

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ
«НАУЧНО – ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ БИОМЕДИЦИНСКОЙ ХИМИИ
ИМЕНИ В.Н. ОРЕХОВИЧА»

На правах рукописи

Вавилов Никита Эдуардович

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

1.5.4. – Биохимия

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель:

д.б.н., профессор РАН

Згода Виктор Гаврилович

Москва 2023

Оглавление

СПИСОК ИСПОЛЬЗУЕМЫХ СОКРАЩЕНИЙ	4
1. ВВЕДЕНИЕ, ЦЕЛЬ И ЗАДАЧИ	7
1.1 Актуальность проблемы, цель и задачи	7
2. ОБЗОР ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ	12
2.1 История масс – спектрометрии	12
2.2 От генома к протеому человека	13
2.3 Проблемы регистрации белков при МС анализе, ограничение чувствительности прибора.....	22
2.3.1 Масс анализатор типа Orbitrap	22
2.3.2 Shotgun анализ	23
2.3.3 Метод BoxCar.....	25
2.3.4 Метод DIA	25
2.3.5 Другие способы преодоления ограничений метода Shotgun.....	26
2.3.6 Чувствительность приборов с Orbitrap масс анализатором	27
2.3.7 Квадрупольный масс анализатор	27
2.3.8 SRM SIS анализ.....	28
2.3.9 Чувствительность приборов типа тройной квадруполь.....	29
2.4 Методы фракционирования, используемые в современной протеомике	30
2.4.1 Фракционирование клеточных органелл	30
2.4.2 Фракционирование белков клетки.....	33
2.4.3 Хроматографические методы фракционирования	37
3. МАТЕРИАЛЫ И МЕТОДЫ.....	41
3.1 Описание образцов печени человека	41
3.2 Выделение белка и ферментативный гидролиз трипсином	41
3.3 Разделение пептидов обращенно-фазовой хроматографией в щелочных условиях	42
3.4 Идентификация белкового состава образца с помощью тандемной масс спектрометрии (Shotgun) с использованием детектора высокой точности (Orbitrap).43	
3.5 Выбор пептидов для регистрации белков, кодируемых генами 18 хромосомы человека.....	46
3.6 Твердофазный синтез пептидов.....	46
3.7 Таргетный протеомный анализ с использованием изотопно меченых пептидов в качестве внутренних стандартов (SRM)	46

4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ	49
4.1 Ограничения методов протеомных технологий (SRM).....	49
4.2 Ограничения методов протеомных технологий (Shotgun).....	52
4.3 Разработка и оценка эффективности методики 2D Shotgun на модельном объекте HepG2	54
4.4 GO анализ белков, идентифицированных при Shotgun 1D и Shotgun 2D анализе по клеточной локализации белков	65
4.5 Анализ распределения обнаруженных белков по биологическим путям по версии базы данных Reactome	66
4.6 Анализ 20 клеточных линий методом панорамной масс спектрометрии для подбора пептидов с оптимальными параметрами эффективности ионизации для таргетного анализа	70
4.7 Разработка и оценка эффективности методики 2D SRM на модельном объекте HepG2	71
4.8 Таргетный протеомный анализ (SRM SIS) белков, кодируемых генами 18 хромосомы человека клеточной линии HepG2	74
4.9 Транскриптомный анализ клеточной линии HepG2	77
4.10 Анализ эффективности ионизации внутренних стандартов, используемых для регистрации белков, кодируемых генами 18 хромосомы человека	80
4.11 Оценка стабильности и воспроизводимости метода двумерного фракционирования на примере клинических образцов нормальной печени человека	81
4.12 Таргетный протеомный анализ (SRM SIS) белков, кодируемых генами 18 хромосомы человека образцов нормальной печени человека	87
4.13 Анализ показателей биологической вариации транскриптомных и протеомных данных в образцах нормальной печени человека	93
5. ЗАКЛЮЧЕНИЕ	95
6. ВЫВОДЫ.....	97
7. СПИСОК ЛИТЕРАТУРЫ	98
8. БЛАГОДАРНОСТИ.....	112
9. ФИНАНСИРОВАНИЕ	112

СПИСОК ИСПОЛЬЗУЕМЫХ СОКРАЩЕНИЙ

2DE – двумерный электрофорез в полиакриламидном геле (от англ. two-dimensional gel electrophoresis)

C-HPP – Хромосомотцентричный Проект Протеома Человека (от англ. Chromosome Centric Human Proteome Project)

cSNP – кодирующие однонуклеотидные полиморфизмы (от англ. coding Single Nucleotide Polymorphism)

DDA – анализ зависимый от данных, синоним Shotgun (от англ. Data Dependent Acquisition)

DIA – анализ независимый от данных (от англ. Data Independent Analysis)

ESI – ионизация методом электрораспылением (от англ. Electrospray Ionization)

FDR – отсечение для модели анализа данных, показывает процент ложноположительных идентификаций белков (от англ. false discovery rate)

GO – словарь терминов, который описывает гены и их продукты с точки зрения функции, локализации, биологических процессов и т.д. (от англ. Gene Ontology)

HUPO – проект протеом человека (от англ. Human Proteome Project)

iBAQ – основанная на интенсивности абсолютная количественная оценка (от англ. intensity Based Absolute Quantification)

ICR – ион циклотронный резонанс (от англ. Ion Cyclotron Resonance)

IEC – ионообменная хроматография (от англ. ion exchange chromatography)

IEF – изоэлектрофокусирование (от англ. isoelectric focusing)

miRNA – микро РНК (от англ. micro RNA)

MS – масс спектрометрия (от англ. Mass Spectrometry)

m/z – масс зарядовое состояние (от англ. mass to charge ratio)

ppm – одна миллионная доля (от англ. pars per million)

RP – обращенно фазовая хроматография (от англ. reversed phase)

siRNA – малая интерферирующая РНК (от англ. small interfering RNA)

SCX – сильный катионообменник (от англ. Strong Cation Exchange)

SDS-PAGE – одномерный электрофорез в полиакриламидном геле в присутствии SDS (от англ. sodium dodecyl sulfate one-dimensional polyacrylamide gel electrophoresis)

SNP – однонуклеотидные полиморфизмы (от англ. Single Nucleotide Polymorphism)

SRM – мониторинг выбранных реакций (от англ. Selected Reaction Monitoring)

SRM SIS - мониторинг выбранных реакций с использованием изотопно-меченных стандартов (от англ. Selected Reaction Monitoring Stable Isotope Labelled Standard)

TOF – времяпролетный масс анализатор (от англ. Time Of Flight)

UPS1 – Универсальный белковый стандарт 1 (от англ. Universal Protein Standard 1)

ВИЧ – вирус иммунодефицита человека

ВЭЖХ – высокоэффективная жидкостная хроматография

ДМСО – диметилсульфоксид

ДНК – дезоксирибонуклеиновая кислота

ГХ/МС – методика газовой хроматографии с последующей детекции с помощью масс спектрометрии

ЖХ – жидкостная хроматография

ЖХ/УФ - жидкостная хроматография, с детекцией ультрафиолетом

ЖХ/МС – жидкостная хроматография, с детекцией методом масс спектрометрии

МС – масс спектрометр

МС1 – масс спектрометрическое сканирование родительских ионов

МС2 – масс спектрометрический анализ дочерних ионов, полученных в результате фрагментации родительского пептидного иона

ПТМ – посттрансляционные модификации

РНК – рибонуклеиновая кислота

1. ВВЕДЕНИЕ, ЦЕЛЬ И ЗАДАЧИ

1.1 Актуальность проблемы, цель и задачи

Проблема чувствительности в современной протеомике особенно остро стоит при исследовании комплексных биологических образцов: сыворотка крови человека, клеточные линии и ткани человека. Основным методом масс спектрометрического анализа является тандемная масс спектрометрия, которая предполагает регистрацию пептидных ионов после элюции с хроматографической колонки. Протеом человека содержит около 20359 белков и 42329 изоформ, что при гидролизе может дать более 3280422 пептидов по версии базы данных neXtProt [1]. Кроме того, белки в клетках и тканях организма находятся в разной концентрации от 10^{-6} М до 10^{-18} М и ниже, демонстрируя высокий динамический диапазон концентраций [2, 3]. Совместная элюция и ионизация пептидных ионов приводит к искажению и супрессии сигнала низкоинтенсивных пептидных ионов на фоне высокоинтенсивных ионов, которые получают при гидролизе белков с высокой концентрацией [4, 5]. Показано, что тандемный масс спектрометрический анализ биологического образца может зарегистрировать около 15000 пептидов. Однако анализ всех изотопных кластеров, включая те, которые не прошли отбор и дочерние ионы не регистрировались в анализе, показал, что в одном образце можно обнаружить 115000 пептидных ионов [6].

Таким образом, тандемный масс спектрометрический анализ регистрирует лишь 15% тех белков, которые действительно находится в образце. Принципиально иной подход демонстрируют методы таргетной протеомики [7]. Эти методы отличаются тем, что исследователь еще перед началом эксперимента выбирает, какие именно пептидные ионы будет регистрировать прибор. В таком случае удастся добиться более высоких показателей чувствительности, что позволяет зарегистрировать пептиды, анализ которых в тандемном масс спектрометрическом анализе затруднен. Например, при анализе белков, кодируемых генами 18 хромосомы человека, в печени методами тандемной масс спектрометрии удалось зарегистрировать 18 белков, а таргетными – 68 [8]. Однако даже 68 белков составляют всего лишь 20% всех белков, кодируемых генами 18 хромосомы человека. Таким образом, решение проблемы ограничений протеомных методов, позволит увеличить чувствительность анализа и тем самым увеличить глубину покрытия протеома 18 хромосомы человека.

Представленная работа направлена на изучение причин ограничения протеомных методов и поиска путей решения проблемы с использованием масс-спектрометрических методов анализа, как таргетных, так и панорамных. Причины ограничений методов изучены на упрощенной белковой системе, которая представляет собой универсальный протеомный стандарт 1 (UPS1). В ходе исследования разработаны методы масс-спектрометрической детекции белков с использованием дополнительного предварительного фракционирования образца и последующим анализом каждой фракции. Данные методы апробированы и охарактеризованы на биологической системе HepG2, клеточной линии гепатоцеллюлярной карциномы человека. В результате анализа панорамными методами зарегистрировано 33 белка, кодируемых генами 18 хромосомы человека, а таргетными методами обнаружено 94 белка. В дальнейшем разработанный метод был использован при анализе клинических образцов печени человека от здоровых доноров для оценки стабильности и воспроизводимости измерений.

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 (2D SRM) 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 разработанного метода обнаружены и измерены точные концентрации 129 белков, кодируемых генами 18 хромосомой 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 удастся зарегистрировать стандартными протеомными методиками. Разработан новый подход, с помощью которого можно обнаружить белки, не детектируемые при исследовании образцов методами таргетного и панорамного МС анализа. Показано, что он способен обнаруживать 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 представляет большую биологическую значимость. Белки, обнаруживаемые при помощи нового метода, позволяют получить более полную и осмысленную информацию об исследуемой биологической системе по сравнению со стандартной методикой 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. Показаны ограничения таргетного протеомного метода (SRM анализа). На стандартном объекте, упрощенной белковой системе UPS1 показана зависимость детектируемости белков от концентрации, наличия высокого динамического диапазона белков на фоне целевых белков системы UPS1. Также показана зависимость чувствительности SRM анализа от выбора протеотипического пептида в качестве внутреннего стандарта.
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. Это 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 метода и увеличения глубины покрытия протеома человека и белков, кодируемых генами 18 хромосомы человека. Свойства пептидов, определяющие эффективность их ионизации, являются ключевыми свойствами, определяющими 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.

2. ОБЗОР ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ

2.1 История масс – спектрометрии

История масс-спектрометрии берёт своё начало в 1919 году, когда студент Томсона Френсис Уильям Эстон сконструировал первый полностью функциональный прибор в лаборатории в Кембридже. С помощью первого масс-спектрометра удалось идентифицировать изотопы хлора, брома, и криптона. Это доказывало, что природные химические элементы состоят из комбинации нескольких стабильных изотопов. В 1921 году Ф.В. Эстон получает Нобелевскую премию по химии за открытие изотопного распределения. Также, Эстон сформулировал правило целых чисел, которое гласит, что массы изотопов одного элемента, кратны атомной массе водорода [9].

К 1940-ым годам масс-спектрометры стали коммерчески доступными приборами. Однако в то время их использовали только для контроля процессов производства, то есть структуры всех химических соединений были заранее известны, а с помощью масс-спектрометрии проводился лишь контроль их количественного содержания. В то время наиболее интересным вопросом являлось установление взаимосвязи полученного масс-спектра с химической структурой вещества, то есть масс-спектрометрическая идентификация компонентов образца. Наиболее известными химиками, разрабатывающими масс-спектрометрические методы идентификации органических молекул по спектру их фрагментации, были исследователи из США: Фред МакЛафферти, Клаус Биemanн, Карл Джерасси [10, 11, 12]. Именно они заложили основы для современных биологических масс-спектрометрических исследований. В послевоенные годы в течение 1950-ых годов Роланд Голк и Фред МакЛафферти соединили в единую систему масс-спектрометр и газовый хроматограф [13]. Это позволило увеличить чувствительность и разрешающую способность метода. Система ГХ/МС позволяет анализировать комплексные химические смеси и идентифицировать структуру соединений. Одна из первых успешных попыток идентификации пептидов в системе ГХ/МС предпринята Клаусом Биemanном. Он модифицировал пептиды до соответствующих полиаминоспиртов, что позволило увеличить летучесть соединений и провести хроматографию и ионизацию в газовой фазе. В таком виде производные пептидов подвергались ионизации электронами и фрагментировались достаточно предсказуемо, чтобы выявить их аминокислотную последовательность [14]. Имеющиеся на тот момент методы ионизации разрушали

молекулы биополимеров, а интерпретация получаемых масс спектров крайне затруднительна. Таким образом, необходимо было разработать более мягкие подходы, которые не разрушают молекулы пептидов, белков и нуклеиновых кислот.

В 1980-ых Джоном Беннетом Фенном был разработан метод электрораспыления и в 2002 году он получил Нобелевскую премию по химии. Данный подход позволяет переводить биополимеры из жидкой фазы в газовую, не разрушая их структуру [15]. Таким образом, стало возможным объединение жидкостного хроматографа и масс спектрометра в единую систему, что уменьшает время, необходимое для анализа биологического образца и упрощает пробоподготовку, а также увеличивает чувствительность и производительность метода. На современном этапе развития масс спектрометрии параллельно развивается приборная база как хроматографических систем, которые могут работать при давлении до 1200 бар и обеспечивать бесперебойный поток жидкости в нанолитровом диапазоне, так и конструкции масс спектрометров, что обеспечивает возможность детекции десятков тысяч пептидов биологического образца за один цикл работы прибора. В связи с этим масс спектрометрические методы детекции широко распространены благодаря их высокой надежности, точности, чувствительности и селективности. Они используются в различных областях: судебной медицине, для контроля над допингом в спорте, в пищевой промышленности, в исследовании космоса, а также в биомедицинских и фармацевтических исследованиях и многих других областях науки и индустрии [16, 17, 18].

2.2 От генома к протеому человека

В ходе проекта «Геном человека» (The Human Genome Project, HGP) впервые полностью расшифрована нуклеотидная последовательность ДНК человека [19]. Проект стартовал в 1990 году под руководством Джеймса Уотсона и считается крупнейшим международным сотрудничеством в биологии. На данный момент проект уже окончен, геном человека секвенирован, но остаются участки, нуждающиеся в дополнительной аннотации или труднодоступные участки гетерохроматина, которые не расшифрованы до настоящего времени. Такие участки составляют до 8% генома человека. При анализе данных, полученных в ходе проекта, предсказано, что количество генов, которые непосредственно кодируют белковый продукт, составляет около 20000. Однако на момент окончания проекта «Геном человека» далеко не все

предсказанные белки были зарегистрированы в эксперименте. Еще меньше известно о функции белков, как уже зарегистрированных, так и тех, которые еще предстоит обнаружить.

Закономерным продолжением проекта «Геном человека» является проект «Протеом человека». Предпосылками для запуска проекта послужили, с одной стороны, создание хорошо охарактеризованной карты генома человека и доступность углубленной транскриптомики, а с другой стороны, развитие масс-спектрометрических методов для исследования биологических систем. Эти достижения и технологии позволили организовать масштабный протеомный проект – Human Proteome Project (HPP). Это международный проект, организованный Human Proteome Organization (HUPO) целью которого является исследование и описание протеома человека. Проект должен расширить знания о белках, составляющих протеом человека, а сопоставление протеомных данных с геномными и транскриптомными данными позволит наиболее полно охарактеризовать изучаемые биологические системы. Предполагается, что проект расширит знания о такой сложной и динамической структуре как протеом человека, а с помощью полученных данных будут выявлены новые механизмы взаимодействия белков, обнаружены не охарактеризованные белки и установлена их функция. Также в протеоме человека присутствуют варианты белков, образованные альтернативным сплайсингом и полученные в результате однонуклеотидного полиморфизма в последовательности гена, кодирующего белок (cSNP). Таким образом, белков, кодируемых 20000 генами человека с учетом изоформ гораздо больше, по разным оценкам их может быть более 100000. На начало проекта в 2010 году было зарегистрировано около 14000 белков протеома человека, информация об остальных предполагаемых белках вовсе отсутствовала. К тому же, даже зарегистрированные белки не были полностью описаны, отсутствовала информация о возможных посттрансляционных модификациях (ПТМ), внутриклеточной локализации белков, тканеспецифичности, функции и концентрации белков в различных органах и тканях человека, а также о взаимодействиях с другими белками протеома человека [20]. Завершение этого проекта улучшит понимание биологии человека на клеточном уровне и заложит основу для разработки диагностических, прогностических, терапевтических и профилактических медицинских приложений.

Проект «Протеом человека» человека разделился на множество инициатив, одна из которых представляет собой хромосомоцентричный проект протеома человека (С-НРР). Основная цель С-НРР (Chromosome-centric Human Proteome Project) – эффективная интеграция протеомных данных в структуру генома каждой хромосомы, что приведет к углублению знаний о сложных биологических системах и облегчит доступ к данным о белках человека. Данный подход предложен Янг-Ки Пайком и его коллегами на нескольких встречах HUPO (Амстердам, август 2008 г.; Торонто, сентябрь 2009 г.; Сидней, сентябрь 2010 г.) [21]. Хромосомоцентричный проект протеома человека подразумевал разделение протеома по хромосомам между странами участниками проекта. Первой была выбрана 13 хромосома командой исследователей из Кореи во главе с Янг-Ки Пайком, инициатором проекта. После этого оставшиеся хромосомы были распределены между остальными странами, участвующими в проекте. Исследователи из России под руководством академика Александра Ивановича Арчакова остановили свой выбор на 18 хромосоме человека. Российская группа занималась не только протеомом, но и транскриптомом, так как для части белок-кодирующих генов не зарегистрированы даже соответствующие транскрипты, кодируемые генами 18 хромосомы человека. Основные биологические объекты, которые использовались в исследованиях: сыворотка крови человека, клетки линии НерГ2, клетки печени человека. Несмотря на то, что 18 хромосома человека сравнительно небольшая, на ней находится всего 264 белок кодирующих гена, они участвуют во многих важных биологических процессах. Например, P25705-АТРА_HUMAN (альфа субъединица АТФ-синтазы) является частью комплекса, катализирующего реакцию синтеза АТФ в митохондриях, обеспечивает клетку энергией. Также, генами 18 хромосомы человека кодируются белки, участвующие в передаче сигнала в метаболическом пути TGF- β , что приводит к активации транскрипционного фактора SMAD2/SMAD4. Данный метаболический путь может играть роль супрессора при колоректальной карциноме. Также белки, кодируемые генами 18 хромосомы, принимают участие в процессах регуляции клеточного цикла, метаболизма фосфолипидов, апоптозе, адгезии, дифференцировки клеток, аутофагии и так далее. Наиболее многочисленные 20 категорий биологических процессов, в которых участвуют белки, кодируемые генами 18 хромосомы человека, представлены на рисунке 1.

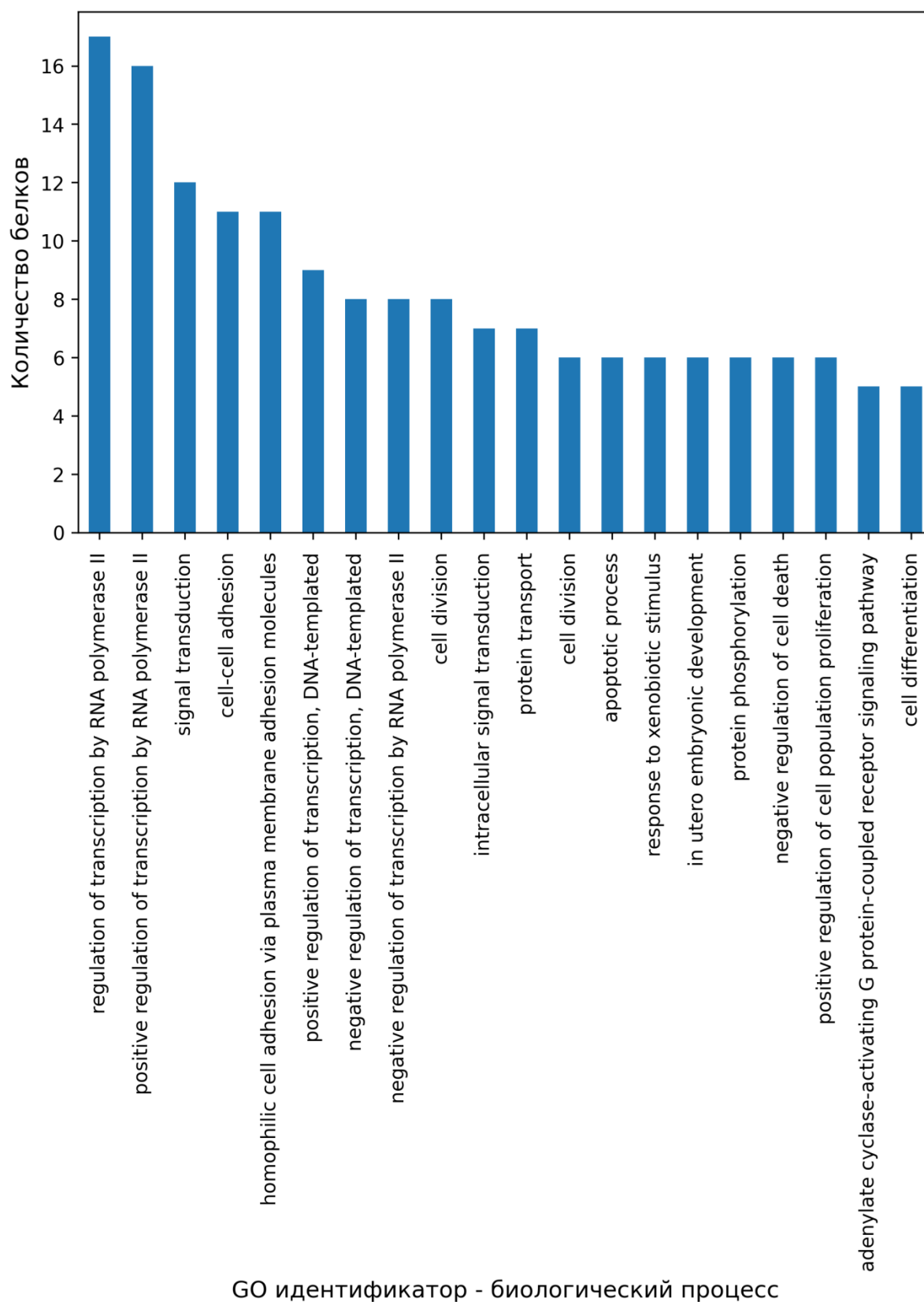


Рис. 1 Распределение белков, кодируемых генами 18 хромосомы человека, по GO идентификаторам, представляющим биологические процессы, в которых участвуют белки. Столбец показывает количество белков, которые участвуют в процессе.

Проект состоит из четырех основных направлений: биоинформатическое (обработка и хранение данных), масс спектрометрическое (масс спектрометрическая детекция белков), иммунологическое (детекция белков с помощью специфичных антител), патологическое (исследование заболеваний). Иммунологическое направление предназначено для продвижения иммунологических подходов к исследованию и анализу протеома человека и интеграции с другими методиками исследования белков. Задачи, которые решаются в рамках этого направления – это создание открытого каталога валидированных антител из многих различных источников и для множества иммуноанализов, создание базы данных ресурсов по антителам (antibodypedia.org), где в свободном доступе находится вся информация, накопленная в ходе проекта по антителам к человеческим белкам [22]. В ближайшей перспективе планируется сфокусироваться на идентификации белков и регистрации экспрессии и внутриклеточной локализации белков. Как минимум один продукт каждого белок кодирующего гена будет использован в рамках проекта протеома человека для разработки соответствующих антител, результаты будут храниться в специально разработанной базе данных Human Protein Atlas [23]. На данный момент (HPA версии 22, обновление 2022 года) в базе данных насчитывается около 27000 антител к 17200 белков протеома человека для использования в иммуногистохимии и иммунофлюоресценции. Долгосрочной целью проекта является построение полной карты экспрессии и внутриклеточной локализации белков в тканях и органах человека на разных стадиях развития и при разных физиологических и патологических состояниях. Планируется накопление и анализ данных о специфичности разработанных антител и потенциала использования в различных иммуноанализах: вестерн блоттинг, ELISA и других. Также одной из задач является сравнение и документация полученных данных о зарегистрированных концентрациях белков с помощью антител и масс спектрометрических методов анализа. Задачами масс спектрометрической части является разработка более чувствительных методов анализа, разработка критериев оценки качественного определения “missing” белков (белки для которых предсказано наличие белок-кодирующих генов, но которые еще не были зарегистрированы в эксперименте), качественное и количественное определение белков в тканях человека и в клеточных линиях. Ключевой целью является разработка наборов протеотипических пептидов для SRM анализов и доступ к ним для всех желающих, в том числе и пептидов, меченных стабильными изотопами, которые

позволяют провести абсолютный количественный анализ с высокой точностью и воспроизводимостью. В краткосрочной перспективе планируется создать библиотеку пептидов хотя бы для одного продукта каждого гена человека. Данная цель будет достигнута с помощью исследования спектров фрагментации пептидов на различных масс спектрометрах и накопления данных об интенсивностях фрагментных ионов в спектре [24]. В среднесрочной перспективе планируется расширение библиотеки SRM пептидов. В нее будут включены пептиды для анализа сплайс вариантов, одноаминокислотных полиморфизмов и возможно большинства известных посттрансляционных модификаций. В настоящее время создана специализированная база данных SRMAthlas (версия БД от 01-2023), содержащая 17242 белка протеома человека, для каждого из которых представлено по несколько протеотипических пептидов [25, 26].

Биоинформатическая часть призвана объединять и хранить полученные данные, а также разрабатывать новые алгоритмы их обработки. Рабочая группа проекта протеом человека решила использовать UniProtKB/SwissProt, PRIDE, PeptideAtlas, GPMDB, и Human Protein Atlas базы данных и инфраструктуру ProteomeExchange для обмена данными. Таким образом, проект «Протеом человека» опирается на уже существующие ресурсы, которые будут улучшаться и расширяться для удовлетворения специфических нужд проекта. Кроме того, разработан специальный ресурс, который объединяет в себе знания, накопленные в ходе проекта «Протеом человека», в частности результаты масс спектрометрических измерений и иммуноанализов с полным спектром функциональных и структурных данных о белках человека – это база данных peXtProt. Проект «Протеом человека», кроме основных направлений деятельности включает в себя много различных инициатив. Например, некоторые научные группы прицельно изучают субпротеомы определенных органов или биологических жидкостей: протеом плазмы, протеом печени, протеом мозга, протеом почки/мочи, протеом скелетных мышц. Также некоторые участники проекта сфокусировались на определенных заболеваниях: злокачественные опухоли, сердечно-сосудистые заболевания, инфекционные заболевания, ревматоидные заболевания. В том числе существуют ответвления, изучающие отдельные органеллы, например, митохондрии или сфокусированные на определенном направлении протеомики – иммунопептидомика и гликопротеомика [27, 28, 29].

Всего база знаний neXtProt насчитывает 19750 предполагаемых белковых последовательностей (версия БД от 2022-08-18) с указанием уровня доказательств по фактическому существованию белка (PE1-PE5). Из них точно известно, что на белковом уровне зарегистрировано 18407 продуктов генов (PE1), обнаружено на уровне транскриптома, но еще не подтверждено на протеомном уровне - 1135 (PE2), предсказано на основе гомологии с другими организмами - 195 (PE3), биоинформатически предсказано наличие белок кодирующих генов - 13 (PE4), информация о том, кодируют ли данные гены белки отсутствует - 609 (PE5). Таким образом “missing” называют белки уровня PE2-PE4, всего на данный момент существует 1343 “missing” белков, динамика регистрации белков с начала проекта показана на рисунке 2.

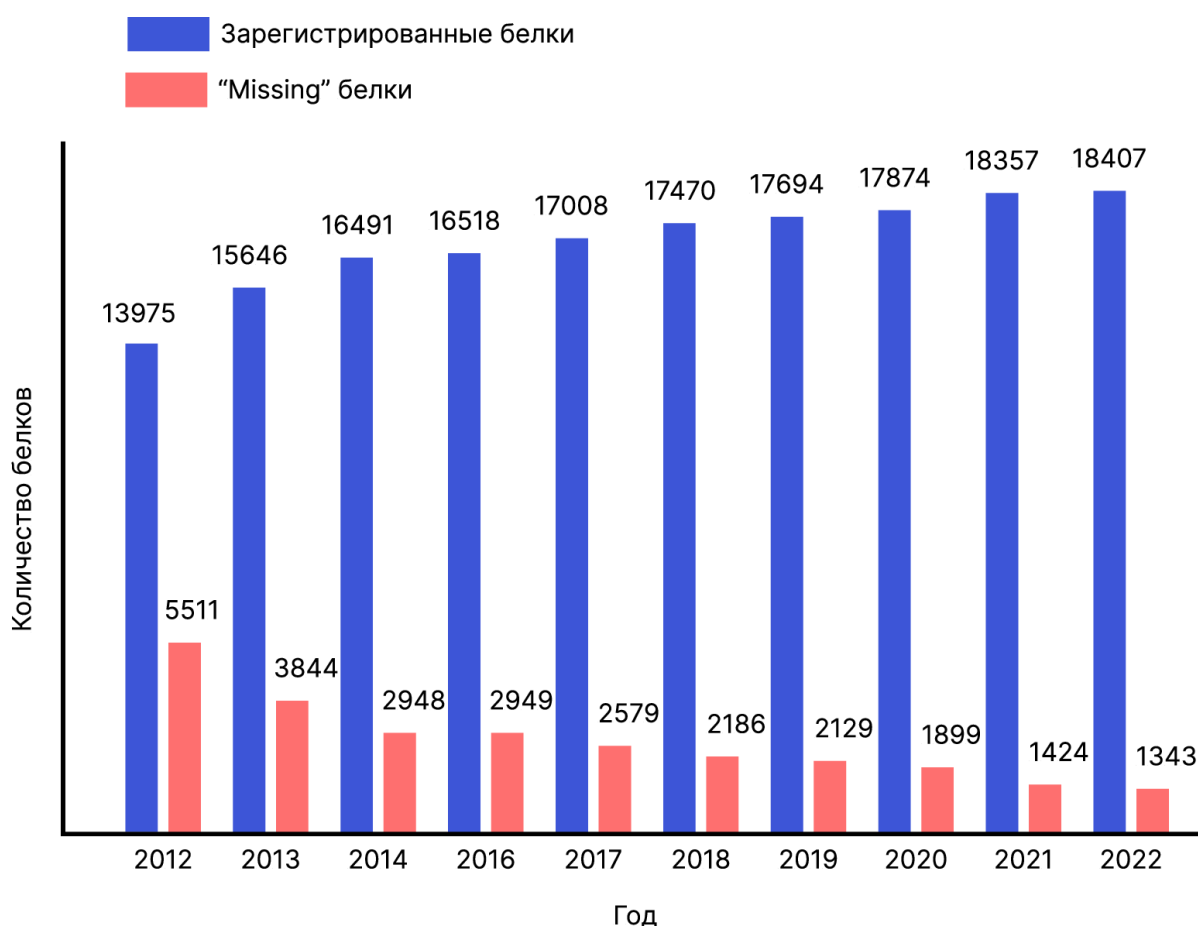


Рис. 2 Динамика регистрации “missing” белков (PE2+PE3+PE4) по годам в ходе проекта «Протеом человека».

