

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА

На правах рукописи



ПЛЮЩЕНКО ИВАН ВИКТОРОВИЧ

**РАЗРАБОТКА МЕТОДОВ ОБРАБОТКИ ДАННЫХ
НЕЦЕЛЕВОГО МЕТАБОЛОМНОГО АНАЛИЗА
МЕТОДОМ ЖИДКОСТНОЙ ХРОМАТОМАСС-СПЕКТРОМЕТРИИ
ДЛЯ ПОИСКА БИОМАРКЕРОВ**

Специальность – 02.00.02 – Аналитическая химия

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата химических наук

Москва – 2022

Работа выполнена на кафедре аналитической химии Химического факультета МГУ имени М.В.Ломоносова.

Научный руководитель **Родин Игорь Александрович**

Доктор химических наук

Официальные оппоненты: **Кирсанов Дмитрий Олегович**

Доктор химических наук

Институт химии ФГБОУ ВО «Санкт-Петербургский государственный университет», профессор

Вирюс Эдуард Даниэлевич

Доктор химических наук

Научно-исследовательский институт общей патологии и патофизиологии, ведущий научный сотрудник

Франкевич Владимир Евгеньевич

Доктор физико-математических наук

ФГБУ «Национальный Медицинский Исследовательский Центр акушерства, гинекологии и перинатологии им. В.И. Кулакова» Минздрава Российской Федерации, заведующий отделом системной биологии в репродукции

Защита диссертации состоится «16» марта 2022 года в 15 часов 00 минут на заседании диссертационного совета МГУ.02.05 Московского государственного университета имени М.В.Ломоносова по адресу: 119991, г. Москва, ГСП-1, Ленинские горы, д.1, стр. 3, МГУ имени М.В.Ломоносова, Химический факультет, аудитория 337.

E-mail: dissovet02.00.02@mail.ru

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В.Ломоносова (Ломоносовский просп., д.27) и на сайте ИАС «ИСТИНА»: <https://istina.msu.ru/dissertations/428601911/>

Автореферат разослан «10» февраля 2022 года.

Ученый секретарь

диссертационного совета,

кандидат химических наук

И.А. Ананьева

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Метабомика изучает содержание низкомолекулярных соединений, совокупность которых отражает химические превращения при физиологических процессах организма. Ключевым методом метаболомного анализа является нецелевое, или ненаправленное профилирование, проводимое методом жидкостной хроматомасс-спектрометрии. Ненаправленное профилирование означает проведение анализа без списка целевых соединений. В этом случае регистрируют хроматографический профиль, характеризующийся детектированием максимального числа пиков, чтобы в дальнейшем выявить статистически значимые компоненты. Чаще всего целью нецелевого метаболомного профилирования становится поиск биомаркеров заболеваний.

Актуальность темы

Многопараметрический по своей природе метаболомный анализ генерирует большой массив данных, что подразумевает использование автоматизированного программного обеспечения для обработки и анализа данных и проведения измерений. Несмотря на большее количество существующих вычислительных решений и инструментов, для многих этапов метаболомного эксперимента все еще не проведено комплексного систематического сравнения и оценки возможностей существующих подходов и алгоритмов. Дополнительной проблемой является отсутствие программной реализации в рамках единого вычислительного инструмента. Кроме того, существующие подходы не предлагают единую универсальную схему для обработки данных метаболомного анализа. По этой причине, отбор оптимального набора биомаркеров для диагностических целей остается научной задачей, которую каждый раз приходится решать исследователям.

Применение масс-спектрометрии низкого разрешения для целей ненаправленного метаболомного исследования и поиска биомаркерных сигналов для диагностики ограничено, в основном, целевым анализом. При этом существующие подходы к обработке данных и программные решения разработаны для инструментов высокого разрешения. В то время как использование инструментов низкого разрешения позволило бы увеличить доступность метаболомного анализа.

Одним из ключевых этапов при обработке метаболомных данных является преобразование сигнала детектора. В это понятие входит проведение процедуры коррекции сигнала для удаления дрейфа сигнала масс-спектрометрического детектора

и батч-эффекта. Дрейф сигнала связан с искажением отклика детектора при анализе биологических образцов в аналитической последовательности вследствие загрязнений источника ионов и ионной оптики. Батч-эффект проявляется как зависимость получаемых данных от технических факторов (номер аналитической последовательности или порядок вколов в хроматограф), а не биологической информации. Другой разновидностью преобразования сигнала является оценка и учет влияния биологических факторов фенотипической изменчивости перед проведением статистического анализа. Разработка и внедрение в практику эффективных методов коррекции сигнала и учета биологических факторов позволит увеличить достоверность результатов метаболомного анализа.

Цель работы

Развитие методологических аспектов метаболомного анализа и разработка методов обработки и анализа метаболомных данных, получаемых при проведении ненаправленного профилирования с использованием жидкостной хроматомасс-спектрометрии низкого и высокого разрешения.

Достижение поставленной цели заключалось в решении следующих **задач**:

1. Предложить способ анализа метаболомных данных, позволяющий выделять набор биомаркеров и провести его апробацию.
2. Протестировать возможности применения жидкостной хроматомасс-спектрометрии низкого разрешения для целей ненаправленного метаболомного профилирования.
3. Разработать методы коррекции сигнала для эффективного устранения дрейфа сигнала и батч-эффекта, а также способы учета факторов фенотипической изменчивости.
4. Разработать многофункциональный программный инструмент для анализа и обработки данных метаболомного профилирования, полученных методом жидкостной хроматомасс-спектрометрии.
5. Провести экспериментальную апробацию разработанных подходов на наборах образцов мочи с целью выявления потенциальных биомаркеров рака мочевого пузыря и биомаркерных сигналов колоректального рака.

Научная новизна

Общая научная новизна работы заключается в создании и экспериментальном обосновании новых способов обработки и анализа данных, получаемых при ненаправленном хроматомасс-спектрометрическом метаболомном профилировании.

Предложен и апробирован на двух независимых наборах данных способ выявления биомаркерных сигналов при ненаправленном метаболомном профилировании методом жидкостной хроматомасс-спектрометрии низкого разрешения.

Разработаны методы коррекции для устранения батч-эффекта и дрейфа сигнала масс-спектрометрического детектора, основанные на алгоритмах машинного обучения, превосходящие описанные ранее методы при многопараметрической оценке.

Предложены нелинейные методы учета факторов фенотипической изменчивости, отличающиеся высокой эффективностью при многоклассовой классификации.

Предложен способ анализа метаболомных данных для автоматического отбора значимых биомаркеров с применением сочетания метода коррекции сигнала и алгоритмов машинного обучения с одномерными статистическими тестами для снижения размерности. Способ апробирован на более чем тридцати наборах данных из открытых источников.

Практическая значимость

Разработан программный инструмент для комплексного анализа и обработки данных метаболомного профилирования методом жидкостной хроматомасс-спектрометрии. Программное обеспечение превосходит описанные ранее инструменты в части проведения коррекции сигнала, учета факторов фенотипической изменчивости и отбора переменных и доступно в сети Интернет ¹.

Разработаны алгоритмы обработки и анализа метаболомных данных. Программные реализации доступны исследователям для использования при решении задач метаболомного анализа ^{2,3}.

Предложены наборы потенциальных биомаркеров рака мочевого пузыря с указанием регуляторных путей и потенциальных биомаркерных сигналов для диагностики колоректального рака.

Хранилища метаболомных данных дополнены данными хроматомасс-спектрометрического профилирования образцов мочи при проведении эксперимента по поиску потенциальных биомаркеров рака мочевого пузыря ⁴ и биомаркерных сигналов

¹ <https://plyush1993.github.io/OUKS/>

² https://github.com/plyush1993/E5S_metabolomics_workflow

³ <https://github.com/plyush1993/Multistudy-experiments-with-Multiple-batches>

⁴ <http://doi.org/10.21228/M8ZT4C>

для диагностики колоректального рака ⁵. Предоставленные данные могут быть использованы научным сообществом для дальнейшего развития методологии метаболомного анализа.

Положения, выносимые на защиту

1. Способ обработки и анализа метаболомных данных для выделения набора биомаркеров, основанный на сочетании алгоритма обработки сигнала, статистических тестов и методов машинного обучения, и адаптированный для учета биологических факторов.

2. Способ выявления биомаркерных сигналов при проведении метаболомного исследования методом жидкостного хроматомасс-спектрометрического ненаправленного профилирования низкого разрешения, апробированный на двух выборках образцов для диагностики колоректального рака.

