

Схема проведения идентификации белка по его пептидному фингерпринту



Вырезание отдельных белковых пятен (>20 нг/мм²)
Трипсинолиз белков в фрагментах геля
Получение масс-спектра



Mascot: Peptide Mass Fingerprint

Your name Email

Search title

Database

Taxonomy

Enzyme Allow up to missed cleavages

Fixed modifications: Acetyl (K), Acetyl (N-term), Amide (C-term), Biotinylated (K), Biotinylated (N-term)
Variable modifications: Oxidation (M), PEO Biotin (C), Phospho (ST), Phospho (Y), Propionamide (C)

Protein mass kDa Peptide tol. ppm

Mass values MH⁺ M_r Monoisotopic Average

Data file Browse...

Query
NB Contents of this field are ignored if a data file is specified.

Overview Report top hits

Start Search ...

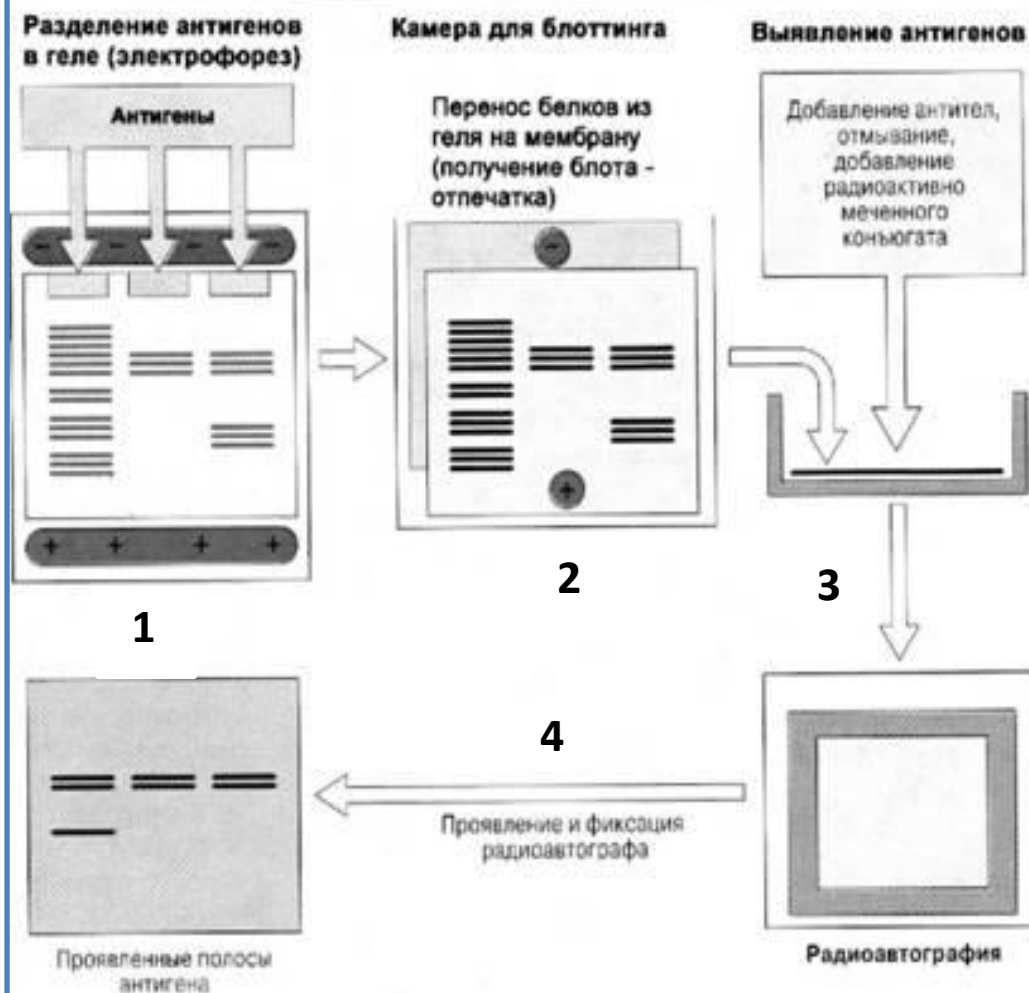
Reset Form

• Масс-спектрометрия позволяет определять пост-трансляционные модификации

Методы протеомики

5) Белковый иммуноблот

Процедура иммуноблоттинга



Вестерн-блоттинг (вестерн-блот, белковый иммуноблот, **Western blot**) — метод для определения специфичных белков в образце.

