



Сравнительный анализ методов изучения протеома плазмы крови

Смирнова Екатерина
Мальков Данил
Веретененко Ирина
Луппов Даниил
Чернова Анастасия
Столярова Ангелина

Синицын Павел
Max Planck Institute of Biochemistry

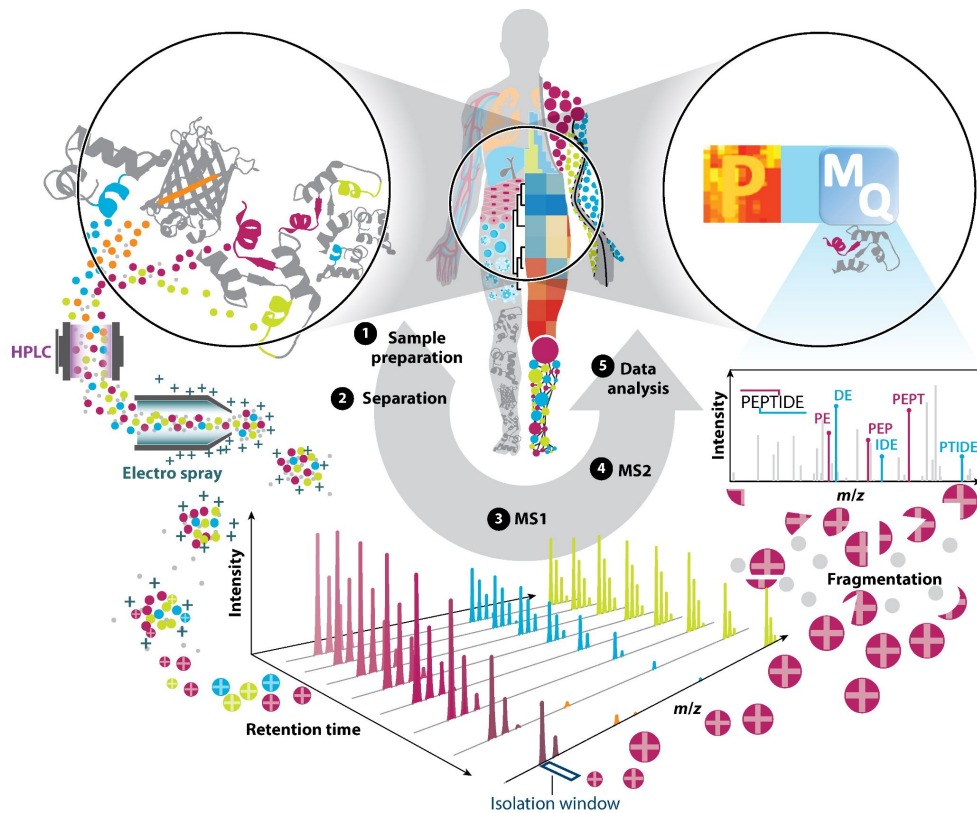


Цели и задачи проекта

Проанализировать данные двух мега-исследований ([DDA](#) и [DIA](#)) с 1000 образцами плазмы крови

- Картировать протеомные данные на предсказания AlphaFold2 и найти закономерности покрытия
- Сравнить статистику пропущенных значений (Missing Values) между технологиями. Сравнить вариацию

Что такое shotgun протеомика?