3. Метод коррекции сигнала на основе алгоритма градиентного бустинга, позволяющий устранить дрейф сигнала масс-спектрометрического детектора и батч-эффект в данных метаболомного профилирования, полученных методом жидкостной хроматомасс-спектрометрии.

4. Программное обеспечение для анализа и обработки метаболомных данных, протестированное при проведении ненаправленного метаболомного профилирования для поиска потенциальных биомаркеров рака мочевого пузыря.

Степень достоверности

Достоверность полученных результатов обеспечивалась использованием современных методик и средств обработки результатов экспериментов, анализом реальных образцов, а также применением современного хроматографического и масс-спектрометрического оборудования. На момент проведения измерений все используемое оборудование имело актуальное свидетельство о периодической поверке.

Соответствие паспорту научной специальности

Диссертационная работа соответствует паспорту специальности 02.00.02 – Аналитическая химия по областям исследований:

- методы химического анализа (химические, физико-химические, атомная и молекулярная спектроскопия, хроматография, рентгеновская спектроскопия, масс-спектрометрия, ядерно-физические методы и др.);
- математическое обеспечение химического анализа;

⁵ <http://doi.org/10.21228/M8X110>

- клинический анализ.

Апробация результатов исследования

Основные результаты работы представлены на следующих конференциях:

2021 год: VI Всероссийский Симпозиум «Разделение и концентрирование в Аналитической химии и Радиохимии» с международным участием, Краснодар, Россия, 26 сентября - 2 октября 2021.

2020 год: IV Всероссийская Конференция с международным участием "Аналитическая хроматография и капиллярный электрофорез", Краснодар, Россия, 28 сентября - 2 октября 2020.

2019 год: VIII Всероссийская Конференция с международным участием «Масс-спектрометрия и ее прикладные проблемы», Институт физической химии и электрохимии им. А.Н. Фрумкина РАН, г. Москва, Россия, 14-18 октября 2019; III Всероссийская Конференция по аналитической спектроскопии с международным участием, г. Туапсе, Россия, 29 сентября - 5 октября 2019; Mass Spectrometry & Advances in the Clinical Lab 2019 EU 6th European Congress, Зальцбург, Австрия, 22-26 сентября 2019.

2018 год: V Всероссийский Симпозиум «Разделение и концентрирование в Аналитической химии и Радиохимии» с международным участием, Краснодар, Россия, 7-13 октября 2018; XXII International Mass Spectrometry Conference, Флоренция, Италия, 26-31 августа 2018.

Публикации

По материалам работы опубликовано 10 печатных работ, в том числе 3 статьи в рецензируемых научных изданиях, индексируемых международными базами данных (Web of Science, Scopus, RSCI) и рекомендованных в диссертационном совете МГУ по специальности 02.00.02 – «Аналитическая химия», и 7 тезисов докладов на российских и международных конференциях.

Личный вклад автора

Личный вклад автора заключался в поиске, систематизации и анализе данных литературы по теме работы, планировании, постановке и проведении экспериментов, обработке и интерпретации полученных результатов, а также в подготовке к публикации результатов проведенных исследований. Экспериментальные результаты получены, обработаны и представлены автором лично. Во всех опубликованных работах вклад автора является определяющим.

Структура и объем работы

Полный текст работы состоит из 5 глав и включает 200 страниц, в том числе 36 рисунков и 24 таблицы. Список литературы содержит 503 наименования.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность выбранной темы и поставлены цель и задачи исследования, раскрыта научная новизна работы и показана ее практическая значимость, публикации по теме исследования, степень ее апробации, структура и объем.

Первая глава представляет собой обзор литературы, в котором систематически изложено и критически обсуждено современное состояние основных этапов ненаправленного жидкостного хроматомасс-спектрометрического профилирования.

В разделе «Режимы хроматографического разделения» обобщены последние достижения и традиционные подходы в хроматографическом разделении компонентов метаболома человека. В разделе «Режимы масс-спектрометрического детектирования» обсуждаются характеристики масс-спектрометров и режимы детектирования. Раздел «Проведение пробоотбора и пробоподготовки» посвящен вариантам проведения подготовки биологических проб и условиям их хранения. Раздел «Обеспечение качества» содержит ключевую информацию по вопросам контроля качества, тестам стабильности и готовности хроматомасс-спектрометрической системы. В разделе «Первоначальная обработка» обсуждаются ключевые подходы в обработке сырых данных до получения таблицы пиков и заполнение пропущенных значений. Раздел «Преобразования сигнала» посвящен систематическому изложению алгоритмов коррекции сигнала детектора, нормализации, учета биологических факторов и фильтрации пиков. Разделы «Аннотирование пиков» и «Идентификация метаболитов» содержат обзор подходов к проведению аннотации сигналов и идентификации метаболитов соответственно. В разделе «Статистический анализ» рассмотрены ключевые методы одномерного и многомерного статистического анализа, особое внимание уделено отбору переменных и снижению размерности.

Вторая глава включает описание используемых в работе химических реактивов, биологических образцов, материалов и оборудования, а также условия и методики проведения профилирования, пробоподготовки, процедуры контроля качества, tandemных масс-спектрометрических измерений и вычислений.

Хроматографическое разделение проводили на обращенно-фазовой колонке Aquity UPLC BEH C18 (100 × 2.1 мм), диаметр зерна сорбента 1.7 мкм («Waters», США) с предколонкой UHPLC Guard 3 PK Zorbax SB-C18 (5 × 2.1 мм), диаметр зерна сорбента 1.8 мкм («Agilent Technologies», США).

Профилирование образцов мочи в экспериментах по определению биомаркеров рака мочевого пузыря (РМП) и первой партии образцов для исследования биомаркеров колоректального рака (КРР) проводили на жидкостной хроматомасс-спектрометрической (ВЭЖХ-МС) системе, состоящей из хроматографа LC 20 AD («Shimadzu», Япония) и гибридного масс-спектрометра типа «ионная ловушка – времяпролетный масс-анализатор (ИЛ-ВПМА)» IT-TOF («Shimadzu», Япония), оснащенного источником электрораспылительной ионизации. На этой же системе проводили регистрацию масс-спектров второго поколения ионов (МС²) в режиме сканирования ионов-продуктов.

Профилирование образцов мочи в экспериментах по определению биомаркерных сигналов для диагностики КРР проводили на ВЭЖХ-МС системе, состоящей из хроматографа 1290 Infinity II («Agilent Technologies», США) и tandemного квадрупольного масс-спектрометра (QQQ) 6470 («Agilent Technologies», США), оснащенного источником электрораспылительной ионизации JetSpray («Agilent Technologies», США).

Третья глава посвящена созданию алгоритма обработки метаболомных данных для снижения размерности и выделению наиболее информативного набора биомаркеров.

Проводилась оптимизация условий программы градиентного элюирования, степени разбавления пробы, объема ввода, диапазона масс и режима генерации ионов. Определены условия кондиционирования системы в выбранных условиях и обработки таблицы пиков.

После проведения профилирования в выбранных условиях (масс-спектрометр типа ИЛ-ВПМА) сорока образцов мочи трех клинических групп пациентов (ЦГ – целевая группа (диагноз КРР), ЦПП – после операции, КГ – контрольная группа) получена таблица пиков. После заполнения пропущенных значений (ЗПЗ) по алгоритму половины минимального значения были протестированы одиннадцать методов коррекции сигнала. Оптимальный метод выбирали по минимальному значению относительного стандартного отклонения (ОСО) пиков в образцах контроля качества (КК) среди всех аналитических последовательностей (батчей), а также при оценке

точности классификации четырьмя методами машинного обучения (МО). Для оценки эффективности моделей применяли 5-блочную кросс-валидацию с тремя повторами, оптимизация проводилась по показателю точности классификации. Были выбраны четыре модели МО и в каждой один параметр для оптимизации: метод к-ближайших соседей (KNN), параметр – число соседей, метод частных наименьших квадратов (PLS) - число компонент, машина опорных векторов (SVM) - параметр C стоимость штрафа, случайный лес (RF) - число предикторов для дерева. Оптимальным алгоритмом для экспериментального набора данных среди 11 методов коррекции стал EigenMS. Оценка точности более 95 % при кросс-валидации, процент пиков с ООС менее 30 % более 70 %. Результаты подтверждены на двух независимых наборах данных в виде таблицы пиков из открытых источников данных.

