

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

На правах рукописи

ЗОРИНА ЕЛЕНА СЕРГЕЕВНА

ПРОТЕОФОРМНОЕ ПРОФИЛИРОВАНИЕ ТКАНИ ПЕЧЕНИ ЧЕЛОВЕКА В
НОРМЕ И ПРИ ГЕПАТОЦЕЛЛЮЛЯРНОМ РАКЕ С ИСПОЛЬЗОВАНИЕМ
ДВУМЕРНОГО ГЕЛЬ-ЭЛЕКТРОФОРЕЗА И МАСС-СПЕКТРОМЕТРИИ

1.5.4. – Биохимия

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата биологических наук

Научный руководитель:
доктор биологических наук
Нарыжный Станислав Николаевич

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	4
1. ОБЗОР ЛИТЕРАТУРЫ	9
1.1. Протеоформы	9
1.2. Протеом печени человека	11
1.3. Транскриптом печени человека.....	13
1.4. Интеграция транскриптомных и протеомных исследований	15
1.5. Роль протеоформ при различных заболеваниях.....	18
1.6. Современные методы диагностики ГЦР	19
1.6.1. Биомаркеры ГЦР в клинической практике	20
1.6.2. Молекулярно-генетические маркеры на основе геномного секвенирования.....	22
1.7. Модели клеточных линий гепатоцеллюлярного рака и их сравнение.....	23
1.8. Методы исследования белков для изучения разнообразия протеоформ	25
1.8.1. Масс-спектрометрические методы.....	26
1.8.2. Двумерный гель-электрофорез (2DE)	29
1.8.3. Иммуноблоттинг и иммунохимические методы.....	32
1.8.4. Биоинформатический анализ	33
1.8.4.1. Базы данных протеоформ человека	36
1.9. Заключение литературного обзора	39
2. МАТЕРИАЛЫ И МЕТОДЫ	41
2.1. Характеристика образцов.....	42
2.1.1. Клеточная культура HepG2	42
2.1.2. Образцы ткани печени	42
2.2. Предварительная подготовка образцов	43
2.2.1. Экстракция белков из клеток HepG2.....	43
2.2.2. Экстракция белков из ткани печени человека.....	43
2.3. Подготовка проб для протеомного профилирования	44
2.3.1. Протокол подготовки проб с использованием центрифужных концентраторов.....	44
2.3.2. Двумерный гель-электрофорез	45
2.4. Гидролитическое расщепление белков	47
2.4.1. Гидролиз с использованием центрифужных концентраторов	47
2.4.2. Гидролиз белков из геля	48
2.5. Масс-спектрометрический анализ (LC-ESI-MS/MS)	48
2.6. Биоинформатический анализ данных.....	49
2.7. Нормализация данных секционного протеомного профилирования	50
2.8. Анализ и статистическая обработка данных	51

3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ	52
3.1. Протеомное профилирование белков нормальной ткани печени и белков клеточной линии HepG2.....	52
3.1.1. Панорамное протеомное профилирование	52
3.3.2. Секционное протеомное профилирование	55
3.3.3. Комбинирование результатов секционного и панорамного протеомного профилирования	60
3.4. Протеомное профилирование белков в образцах пациентов с ГЦР	63
3.4.2. Панорамное протеомное профилирование	64
3.4.3. Секционное протеомное профилирование	69
3.2.3. Комбинирование результатов панорамного и секционного протеомного профилирования	72
3.5. Сравнительный анализ белков клеточной линии HepG2, нормальной ткани печени и образцов ГЦР.....	77
3.5.2. ПТМ потенциальных биомаркеров ГЦР	81
3.3.2. Функциональная аннотация потенциальных биомаркеров ГЦР.....	92
ЗАКЛЮЧЕНИЕ	96
ВЫВОДЫ	99
ФИНАНСИРОВАНИЕ.....	100
БЛАГОДАРНОСТИ	101
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.....	102
СПИСОК ТЕРМИНОВ	103
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	104
ПРИЛОЖЕНИЕ 1. Описание образцов ткани печени.....	118
ПРИЛОЖЕНИЕ 2. Полный список белков идентифицированных двумя методами – секционным и панорамным протеомным профилированием.....	119
ПРИЛОЖЕНИЕ 3. Двумерные электрофореграммы образцов от пациентов с ГЦР	120
ПРИЛОЖЕНИЕ 4. Сравнение данных, полученных для 52 белков с наиболее измененными паттернами протеоформ с базой данных Plasma Proteome Database (PPD).....	122
ПРИЛОЖЕНИЕ 5. Тысяча типичных паттернов протеоформ, полученных для всех типов анализированных образцов (клеточная линия HepG2, нормальная ткань печени, опухолевая ткань печени, контрольная ткань печени).....	126