Неполное покрытие

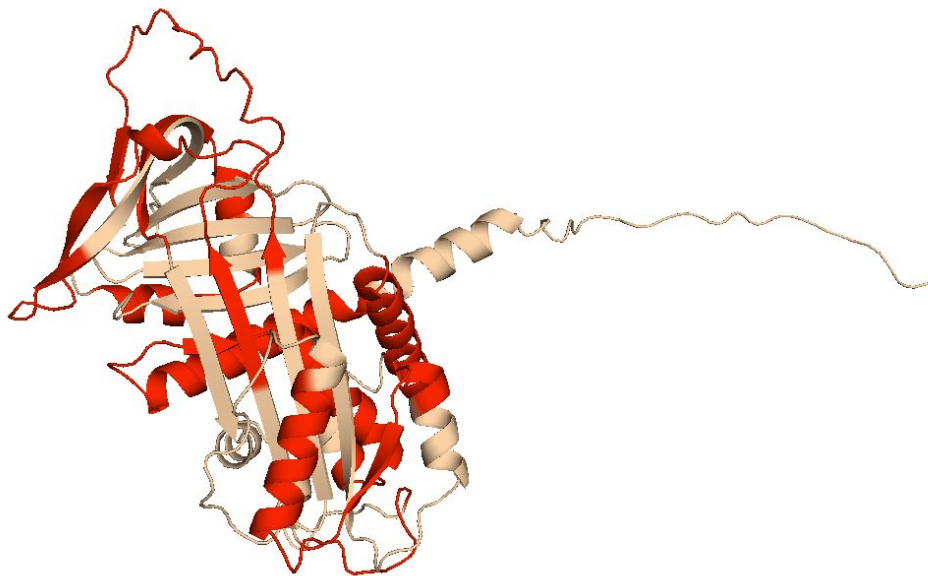
<https://alphafold.ebi.ac.uk/>

хорошее покрытие
85%

не очень хорошее покрытие
45%



Serotransferrin
(P02787 Uniprot)



Corticosteroid-binding globulin
(P08185 Uniprot)

Какие предпочтения у трипсина?

Постановка задачи

Трипсин - разрезает белок
по ARG/LYS

cleavage

not cleavage

Дано:

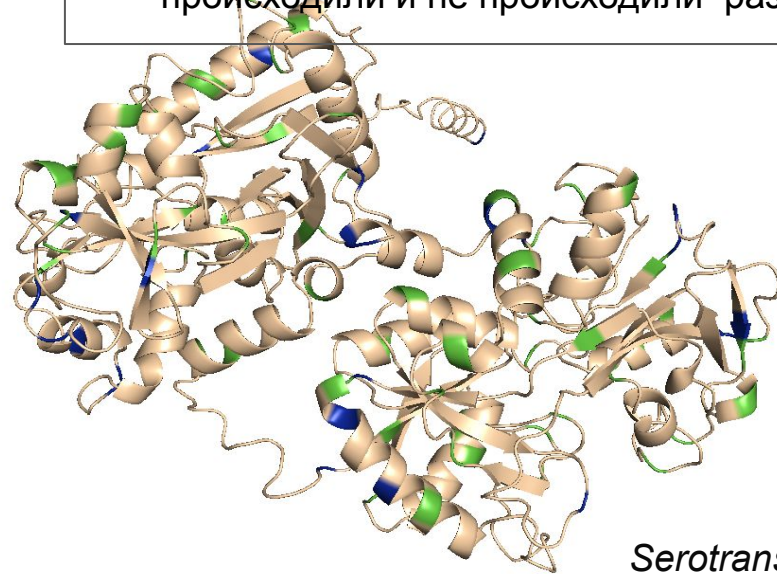
- fasta - пептиды, на которые нарезан белок
- pdb - структура белка

Цель:

- Построить распределение расстояний от cleaved/not cleaved ARG/LYS до поверхности белка
- Гипотеза: трипсин разрезает ARG/LYS близко к поверхности белка => чем больше белок, тем более вероятны различия cleaved/not cleaved

Алгоритм решения

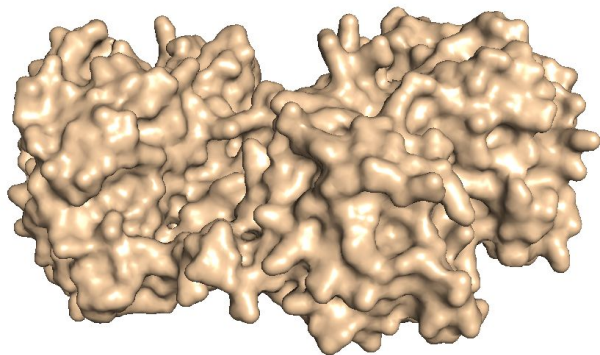
1. Находим положение пептидов в белке
↓
знаем номера ARG/LYS, по которым происходили и не происходили разрезы