На следующем этапе разрабатывали алгоритм проведения статистического анализа для снижения размерности и выделения наиболее информативного набора метаболитов. Для этого кроме экспериментального набора данных также использовали 36 наборов данных в виде таблицы пиков из 21 исследования из открытых источников данных с различными характеристиками. Схема анализа приведена на рис. 1.

Сначала к таблице данных (после применения метода EigenMS) применяли одномерный статистический фильтр (ОСФ). ОСФ – представляет собой последовательное применение теста нормальности Шапиро-Уилка, проверки гомогенности дисперсии по Бартлетту и соответствующего теста сравнения средних (Уилкоксона, Уэлча и Стьюдента) с поправкой на множественные сравнения по алгоритму Бонферрони- Холма. Отбирались переменные с р-значением менее 0.05 (хотя бы между любыми двумя группами). Затем настраивали четыре модели машинного обучения и отбирали первые пятьдесят важных предикторов в каждом алгоритме. Все двести переменных объединяли в один список и оставляли только те из них, которые встречаются дважды (частота 50 %). Эта процедура была названа фильтрация стабильных предикторов (СПФ). К отобранному набору биомаркеров далее применяли процедуру рекурсивного отбора переменных (РОП). РОП проводили по методу обратного исключения алгоритмами наивного байесовского классификатора (НБК) или RF. Значение площади под ROC (рабочая характеристика приёмника) кривой (AUROC) вычисляли для данных до начала статистической обработки, после ЗПЗ и ОСФ, СПФ и РОП для контроля процедуры снижения размерности. В конце проводили проекцию данных (выбранных биомаркеров) в пространстве главных компонент (метод главных компонент, МГК) и на дендрограмме (иерархический кластерный

анализ, ИКА). Таким образом, качество отбора переменных проверялось анализом ROC кривых и многомерным проецированием.

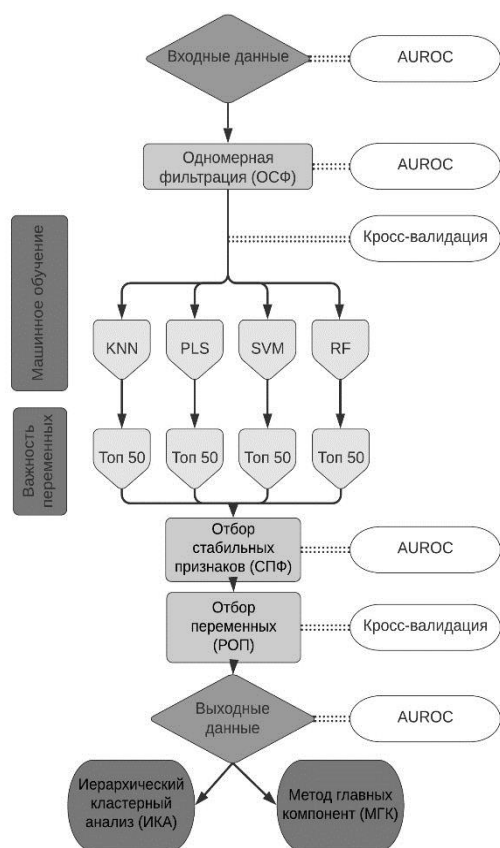


Рис.1. Схема отбора переменных

метку класса в сырых данных. Правильная кластеризация приводит к одному и тому же цвету во всех случаях.

Анализируя результаты проецирования для экспериментальных данных видно, что точки данных в сыром наборе данных после ОСФ (часть А на рис. 2,3) сильно перекрываются для групп ЦГ и ЦГП и в целом образуют два облака данных. Первые два этапа обработки (ОСФ, EigenMS, часть В на рис. 2,3) приводят к полному и правильному распределению точек данных на три клинические группы в пространствах главных компонент и евклидовом. Следующие вычисления (СПФ, РОП) направлены на снижение размерности при сохранении способности к классификации. График счетов МГК и дендрограмма ИКА (часть С на рис. 2,3) демонстрируют полное разделение точек данных на экспериментальные группы после применения полного цикла операций алгоритма. Число предикторов сократилось более чем в сто раз (с 3441 до 25).

Результаты многомерного моделирования методами ИКА и МГК для всех наборов данных доступны по адресу ⁶. Для каждого рисунка части (А, В, С) представляют собой набор данных на разных этапах применения алгоритма: А – сырые данные после ОСФ, В – после EigenMS и ОСФ, С – после всех этапов (ОСФ, EigenMS, СПФ, РОП). Цвет эллипса на графиках счетов МГК и точек связан с меткой в сыром наборе данных и вычисляется автоматически. Уровень значимости 95%. На графиках дендрограмм ИКА цвет ветвей и соответствующий прямоугольник (разделяющий дерево по заданному количеству групп) отражает метку класса, определенную метрикой ИКА. Цвет листьев (текст) отражает

⁶https://www.rsc.org/suppdata/d0/ay/d0ay00204f/d0ay00204f1.rar?_ga=2.50409569.973419256.1637425306-609784116.1558564539

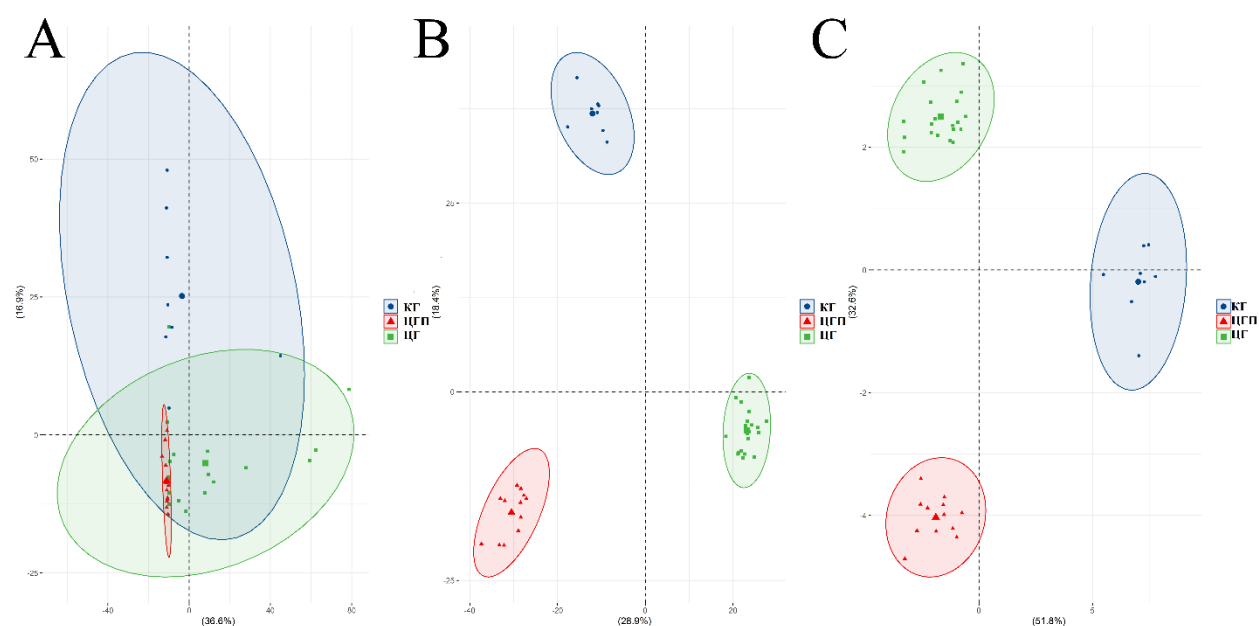


Рис. 2. График счетов МГК для экспериментальных данных. Применено масштабирование и центрирование. ЦГ – целевая группа, ЦГП – после операции, КГ – контрольная группа

