

*На правах рукописи*

**Вавилов Никита Эдуардович**

**ДЕТЕКЦИЯ НИЗКОПРЕДСТАВЛЕННЫХ БЕЛКОВ 18 ХРОМОСОМЫ  
МЕТОДАМИ ТАРГЕТНОЙ ПРОТЕОМИКИ**

1.5.4. – биохимия

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени

кандидата биологических наук

Москва, 2023

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича» (ИБМХ)

**Научный руководитель:** доктор биологических наук, профессор РАН  
**Згода Виктор Гаврилович**

**Официальные оппоненты:** **Шевченко Валерий Евгеньевич,**  
доктор биологических наук, профессор,  
ФГБУ «Национальный медицинский  
исследовательский центр онкологии имени  
Н.Н. Блохина» Министерства здравоохранения  
Российской Федерации, главный научный консультант

**Лазарев Василий Николаевич,**  
доктор биологических наук, доцент,  
ФГБУ «Федеральный научно-клинический центр  
физико-химической медицины имени академика  
Ю.М. Лопухина Федерального медико-биологического  
агентства», заместитель Генерального директора по  
научной работе, руководитель лаборатории

**Ведущая организация:** Федеральное государственное бюджетное учреждение  
науки «Федеральный исследовательский центр  
питания, биотехнологии и безопасности пищи»

Защита состоится «5» октября 2023 г. в 11.00 часов на заседании диссертационного совета 24.1.172.01 (Д 001.010.01) на базе Федерального государственного бюджетного научного учреждения «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича» по адресу: 119121, Москва, ул. Погодинская, д. 10, стр. 8.

С диссертацией можно ознакомиться в библиотеке ИБМХ и на сайте [www.ibmc.msk.ru](http://www.ibmc.msk.ru).

Автореферат разослан «\_\_» \_\_\_\_\_ 2023 г.

Ученый секретарь диссертационного совета,  
кандидат химических наук



Карпова Е.А.

# 1. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

## 1.1. Актуальность проблемы, цель и задачи

Проект «Протеом человека» стартовал в 2010 году на международном конгрессе HUPO в Сиднее. Целью проекта являлось объединение усилий ученых всего мира для поиска и описания белков, кодируемых генами человека. В дальнейшем для разделения усилий по странам предложен хромосоמוцентричный проект «Протеом человека» – C-HPP (Chromosome-centric Human Proteome Project). Исследователи из России под руководством академика А. И. Арчакова провели биоинформационный анализ по уровню медицинской значимости белков, кодируемых генами человека. На основе полученной информации в качестве объекта исследований были выбраны продукты генов 18 хромосомы человека. Организованный консорциум Российской части проекта занимался исследованием не только белков, но и транскриптов, так как для некоторых белок-кодирующих генов не были зарегистрированы соответствующие последовательности мРНК, кодируемые генами 18 хромосомы человека. В ходе проекта разработан таргетный метод регистрации белков, который позволяет детектировать целевые белки в биоматериале человека с высокой селективностью и чувствительностью. С помощью разработанной методики измерены абсолютные концентрации белков, кодируемых генами 18 хромосомы человека, в клетках линии НерG2, клетках печени человека и сыворотке крови человека. Диапазон измеренных концентраций составлял от  $10^{-6}$  М до  $10^{-18}$  М. Наиболее высоких показателей чувствительности от  $10^{-14}$  М до  $10^{-18}$  М, удалось достичь с помощью методики необратимого связывания аналитов в растворе.

Проблема чувствительности в современной протеомике особенно остро стоит при исследовании комплексных биологических образцов: сыворотки крови человека, клеточных линий и тканей человека. Протеом человека содержит около 20359 белков, а с учетом всех возможных мутаций, посттрансляционных модификаций, одноаминокислотных полиморфизмов, по разным оценкам содержит от 0,6 до 6 миллионов белков (Ponomarenko et al., 2020). Кроме того, белки в клетках и тканях организма находятся в разной концентрации от  $10^{-6}$  М до  $10^{-12}$  М и ниже, что накладывает дополнительные ограничения и затрудняет их детекцию в низкой концентрации (Millioni et al., 2011). Показано, что с помощью методики панорамного масс-спектрометрического анализа в комплексном биологическом образце клеточной линии человека регистрируется около 15000 пептидов. Однако, анализ всех изотопных

кластеров показал, что в одном образце можно обнаружить около 115000 пептидных ионов (Michalski et al., 2011). Таким образом, панорамный масс-спектрометрический анализ регистрирует лишь 13% тех белков, которые действительно находятся в образце. Принципиально иной подход демонстрируют методы таргетной протеомики (Vavilov et al., 2020). Данные методики анализа характеризуются тем, что исследователь еще перед началом эксперимента выбирает пептидные ионы, которые будет регистрировать прибор. При таком подходе удастся добиться более высоких показателей чувствительности, что позволяет зарегистрировать те белки биологического образца, которые не были идентифицированы при панорамном масс-спектрометрическом анализе.

В настоящей работе представлены результаты, которые демонстрируют основные факторы, лимитирующие чувствительность протеомных методов, а также проведен поиск путей преодоления этих факторов с использованием биохимических методов фракционирования и масс-спектрометрических методов анализа. Причины ограничений протеомных методов изучены на стандартном объекте - упрощенной белковой системе, которая представляет собой универсальный протеомный стандарт, состоящий из эквимольной смеси 47 очищенных белков человека (UPS1). В ходе исследования разработаны методы масс-спектрометрической детекции белков с использованием предварительного фракционирования образца и последующим анализом каждой фракции. Методика апробирована и охарактеризована на биологической системе HepG2, клеточной линии гепатоцеллюлярной карциномы человека. В дальнейшем метод использован для оценки стабильности и воспроизводимости измерений при анализе клинических образцов печени человека от здоровых доноров.

**Цель** данной работы – определить факторы, лимитирующие детекцию белков при протеомных методах анализа, и разработать методы для высокочувствительной масс-спектрометрической детекции низкокопийных белков, кодируемых генами 18 хромосомы человека.

Для достижения указанной цели поставлены следующие **задачи**:

1. Определить факторы, лимитирующие детекцию белков, при использовании таргетного масс-спектрометрического анализа на примере стандартного объекта - упрощенной белковой системы UPS1.

2. Разработать методы преодоления факторов, лимитирующих детекцию белков, и апробировать их на биологическом объекте – клеточной линии HepG2.
3. Оценить эффективность метода и провести анализ протеомных и транскриптомных данных, полученных на клетках линии HepG2.
4. Применить высокочувствительный метод таргетного масс-спектрометрического анализа с использованием двумерного фракционирования пептидов для оценки стабильности количественного анализа и детекции низкокопийных белков, кодируемых генами 18 хромосомы человека, в клинических образцах печени от здоровых доноров.

## **1.2. Научная новизна работы**

Определены факторы, лимитирующие детекцию низкокопийных белков в биологических образцах. Впервые разработан и применен высокочувствительный метод таргетного протеомного анализа (2D SRM) с применением двумерного фракционирования комплексных биологических образцов. Детально изучены возможности метода, его преимущества перед стандартными одномерными методами анализа на примере клеточной линии HepG2. Показана возможность обнаружения белков, которые не детектируются с помощью технологий высокопроизводительного транскриптомного анализа RNA-seq (Illumina). В работе с помощью нового метода обнаружены и измерены точные концентрации 129 белков, кодируемых генами 18 хромосомы человека, в гетерогенных клинических образцах печени человека.

## **1.3. Теоретическая и практическая значимость работы**

В работе показана недостаточная чувствительность методов современного протеомного анализа при детекции белков в комплексных биологических образцах. В исследовании было обнаружено, что биологические системы содержат гораздо большее количество белков, чем удается зарегистрировать стандартными протеомными методиками. Разработан новый подход, с помощью которого можно обнаружить белки, которые не детектируются при исследовании образцов методами стандартного таргетного и панорамного МС анализа. Показано, что он позволяет обнаружить очень низкие концентрации белка в клетке, до 100 копий на клетку. Данные низкокопийные белки, в основном, участвуют в таких важных процессах в клетке как передача сигнала, поэтому их детекция является биологически значимой. Белки, обнаруживаемые при помощи нового метода, позволяют получить более

полную и осмысленную информацию об исследуемой биологической системе по сравнению со стандартными методиками МС анализа.