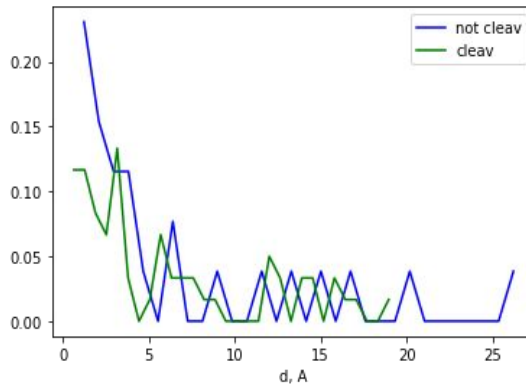
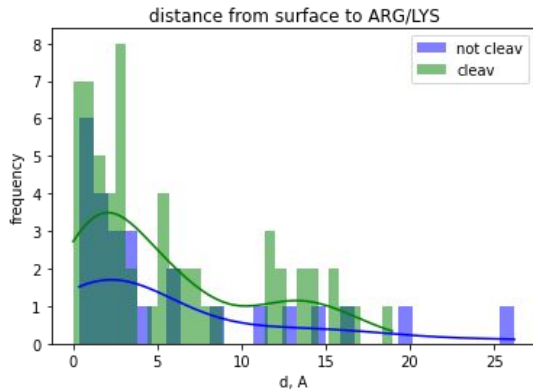
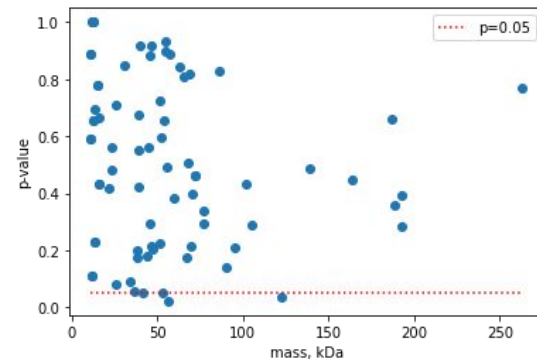
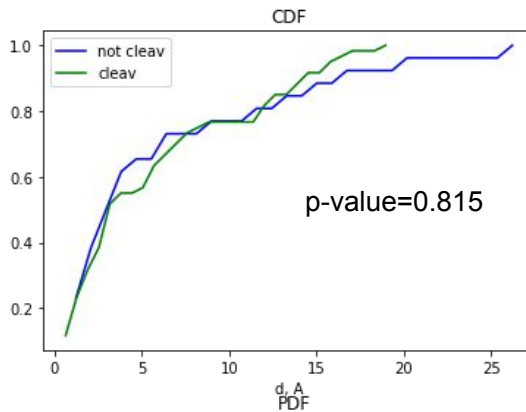
Serotransferrin
(P02787 Uniprot)

Какие предпочтения у трипсина?

2. Строим поверхность белка, находим расстояния от всех ARG/LYS до поверхности



3. Тест Колмогорова-Смирнова:
H₀: cleav и not_cleav из одного распределения



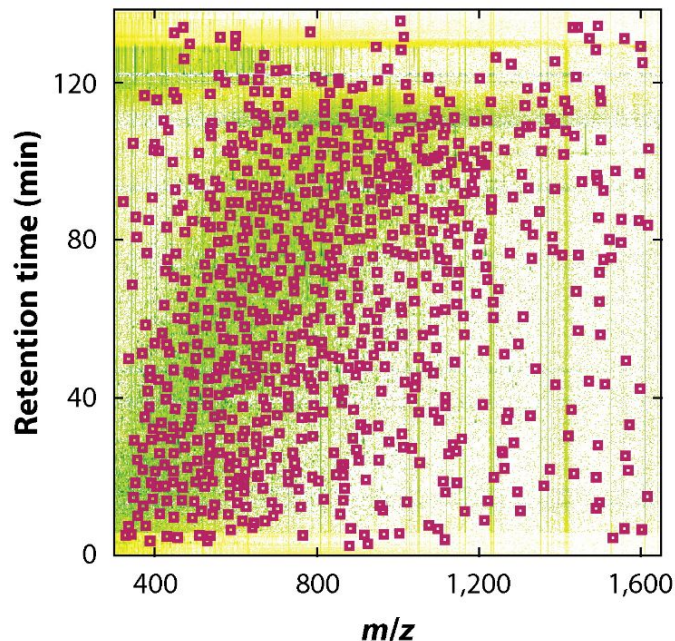
Итог:

- p-value > 0.05
- нет ↓ p-value при ↑ массы белка

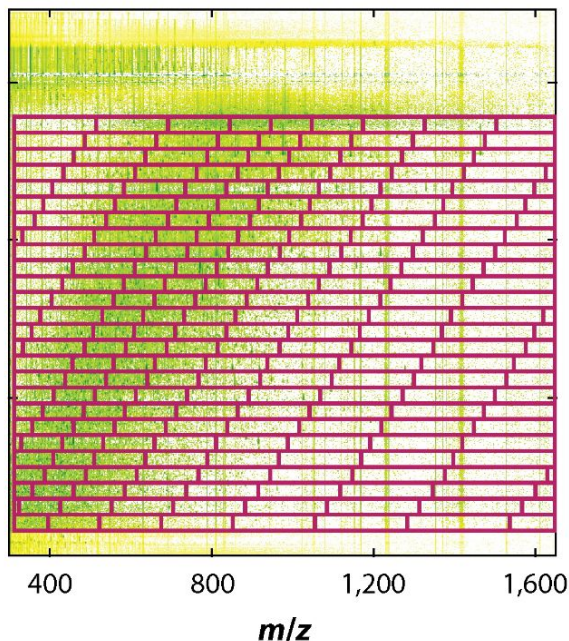
нельзя сказать, что у трипсина
есть предпочтения!

DDA или DIA?