В начале проекта были разработаны общепринятые правила валидации белков, особенно важно это было для подтверждения существования “missing” белков:

необходимо было зарегистрировать минимум 2 пептида белка длиной не короче 7 аминокислотных остатков и при FDR на уровне белковой последовательности не выше 1%. Однако такие критерии были недостаточно строгими. В 2014 году вышла масштабная работа по анализу широкого спектра органов и тканей человека, в том числе зародышевых, где было обнаружено несколько сотен “missing” белков. После тщательного анализа данных этой работы в ручном режиме не было обнаружено достаточных свидетельств для валидации заявленных “missing” белков. После этого протеомное сообщество запросило переработку критериев идентификации “missing” белков [30]. По новым правилам для валидации уже известных PE1 белков достаточно было пройти отсечение в 1% FDR на уровне белка. Для “missing” белков были разработаны новые критерии: необходимо было зарегистрировать 2 уникальных, неперекрывающихся протеотипических пептида длиной минимум 9 аминокислотных остатков, спектры которых полностью соответствовали синтетическим пептидам. В результате все данные о “missing” белках были переработаны в соответствии с новыми стандартами качества, что привело к уменьшению общего количества PE1 белков на 485 к релизу 2016 года и их перевод в категории PE2, PE3, PE4. Новые стандарты качества существенно замедлили детекцию новых “missing” белков с 2014 года. Однако, несмотря на это, уже более 90% протеома человека зарегистрировано, но все еще не описаны функции 1254 белков. Поэтому в 2017 году была запущена инициатива CP50. Порядка 90% исследований фокусируются на группе из 10% наиболее описанных белков, поэтому исследователям по всему миру было предложено исследовать белки с неизвестной функцией [31].

Совокупность генов локализованных на 18 хромосоме человека насчитывает 264 белок кодирующих гена, из которых 257 находится на уровне PE1, а 7-PE2. Таким образом, на 18 хромосоме человека содержится 7 генов, которые могут кодировать “missing” белки. Используемые в данной работе методы также могут быть направлены на детекцию “missing” белков, так как они вероятнее всего экспрессируются в клетках в ультранизких концентрация (10^{-12} М и ниже) и их возможно обнаружить только с помощью самых чувствительных аналитических методов.

Поиск ранее не зарегистрированных белков предполагается вести в любом биологическом материале человека, в том числе, чтобы составить карту экспрессии белков и определить тканеспецифичность. Однако биологический материал обладает

высокой степенью вариабельности протеома, наивысшая степень которой зарегистрирована в протеоме сыворотки крови человека. Она является очень привлекательным материалом для исследования, так как относительно доступна и может быть получена малоинвазивным способом. В теории в этом виде биоматериала могут встречаться все белки протеома человека, так как сыворотка крови омывает все органы и ткани человека [32]. Однако большую часть белков сыворотки крови человека составляют хорошо известные белки, присутствующие в высокой концентрации, например 55% всех белков плазмы составляет сывороточный альбумин человека (NX_P02768), 38% - глобулины, 6% - фибриноген (NX_P02671). В оставшемся 1% и предполагается обнаружение всех белков протеома человека [33]. В рамках C-HPP группой российских ученых проведено исследование сыворотки крови кандидатов в космонавты. Проводилось исследование уровня экспрессии белков, кодируемых генами 18, 13, Y и митохондриальной хромосомой. Исследовано наличие и уровень экспрессии 619 белков, которые потенциально могли быть зарегистрированы в сыворотке крови. Всего зарегистрировано 205 белков хотя бы в одном из 54 образцов. Однако стабильно обнаруживались в каждом образце с низкой межиндивидуальной вариабельностью всего лишь 30 белков [34]. В плазме также зарегистрирован самый высокий динамический концентрационный диапазон в 9 порядков и возможно даже выше [35]. Все эти факторы не позволяют зарегистрировать существенное число белков в сыворотке крови человека, использование стандартного панорамного протеомного анализа позволяет зарегистрировать всего несколько сотен белков, когда в клеточном материале обычно регистрируется порядка 2-3 тысяч белков. В клеточном материале не наблюдается преобладания нескольких белков в высокой концентрации, как это показано в сыворотке крови, а также динамический концентрационный диапазон снижен и составляет 6 порядков [36]. В рамках проекта исследовались концентрации белков, кодируемых генами 18 хромосомы в клетках линии HepG2 и ткани печени. Таргетным методом SRM SIS зарегистрировано 66 и 73 белка соответственно, в то время как в сыворотке в 75% биологических образцов регистрируется 8-12 белков, кодируемых генами 18 хромосомы человека [34]. Поэтому “missing” белки более вероятно найти именно в клеточных линиях или тканях человека.

2.3 Проблемы регистрации белков при МС анализе, ограничение чувствительности прибора

Главная проблема масс спектрометрических методов – низкая чувствительность прибора. Современные наиболее чувствительные приборы и методики способны регистрировать белки в концентрации 10^{-11} М или аттмолярном количестве. Однако такая чувствительность показана на упрощенной белковой системе UPS2, которая имитирует диапазон концентраций белков от 10^{-6} М до 10^{-11} М (0.5 фмоль – 50 пмоль). При добавлении клеточного лизата *E.coli* чувствительность снижается на 2-3 порядка концентрации [37]. Таким образом, возникает проблема высокого динамического диапазона концентраций в биологическом образце. Клеточный лизат *E.coli* содержит большое количество белков в высокой концентрации, которые сложно зарегистрировать одновременно, и уже в такой системе детекция низкоконцентрированных белков затруднена, поэтому наблюдается снижение количества детектированных белков набора UPS2. Современными и наиболее распространенными масс спектрометрами, которые используются в протеомике, являются сложные гибридные приборы, оснащенные орбитальной ловушкой (Orbitrap) в качестве масс анализатора и приборы типа тройной квадруполь (QQQ).

2.3.1 Масс анализатор типа Orbitrap

В 1990-ых годах нашим соотечественником Макаровым, возглавлявшим группу физиков, разработан масс анализатор Orbitrap. Масс анализатор изобретен после ICR и унаследовал определенные характерные черты своих предшественников. Функции предыдущих моделей соединенные в новом приборе дали исследователям новые аналитические возможности. Также они помогли устранить некоторые ограничения предыдущих моделей, такие как необходимость в сверхпроводящем магните в ICR, динамический диапазон детекции в TOF анализаторах. Главным отличием от ICR является отсутствие магнитного поля, что очень удешевляет содержание приборов с масс анализатором типа Orbitrap, а также по сравнению с ICR у Orbitrap масс анализаторов выше разрешающая способность в области больших значений отношения массы к заряду [38].

Orbitrap масс анализатор представляет собой ловушку для ионов, которые после инъекции, при воздействии постоянного электрического поля осциллируют в пространстве между внутренним и внешним электродом вдоль внутреннего электрода.

Внешние электроды имеют форму чашек, обращенных друг к другу. Они электрически изолированы тонким зазором, закрепленным центральным кольцом диэлектрика. Веретенообразный центральный электрод удерживает ловушку вместе и выравнивает ее с помощью диэлектрических концевых прокладок. Когда напряжение подается между внешним и центральным электродами, результирующее электрическое поле строго линейно вдоль оси, следовательно, колебания вдоль этого направления будут чисто гармоническими. В то же время радиальная составляющая поля сильно притягивает ионы к центральному электроду. Ионы инжектируются в объем между центральным и внешним электродами по касательной через специально обработанную прорезь с компенсационным электродом («дефлектором») в одном из внешних электродов. При приложении напряжения между центральным и внешним электродами радиальное электрическое поле изгибает траекторию иона к центральному электроду, в то время как тангенциальная скорость создает противодействующую центробежную силу. При правильном выборе параметров ионы движутся внутри ловушки почти по круговой спирали, очень похожей на траекторию движения планеты в солнечной системе. В то же время осевое электрическое поле, вызванное особой конической формой электродов, толкает ионы к самой широкой части ловушки, инициируя гармонические осевые колебания. Затем внешние электроды используются в качестве приемных пластин для обнаружения тока этих осевых колебаний. В этом случае так же, как и в ICR применяются преобразования Фурье для преобразования сложных волновых функций движения ионов в отношение массы к заряду. Одно из самых важных достижений, которое сделало возможным использование Orbitrap на практике - изобретение C-trap. Ионная ловушка в форме буквы С находится перед Orbitrap и последовательно соединена с ним. Она служит накопителем ионов, а по достижению заданного количества ионов происходит импульсная инжекция накопленных ионов в Orbitrap из C-trap. Это изобретение сделало возможным подключение любого устройства передачи ионов и любого метода фрагментации к Orbitrap масс анализатору [39].

2.3.2 Shotgun анализ

Стандартный панорамный МС анализ выполняется следующим образом: пептиды элюируют с обращенно фазовой колонки и попадают в источник ионизации, где переходят в газовую фазу и приобретают заряд. В приборе заряженные частицы аккумулируются в 3D ионной ловушке и при накоплении определенного числа зарядов

в ловушке инжестируются в Orbitrap. Таким образом, происходит сканирование на уровне MS1, или измерение m/z родительских ионов, то есть цельных пептидов в широком окне m/z , обычно в диапазоне 350-1500 m/z . Прибор анализирует MS1 скан на наличие потенциальных пептидных ионов, одним из главных критериев является прохождение отсечения по интенсивности родительского иона. При наличии таких сигналов происходит инициация накопления и фрагментации узкого диапазона масс, чаще всего не более 2 m/z , в котором был зарегистрирован интенсивный ион. После этого происходит регистрация спектра всех полученных дочерних ионов или MS2 сканирование. Именно на основе MS2 сканов в дальнейшем происходит идентификация пептидов и белков поисковыми машинами при обработке данных [40]. Таким образом, методика позволяет определить качественный состав белков образца, достаточно зная исследуемый организм и последовательности белков для обработки данных. Также существует множество алгоритмов, которые позволяют провести полуколичественный анализ полученных данных и, например, сравнить экспрессию белков в контрольной и опытной группе образцов. В настоящий момент разработано большое количество алгоритмов расчета полуколичественных показателей, которые основаны на подсчете количества MS2 спектров пептидов, интенсивности MS1 пика и так далее [41, 42].

Однако, масс спектрометры типа Orbitrap в режиме стандартного панорамного сканирования ограничены в своей чувствительности из-за неэффективного сканирования на уровне MS1 при анализе комплексных биологических образцов, таких как: цельные клеточные лизаты, ткани человека, сыворотка крови человека. Его эффективность снижена вследствие конструктивных особенностей прибора и высокого динамического диапазона концентраций в биологических образцах. Ионная ловушка (C-trap) способна накапливать $1 \cdot 10^6$ зарядов. Эта особенность является причиной регистрации наиболее интенсивных ионов, так как в ловушке при каждом сканировании накапливаются ионы из широкого диапазона сканирования MS1. Таким образом, ловушку быстро заполняют наиболее интенсивные ионы, а низкоинтенсивные не успевают накопиться в ловушке до определенной интенсивности, чтобы прибор принял решение об их изоляции и проведении MS2 сканирования [43].

2.3.3 Метод VoxCar

Предпринимаются попытки преодолеть проблему нерационального использования C-trap и создать новые методы. Одним из таких методов является VoxCar [44]. Исследователи, разработавшие его пошли по пути улучшения эффективности MS1 анализа. Они установили, что во время элюции пептидов время накопления ионов в ловушке всегда минимально. Поэтому они решили заполнять ловушку не сразу всем широким диапазоном ионов от 350 m/z до 1500 m/z, а заполнять его последовательно узкими полосами m/z с помощью квадруполя, которые охватывают $\frac{1}{2}$ или $\frac{1}{3}$ всего диапазона m/z. В таком случае диапазон сканирования искусственно поделен на 20 частей, каждая часть представляет собой 57,5 m/z узкую полосу всего спектра. Каждая такая полоска, например 350-407,5 m/z заполняется на 100000 зарядов, далее заполняется 465-522,5 m/z на 100000 зарядов, и так далее еще 8 окон, пока в C-trap не накопится 1000000 зарядов. После этого ионы инжектируются в Orbitrap и происходит MS1 скан. В следующий MS1 скан заполняются другие окна с m/z, которые не были охвачены в предыдущем скане. В последующем выбор ионов на MS2 сканирование происходит также как и при стандартном панорамном MS анализе, но выбор ионов значительно больше, так как MS1 соотношение сигнал/шум значительно выше в таких сканах по сравнению со стандартным сканом по всему диапазону. Таким образом, разделение всего полотна широкого диапазона MS1 на небольшие участки позволяет низкоинтенсивным пептидным ионам накопиться в ловушке и в последующем зарегистрировать их с помощью MS2. В результате такая методика дает в 2 раза больше белковых идентификаций, чем стандартный панорамный MS анализ. Для VoxCar не требуется спектральной библиотеки, кроме стандартной базы данных с последовательностью белков. Принципиально это тот же самый панорамный MS анализ, но с усовершенствованной системой отбора ионов на MS2 анализ, однако такой метод не предусмотрен разработчиками ПО для приборов и разработчики VoxCar сообщали о сбоях в электронике прибора. В дальнейшем они разработали специальное программное обеспечение, которое позволяет использовать данный метод без перебоев в работе прибора.

2.3.4 Метод DIA

Совершенно иной подход используется при DIA, однако получены похожие результаты [45]. В этой методике появилась идея полностью отказаться от MS1 сканирования. В таком случае встает проблема определения массы родительского

иона. Эта проблема решается тем, что MS2 скан проходит в узком диапазоне масс. Для одного MS2 сканирования используют небольшое окно m/z , например 20 m/z . За один цикл работы прибора накапливается 10^6 зарядов в диапазоне 400-420 m/z и все, что накопилось, подвергается фрагментации и MS2 сканированию. Все что мы знаем об уровне MS1 это то, что в этом MS2 скане находятся фрагменты родительских ионов m/z которых находится в диапазоне 400-420 m/z . Такой точности недостаточно, чтобы правильно определить последовательность пептида. Поэтому перед началом такого анализа создают спектральную библиотеку с помощью стандартного панорамного MS анализа [46]. При этом такая библиотека должна состоять из большего числа белков и пептидов, чем обычно получают при стандартном панорамном MS анализе. В этом случае пользуются методами биохимического фракционирования образца. Тогда будет получен более глубокий протеом исследуемого объекта и после DIA анализа возможно будет соотнести времена удерживания пептидов и интенсивности и определить последовательность. Такой метод более трудоемкий, чем BoxCar, однако сохраняет машинное время и экономит количество образца по сравнению с биохимическим фракционированием. Однако сейчас появляется программное обеспечение, которое позволяет проводить DIA анализ без спектральной библиотеки [47].

2.3.5 Другие способы преодоления ограничений метода Shotgun

Один из способов увеличения чувствительности-последовательно элюировать образец с хроматографической колонки при масс спектрометрическом анализе различными растворителями (AcN, MeOH, Acetone). Таким образом, можно хроматографически разделить пептиды, которые при элюции ацетонитрилом коэлюируют и получить новые идентификации. Однако, эффективность методики относительно низкая, увеличение чувствительности составляет 10% [48]. Добавление ДМСО в количестве 5% в хроматографическую фазу, увеличивает число идентификаций на 20%. Механизм этого явления до конца не ясен, однако увеличивается интенсивность пептидных ионов в присутствии ДМСО, что позволяет получить большее количество идентификаций и улучшает стабильность количественного анализа [49, 50]. Эффективность этих методов ограничена, однако добавление ДМСО в хроматографическую фазу можно легко скомбинировать с методиками BoxCar или DIA, что в результате может увеличить число зарегистрированных белков.

2.3.6 Чувствительность приборов с Orbitrap масс анализатором

Так как стандартный панорамный МС анализ не предполагает регистрации абсолютных концентраций белков, то чувствительность обычно измеряют по количеству зарегистрированных белков в биоматериале. Один из последних приборов в линейке Orbitrap Exploris 480 показывает чувствительность около 3000 белков, используя нанограммовые количества стандартного клеточного лизата Hela в качестве биологического образца, за 60 минут хроматографического градиента. Однако некоторые исследователи используют синтетические пептиды для измерения предела детекции. Так для определенных пептидов измерен предел чувствительности на уровне аттмолярных и даже субаттмолярных концентраций с помощью DIA анализа [51]. Эти данные были получены на упрощенной системе из нескольких пептидов, а не на фоне комплексного биологического образца. Также в одном из исследований использовалась упрощенная белковая система UPS2, которая состоит из 48 белков и имитирует динамический диапазон концентраций от 500 аттомоль до 50 пмоль. Было показано, что прибор в DIA режиме способен зарегистрировать только 2 наиболее концентрированные группы белков с минимальной концентрацией в 140 фмоль [52]. В одной из работ была измерена чувствительность прибора при добавлении известной концентрации моноклонального антитела на фоне лизата из *S.cervisiae*. Исследуемое антитело было зарегистрировано в наименьшей концентрации 67,6 аттомоль или 10 пикограмм [53].

2.3.7 Квадрупольный масс анализатор

Квадрупольный масс анализатор является масс фильтром и пропускает только ионы определенного m/z . Электрическое радиочастотное поле используется для направления потока заряженных частиц вдоль центральной оси квадруполя. Накладываемое поле постоянного тока используется для избирательной дестабилизации определенных ионов и их удаления из квадруполя. Сила обоих полей может быть настроена так, чтобы только небольшой диапазон m/z имел стабильную траекторию движения вдоль квадруполя. Все ионы вне этого диапазона будут удалены, а частицы внутри этого диапазона будут проходить далее. Квадруполи, гекса и октаполи могут использоваться в качестве одной из деталей сложных гибридных масс спектрометров для изоляции родительского иона. Такая модификация используется для накопления и выделения одного иона и его последующей фрагментации. Также используются 3 квадруполя, соединенных последовательно. Первый квадруполь используется для изоляции

родительского иона, второй как ячейка соударения для фрагментации иона, третий – для изоляции дочернего иона. Такая конфигурация прибора используется для таргетного протеомного анализа и мониторинга множественных реакций (SRM). Такая конструкция и принцип работы делает тройной квадруполь очень специфичным, но при этом точность масс такого прибора низкая-0,4 m/z [54].

2.3.8 SRM SIS анализ

Концепция SRM анализа отличается от Shotgun тем, что исследователь заранее должен выбрать и явно указать при подготовке анализа интересующие его аналиты. Для регистрации пептидов выбирают обычно несколько наиболее интенсивных фрагментных ионов и проводят оптимизацию параметров каждого перехода: энергию соударения, ускоряющий потенциал после ячейки соударения. Также для абсолютного количественного анализа необходимы внутренние стандарты. Для SRM SIS используются меченные стабильными изотопами пептиды с последовательностью идентичной эндогенному пептиду образца. Регистрация стандарта и эндогенного пептида параллельно в ходе анализа позволяет провести абсолютный количественный анализ содержания белка опосредованно по концентрации соответствующих пептидов. Причина, по которой измерения абсолютного количественного содержания пептидов проводятся на приборе типа тройной квадруполь в его высокой стабильности и воспроизводимости измерений.

При этом в приборах, на которых традиционно выполняется этот вид анализа (тройной квадруполь) нет накопительных элементов с ограниченной емкостью, поэтому даже самый слабый сигнал от аналита в пределах возможности прибора будет зарегистрирован [55]. Таким образом, аналитическая чувствительность SRM метода значительно превышает стандартный Shotgun анализ. В молярных концентрациях это составляет 3 порядка, если Shotgun обычно регистрирует белки до наномолярных концентраций, то SRM способен детектировать белки в фемтомолярных концентрациях и в отдельных случаях даже ниже [56, 57, 58]. Однако SRM анализ полностью не может заменить Shotgun, так как его мультиплексность ограничена. В SRM анализе наиболее высокая эффективность достигается при полном хроматографическом разделении всех аналитов, добиться чего не всегда так просто. В случае если интересующие соединения коэлюируют эффективность анализа снижается и в таком режиме целесообразно регистрировать до 500 аналитов, в то время как при

стандартном панорамном МС анализе можно идентифицировать несколько тысяч белков, что в среднем составляет 20000 пептидов за один рабочий цикл прибора [59].

2.3.9 Чувствительность приборов типа тройной квадруполь

Пока большинство исследователей отмечает, что конфигурация прибора типа тройной квадруполь и анализ в режиме мониторинга множественных реакций является лучшим методом по чувствительности и специфичности. Так в одной из работ по регистрации уровня 6 пептидов сыворотки крови была показана чувствительность на уровне от 10 фмоль до 500 аттомоль [60]. Было показано, что прибор QTRAP 5500 превосходит по чувствительности прибор с масс детектором типа TOF. В другой работе было продемонстрировано, что тройной квадруполь Agilent 6495 способен регистрировать 4 аттомоль изотопно-меченного пептида на фоне биологической матрицы из белков плазмы крови человека [61].

Наиболее чувствительные приборы, которые используются в протеомике, оснащены источником ионизации электрораспылением. При хроматографическом разделении комплексных биологических образцов происходит коэлюция и затем коионизация большого количества пептидов одновременно. Это приводит к конкуренции за протон и подавлению и искажению сигнала низкоинтенсивных пептидных ионов [62]. Это проблема методики LC-ESI-MS, которая наблюдается в каждом приборе, который объединен в систему с жидкостным хроматографом. Кроме того существует проблема чувствительности детектора масс-спектрометра и недостаточной скорости сканирования. В будущем с развитием приборной базы и совершенствованием методики ионизации и хроматографических методов ожидается решение этих проблем. Таким образом, ограничение чувствительности прибора ожидается преодолеть с помощью новых инженерных открытий, а с точки зрения биохимии их решить невозможно. Однако существует проблема высокого динамического диапазона концентраций в биологических образцах, и решить эту задачу возможно с помощью методов фракционирования.

2.4 Методы фракционирования, используемые в современной протеомике

2.4.1 Фракционирование клеточных органелл

Протеомы клеточных компартментов и органелл могут быть выделены с помощью нескольких подходов, наиболее часто используемой стратегией является фракционирование внутриклеточных компонентов [63, 64]. Процедура фракционирования внутриклеточных компонентов включает разрушение клеток и тканей с использованием механических, химических или физических методов. Для успешного фракционирования внутриклеточных структур при выделении важно разрушить мембрану клетки, но сохранить целостность внутриклеточных органелл. Обычные методы фракционирования внутриклеточных компонентов, как правило, предусматривают серию из нескольких стадий центрифугирования для выделения различных популяций клеточных компартментов или органелл на основе их массы и/или плотности путем дифференциального центрифугирования и центрифугирования в градиенте плотности [65].

Дифференциальное центрифугирование

Дифференциальное центрифугирование осуществляется посредством последовательного центрифугирования гомогената клеток или тканей для получения внутриклеточных органелл, таких как ядра, митохондрии и лизосомы. Этот метод разделения основан на разности в размере и плотности компартментов клетки, крупные и более плотные органеллы осаждаются при более низких центробежных силах. Однако этот метод имеет низкую разрешающую способность. Таким образом, с помощью дифференциального центрифугирования сложно получить чистые фракции определенных органелл. В полученных фракциях часто будет наблюдаться смесь компартментов, которые имеют сходные скорости седиментации.

Центрифугирование в градиенте плотности

Более широко применяемый метод центрифугирования в градиенте плотности разделяет органеллы с помощью непрерывных или ступенчатых градиентов с использованием сред различной осмолярности, вязкости или плотности. Как правило, готовится среда, на которую наносится клеточный гомогенат и проводится центрифугирование. При центрифугировании органелла фокусируется в градиенте, где ее плотность равна плотности окружающей среды. Таким образом, скорость осаждения

органелл в градиенте плотности зависит от разности плотности органелл и среды, что, в свою очередь, определяется биохимическим составом органелл, соотношением липидов и белков, размером и формой [66]. Например, митохондрии и эндоплазматический ретикулум имеют высокую плотность из-за высокого содержания белка, тогда как эндосомы имеют низкую плотность так, как их мембраны богаты липидами. Сахароза является наиболее часто используемой средой для центрифугирования в градиенте плотности, поскольку сахароза биологически инертна и доступна. Также для дифференциального центрифугирования используют фиколл, перколл, никоденз и мертизамид [67]. Градиенты плотности делятся на непрерывные и ступенчатые. При использовании непрерывного градиента плотность изменяется линейно, в то время как ступенчатый градиент представляет собой фиксированные участки с различной плотностью. При центрифугировании на ступенчатом градиенте обогащение органеллами происходит на границах раздела фаз. С другой стороны, при центрифугировании в непрерывном градиенте происходит равновесное разделение, при котором органеллы распределяются по всему градиенту и фокусируются в своих точках равновесия. Это обеспечивает лучшее разрешение органелл, чем при использовании ступенчатых градиентов. Однако приготовление непрерывного градиента может занять много времени и потребовать специального оборудования. Центрифугирование в градиенте плотности часто используется исследователями, поскольку этот метод хорошо описан и применяется во многих протеомных исследованиях органелл [68].

Дифференциальное фракционирование детергентом

Метод дифференциального фракционирования детергентом сочетает в себе центрифугирование и использование различных детергентов для экстракции определенных органелл. Фракционирование детергентом является простой и надежной методикой, которая позволяет разделить органеллы соответственно пяти внутриклеточным компартментам: цитозоль, мембранные органеллы, ядерная фракция, цитоскелет и матрикс [69, 70]. Она заключается в последовательной инкубации в растворах различных детергентов. Обычно в качестве детергентов используются растворы дигитонина/ЭДТА, тритона X-100/ЭДТА, tween/дезоксихолата, после каждой инкубации необходима стадия центрифугирования и отбора супернатанта. Группой авторов был разработан новый надежный метод фракционирования детергентом с высокой эффективностью и селективностью, реагенты для методики исследователи не

раскрывают, а набор для выделения органелл распространяется на коммерческой основе. Его высокие показатели были подтверждены результатами 2D гель электрофореза, морфологическими, иммунологическими, ферментативными анализами.

Электрофорез в свободном потоке

Выделение органелл с помощью электрофореза в свободном потоке основано на их различиях в плотности заряда или изоэлектрических точках. Также электрофорез в свободном потоке используется для разделения белков и пептидов [71]. Во время разделения буфер-носитель непрерывно и ламинарно протекает в узком пространстве между двумя охлаждающими пластинами, при этом электрическое поле прикладывается перпендикулярно ламинарному потоку буфера и образца, что приводит к передвижению и разделению молекул образца в соответствии с поверхностным зарядом/изоэлектрической точкой [72]. Некоторые из преимуществ электрофореза в свободном потоке включают в себя широкий выбор pH, проводимости и состава растворов электролитов и степень покрытия образца (приблизительно около 95%), которая достигается за счет введения образцов в виде тонкой пленки в растворе, что исключает использование матриц, таких как полиакриламидный гель в электрофорезе. Непрерывное фракционирование и сбор образцов на другом конце камеры для электрофореза хорошо подходит для препаративного разделения органелл. Кроме того, очищенные органеллы сохраняют свою целостность и функциональность, что делает данную методику привлекательной не только для протеомного анализа, но и для функциональных исследований. Одна из проблем электрофореза в свободном потоке заключается в том, что некоторые органеллы или клеточные структуры имеют сходные изоэлектрические точки, что может привести к их совместной миграции. Например, было обнаружено, что при разделении органелл нейтрофилов фракция, в которой ожидалось зарегистрировать секреторные везикулы, была контаминирована ЭР и митохондриями, вероятно, в результате сходства изоэлектрических точек. Похожим образом, митохондриальные и микросомальные белки часто регистрируются в препаратах пероксисом и, следовательно, для их разрешения требуются другие ортогональные подходы [73].

Иммуноаффинное выделение органелл

Иммуноаффинная очистка является мощным методом выделения органелл с высокой специфичностью и большим выходом [74]. Два основных метода, используются для выделения органелл: аффинная очистка и иммунопреципитация. Они основаны на связывании между лигандами (антителами), иммобилизованными на твердых носителях, и их мишенями (органеллами). Лигандами могут являться не только антитела, но также используются химические лиганды, комбинаторные библиотеки пептидов, аптамеры [75]. Для протеомного анализа аффинная очистка является хорошим вариантом в качестве первого этапа фракционирования из-за возможности проводить повторную очистку, следовательно, обогащение органеллы. Ограничениями аффинной очистки являются ее высокая стоимость и большая продолжительность эксперимента, которая необходима для эффективной очистки. Однако для сложных биологических образцов использование только аффинной очистки может привести к неудовлетворительному результату при последующем протеомном анализе. Поэтому, чтобы увеличить эффективность фракционирования, часто используют иммуноочистку в комбинации с другими методами. Например, было показано, что центрифугирование в градиенте плотности никоденза с последующей иммуноаффинной очисткой позволяет выделить пероксисомы из печени крысы [76]. Похожая стратегия выделения использовалась для обогащения секреторных синаптических пузырьков из синапсом [77].

Сами по себе методы фракционирования органелл скорее подходят для определения локализации белков. Однако в комбинации с другими методами фракционирования белков и пептидов могут значительно снизить динамический диапазон концентраций белков в биологическом образце, улучшить соотношение сигнал/шум при МС анализе и позволить регистрировать низкокопийные белки тканей и клеток человека.

2.4.2 Фракционирование белков клетки

Для эффективного разделения белков клетки необходимо выделить белки с полным разрушением клеток и внутриклеточных структур. Однако для получения более глубоких данных о белковом составе клеточных органелл, можно использовать фракционирование клеточных структур, затем выделить белки и фракционировать их другими методами.

Одномерный электрофорез в полиакриламидном геле в присутствии SDS (SDS-PAGE)

В протеомике эта методика часто используется для фракционирования белков клетки. Часто одномерный электрофорез в полиакриламидном геле в присутствии SDS используют при определении дифференциальной экспрессии клеток, валидации методов деплеции белков и поиске биомаркеров [78].