Используемый в работе подход может быть применен к исследованию любых комплексных биологических образцов, где встречается проблема большого динамического концентрационного диапазона.

#### **1.4. Личный вклад автора**

Работа выполнена в лаборатории системной биологии ИБМХ им. В.Н. Ореховича в период с 2018 по 2022 год. Автор проводил поиск и анализ литературы, планирование экспериментов, пробоподготовку к протеомному анализу, масс-спектрометрический анализ на системах ВЭЖХ/МС: Orbitrap, Triple Quad, и хроматографический анализ на системе ЖХ/УФ, а также биоинформатический анализ полученных протеомных данных и подготовку результатов работы к публикации. Автор благодарит коллег за синтез изотопно-меченных пептидных стандартов, работу с клеточными культурами.

#### **1.5. Основные положения, выносимые на защиту**

1. Ограничения таргетных протеомных методов заключаются в лимитирующих факторах, которые ограничивают детектируемость целевых белков на фоне биологической матрицы. Наиболее значимыми факторами являются: концентрация целевых белков, наличие высокого динамического диапазона концентраций белков матрицы, выбора протеотипического пептида в качестве внутреннего стандарта для детекции целевого белка.
2. Разработанный протеомный подход с использованием дополнительной стадии фракционирования пептидов снижает комплексность образца и позволяет исследовать не все белки биологического образца в течение одного анализа, а определенную обогащенную фракцию. Такая методика позволяет преодолеть ограничения стандартных методов протеомного анализа и регистрировать низкокопийные белки протеома человека и белки, кодируемые генами 18 хромосомы человека.
3. В работе предложены варианты дальнейшего улучшения методики и увеличения глубины покрытия протеома человека и белков, кодируемых генами 18 хромосомы человека. Свойства пептидов, определяющие эффективность их ионизации, являются ключевыми свойствами, от которых зависит чувствительность таргетного анализа при разработке метода SRM SIS анализа.

## **1.6. Степень достоверности и апробация результатов**

Достоверность полученных результатов обеспечена использованием методов обработки и оценки данных, соответствующих современным научным критериям. Обсуждение результатов проведено с учетом актуальных исследований, опубликованных в области биохимии, протеомики и масс спектрометрии. Положения и выводы, изложенные в диссертации, подтверждены публикациями и следуют из результатов исследований, проведенных диссертантом. Основные положения работы опубликованы в рецензируемых научных изданиях, а также представлены в виде постерных докладов на конгрессе международной организации «Протеом человека» (HUPO 2021), конференции «Федерация европейских биохимических сообществ» (FEBS 2021).

## **1.7. Публикации**

По теме диссертационной работы опубликовано 6 работ, из которых 4 статьи представлены в рецензируемых научных журналах и 2 публикации находятся в трудах конференций.

## **1.8. Объем и структура диссертации**

Диссертация изложена на 112 страницах машинописного текста, содержит 4 таблицы и 36 рисунков. Состоит из следующих глав: список используемых сокращений, введение, цель и задачи, обзор литературных источников, материалы и методы, результаты и обсуждение, заключение, выводы, список литературы, а также включает благодарности, финансирование.

## **2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ**

В данной работе выполнен поиск белков, кодируемых генами 18 хромосомы человека, с помощью панорамных и таргетных протеомных методов. Основные объекты исследования: упрощенная белковая система UPS1 (Sigma-Aldrich, США), клетки линии HepG2 и образцы ткани печени человека от здоровых доноров.

Для белков системы UPS1 разработана соответствующая SRM методика с оптимизацией параметров масс-спектрометрического анализа для каждого индивидуального пептида. Затем проведен анализ белковой системы на фоне биологического образца в виде гидролизата белков *E. coli* для имитации биологических условий, а также в отсутствие гидролизата *E. coli*. Для эксперимента использовали несколько разведений UPS1 от  $10^{-9}$  М до  $10^{-13}$  М. Также, для образцов в концентрации

$10^{-12}$  М проведено концентрирование в 100 раз и показана возможность улучшения результатов анализа. Один из образцов, который наиболее близко имитирует высокий динамический диапазон концентраций, был фракционирован с помощью обращенно-фазовой хроматографии в щелочных условиях. Затем каждая фракция была проанализирована на предмет наличия белков системы UPS1.

Следующим этапом работы являлась разработка методики, которая повышает производительность масс-спектрометрического анализа образцов клеток линии HepG2. На этом этапе проводился комплексный анализ с помощью методов панорамной и таргетной масс-спектрометрии. Панорамный анализ выполнен на масс спектрометре Q-Exactive HF с масс анализатором типа Orbitrap, оборудованный pESI источником ионизации, объединенный в систему с UHPLC хроматографом, оборудованным нано насосом, автосемплером, термостатом колонки, загрузочным микро насосом, аналитической колонкой (Acclaim Pep-Map RSLC, Thermo Fisher Scientific, Rockwell, IL, USA). Во время анализа скорость потока составляла 0,3 мкл/мин, для элюции пептидов с колонки использовался линейный градиент раствора 0,1% муравьиной кислоты в ddH<sub>2</sub>O и раствора 80% ацетонитрила, 0,1% муравьиной кислоты в ddH<sub>2</sub>O. Таргетный анализ выполнен на масс-спектрометре Triple Quad 6495 Agilent, оборудованный Dual Jet Source источником ионизации, объединенный в единую систему с хроматографом Agilent 1200 Series, оборудованным насосом с нормальным потоком, автосемплером, термостатом колонки, аналитической колонкой Eclipse C18 (2.1 × 50 mm, 1.8 μm particles size, Agilent, Palo Alto, CA, USA). Во время анализа скорость потока составляла 300 мкл/мин, для элюции пептидов с колонки использовался линейный градиент раствора 0,1% муравьиной кислоты, 0,01% трифторуксусной кислоты в ddH<sub>2</sub>O и раствора 80% ацетонитрила, 0,1% муравьиной кислоты, 0,01% трифторуксусной кислоты в ddH<sub>2</sub>O. В качестве внутренних стандартов для таргетного анализа использованы синтетические пептиды, которые содержат аминокислоты, меченные стабильными изотопами. Вторая часть работы заключалась в разработке и применении технологии двумерного фракционирования и оценке ее эффективности. Первое фракционирование белков клеточной линии HepG2, гидролизованных трипсином выполнено на хроматографической системе Agilent 1200 Series, оборудованной коллектором фракций. Разделение пептидов проводилось в щелочных условиях на хроматографической колонке со стационарной фазой C18 (4,6 x 250 мм, 5 мкм, Waters XBridge C18, Ireland). Во время анализа проходил сбор фракций, суммарно получено 24 фракции за один цикл работы прибора. Затем каждая фракция

проанализирована панорамным и таргетным методами. Также для клеточной линии HepG2 проведен транскриптомный анализ с применением платформы Illumina для анализа сходимости транскриптомных и протеомных данных. Также проведена оценка эффективности ионизации пептидов, используемых в качестве изотопно-меченных стандартов. Эффективность ионизации оценивали путем добавления пептидных стандартов в гидролизат белков клеток линии HepG2 в нескольких разведениях до полного исчезновения сигнала.

В дальнейшем проведен такой же эксперимент с применением двумерного фракционирования и без применения двумерного фракционирования на сложном гетерогенном биологическом материале – ткани печени человека от здоровых доноров. Технология опробована на трех биологических образцах с целью оценки биологической вариабельности белков и соответствующих им транскриптов.