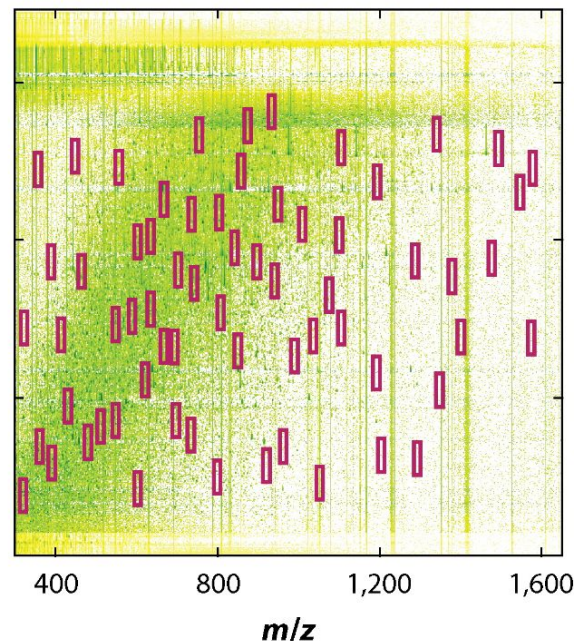
a Data-dependent acquisition



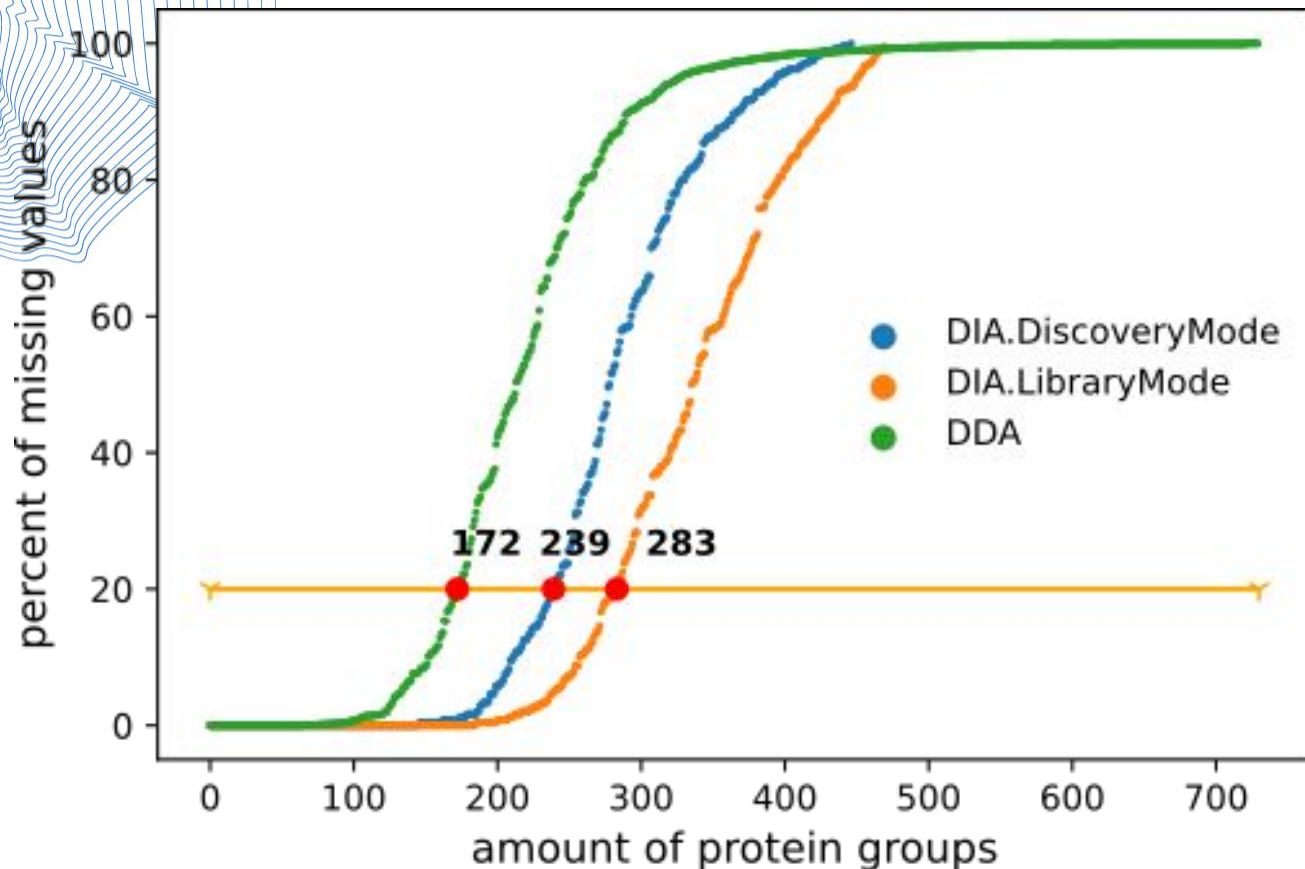
b Data-independent acquisition



c Targeted



Статистика пропущенных значений

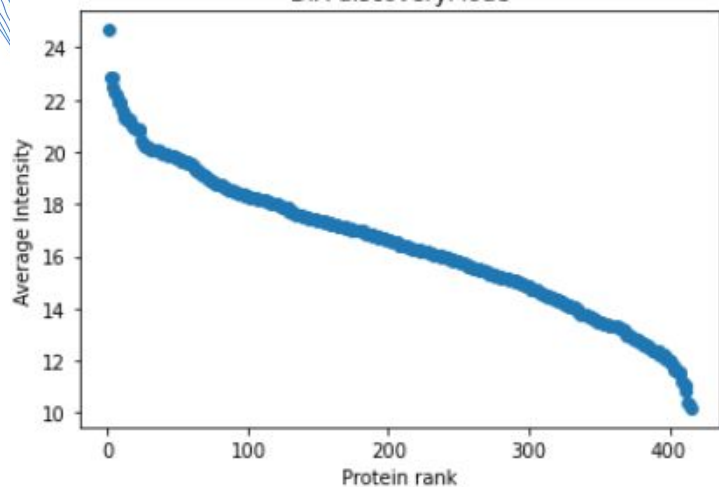


DDA дает большое разнообразие пептидов, но и множество пропущенных значений