SDS (додецил-сульфат натрия) представляет собой детергент, который связывается с белками в эквимольном соотношении. При добавлении в образец SDS и восстанавливающего агента, который расщепляет дисульфидные S-S связи, белки утрачивают свою трехмерную структуру и денатурируют в более линейные цепи с отрицательным зарядом, пропорциональным длине полипептидной цепи. Это в значительной степени устраняет влияние собственного заряда белка, и его изначальной формы на электрофоретическую подвижность и белки будут разделяться по молекулярной массе. Низкомолекулярные белки будут иметь высокую подвижность, в то время как высокомолекулярные-низкую. Более точно молекулярную массу можно установить визуально при наличии образца набора белков стандартов с известной молекулярной массой. После разделения и окраски гель обычно нарезается на небольшие полоски, проводится ферментативный гидролиз в геле и масс-спектрометрическая идентификация белков.

Преимущества SDS-PAGE включают полную солиubilизацию белка с помощью SDS. В ходе гель-электрофореза из белкового экстракта удаляются детергенты, буферы и соли, которые могут мешать масс-спектрометрическому анализу. Кроме того, SDS-PAGE позволяет провести промежуточный уровень контроля наличия и разделения белков при окрашивании геля.

С помощью SDS-PAGE фракционирования и последующей масс-спектрометрической идентификации были выявлены 26 дифференциально экспрессируемых белка в клетках гладкой мускулатуры стимулированных ангиотензином II по сравнению с контролем [79]. Также методика была использована при изучении секретома раковых клеток и показано увеличение уровня остеопонтина при прогрессии опухоли [80].

Двумерный гель-электрофорез (2DE)

Двумерный гель-электрофорез включает в себя две последовательные стадии фракционирования белков. На первой стадии используется метод изоэлектрофокусировки белков (IEF), в ходе которого белки разделяются на основе разности поверхностного заряда. Однако IEF может использоваться не только для разделения белков, но и пептидов. IEF не обязательно должна быть частью методики 2DE, а может использоваться как отдельная самостоятельная методика фракционирования. Первое разделение в 2DE происходит в полосе геля, который образуется в результате полимеризации мономеров акриламида, связанных бис-акриламидом с молекулами иммобилина. Иммобилины различных pK_a могут создавать фиксированный градиент pH внутри акриламидного геля. Они широко используются в 2DE, поскольку полученный градиент IEF очень стабилен во времени и в сильном электрическом поле [81]. При приложении электрического поля, отрицательно заряженные белки движутся к аноду, а положительно заряженные молекулы движутся к катоду. Когда белки разделяются в соответствии с их изоэлектрической точкой, общий суммарный заряд равен нулю, и белок фокусируется в определенной точке геля. Фокусировка достигается с помощью специального устройства, способного выдавать до 8000 или 10 000 В, но с ограничением силы тока (максимум 50 мкА/полоска) для уменьшения нагрева. Также разработаны приборы, которые позволяют проводить разделение в двухфазной системе с верхней жидкой фазой, которая разделена на компартменты и нижней в виде геля. Фракционирование происходит в геле, где белки разделяются на основании разности поверхностных зарядов, однако затем могут быть экстрагированы в водную фазу для дальнейшего анализа. IEF — это не только метод разделения с высоким разрешением и высокой производительностью, он также предоставляет дополнительную физико-химическую информацию, а именно изоэлектрическую точку белка. Полученные значения pI можно использовать в качестве критерия фильтрации и валидации данных при анализе результатов МС исследования [82].

Вторая стадия двумерного гель-электрофореза заключается в разделении белков по молекулярной массе, принцип которого описан выше в разделе SDS-PAGE. На следующих этапах происходит визуализация белков в геле с помощью красителей. Наиболее часто используемыми красителями являются Кумасси, коллоидный Кумасси, который окрашивает белки в голубой цвет, или нитрат серебра. Они имеют

различную чувствительность: 50, 10, 0,5 нг белка/пятно соответственно. На следующем этапе гель нарезается на кусочки или вырезаются белковые пятна и проводится ферментативный гидролиз с последующей масс-спектрометрической детекцией белков. Методика двумерного гель-электрофореза позволяет проводить количественную оценку содержания белков, регистрировать изоформы белка и посттрансляционные модификации в широком диапазоне биологических систем.

В проекте «Протеом человека» кроме регистрации всех “missing” белков также стоит задача определить варианты белков, которые образуются в результате альтернативного сплайсинга и посттрансляционных модификаций. Двумерный гель-электрофорез как нельзя лучше подходит для решения второй задачи [83]. С помощью методики 2DE можно детектировать изоформы белка в виде отдельных пятен на геле и детектировать сплайс варианты или посттрансляционные модификации при МС анализе. Так двумерный гель-электрофорез был использован для регистрации изоформ альфа амилазы слюны человека. Было зарегистрировано 140 пятен, которые были идентифицированы на МС как альфа амилаза. Тщательный анализ показал наличие нескольких групп изоформ со специфическими характеристиками последовательности, которые определяли биологические активности белка [84]. Технология 2DE также подходит для профилирования фосфопротеома. Было показано изменение профиля фосфорилирования митохондриальных белков сердца в результате активности протеинкиназы. Среди 46 субъединиц комплекса митохондриальной дыхательной цепи НАДН: убихинооксидоредуктаза (комплекс I) субъединица массой 42 кДа фосфорилируется в стационарном состоянии. Однако воздействие на митохондрии цАМФ-зависимой протеинкиназы увеличивает фосфорилирование этой субъединицы 42 кДа и также приводит к фосфорилированию *de novo* субъединицы 18 кДа [85].

Фракционирование белков является надежным и широко используемым методом для регистрации белков протеома человека. Однако при дальнейшем анализе с помощью LC-ESI-MS для идентификации максимального числа пептидов более целесообразно фракционировать гидролизат белков для максимального обогащения фракций пептидами. Один и тот же белок может присутствовать в клетке в разных состояниях. Например, в случае модификации белка может измениться его электрофоретическая подвижность, а модифицированная форма может оказаться в другой полосе геля. Это усложняет детекцию пептидов этого белка, так как он

разделен на разные фракции, а соответственно и пептиды и эффективность регистрации с помощью МС в таком случае падает.

2.4.3 Хроматографические методы фракционирования

Ионообменная хроматография (ИЕС)

Методика ИЕС позволяет разделять пептиды и белки в соответствии с их электрическим зарядом [86]. Механизм удерживания на колонке заключается в электростатическом притяжении между аналитами и функциональными группами стационарной фазы, которые имеют противоположные заряды [87]. В катионообменной хроматографии отрицательные функциональные группы притягивают положительно заряженные аналиты в кислом pH среды, тогда как в анионообменной хроматографии положительные функциональные группы имеют сродство к отрицательно заряженным аналитам в щелочном pH. Каждый из этих методов можно разделить на два подтипа: сильные (катионный [SCX] и анионный [SAX] обменники) и слабые (катионный [WCX] и анионный [WAX] обменники), в зависимости от типа связанной функциональной группы стационарной фазы. Сильные группы ионизируются при всех рабочих значениях pH, в то время как слабые группы могут быть как заряженными, так и незаряженными в зависимости от значения pH подвижной фазы [88]. Функциональные группы чаще всего ковалентно связаны с диоксидом кремния и синтетическими полимерными смолами, хотя иногда используются некоторые неорганические материалы. Основным недостатком использования диоксида кремния является его нестабильность при значениях pH выше 8 и ниже 2. В ионообменной хроматографии для элюции пептидов используется градиент мобильной фазы с постепенно увеличивающейся ионной силой. Наиболее часто в протеомике используется вариант ИЕС SCX с градиентом соли. При увеличении концентрации соли в эксперименте с SCX аналиты постепенно элюируют из-за увеличения ионной силы мобильной фазы и уменьшения электростатических взаимодействий аналитов и стационарной фазы.

Так при анализе протеомного состава семенников *Rattus norvegicus* (норвежская крыса) в качестве фракционирования использовали SCX и последующий анализ пептидов LC-MS/MS. Использование двух последовательных методов фракционирования позволило зарегистрировать 2556 уникальных белков и 6773 пептидов [89].

Обращенно-фазовая (RP) хроматография

Обращенно-фазовая хроматография является наиболее широко используемым методом фракционирования в протеомике, так как большинство ESI-LC-MS систем в качестве метода онлайн фракционирования пептидов используют именно эту методику. Она позволяет разделять молекулы в растворе на основе гидрофобных свойств. В случае пептидов более полярные пептиды элюируются первыми, в то время как менее полярные пептиды сильнее взаимодействуют с гидрофобными группами стационарной фазы [90]. Обращенно-фазовая хроматография широко применяется для разделения пептидов из-за совместимости с образцами, растворимыми в воде и универсальности механизма удерживания, позволяющего модифицировать механизм разделения, изменением pH подвижной фазы, органическими модификаторами или добавками [91]. Обращенно-фазовая хроматография при кислых значениях pH универсальна и проста в использовании, она чаще всего применяется для разделения пептидов в первом измерении. Чтобы увеличить ортогональность фракционирования при двух последовательных разделениях с помощью RP был изменен pH подвижной фазы, что привело к значительным улучшениям. Из-за физико-химических свойств пептидов, получаемых при ферментативном гидролизе трипсином, pH может быть мощным инструментом для изменения селективности разделения RP за счет использования высокого pH в первом измерении и низкого pH во втором [92]. Это стало возможным благодаря тому, что была разработана стационарная фаза C18, которая оставалась стабильной при высоких значениях pH. Обращенно-фазовая хроматография с высокими значениями pH с использованием триэтиамина (pH 10,6) без использования солей позволяет концентрировать фракции с помощью упаривания перед последующим разделением на обращенной фазе в кислых условиях и получать чистые образцы пептидов без необходимости дополнительных процедур очистки от солей [93].

Было показано, что использование разделения на обращенной фазе в щелочных условиях перед МС анализом позволяет зарегистрировать на 13% больше уникальных пептидов и белков бактериального протеома *Corynebacterium glutamicum*, чем при фракционировании с помощью SCX [94]. Также RP фракционирование отличается более высокой эффективностью, так как на анализ фракций после RP было потрачено всего 124 часа, а для SCX 176 часов. Таким образом, в данной работе для

фракционирования биологических образцов человека в качестве основной методики анализа был выбран подход RP-LC(RP)-ESI-MS.

Основной задачей исследований протеома является получение полного протеома биологической системы, однако пока этого не удастся достичь в полной мере. В одной из работ проведено сравнение SDS-PAGE, SCX, IEF фракционирования примере цельноклеточного лизата *E.coli* [95]. В результате масс спектрометрического анализа фракций установлено, что всего 6424 уникальных пептида было обнаружено при анализе кусочков геля после разделения с помощью SDS-PAGE, 6731 при разделении с помощью SCX, а 4221 - после разделения с помощью IEF. Интересным наблюдением является то, что 3624 уникальных пептида обнаружены как при разделении с помощью SDS-PAGE так и при фракционировании SCX. Однако 2618 пептидов зарегистрировано только при фракционировании SDS-PAGE, а 3107 уникальных пептида обнаружено только при разделении с помощью SCX. Каждый метод дополняет картину протеома *E.coli* и ни один не является универсальным для получения наиболее полной информации о биологической системе. Возможно, для получения всеобъемлющей информации о протеоме необходимо использовать не двумерное, а трехмерное фракционирование клеток, которое включает исследование субпротеомов ее органелл. Схема таких вариантов разделения показана на рисунке 3. Также, одной из проблем является МС детекция пептидов, по которым идентифицируют белки образца, так как покрытие последовательности белка редко достигает даже 50%, что затрудняет количественный анализ, регистрацию изоформ и посттрансляционных модификаций. Возможно, в будущем будет разработана технология МС анализа интактных белков комплексного биологического образца, которая позволит преодолеть эти ограничения и углубить наши представления об исследуемых биологических системах.

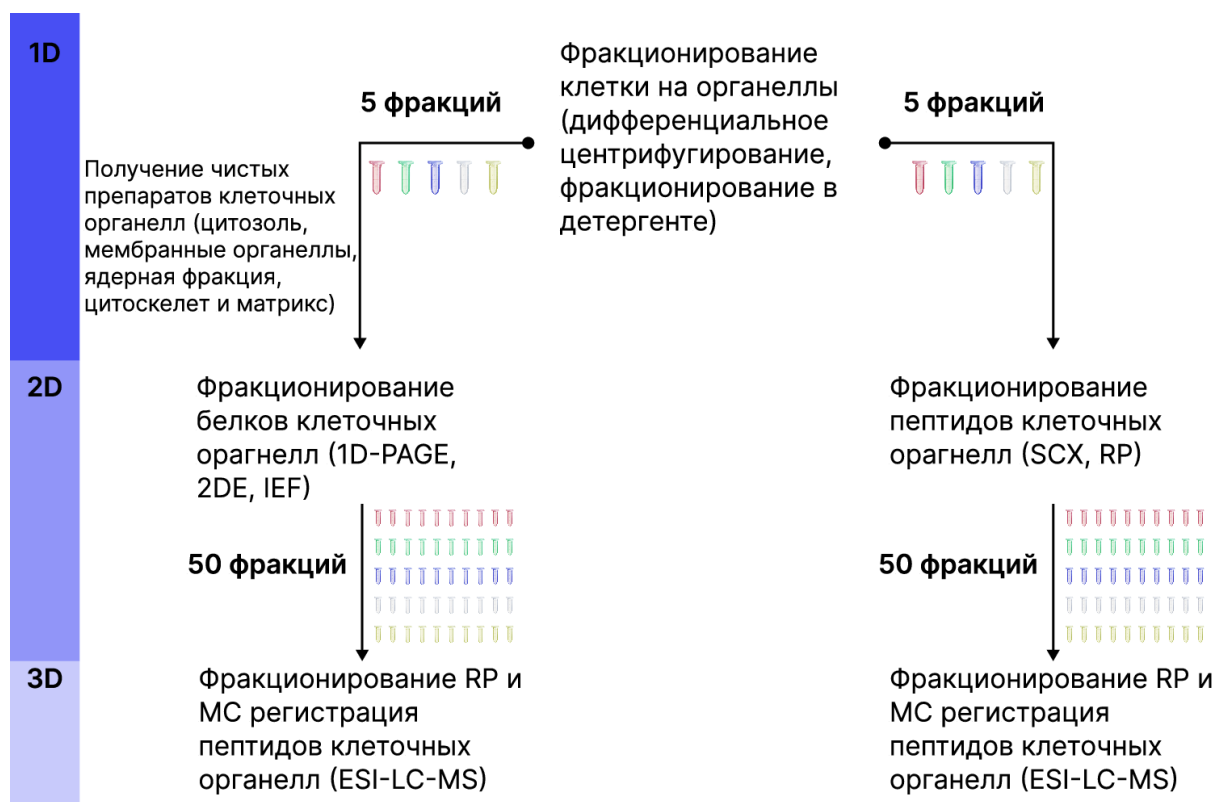


Рис. 3 Гипотететические схемы экспериментов с 3D фракционированием для получения всеобъемлющей информации о протеоме клеток.

3. МАТЕРИАЛЫ И МЕТОДЫ

3.1 Описание образцов печени человека

Образцы печени человека собраны при вскрытии от 3 доноров-мужчин (далее обозначенных как доноры №1, №3 и №5) в возрасте 65, 38 и 54 лет с одобрения Этического комитета Национального института медицины им. РНИМУ им. Н.И. Пирогова (протокол №3; 15 марта 2018 г.) с информированного согласия представителей донора. Доноры не инфицированы ВИЧ и гепатитом, а срезы не имели гистологических признаков заболеваний печени. Посмертно удаленные образцы немедленно помещали в раствор для стабилизации РНК RNAlater (Thermo Fisher Scientific, США) и хранили при -20 °С до дальнейшего использования [96].

3.2 Выделение белка и ферментативный гидролиз трипсином

Приготовление лизирующего буфера

В пробирку добавляют 50 мкл раствора 8М мочевины, 10 мкл раствора 1М натрий фосфатного буферного раствора pH 7,4, 7,5 мкл раствора 4М хлорида натрия, 10 мкл раствора 10% дезоксихолата натрия, 15 мкл раствора 100% ацетонитрила, 4 мкл 500 мМ раствора трискарбоксииэтилфосфина (Thermo Scientific, США), доводят дважды деионизованной водой до общего объема 100 мкл. В результате получен раствор 2М мочевины, 100 мМ натрий фосфатного буферного раствора pH 7,4, 300 мМ хлорида натрия, 1% дезоксихолата натрия, 15% ацетонитрила, 20 мМ трискарбоксииэтилфосфина в дважды деионизованной воде.

Выделение белка

Предварительно клетки необходимо промыть от остатков среды для инкубации с помощью 100 мМ натрий фосфатного буфера с pH раствора 7,4. Для этого добавляют 1 мл буфера в клетки и центрифугируют 10 минут при 4°C, затем отбирают супернатант и повторяют процедуру 2 раза. В пробирку, содержащую клеточный материал, вносят 20 мкл лизирующего буфера. Полученную смесь подвергают обработке ультразвуком для разрушения клеточной мембраны и ДНК. Раствор нагревают до 60°C для восстановления дисульфидных связей в молекулах белка (-S-S-) в течение 20 минут. После этого во флакон вносят алкилирующий агент в конечной концентрации 50 мМ (2-хлорацетамид) и инкубируют в отсутствии света 30 минут.

Ферментативный гидролиз трипсином.

Раствор белка разводят до конечного объема 200 мкл 100 мМ триэтиламмония бикарбонатом pH 9,0. После этого в образец вносят 1 мкг свиного трипсина (Promega, USA) и инкубируют 8 часов при температуре 38°C. Гидролиз останавливают внесением 3 мкл концентрированной муравьиной кислоты и центрифугируют при температуре 4°C и 12000g в течение 15 минут. По истечении этого времени из образцов необходимо аккуратно отобрать супернатант и перенести в чистые стеклянные вставки объемом 250 мкл (Agilent, USA). Упарить образец до минимального объема, избегая полного высушивания и перерастворить в 20 мкл 0,1% растворе муравьиной кислоты в дважды деионизованной воде.

3.3 Разделение пептидов обращенно-фазовой хроматографией в щелочных условиях

Хроматографические измерения проводятся согласно инструкции производителя при следующих условиях. Для анализа используется образец пептидов, полученный при гидролизе белка, выделенного из клеток. Разделение осуществляется на стационарной фазе (4,6 x 250 мм, 5 мкм, Waters XBridge C18, Ireland). При анализе использовалась система следующей конфигурации: хроматографическая система Agilent 1200 Series с УФ детектором, которая состоит из блока дегазатора (G1379B), насосного блока (G2226A), блока автосемплера (G4226A), коллектора фракций (G1346C), термостатируемого компартмента хроматографической колонки (G1330B), блока УФ детектора (G1316A), согласно документации узлы системы выдерживают давление до 1200 бар.

Состав хроматографической фазы «А»: 15 мМ ацетат аммония, дважды деионизованная вода, pH 9,0, с добавлением гидроксида аммония для достижения высоких значений pH.

Состав хроматографической фазы «Б»: 80% ацетонитрил, дважды деионизованная вода, pH 9,0, с добавлением гидроксида аммония для достижения высоких значений pH.

Время, (мин)	Подача хроматографической фазы «Б», (%)	Скорость потока, (мкл/мин)
--------------	--	-------------------------------

0	2	750
3	2	750
4	12	750
40	40	750
42	98	750
47	98	750
50	2	750
60	2	750

Таблица 1 Параметры градиента элюции пептидов с хроматографической колонки Waters XBridge C18.

Изменения подачи хроматографических фаз происходят в линейном режиме. Сбор фракций начинается с 3 минуты градиента и продолжается до 50 минуты. Коллектор фракций работает в режиме сбора по времени, каждые 2 минуты начинается сбор новой фракции. Фракции собираются в пустые чистые полипропиленовые пробирки, объем одной фракции 1500 мкл. Пост-аналитическое время уравнивания системы составляет 10 минут при скорости потока 750 мкл/мин. В ходе метода записываются сигналы поглощения смеси при длинах волн 210 нм и 280 нм для контроля сбора фракций при референсной длине 360 нм. По окончании анализа фракции необходимо перенести в вакуумный концентратор и упарить до минимального объема, не допуская полного высушивания, перерастворить в 20 мкл 0,1% муравьиной кислоты в дважды деионизованной воде и перенести в чистые стеклянные пробирки.

3.4 Идентификация белкового состава образца с помощью тандемной масс спектрометрии (Shotgun) с использованием детектора высокой точности (Orbitrap)

Разделение пептидов обращенно-фазовой хроматографией в кислых условиях.

Для анализа используется хроматографическая система Ultimate 3000 RSLCnano («Thermo Scientific», США), которая состоит из: нано насоса высокого давления (до 800 бар), загрузочного насоса высокого давления (до 500 бар), термостата автоматизированного инжектора, термостата хроматографической колонки с

интегрированным перепускным 10-ти канальным клапаном высокого давления (до 1000 бар) переключения потоков, автоматизированного инжектора с установленной петлей до 20 мкл и 6-ти канальным клапаном высокого давления (до 1000 бар). Обогащающая колонка Acclaim μ -Precolumn (0,5 мм х 3 мм, размер частиц 5 мкм, «Thermo Scientific», США), аналитическая колонка Acclaim Пермар® C18 (75 мкм х 150 мм, 2 мкм размер частиц, «Thermo Scientific», США). Хроматографическое разделение должно осуществляться при следующих параметрах: образец загружается на обогащающую колонку Acclaim μ -Precolumn в положение клапана 1-2 (0,5 мм х 3 мм, размер частиц 5 мкм, «Thermo Scientific», США) загрузочным насосом при скорости потока 10 мкл/мин в течение 4 минут, на всем протяжении анализа насос работает в изократическом режиме при скорости потока 10 мкл/мин с использованием хроматографической фазы «С» в качестве подвижной фазы (5% ацетонитрил, 0,1% муравьиной кислоты в деионизованной воде). Затем клапан переключается в положение 10-1 и направляет поток nano насоса через обогащающую колонку на аналитическую колонку при скорости потока 0,3 мкл/мин. Пост аналитическое время уравнивания колонки – 10 минут.

Состав хроматографической фазы «А» (0,1% раствор муравьиной кислоты в дважды деионизованной воде).

Состав хроматографической фазы «Б» (80% ацетонитрил, 0,1% раствор муравьиной кислоты в дважды деионизованной воде).

Изменения подачи хроматографических фаз происходят в линейном режиме.

Время, (мин)	Подача хроматографической фазы «Б», (%)	Скорость потока, (мкл/мин)
0	2	0,3
10	2	0,3
73	35	0,3
75	98	0,3
77	98	0,3

80	2	0,3
90	2	0,3

Таблица 2 Параметры градиента элюции пептидов с хроматографической колонки Acclaim Пермар® C18.

Для анализа использован масс спектрометр Thermo Scientific Q Exactive HF-X, оборудованный наноэлектроспрейным источником ионизации (nESI). Прибор работал в режиме положительной ионизации с доминантным зарядовым состоянием 2+, зарядовые состояния более 6+ и менее 2+ исключались из последующего анализа. Диапазон окна full-MS сканирования регистрации родительских ионов составлял 310–1500 m/z (MS1) с разрешающей способностью 120000 при m/z 200, объем накопления ионов в ловушке установлен как 1E6, время накопления ионов в ловушке составляло максимум 50 мс. Спектры дочерних ионов регистрировались в окне диапазона от 150 m/z как фиксированная первая масса, максимальная масса динамически вычислялась на основе массы и заряда родительского иона. Разрешающая способность для регистрации дочерних ионов установлена на 15000 при m/z 200, объем накопления ионов в ловушке составлял 1E5, время накопления ионов в ловушке составляло максимум 150 мс. Родительские ионы изолированы в узком окне диапазоном 2,0 m/z с отступом в 0,5 m/z для лучшего захвата изотопов. Ранее изолированные родительские ионы исключались из последующего анализа на 20 секунд после сканирования его дочерних ионов с точностью масс в 2 ppm. Максимальное количество изолированных ионов за один цикл работы прибора составляло 25 с нормализованной энергией соударения в 29 единиц. Полученные масс спектрометрические данные с расширением .raw процессированы программным обеспечением MaxQuant (v 1.6.3.4) с встроенным поисковым алгоритмом Andromeda [97]. База данных последовательностей белков полного протеома человека загружена из Uniprot (20364 последовательностей белков) с добавлением обращенных последовательностей и часто встречаемых контаминирующих последовательностей для применения подхода target decoy [98]. Метод применялся для расчета параметра FDR (False Discovery Rate), в качестве отсечения принимался параметр FDR 1% для регистрации белков. Карбамидометилирование цистеинов установлено как фиксированная модификация, N-концевое ацетилирование и окисление метионинов установлено как варибельная модификация для поиска пептидов. Точность масс для дочерних ионов составляла 20

ppm и 5 ppm для родительских ионов при обработке экспериментальных спектров с целью идентификации пептидов. Относительный количественный анализ белков, зарегистрированных при анализе, выполнен с помощью расчета параметра iBAQ (intensity Based Absolute Quantification) [99].

3.5 Выбор пептидов для регистрации белков, кодируемых генами 18 хромосомы человека

В качестве внутренних стандартов для SRM SIS анализа выбраны уникальные пептиды для каждого белка, которые встречаются в геноме только в этом белке. Пептиды выбирали по следующим критериям: длина 9-25 аминокислотных остатков, полностью специфичный гидролиз трипсином с обоих концов пептида, отсутствие пропущенных сайтов гидролиза. При отсутствии возможности выбора более длинного специфичного пептида использовались пептиды короче 9 аминокислотных остатков. Пептиды выбраны на основе биоинформатического анализа данных, полученных в ходе хромосомотцентричного проекта протеома человека, а также открытых источников данных.

3.6 Твердофазный синтез пептидов

Выбранные в качестве внутренних стандартов пептиды синтезированы с использованием методики твердофазного синтеза на синтезаторе Overture (Protein Technologies, USA) или роботизированной станции Hamilton Microlab STAR согласно ранее опубликованному методу [100]. Изотопно-меченый лейцин ($^{13}\text{C}_6$, ^{15}N), аргинин ($^{13}\text{C}_6$, $^{15}\text{N}_4$), лизин ($^{13}\text{C}_6$, $^{15}\text{N}_2$), или серин ($^{13}\text{C}_3$, $^{15}\text{N}_1$) использовались в качестве изотопной метки в пептидах при синтезе. После синтеза пептиды очищены от примесей и побочных продуктов обращенно фазовой хроматографией, а их концентрации измерены с помощью аминокислотного анализа с детекцией флуоресцентного сигнала аминокислот после кислотного гидролиза пептидов.

3.7 Таргетный протеомный анализ с использованием изотопно меченых пептидов в качестве внутренних стандартов (SRM)

Разделение пептидов обращенно-фазовой хроматографией в кислых условиях

Для разделения пептидов используется хроматографическая система Infinity 1290 («Agilent», США), которая состоит из: бинарного четырехканального насоса высокого давления с нормированием скорости потока подвижной фазы до 5 мл/мин (G4220A),

термостата автоматизированного инжектора (G1330B), термостата хроматографической колонки с интегрированным перепускным 6-ти канальным клапаном переключения потоков (до $p=1200$ бар) (G1316C), автоматизированного инжектора с установленной петлей до 20 мкл (G4226A), колонка Zorbax Eclipse Plus RRHD с геометрией $10 \times 2,1$ мм, размер частиц 1,8 мкм, размер пор 60Å. Хроматографическое разделение должно осуществляться при следующих параметрах: образец загружается на аналитическую колонку Zorbax Eclipse Plus RRHD, на всем протяжении анализа насос работает в изократическом режиме при скорости потока 300 мкл/мин с использованием смеси хроматографических фаз «А» и «Б» в качестве подвижной фазы. Пост аналитическое время уравнивания колонки - 10 минут.

Состав хроматографической фазы «А» (0,1% раствор муравьиной кислоты, 0,01% раствор трифторуксусной кислоты в дважды деионизованной воде).

Состав хроматографической фазы «Б» (80% ацетонитрил, 0,1% раствор муравьиной кислоты, 0,01% раствор трифторуксусной кислоты в дважды деионизованной воде).

Изменения подачи хроматографических фаз происходят в линейном режиме.

Время, (мин)	Подача хроматографической фазы «Б», (%)	Скорость потока, (мкл/мин)
2	11	300
50	40	300
52	67	300
53	97	300
58	97	300
60	3	300

Таблица 3 Параметры градиента элюции пептидов с хроматографической колонки Zorbax Eclipse Plus RRHD.

Для анализа использовался масс спектрометр 6495 Triple Quad (Agilent, «США»), оборудованный источником ионизации ESI Jet Stream. Температура осушающего газа в источнике (азот) – 160°C; скорость потока осушающего газа – 14 Л/мин; температура фокусирующего газа в источнике (азот) – 200°C; - скорость потока фокусирующего газа – 11 Л/мин; напряжение на капилляре – 3600 В; фокусирующее напряжение – 750В; давление на распылителе – 1,237 атмосфер; потенциал на ионной воронке низкой компрессии – 60В; потенциал на ионной воронке высокой компрессии – 150В; напряжение на усилителе электронного умножителя детектора – $\Delta EMV = +470V$; изоляция прекурсорных и фрагментных ионов в режиме – Unit; ширина изоляции прекурсорных и фрагментных ионов – $\Delta m/z = \pm 0,2$ amu и толерантность $\Delta m/z = \pm 0,15$ amu; детекция в режиме динамического мониторинга множественных реакций. Источник работал в режиме положительной ионизации. Прекурсорные и фрагментные ионы изолированы в первом и третьем квадруполе соответственно в узком диапазоне $\pm 0,65$ u и в рамках установленного для каждого иона времени удерживания. Энергии соударения и энергии ускорения ячейки оптимизированы для каждой транзиции в методе. Полное время цикла составляло 1500 мс с динамическим временем анализа транзиции в зависимости от количества конкурентных транзиций. Полученные данные обработаны с помощью программного обеспечения Skyline (v. 3.7.0.113117) [101]. Сигнал от эндогенного пептида засчитывался как детектированный, если разница между относительными интенсивностями трех транзиций эндогенного и изотопно меченого пептида не превышала 25%, хроматографические профили транзиций эндогенного и стандартного пептидов совпадали. Для каждого пептида, использованного в анализе, построена калибровочная кривая и определен предел детекции каждого пептида в чистом растворе. При анализе данных, в качестве зарегистрированных принимались только те сигналы эндогенных пептидов, интенсивность которых была выше минимального сигнала (предела детекции) для данного пептида. Если сигнал эндогенного пептида был ниже предела детекции, однако, воспроизводился во всех 3 повторах, такой сигнал принимался как зарегистрированный.

4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

4.1 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. Чтобы доказать это был поставлен эксперимент со стандартным объектом UPS1. Раствор гидролизата белков UPS1 был 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 можно зарегистрировать 100% белков образца, которые находятся в концентрации 10^{-9} М (рисунок 4).