ВВЕДЕНИЕ

Актуальность темы исследования

Гепатоцеллюлярный рак (ГЦР), или гепатоцеллюлярная карцинома (ГЦК), является первичным раком печени, то есть формируется из печеночных клеток (гепатоцитов) и входит в число шести наиболее часто диагностируемых злокачественных опухолевых заболеваний печени. При этом общая пятилетняя выживаемость коррелирует со стадией заболевания и при более поздних стадиях составляет не более 20 % [1]. К основным факторам риска развития ГЦР относят цирроз печени, вирусные гепатиты В и С, алкогольный и неалкогольный стеатогепатит, ряд наследственных заболеваний (дефицит альфа-1-антитрипсина, тирозинемии и гематохроматоз), прием лекарственных препаратов, содержащих стероидные гормоны, вызывающие токсические повреждения печени [2]. Помимо перечисленных выше заболеваний важным фактором риска развития ГЦР является жировая болезнь печени, которая может быть вызвана чрезмерным употреблением алкоголя (алкогольная болезнь печени) или другими факторами (неалкогольная болезнь печени). В настоящее время наблюдается общемировая тенденция увеличения заболеваемости ГЦР на фоне неалкогольной жировой болезни печени (НАЖБП) [3–5].

В мировой клинической практике для скрининга групп риска, течения заболевания и прогноза выживаемости используется опухолеспецифичный маркер – альфа-фетопроtein (АФП). Предложенный Ю.С. Татариновым в 1964 году альфа-фетопроteinовый тест для выявления сывороточных фракций эмбриональных альфа-глобулинов в качестве диагностического маркера ГЦР используется до сих пор в клинической практике [6]. Специфичность АФП составляет примерно 90% при 60% чувствительности, что объясняется повышением уровня АФП у части пациентов без признаков ГЦР и наоборот. Тем не менее, АФП остается широко используемым биомаркером из-за его низкой стоимости, простоты измерения и широкой доступности.

Следует отметить, что ГЦР выявляется также у пациентов, не входящих в группу риска, а чувствительность и специфичность используемых маркеров оказывается недостаточной как для ранней диагностики, так и для широкого скрининга. В связи с этим актуален поиск новых, более специфичных и чувствительных маркеров, на роль которых претендуют в том числе и белки, содержание которых повышается или понижается в опухолях.

Особенностью значительной части белков человека является их существование в нескольких или многих модификациях – протеоформах. Например, упоминавшийся выше АФП существует в нескольких вариантах гликозилированных форм (протеоформах). При этом большую чувствительность и специфичность (по некоторым исследованиям до 95%) показывают применяемые в клинической практике тесты с измерением фукозилированной формы АФП (АФП-L3) [7]. Еще одним маркером, который также присутствует у 50-60% пациентов с ГЦР, является дез-гамма-карбокси-протромбин (ДГП или PIVKA-II). Он представляет собой аномальный протромбин, который образуется при дефиците витамина К или нарушении его метаболизма в печени. При ГЦР клетки печени теряют способность карбоксилировать протромбин, что приводит к повышению уровня в крови его немодифицированной формы [8]. Рассмотренные примеры по АФП и ДГП, показывают, что измерение содержания конкретных протеоформ обеспечивают лучшие результаты анализа [7,9]. При этом, комбинация АФП, ДГП и АФП-L3 позволяет достигнуть чувствительности 94% и специфичности более 97% [10].

Основная сложность изучения протеоформ состоит в том, что для получения информации о них с использованием масс-спектрометрических (МС) методов необходимы целые очищенные белки. Поэтому наибольший интерес представляет сочетание различных методов разделения белков с последующей МС. За последние десятилетия использование комбинации двумерного электрофореза (two-dimensional electrophoresis, 2DE) с панорамной масс-спектрометрией стало перспективным подходом для изучения протеоформ, в

котором эффективно реализуются возможности данных методов. В результате, для каждого из представленных белков в образце, за счет предварительного фракционирования, можно получить информацию о протеоформных паттернах при различных состояниях. Сравнительный анализ таких паттернов в норме и при ГЦР, позволяет сократить трудозатраты для поиска специфических протеоформ как потенциальных онкомаркеров. Поэтому получение информации о протеоформах и их профилях особенно актуально, учитывая многочисленные изменения внутри протеома человека при раковых заболеваниях.

Цель и задачи исследования

Целью данного исследования являлось на основании различия уровней белков, а также паттернов их протеоформ определить белковые сигнатуры, специфичные для злокачественных клеток печени человека.

Для достижения данной цели были поставлены следующие задачи:

1. Получить протеомные профили для нормальной ткани печени и злокачественных клеток печени, используя методы панорамного и секционного протеомного профилирования.
2. Определить специфичные и общие протеоформные паттерны для всех проанализированных типов образцов.
3. Провести сравнительный анализ результатов панорамного и секционного протеомного профилирования с целью выявления специфичных для опухолевой ткани печени белковых сигнатур.