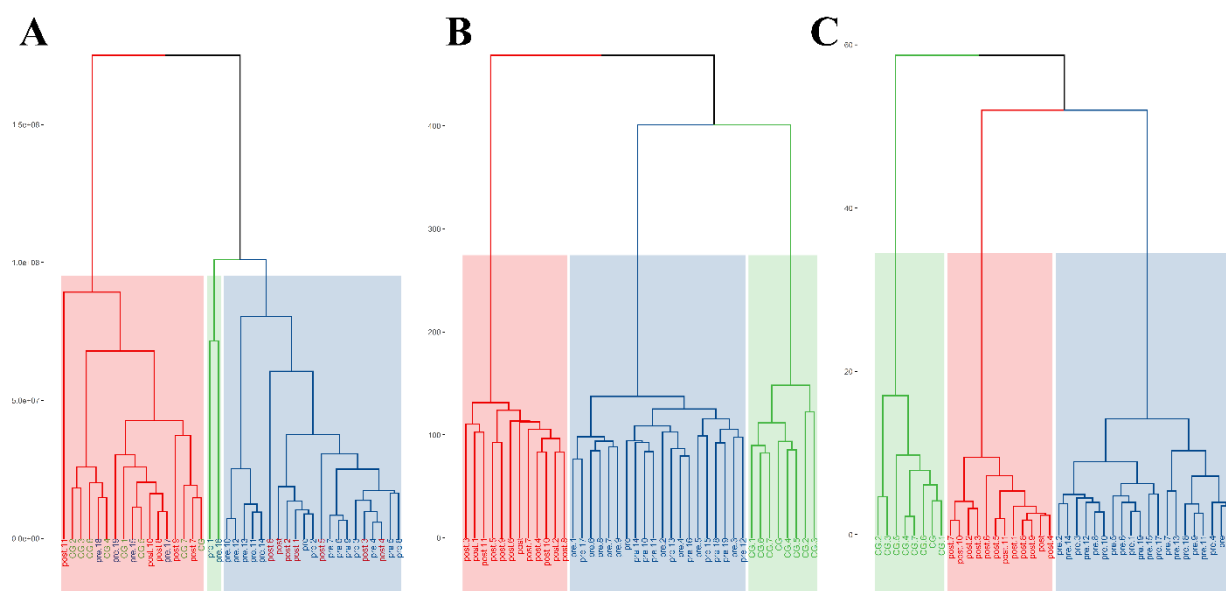


Рис. 3. Дендрограмма ИКА для экспериментальных данных. Евклидово расстояние и алгоритм объединения по Уорду («ward.D2»)

Более того, среднее значение AUROC изменилось с 0.683 до 1.0 после СПФ. Можно заключить, что данные правильно кластеризуются на графиках МГК и ИКА при снижении числа переменных и высоком значении AUROC.

Также достигнута высокая средняя оценка точности классификации при кросс-валидации для всех 36 наборов данных после EigenMS (более 90 %), следовательно, EigenMS можно успешно применять для коррекции сигнала и также как этап обработки данных перед проведением статистического анализа. Также наблюдался постоянный рост значения AUROC (до 1.0) и постепенное уменьшение числа переменных. Только

шесть наборов данных требовали некоторой корректировки алгоритма (рис.1). Были сформулированы правила для настройки алгоритма в случае, если параметры (рис. 1) не приводят к полному разделению точек данных на группы: контроль точности классификации и одного значения в СПФ (число предикторов) и РОП (выбор алгоритма). Таким образом, было показано, что наборы данных из разных источников с различными характеристиками могут быть обработаны единым образом (оценка точности при кросс-валидации $> 90 \%$, AUROC > 0.8 во всех случаях после всех этапов обработки). Основными отличительными способностями предлагаемого алгоритма являются минимальное количество операций и значительное снижение размерности (в 10 раз и более) с многомерной визуализацией.

Далее был проведен учет фенотипических факторов. В качестве иллюстрации выбраны семь из проанализированных наборов данных, которые содержат биологические факторы и/или несколько типов экспериментальных групп. Процедуру можно выполнить напрямую на этапе коррекции EigenMS (в аргументе «treatment» функции «eig_norm1»). В ряде случаев количество групп оставалось постоянным для статистической обработки и было изменено только при коррекции EigenMS. Оценка значения точности при кросс-валидации $> 90 \%$, средние значения AUROC > 0.9 , уменьшение размерности > 10 раз для всех семи наборов данных после всех этапов обработки. За исключением некоторых несоответствий в нескольких случаях, было достигнуто полное разделение точек данных на исходные экспериментальные группы методами ИКА и МГК. Стоит отметить, что большое количество многоуровневых коэффициентов не может быть применено в коррекции EigenMS (максимум два фактора с количеством уровней, равным половине от общего числа образцов, для функции «eig_norm2»). Это ограничение можно обойти с помощью пакета lmm2met (построение смешанных линейных моделей (СЛМ) для каждого метаболита с биологическими факторами в качестве ковариат). Были проведены вычисления для набора данных из пакета lmm2met. Учет факторов в EigenMS был проведен с теми же фиксированными факторами, как и для lmm2met. Затем обе таблицы данных сравнивались в статистической части алгоритма (ОСФ, СПФ, РОП). Около 70 % признаков были одинаковыми после СПФ для таблицы данных после EigenMS в сравнении с матрицей после lmm2met и 100 % предикторов совпадали после РОП для данных после EigenMS. Результаты позволяют предположить, что EigenMS как этап предварительной обработки для учета биологических факторов сопоставим с методом lmm2met.

Таким образом, предложен алгоритм для обработки метаболомных ВЭЖХ-МС данных от таблицы пиков до набора биомаркеров. Проверка разработанного алгоритма основана на тестировании репрезентативного количества метаболомных экспериментов (36 наборов данных в 21 исследовании) из общедоступных открытых источников данных. Также описана процедура учета биологических факторов. Программная реализация доступна в сети Интернет.

Четвертая глава посвящена созданию способа проведения метаболомного исследования методом жидкостного хроматомасс-спектрометрического ненаправленного профилирования низкого разрешения с двумя наборами данных.

Были проанализированы две выборки образцов (эксперимент по определению биомаркеров KPP) в оптимизированных ранее условиях. Партии анализировались с интервалом в шесть месяцев. Первая когорта (сорок образцов от трех групп пациентов (ЦГ, ЦГП, КГ)) использовалась как тестовая выборка для проверки алгоритма. Вторая когорта (сорок восемь образцов от двух групп пациентов (ЦГ, КГ)) использовалась как обучающая выборка для настройки алгоритма. После проведения профилирования (масс-спектрометр типа QQQ) образцов мочи из двух выборок были получены две таблицы пиков для каждого эксперимента. После ЗПЗ по алгоритму половины минимального значения были протестированы семь методов коррекции сигнала (по критериям ОСО и точности классификации). Вычисление ОСО и построение моделей МО аналогично главе 3. Расчет проводился для обучающей выборки.

Оптимальным алгоритмом среди 7 методов также оказался EigenMS (оценка точности при кросс-валидации > 85 %, процент пиков с ОСО < 30 % более 65 %). После EigenMS был применен алгоритм статистической обработки (рис. 1). Однако при выделении признаков алгоритмами РОП процесс снижения размерности был не эффективен. По этой причине был использован алгоритм vscc вместо РОП. Число переменных сократилось более чем в сто раз (с 4174 до 11), среднее значение AUROC изменилось с 0.550 до 0.908 после vscc.

После отбора переменных в обучающем наборе, провели ЗПЗ и расчет коррекции EigenMS для таблицы пиков тестового набора. Затем вручную отбирали биомаркерные сигналы в тестовом наборе из выбранных по алгоритму vscc в обучающем наборе. Выбор осуществляли по значению m/z (с точностью 0.1 Да) и временам удерживания (с точностью 0.3 мин). Из 11 биомаркерных сигналов, выбранных после проведения статистического анализа в обучающем наборе, удалось обнаружить 6 сигналов в таблице пиков тестового набора данных. Полученный усеченный набор данных

использовали для дальнейшей валидации. Диаграммы размаха для каждого соединения в тестовом наборе данных приведены на рис.4.

Объем выборки слишком мал для надежной оценки распределения признаков между группами в популяции, однако, видно, что во всех случаях 25 и 75 процентиля отличаются по крайней мере между двумя группами. Для двух биомаркерных сигналов (509.7/25.1708 и 615.5/24.461) – между всеми группами.

После применения всех этапов алгоритма, оценка точности классификации при кросс-валидации $> 90 \%$ и среднее значение AUROC > 0.9 для тестового набора данных. Завершающим этапом проверки применимости разработанного способа являлось многомерное проецирование методами МГК и ИКА (рис. 5).