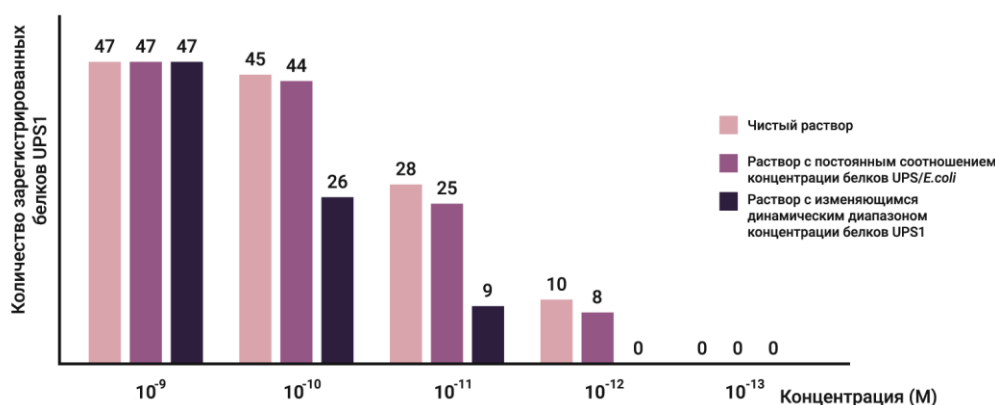


Рис. 4 Раститровка UPS1 в чистом виде, с постоянным отношением сигнал/шум, с уменьшающимся соотношением сигнал/шум.

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. Так 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 удается зарегистрировать ни одного белка смеси UPS1, а 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 5 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.

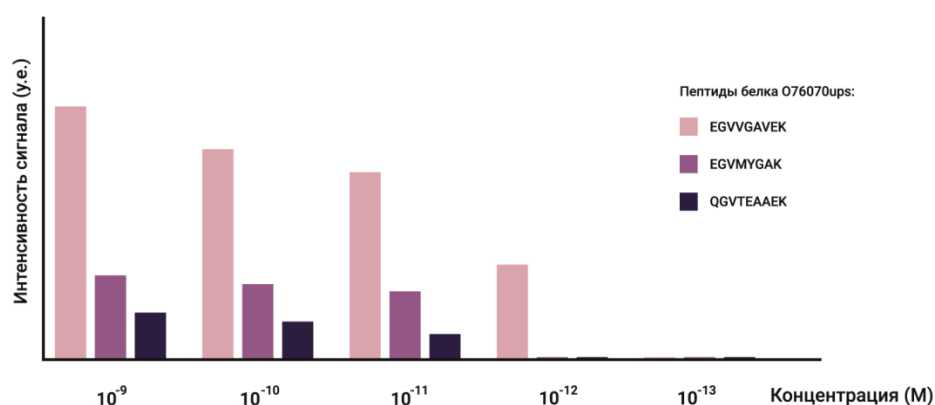


Рис. 5 Зависимость величины сигнала 3 пептидов белка O76070ups от концентрации белка.

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 провели 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 6). 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 разделения. В результате таргетного анализа каждой фракции зарегистрировано 26 белков UPS1. Схема эксперимента с фракционированием приведена на рисунке 7.

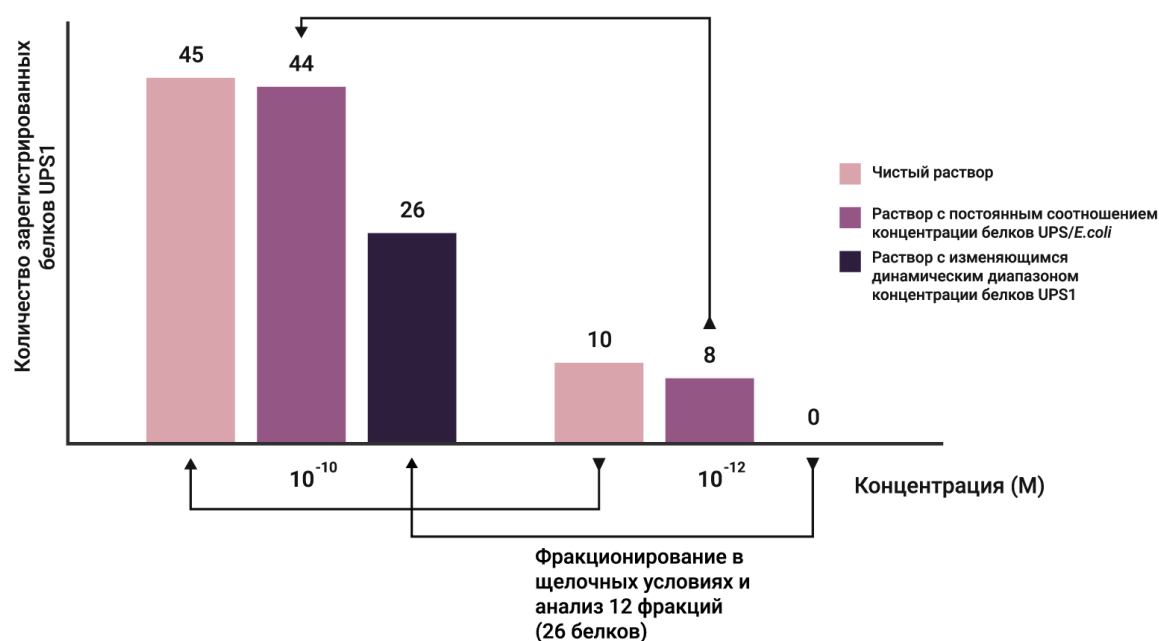


Рис. 6 Схема эксперимента с концентрированием белков смеси UPS1.

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.

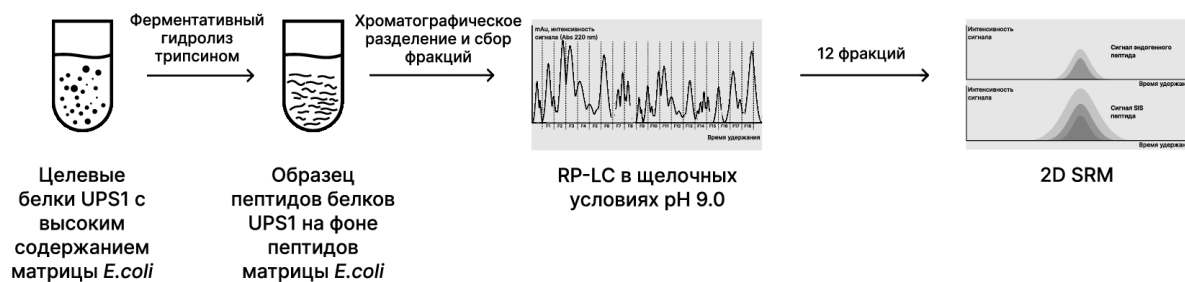


Рис. 7 Схема эксперимента с фракционированием образца белков смеси UPS1 с высоким содержанием биологической матрицы *E.coli*.

4.2 Ограничения методов протеомных технологий (Shotgun)

Методика панорамного МС анализа масс (Shotgun) принципиально отличается от методики таргетного анализа, и продемонстрировать ограниченность методики необходимо иначе. В панорамном анализе прибор будет регистрировать те пептидные ионы, которые прошли отбор по интенсивности. Те ионы, которые превысили порог отсечения по интенсивности, будут направлены на фрагментацию и регистрацию дочерних ионов. Однако те ионы, интенсивность которых не превышает выставленный порог отсечения так и не будут фрагментированы, таким образом им не будут присвоены пептидные последовательности из базы данных в ходе работы поискового алгоритма. Тем не менее, такие ионы, которые попали в поле зрения масс спектрометра, с большой долей вероятности являются именно пептидными ионами. Об этом можно судить исходя из того, что эти ионы удерживаются на хроматографической колонке C18, а также, если их зарядовое состояние равно 2+ и выше. С помощью зарядового состояния можно отсечь большую часть низкомолекулярных химических соединений, некоторые из которых могут удерживаться на хроматографической колонке C18, но чаще всего имеют зарядовое состояние +1. Пептиды, которые получены при гидролизе трипсином, имеют как минимум 2 группы, способные нести протон – N-концевая аминогруппа и аминогруппа лизина или гуанидиновая группа аспарагина. Также, чтобы исключить ошибку при подсчете незарегистрированных ионов проверяется наличие природного изотопного распределения иона и образование хорошо различимого хроматографического пика. Таким образом, для обозначения проблемы ограниченности методики панорамного МС анализа необходимо подсчитать количество зарегистрированных и не зарегистрированных пептидных ионов и их соотношение при анализе комплексного

биологического образца. Для демонстрации был выбран модельный объект – клетки гепатоцеллюлярной карциномы печени человека HepG2. На основе не обработанных масс спектрометрических данных было посчитано количество уникальных изотопных кластеров с зарядовым состоянием 2+ и более. Всего было обнаружено 79025 изотопных кластеров, но пептидная последовательность присвоена только 19153 из них (рисунок 8).

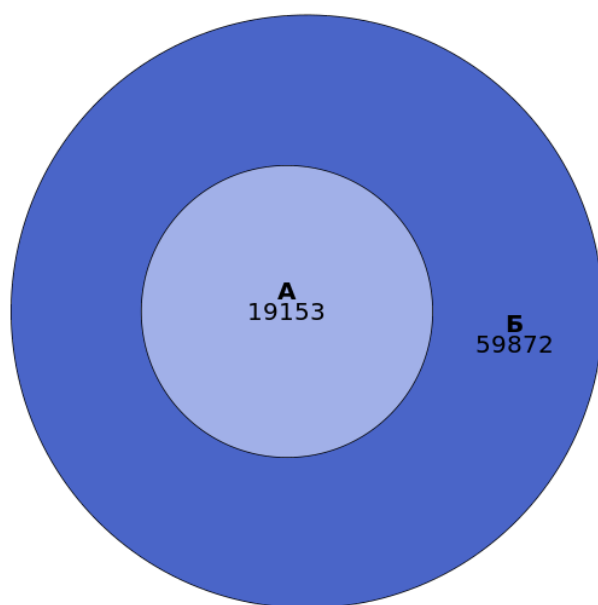


Рис. 8 Диаграмма Эйлера-Венна, подмножество А показывает количество изотопных кластеров - 19153, которым присвоена пептидная последовательность, подмножество Б показывает количество изотопных кластеров - 59872, которым не присвоена пептидная последовательность, всего зарегистрировано 79025 изотопных кластеров (подмножество А полностью входит в подмножество Б)

Прибор регистрирует всего ~25% пептидных ионов, которые попадают в его поле зрения. Наиболее простой выход из ситуации заключается в снижении порога отсечения по интенсивности иона. В результате низкоинтенсивные ионы будут долго накапливаться, а спектр фрагментации дочерних ионов получится малоинформативным. Такой спектр с высокой долей вероятности будет ошибочно сопоставлен с теоретическим спектром из базы данных или не сопоставлен вообще. В итоге происходит снижение общего количества идентифицированных изотопных кластеров из-за дефицита времени при МС анализе.

Таким образом, эти ограничения не позволяют эффективно зарегистрировать низкоинтенсивные пептидные ионы комплексного биологического образца. Следовательно, необходимо разработать методику, которая позволит обогащать

образец низкоинтенсивными пептидным ионами до такой степени, что их можно будет эффективно детектировать.

4.3 Разработка и оценка эффективности методики 2D Shotgun на модельном объекте HepG2

Для преодоления вышеописанных ограничений была разработана методика, которая включает в себя дополнительную стадию фракционирования пептидов – обращенно фазовую хроматографию в щелочных условиях. Такой же вариант хроматографии используется при масс спектрометрическом анализе, но в кислых условиях для генерации положительно заряженных ионов. Механизм фракционирования в RP не отличается в кислых и щелочных условиях. Однако при модификации pH подвижной фазы изменяются физико-химические свойства пептидов, и поэтому они по-разному взаимодействуют с колонкой. Следовательно, достигается ортогональность двух последовательных фракционирований RP-RP. Таким образом, последовательное фракционирование сначала в щелочных условиях, а затем в кислых условиях при МС анализе является эффективным и должно способствовать обогащению низкоинтенсивными пептидами и увеличению количества identifications. Схема такого эксперимента приведена на рисунке 9.

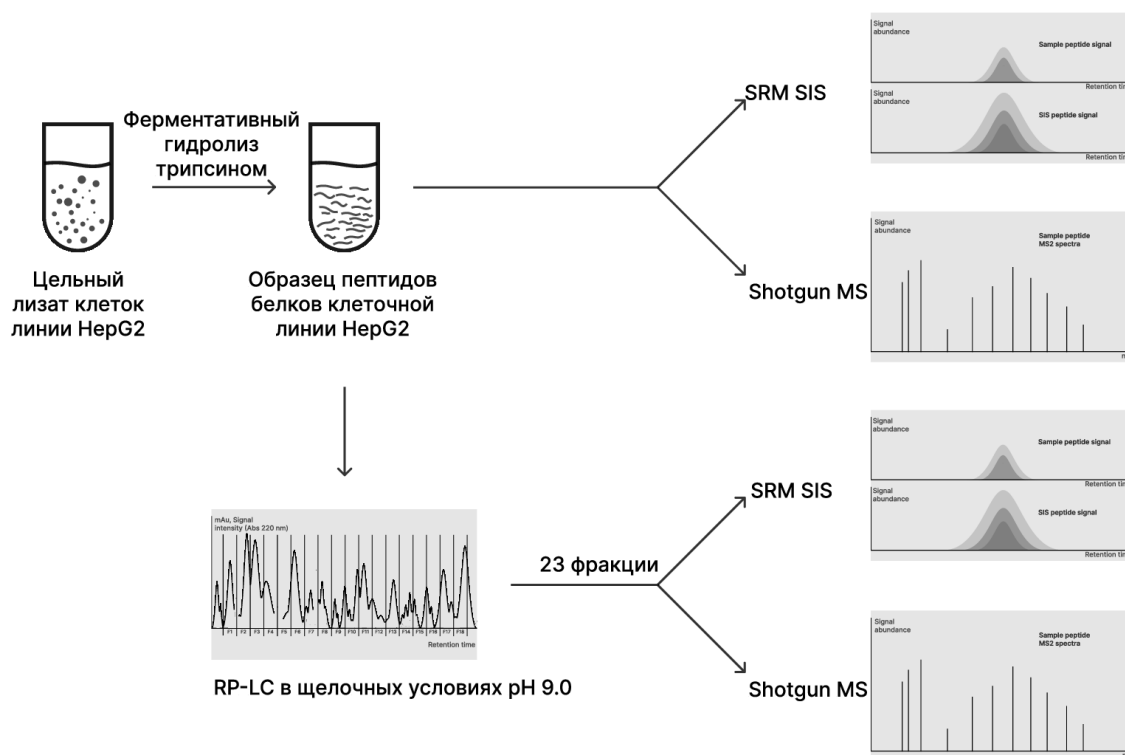


Рис. 9 Схема эксперимента по разработке методики анализа с двумерным фракционированием

На первом этапе необходимо было оптимизировать градиент при фракционировании в щелочных условиях таким образом, чтобы в каждой фракции регистрировалось одинаковое количество пептидов. Было протестировано несколько вариантов градиентов элюции, и в итоге достигли поставленной цели. Основное хроматографическое разделение пептидов при высоких значениях pH происходит с 4 по 19 фракцию. В этих фракциях наблюдается высокий процент идентифицированных МС/МС сканов и процент зарегистрированных изотопных кластеров (Рис. 10).

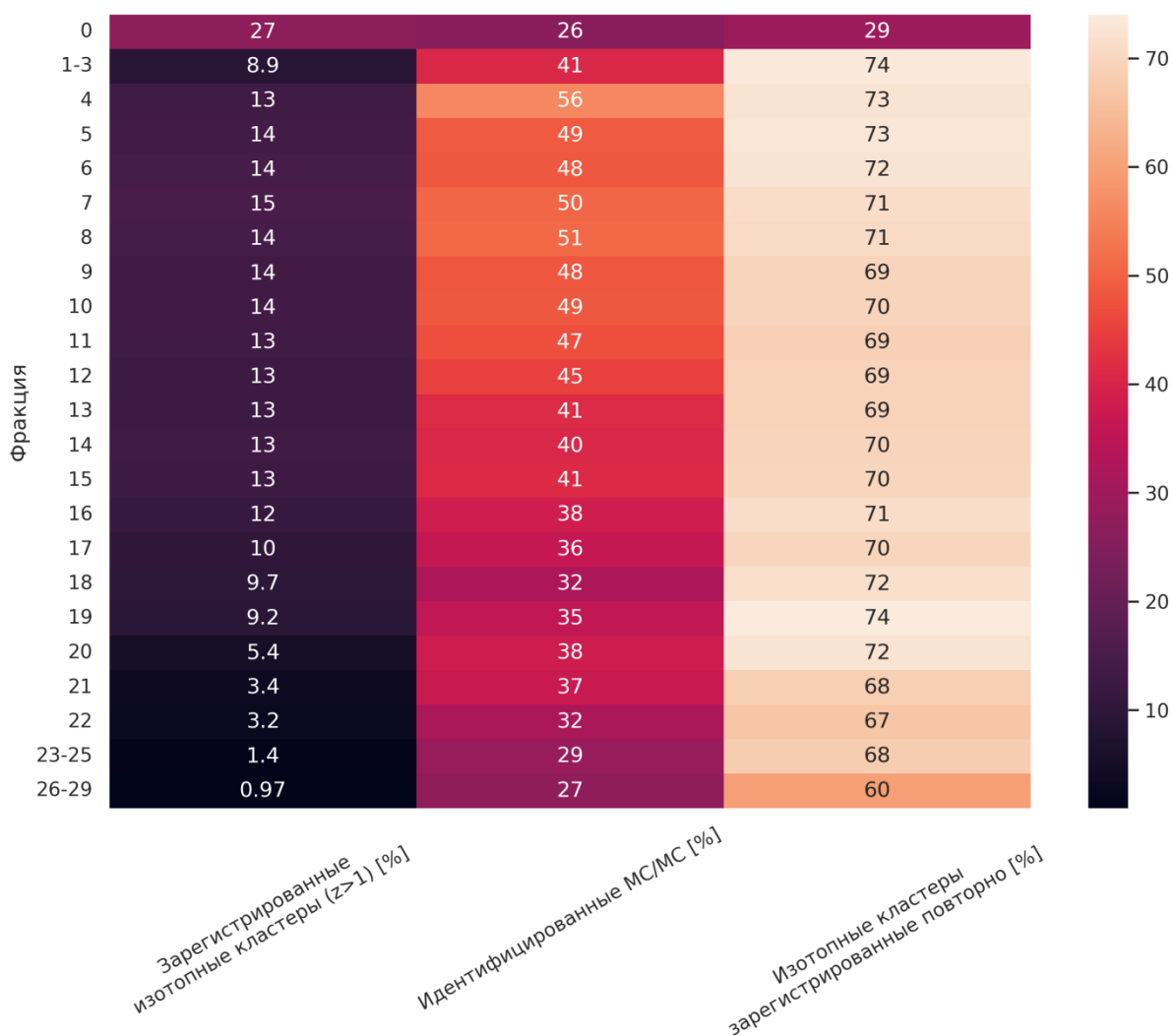


Рис. 10 Распределение количества идентифицированных МС/МС и зарегистрированных изотопных кластеров по фракциям в процентах, собранных после обращенно фазовой хроматографии в щелочных условиях, где 0 соответствует образцу цельного лизата.

Фракции, которые элюируют в начале и конце градиента (1-3 и 20-29), были обеднены пептидами. Однако их удобно использовать в качестве контроля. Таким образом, низкое количество зарегистрированных пептидов в этих фракциях отражает начало и конец элюции пептидов с колонки, а соответственно границы фракционирования, что позволяет проводить контроль качества разделения пептидов в щелочных условиях разделения. Пептидный состав каждой фракции был значительно упрощен относительно исходного комплексного образца, так как при МС анализе была увеличена частота повторной фрагментации одних и тех же ионов (70%), чего не наблюдается при анализе цельных лизатов в связи с дефицитом времени (29%).

В конечном итоге во фракционированном образце удалось идентифицировать 60277 изотопных кластеров, что в 3 раза больше, чем при анализе комплексного образца HepG2. Однако всего обнаружено 857710 изотопных кластера, при среднем количестве по 40843 кластера на каждую фракцию. Это говорит о том, что методика позволяет зарегистрировать большее количество изотопных кластеров, а также выявляет еще большее количество кластеров, которые не были идентифицированы как пептиды с определенной последовательностью. По сравнению со стандартным одномерным анализом, в котором зарегистрировано 79025 изотопных кластеров, методика с применением двумерного фракционирования выявляет почти в 11 раз больше изотопных кластеров. Поэтому количество зарегистрированных изотопных кластеров в процентах в каждой фракции ниже в 2-3 раза относительно образца цельного лизата. Однако процент идентифицированных МС/МС в каждой фракции выше относительно комплексного образца, предположительно это связано с увеличением отношения сигнал/шум во фракциях, таким образом, могли быть получены более качественные спектры фрагментации родительских ионов.

В качестве сложного биологического объекта выбрана клеточная линия HepG2, которая наряду с нормальной печенью человека и сывороткой крови человека является официальным объектом исследований в программе российской части хромосомотцентричного проекта «Протеом человека». При одномерном панорамном протеомном исследовании 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, идентифицировано **32509** уникальных пептидов, которые картированы на **3312** белковых последовательностях. При анализе данных учитывались только те белки, которые имели более 2-х детектированных пептидов на последовательность белка.

Хромосома	Количество белковых идентификаций (комплексный образец)	Процентное соотношение к общему количеству белок кодирующих генов, расположенных на хромосоме (комплексный образец) (1453)	Количество белковых идентификаций (фракционированный образец)	Процентное соотношение к общему количеству белок кодирующих генов, расположенных на хромосоме (фракционированный образец) (3312)
1	142	7,2	334	16,8
2	123	9,8	253	20,1
17	103	9,1	209	18,5
19	97	7,0	207	15,0
11	84	6,7	193	15,4
12	83	8,4	189	19,1
3	77	7,4	195	18,8
5	70	8,3	155	18,3
6	68	7,0	177	18,1
16	62	7,7	154	19,2
10	58	8,2	136	19,2
14	57	8,1	117	16,7
20	57	10,9	110	21,0
7	57	6,0	148	15,6
4	51	7,0	113	15,4
9	51	6,8	126	16,9
X	42	5,3	109	13,7
8	42	6,4	104	15,8
22	36	7,7	76	16,2
15	34	6,0	85	15,0
13	19	6,1	48	15,3
18	18	6,9	33	12,6
21	13	6,0	27	12,5
Unplaced	4	0,4	6	0,6
MT	1	6,7	1	6,7

Таблица 4 Распределение идентифицированных белков по принадлежности гена, кодирующего белок к соответствующей хромосоме, а также вклад в процентное соотношение идентифицированных белков-кодирующих генов, расположенных на каждой хромосоме.

Данные, представленные в таблице 4, свидетельствуют о том, что при предварительном фракционировании пробы количество идентифицированных белков при панорамном сканировании увеличивается в 2,3 раза, что справедливо как для общего числа белковых идентификаций (**1453** в комплексном образце – **3312** в фракционированной пробе), так и для количества белковых идентификаций для каждой хромосомы. В среднем увеличение количества идентификаций по каждой хромосоме составляет $2,2 \pm 0,4$ раза. Из общего правила выделяются группа Unplaced (4 идентификации в комплексном образце, 6 – в фракционированном) – это категория белков, которая по версии базы данных Uniprot имеет неуточненную локализацию генов и низкий уровень аннотации на уровне белка, что означает, что белки этой группы редко обнаруживаются в эксперименте (всего эта группа в 2023 году насчитывает 973 белка). Также был обнаружен всего 1 белок, кодируемый митохондриальным геномом. Предположительно, это связано с тем, что эта кольцевая хромосома кодирует 15 белков, которые являются компонентами дыхательной цепи митохондрий с мембранной локализацией. Детекция мембранных белков составляет отдельную проблему в протеомике в связи с тем, что такие белки сложно солюбилизовать при пробоподготовке, поскольку в растворе они образуют агрегаты. Также в белках, кодируемых митохондриальным геномом, наблюдается небольшое количество сайтов для гидролиза трипсином. Например, в NADH-убихинон-оксидоредуктазе цепи 6 P03923 (NU6M_HUMAN) на 174 аминокислоты всего 3 аргинина и 2 лизина. Один из сайтов гидролиза заблокирован пролином, при гидролизе двух других будут получены пептиды длиной 107 и 29 аминокислот, и их масс-спектриметрическая регистрация будет затруднена ввиду большой длины и массы. При специфическом гидролизе трипсином остается 2 пептида длиной в 14 и 22 аминокислоты, и их непременно необходимо зарегистрировать оба для идентификации этого белка, что очень маловероятно.

В связи с этим можно предположить, что увеличение количества идентификаций не распространяется на мембранные белки, однако, чтобы это выяснить, необходимо провести более подробный GO анализ локализации всех обнаруженных белков.

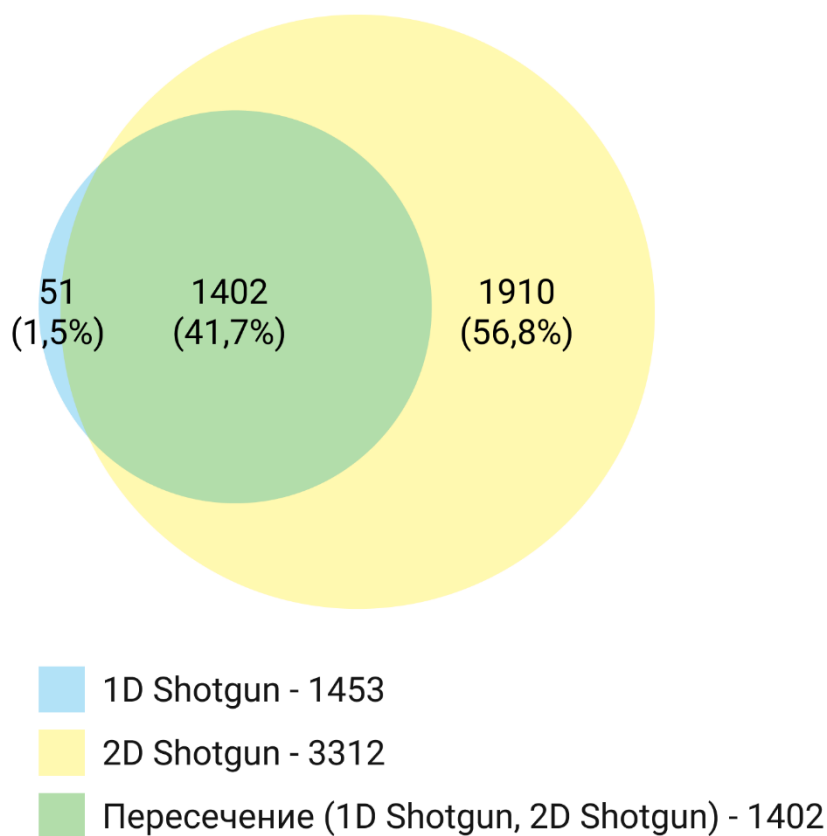


Рис. 11 Диаграмма Эйлера-Венна, сравнение множеств белков, обнаруженных при панорамном сканировании не фракционированной пробы (1D Shotgun) и фракционированной пробы (2D Shotgun).

При панорамном анализе без предварительного фракционирования (1D Shotgun) было зарегистрировано 1453 белка, в 2D Shotgun анализе – 3312 белков. При этом, общее количество зарегистрированных белков составляет 1402, а всего в клеточной линии HepG2 было зарегистрировано 3363 белка. Уникальными идентификациями для множества 1D Shotgun является 51 белок, а для 2D Shotgun - 1910. Часть белков идентифицируется исключительно в комплексном образце. Это связано с потерей части образца, что неизбежно происходит в ходе дополнительных манипуляций перед анализом, а именно, при фракционировании в щелочных условиях (рисунок 11). Однако при этом происходит обогащение фракций пептидами, и несмотря на потерю

51 белка при 2D Shotgun анализе новых уникальных белков было зарегистрировано почти в 40 раз больше (1910), чем потеряно.

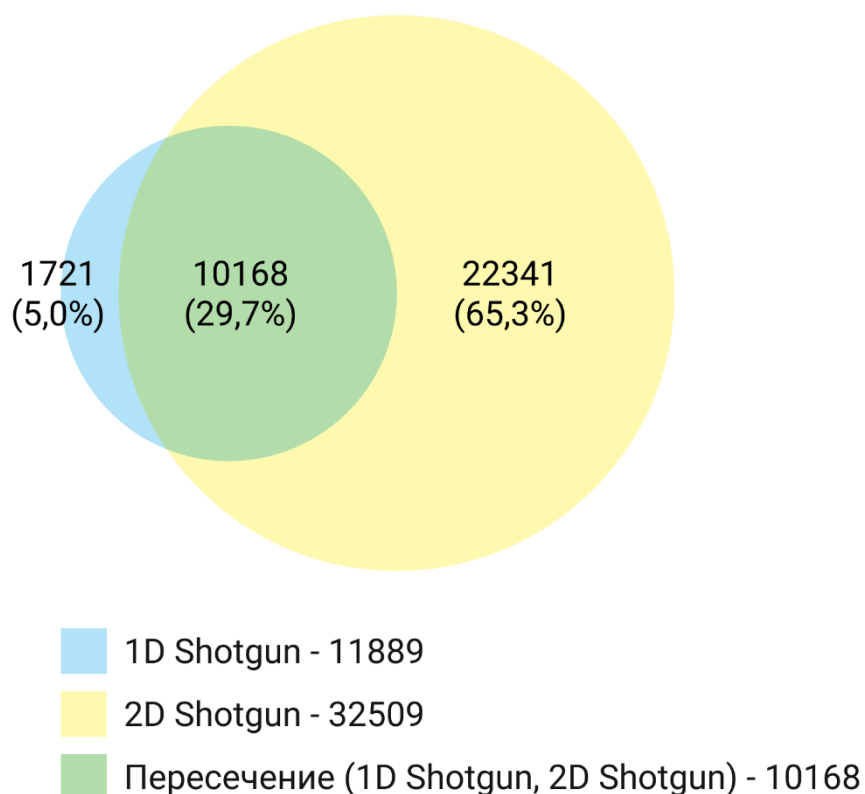


Рис. 12 Диаграмма Эйлера-Венна, сравнение множеств уникальных пептидов, обнаруженных при панорамном сканировании не фракционированной пробы (1D Shotgun) и фракционированной пробы (2D Shotgun).

Часть пептидов идентифицируется исключительно в не фракционированном образце (1D Shotgun, 1721 уникальных пептида), однако для комплексного образца, как показано на рисунке 11, уникальными являются всего лишь 51 белковых идентификаций. Для одних и тех же белков в исходном комплексном и фракционированном образце идентифицированы различные по своей аминокислотной последовательности пептиды (рисунок 12). Например, для белка P62910 комплексного образца идентифицировано 3 пептида с аминокислотной последовательностью: **ILMPNIGYGSNK**, **GQILMPNIGYGSNK**, **HMLPSGFR**, а для фракционированной пробы идентифицировано 6 пептидов, 2 из которых совпадают, а также зарегистрировано 4 новых пептида с иной аминокислотной последовательностью: **KFLVHNVK**, **HMLPSGFR**, **AALRPLVKPK**, **LRPLVKPK**, **GQILMPNIGYGSNK**,

LLTGISAQQNR. Это указывает на значительные потери пептида **ILMPNIGYGSNK** в ходе дополнительных манипуляций с образцом в случае использования предварительного фракционирования в щелочных условиях. Общий пул пептидов для фракционированного и комплексного образца составил 10168 пептидов. Таким образом, благодаря предварительному фракционированию удалось идентифицировать дополнительно 22341 уникальных пептидов.