Научная новизна работы

В работе был применен разработанный в нашей лаборатории подход к инвентаризации протеоформ различных биологических образцов, а именно секционное протеомное профилирование, которое включает в себя 2DE с последующей панорамной масс-спектрометрией. При этом в анализ попадают все разделенные 2DE белки, а не только те, которые визуализированы окрашиванием.

Это позволяет максимально детализировать молекулярный состав изучаемого образца и получить данные о протеомных профилях и их протеоформных паттернах. Данный подход может использоваться как самостоятельный метод поиска протеоформных паттернов, так и в качестве дополнения к данным масс-спектрометрического протеомного анализа. Для образцов гепатоцеллюлярной опухоли, полученных от пациентов, были показаны количественные и качественные изменения как на уровне протеомных профилей, так и на уровне протеоформных паттернов. Предложен ряд потенциальных прогностических белков-маркеров, составляющих специфичные белковые сигнатуры опухолевой ткани печени при ГЦР.

Теоретическая и практическая значимость работы

В проекте для инвентаризации и визуального представления протеомных профилей по различным образцам (нормальным и раковым тканям и клеточным линиям) в дополнение к панорамному протеомному профилированию используется секционное протеомное профилирование. Практическая значимость данного исследования заключается в том, что с помощью такого подхода можно получить не только протеомный профиль, характеризующий тип образца, но и более подробную информацию о перестройках протеома на уровне протеоформ, которые происходят при канцерогенезе, что может быть использовано для идентификации новых биомаркеров.

Положения, выносимые на защиту:

1. Использование секционного протеомного профилирования значительно увеличивает количество обнаруженных белков по сравнению с панорамным протеомным профилированием.
2. Секционное протеомное профилирование позволяет детализировать молекулярный состав исследуемого типа материала с учетом физико-химических параметров белковых продуктов одного гена.

3. Существуют специфичные протеоформные паттерны, которые могут быть использованы для протеотипирования образцов в норме и при патологии.

Личный вклад автора

Автор диссертации принимала непосредственное участие в планировании экспериментов, подготовке проб для секционного 2DE и протеомного анализа данных, биоинформатической интерпретации результатов и подготовке публикаций. Работа была выполнена в лаборатории анализа постгеномных данных ИБМХ в период с 2017 по 2024 год.

Степень достоверности и апробация результатов

Степень достоверности диссертации обеспечена использованием современных методов исследования. Основные положения работы были опубликованы в рецензируемых научных изданиях, а также были представлены в виде постерных докладов на международной конференции «Клиническая протеомика. Постгеномная медицина» (ClinProt 2017, Москва), молодежном научном форуме «Open Science» (Гатчина, 2017), конгрессе Европейской протеомной ассоциации (EuPA 2019, Потсдам), IV Всероссийской конференции с международным участием «Опухолевые маркеры» фундаментальные и клинические аспекты (Горно-Алтайск, 2023). Устные сообщения диссертантом представлялись на молодежном научном форуме «Open Science» (Гатчина, 2018), международном конгрессе «Биотехнология: состояние и перспективы развития. Науки о жизни» (Москва, 2018), на II и III Объединенном научном форуме VI и VII съезда физиологов СНГ, VI и VII съезда биохимиков России, IX и X Российском симпозиуме «Белки и пептиды» (Сочи, 2019 и 2022), IV Сеченовском международном биомедицинском саммите (SIBS-2020), Сеченовский Университет (Москва, 2020), Всероссийской конференции с международным участием «Биомедицинская химия: наука и практика» (Москва, 2024).

По теме диссертации опубликовано 15 печатных работ – 5 статей в рецензируемых научных журналах и 10 в трудах конференций.

1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Протеоформы

Протеом человека является динамической системой, которая может изменяться в зависимости от различных факторов, включая генетические вариации, окружающую среду, возраст и состояние здоровья. По данным проекта «Протеом человека», обнаруженное количество генов, кодирующих белки, составляет 19823 (по состоянию на июнь 2024 года). Согласно теоретическим предсказаниям, весь протеом человека оценивается диапазоном от сотен тысяч до нескольких миллионов различных белковых молекул – протеоформ [11–13]. Но при этом остается открытым вопрос – какая часть этих протеоформ образуется на самом деле, как они могут распределяться по клеткам и какие функциональные последствия они имеют?

Как было предложено в 2013 году консорциумом «Top-down proteomics» (CTDP), термин «протеоформа» должен использоваться для белкового продукта одного гена, содержащего изменения, вызванные за счет однонуклеотидного полиморфизма (ОНП) или альтернативного сплайсинга или посттрансляционных модификаций (ПТМ) [14]. Таким образом, протеоформа представляет собой полипептид, имеющий заданную аминокислотную последовательность с определенным набором ПТМ или их отсутствием. На Рисунке 1 представлено схематическое образование протеоформ.