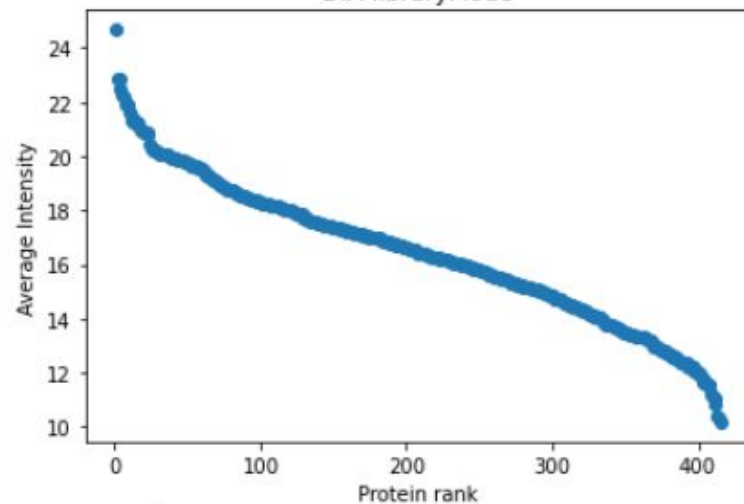
DIA напротив предоставляет меньшее число пептидов, но количество проанализированных выше

Average Intensity of plasma proteins

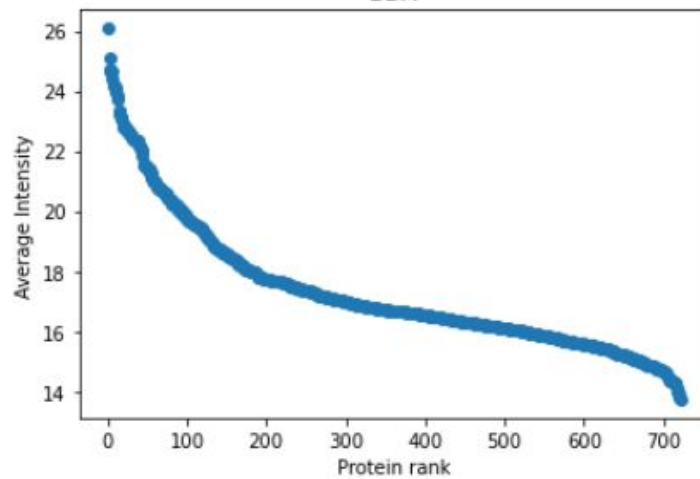
DIA discoveryMode



DIA libraryMode

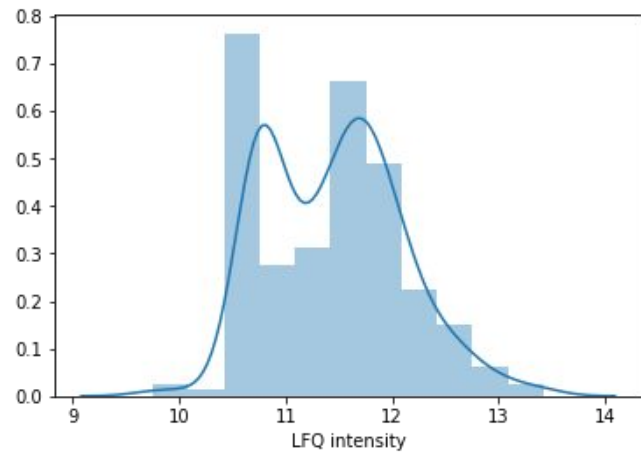
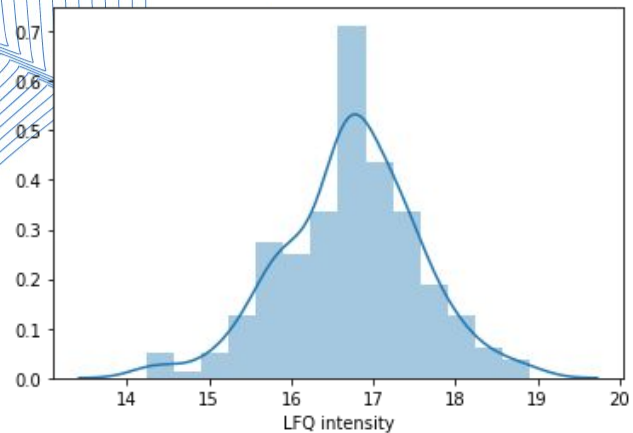