В теории при хроматографическом разделении в щелочных условиях должно произойти обогащение всех пептидов и значительных потерь быть не должно. Однако полученные данные свидетельствуют о том, что 1721 пептид зарегистрирован только в комплексном образце. Это требует дополнительного анализа, так как существует вероятность, что при фракционировании часть градиента, где еще элюировали пептиды была искусственно обрезана временными границами сбора фракций. Времена удерживания каждого конкретного пептида при фракционировании в щелочных условиях установить невозможно, так как ультрафиолетовый детектор не дает такого разрешения и чувствительности, однако можно проанализировать времена удерживания пептидов, которые регистрируются только в 1D Shotgun, общей фракции пептидов и пептидов, которые регистрируются только в 2D Shotgun.

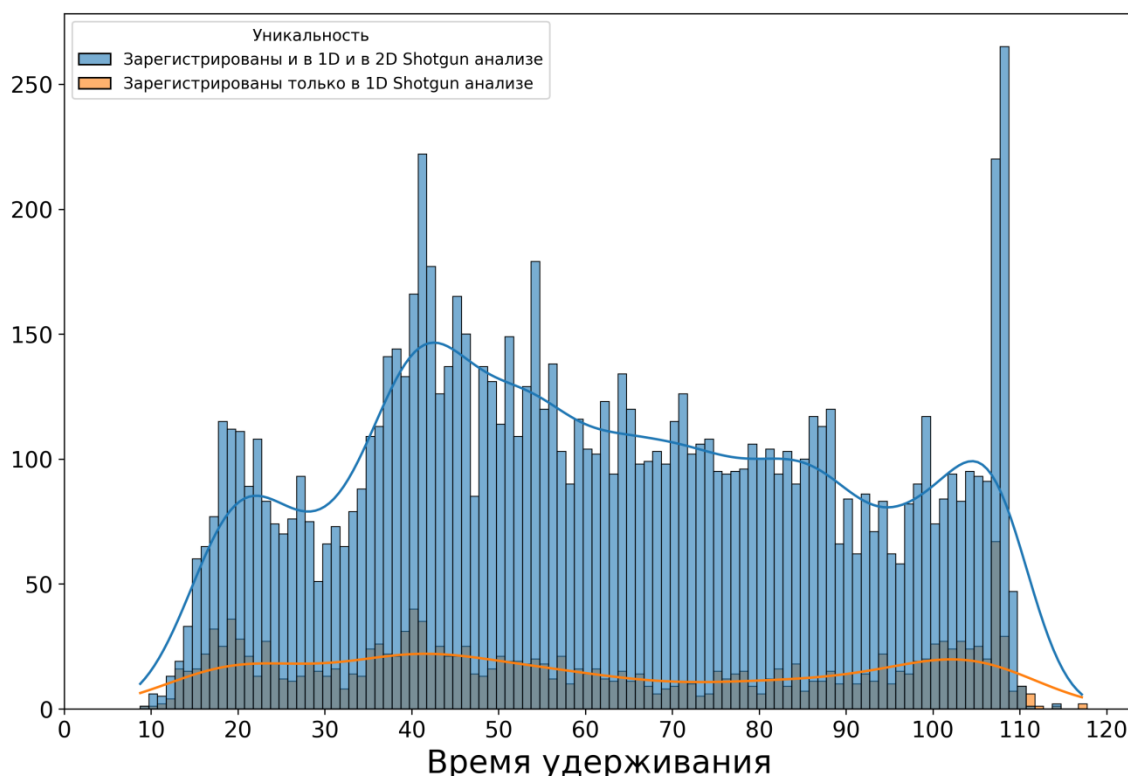


Рис. 13 Гистограмма распределения времени удерживания пептидов, которые зарегистрированы и в 2D Shotgun и в 1D Shotgun, а также пептидов, зарегистрированных только в 1D Shotgun.

На рисунке 13 показано, что время удерживания общих для 1D Shotgun и 2D Shotgun и уникальных для 1D Shotgun пептидов распределяется равномерно по всей длине хроматографического градиента. В общих чертах форма распределения уникальных для 1D Shotgun пептидов повторяет форму распределения общих пептидов. Тот факт, что распределение уникальных для 1D Shotgun пептидов равномерно, позволяет предположить, что градиент не обрезан искусственно при разработке методики фракционирования в щелочных условиях. Таким образом, потери пептидов при разделении в щелочных условиях происходят на протяжении всего хроматографического градиента. Для того чтобы это выяснить необходимо было рассмотреть распределение времен удерживания пептидов, которые были зарегистрированы обоими методами одновременно и уникальные идентификации сделанные только при 2D Shotgun анализе.

На рисунке 14 показано, что распределение общих и уникальных для 2D Shotgun пептидов также равномерно и имеет одинаковую форму. Несмотря на то, что

разделение в щелочной среде должно проходить в ином направлении, все равно наблюдается перемещение верхней точки распределения от гидрофильной части (начала градиента) к гидрофобной части (концу градиента). Особенно хорошо это видно, если сравнить две фракции, отстоящие далеко друг от друга. Например, во фракции 1-3 максимальное количество идентифицированных пептидов наблюдается на 20 минуте. В следующих фракциях оно постепенно смещается к центру градиента, а начиная с фракции 17 находится в более гидрофобной области на 50 минуте. Общие пептиды и уникальные пептиды для 2D Shotgun присутствуют в каждой фракции. Это означает, что потери действительно происходят на протяжении всего хроматографического разделения пептидов в щелочных условиях. Причиной этого может быть несовершенство системы сбора фракций. Коллектор переходит к сбору новой фракции каждые 2 минуты метода. В этих точках переключения происходит элюция большого числа пептидов одновременно, как и на протяжении всего хроматографического градиента. Таким образом, часть хроматографического пика пептида будет собрана в одну фракцию, а часть в другую. При этом фракция не обогащается этим пептидом, а наоборот происходит его размытие. Степень обогащения пептидов можно подсчитать, сравнив нормализованные интенсивности общих пептидов для 1D Shotgun и 2D Shotgun анализа. В среднем интенсивность пептидных ионов выше в 22,9 раз в 2D Shotgun, при этом, у 10128 пептидов интенсивность увеличилась по сравнению с 1D Shotgun, а всего у 40 уменьшилась.

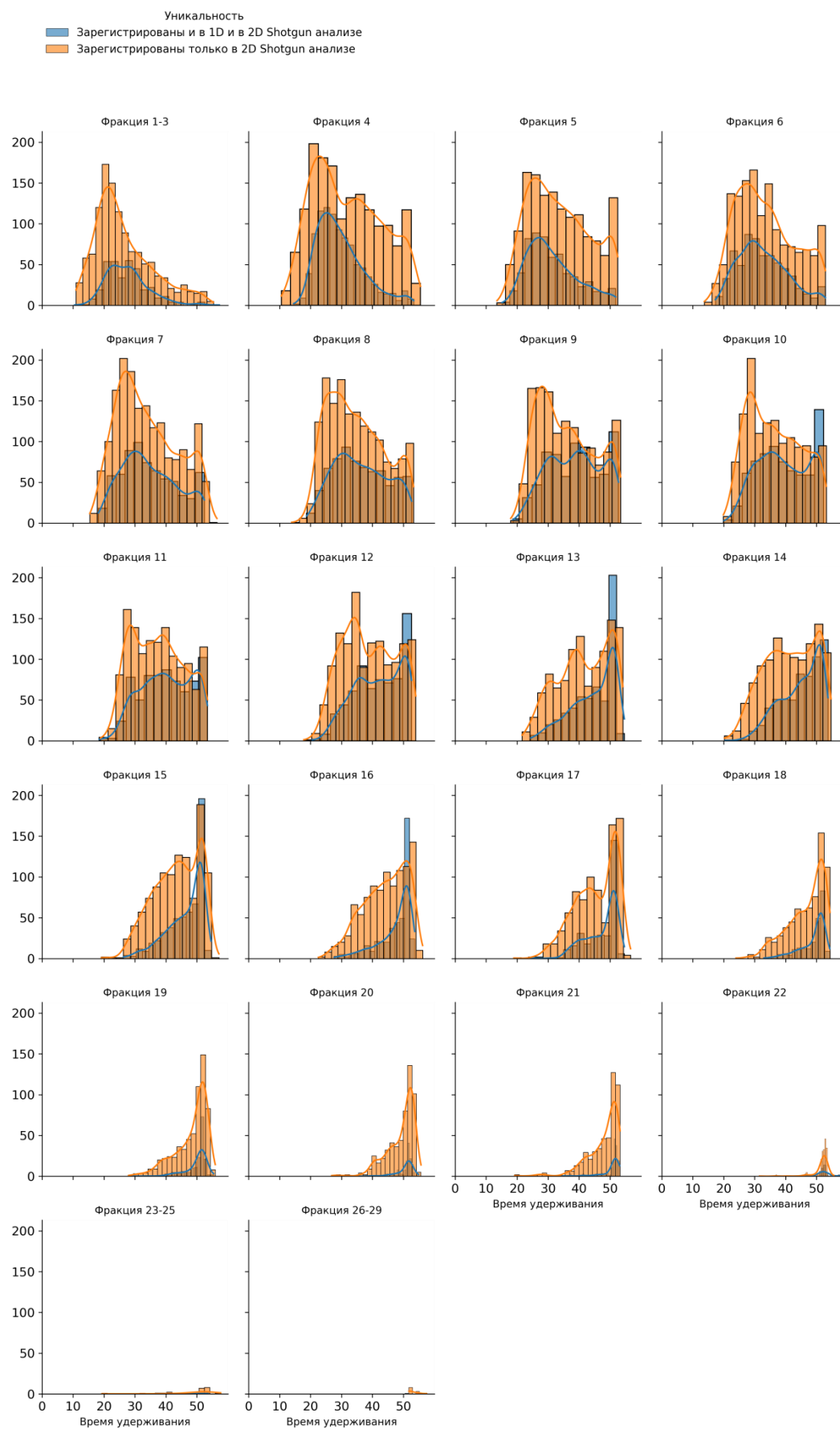


Рис. 14 Распределение пептидов, обнаруженных во фракциях после хроматографического разделения в щелочных условиях. Распределение пептидов уникальных для 2D Shotgun (оранжевые), а также пептидов, которые обнаружены и в 1D Shotgun и в 2D Shotgun (синие).

При 1D Shotgun анализе обнаружено 18 белков, кодируемых генами 18 хромосомы, при применении метода 2D Shotgun – 33. Все белки, зарегистрированные в комплексном образце, регистрируются также и во фракционированном, поэтому общая сумма зарегистрированных белков – 33. При стандартном панорамном протеомном анализе клеточных линий человека обычно детектируется от 20 до 30 белков, кодируемых генами 18 хромосомы человека.

4.4 GO анализ белков, идентифицированных при Shotgun 1D и Shotgun 2D анализе по клеточной локализации белков

При оценке новой технологии анализа проведен сравнительный анализ GO категорий зарегистрированных белков. Цель данного анализа заключалась в сравнении количества зарегистрированных белков, принадлежащих одинаковым категориям GO при 1D и 2D анализе и выяснить, присутствует ли обогащение белками в определенной категории или нет.

Всего было обнаружено 913 и 1200 уникальных GO категорий в белках, зарегистрированных в 1D Shotgun и 2D Shotgun анализе, соответственно. Однако некоторые GO относятся к очень узкоспецифичным группам, в которых присутствует несколько или всего один белок протеома человека. Например, только один белок Q12797 относится к группе GO:0032541: cortical endoplasmic reticulum, что означает одну из зарегистрированных клеточных локализаций белка. При анализе GO идентификаторов обнаружено, что большинство полученных групп включают в себя не более 6 и 8 белков для 1D Shotgun и 2D Shotgun соответственно (75% всех GO идентификаторов). Для более точного анализа обогащения отобраны только наиболее многочисленные группы, начиная от 20 белков для одного GO идентификатора. При фильтрации получено всего 85 общих групп и для 1D Shotgun и для 2D Shotgun. Был посчитан средний коэффициент обогащения для данных групп, который составил $1,94 \pm 0,45$ с минимальным значением в выборке 1,04, а максимальным 3,03. Эти данные позволяют сделать вывод о том, что основные группы белков обогащаются равномерно и обогащения специфических групп белков не происходит. Наиболее обширная группа белков-цитозольные белки. Это объясняется тем, что многие метаболические процессы

происходят в цитозоле, например, гликолиз. Однако также и процессы передачи сигнала часто начинаются в цитозоле, результатом чего является активированный транскрипционный фактор, который транслоцируется в ядро (P17252), например РКС участвует в каскаде фосфорилирования и активации транскрипционного фактора NF-κB. Следующей группой являются белки с ядерной локализацией. В ядре находится большое количество гистоновых белков (Q71UI9) и ферментов, которые их модифицируют (O14929), а также ядерные рибонуклеопротеины (P09661), а также транскрипционные факторы (P17480). Кроме того в наиболее обширные группы белков также входят мембранные и митохондриальные белки.

Таким образом, можно говорить о том, что в ходе 2D Shotgun анализа специфического обогащения определенных групп белков с GO категориями не происходит. Показано, что рост числа зарегистрированных белков происходит равномерно по всем группам (рисунок 15).

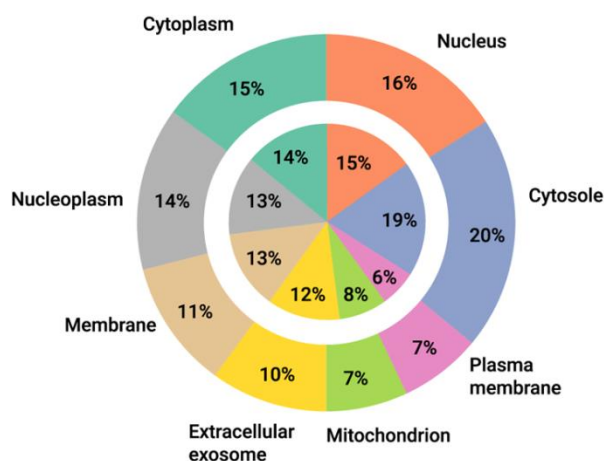


Рис. 15 Локализация и процентное содержание зарегистрированных белков в клетках линии HepG2 в 1D Shotgun (внутренний круг) и в 2D Shotgun (внешний круг).

4.5 Анализ распределения обнаруженных белков по биологическим путям по версии базы данных Reactome

При GO-аннотации происходит анализ отдельных групп белков, однако более интересным подходом является переход от набора зарегистрированных белков к конкретным метаболическим процессам в клетке, что позволяет сделать использование базы данных Reactome [102]. Обнаруженные белки в 1D Shotgun и 2D Shotgun были

картированы на метаболические пути и было проведено сравнение, насколько больше белков зарегистрировано в каждом метаболическом пути. Всего для 1D Shotgun анализа обнаружено 1624 метаболических пути, а для 2D Shotgun анализа - 1931.

Анализ распределения количества обнаруженных белков в метаболических путях показал, что в 75% случаев для 1D анализа найдено не более 4 белков. Поэтому данное значение было использовано как отсечение по минимальному количеству белков в метаболическом пути. После фильтрации было получено всего 810 общих групп и для 1D Shotgun и для 2D Shotgun. Был посчитан средний коэффициент обогащения для метаболических путей, который составил $1,94 \pm 0,76$ с минимальным значением в выборке 0,4, а максимальным 6,60. Эти данные позволяют сделать вывод о том, что основные метаболические пути обогащаются равномерно и обогащения специфических путей по количеству зарегистрированных белков не происходит. Также проанализирована доля покрытия метаболического пути, которая в 1D Shotgun анализе в среднем составляет $0,26 \pm 0,18$, а в 2D Shotgun в среднем составляет $0,43 \pm 0,21$ (рисунок 16). Наиболее многочисленными по составу пути относятся к общему метаболизму, метаболизму белков, проведению сигнала, метаболизму РНК, то есть к основным процессам жизнедеятельности клетки. На таком более высоком уровне организации как метаболический путь сохраняется тенденция повышения степени покрытия путей, обнаруженных при 2D анализе по сравнению с 1D. Также при 2D анализе выявлено 54 новых уникальных пути, которые в 1D Shotgun не обнаружены. Средняя степень покрытия новых путей, обнаруженных в 2D анализе, составляет $0,34 \pm 0,18$, однако, они также относятся к общему клеточному метаболизму и процессам транскрипции.

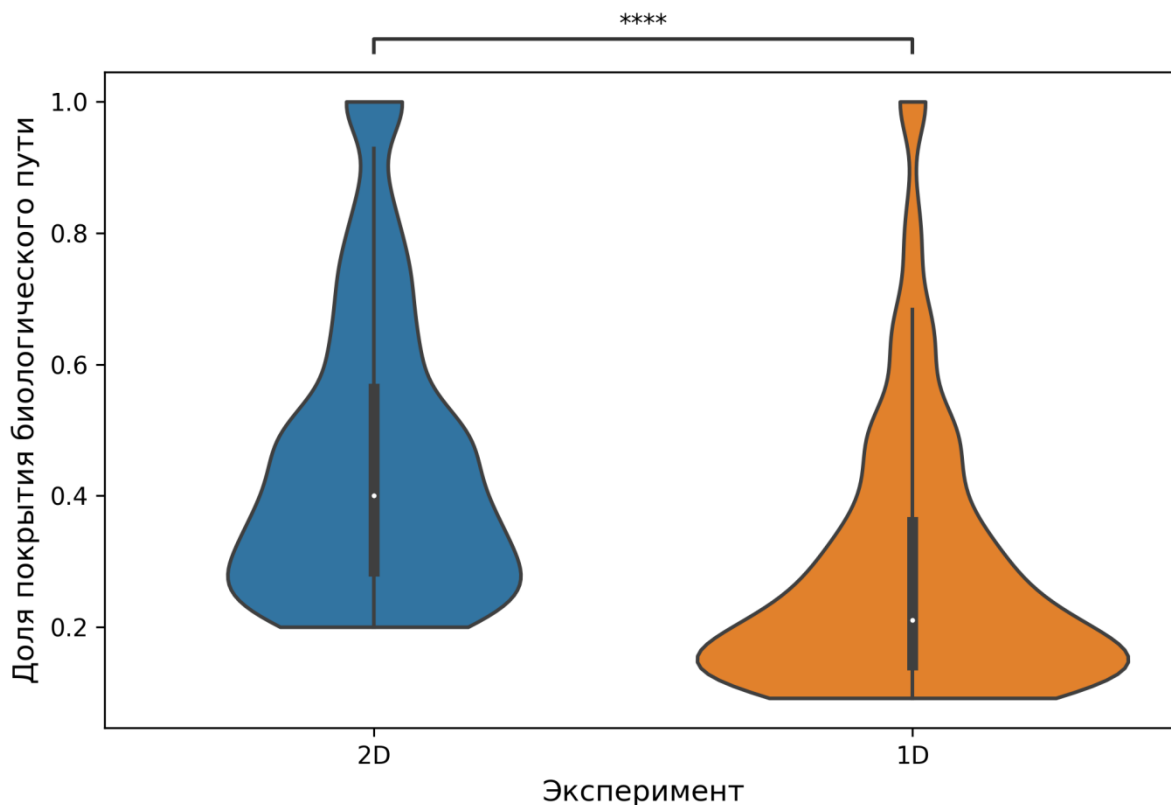


Рис. 16 Распределение доли покрытия биологических путей, зарегистрированных в 1D и 2D анализе (Reactome). Звездами показан результат t-теста с р-уровнем значимости $7,6e^{-90}$. При использовании 2D анализа наблюдается значимое увеличение доли покрытия белков метаболического пути. ****-р-уровень значимости $\leq 0,00001$.

Проведенный анализ показывает, что в наиболее многочисленных категориях, в которые попадает не менее 20 белков, увеличение количества обнаруженных белков происходит равномерно (рисунок 17). Таким образом, применение технологии двумерного фракционирования совместно с Shotgun анализом дает возможность глубже оценить и проанализировать протеом клеток линии HepG2. Полученные результаты могут быть использованы при планировании эксперимента. То есть, на основе 1D Shotgun анализа (GO-аннотации, анализа метаболических путей), можно сделать прогноз о том какие группы белков будут обогащаться при 2D Shotgun анализе, и принять решение о целесообразности проведения эксперимента с двумерным фракционированием.

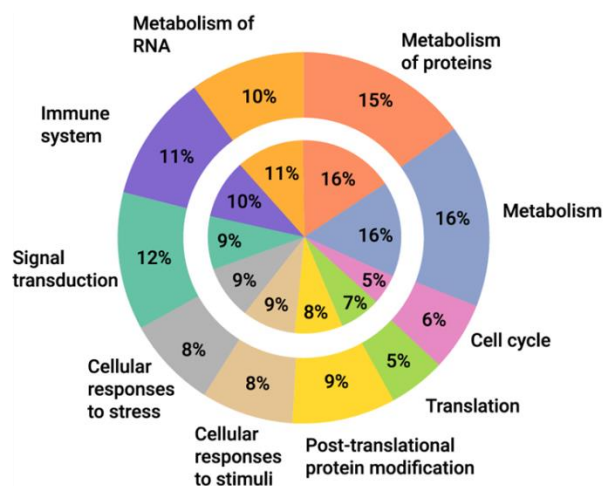


Рис. 17 Функции и процентное содержание зарегистрированных белков в клетках линии HepG2 в 1D Shotgun (внутренний круг) и в 2D Shotgun (внешний круг).

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.

Стандартный вариант подготовки к таргетному анализу – детекция максимального количества белков с помощью Shotgun анализа и выбор белков-мишеней. Среди зарегистрированных уникальных пептидов, полученных в результате гидролиза трипсином, необходимо выбрать наиболее оптимальные для синтеза изотопно-меченных стандартов и разработать таргетный SRM SIS метод анализа. Это общепринятый подход, но при таком подходе нет возможности детектировать больше белков, кодируемых генами 18 хромосомы человека, чем обнаруживается при Shotgun анализе. В этой работе был использован другой подход, который уже зарекомендовал себя. Известно, что с помощью SRM анализа можно детектировать низкие концентрации белка, вплоть до 10^{-12} М, в то время как предел детекции для Shotgun анализа составляет 10^{-9} М. Поэтому был разработан таргетный метод анализа, который покрывает все 264 белка, кодируемые генами 18 хромосомы человека.

4.6 Анализ 20 клеточных линий методом панорамной масс спектрометрии для подбора пептидов с оптимальными параметрами эффективности ионизации для таргетного анализа

На рисунке 18 показан результат UMAP анализа: уменьшение размерности транскриптомных данных до двух условных переменных. На наибольшем удалении находятся 2 кластера клеточных линий: суспензионные и адгезивные в левой части графика и правой соответственно. Для регистрации максимального количества уникальных белков, кодируемых генами 18 хромосомы человека, были выбраны 20 клеточных линий не похожих друг на друга по транскриптомному профилю.

Суспензионные клеточные линии делятся на 2 группы лимфоидного и миелоидного происхождения, адгезивные в свою очередь подразделяются на группы в соответствии с тканевым происхождением. Для анализа использовали часть суспензионных и часть адгезивных клеточных линий различного тканевого происхождения: SW620, SKBR3, HL-60, Molt-4, PC-3, Daudi, CCRF-CEM, BT-474, LN-CAP, DU-145, MDA-MB-231, U937, HCT, HBL-100, huh-7, HT-29, Raji, Kasumi-1, THP-1.

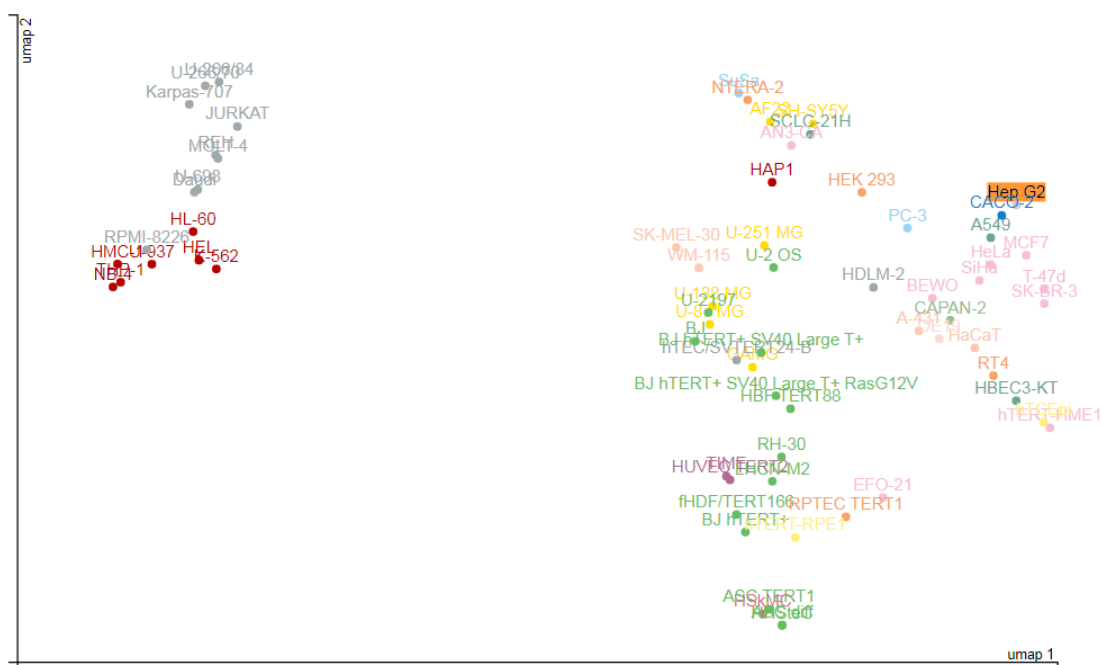


Рис. 18 Результат UMAP анализа транскриптомных данных 69 клеточных линий (HerG2 выделена оранжевым в правом верхнем углу).

Shotgun анализ выявил 53 белка, кодируемых генами 18 хромосомы человека, в 20 клеточных линиях. Всего было детектировано 399 пептидов белков, кодируемых генами 18 хромосомы человека. Они были проверены на уникальность и для каждого белка выбрано по одному протеотипическому пептиду, который имел наилучшие показатели интенсивности сигнала. Для оставшихся белков, кодируемых генами 18 хромосомы человека, протеотипические пептиды были выбраны из открытых источников протеомных данных.

4.7 Разработка и оценка эффективности методики 2D SRM на модельном объекте HerG2

В SRM анализе белков, кодируемых генами 18 хромосомы человека, было использовано 264 стандартных изотопно-меченных пептида. При анализе происходит регистрация пары пептидов: стандартного и эндогенного. Поэтому всего такой SRM анализ будет регистрировать 528 пептидов. В таком режиме работы прибор может регистрировать до 200 транзиций за один рабочий цикл – 1500 мс. Для каждого пептида при разработке метода было выбрано по 3 транзиции. Одна транзиция выбиралась как ион-квантор для количественного анализа, а 2 другие как ионы-спецификаторы, для специфичной идентификации пептида. Таким образом, всего получилось 1584 транзиции в методе.

Чувствительность методики SRM анализа напрямую зависит от времени, которое прибор тратит на регистрацию одной транзиции. На рисунке 19 в начале градиента, на 3 минуте находится пик, означающий, что в эту минуту регистрируется 120 транзиций одновременно, а значит, прибор регистрирует одну транзицию в течение 12,5 мс. В начале и середине градиента в среднем приходится по 60 транзиций на каждую минуту, что составляет 25 мс на одну транзицию. Производитель прибора заявляет, что в одно временное окно должно регистрироваться минимум транзиций, а лучше всего одна для лучшего результата и чувствительности. В нашем случае минимум составляет пара: эндогенный пептид/стандартный пептид, а значит 6 транзиций, что составляет 250 мс на транзицию. Разрешить такое количество пептидов за 60 минут хроматографического градиента не представляется возможным, однако разделение образца на фракции позволяет значительно снизить количество регистрируемых в одно время транзиций.

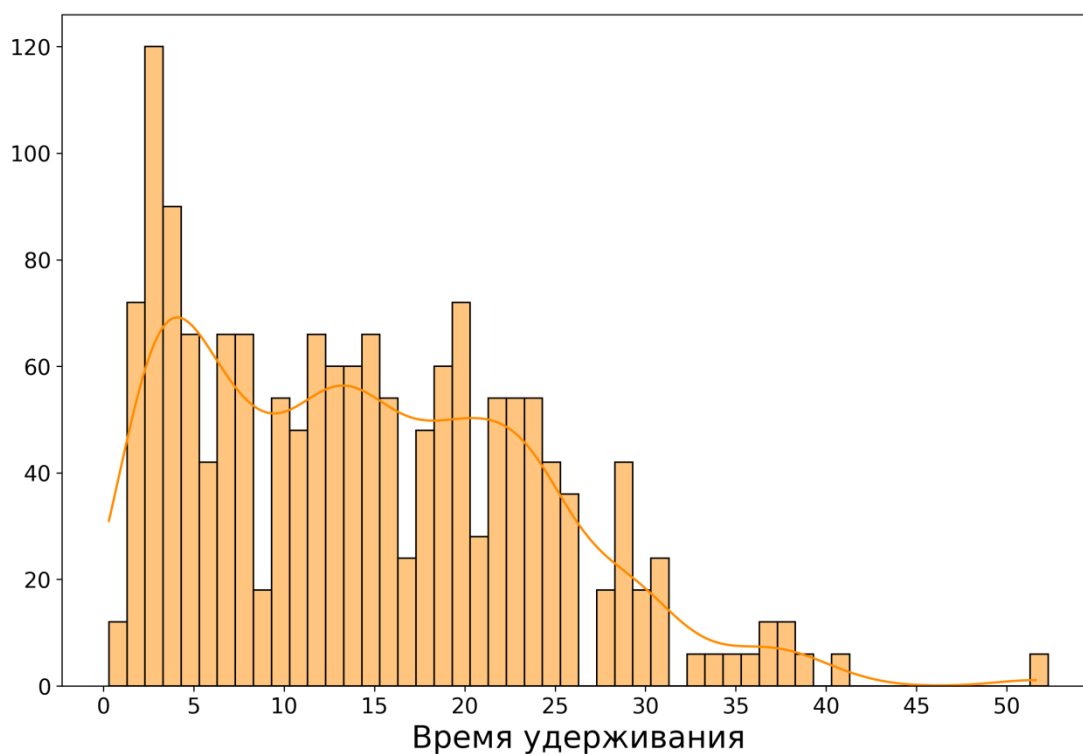


Рис. 19 Распределение регистрируемых транзиций в SRM методике по скринингу белков, кодируемых генами 18 хромосомы человека на хроматографическом градиенте.