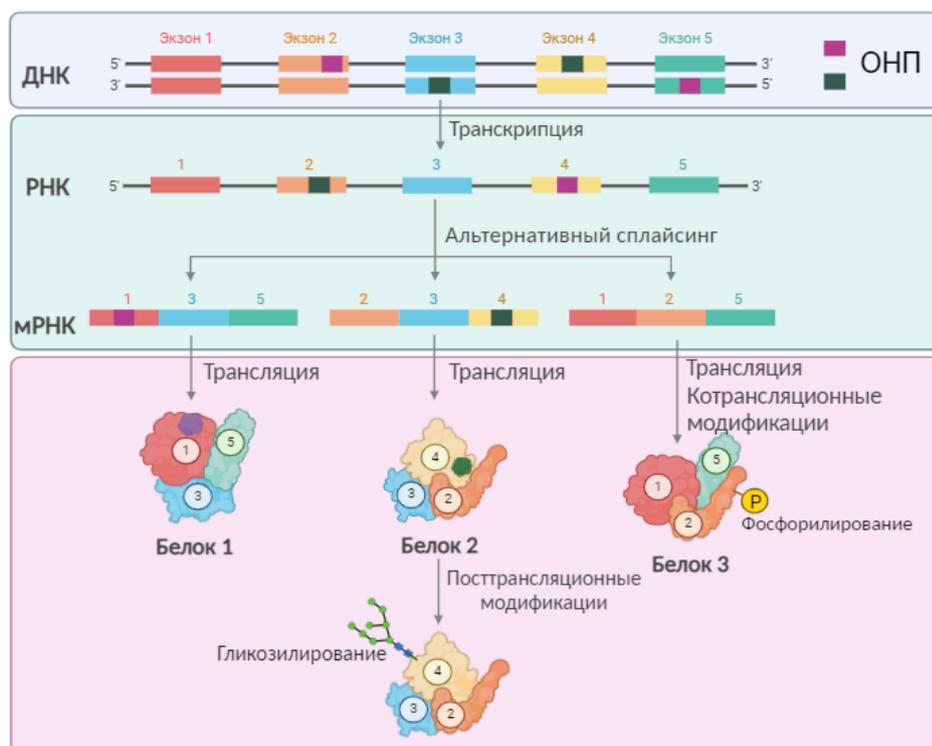


Рисунок 1. Схематическое представление образования протеоформ. Адаптировано из шаблона «Gene Splicing» от BioRender.com (2024).
ОНП – однонуклеотидный полиморфизм.

Согласно центральной догме молекулярной биологии, сформированной Френсисом Криком в 1958 году, информацию, закодированную в нуклеиновых кислотах (ДНК или РНК), нельзя обратить или перенести обратно из белка в нуклеиновую кислоту. Это предполагает, что переход генетической информации является однонаправленным от ДНК к РНК и белку. Следовательно, ДНК транскрибируется в РНК, а РНК транслируется в белок [15]. Во время трансляции образовывается множество функционально различающихся белков, но при этом часто встречается, что один и тот же белок может выполнять различные функции в клетке одного организма путем образования ПТМ. Таким образом, ПТМ это процесс регуляции на белковом уровне, который может происходить как в ходе трансляции (т.н. котрансляционные), так и после сборки и транспортировки в место локализации для активации или инактивации каталитической активности белка. В общедоступных базах данных (к примеру, SwissProt) количество сайтов ПТМ для одного белка может варьироваться от 0 до более 90 сайтов. Однако распределение ПТМ неравномерно – около 75% записей содержат два или меньше

аннотированных ПТМ, и только пять записей имеют более 90 аннотированных ПТМ. Всего известно более 400 различных типов ПТМ [16]. Среди них самые распространенные: фосфорилирование остатков серина, треонина и тирозина, ацетилирование, метилирование и убиквитинирование лизина. Наиболее часто ПТМ происходят в боковых цепях аминокислот за счет ферментов, составляющих около 5% протеома, которые добавляют или удаляют функциональные группы, белки, липиды или углеводы. Также ПТМ возникают за счет ограниченного протеолиза, при котором происходит расщепление пептидных связей с целью удаления специфических последовательностей или регуляторных субъединиц. Многие белки могут также модифицироваться через автокаталитические домены. Все эти механизмы способствуют разнообразию протеоформ и обеспечивают белкам большую гибкость и функциональную разнообразность [17].

Следует отметить, большинство методов протеомного анализа, используемых на сегодняшний день, не могут в достаточной мере одновременно различать все протеоформы, присутствующие в исследуемом типе материала, либо могут различать лишь небольшую их часть. Такие ограничения могут быть обусловлены как аналитическим пределом обнаружения, так и тем, что при анализе сложных белковых смесей, очень непросто обнаружить множество различных ПТМ не направленными методами исследований, к примеру, такими, как использование синтетических изотопно-меченых пептидов в качестве внутренних стандартов при мониторинге множественных реакций (multiple reaction monitoring, MRM).