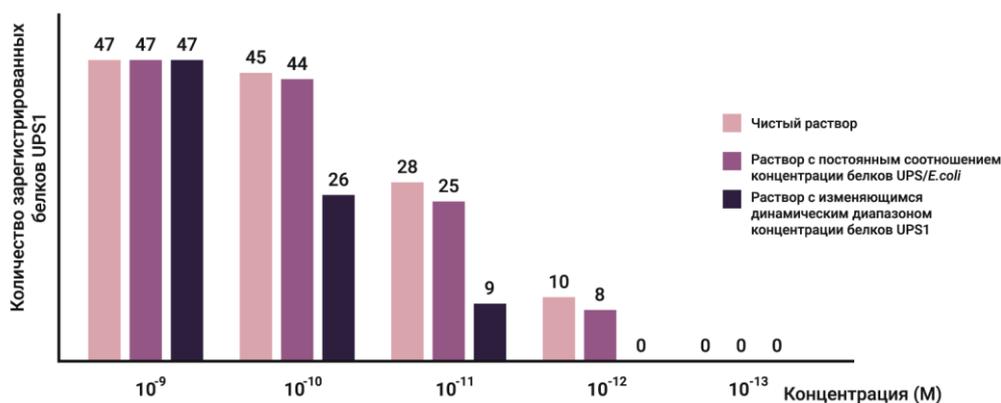
### **3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ**

#### **3.1. Ограничения методов протеомных технологий (SRM)**

Причинами низкого уровня идентификации белков протеома человека является недостаточная чувствительность и ограничения по динамическому диапазону измерений для протеомных методов при анализе комплексных биологических образцов. Для подтверждения этого был поставлен эксперимент со стандартным объектом UPS1. Раствор гидролизата белков UPS1 приготовлен в разведениях от  $10^{-7}$  М до  $10^{-13}$  М в виде трех модельных образцов:

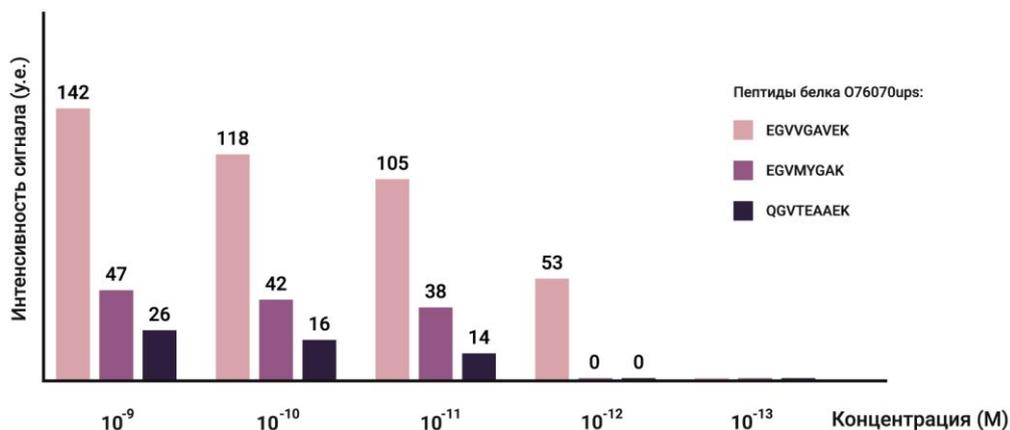
1. Чистый раствор гидролизата белков UPS1.
2. Раствор с постоянным соотношением концентраций белков UPS1/*E. coli* (раствор гидролизата белков UPS1 на фоне гидролизата белков *E. coli*, при этом концентрация UPS1 и *E. coli* снижалась равномерно в ходе приготовления разведений).
3. Раствор с изменяющимся динамическим диапазоном концентрации белков UPS1 (концентрация белков триптического лизата *E. coli* оставалась постоянной, при этом концентрация UPS1 снижалась в ходе приготовления разведений).

Последняя модель наиболее реально имитирует биологический образец. Было показано, что при таргетном протеомном анализе гидролизата UPS1 регистрируется 100% белков образца, которые находятся в концентрации  $10^{-9}$  М (рисунок 1).



**Рисунок 1.** Количество зарегистрированных белков системы UPS1 в зависимости от концентрации целевых белков в модельных образцах.

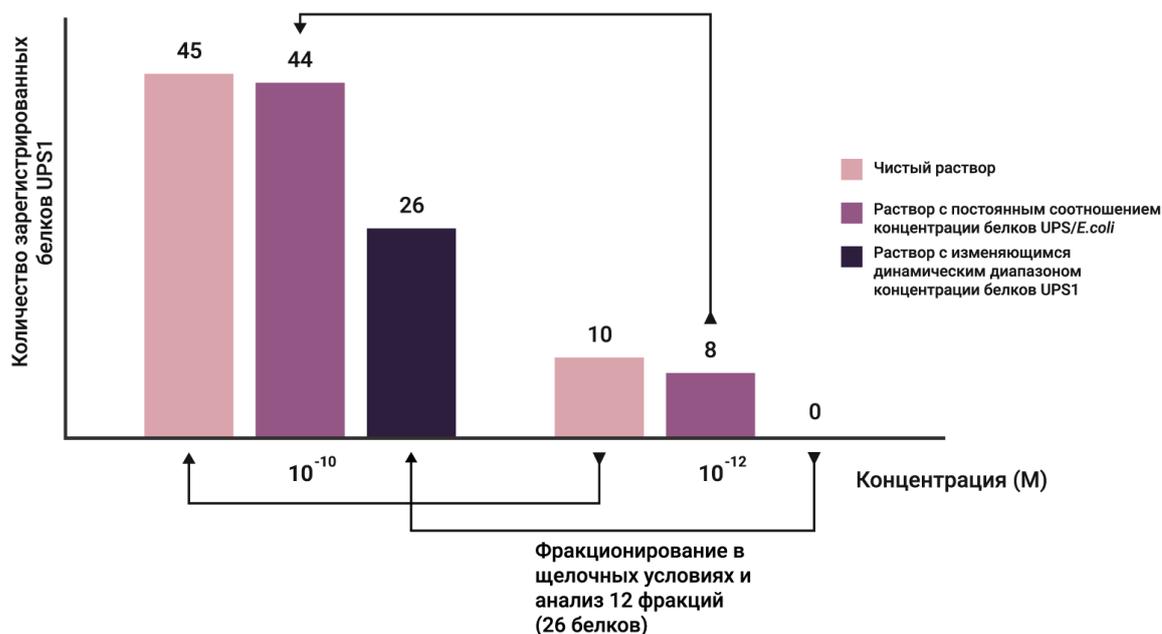
Однако при уменьшении концентрации белков UPS1 до  $10^{-10}$  М количество идентификаций снижается: наиболее резкое снижение наблюдается в модельном образце 3, который имитирует реальную биологическую систему. Число зарегистрированных белков в этой реконструированной системе составило 26, что составляет ~50% белков набора UPS1. При дальнейшем снижении концентрации целевых белков наблюдается уменьшение количества зарегистрированных белков, а при концентрации  $10^{-13}$  М не удастся зарегистрировать ни одного белка смеси UPS1. В модельном образце 3, имитирующем биологический образец невозможно зарегистрировать сигнал белков набора UPS1 уже при концентрации  $10^{-12}$  М. Также важным фактором детектируемости белка при таргетном анализе является выбор соответствующего пептида в качестве стандарта с минимальным пределом детекции. Так на рисунке 2 показано, что все 3 пептида белка O76070ups регистрируются в концентрации  $10^{-11}$  М, однако только один из пептидов детектируется при концентрации белка на порядок ниже  $10^{-12}$  М.



**Рисунок 2.** Зависимость величины сигнала 3 пептидов белка O76070ups от концентрации белка.

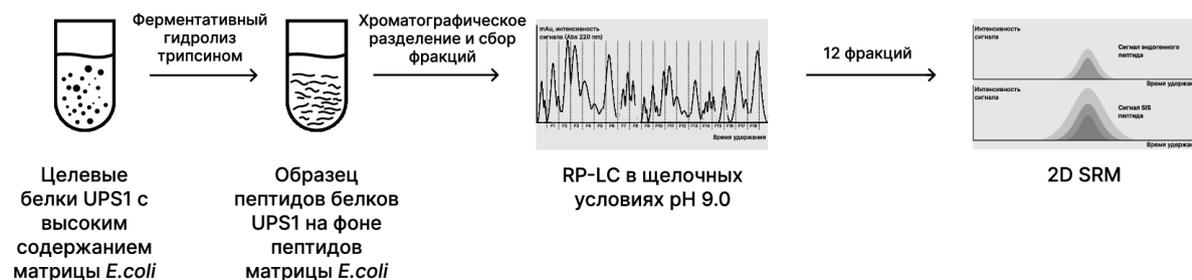
Таким образом, на детектируемость белков смеси UPS1 кроме чувствительности прибора оказывает влияние высокий динамический диапазон концентраций белков *E. coli*, а также использование пептидов с низким пределом детекции в качестве внутреннего стандарта для таргетного анализа.