Для разработки SRM метода, совмещенного с фракционированием в щелочных условиях, был определен пептидный состав каждой фракции после фракционирования стандартных пептидов и разработаны SRM методики под каждую конкретную фракцию. На рисунке 20 показано распределение транзиций по фракциям, из которого следует, что количество транзиций, регистрируемых в минуту, значительно уменьшилось, по сравнению с полным методом сканирования. В первых фракциях (1-3-8) наблюдается пик в начале градиента в гидрофильной области. Однако он составляет всего 60-80 транзиций, что в 2 раза ниже, чем в методе без фракционирования. А в остальных частях градиента количество регистрируемых транзиций в минуту распределяется достаточно равномерно и редко превышает 20, что в 3 раза ниже, чем в методе без фракционирования. Таким образом, разработка отдельных SRM методик для каждой фракции позволяет увеличить время на регистрацию одной транзиции в 2-3 раза, до 50-75 мс на одну транзицию при общем времени цикла 1500 мс.

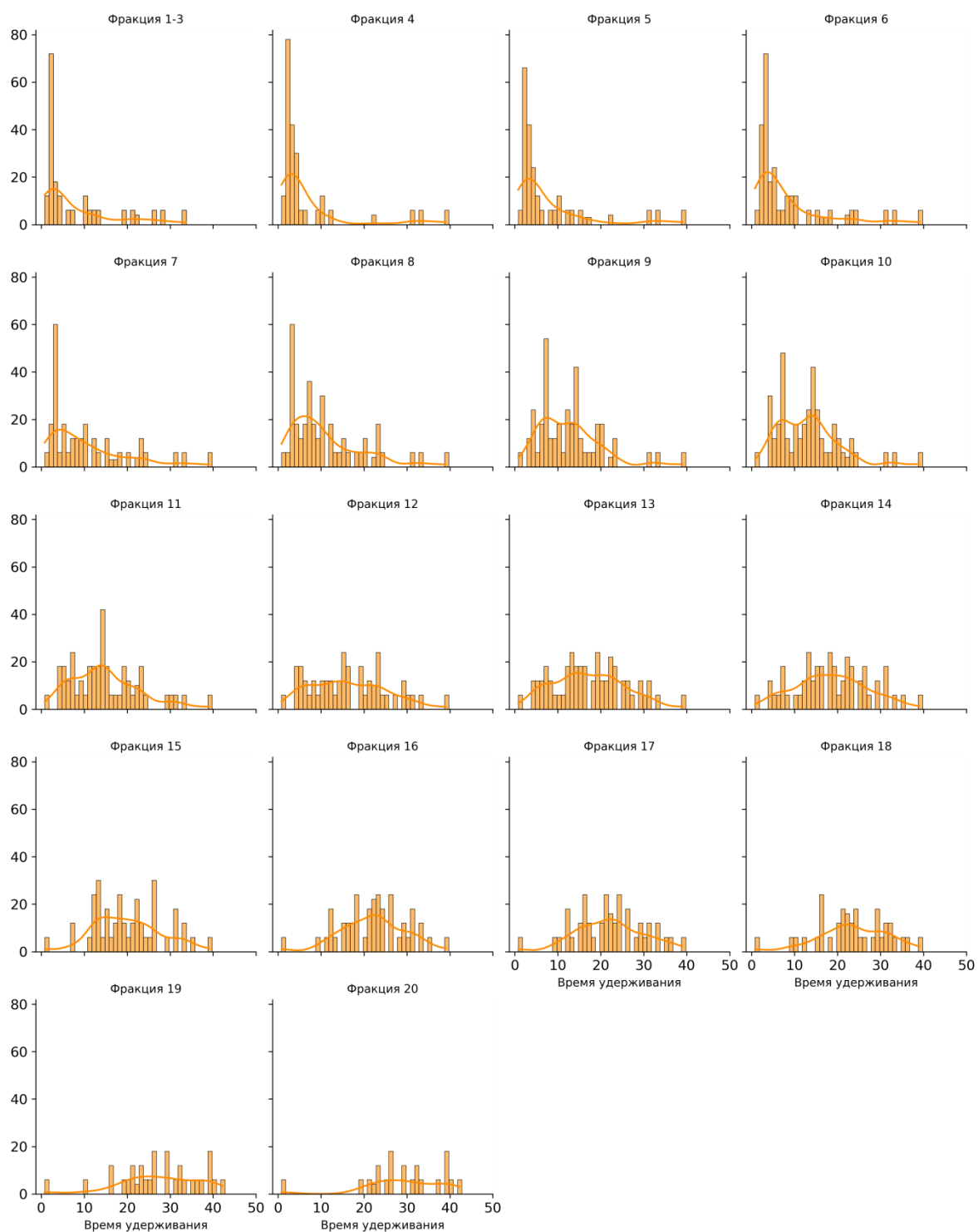


Рис. 20 Распределение регистрируемых транзиций в SRM методике по скринингу белков, кодируемых генами 18 хромосомы человека на хроматографическом градиенте в каждой фракции, полученной после хроматографического разделения пептидов в щелочных условиях.

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

Для таргетного SRM SIS метода был приготовлен образец пептидов из 3 миллионов клеток линии HepG2 аналогичным образом. В часть пробы, которая соответствует 100 тысячам клеток, были добавлены протеотипические пептиды, меченные стабильными изотопами, соответствующие белкам, которые кодируются генами 18 хромосомы человека. Полученный образец был использован для SRM SIS анализа. Таким образом, в комплексном образце проведён направленный поиск пептидов белков, кодируемых генами 18 хромосомы человека, общим методом для всех пептидов.

Перед фракционированием пробы пептидов, полученных из 3 миллионов клеток линии HepG2, в образец также были добавлены протеотипические пептиды, меченные стабильными изотопами, соответствующие белкам, кодируемым генами 18 хромосомы человека. На первом этапе проводили хроматографическое разделение смеси пептидов, полученных из 3 миллионов клеток линии HepG2 и изотопно – меченых пептидов. Затем каждая фракция была проанализирована SRM методикой, которая создавалась для данной фракции.

В одномерном 1D SRM анализе зарегистрировано 68 белков, в двумерном 2D SRM – 88, общее множество белков составляет 62, а всего зарегистрировано 94 белка, кодируемых генами 18 хромосомы человека. В 1D SRM анализе зарегистрировано 6 уникальных белков, а в 2D SRM анализе 26 уникальных белков (рисунок 21).

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

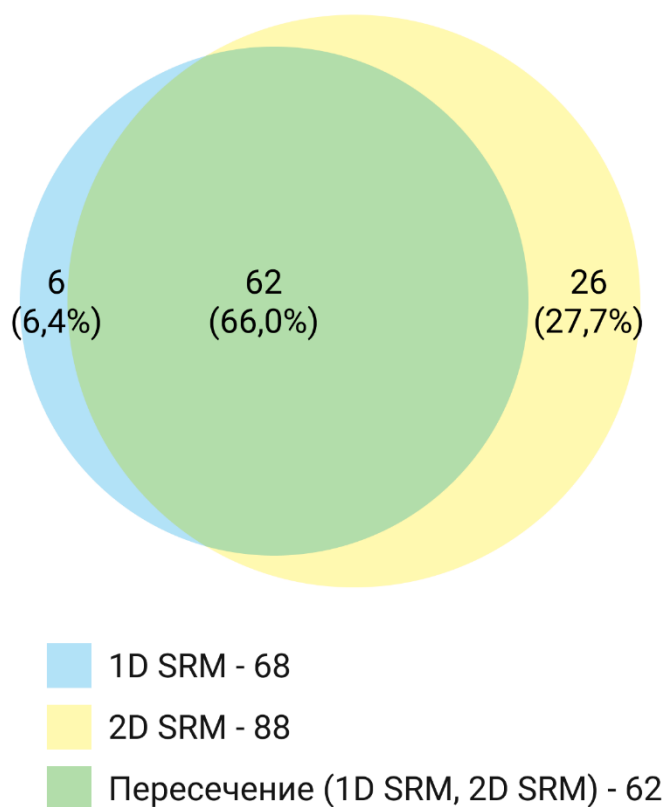


Рис. 21 Диаграмма Эйлера-Венна, сравнение множеств белков, обнаруженных при одномерном сканировании не фракционированной пробы (1D SRM) и фракционированной пробы (2D SRM).

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 22), 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 методом. Несмотря на то, что SRM метод в 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.

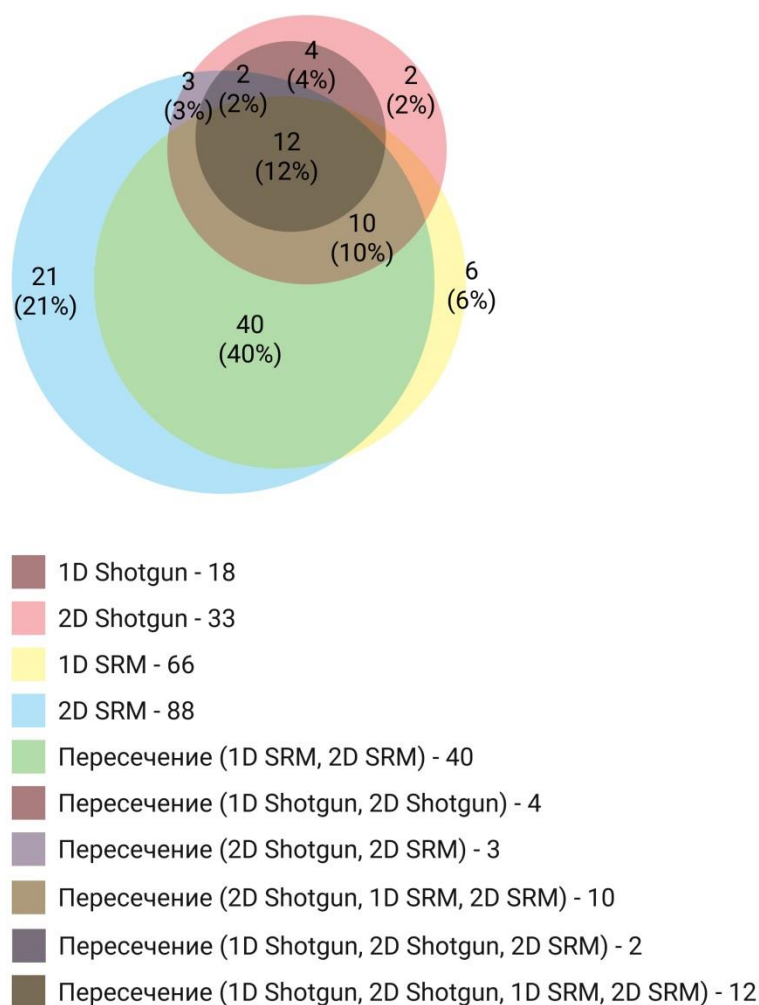


Рис. 22 Диаграмма Эйлера-Венна, сравнение множеств белков, кодируемых генами 18 хромосомы человека, обнаруженных при одномерном сканировании не фракционированной пробы (1D Shotgun, 1D SRM) и фракционированной пробы (2D Shotgun, 2D SRM).

Диапазон концентраций зарегистрированных белков лежит в пределах от 941000 копий белка на клетку до 100 копий на клетку и их можно условно разделить на три группы: высокое содержание, среднее и малое. К белкам в высокой концентрации относятся: P62314 – играет роль в сплайсинге пре – мРНК как коровый компонент

комплекса SMN-Sm, который опосредует сборку малых ядрышковых рибонуклеопротеинов, P18621 – компонент большой рибосомальной субъединицы, Q8TDN4 – увеличивает степень фосфорилирования тирозина циклин – зависимой киназы с помощью не рецепторных тирозинкиназ, таких как CDK5 [103]. Данные белки отвечают за структурные функции и регуляцию клеточного цикла. К белкам в средней концентрации относятся: P49257 – манноза – специфичный лектин, может узнавать сахарные остатки гликопротеинов, гликолипидов или гликозилфосфатидил инозитольные якоря и вовлечен в сортировку и переработку белков и липидов, его комплекс с MCFD2 формирует рецептор, транспортирующий белки из ЭПР в комплекс Гольджи, P25705 – митохондриальная мембранная АТФ – синтаза, катализирует синтез АТФ из АДФ, в присутствии градиента протонов между мембранами, который генерируется электрон транспортными комплексами дыхательной цепи, Q15555 – может участвовать в процессе полимеризации микротрубочек, миграции клетки [104, 105]. Описанные белки отвечают за базовые клеточные функции – миграцию, рецепторную и метаболизм. К белкам в низкой концентрации относятся: P07947 – нерецепторная тирозинкиназа, участвует в регуляции клеточного роста, выживания, апоптоза, клеточной адгезии, дифференцировки, фосфорилируется рецепторными тирозинкиназами EGRF, PDGFR, CSF1R, O14730 – участвует в регуляции зависимого от интерферона 1 иммунного ответа, который играет важную роль в иммунном ответе на вирусную ДНК и РНК, Q13464 – протеинкиназа, которая играет ключевую роль в регуляции актинового цитоскелета и клеточной полярности [106, 107, 108]. Данные белки отвечают в основном за передачу сигнала в клетке.

4.9 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 23). Полученная гистограмма свидетельствует о том, что наибольшее пересечение в процентном отношении белков и соответствующих им транскриптов обнаруживается при отсечении по RPKM ≥ 1 (52,2%), при котором регистрируются 145 транскриптов, кодируемых генами 18 хромосомы человека.

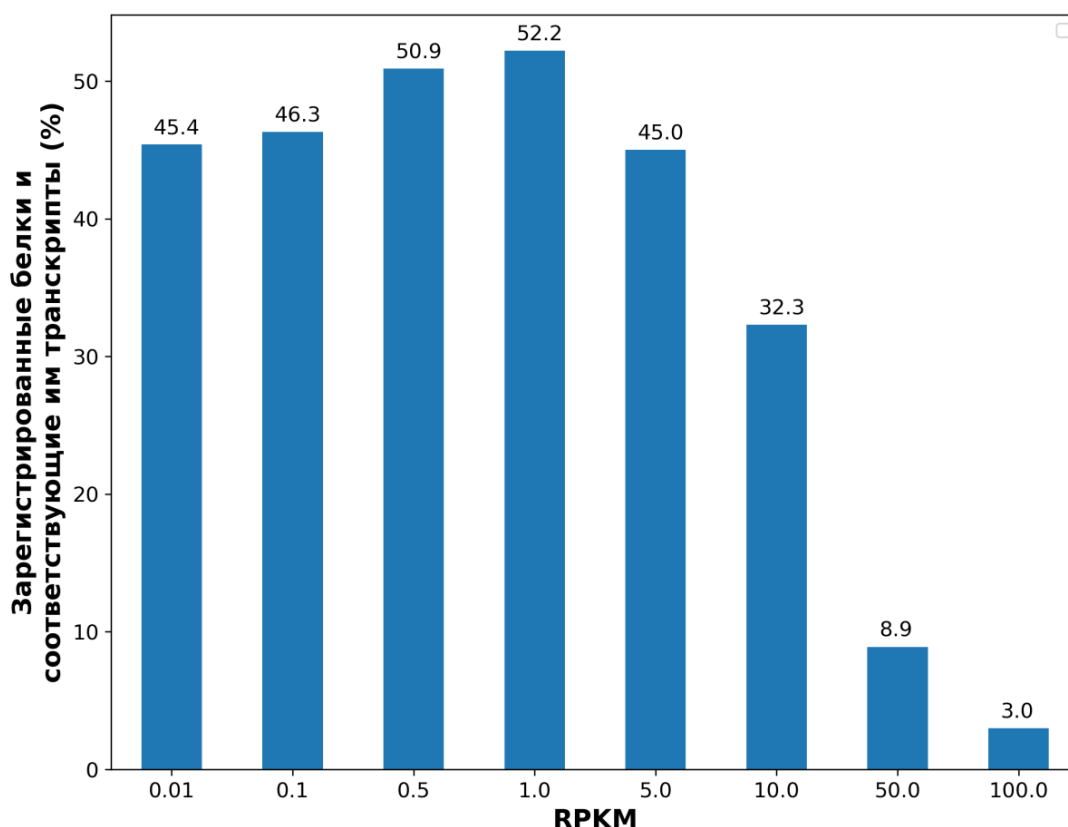


Рис. 23 Зависимость процента совместно зарегистрированных транскриптов и соответствующих им белков, кодируемых генами 18 хромосомы человека от отсечения по RPKM.

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 24 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 [109]. Так, период полураспада молекулы РНК в клетках млекопитающих варьирует от нескольких минут до нескольких часов [110, 111]. При этом самый малый зарегистрированный период полураспада белка составляет несколько часов, а наиболее стабильные не деградируют более 1000 часов [112]. 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 [113, 114]. Поэтому, при достаточно высокой

эффективности трансляции возможна ситуация, когда белок присутствует в клетке в концентрации, достаточной для детекции, а мРНК при этом отсутствует.

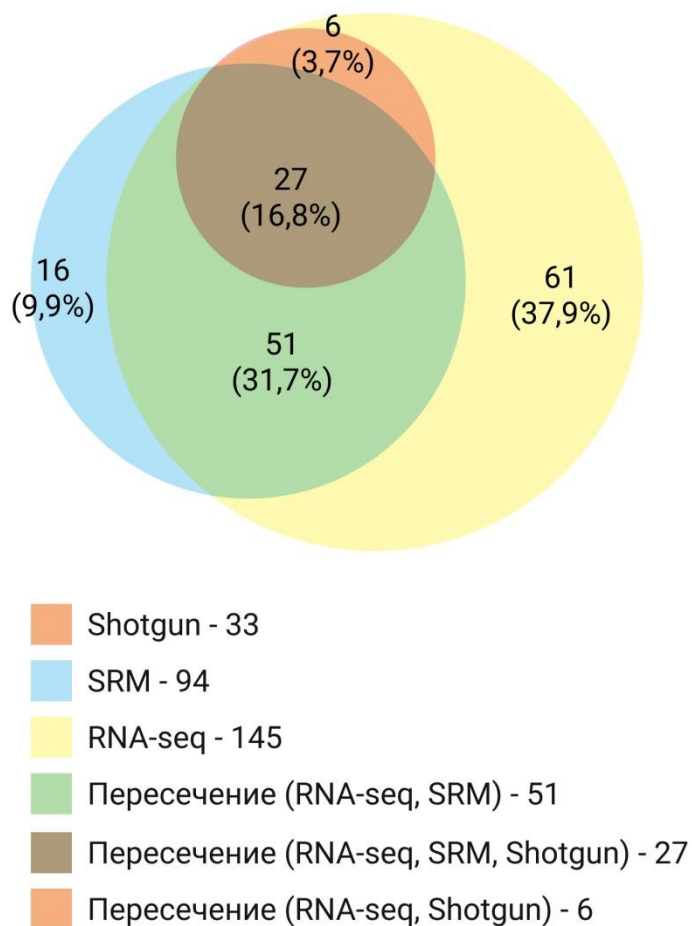


Рис. 24 Диаграмма Эйлера-Венна, представляет собой три множества, зарегистрированные в образце НерG2 транскрипты – RNA-seq, белки, кодируемые генами 18 хромосомы человека, зарегистрированные методом Shotgun, белки, зарегистрированные методом SRM.

Несмотря на отсутствие транскрипта, достоверность идентификации этих белков не вызывает сомнений. Так, на рисунке 25 представлены SRM-хроматограммы пептида YIHSANVLHR (белок P31152) и его изотопно-меченного стандарта. Как видно из представленных данных, форма и порядок транзиций и их время удерживания полностью совпадают, а их идентичность не вызывает сомнений.

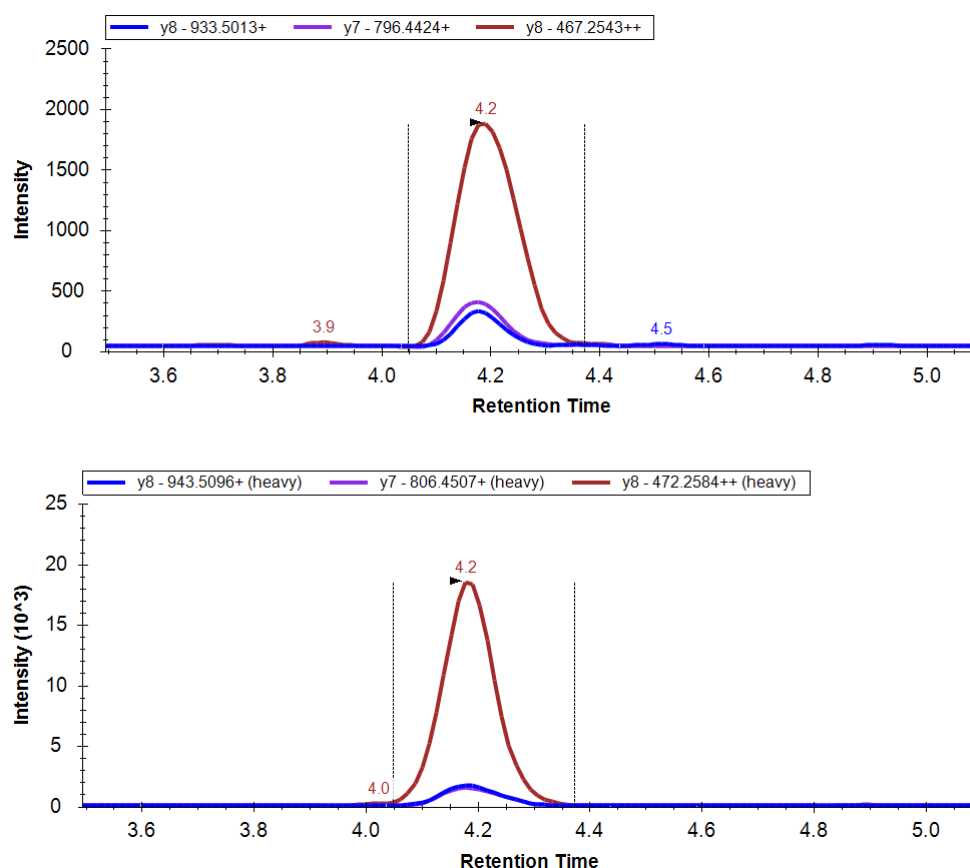


Рис. 25 Хроматографические пики переходов изотопно меченного пептидного стандарта YIHSANVLHR (нижний ряд) и его природного аналога из образца HepG2 (верхний ряд).

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

4.10 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 детекции 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 26 представлены результаты измерения эффективности ионизации пептидов, проведенные на фоне биологической матрицы клеток линии HepG2. С помощью алгоритма кластеризации k-

средних стандартные пептиды были разделены на 3 условные группы: с высокой эффективностью ионизации, средней и низкой. На рисунке 26 цветом выделены точки, каждая из которых обозначает один белок и соответствующий транскрипт, зарегистрированный в клетках линии HepG2. 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%. Таким образом, чем ниже эффективность ионизации протеотипического пептида, тем меньше вероятность зарегистрировать белок, соответствующий транскрипт которого детектируется в образце.

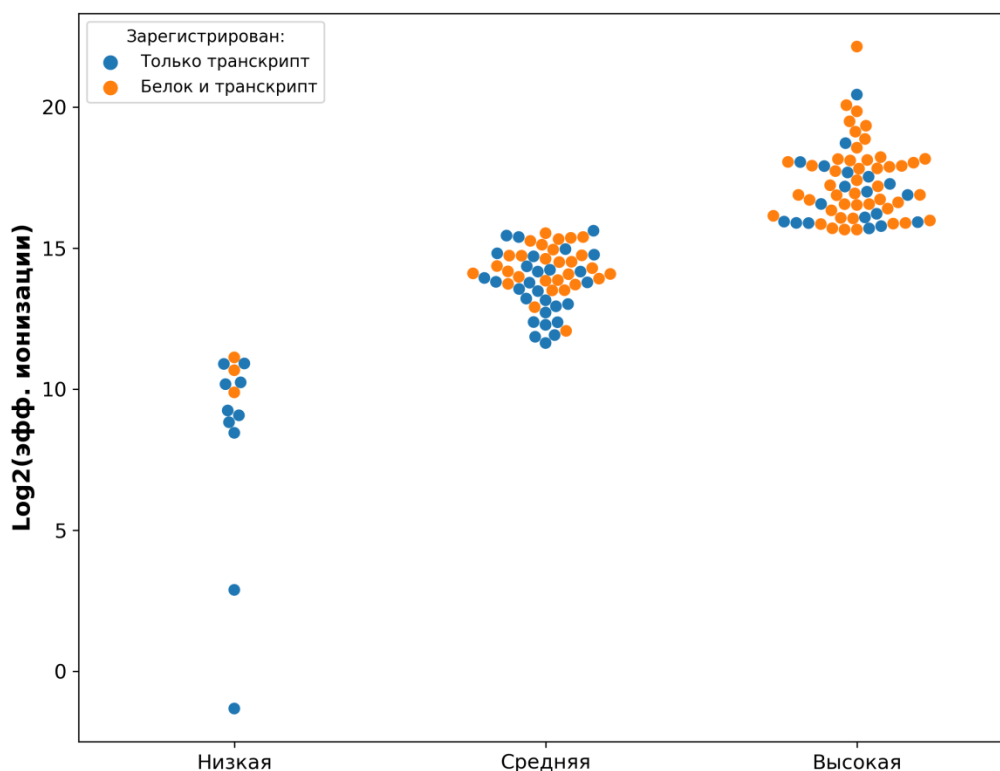


Рис. 26 Эффективность ионизации стандартных пептидов для таргетного анализа на фоне биологической матрицы пептидов клеточной линии HepG2, цветом выделены белки, для которых был зарегистрирован и белок и соответствующий ему транскрипт или только транскрипт.

4.11 Оценка стабильности и воспроизводимости метода двумерного фракционирования на примере клинических образцов нормальной печени человека

Для проведения эксперимента использовали 3 образца печени человека. Этапы выполнения работы полностью повторяли вышеописанные для образца HepG2. Для

данных образцов были получены похожие закономерности, которые получены для клеток линии HepG2. Показано, что в среднем в образце печени человека 1D Shotgun анализом можно обнаружить 21 ± 3 белка в каждом образце, а для 2D Shotgun анализа – 36 ± 5 белков. В 2D анализе обнаружено в 1,5 раза больше белков.

В 1D Shotgun анализе зарегистрировано 26 белка, в 2D Shotgun анализе – 48 белков. При этом общее множество зарегистрированных белков составляет 25, а всего зарегистрировано 49 белков, кодируемых генами 18 хромосомы человека в клетках печени. Уникальными идентификациями для множества 1D Shotgun является 1 белок, а для 2D Shotgun 23 белка. Один белок идентифицируется исключительно в комплексном образце (1D Shotgun), это связано с потерей части образца, что неизбежно происходит в ходе дополнительных манипуляций перед анализом, при фракционировании в щелочных условиях (рисунок 27). Однако большое количество - 23 уникальных белка - относится к 2D Shotgun, что свидетельствует об эффективности методики двумерного фракционирования.

При применении методики 2D Shotgun дополнительно зарегистрировано 23 уникальных белка печени, кодируемых генами 18 хромосомы человека, а использование методики в эксперименте с клеточной линией HepG2 позволило зарегистрировать дополнительно 15 белков. Таким образом, качественный анализ показывает, что на клинических образцах печени человека предложенный метод работает и показывает эффективность, сравнимую с результатами, полученными на клеточной линии HepG2.

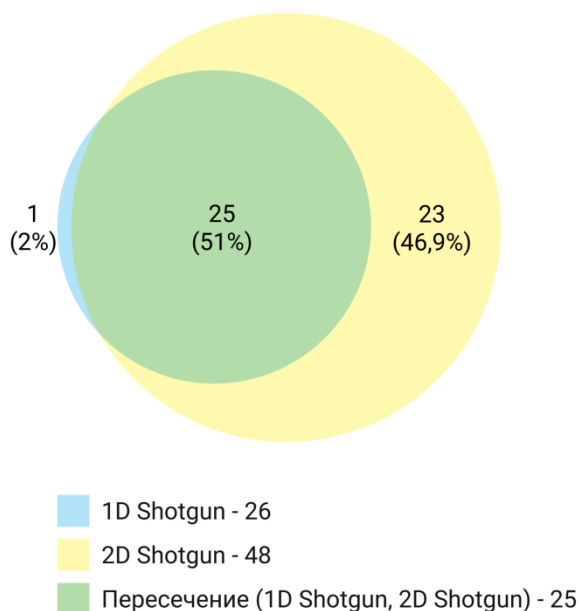


Рис. 27 Диаграмма Эйлера-Венна, количество идентифицированных белков, кодируемых генами 18 хромосомы человека в 1D Shotgun анализе и 2D Shotgun.

Далее в ходе работ была проведена оценка стабильности и воспроизводимости методики 1D Shotgun анализа. Всего в трех образцах нормальной печени человека зарегистрировано 26 белков, кодируемых генами 18 хромосомы человека (рисунок 28). Большинство белков, 17 из 26, регистрируется в каждом образце, однако пять белков зарегистрировано только в первом образце и по одному во втором и третьем. Таким образом, вариабельность по количеству зарегистрированных белков между образцами составляет 34,6%. Для относительного количественного анализа был использован критерий iBAQ. Он позволяет оценить соотношение концентраций белка в образце и рассчитывается следующим образом: берется сумма интенсивностей обнаруженных пептидов и делится на количество теоретических пептидов, которые могли быть получены при гидролизе для данного конкретного белка. Для нормализации значений iBAQ использовалась сумма всех значений iBAQ в образце. Каждое значение iBAQ для белков затем делилось на сумму всех iBAQ и получается riBAQ. Он используется для сравнения концентраций разных образцов между собой.

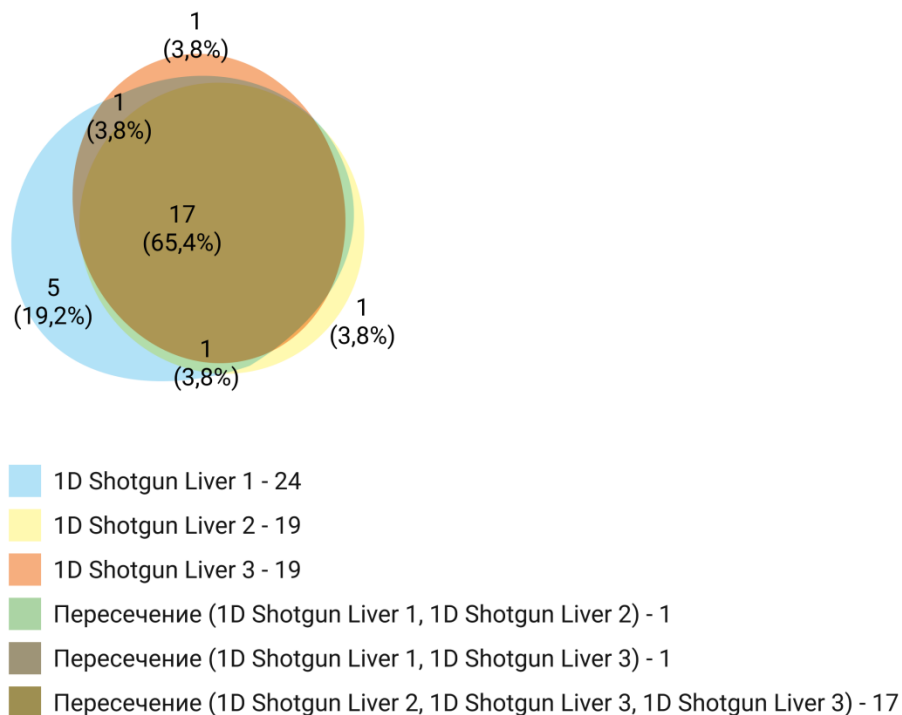


Рис. 28 Диаграмма Эйлера-Венна, количество идентифицированных белков, кодируемых генами 18 хромосомы человека в 1D Shotgun анализе и 2D Shotgun в каждом образце печени.