1.2. Протеом печени человека

Значительные успехи в области изучения протеома печени человека стали более систематическими и организованными в рамках проекта «Протеом человека», хотя исследования отдельных аспектов протеома печени человека начались гораздо раньше в рамках различных научных исследований и проектов, которые начали активно развиваться в начале 2000-х годов. Примером одного из ранних проектов, сфокусированных на изучении протеома печени человека,

может служить проект «Протеом печени человека» (Human liver proteome project, HLPP) [18]. В его рамках была создана база данных протеома печени человека (Human Liver Proteome Database, HLPD), содержащая количественную информацию о белках четырех основных типов клеток печени: гепатоцитов, синусоидальных эндотелиальных клеток, звездчатых клеток печени и клеток Купфера. В дополнение к базовым протеомным данным HLPD включает данные пациентов с циррозом печени и неалкогольным стеатогепатитом (НАСГ), демонстрирующие нарушение регуляции в печени и плазме как на белковом уровне, так и на уровне метаболических путей, что может указывать на области потенциального терапевтического вмешательства [19].

Как отмечалось ранее, одними из факторов риска развития ГЦР являются вирусы гепатита В и С. Риск развития при их наличии составляет 54% и 31%, соответственно [20]. В исследовании Kim и соавторов (2003), было показано что при сравнительном анализе с помощью комбинации 2DE и времяпролетной масс-спектрометрией с матрично-активированной лазерной десорбцией/ионизацией (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, MALDI-TOF MS) имеются различия между образцами пациентов с ГЦР, вызванным вирусом гепатита В от ГЦР, вызванным вирусом гепатита С [21]. Кроме того, был выявлен ряд белков, уровни экспрессии которых существенно различались между опухолевой и неопухолевой тканью, не связанной с вирусной этиологией. Это свидетельствует о том, что протеомный состав опухоли ГЦР тесно связан с этиологическими факторами (Рисунок 2) [21].

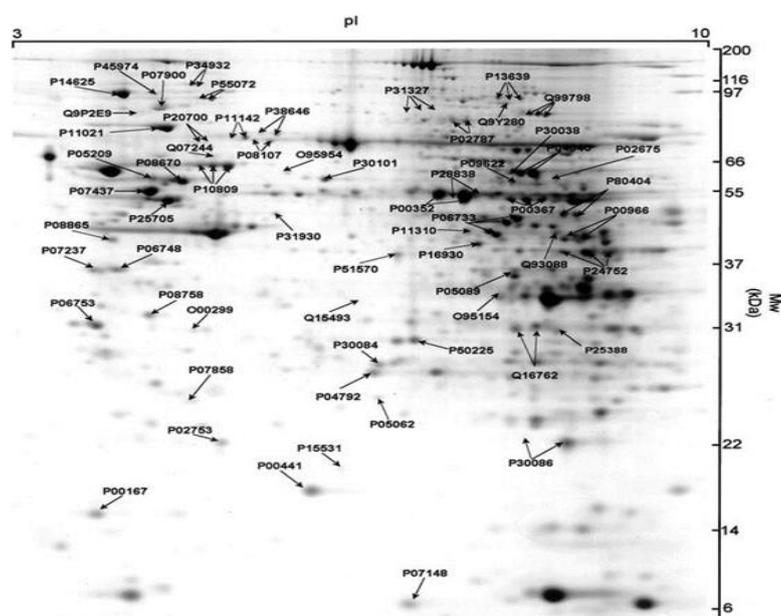


Рисунок 2. 2DE-карта белков печени человека, полученная для опухолевой ткани ГЦР без вирусной этиологии. Белковые пятна, демонстрирующие значительные изменения, отмечены стрелкой и номером доступа SWISS-PROT [21].

По оси абсцисс – изоэлектрическая точка (pI), ось ординат – молекулярная масса (кДа).

1.3. Транскриптом печени человека

Анализ транскриптома печени человека позволяет получить важную информацию об уровнях экспрессии и регуляции генов при различных состояниях, которая в дальнейшем используются для выявления генов или наборов генов, отличающихся по уровням экспрессии между нормальным и патологическим состоянием. Таким образом, дифференциально экспрессируемые гены (ДЭГ) могут служить как потенциальными биомаркерами для диагностики, прогнозирования или мониторинга эффективности лечения, так и представлять ценную информацию о молекулярных механизмах развития и прогрессирования заболевания.

Существует ряд международных инициатив, таких как ENCODE (Encyclopedia of DNA Elements) и GTEx (Genotype-Tissue Expression), в рамках которых активно секвенируют транскриптомы различных тканей множества организмов, в том числе человека, создавая обширные наборы транскриптомных данных. Помимо международных инициатив существуют и открытые репозитории данных, таких как Gene Expression Omnibus (GEO) и ArrayExpress,

которые активно используются для дальнейших биоинформатических исследований. Ниже приведены примеры биоинформатических работ, в которых были использованы данные из репозитория GEO для выявления ДЭГ.