1. На первом этапе используют электрофорез белков в полиакриламидном геле для разделения денатурированных полипептидов по массе
2. Белки переносят на нитроцеллюлозную мембрану
3. Затем проводят детекцию с использованием антител и меченного конъюгата
4. Радиоавтография

Другие методы для идентификации белков на двумерной карте

- **Комиграция с известными белками** – для этого требуется набор стандартов чистых белков
- **Специфическое связывание с мечеными лигандами** – основано на способности белков связывать другие белки или лиганды с последующим выявлением образовавшихся комплексов (например, GTP-связывающие белки и т.д.).

Оборудование для протеомной лаборатории с 2D электрофорезом



Практическая протеомика –

новые перспективы для медицины, ветеринарии, фармакологии, пищевой промышленности и других прикладных областей

- Разработка новых диагностических тестов
- Выявление новых молекулярных мишеней для лекарственных препаратов, на основе сравнения протеомных карт нормальных и патологических тканей
- Сравнение протеомов опухолевых и нормальных клеток для диагностики и лечения
- Создание протеомных карт крови для определения причины заболевания

Практическая протеомика:

- **Уточнение аннотации генома:**

Исследование протеома позволяет обнаружить все варианты альтернативного сплайсинга

- **Сравнительная протеомика:**

- Сравнение протеомов позволяет выявить общие белки для разных организмов и белки, которые обуславливают различия фенотипов
- Такой анализ необходим для понимания путей эволюции
- При помощи сравнительной протеомики выявляют белки в ответ на действие агенов (например, белки насекомых-вредителей в ответ на обработку инсектицидами для практических целей)

Протеомика

Структурная
протеомика

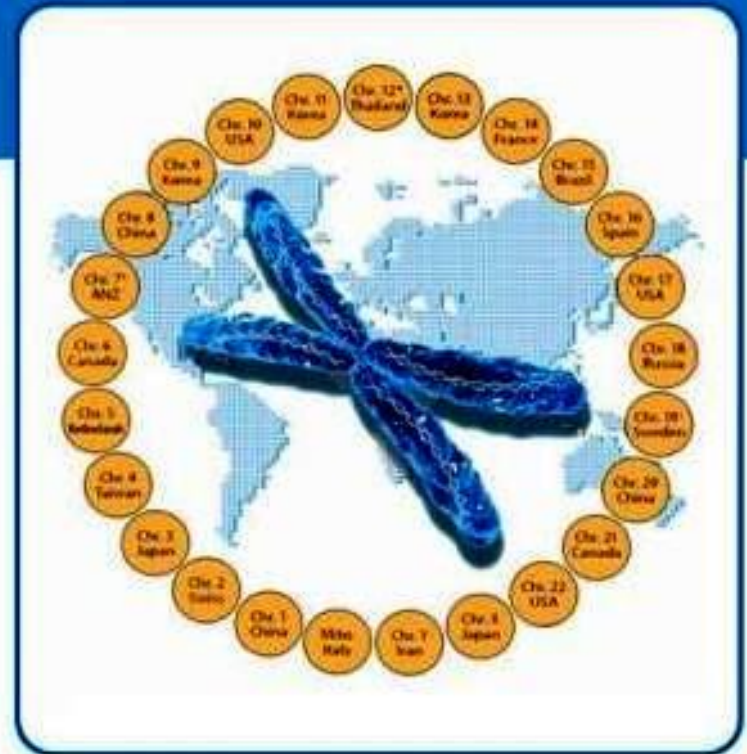
Функциональная
протеомика

Прикладная
протеомика

Международный проект «Протеом человека», 2012-2022 гг.

В ПРОДОЛЖЕНИЕ ПРОЕКТА
«ГЕНОМ ЧЕЛОВЕКА»

20 СТРАН ПРИНИМАЮТ УЧАСТИЕ
В СОЗДАНИИ КАТАЛОГА
БЕЛКОВ, ЗАКОДИРОВАННЫХ В
ГЕНОМЕ ЧЕЛОВЕКА



The call of the human proteome, *Nature Methods*, 2010

Chromosome(s)α	Country(s)α
1α	Chinaα
2α	Switzerlandα
3α	Japanα
4α	Taiwanα
5α	Netherlandsα
6α	Canadaα
7α	Australia and New Zealand
8α	Chinaα
9α	Koreaα
10α	United Statesα
11α	Koreaα
12α	Thailandα
13α	Koreaα
14α	Franceα
15α	Brazilα
16α	Spainα
17α	United Statesα
18α	Russiaα
19α	Swedenα
20α	Chinaα
21α	Canadaα
22α	United Statesα
Mitochondriaα	Italyα
Yα	Iranα
Xα	Japanα

**Международный
проект «Протеом
человека», который
стартовал
23 сентября 2010
года в Сиднее:**

**Human Proteome
Organization -
HUPO
Human Proteome
Project - HPP**



Россия идентифицирует все белки 18 хромосомы в проекте HUPO