DDA

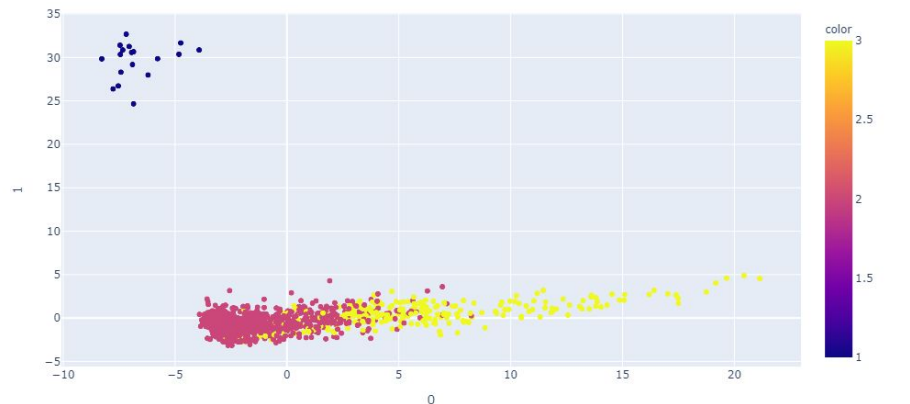


PCA plot и иерархическая кластеризация масс-спектрометрических данных

Подготовка данных



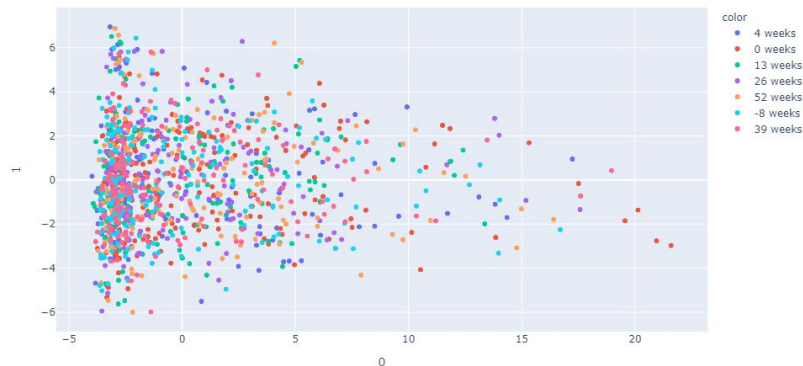
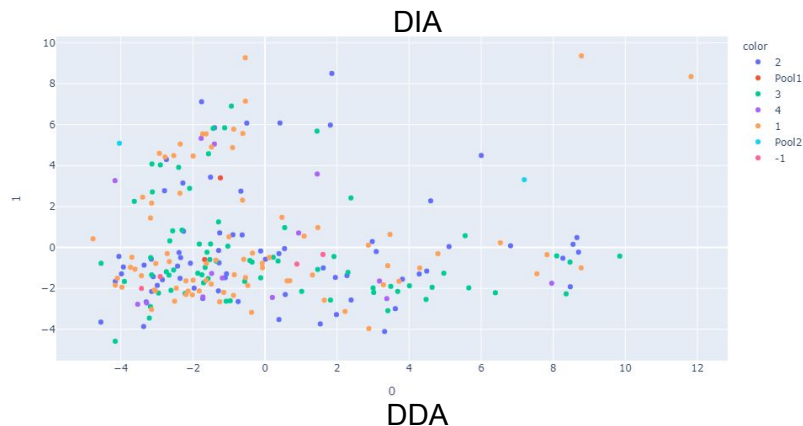
Batch в DDA



	source name	characteristics[organism]	characteristics[organism part]	characteristics[individual]	characteristics[cell type]	characteristics[ancestry category]	charac
assay name							
doubletop14_f1_1	PXD004242-F1 plasma at not applicable weeks	Homo sapiens	blood plasma		F1	not applicable	not available
doubletop14_f1_2	PXD004242-F1 plasma at not applicable weeks	Homo sapiens	blood plasma		F1	not applicable	not available
doubletop14_f1_3	PXD004242-F1 plasma at not applicable weeks	Homo sapiens	blood plasma		F1	not applicable	not available