В работе Li и соавторов (2020) были использованы данные о 320 образцах ГЦР и 270 образцах нормальных тканей печени. Суммарно им удалось идентифицировать 89 ДЭГ (31 ген с повышенной и 58 с пониженной регуляцией) [22]. Функциональная характеристика ДЭГ по трем категориям: биологические процессы, молекулярные функции и клеточная локализация позволила выявить, что большее количество генов было включено в тримеризацию коллагеновых цепей, сборку коллагеновых фибрилл и деградацию коллагена. Были определены клеточные компоненты, которые являются важными факторами развития и прогрессирования заболевания. Одним из них является внеклеточный матрикс, обеспечивающий физический каркас для окружающих клеток, связывания факторов роста и регулирования поведения клеток. Также были выявлены ключевые сигнальные пути, участвующие в канцерогенезе, а именно пути WNT/ β -катенин, путь клеточного цикла p53, окислительный стресс, пути PI3K/AKT/MTOR и RAS/RAF/MAPK. Анализ белок-белковых взаимодействий выявил пять белок-кодирующих генов с повышенной экспрессией в образцах ГЦР, а именно, остеопонтин (*SPP1*), цепь коллагена альфа-2(I) (*COL1A2*), инсулиноподобный фактор роста I (*IGF1*), липопротеин A (*LPA*) и галектин-3 (*LGALS3*). Авторы полагают, что данные белок-кодирующие гены следует рассматривать в качестве потенциальных биомаркеров ГЦР и возможных мишеней для терапевтического воздействия [22].

Известно, что к основным факторам риска развития ГЦР относится цирроз печени, поэтому наборы данных GEO также могут быть использованы для изучения молекулярных механизмов преобразования цирроза печени в ГЦР и выявления ключевых генов данных процессов. Например, при исследовании наборов данных микрочипов мРНК GSE89377, GSE17548, GSE63898 и GSE54236 было идентифицировано 58 ДЭГ между ГЦР и циррозом печени. Среди этих ДЭГ

были определены три т.н. хаб-гена: ингибитор циклинзависимой киназы 3 (*CDKN3*), цитохром P450 2C9 (*CYP2C9*) и фосфатидилхолин-стерол ацилтрансфераза (*LCAT*), являющиеся наиболее важными в ходе трансформации цирроза печени в ГЦР. Экспрессия гена *CDKN3* и кодируемого им белка *CDKN3* значительно увеличивается в тканях ГЦР по сравнению с циррозом и нормальными тканями, а также повышается при анапластической олигоастроцитоме, лейкемии и саркоме. Уровни *CDKN3* коррелируют со степенью развития опухоли, наличием/отсутствием инфекции вирусом гепатита, микросателлитами и сосудистой инвазией у пациентов с ГЦР. *CDKN3* влияет на пути связанные с клеточным циклом, репликацией ДНК, мейозом ооцитов и репарацией ошибок, регулируя клеточный цикл от фазы G1 к фазе S при ГЦР [23].

В исследовании 2022 года был проведен сравнительный анализ профилей экспрессии генов в 81 образцах, полученных от пациентов с ГЦР и 10 здоровых доноров [24]. Выполненный на наборе данных микрочипов мРНК GSE62232 анализ позволил идентифицировать 598 ДЭГ, включая десять хаб-генов (*CCNB1*, *CCNB2*, *CDK1*, *BUB1*, *NDC80*, *BUB1B*, *NCAPG*, *MAD2L1*, *CDC20* и *CCNB2*). Функциональный анализ генов и кодируемых ими белков, выявил их участие в различных биологических процессах, связанных с регуляцией клеточного цикла, митозом, организацией хромосом и делением ядра. На мышцах инбредной линии BALB/c было показано, что нокаут гена *CCNB1* с помощью микроРНК-144 ингибирует миграцию, инвазию и пролиферацию клеток ГЦР, что делает его потенциальной терапевтической мишенью для лечения ГЦР. При анализе выживаемости высокая экспрессия белков, кодируемых генами *CCNB1*, *CCNB2*, *CDK1*, *BUB1*, *BUB1B*, *NCAPG*, *MAD2L1* и *CDC20* коррелирует со снижением общей выживаемости у пациентов с ГЦР [24].