Для проверки предположения о значении динамического диапазона, как лимитирующего фактора детекции белков было проведено обратное концентрирование образцов. Чистый раствор и раствор с постоянным соотношением концентраций белков UPS1/*E. coli* были сконцентрированы в 100 раз. В результате удалось восстановить количество идентификаций до исходного уровня 45 и 44 идентификации соответственно (рисунок 3). Для образца с изменяющимся динамическим диапазоном концентрации белков UPS1 (модельный образец 3) концентрирование не приводит к соответствующим результатам, так как не решает проблему высокого динамического диапазона, вместе с белками UPS1 одновременно происходит концентрация матрицы. Для преодоления этого ограничения раствор UPS1 с высоким содержанием матрицы *E. coli* фракционировали методом высокоэффективной жидкостной хроматографии на обращенной фазе в щелочных условиях элюции. В результате таргетного анализа каждой фракции зарегистрировано 26 белков UPS1. Схема эксперимента с фракционированием приведена на рисунке 4.



**Рисунок 3.** Схема эксперимента с концентрированием белков смеси UPS1.

Фракционирование в щелочных условиях разделения позволяет преодолеть ограничения чувствительности протеомных методов и снизить динамический диапазон концентраций белков. Таким образом, оно может быть использовано для анализа сложных биологических образцов.



**Рисунок 4.** Схема эксперимента с фракционированием образца белков смеси UPS1 с высоким содержанием биологической матрицы *E.coli*.

### 3.2 Разработка и оценка эффективности методики 2D Shotgun на модельном объекте клеточной линии HepG2

Для преодоления вышеописанных ограничений была разработана методика, которая включает в себя дополнительную стадию фракционирования пептидов – обращено фазовую хроматографию в щелочных условиях.

В качестве сложного биологического объекта выбрана клеточная линия HepG2, которая наряду с клетками печени и сывороткой крови человека является

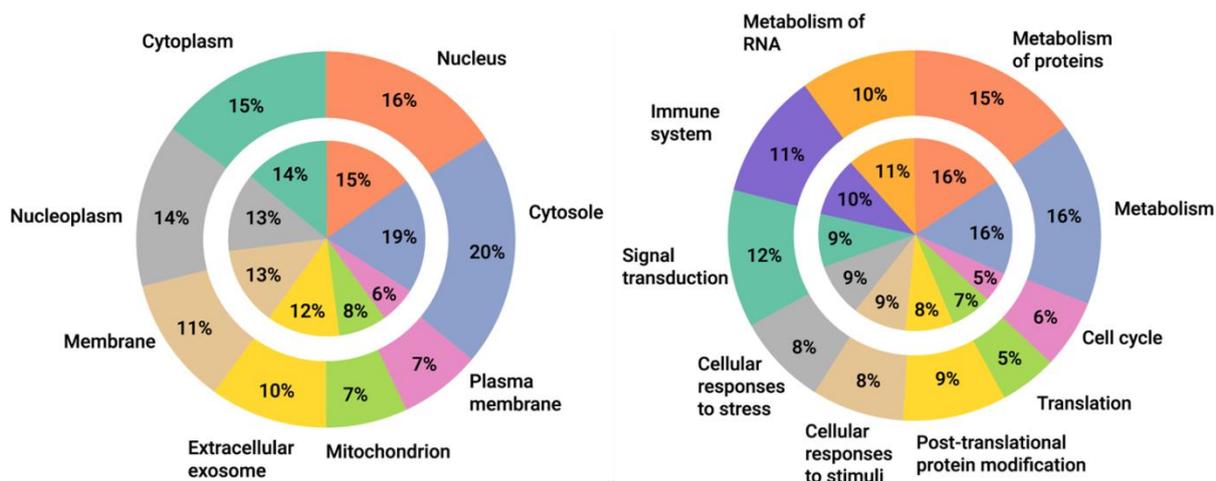
официальным объектом исследований в программе Российской части хромосомотричного проекта «Протеом человека». При одномерном панорамном протеомном исследовании (1D Shotgun) образца HepG2 было идентифицировано **11889** уникальных пептидов, которые картированы на **1453** белковых последовательностях.

Те же препараты гидролиза белков HepG2 использовали для фракционирования на обращенной фазе в щелочных условиях разделения. Всего было получено 23 фракции. При панорамном сканировании фракций (2D Shotgun) идентифицировано **32509** уникальных пептидов, которые картированы на **3312** последовательностях белков. При анализе данных учитывались только те белки, для которых было зарегистрировано более 2-х уникальных пептидов на последовательность белка.

При фракционировании образца количество идентифицированных белков в ходе панорамного сканирования увеличивается в 2,3 раза, что справедливо как для общего числа белковых идентификаций (**1453** в комплексном образце – **3312** во фракционированной пробе), так и для количества идентификаций белков для каждой хромосомы. Обнаруженные белки в 1D Shotgun и 2D Shotgun анализе картированы на метаболические пути базы данных Reactome (Gillespie et al., 2022).

Средний коэффициент обогащения для метаболических путей составляет  $1,94 \pm 0,76$ . Полученные данные позволяют сделать вывод о том, что методика двумерного фракционирования позволяет глубже изучить биологические процессы клетки. Наиболее представленными по белковому составу пути являлись: общий метаболизм, метаболизм белков, проведение сигнала, метаболизм РНК, то есть основные процессы жизнеобеспечения клетки. Также проведен анализ обогащения белков по клеточной локализации. В группах белков с цитоплазматической, ядерной, мембранной или иной локализацией не наблюдается значимого обогащения в 2D относительно 1D Shotgun анализа (рисунок 5).

Таким образом, разработанная методика 2D фракционирования на модельном объекте HepG2 показала увеличение чувствительности по сравнению с 1D в 2,3 раза по количеству зарегистрированных белков. При этом с помощью 1D Shotgun анализа обнаружено 18 белков, кодируемых генами 18 хромосомы человека, а при применении метода 2D-33. Двумерный анализ позволяет дополнительно обнаружить 15 белков, кодируемых генами 18 хромосомы человека. Полученные данные позволяют сделать вывод о том, что применение двумерного фракционирования совместно с SRM анализом позволит увеличить покрытие белков, кодируемых генами 18 хромосомы человека.

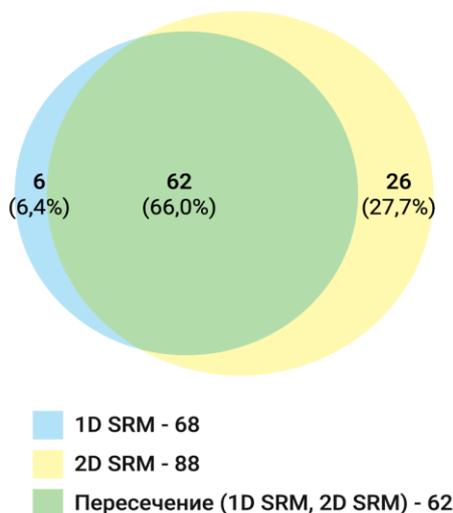


**Рисунок 5.** Клеточная локализация (левый рисунок) и функции белков (правый рисунок), зарегистрированных в 1D Shotgun анализе (внутренний круг) и 2D Shotgun (внешний круг) анализе клеточной линии HepG2. В абсолютном количестве в 2D Shotgun анализе было зарегистрировано в 2,3 раза больше белков, однако процентное соотношение зарегистрированных белков по группам остается равным.