Также была проведена оценка стабильности и воспроизводимости методики 2D Shotgun анализа. Всего в трех образцах нормальной печени человека было зарегистрировано 48 белков, кодируемых генами 18 хромосомы человека (рисунок 29). Большинство белков – 24, регистрируется в каждом образце, однако семь белков зарегистрировано только в первом образце, а также два и четыре во втором и третьем образцах соответственно. Таким образом, вариабельность по количеству зарегистрированных белков между образцами составляет 50%. В этом случае вариация заметно выше, чем при 1D Shotgun анализе, однако и суммарное количество обнаруженных белков и белков, обнаруженных в каждом образце выше 48 и 24 соответственно.

Таким образом, 2D анализ позволяет глубже проанализировать образцы, увеличить количество зарегистрированных белков, кодируемых генами 18 хромосомы человека, как общих для каждого образца, так и индивидуальных. Это помогает лучше понять индивидуальные различия образцов на количественном и качественном уровнях.

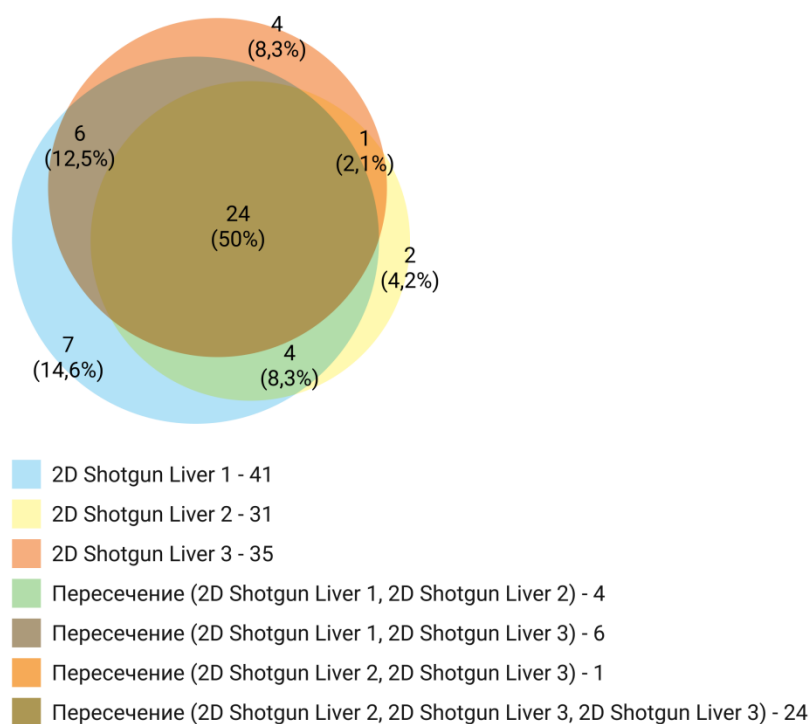


Рис. 29 Диаграмма Эйлера-Венна, количество идентифицированных белков, кодируемых генами 18 хромосомы человека в 2D Shotgun анализе в каждом образце печени.

Для оценки стабильности количественного анализа было рассчитано стандартное отклонение в процентах для относительных концентраций или коэффициент вариации (riBAQ) для 15 белков, кодируемых генами 18 хромосомы человека, которые обнаружены в каждом образце и в 1D Shotgun анализе и в 2D Shotgun.

На рисунке 30 показано, что коэффициент вариации для белков, измеренных методами 1D Shotgun и 2D Shotgun не отличается и в среднем составляет 40%. Таким образом, можно сделать вывод, что методика 2D Shotgun обеспечивает достаточный уровень стабильности и воспроизводимости измерений и не вносит дополнительной неопределенности в данные количественного анализа, а по точности соответствует 1D Shotgun анализу.

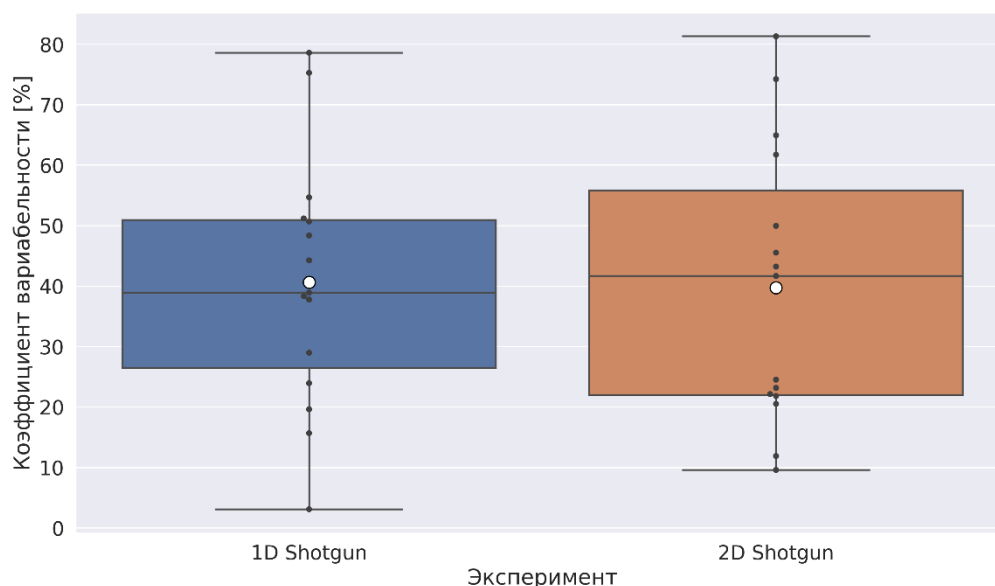


Рис. 30 Сравнение распределений коэффициентов вариабельности в процентах измеренных значений относительной концентрации iBAQ для белков, зарегистрированных в каждом образце и в 1D Shotgun и в 2D Shotgun анализе (среднее значение обозначено белым кругом).

Полученные концентрации белков, кодируемых генами 18 хромосомы человека, были сгруппированы с помощью алгоритма кластеризации k-means и разбиты на 3 кластера как белки в высокой концентрации, средней и низкой концентрации (рисунок 31). В группу с высокой концентрацией определены белки, отвечающие за получение энергии клеткой и регуляцию сокращения гладкой мускулатуры клеток: P42765 (THIM_HUMAN), P00167 (CYB5_HUMAN), P25705 (ATPA_HUMAN), Q96KP4 (CNBP2_HUMAN), P24844 (MYL9_HUMAN). Группа белков в средней концентрации представляет собой структурные белки и белки, которые отвечают за общие метаболические и рецепторные функции клетки: P49257 (LMAN1_HUMAN), Q9P0L0 (VAPA_HUMAN), P18621 (RL17_HUMAN), Q9Y4W6 (AFG32_HUMAN), P22830 (HEMH_HUMAN), Q7LBR1 (CHM1B_HUMAN), P02766 (TTHY_HUMAN). Группу белков в низкой концентрации представляют протеинкиназы и белки, отвечающие за межклеточные взаимодействия: Q13464 (ROCK1_HUMAN), Q96EN8 (MOCOS_HUMAN), Q14126 (DSG2_HUMAN).

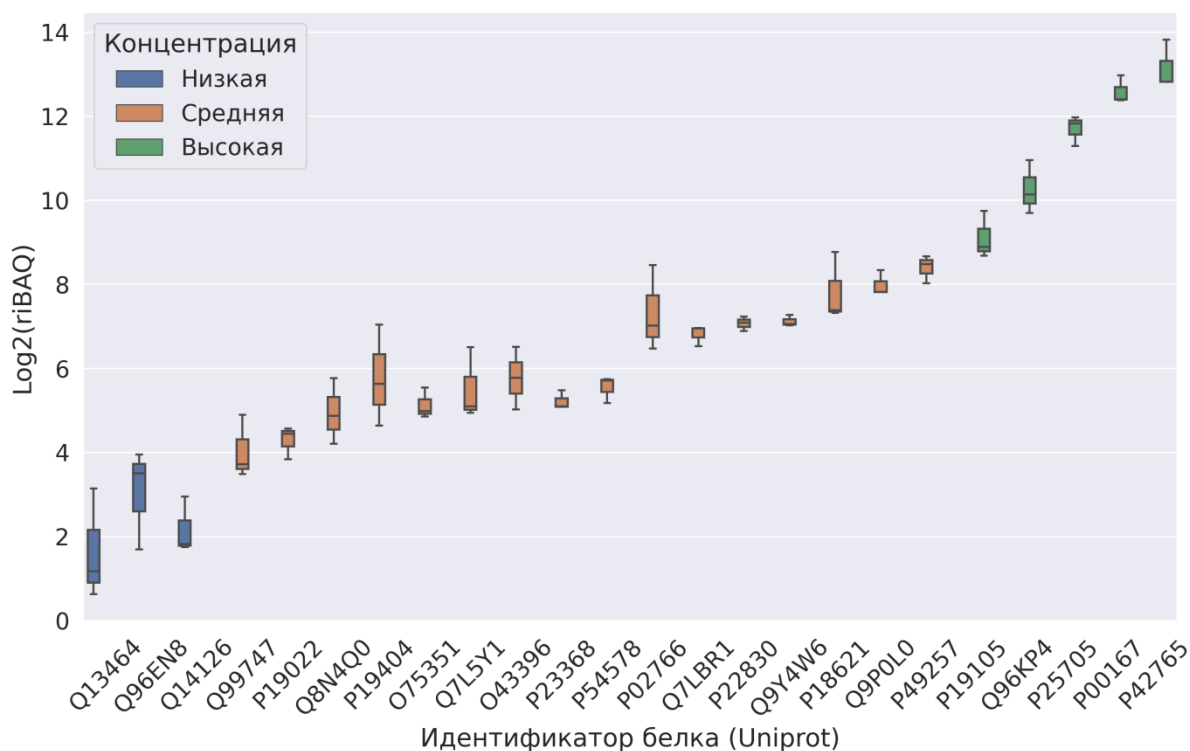


Рис. 31 Относительные концентрации белков, кодируемых генами 18 хромосомы человека обнаруженные в каждом образце печени человека, сгруппированные с помощью кластеризации на группы по двоичному логарифму относительной концентрации, рассчитанной по алгоритму iBAQ.

4.12 Таргетный протеомный анализ (SRM SIS) белков, кодируемых генами 18 хромосомы человека образцов нормальной печени человека

Проведена оценка стабильности и воспроизводимости методики 1D SRM анализа. Всего в трех образцах нормальной печени человека было зарегистрировано 66 белков, кодируемых генами 18 хромосомы человека (рисунок 32). Большинство белков – 40, регистрируется в каждом образце, однако 5 белков зарегистрировано только в первом образце и по 6 во втором и третьем. Таким образом, вариабельность между образцами составляет 39,4%. Для количественного анализа использовались абсолютные концентрации, полученные при расчете отношения площади под кривой сигнала стандартного изотопно-меченого пептида и эндогенного пептида образца.

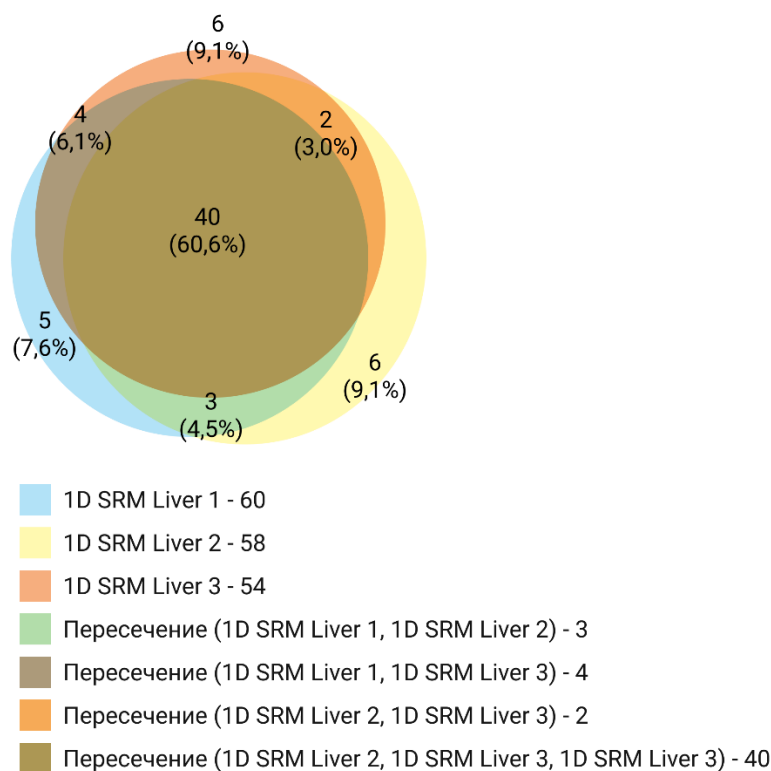


Рис. 32 Диаграмма Эйлера-Венна, количество идентифицированных белков, кодируемых генами 18 хромосомы человека в 1D SRM анализе в каждом образце печени.

Показатели стабильности 2D SRM анализа схожи с показателями методики 1D SRM анализа. Всего в трех образцах нормальной печени человека зарегистрировано 110 белков, кодируемых генами 18 хромосомы человека (рисунок 33). Большинство белков – 68, регистрируется в каждом образце, однако 7 белков зарегистрировано только в первом образце и 9 во втором и 6 белков в третьем образце. Таким образом, вариабельность между образцами составляет 38,2%. Так вариабельность между образцами в 1D и 2D SRM анализе составляет 39,4% и 38,2% соответственно, однако в абсолютном количестве в 2D анализе регистрируется на 28 белков больше в каждом образце, что позволяет более детально охарактеризовать изучаемую биологическую систему.

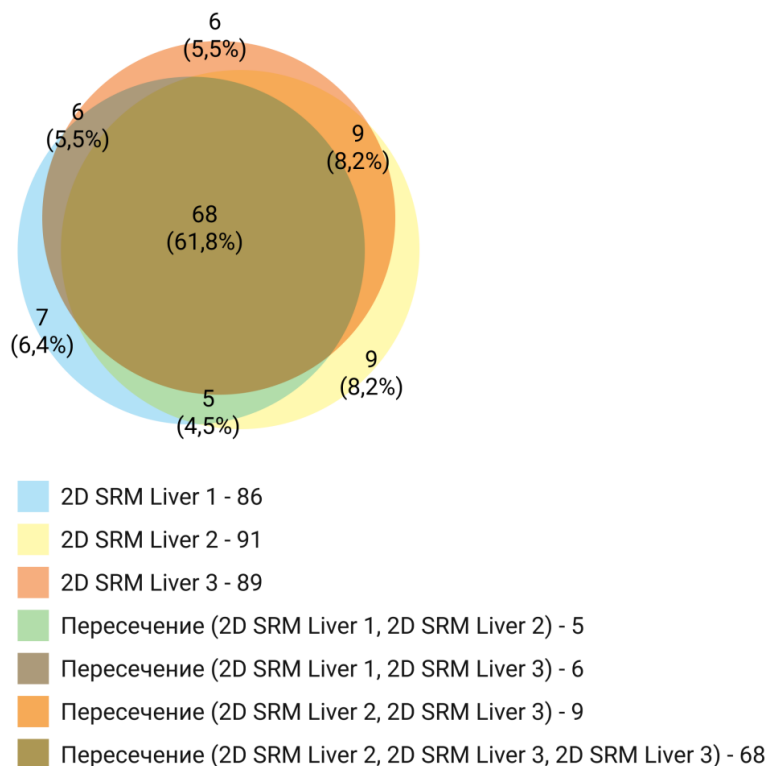


Рис. 33 Диаграмма Эйлера-Венна, количество идентифицированных белков, кодируемых генами 18 хромосомы человека в 1D SRM анализе в каждом образце печени.

Также для группы белков, которые регистрируются в каждом образце печени и 2D SRM и 1D SRM методиками проведен анализ стабильности детектируемых концентраций белков, кодируемых генами 18 хромосомы человека (рисунок 34). Показатели коэффициента вариабельности регистрируемых концентраций белков в среднем составляют 30% как для 1D, так и для 2D анализа. Таким образом, методика 2D SRM обеспечивает достаточный уровень стабильности и воспроизводимости измерений и не вносит дополнительной неопределенности в данные количественного анализа, а по точности соответствует 1D SRM анализу.

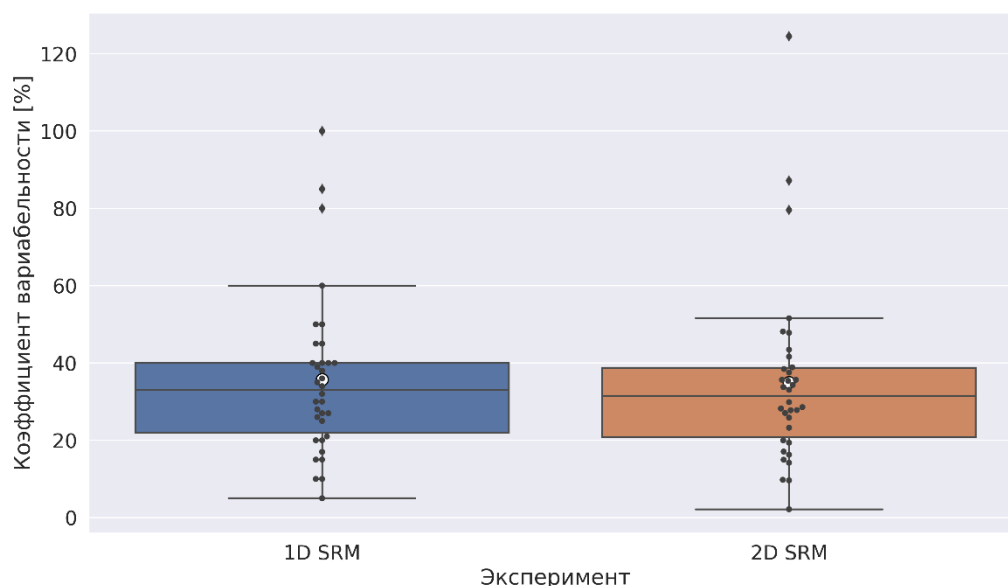
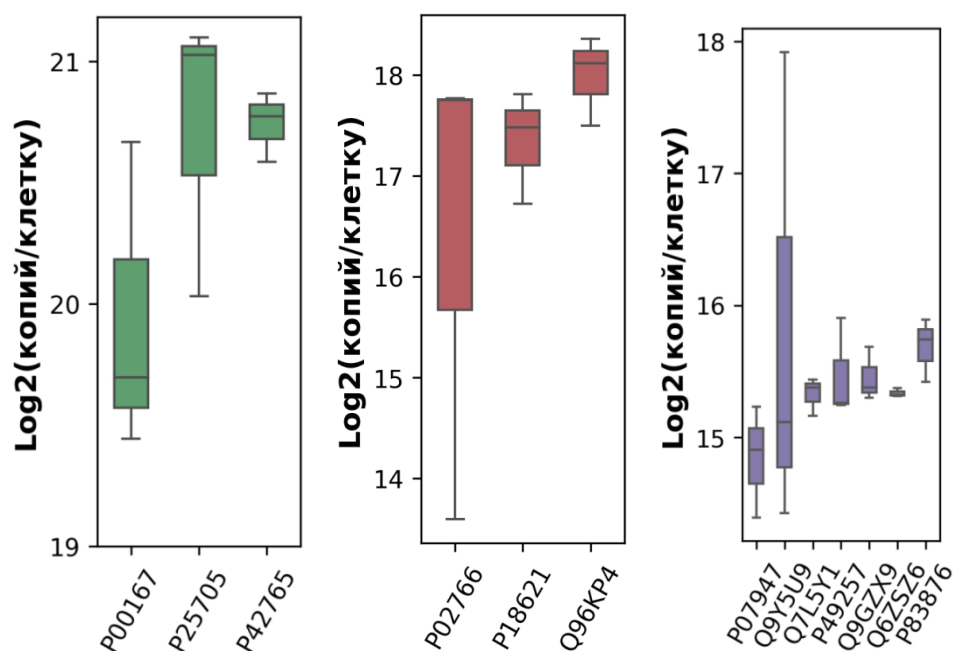


Рис. 34 Сравнение распределений коэффициентов варибельности в процентах измеренных значений количества копий на клетку для белков, зарегистрированных в каждом образце и в 1D SRM и в 2D SRM анализе (среднее значение обозначено белым кругом).

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 35). 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 копийность которых ниже 1000 на клетку с минимально 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) нерецепторная 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 Р 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.



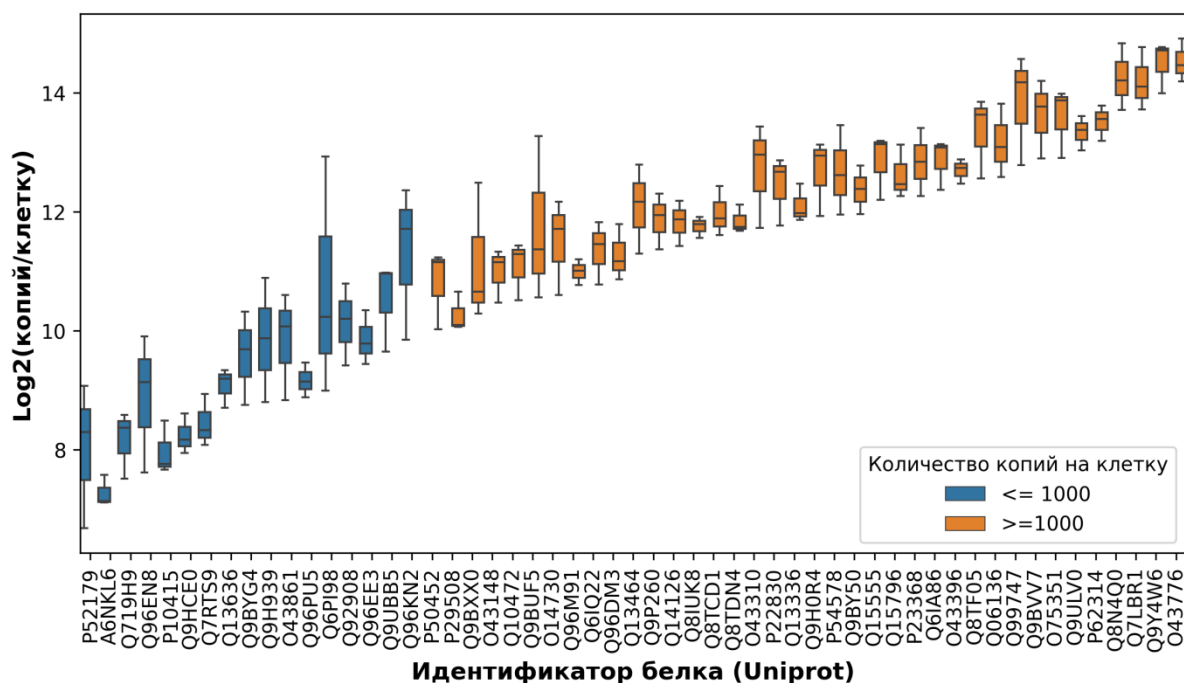


Рис. 35 Содержание в клетке зарегистрированных белков в количестве копий на клетку, разделенные на кластеры по содержанию в клетке.

Всего в печени зарегистрировано больше белков, кодируемых генами 18 хромосомы человека - 129, чем в клеточной линии HepG2 - 100, этому явлению может быть несколько объяснений. Во-первых, биологическая вариабельность: в каждом образце обнаружены уникальные для данного образца белки. Во-вторых, печень – сложный гетерогенный орган, который состоит из нескольких типов клеток. Четыре основных типа клеток печени — гепатоциты (HCs), звездчатые клетки печени (HSCs), клетки Купфера (KCs) и синусоидальные эндотелиальные клетки печени (LSECs) — взаимодействуют в пространстве и времени, формируя и поддерживая функции печени. Гепатоциты составляют 70% популяции всех клеток печени. Остальная часть популяции состоит из непаренхиматозных клеток, а именно LSEC, KC и HSC [115]. Являясь паренхиматозной частью печени, HCs в основном участвуют в основных функциях печени, включая метаболизм липидов, метаболизм лекарственных средств и секрецию факторов свертывания и комплемента [116]. Клетки Купфера, которые составляют одну треть непаренхиматозных клеток в печени, осуществляют иммунный надзор [117]. Хотя звездчатые клетки составляют только 5% клеток печени, они играют центральную роль в запасании витамина А и липидов [118]. Эндотелиальные клетки, которые составляют большую часть (50%) непаренхиматозных клеток печени, отделяют гепатоциты от синусоидального просвета [119].

4.13 Анализ показателей биологической вариации транскриптомных и протеомных данных в образцах нормальной печени человека

Для анализа были использованы данные высокопроизводительного РНК секвенирования полученные с помощью платформы Illumina 3 образцов нормальных клеток печени человека, сгруппированных по значениям RPKM с помощью алгоритма кластеризации k-средних на 4 группы: высокие значения (среднее значение по кластеру 94297 RPKM), средние (среднее значение по кластеру 8495 RPKM), низкие (среднее значение по кластеру 1171 RPKM), ультранизкие (среднее значение по кластеру 150 RPKM). Протеомные данные представляют собой зарегистрированные концентрации белков, кодируемых генами 18 хромосомы человека, в 3 образцах нормальной печени человека с помощью методики 2D SRM SIS.

Анализ биологической вариации показал, что существует группа с низким коэффициентом вариабельности, как белков, так и транскриптов значения коэффициента составляют менее 50%, нижняя левая часть графика. 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 36). В эту группу вошли такие важные для выживания и жизнедеятельности клетки белки как: Q13336 (транспортер мочевины 1), O43310 (CBP80/20-зависимый фактор инициации трансляции), Q6IQ22 (Ras-родственный белок Rab-12, регулирует внутриклеточный мембранный транспорт везикул), P10415 (регулятор апоптоза bcl-2), Q9UBB5 (Метил-CpG-связывающий домен белок 2, реперессор транскрипции). Также на графике (рисунок 36) показано, что данные транскриптомного анализа показывают значительно более высокие значения коэффициента вариации, чем данные протеомного анализа, что может объясниться значительно более низкими значениями периода полураспада транскриптов относительно соответствующих молекул белка.

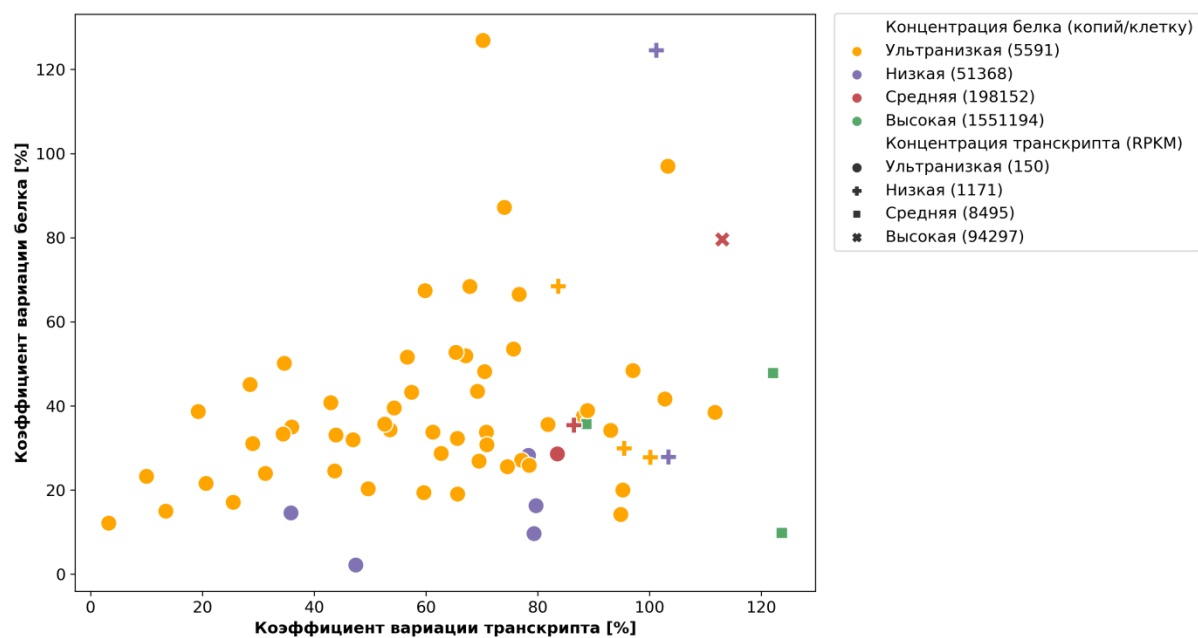


Рис. 36 Анализ коэффициентов вариации белков и соответствующих им транскриптов зарегистрированных в каждом образце печени человека.

5. ЗАКЛЮЧЕНИЕ

Основной проблемой таргетного анализа комплексных биологических образцов является низкая чувствительность прибора и 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, количество которых может достигать до 6 миллионов являются матрицей. В таком случае целесообразно снизить эффект матрицы с помощью фракционирования. Таким образом, разделение исходного комплексного образца на 20 фракций дает значимое снижение динамического диапазона белков. Так, одна фракция содержит не 6 миллионов белков а 300 тысяч, что оказывает значимый эффект и увеличивает количество регистрируемых белков по сравнению со стандартным одномерным масс-спектрометрическим анализом. В работе показано, что белки в ультранизкой концентрации составляют большую часть протеома 18 хромосомы в клетках печени человека (56 из 69 белков), 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 высокочувствительной детекции низкокопийных белков недостаточно снизить эффект матрицы в образце. На примере пептидов стандартной системы UPS1 и в последующем анализе при помощи стандартных протеотипических пептидов, 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 такого метода ограничена, и, следовательно, для каждого белка не удастся найти оптимальный пептид для таргетного анализа. Поэтому актуальной задачей является создание предсказательной модели, которая могла бы выявлять пептиды с высокой эффективностью ионизации на основе полученных данных панорамной масс-спектрометрии. Особенно полезным это предсказание может оказаться при детекции белков протеома 18 хромосомы печени,

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.

Применение методов с двумерным фракционированием не позволяет достигнуть чувствительности транскриптомных технологий – 145 зарегистрированных транскриптов, кодируемых генами 18 хромосомы против 100 соответствующих белков в клеточной линии HepG2. Наличие молекул мРНК, зарегистрированных в образце, говорит о том, что возможно наличие соответствующего белка. Однако существуют данные об очень низкой корреляции транскрипции трансляции эукариотических клеток. Это можно объяснить большим количеством различных вариантов регуляции трансляции. В клетке существует несколько классов РНК, которые напрямую ингибируют трансляцию: miRNA, siRNA [120]. Кроме РНК существуют также и белки, комплементарные определённым транскриптам и блокирующие трансляцию [121]. Таким образом, есть основания полагать, что транскриптом не является зеркальным отражением протеома, т.е. ситуация когда в клетке присутствует транскрипт и отсутствует соответствующий белок вполне нормальна. В этом направлении, возможно, будет достигнут прогресс в рамках исследования транслятома, то есть только тех молекул РНК, которые связаны с рибосомами и вероятнее всего транслируются [122].