1.4. Интеграция транскриптомных и протеомных исследований

Одним из важных аспектов омиксных исследований является их взаимная интеграция. Примером базы данных, которая сочетает транскриптомные и протеомные данные, является «The Human Protein Atlas (HPA)». Она содержит не

только транскриптомные, протеомные данные, полученные для различных видов тканей и заболеваний, а также данные о субклеточной локализации и тканевой специфичности. В частности, для ГЦР содержатся данные, полученные от 365 пациентов (119 женщин и 246 мужчин) с различными стадиями заболевания (1 стадия – 170 пациентов, 2 стадия – 84 пациента, 3 стадия – 83 пациента, 4 стадия – 4 пациента, и 24 пациента с отсутствующей информацией о стадии). Используя транскриптомный подход, определены 2615 белок-кодирующих генов, увеличение экспрессии которых связано с неблагоприятным прогнозом общей пятилетней выживаемости при раке печени, из них 569 белок-кодирующих генов высоко экспрессируются только при раке печени по сравнению с другими видами рака. Среди высоко экспрессируемых белок-кодирующих генов для 20 было показано, что их высокие уровни экспрессии более чем на 30% снижали общую пятилетнюю выживаемость. С благоприятным прогнозом общей пятилетней выживаемости связано увеличение экспрессии 263 генов. Иммуногистохимическим методом показано, что они имеют дифференцированный характер экспрессии при раке печени (Таблица 1). Для удобства поиска названия генов и кодируемых ими белков, а также возможных ПТМ, здесь и далее они приведены на языке оригинала (английский) во избежание разночтений.

Таблица 1. Топ 20 белок-кодирующих генов, связанных с неблагоприятным прогнозом выживаемости.

Ген	Название	мРНК¹	p-значение
<i>SFPQ</i>	Splicing factor proline and glutamine rich	18,7	7,91e-13
<i>G6PD</i>	Glucose-6-phosphate dehydrogenase	13,5	9,45e-13
<i>KIF20A</i>	Kinesin family member 20A	2,9	2,14e-12
<i>KDM1A</i>	Lysine demethylase 1A	7,0	4,47e-12
<i>TRMT6</i>	Trna methyltransferase 6 non-catalytic subunit	3,6	4,70e-12
<i>PSMD1</i>	Proteasome 26S subunit, non-ATPase 1	17,6	8,63e-12
<i>GTPBP4</i>	GTP binding protein 4	5,6	8,85e-12
<i>CCT4</i>	Chaperonin containing TCP1 subunit 4	40,9	9,37e-12
<i>CAD</i>	Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	4,2	1,72e-11
<i>RBM28</i>	RNA binding motif protein 28	1,1	2,72e-11
<i>HILPDA</i>	Hypoxia inducible lipid droplet associated	3,2	4,30e-11
<i>CCT5</i>	Chaperonin containing TCP1 subunit 5	20,6	6,57e-11
<i>YARS1</i>	Tyrosyl-tRNA synthetase 1	9,3	6,97e-11
<i>INTS13</i>	Integrator complex subunit 13	4,7	1,17e-10
<i>GTSE1</i>	G2 and S-phase expressed 1	1,4	1,90e-10
<i>KPNA2</i>	Karyopherin subunit alpha 2	19,8	2,12e-10
<i>SNX5</i>	Sorting nexin 5	8,3	2,30e-10
<i>TTK</i>	TTK protein kinase	1,2	3,84e-10
<i>CDCA8</i>	Cell division cycle associated 8	3,8	4,93e-10
<i>HDAC2</i>	Histone deacetylase 2	2,6	5,39e-10

¹ Значения суммарных уровней экспрессии транскриптов с нормализацией FPKM (Fragments Per Kilobase of transcript per Million / Фрагменты на килобазу транскрипта на миллион сопоставленных чтений).

1.5. Роль протеоформ при различных заболеваниях

В настоящее время существует множество работ, которые показывают важность изучения протеоформ при различных заболеваниях. Известно, что разные протеоформы могут способствовать развитию и прогрессии опухолей, к примеру, специфические изоформы киназ, такие как BRAF (serine/threonine-protein kinase B-raf) или PIK3CA (phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform). Их активация при мутациях приводит к неконтролируемому клеточному росту и развитию таких видов рака, как колоректальный рак, меланома и папиллярный рак щитовидной железы [25,26]. Также известны протеоформы, участвующие в патогенезе нейродегенеративных заболеваний. В частности, с использованием МС и иммунодетекции были выявлены и охарактеризованы протеоформы, образованные различными видами ПТМ (фосфорилирование, N-концевое ацетилирование, окисление метионина и окисление цистеина) для 51 белка-мишени при болезни Паркинсона. При этом показано, что некоторые из них могут быть использованы для функциональной характеристики и дифференциальной диагностики болезни Паркинсона среди других нейродегенеративных заболеваний [27]. В недавнем исследовании Noog и соавторов было показано, что существуют 33 протеоформы белка амилоид-бета (A β), связанные с быстрым прогрессированием болезни Альцгеймера [28]. Важную роль при ишемических болезнях сердца играют измененные уровни фосфорилирования и экспрессии сердечного тропонина I (*TNNI3*) и гомолога энigma 2 (enigma homologue 2, *ENH2*), а также увеличение фосфорилирования мышечного белка MLIM (Muscle LIM protein, *CSRP3*) и кальсарцина-1 (*Cal-1*) при ишемической кардиомиопатии [29].