График счетов первых двух главных компонент для усеченного набора данных тестового набора (рис. 5 А) позволяет визуально оценить расположение точек данных. Наблюдения каждой группы расположены близко друг к другу и пространственно отделены от других групп и, хотя имеется несколько выбросов, большую часть образцов можно верно кластеризовать по графику. Дендрограмма ИКА наблюдений (рис. 5 В) позволяет однозначно и правильно кластеризовать все образцы по трем экспериментальным группам. Разница высоты расщепления дендрограммы для двух и трех классов менее 5 единиц. Разница высоты расщепления дендрограммы для трех и четырех классов более 20 единиц. Дальнейшее разбиение на классы дает разницу менее 5-10 единиц. Даже не имея изначальной гипотезы о числе групп в наборе данных, структура дендрограммы указывает, что наиболее предпочтительное число классов равно трем, как и в исходном наборе данных.

Предложен способ определения потенциальных биомаркерных сигналов, методом ненаправленного профилирования с использованием ВЭЖХ-МС низкого разрешения. Отбор переменных оценивался методами МО, вычислением AUROC и многомерным проецированием. Способ разработан и апробирован на двух независимых наборах данных (обучающем и тестовом). Программная реализация доступна в сети Интернет.

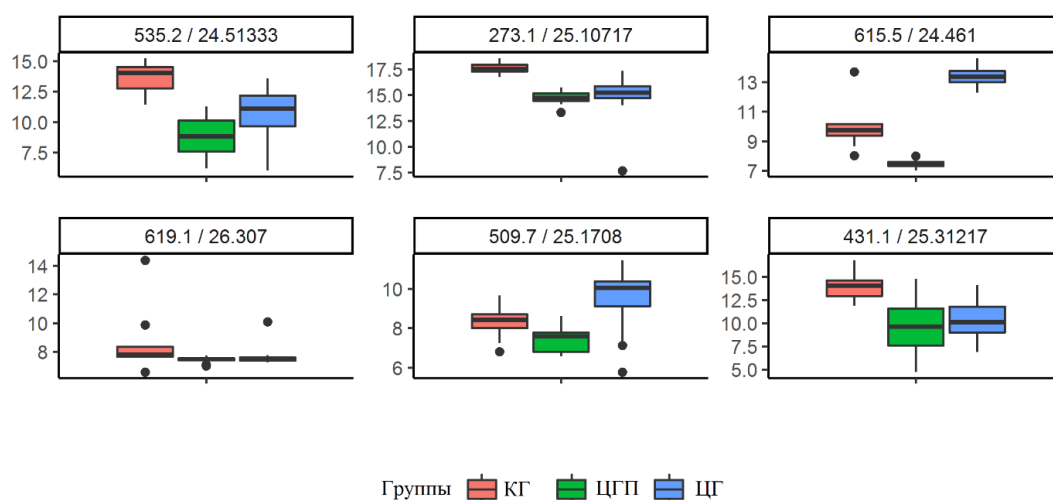


Рис.4. Диаграммы размаха биомаркерных сигналов из тестового набора. Обозначения предикторов: значение m/z / время удерживания, мин. На диаграммах приведены значения нормализованной площади пиков в у.е.

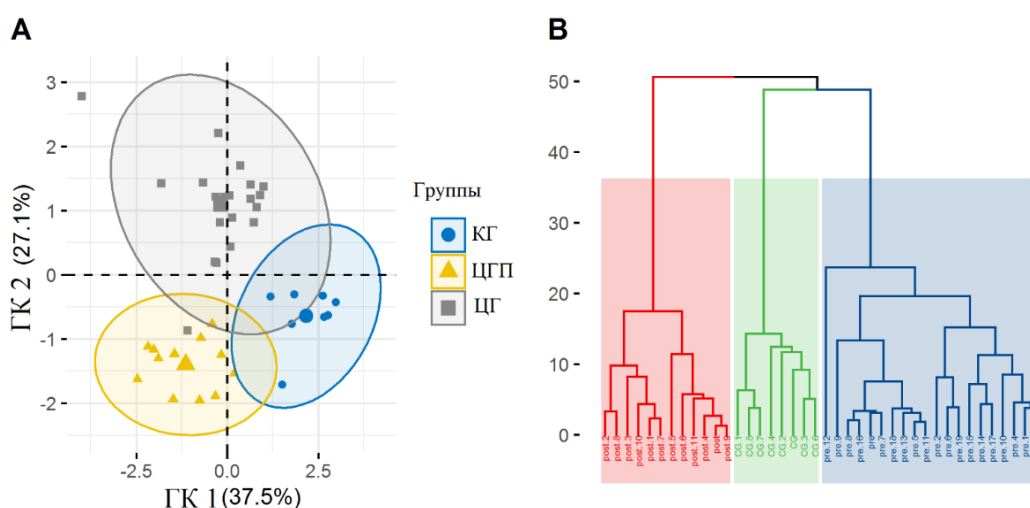


Рис. 5. Многомерное проецирование для эксперимента 1. А – график счетов МГК. В – дендрограмма ИКА (расстояние Манхэттена, объединение по алгоритму Уорда (ward.D2))

Пятая глава посвящена созданию и апробации программного инструмента для анализа данных нецелевого жидкостного хроматомасс-спектрометрического профилирования.

Было разработано программное обеспечение, которое можно рассматривать как дополнение к существующим инструментам, сфокусированное на обработке данных ненаправленного ВЭЖХ-МС профилирования в режиме сканирования. Для этого интегрировано 124 пакета на языке R в готовой для использования унифицированной манере. Программное обеспечение представляет собой коллекцию строк кода (скриптов) из девяти частей (вычислительных этапов) с комментариями, инструкцией и прочими дополнительными материалами. Разработанный программный инструмент был назван OUKS и свободно доступен в сети Интернет. Общая архитектура OUKS

спроектирована с возможностью расширения, изменения и настройки для пользовательских целей, а также быстрой разработки конвейера строк кода. Проведено сравнение разработанного программного обеспечения с другими инструментами ⁷. OUKS превосходит существующие решения в части отбора переменных, коррекции сигнала, учета факторов фенотипической изменчивости и обеспечивает исчерпывающую оценку и сравнение специфичных для метаболомики методов обработки сигнала. Кроме того, предложены новые методы коррекции сигнала и учета биологических факторов.

Новые методы коррекции разработаны на основе использования регрессионных алгоритмов KNN (QC-KNN), деревьев решений (QC-DT), бэггинга (QC-BT) и двух алгоритмов градиентного бустинга (QC-XGB, QC-CTB). Кратко, фактор коррекции в новых методах коррекции определялся для каждого метаболита отдельно путем построения регрессионной модели на образцах КК и последующем предсказании зависимой переменной для экспериментальных образцов. Всего реализовано пять новых методов коррекции сигнала (QC-KNN, QC-DT, QC-BT, QC-XGB, QC-CTB), каждый в трех режимах (одноэтапный, батчевый и двухпараметрический) и все в двух модификациях (деления, вычитания). Итого, тридцать уникальных алгоритмов коррекции дрейфа сигнала.

Методы учета биологической изменчивости реализованы на основе алгоритмов смешанных линейных моделей (СЛМ), линейной модели (ЛМ), обобщенных аддитивных моделей (GAM), обобщенных аддитивных смешанных моделей (GAMM), градиентного бустинга (GBM) и градиентного бустинга со смешанными эффектами (GBMM). Алгоритмы на основе аддитивных моделей и градиентного бустинга позволяют учитывать нелинейные зависимости и предложены впервые. Кратко, нескорректированные интенсивности (зависимая переменная) и ковариаты (независимые переменные: фиксированные и случайные эффекты) загружались в качестве входных данных в модель, предсказанные значения являлись скорректированными интенсивностями.

Для визуализации возможностей алгоритмов учета биологических факторов на рис. 6 приведены графики счетов МГК для сырых данных (А), после применения алгоритма СЛМ (Б), после GAMM (С) для данных профилирования из открытых источников данных с шестью группами. Моделирование по алгоритму СЛМ

⁷ <https://plyush1993.github.io/OUKS/Comparison.html>

распределяет точки данных в пространстве по группам, однако разделяющие границы классов линейные и узкие, эллипсы 95 % вероятности во многих случаях пересекаются. Сами облака точек данных по группам широкие и разрежены в пространстве. При GAMM моделировании облака данных сжимаются в окружности с высокой частотой заполнения, эллипсы также сжимаются и не пересекаются между собой. Разделяющие границы между облаками данных широкие и нелинейные. Таким образом, можно заключить, что новые нелинейные алгоритмы учета фенотипических факторов (по крайней мере GAMM моделирование) могут быть эффективны для многомерного проецирования данных и моделирования сигнала в случае многоклассовой задачи.