### 3.3 Таргетный протеомный анализ (SRM SIS) белков клеточной линии HepG2, кодируемых генами 18 хромосомы человека

Применение методов таргетного анализа позволило суммарно обнаружить 94 белка, кодируемых генами 18 хромосомы человека (рисунок 6). С помощью метода 1D SRM обнаружено 68 белков. Применение метода двумерного фракционирования (2D SRM) увеличило число идентификаций на 26.

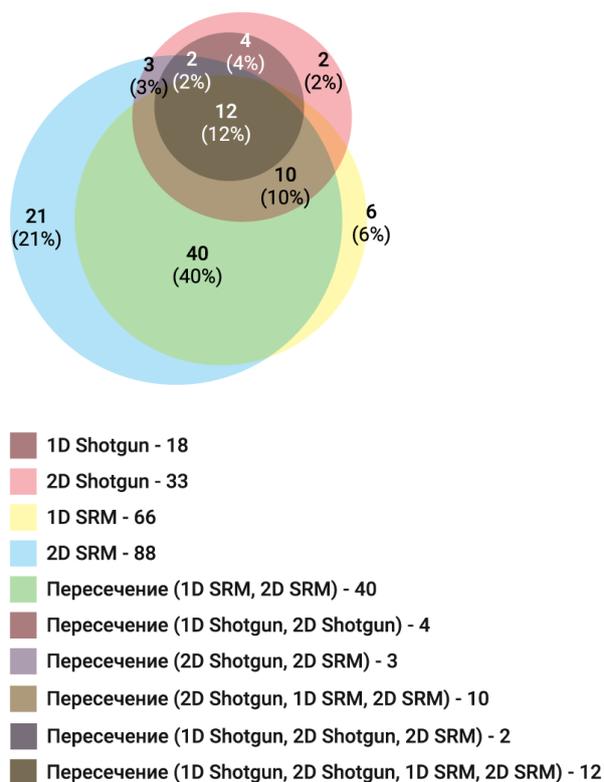
Чувствительность метода с применением двумерного фракционирования (2D SRM) выше в 1,33 раза по количеству идентифицированных белков по сравнению с методом одномерного анализа (1D SRM).



**Рисунок 6.** Диаграмма Эйлера-Венна, сравнение множеств белков, обнаруженных при одномерном сканировании не фракционированной пробы (1D SRM) и фракционированной пробы (2D SRM).

Чувствительность SRM анализа позволяет зарегистрировать белки в концентрации до 100 копий/клетку. Однако за счет обогащения фракций соответствующими пептидами в двумерном анализе зарегистрировано 26 белков, которые не регистрируются 1D SRM анализом. Концентрация этих белков варьирует в диапазоне от 50000 до 100 копий/клетку. Среди них присутствуют такие важные для клетки белки как P31152 (митоген активируемая протеинкиназа 4) и Q92908 (транскрипционный фактор GATA-6), которые играют роль в модуляции экспрессии генов. Таким образом, использование двумерного фракционирования позволяет получить важную биологическую информацию, которая недоступна при использовании только методики одномерного SRM анализа.

В совокупности, использование всех протеомных подходов анализа белков клеточной линии HepG2 позволило зарегистрировать 100 белков, кодируемых генами 18 хромосомы человека (рисунок 7), что составляет 38% белок-кодирующих генов 18 хромосомы человека. Уникальные идентификации, детектированные только SRM анализом, составили 67 белков. Также 27 белков составляют общий пул, идентифицированный и панорамным и таргетным методами. Несмотря на то, что SRM метод 2,8 раз более чувствительный (по количеству идентифицированных белков), целесообразно использование и панорамного метода, т.к. он не ограничен анализом уникальных протеотипических пептидов. В данном случае обнаружено 6 белков, кодируемых генами 18 хромосомы человека исключительно методикой Shotgun анализа.

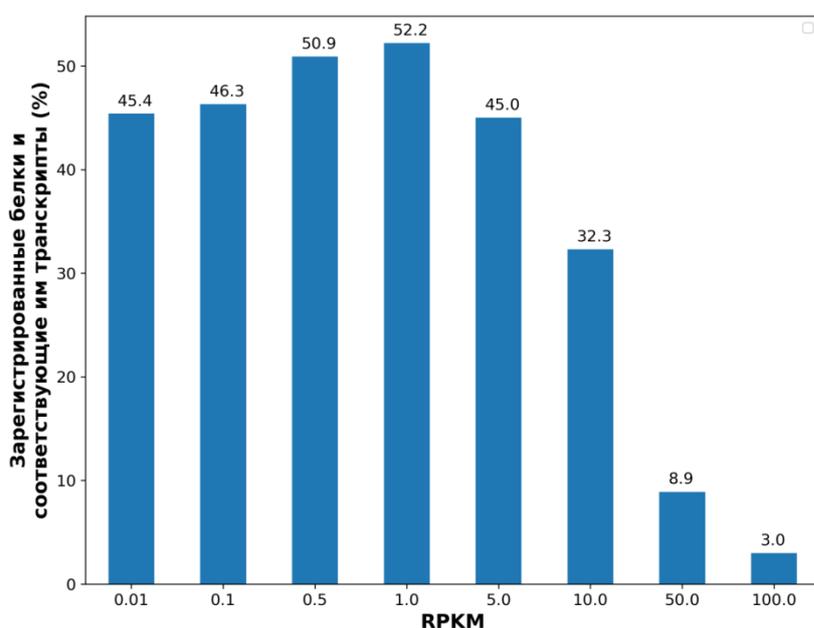


**Рисунок 7.** Диаграмма Эйлера-Венна, сравнение множеств белков, кодируемых генами 18 хромосомы человека, обнаруженных при одномерном сканировании комплексного образца (1D Shotgun, 1D SRM) и фракционированной пробы (2D Shotgun, 2D SRM).

### 3.4 Транскриптомный анализ клеточной линии HepG2

Транскриптомные данные, которые используются в работе для анализа, содержат результаты высокопроизводительного РНК – секвенирования, выполненного на платформе Illumina.

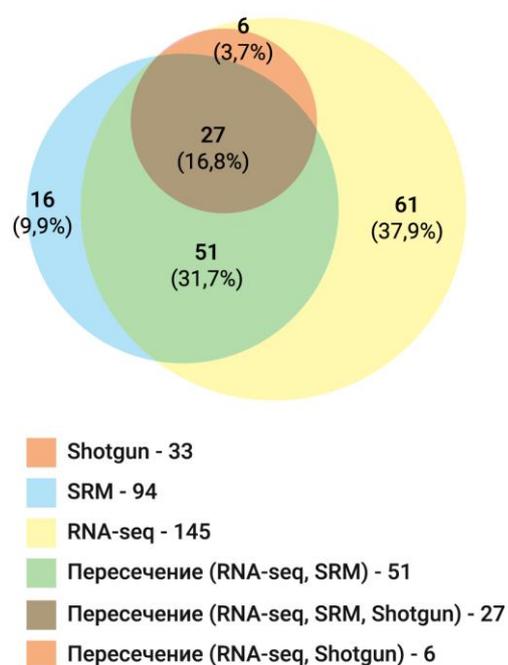
В протеомике существуют общепринятые стандарты по валидации идентификаций белков, а для транскриптомных данных такой стандарт отсутствует. Поэтому была проведена оценка сходимости протеомных и транскриптомных данных при разных отсечениях по RPKM (рисунок 8). Полученная гистограмма свидетельствует о том, что наибольшее пересечение в процентном отношении белков и соответствующих им транскриптов обнаруживается при отсечении по RPKM  $\geq 1$  (52,2%), при котором регистрируются 145 транскриптов, кодируемых генами 18 хромосомы человека.



**Рисунок 8.** Зависимость процента совместно зарегистрированных транскриптов и соответствующих им белков, кодируемых генами 18 хромосомы человека, от отсечения по RPKM.