PCA plot и иерархическая кластеризация масс-спектрометрических данных

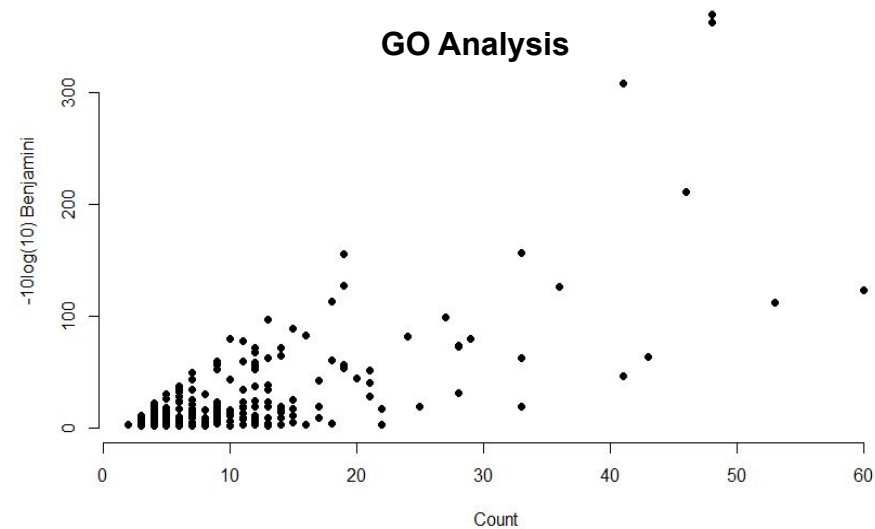
Процессинг



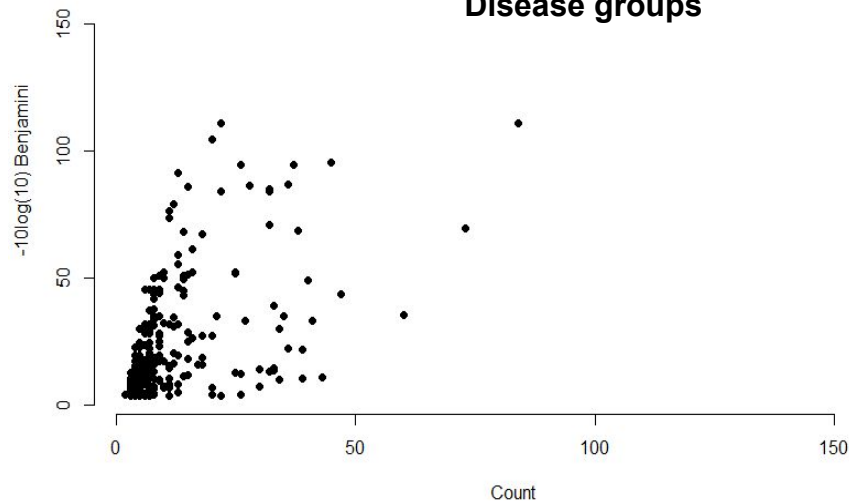
Выводы

1. Проверка и подготовка данных это важно
2. Индивидуальные отличия каждого пациента значительно влияют на белковый состав, делая невозможным разделение образцов на pca_plot-е
3. Поиск данных - важная составляющая работы биоинформатика





Disease groups



Category	Term	Count	%	PValue	Genes
GAD_DISEASE	Type 2 Diabetes edema rosiglitazone	164	24.773414	1.987971e-22	89, 7045, 7169,
GAD_DISEASE	null	84	12.688822	4.686090e-15	89, 7167, 5265,
GAD_DISEASE	Alzheimer's disease	73	11.027190	6.496917e-10	89, 9188, 6277,
GAD_DISEASE	Chronic renal failure Kidney Failure, Chronic	60	9.063444	4.400211e-06	5782, 5345, 348
GAD_DISEASE	Lung Cancer	47	7.099698	5.668734e-07	929, 9782, 5265
GAD_DISEASE	atherosclerosis	45	6.797583	4.154923e-03	929, 5360, 5265
GAD_DISEASE	Schizophrenia	43	6.495468	5.470578e-03	9562, 8991, 544
GAD_DISEASE	Myocardial Infarction	41	6.193353	8.082206e-06	929, 57180, 526
GAD_DISEASE	diabetes, type 2	40	6.042296	1.240325e-07	5184, 929, 5444