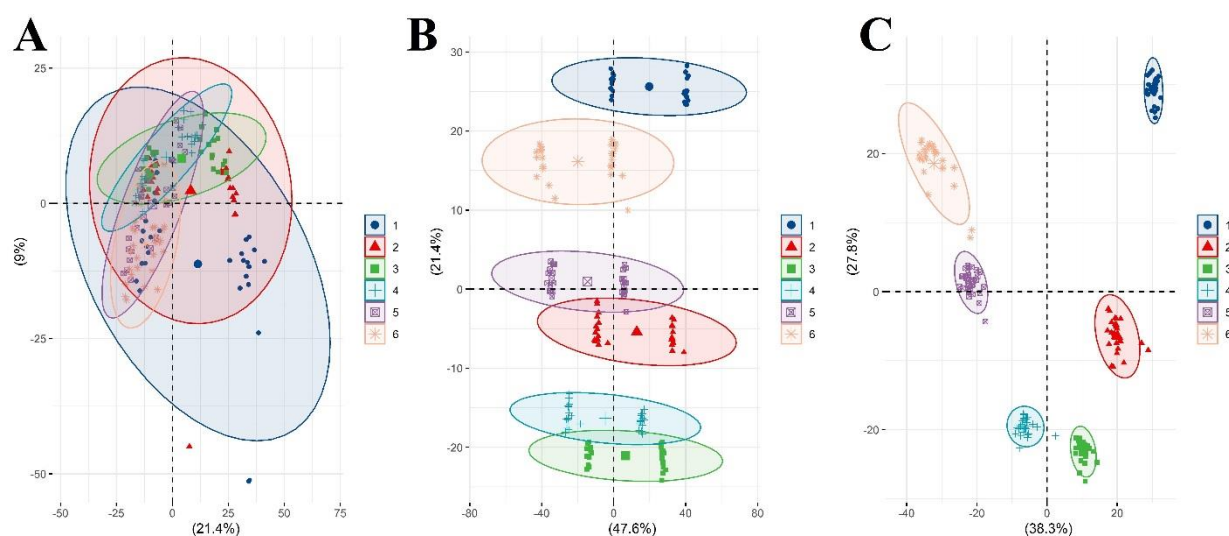


Рис. 6. Применение различных алгоритмов учета биологических факторов

Для визуализации возможностей OUKS проведено нецелевое ВЭЖХ-МС профилирование (масс-спектрометр типа ИЛ-ВПМА) ста двадцати четырех образцов мочи двух групп пациентов для поиска биомаркеров РМП в условиях аналогичных оптимизированным. После проведения профилирования были оптимизированы параметры автоматической разметки пиков и выравнивания в XCMS. Затем проводили поиск артефактных сигналов и подбирали оптимальный метод ЗПЗ из девяти алгоритмов по нормализованной среднеквадратичной ошибке, сумме квадратов остатков в прокрустовом анализе и коэффициенту корреляции. Оптимальным методом стало ЗПЗ по алгоритму RF. На следующем этапе проводили обработку сигнала, включающую в себя последовательное применение: коррекции сигнала, аннотации, фильтрации пиков, нормализации и учета биологических факторов.

В качестве критериев качества, используемых для выбора оптимального метода коррекции применяли шесть числовых показателей (среднее расстояние Бхаттачария для центров кластеров образцов КК, средняя мера несхожести для образцов КК в ИКА,

σ значение в gPCA, оценка вариативности методом PVCA, доля в % признаков с р-значением менее 0.05 в однофакторном дисперсионном анализе, доля признаков с ОСО менее 30 % для образцов КК). При сравнении по шести числовым показателям, а также визуальной оценке по диаграммам рассеяния (рис. 7) и графикам счетов МГК (рис. 8) оптимальным методом был признан разработанный алгоритм на основе градиентного бустинга (QC-XGB). Был протестирован сорок один метод коррекции сигнала.

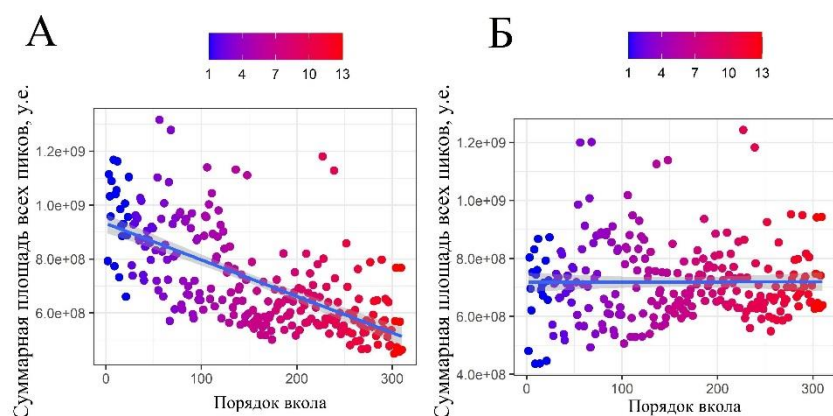


Рис. 7. Устранение дрейфа сигнала

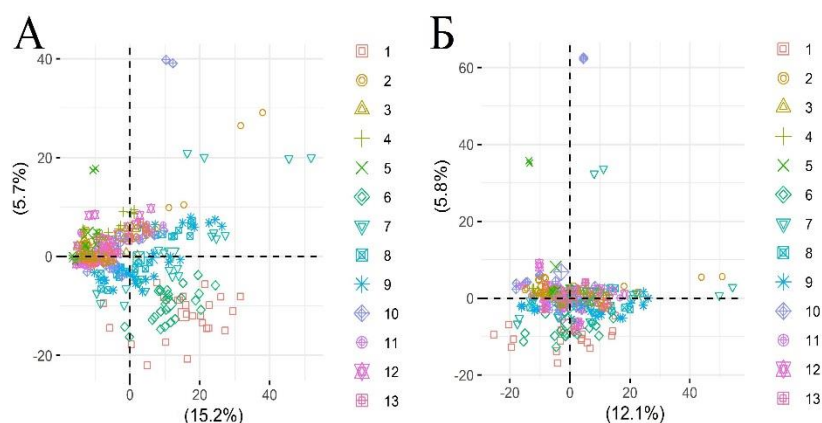


Рис. 8. Устранение батч-эффекта

7-8). Аннотацию сигналов в таблице пиков проводили по алгоритмам CAMERA, RAMClustR, xMSannotator и mWISE. Затем применили фильтрацию по ОСО и были удалены все не аннотированные сигналы. В завершении было проведено усреднение сигналов для повторных измерений.

Протестировано пять методов нормализации (общая сумма полезного сигнала, вероятностная нормировка частного, квантильная нормализация, медианный диапазон изменения и нормализация по креатинину). Распределение данных сравнивалось с сырыми данными и после QC-XGB на коробочных диаграммах относительной логарифмированной интенсивности. Медианное значение было близко к нулю при

Было достигнуто успешное устранение дрейфа сигнала (рис. 7), выражающегося в отрицательной корреляции общей интенсивности от номера вкола, и батч-эффекта (рис. 8), выражающегося в кластеризации точек данных в пространстве главных компонент по номеру аналитической последовательности, после применения QC-XGB коррекции (части Б на рис. 7-8) в сравнении с сырыми данными (части А на рис.

малой дисперсии вокруг медианного значения для сырых данных и после QC-XGB коррекции, поэтому нормализацию не применяли.

Для учета биологических факторов фенотипические данные (экспериментальная группа (КГ или РМП (ЦГ)), возраст, пол) и технические данные (номер аналитической последовательности) учтены как ковариаты, а интенсивность метаболитов (после QC-XGB и фильтрации) как зависимая переменная при моделировании по алгоритмам: СЛМ и линейным моделям. Для оценки эффективности моделей применяли десятиблочную кросс-валидацию с десятью повторами, оптимизация проводилась по показателю точности классификации, для каждой модели оптимизировали один параметр. Выбраны четыре модели МО: ближайшие сжатые центроиды (РАМ), PLS, SVM, RF. Использование СЛМ приводит к резкому увеличению точности вплоть до 100 %, поэтому применена корректировка биологических факторов по алгоритму СЛМ.

Завершающий этап исследования – поиск оптимального набора потенциальных веществ-биомаркеров методами статистического анализа. На основании предположения о линейной разделимости двух экспериментальных групп, сделанном при предварительном анализе графика счетов МГК, было решено проводить отбор переменных по пяти критериям задачи бинарной классификации: 1) признаки с p -значением менее 0.05 после ОСФ; 2) признаки с p -значением менее 0.05 в модерированном t -тесте с поправками на множественные сравнения; 3). диапазон изменения (fold change) больше 1.0 в абсолютном значении; 4). важность предиктора (VIP значение) в модели PLS более 1.0; 5). AUROC более 0.8. Отобраны восемь метаболитов по этим критериям.