Необходимо отметить, что при анализе белков клеточной линии HepG2 были использованы транскриптомные данные в качестве гипотезы о наличии белка в клетке. Однако, как показано на рисунке 9 наблюдаются отклонения в обе стороны - зарегистрированы белки, для которых нет соответствующей мРНК и, с другой стороны, отсутствуют белки при наличии мРНК. Факт детекции белков, для которых отсутствует соответствующая мРНК показан ранее, и объясняется различием периода полураспада мРНК и белка (Reimegård et al., 2019). Так, период полураспада молекулы

РНК в клетках млекопитающих варьирует от нескольких минут до нескольких часов (Shyu et al., 1989). При этом, самый малый зарегистрированный период полураспада белка составляет несколько часов, а наиболее стабильные не деградируют более 1000 часов (Mathieson et al., 2018). Например, для белков, P38405, Q13636, Q14574 период полураспада мРНК составляет от нескольких минут до десятков секунд (0,2, 0,016, 0,014 часа, соответственно). При этом период полураспада для соответствующих белковых молекул составляет 31, 51, 100 часов соответственно (Bevilacqua et al., 2003). Поэтому, при достаточно высокой эффективности трансляции возможна ситуация, когда белок присутствует в клетке в концентрации, достаточной для детекции, а мРНК при этом отсутствует.



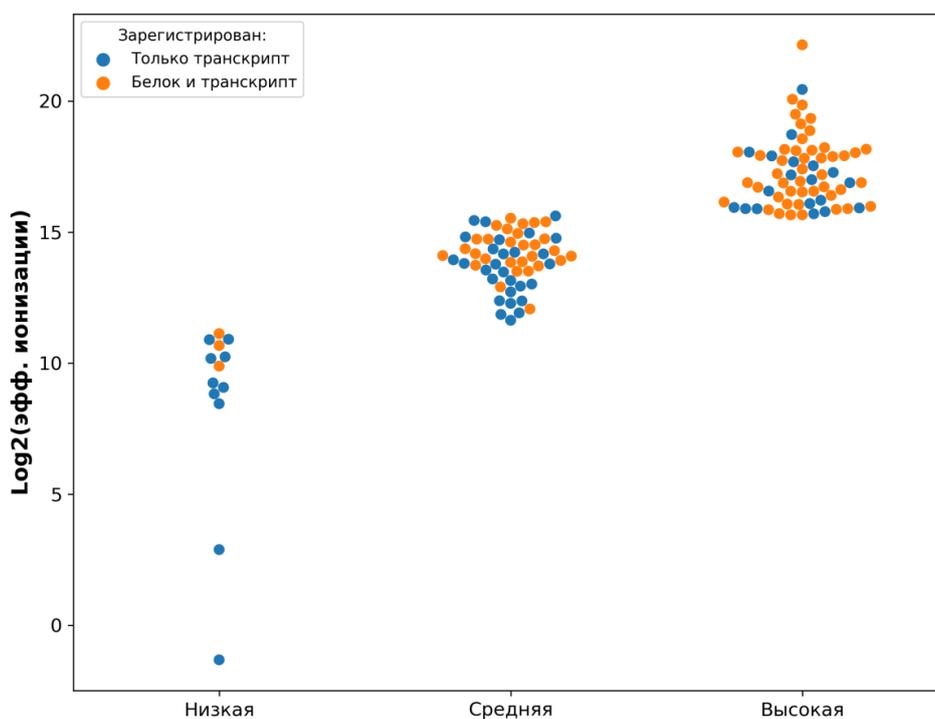
**Рисунок 9.** Диаграмма Эйлера-Венна, представляет собой три множества, зарегистрированные в образце HepG2 транскрипты – RNA-seq, белки, кодируемые генами 18 хромосомы человека, зарегистрированные методом Shotgun, белки, зарегистрированные методом SRM.

Таким образом, использование всех методов протеомного анализа позволяет зарегистрировать 100 белков, что составляет 69% зарегистрированного транскриптома и 38% всех белков кодирующих генов, расположенных на 18 хромосоме человека.

### 3.5 Анализ эффективности ионизации внутренних стандартов, используемых для регистрации белков, кодируемых генами 18 хромосомы человека

Одной из причин регистрации транскрипта на фоне отсутствия детекции белка может быть недостаточная чувствительность МС анализа, обусловленная низкой эффективностью ионизации протеотипического пептида, выбранного для мониторинга соответствующего белка в таргетном анализе. Под эффективностью ионизации понимается величина регистрируемого сигнала при масс-спектрометрическом анализе

исследуемых пептидов в одинаковой концентрации. На рисунке 10 представлены результаты измерения эффективности ионизации пептидов, проведенные на фоне биологической матрицы клеток линии HerG2. С помощью алгоритма кластеризации k-средних стандартные пептиды были разделены на 3 условные группы: с высокой эффективностью ионизации, средней и низкой. На рисунке 10 цветом выделены точки, каждая из которых обозначает один белок и соответствующий транскрипт, зарегистрированный в клетках линии HerG2. В группе с высокой эффективностью ионизации пептидов было зарегистрировано больше всего транскриптов и соответствующих им белков 70,8%, а в группе со средней и низкой эффективностью 50,9% и 23,1% (рисунок 10). Таким образом, чем ниже эффективность ионизации протеотипического пептида, тем меньше вероятность зарегистрировать белок, соответствующий транскрипт которого детектируется в образце.



**Рисунок 10.** Эффективность ионизации стандартных пептидов для таргетного анализа на фоне биологической матрицы пептидов клеточной линии HerG2, цветом выделены белки, для которых был зарегистрирован и белок и соответствующий ему транскрипт или только транскрипт.

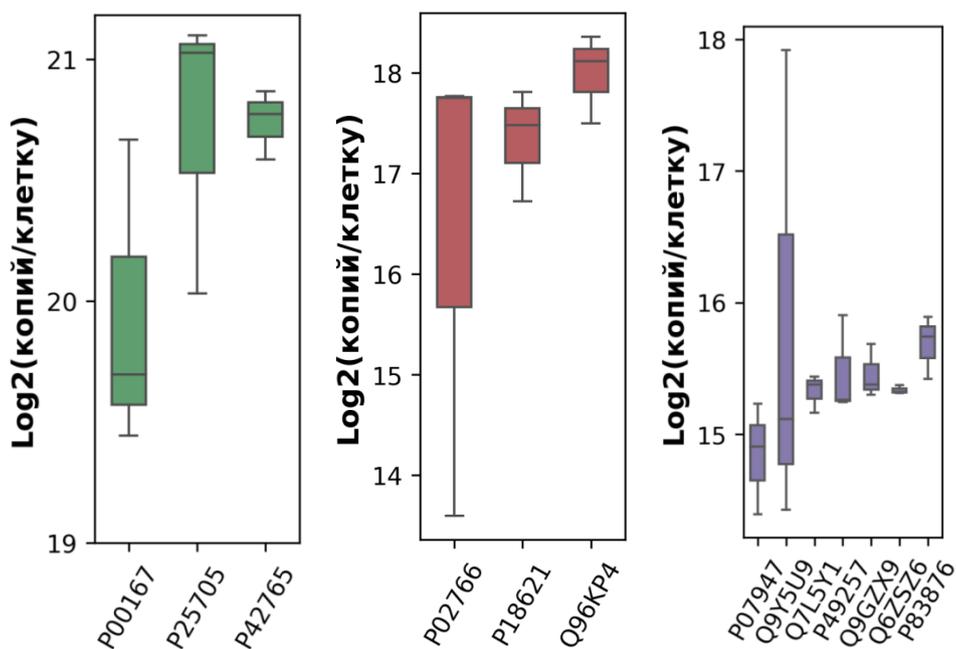
### 3.6 Применение разработанного метода 2D фракционирования для анализа клинических образцов печени человека