6. 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 HerG2 позволило увеличить количества идентификаций белков в 2 раза и на 33% панорамного и таргетного анализа, соответственно. В ходе 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 нормальной 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 детектировано 6599 белков, что составляет 32% протеома человека, в том числе 54% белков (142), кодируемых генами 18 хромосомы человека.

7. СПИСОК ЛИТЕРАТУРЫ

1. Zahn-Zabal, M., Michel, P. A., Gateau, A., Nikitin, F., Schaeffer, M., Audot, E., Gaudet, P., Duek, P. D., Teixeira, D., de Laval, V. R., Samarasinghe, K., Bairoch, A., & Lane, L. The neXtProt knowledgebase in 2020: Data, tools and usability improvements. // *Nucleic Acids Research*. 2020. Vol. 48(D1). P. 328–334.
2. Millionsi, R., Tolin, S., Puricelli, L., Sbrignadello, S., Fadini, G. P., Tessari, P., & Arrigoni, G. High Abundance Proteins Depletion vs Low Abundance Proteins Enrichment: Comparison of Methods to Reduce the Plasma Proteome Complexity. // *PloS one*. 2011. Vol. 6, №5. e19603.
3. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., & Aebersold, R. (2009). Full Dynamic Range Proteome Analysis of *S. cerevisiae* by Targeted Proteomics. // *Cell*. 2009. Vol. 138, №4. P. 795–806.
4. Furey, A., Moriarty, M., Bane, V., Kinsella, B., & Lehane, M. Ion suppression: A critical review on causes, evaluation, prevention and applications. // *Talanta*. 2013. Vol. 115, P. 104–122.
5. Mei, H., Hsieh, Y., Nardo, C., Xu, X., Wang, S., Ng, K., & Korfmacher, W. A. Investigation of matrix effects in bioanalytical high-performance liquid chromatography/tandem mass spectrometric assays: Application to drug discovery. // *Rapid Communications in Mass Spectrometry*. 2003. Vol. 17, №1. P. 97–103.
6. Michalski, A., Cox, J., & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. // *Journal of Proteome Research*. 2011. Vol. 10, №4. P. 1785–1793.
7. Picotti, P., Bodenmiller, B., & Aebersold, R. Proteomics meets the scientific method. // *Nature Methods*. 2013. Vol. 10, №1. P. 24–27.
8. 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.

9. Squires, G. Francis Aston and the mass spectrograph. // J. Chem. Soc.. 1998. Vol. 23. P. 3893–3899.
10. McLafferty, F. Mass Spectrometric Analysis. Molecular rearrangements. // Anal. Chem. 1959. Vol. 31, №1. P. 82–87.
11. Biemann, K. On the ability of the Viking gas chromatograph–mass spectrometer to detect organic matter. // Proceedings of the National Academy of Sciences. 2007. Vol. 104, №25. P. 10310–10313.
12. Fenselau, C. C. Carl Djerassi (1923 – 2015): an Obituary for Chemists. // Journal of The American Society for Mass Spectrometry. 2015. Vol. 26, №4. P. 535–537.
13. Gohlke, R. S., & McLafferty, F. W. Early gas chromatography/mass spectrometry. // Journal of the American Society for Mass Spectrometry. 1993. Vol. 4, №5. P. 367–371.
14. Biemann, K., Gapp, G., & Seibl, J. Application of mass spectrometry to structure problems. Amino acid sequence in peptides. // Journal of the American Chemical Society, 1959. Vol. 81, № 9. P. 2274–2275.
15. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. Electrospray ionization–principles and practice. // Mass Spectrometry Reviews. 1990. Vol. 9, №1. P. 37–70.
16. Wood, M., Laloup, M., Samyn, N., del Mar Ramirez Fernandez, M., de Bruijn, E. A., Maes, R. A. A., & De Boeck, G. Recent applications of liquid chromatography-mass spectrometry in forensic science. // Journal of Chromatography. 2006. A, Vol. 1130, №1 Spec. Iss. P. 3–15.
17. Ojanperä, I., Kolmonen, M., & Pelander, A. Current use of high-resolution mass spectrometry in drug screening relevant to clinical and forensic toxicology and doping control. // Analytical and Bioanalytical Chemistry. 2012. Vol. 403, №5. P. 1203–1220.
18. Peters, R. J. B., Stolker, A. A. M., Mol, J. G. J., Lommen, A., Lyris, E., Angelis, Y., Vonaparti, A., Stamou, M., Georgakopoulos, C., & Nielen, M. W. F. Screening in veterinary drug analysis and sports doping control based on full-scan, accurate-mass spectrometry. // TrAC - Trends in Analytical Chemistry. 2010. Vol. 29, №11. P. 1250–1268.

19. Li, M., Wang, I. X., Li, Y., Bruzel, A., Richards, A. L., Toung, J. M., & Cheung, V. G. Widespread RNA and DNA sequence differences in the human transcriptome. // *Science*. 2011. Vol. 333, № 6038, P. 53–58.
20. The call of the human proteome. // *Nature Methods*. 2010. Vol. 7, № 9. P. 661.
21. Hancock, W., Omenn, G., Legrain, P., & Paik, Y. K. Editorial: Proteomics, human proteome project, and chromosomes. // *Journal of Proteome Research*. 2011. Vol. 10, № 1. P. 210.
22. Keshishian, H., Burgess, M. W., Specht, H., Wallace, L., Clauser, K. R., Gillette, M. A., & Carr, S. A. Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. // *Nature protocols*. 2017. Vol. 12, №8. P. 1683–1701.
23. Berglund, L., Björling, E., Oksvold, P., Fagerberg, L., Asplund, A., Szigartyo, C. A., Persson, A., Ottosson, J., Wernérus, H., Nilsson, P., Lundberg, E., Sivertsson, A., Navani, S., Wester, K., Kampf, C., Hober, S., Pontén, F., & Uhlén, M. A genecentric Human Protein Atlas for expression profiles based on antibodies. // *Molecular & cellular proteomics: MCP*. 2008. Vol. 7, №10. P. 2019–2027.
24. Anderson, L., & Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. // *Molecular & cellular proteomics : MCP*. 2006. Vol. 5, №4. P. 573–588.
25. Picotti, P., Lam, H., Campbell, D., Deutsch, E. W., Mirzaei, H., Ranish, J., Domon, B., & Aebersold, R. A database of mass spectrometric assays for the yeast proteome. // *Nature methods*. 2008. Vol. 5, №11. P. 913–914.
26. Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., Wenschuh, H., & Aebersold, R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. // *Nature methods*. 2010. Vol. 7, №1. P. 43–46.
27. Meyer, H. E., Klose, J., & Hamacher, M. HBPP and the pursuit of standardisation. // *The Lancet. Neurology*. 2003. Vol. 2, №11. P. 657–658.
28. Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D.

- W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Hanash, S. M. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. // *Proteomics*. 2005. Vol. 5, №13. P. 3226–3245.
29. He F. Human liver proteome project: plan, progress, and perspectives. // *Molecular & cellular proteomics: MCP*. 2005. Vol. 4, №12. P. 1841–1848.
30. Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabudhe, N. A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L. D., Pandey, A. A draft map of the human proteome. // *Nature*. 2014. Vol. 509, №7502. P. 575–581.
31. Paik, Y. K., Lane, L., Kawamura, T., Chen, Y. J., Cho, J. Y., LaBaer, J., Yoo, J. S., Domont, G., Corrales, F., Omenn, G. S., Archakov, A., Encarnación-Guevara, S., Lui, S., Salekdeh, G. H., Cho, J. Y., Kim, C. Y., & Overall, C. M. Launching the C-HPP neXt-CP50 Pilot Project for Functional Characterization of Identified Proteins with No Known Function. // *Journal of proteome research*. 2018. Vol. 17, №12, P. 4042–4050.
32. Smith, G. S., Walter, G. L., & Walker, R. M. Clinical Pathology in Non-Clinical Toxicology Testing. // *Haschek and Rousseaux's Handbook of Toxicologic Pathology*. 2013. Vol. 1. P. 565–594.
33. Paul, J., & Veenstra, T. D. Separation of Serum and Plasma Proteins for In-Depth Proteomic Analysis. // *MDPI. Separations*. 2022. Vol. 9, №4. e 9040089.
34. Kopylov, A. T., Ilgisonis, E. V., Moysa, A. A., Tikhonova, O. V., Zavialova, M. G., Novikova, S. E., Lisitsa, A. v., Ponomarenko, E. A., Moshkovskii, S. A., Markin, A. A., Grigoriev, A. I., Zgoda, V. G., & Archakov, A. I. Targeted Quantitative Screening of Chromosome 18 Encoded Proteome in Plasma Samples of Astronaut Candidates. // *Journal of Proteome Research*. 2016. Vol. 15, №11. P. 4039–4046.
35. Alm, T., von Feilitzen, K., Lundberg, E., Sivertsson, Å. & Uhlén, M. A chromosome-centric analysis of antibodies directed toward the human proteome using antibodypedia. // *Journal of Proteome Research*. 2014. Vol. 13, №3. P. 1669–1676.

36. Gosetti, F., Mazzucco, E., Zampieri, D., & Gennaro, M. C. Signal suppression/enhancement in high-performance liquid chromatography tandem mass spectrometry. // *Journal of chromatography*. 2010. A. Vol. 1217, №25. P. 3929–3937.
37. Ilgisonis, E. V., Kopylov, A. T., Ponomarenko, E. A., Poverennaya, E. V., Tikhonova, O.V., Farafonova, T. E., Novikova, S., Lisitsa, A. V., Zgoda, V. G., & Archakov, A. I. Increased Sensitivity of Mass Spectrometry by Alkaline Two-Dimensional Liquid Chromatography: Deep Cover of the Human Proteome in Gene-Centric Mode. // *Journal of Proteome Research*. 2018. Vol. 17, №12. P. 4258–4266.
38. Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., & Cooks, R. G. The Orbitrap: A new mass spectrometer. // *In Journal of Mass Spectrometry*. 2005. Vol. 40, №4. P. 430–443.
39. Zubarev, R. A., & Makarov, A. Orbitrap mass spectrometry. // *Analytical Chemistry*. 2013. Vol. 85, №11. P. 5288–5296.
40. Domon, B., & Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. // *Nature biotechnology*. 2010. Vol. 28, №7. P. 710–721.
41. Rappsilber, J., Ryder, U., Lamond, A. I., & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome research*. 2002. Vol. 12, №8. P. 1231–1245.
42. Ahn, E., Molzahn, L., Glatter, T., & Schmidt, A. Critical assessment of proteome-wide label-free absolute abundance estimation strategies. // *Proteomics*. 2013. Vol. 13, №17. P. 2567–2578.
43. Tang, H. Y., Beer, L. A., Chang-Wong, T., Hammond, R., Gimotty, P., Coukos, G., & Speicher, D. W. A xenograft mouse model coupled with in-depth plasma proteome analysis facilitates identification of novel serum biomarkers for human ovarian cancer. // *Journal of proteome research*. 2012. Vol. 11, №2. P. 678–691.
44. Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J., & Mann, M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. // *Nature methods*. 2018. Vol. 15, №6. P. 440–448.
45. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., & Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent

acquisition: a new concept for consistent and accurate proteome analysis. // Molecular & cellular proteomics: MCP. 2012. Vol. 11, №6. O111.016717.

46. Wang, J. H., Choong, W. K., Chen, C. T., & Sung, T. Y. Calibr improves spectral library search for spectrum-centric analysis of data independent acquisition proteomics. // Scientific reports. 2022. Vol. 12, №1. s41598-022-06026-9.

47. Tsou, C. C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A. C., & Nesvizhskii, A. I. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. // Nature methods. 2015. Vol. 12, №3. P. 258–264.

48. Budamgunta, H., Maes, E., Willems, H., Menschaert, G., Schildermans, K., Kumar, A. A., Boonen, K., & Baggerman, G. Multiple solvent elution, a method to counter the effects of coelution and ion suppression in LC-MS analysis in bottom up proteomics. // Journal of Chromatography. 2019. Vol. 1124. P. 256–264.

49. Strzelecka, D., Holman, S. W., & Eyers, C. E. Evaluation of dimethyl sulfoxide (DMSO) as a mobile phase additive during top 3 label-free quantitative proteomics. // International journal of mass spectrometry. 2015. Vol. 391. P. 157–160.

50. Hahne, H., Pachl, F., Ruprecht, B., Maier, S. K., Klaeger, S., Helm, D., Médard, G., Wilm, M., Lemeer, S., & Kuster, B. DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. // Nature Methods. 2013. Vol. 10, №10. P. 989–991.

51. Cifani, P., & Kentsis, A. High Sensitivity Quantitative Proteomics Using Automated Multidimensional Nano-flow Chromatography and Accumulated Ion Monitoring on Quadrupole-Orbitrap-Linear Ion Trap Mass Spectrometer. // Molecular & cellular proteomics: MCP. 2017. Vol. 16, №11. P. 2006–2016.

52. Geiger, T., Cox, J., & Mann, M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. // Molecular & cellular proteomics: MCP. 2010. Vol. 9, №10. P. 2252–2261.

53. Levy, M. J., Washburn, M. P., & Florens, L. Probing the Sensitivity of the Orbitrap Lumos Mass Spectrometer Using a Standard Reference Protein in a Complex Background. // Journal of proteome research. 2018. Vol. 17, №10. P. 3586–3592.

54. Glish, G. L., Mcluckey, S.-A., Ridley, T. Y., & Cooks, R.-G. A new “hybrid” sector/quadrupole mass spectrometer for mass spectrometry/mass spectrometry. // International Journal of Mass Spectrometry and Ion Physics. 1982. Vol. 41, №3. P. 157-177.
55. Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., Spiegelman, C. H., Zimmerman, L. J., Ham, A. J., Keshishian, H., Hall, S. C., Allen, S., Blackman, R. K., Borchers, C. H., Buck, C., Cardasis, H. L., Cusack, M. P., Dodder, N. G., Gibson, B. W., Held, J. M., Carr, S. A. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. // Nature biotechnology. 2009. Vol. 27, №7. P. 633–641.
56. Bisson, N., James, D. A., Ivosev, G., Tate, S. A., Bonner, R., Taylor, L., & Pawson, T. Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. // Nature biotechnology. 2011. Vol. 29, №7. P. 653–658.
57. Costenoble, R., Picotti, P., Reiter, L., Stallmach, R., Heinemann, M., Sauer, U., & Aebersold, R. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics. // Molecular systems biology. 2011. Vol. 7. an.464.
58. Kiyonami, R., Schoen, A., Prakash, A., Peterman, S., Zabrouskov, V., Picotti, P., Aebersold, R., Huhmer, A., & Domon, B. Increased selectivity, analytical precision, and throughput in targeted proteomics. // Molecular & cellular proteomics: MCP. 2011. Vol. 10, №2. M110.002931.
59. Muntel, J., Gandhi, T., Verbeke, L., Bernhardt, O. M., Treiber, T., Bruderer, R., & Reiter, L. Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. // Molecular omics. 2019. Vol. 15, №5. P. 348–360.
60. Dillen, L., Cools, W., Vereyken, L., Lorreyne, W., Huybrechts, T., de Vries, R., Ghobarah, H., & Cuyckens, F. Comparison of triple quadrupole and high-resolution TOF-MS for quantification of peptides. // Bioanalysis. 2012. Vol. 4, №5. P. 565–579.
61. Wu, L., Bache, N., & Bekker-Jensen, D. B. Robust and Reproducible Protein Quantification in Plasma using the Evosep One and the Agilent 6495 Triple Quadrupole LC/MS // Application note. 2023. P. 1-7.

62. Paul Kebarle and Liang Tang. From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. // *Analytical Chemistry*. 1993. Vol. 65, №22. P. 972-986.
63. Jung, E., Heller, M., Sanchez, J. C., & Hochstrasser, D. F. Proteomics meets cell biology: the establishment of subcellular proteomes. // *Electrophoresis*. 2000. Vol. 21, №16. P. 3369–3377.
64. Huber, L. A., Pfaller, K., & Vietor, I. Organelle proteomics: implications for subcellular fractionation in proteomics. // *Circulation research*. 2003. Vol. 92, №9. P. 962–968.
65. Dreger M. Subcellular proteomics. // *Mass spectrometry reviews*. 2003. Vol. 22, №1. P. 27–56.
66. de Araùjo, M. E., Huber, L. A., & Stasyk, T. Isolation of endocytic organelles by density gradient centrifugation. // *Methods in molecular biology*. 2008. Vol. 424. P. 317–331.
67. Graham, J. M., Ford, T., & Rickwood, D. Isolation of the major subcellular organelles from mouse liver using Nycodenz gradients without the use of an ultracentrifuge. // *Analytical biochemistry*. 1990. Vol. 187, №2. P. 318–323.
68. Graham, J. M., & Rickwood, D. Subcellular fractionation: a practical approach (No. 173). Oxford University Press. 1997. P. 1-360. P. 187.
69. Ramsby, M. L., Makowski, G. S., & Khairallah, E. A. Differential detergent fractionation of isolated hepatocytes: biochemical, immunochemical and two-dimensional gel electrophoresis characterization of cytoskeletal and noncytoskeletal compartments. // *Electrophoresis*. 1994. Vol. 15, №2. P. 265–277.
70. Abdolzade-Bavil, A., Hayes, S., Goretzki, L., Kröger, M., Anders, J., & Hendriks, R. Convenient and versatile subcellular extraction procedure, that facilitates classical protein expression profiling and functional protein analysis. // *Proteomics*. 2004. Vol. 4, №5. P. 1397–1405.
71. Hoffmann, P., Ji, H., Moritz, R. L., Connolly, L. M., Frecklington, D. F., Layton, M. J., Eddes, J. S., & Simpson, R. J. Continuous free-flow electrophoresis separation of cytosolic proteins from the human colon carcinoma cell line LIM 1215: a non two-dimensional gel

electrophoresis-based proteome analysis strategy. // *Proteomics*. 2001. Vol. 1, №7. P. 807–818.

72. Wagner, H., Free-flow electrophoresis. // *Nature*. 1989. Vol. 341. P. 669–670.

73. Weber, G., Islinger, M., Weber, P., Eckerskorn, C., & Völkl, A. Efficient separation and analysis of peroxisomal membrane proteins using free-flow isoelectric focusing. // *Electrophoresis*. 2004. Vol. 25, №12. P. 1735–1747.

74. Ackermann, B. L., & Berna, M. J. Coupling immunoaffinity techniques with MS for quantitative analysis of low-abundance protein biomarkers. // *Expert review of proteomics*. 2007. Vol. 4, №2. P. 175–186.

75. Fang, X., & Zhang, W. W. Affinity separation and enrichment methods in proteomic analysis. // *Journal of proteomics*. 2008. Vol. 71, №3. P. 284–303.

76. Kikuchi, M., Hatano, N., Yokota, S., Shimozawa, N., Imanaka, T., & Taniguchi, H. Proteomic analysis of rat liver peroxisome: presence of peroxisome-specific isozyme of Lon protease. // *The Journal of biological chemistry*. 2004. Vol. 279, №1. P. 421–428.

77. Morciano, M., Burré, J., Corvey, C., Karas, M., Zimmermann, H., & Volkandt, W. Immun isolation of two synaptic vesicle pools from synaptosomes: a proteomics analysis. // *Journal of neurochemistry*. 2005. Vol. 95, №6. P. 1732–1745.

78. Fratantoni, S. A., Piersma, S. R., & Jimenez, C. R. Comparison of the performance of two affinity depletion spin filters for quantitative proteomics of CSF: Evaluation of sensitivity and reproducibility of CSF analysis using GeLC-MS/MS and spectral counting. // *Proteomics*. 2010. Clinical applications, Vol. 4, №6-7. P. 613–617.

79. Albrethsen, J., Knol, J. C., Piersma, S. R., Pham, T. V., de Wit, M., Mongera, S., Carvalho, B., Verheul, H. M., Fijneman, R. J., Meijer, G. A., & Jimenez, C. R. Subnuclear proteomics in colorectal cancer: identification of proteins enriched in the nuclear matrix fraction and regulation in adenoma to carcinoma progression. // *Molecular & cellular proteomics: MCP*. 2010. Vol. 9, №5. P. 988–1005.

80. Piersma, S. R., Fiedler, U., Span, S., Lingnau, A., Pham, T. V., Hoffmann, S., Kubbutat, M. H., & Jiménez, C. R. Workflow comparison for label-free, quantitative secretome

proteomics for cancer biomarker discovery: method evaluation, differential analysis, and verification in serum. // Journal of proteome research. 2010. Vol. 9, №4. P. 1913–1922.

81. Mena, M. L., Moreno-Gordaliza, E., Moraleja, I., Cañas, B., & Gómez-Gómez, M. M. (2011). OFFGEL isoelectric focusing and polyacrylamide gel electrophoresis separation of platinum-binding proteins. // Journal of chromatography. A, Vol. 1218, №9. P. 1281–1290.

82. Carrette, O., Burkhard, P. R., Sanchez, J. C., & Hochstrasser, D. F. State-of-the-art two-dimensional gel electrophoresis: a key tool of proteomics research. // Nature protocols. 2006. Vol. 1, №2. P. 812–823.

83. Ernoul, E., Gamelin, E., & Guette, C. Improved proteome coverage by using iTRAQ labelling and peptide OFFGEL fractionation. // Proteome science. 2008. Vol. 6, e27.

84. Holland, J. W., Deeth, H. C., & Alewood, P. F. Proteomic analysis of kappa-casein micro-heterogeneity. // Proteomics. 2004. Vol. 4, №3. P. 743–752.

85. Hirtz, C., Chevalier, F., Centeno, D., Rofidal, V., Egea, J. C., Rossignol, M., Sommerer, N., & Deville de Périère, D. MS characterization of multiple forms of alpha-amylase in human saliva. // Proteomics. 2005. Vol. 5, №17. P. 4597–4607.

86. Schulenberg, B., Goodman, T. N., Aggeler, R., Capaldi, R. A., & Patton, W. F. Characterization of dynamic and steady-state protein phosphorylation using a fluorescent phosphoprotein gel stain and mass spectrometry. // Electrophoresis. 2004. Vol. 25, №15. P. 2526–2532.

87. Weston A, Brown PR. HPLC and CE – Principles and Practice. Academic Press. 1997. P. 1-280. P. 195.

88. McMaster, M. C. HPLC: a practical user's guide. John Wiley & Sons. 2007. P. 1-256. P. 140.

89. Essader, A. S., Cargile, B. J., Bundy, J. L., & Stephenson, J. L., Jr. A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. // Proteomics. 2005. Vol. 5, №1. P. 24–34.

90. Dong, M. W. Modern HPLC for practicing scientists. John Wiley & Sons. 2006. P. 1-304. P. 50.

91. McCalley D. V. The challenges of the analysis of basic compounds by high performance liquid chromatography: some possible approaches for improved separations. // Journal of chromatography. 2010. Vol. 1217, №6. P. 858–880.
92. Gilar, M., Olivova, P., Daly, A. E., & Gebler, J. C. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. // Journal of separation science. 2005. Vol. 28, №14. P. 1694–1703.
93. Toll, H., Oberacher, H., Swart, R., & Huber, C. G. Separation, detection, and identification of peptides by ion-pair reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry at high and low pH. // Journal of chromatography. 2005. A, Vol. 1079, №1-2. P. 274–286.
94. Delmotte, N., Lasaosa, M., Tholey, A., Heinzle, E., & Huber, C. G. Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. // Journal of proteome research. 2007. Vol. 6, №11. P. 4363–4373.
95. Mostovenko, E., Hassan, C., Rattke, J., Deelder, A. M., van Veelen, P. A., & Palmblad, M. (2013). Comparison of peptide and protein fractionation methods in proteomics. // EuPA Open Proteomics. 2013. Vol. 1. P. 30–37.
96. Deinichenko, K. A., Krasnov, G. S., Radko, S. P., Ptitsyn, K. G., Shapovalova, V. V., Timoshenko, O. S., Ponomarenko, E. A. Human CHR18: “Stakhanovite” Genes, Missing and uPE1 Proteins in Liver Tissue and HepG2 Cells. // Biomedical Chemistry: Research and Methods. 2021. Vol. 4, №1. e00144.
97. Cox, J., & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. // Nature Biotechnology. 2008. Vol. 26, №12. P. 1367–1372.
98. Elias, J. E., & Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. // Methods in molecular biology. 2010. Vol. 604, P. 55–71.
99. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M. (2011). Global quantification of mammalian gene expression control. // Nature. 2011. Vol. 473, №7347. P. 337–342.

100. Hood CA, Fuentes G, Patel H, Page K, Menakuru M, Park JH. Fast conventional Fmoc solid-phase peptide synthesis with HCTU. // *J Pept Sci*. 2008. Vol. 14, №1. P. 97-101.
101. Pino, L. K., Searle, B. C., Bollinger, J. G., Nunn, B., MacLean, B., & MacCoss, M. J. The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. // *Mass spectrometry reviews*. 2020. Vol. 39, №3. P. 229–244.
102. Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., Deng, C., Varusai, T., Ragueneau, E., Haider, Y., May, B., Shamovsky, V., Weiser, J., Brunson, T., Sanati, N., D'Eustachio, P. (2022). The reactome pathway knowledgebase 2022. // *Nucleic Acids Research*. 2020. Vol. 50, №D1. P. 687–692.
103. Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N., & Moore, M. J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. // *RNA*. 2002. Vol. 8, №4. P. 426–439.
104. Nufer, O., Kappeler, F., Guldbrandsen, S., & Hauri, H. P. ER export of ERGIC-53 is controlled by cooperation of targeting determinants in all three of its domains. // *Journal of cell science*. 2003. Vol. 116, №21. P. 4429–4440.
105. Qi, B., & Han, M. Microbial Siderophore Enterobactin Promotes Mitochondrial Iron Uptake and Development of the Host via Interaction with ATP Synthase. // *Cell*. 2018. Vol. 175, №2. P. 571–582.
106. Lee, S. W., Bonnah, R. A., Higashi, D. L., Atkinson, J. P., Milgram, S. L., & So, M. CD46 is phosphorylated at tyrosine 354 upon infection of epithelial cells by *Neisseria gonorrhoeae*. // *The Journal of cell biology*. 2002. Vol. 156, №6. P. 951–957.
107. Shan, J., Wang, P., Zhou, J., Wu, D., Shi, H., & Huo, K. RIOK3 interacts with caspase-10 and negatively regulates the NF-kappaB signaling pathway. // *Molecular and cellular biochemistry*. 2009. Vol. 332, №1-2. P. 113–120.
108. Ishizaki, T., Maekawa, M., Fujisawa, K., Okawa, K., Iwamatsu, A., Fujita, A., Watanabe, N., Saito, Y., Kakizuka, A., Morii, N., & Narumiya, S. The small GTP-binding protein Rho binds to and activates a 160 kDa Ser/Thr protein kinase homologous to myotonic dystrophy kinase. // *The EMBO journal*. 1996. Vol. 15, №8. P. 1885–1893.

109. Reimegård, J., Tarbier, M., Danielsson, M., Schuster, J., Baskaran, S., Panagiotou, S., Dahl, N., Friedländer, M. R., & Gallant, C. J. A combined approach for single-cell mRNA and intracellular protein expression analysis. // *Communications Biology*. 2021. Vol. 4. an.624 (2021).
110. Shyu AB1, Greenberg ME, Belasco JG. The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. // *Genes Dev*. 1989. Vol. 3, №1. P. 60-72.
111. Treisman R. Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5' element and c-fos 3' sequences. // *Cell*. 1985. Vol. 42, №3. P. 889-902.
112. Mathieson, T., Franken, H., Kosinski, J. et al. Systematic analysis of protein turnover in primary cells. // *Nat Commun*. 2018. Vol. 9, №1. P. 1-10.
113. Bevilacqua, Annamaria & Ceriani, Maria & Canti, Gianfranco & Asnaghi, Laura & Gherzi, Roberto & Brewer, Gary & Papucci, Laura & Schiavone, Nicola & Capaccioli, Sergio & Nicolini, Angelo. Bcl-2 Protein Is Required for the Adenine/Uridine-rich Element (ARE)-dependent Degradation of Its Own Messenger. // *The Journal of biological chemistry*. 2003. Vol. 278, №26. P. 23451-23459.
114. Edward Yang, Erik van Nimwegen, Mihaela Zavolan, Nikolaus Rajewsky, Mark Schroeder, Marcelo Magnasco and, James E. Darnell, Jr. Decay Rates of Human mRNAs: Correlation With Functional Characteristics and Sequence Attributes. // *Genome Res*. 2003. Vol. 13, №18. P. 1863-1872.
115. Racanelli, V., & Rehmann, B. The liver as an immunological organ. // *Hepatology*. 2006. Vol. 43, №2. P. 54–62.
116. Jenne, C. N., & Kubes, P. Immune surveillance by the liver. // *Nature immunology*. 2013. Vol. 14, №10. P. 996–1006.
117. Bilzer, M., Roggel, F., & Gerbes, A. L. Role of Kupffer cells in host defense and liver disease. // *Liver international: official journal of the International Association for the Study of the Liver*. 2006. Vol. 26, №10. P. 1175–1186.
118. Geerts A. History, heterogeneity, developmental biology, and functions of quiescent hepatic stellate cells. // *Seminars in liver disease*. 2001. Vol. 21, №3. P. 311–335.

119. Wisse, E., Braet, F., Luo, D., De Zanger, R., Jans, D., Crabbé, E., & Vermoesen, A. Structure and function of sinusoidal lining cells in the liver. // *Toxicologic pathology*. 1996. Vol. 24, №1. P. 100–111.
120. Agafonov DE, Kolb VA, Spirin AS. Ribosome-associated protein that inhibits translation at the aminoacyl-tRNA binding stage. // *EMBO Rep*. 2001. Vol. 2, №5. P. 399-402.
121. Watson, C. N., Belli, A., & Di Pietro, V. Small non-coding RNAs: New class of biomarkers and potential therapeutic targets in neurodegenerative disease. // *Frontiers in Genetics*. 2019. Vol. 10. P. 1–14.
122. Helen A. King, André P. Gerber. Translatome profiling: methods for genome-scale analysis of mRNA translation // *Briefings in Functional Genomics*. 2016. Vol. 15, №1. P. 22–31.

8. БЛАГОДАРНОСТИ

Автор благодарит руководителя и коллег за помощь в проведении работы и подготовке диссертации. В частности, автор благодарен д.б.н., проф. РАН Згоде В.Г. за помощь в проведении масс-спектрометрического анализа и в обработке данных, 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. Калужского Л.К. за предоставление материала *E.Coli*. и коллег из ИБХ, которые предоставили 20 клеточных линий для протеомного анализа.

9. ФИНАНСИРОВАНИЕ

Работа была поддержана грантом Российского научного фонда № 20-15-00410 под руководством академика РАН Арчакова А.И.