Способность к визуальной кластеризации набора данных (после полного цикла обработки данных и отбора восьми переменных) методами ИКА (расстояние Манхэттена, объединение по алгоритму Уорда (ward.D2)) и МГК представлена на рис. 9 (В и Б соответственно). Образцы имеют стойкую тенденцию к верной кластеризации по экспериментальным группам. Общая дифференцирующая способность полного набора предикторов проиллюстрирована на volcano диаграмме (рис. 9, А), которая визуализирует распределение значений диапазона изменений и p -значений в t -тесте среди всех переменных. Способность выбранных потенциальных биомаркеров предсказывать статус заболевания была протестирована с помощью тестового набора

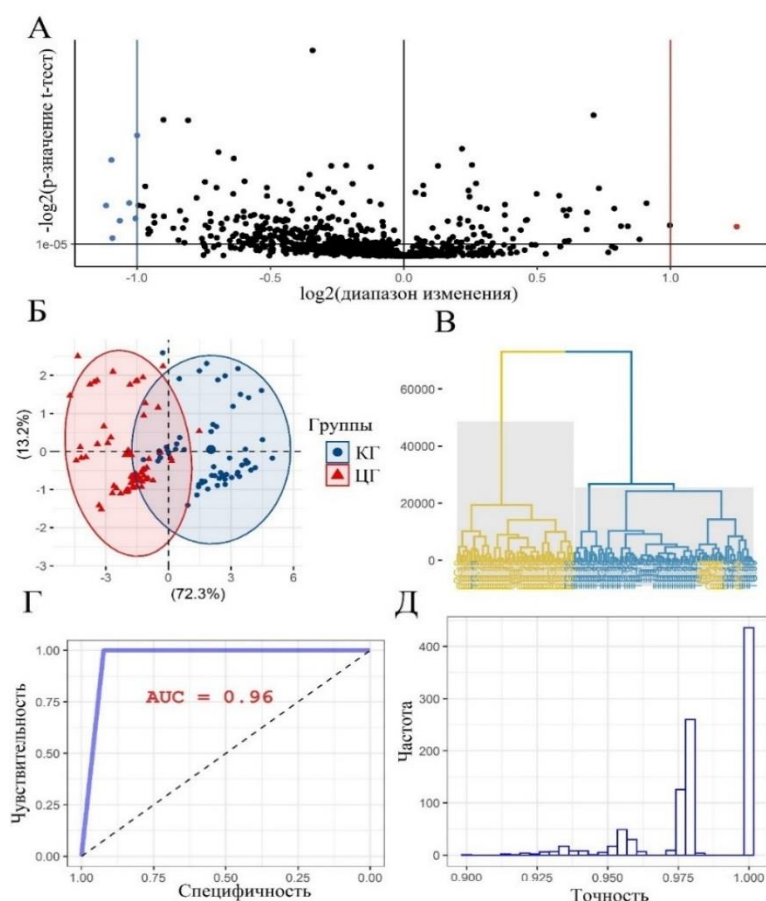


Рис. 9. Статистическая валидация набора биомаркеров. А – volcano диаграмма; **Б** – график счетов МГК; **В** – дендрограмма ИКА; **Г** – ROC кривая и AUROC (AUC); **Д** – гистограмма значений точности классификации

показателю точности классификации. Внутренняя проверка реализована путем расчета распределения значений показателя точности классификации на 1000 бутстрэповских выборках (рис. 9, Д) на полном наборе данных (124 образца). Для этого была построена аналогичная модель SVM. Оценки точности выбранного набора биомаркеров: AUROC > 0.9 для тестового набора данных, а распределение значений показателя точности классификации имеет максимум около значения равного 100 % и всегда > 90 %. Процедуры статистического анализа проводились для таблицы пиков после полного цикла обработки данных.

Все биомаркеры были предварительно идентифицированы в соответствии с трехэтапной последовательной процедурой: изучение экспериментального масс-спектра и результатов аннотации сигналов (во всех случаях предсказан однозарядный протонированный молекулярный ион); поиск по протонированным молекулярным ионам в базах данных HMDB, KEGG, RefMet; интерпретация MS² спектров программами CFM-ID, MetFrag. Информация о биомаркерах, статистические показатели и регуляционные пути представлены в таблице 1.

данных и процедуры бутстрэпа. Внешняя проверка проводилась путем построения модели SVM на обучающем наборе и предсказании на тестовом наборе данных с построением ROC кривой (рис. 9, Г). Исходный набор данных был разделен на две группы случайным образом и с сохранением баланса классов: 100 образцов для обучающего набора и 24 для тестового. Для оценки эффективности модели применяли десяти-блочную кросс-валидацию с десятью повторами, оптимизация проводилась по

Таблица 1. Детальная информация о потенциальных биомаркерах рака мочевого пузыря

Название	Значение <i>m/z</i>	Время удерживания, мин.	Точность определения массы, ppm; Из таблицы пиков (при МС ²)	Брутто- формула	Метаболомный путь	PLS	AUROC	Диапазон изменения	t-тест, р- значение	СПФ, р- значение
Бензил 2-метил-3-оксобутаноат	207.1020	14.1	2 (11)	C ₁₂ H ₁₄ O ₃	Метаболизм аминокислот	2.0	1.0	1.0	2.3×10 ⁻⁵⁰	6.4×10 ⁻²¹
LysoPA(20:4(8Z,11Z,14Z,17Z)/0:0) окисленная форма лизофосфатидной кислоты	475.2466	16.1	2 (4)	C ₂₃ H ₃₉ O ₈ P	Метаболизм жирных кислот/липидов	1.8	1.0	1.1	3.2×10 ⁻⁴⁰	4.8×10 ⁻²⁰
Треонинил- глутамин	248.1296	3.7	22 (24)	C ₉ H ₁₇ N ₃ O ₅	Метаболизм аминокислот	1.6	0.9	1.0	1.4×10 ⁻²²	9.0×10 ⁻²²
QH(2) Убихинон	253.1465	14	12 (2)	C ₁₄ H ₂₀ O ₄	Окисление жирных кислот	1.6	0.9	1.1	1.6×10 ⁻²¹	5.2×10 ⁻¹⁵
Аспарагинил- Глутамин	261.0965	1.3	-87 (-30)	C ₉ H ₁₆ N ₄ O ₅	Метаболизм аминокислот	1.6	0.9	1.2	7.0×10 ⁻¹³	5.8×10 ⁻¹⁵
6-Кето- деcanoилкарнитин	330.2259	14.3	-5 (-1)	C ₁₇ H ₃₁ NO ₅	Окисление жирных кислот	1.4	0.9	1.1	2.9×10 ⁻¹⁵	8.8×10 ⁻¹⁴
2-транс, 4-цис- Декадиеноилкарнитин	312.2150	15.6	-6 (-1)	C ₁₇ H ₂₉ NO ₄	Окисление жирных кислот	1.4	0.8	1.0	3.1×10 ⁻¹⁶	1.3×10 ⁻¹⁰
(6R, 8Z) -6- Гидрокси-3- оксотетрадеценвая кислота	257.1506	18.8	-94 (23)	C ₁₄ H ₂₄ O ₄	Метаболизм жирных кислот/липидов	1.2	0.8	1.1	3.3×10 ⁻⁰⁸	6.0×10 ⁻⁰⁹

Таким образом, разработанное программное обеспечение апробировано на экспериментальных данных, проведена подробная и исчерпывающая процедура обработки данных, выявлены значимые потенциальные биомаркеры, проведена их статистическая валидация, идентификация и изучена биологическая роль.

В заключении описаны основные результаты, полученные в работе.

Исследования, описанные в главах 3 и 4, посвящены разработке способов к выделению набора биомаркеров. Первый этап в обоих подходах – проведение EigenMS коррекции, которая одновременно может рассматриваться как этап обработки данных. После последовательно применяются статистические тесты, строятся модели МО, затем переменные ранжируются по важности, применяется алгоритм автоматического отбора переменных (РОП или vscs). В результате признаков пространство последовательно снижается для выявления информативных предикторов. Контроль процедуры снижения размерности проводился путем расчета значений AUROC и при многомерном проецировании (МГК, ИКА). Эффективность применения EigenMS коррекции – по расчету значений ОСО и оценке точности классификации при кросс-валидации.