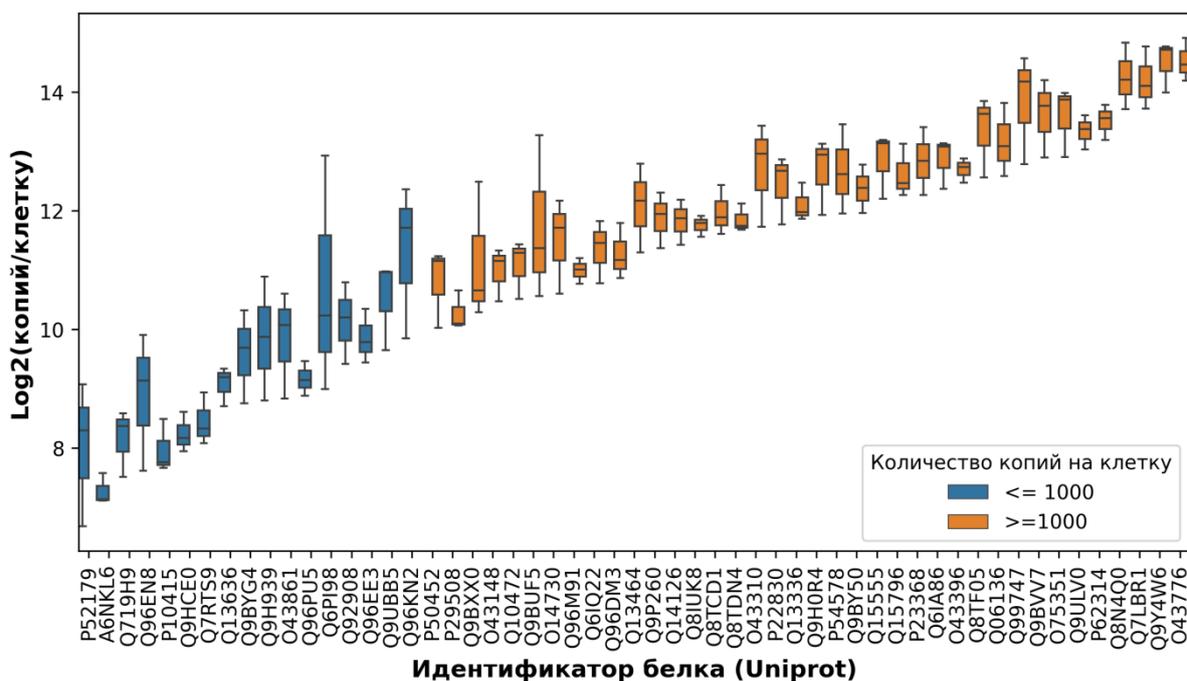
Для проведения измерений использовали 3 образца печени человека. Этапы и условия выполнения работы полностью повторяли вышеописанные для образца HepG2. Полученные результаты показали, что в образцах печени человека 1D Shotgun анализом можно обнаружить в среднем  $23 \pm 3$  белка, кодируемых генами 18 хромосомы человека, в каждом образце, а для 2D Shotgun анализа –  $35 \pm 5$  белков.

При анализе образцов печени с использованием таргетных методов было зарегистрировано 129 белков, кодируемых генами 18 хромосомы человека, при этом 63 и 110 белков с использованием 1D SRM и 2D SRM методик, соответственно.

В ходе исследования были измерены концентрации белков, которые с помощью алгоритма кластеризации k-средних разделили на 4 кластера: белки в высокой концентрации, средней, низкой и ультранизкой концентрации, при этом средние значения концентрации в каждом кластере составляют 1551194 копий/клетку, 198152 копий/клетку, 51368 копий/клетку, 5591 копий/клетку, соответственно (рисунок 11). Среди белков, зарегистрированных в каждом образце печени человека 1D SRM анализом, наименьшее обнаруживаемое число копий/клетку составляет 8000, а в 2D SRM анализе обнаружено 17 белков, копияность которых ниже 1000 на клетку, с минимальным значением в 100 копий/клетку. Группа белков в высокой концентрации состоит из трех белков P25705 (альфа субъединица митохондриальной АТФ синтазы), P42765 (3-кетоацил-КоА тиолаза) и P00167 (цитохром b5), каждый из которых задействован в процессах получения энергии. Группа белков в средней концентрации состоит из трех белков P02766 (транстиретин) белок сыворотки крови, который синтезируется в печени, P18621 (60S рибосомальный белок L17) структурный компонент рибосомы, Q96KP4 (цитозольная неспецифическая дипептидаза) клеточная дипептидаза, которая участвует в метаболизме глутатиона. Группа белков в низкой концентрации состоит из семи белков Q9GZX9 (гомолог белка скрученной гастрюляции 1) точная функция неизвестна, может играть роль в развитии хондроцитов, P49257 (белок ERGIC-53) рецепторный белок, играет роль в транспорте гликопротеинов и гликолипидов от ЭПР в аппарат Гольджи, P07947 (тирозиновая протеинкиназа Yes) нерцепторная тирозинкиназа, которая играет роль в регуляции клеточного цикла и выживания, апоптоза и дифференцировки, Q7L5Y1 (член суперсемейства митохондриальных енолаз 1) метаболизирует L-фукозу, Q6ZSZ6

(антиген NY-CO-33) возможный регулятор экспрессии генов, играет роль в процессах развития, Q9Y5U9 (3-взаимодействующий белок немедленного раннего ответа 1) регулятор секреции ЭПР, P83876 (тиоредоксин подобный белок 4A) участвует в сплайсинге. Группа белков в ультранизкой концентрации состоит из 56 белков P10415 (регулятор апоптоза Bcl-2), Q9HCE0 (гомолог белка эктопических гранул P 5) регулятор аутофагии, Q13636 (родственный-Ras белок Rab-31) регулятор фагоцитоза, Q96PU5 (NEDD4-подобная убиквитин протеин лигаза E3) участвует в убиквитинировании, O14730 (серин/треониновая протеинкиназа RIO3), участвует во врожденных процессах иммунного ответа на дцРНК.

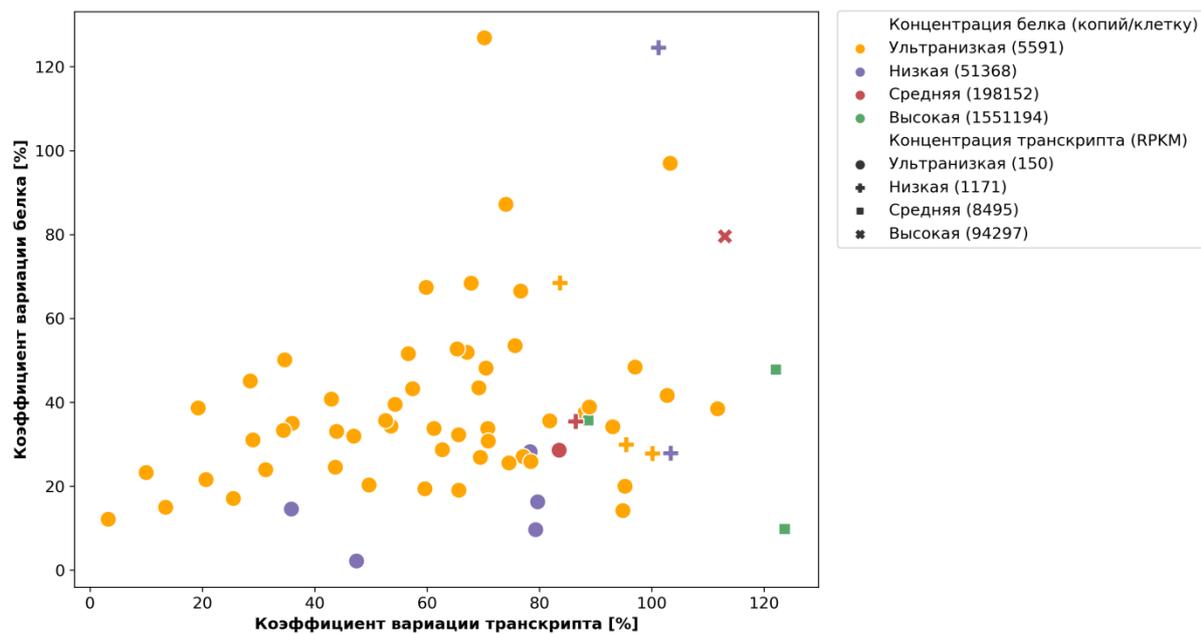




*Рисунок 11. Содержание в клетке зарегистрированных белков в количестве копий на клетку, разделенные на кластеры по содержанию в клетке.*

### 3.7 Анализ показателей биологической вариации содержания транскриптов и белков, кодируемых генами 18 хромосомы человека, в образцах печени

Анализ биологической вариации показал, что существует группа с низкой коэффициентом вариабельности не только белков, но и транскриптов. В эту группу вошли белки с преимущественно низкой концентрацией от 250 до 45000 копий/клетку. Несмотря на низкую концентрацию, коэффициент вариабельности концентрации белков этой группы составляет в среднем 29%, а транскриптов-37%. Таким образом, эта группа из 25 белков составляет стабильно экспрессируемую часть протеома 18 хромосомы человека в печени (рисунок 12). В эту группу вошли такие важные для выживания и жизнедеятельности клетки белки как: Q13336 (транспортер мочевины 1), O43310 (СВР80/20-зависимый фактор инициации трансляции), Q6IQ22 (Ras-родственный белок Rab-12, регулирует внутриклеточный мембранный транспорт везикул), P10415 (регулятор апоптоза bcl-2), Q9UBB5 (Метил-СрG-связывающий домен белок 2, репрессор транскрипции).