Разработан программный инструмент для исчерпывающего анализа и обработки данных метаболомного профилирования методом ВЭЖХ-МС (глава 5). Программное обеспечение интегрирует 124 пакета на языке R. Программный инструмент предлагает уникальный функционал для проведения и тестирования коррекции сигнала детектора и устранения батч-эффекта, учета факторов биологической изменчивости и отбора переменных и бесплатно доступен в сети Интернет. Также разработаны новые методы коррекции сигнала детектора и учета фенотипических факторов. Новый метод коррекции, на основе алгоритма градиентного бустинга, превзошел остальные при мультипараметрической оценке. Показаны преимущества использования нового нелинейного алгоритма учета биологических факторов при анализе метаболомных данных с числом экспериментальных групп больше двух.

Проведено ненаправленное метаболомное профилирование образцов мочи методом ВЭЖХ-МС высокого разрешения для поиска биомаркеров рака мочевого пузыря. При обработке данных был использован разработанный программный инструмент, данные профилирования доступны в сети Интернет. Выполнена подробная оптимизация всех этапов обработки данных, включая статистическую валидацию методами одномерной и многомерной статистики, идентификацию метаболитов и установление биологической роли биомаркеров. В результате предложен список потенциальных биомаркеров рака мочевого пузыря.

ВЫВОДЫ

1. Разработан способ анализа данных метаболомного профилирования. Способ апробирован на экспериментальном наборе данных профилирования, а также на 36 данных профилирования из открытых источников данных. После выполнения всех этапов алгоритма размерность данных снижается в 10 раз и более при оценке точности классификации более 90 % и AUROC более 0.8. Разработанный способ был адаптирован для учета факторов фенотипической изменчивости и проверен на 7 наборах данных.

2. Предложена схема обработки и анализа данных профилирования, полученных методом ВЭЖХ-МС низкого разрешения. Схема включает в себя последовательное использование двух независимых экспериментальных наборов данных. При апробации выявлены потенциальные биомаркерные сигналы для диагностики колоректального рака. После обработки данных достигнута оценка точности классификации более 90 %, AUROC более 0.9 и правильная кластеризация образцов методами многомерного проецирования на независимом наборе данных.

3. Разработано универсальное программное обеспечение, позволяющее обрабатывать и анализировать данные ненаправленного метаболомного профилирования. Программный инструмент включает в себя все этапы от обработки сырых данных до набора веществ-биомаркеров и превосходит описанные ранее программы в части проведения коррекции сигнала, учета факторов биологической изменчивости и отбора переменных.

4. Проведено метаболомное профилирование образцов мочи методом ВЭЖХ-МС высокого разрешения для определения потенциальных биомаркеров рака мочевого пузыря. Оценка точности классификационной модели, построенной по выбранным биомаркерам после обработки данных, более 90 % на 1000 выборок с возвратом и AUROC более 0.9 на тестовом наборе данных.

5. Разработаны новые методы коррекции сигнала и учета биологических факторов. Новый метод коррекции превосходит описанные ранее при мультипараметрической оценке. Показана эффективность нового нелинейного алгоритма учета биологических факторов при решении задачи многоклассовой классификации.

Основные результаты работы изложены в следующих публикациях:

Научные статьи, опубликованные в рецензируемых научных журналах, индексируемых в базах данных Web of Science, Scopus, RSCI и рекомендованных для защиты в диссертационном совете МГУ по специальности 02.00.02 – «Аналитическая химия»:

1. **Плющенко И.В.**, Шахматов Д.Г., Родин И.А. Алгоритм сочетания хромато-масс-спектрометрического ненаправленного профилирования и многомерного анализа для выявления веществ-маркеров в образцах сложного состава // Заводская лаборатория. Диагностика материалов. 2020. Т. 86. № 7. С. 12-19. (Импакт-фактор RSCI – 0.243), 70%. (перевод: **Plyushchenko I.V.**, Shakhmatov D.G., Rodin I.A. Algorithm of Combining Chromatography–Mass Spectrometry Untargeted Profiling and Multivariate Analysis for Identification of Marker Substances in Samples of Complex Composition. // Inorg Mater. 2021. V. 57. №. 14. P. 1397–1403. (Импакт-фактор Web of Science – 0.902, **Q4**), 70%)

2. **Plyushchenko I.**, Shakhmatov D., Bolotnik T., Baygildiev T., Nesterenko P., Rodin I. An approach for feature selection with data modelling in LC-MS metabolomics // Analytical Methods. 2020. V. 12. № 28. P. 3582-3591. (Импакт-фактор Web of Science – 2.716, **Q2**), 60%.

3. **Plyushchenko I. V.**, Fedorova E. S., Potoldykova N. V., Polyakovskiy K. A., Glukhov A. I., Rodin I. A. Omics Untargeted Key Script: R-Based Software Toolbox for Untargeted Metabolomics with Bladder Cancer Biomarkers Discovery Case Study // Journal of Proteome Research. 2021. acs.jproteome.1c00392. (Импакт-фактор Web of Science – 4.352, **Q1**), 60%.

Иные публикации:

4. **Plyushchenko I.**, Bolotnik T., Shakhmatov D., Achkasov S., Sushkov O., Shpigun O., Rodin I. Typical LC-MS metabolomics workflow for profiling urine samples of patients with colorectal cancer / XXII International Mass Spectrometry Conference, Флоренция, Италия, 26-31 августа 2018, P. 411.

5. **Плющенко И.В.**, Болотник Т.А., Шахматов Д.Г., Ачкасов С.И., Сушков О.И., Шпигун О.А., Родин И.А. Хромато-масс-спектрометрическое ненаправленное профилирование для выявления биомаркеров рака / Материалы V Всероссийского симпозиума «Разделение и концентрирование в аналитической химии и радиохимии» с международным участием, Краснодар, Россия, 7-13 октября 2018, С. 167.

6. **Plyushchenko I.V.**, Rodin I.A. Bioassay Classification Study via LC-MS and Machine Learning in Conjunction with Dimensionality Reduction / Mass Spectrometry & Advances in the Clinical Lab 2019 EU 6th European Congress, Зальцбург, Австрия, 22-26 сентября 2019, Р. 30.

7. **Плющенко И.В.**, Родин И.А. Хромато-масс-спектрометрическое ненаправленное профилирование и интеллектуальный анализ данных для выявления потенциальных биомаркеров / Материалы III Всероссийской Конференции по аналитической спектроскопии с международным участием, г. Туапсе, Россия, 29 сентября - 5 октября 2019, С. 57.

8. **Плющенко И.В.**, Родин И.А. Сочетание алгоритмов предобработки данных, машинного обучения и отбора переменных для выбора оптимального признакового пространства в классификации биообразцов методом хроматомасс-спектрометрии / Материалы VIII Всероссийской Конференции с международным участием «Масс-спектрометрия и ее прикладные проблемы», Институт физической химии и электрохимии им. А.Н. Фрумкина РАН, г. Москва, Россия, 14-18 октября 2019, С. 146.

9. **Плющенко И.В.**, Родин И.А. Валидация результатов хроматомасс-спектрометрического ненаправленного профилирования / Материалы IV Всероссийской конференции с международным участием "Аналитическая хроматография и капиллярный электрофорез", Краснодар, Россия, 28 сентября - 2 октября 2020, С. 114.

10. **Плющенко И.В.**, Фёдорова Е.С., Глухов А.И., Родин И.А. Потенциальные неинвазивные биомаркеры рака мочевого пузыря / Материалы VI Всероссийского симпозиума «Разделение и концентрирование в аналитической химии и радиохимии» с международным участием, Краснодар, Россия, 26 сентября - 2 октября 2021, С. 168.

Автор выражает искреннюю признательность и благодарность научному руководителю д.х.н. Родину И.А. за участие и помощь в постановке и решении задач, ценные советы и замечания при подготовке диссертации; д.х.н., доц., проф. РАН Проскурнину М.А. за ценные советы и замечания при подготовке диссертации; инж. Фёдоровой Е.С. за помощь в обсуждении результатов работы и подготовке печатных работ; д.б.н. проф. Глухову А.И. и к.м.н. Шахматову Д.Г. за предоставленные образцы; всем членам лабораторий масс-спектрометрии и хроматографии за постоянные помощь в работе и поддержку.