*Рисунок 12. Анализ коэффициентов вариации белков и соответствующих им транскриптов, зарегистрированных в каждом образце печени человека.*

#### 4. ЗАКЛЮЧЕНИЕ

Основной проблемой таргетного анализа комплексных биологических образцов является наличие высокого динамического диапазона концентраций белков системы. При анализе белков, кодируемых генами 18 хромосомы человека, в клетках печени целевыми белками являются 264 белка, в то время как остальные белки образца, количество которых может достигать до 6 миллионов являются матрицей. В таком случае целесообразно снизить эффект матрицы с помощью фракционирования. Таким образом, разделение комплексного образца на 20 фракций дает значимое снижение динамического диапазона белков. Так одна фракция содержит не 6 миллионов, а 300 тысяч белков, что оказывает значимый эффект и увеличивает количество регистрируемых белков по сравнению со стандартным одномерным МС анализом. В работе показано, что белки в ультранизкой концентрации составляют большую часть протеома 18 хромосомы в клетках печени человека (56 из 69), среди них встречаются биологически значимые белки, например транскрипционные факторы и белки, ответственные за передачу сигнала в клетке. При использовании стандартных методов анализа идентификация этих белков затруднена, а двумерный таргетный анализ позволяет провести детекцию и точную количественную оценку этих белков. Однако для детекции низкокопийных белков недостаточно снизить эффект матрицы в образце. На примере пептидов стандартной системы UPS1 и в последующем на стандартных протеотипических пептидах, используемых для детекции белков, кодируемых генами 18 хромосомы человека, показано, что чувствительность таргетного анализа зависит от эффективности ионизации пептида выбранного в качестве стандарта. Проблема выбора пептида в качестве стандарта заключается в том, что сложно предсказать какой именно пептид белка будет иметь лучшие ионизационные свойства. Параметры эффективности ионизации определяются эмпирически при детекции с помощью панорамной масс-спектрометрии. Однако чувствительность этих методов ограничена, поэтому для каждого белка не удастся найти оптимальный пептид для таргетного анализа. Таким образом, актуальной задачей является создание предсказательной модели, которая могла бы выявлять пептиды с высокой эффективностью ионизации на основе данных панорамной масс-спектрометрии. Особенно полезным это предсказание может оказаться при детекции белков протеома 18 хромосомы, которые не были обнаружены с помощью методики 2D фракционирования, а также для белков, которые еще не зарегистрированы ни в одном эксперименте, так называемых “missing” белков.

## 5. ВЫВОДЫ

1. Определены ограничения таргетных протеомных методов. Для белков стандартной системы UPS1 показана концентрационная зависимость их детектируемости. Предел детекции белков смеси составляет  $10^{-13}$  М в чистом растворе и  $10^{-12}$  М в присутствии белков биологической матрицы (высокого динамического диапазона концентраций белков). Также показано влияние выбора протеотипического пептида на чувствительность таргетного анализа.

2. Применение методики 2D фракционирования на модельной системе клеточной линии HepG2 позволило увеличить количество идентификаций белков на 100% и на 33% для панорамного и таргетного анализа, соответственно. В результате анализа было зарегистрировано 100 белков, кодируемых генами 18 хромосомы человека, что составляет 37,9% белков протеома 18 хромосомы человека и 52,2% зарегистрированного транскриптома.

3. При анализе образцов печени человека показано, что методика двумерного SRM анализа обладает высокими показателями стабильности измерений, что позволило зарегистрировать 69 белков в каждом образце печени человека и оценить точные концентрации белков в клетках, в том числе 17 белков в концентрации менее 1000 копий на клетку. В образцах печени всего зарегистрировано 129 белков, кодируемых генами 18 хромосомы человека.

4. В результате панорамного масс спектрометрического анализа клеточной линии HepG2 и клеток печени человека детектировано **6599** уникальных белков, что составляет **32%** протеома человека. Таргетным анализом было зарегистрировано **142** уникальных белка (**54%**), кодируемых генами 18 хромосомы человека.

## 6. СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых научных журналах:

1) **Vavilov Nikita**, Ilgisonis Ekaterina, Lisitsa Andrey, Ponomarenko Elena, Farafonova Tatiana, Tikhonova Olga, Zgoda Victor and Archakov Alexander, Number of detected proteins as the function of the sensitivity of proteomic technology in human liver cells. // Current Protein & Peptide Science. 2022. Vol. 23, №4. P. 290-298.

2) **Vavilov, N. E.**, Ilgisonis, E. V., Lisitsa, A. V., Ponomarenko, E. A., Farafonova, T. E., Tikhonova, O. V., Zgoda, V. G., & Archakov, A. I. Deep proteomic dataset of human liver samples obtained by two-dimensional sample fractionation coupled with tandem mass spectrometry. // Data in Brief. 2022. Vol. 42, P1-7.

3) **Vavilov, N. E.**, Zgoda, V. G., Tikhonova, O. V., Farafonova, T. E., Shushkova, N. A., Novikova, S. E., Yarygin, K. N., Radko, S. P., Ilgisonis, E. V., Ponomarenko, E. A., Lisitsa, A. V., & Archakov, A. I. Proteomic Analysis of Chr 18 Proteins Using 2D Fractionation. // Journal of Proteome Research. 2020. Vol. 19, №12. P. 4901–4906.

4) Ilgisonis, E., **Vavilov, N.**, Ponomarenko, E., Lisitsa, A., Poverennaya, E., Zgoda, V., Radko, S., & Archakov, A. Genome of the Single Human Chromosome 18 as a “Gold Standard” for Its Transcriptome. // Frontiers in Genetics. 2021. Vol. 12. P. 1–7.

**Работы, опубликованные в сборниках материалов научных конференций:**

5) **Vavilov Nikita**, Dr Ekaterina Ilgisonis, Dr Andrey Lisitsa, Dr Elena Ponomarenko, Dr Tatiana Farafonova, Dr Olga Tikhonova, Dr Victor Zgoda, Dr Alexander Archakov Number of Detected Proteins as the Function of the Sensitivity of Proteomic Technology in Human Liver Cells // HUPO ReCONNECT 2021 Congress P. 294.

6) **Vavilov N.**, A. Kopylov, T. Farafonova, N. Solovieva, A. Archakov, V. Zgoda Increased sensitivity of MS - based proteomics methods obtained by sample fractionation technique // FEBS Open Bio 11 (Suppl. 1) 2021 P. 180.

## **БЛАГОДАРНОСТИ**

Автор благодарит руководителя и коллег за помощь в проведении работы и подготовке диссертации. В частности, автор благодарен к.б.н. Фарафоновой Т.Е. за синтез внутренних стандартов для SRM SIS, д.б.н. проф. Ярыгина К.Н. и к.м.н. Вахрушева И.В. за работу с клеточной линией HepG2, к.б.н. Курбатова Л.К. за предоставление материала *E. coli* и коллег из ИБХ, которые предоставили 20 клеточных линий для анализа.