Аберрантное гликозилирование является общей характеристикой раковых клеток и может влиять на различные стадии опухолевой прогрессии. Одним из важных изменений в гликозилировании опухоли является коровое фукозилирование (добавление фукозы через альфа-1,6-гликозидную связь 1-го остатка N-ацетилглюкозамина), которое катализирует фермент

фукозилтрансфераза 8 (*FUT8*) [30]. Данный фермент активируется при различных типах рака, включая меланому, рак печени, легких, толстой кишки, яичников, простаты, молочной железы, щитовидной железы и поджелудочной железы [31]. Увеличение уровня экспрессии гена *FUT8* и кодируемого им белка в опухолевых тканях и метастатических клетках может свидетельствовать о более агрессивном характере опухоли, а также ассоциироваться с повышенной частотой рецидивов и снижением общей выживаемости пациентов [32].

Как уже было ранее упомянуто, АФП является клиническим маркером ГЦР. Но ввиду повышения его уровня не только при ГЦР, но и при других заболеваниях, было проведено исследование гликозилирования АФП и его роли в дифференциации цирроза печени от ГЦР [7]. Авторами был разработан метод обнаружения и количественной оценки в образцах сыворотки относительного содержания гликоформ АФП, особенно АФП-L3. Этот метод включает иммунное обогащение эндогенного АФП с последующей жидкостной хроматографией в сочетании с анализом масс-спектрометрией высокого разрешения (Liquid chromatography-high resolution mass spectrometry, LC-HRMS). На основании данных о возможных модификациях гликанами в сыворотке крови пациентов они выделили 12 различных гликоформ (G_0 , G_0F , G_1 , G_1F , G_1S_1 , G_1FS_1 , G_2 , G_2FF , G_2S_2 , G_2S_1 , G_2FS_1 и G_2FS_2). Исследование показало, что часто наблюдаемыми гликоформами АФП без фукозилирования ядра (АФП-L1) являются G_2S_2 и G_2S_1 , тогда как распространенными гликоформами с фукозилированным ядром (АФП-L3) являются G_2FS_1 и G_2FS_2 . В исследовании также рассматривалось фосфорилирование как дополнительная ПТМ при анализе интактных протеоформ АФП [7].

1.6. Современные методы диагностики ГЦР

Первичная диагностика ГЦР включает в себя следующие исследования: общий анализ крови, анализ на наличие антител к вирусу гепатита В и С, ультразвуковое исследование. Дальнейшие скрининговые мероприятия проводятся по результатам первичных исследований. В случае положительного

результата на наличие антител к вирусу гепатита В или С проводят анализ на опухолеспецифичный маркер АФП, мультиспиральную компьютерную томографию (МСКТ) и магнитно-резонансную томографию (МРТ). Повышение качества лучевой диагностики МСКТ и МРТ может достигаться путем замещения ручной инъекции контраст-вещества на болюсное контрастирование (введение йодсодержащего контраст-вещества с помощью шприца-инжектора). Для достоверного подтверждения гистологического типа опухоли выполняют чрескожную игольную биопсию печени [33]. Из-за отсутствия ярко выраженных симптомов ГЦР чаще всего диагностируется на более поздних стадиях и имеет неблагоприятный прогноз общей выживаемости ввиду ограниченных вариантов методов лечения.

1.6.1. Биомаркеры ГЦР в клинической практике

Современная медицина сильно зависит от «надежных» биомаркеров для скрининга, мониторинга заболеваний и ответа на терапию, прогнозирования и безопасности лечения. Глоссарий BEST (Biomarkers, Endpoints, and other Tools) рабочей группы FDA-NIH (Food and Drug Administration and The National Institutes of Health)² определяет биомаркер как параметр, который отражает как нормальные биологические процессы, так и патогенные процессы, или биологические реакции на воздействие или вмешательство (включая терапевтические). «Идеальный биомаркер» должен отвечать ряду критериев: обладать высокой специфичностью и чувствительностью, коррелировать с тяжестью заболевания и получаемой терапией, детектироваться в периферических жидкостях, а также быть доступным для рутинного анализа.

В соответствии с национальными клиническими рекомендациями для постановки диагноза пациентам определяют содержание опухолеспецифичного маркера АФП в сыворотке крови. Результат анализа на АФП считается положительным, если уровень АФП >100 нг/мл или если он увеличивается на 7

² www.ncbi.nlm.nih.gov/books/NBK326791

нг/мл в месяц по результатам трех последовательных измерений [34]. Несмотря на то, что специфичность АФП составляет более 90%, его чувствительность сравнительно невысока и варьирует в пределах 39-65% в зависимости от базового уровня [35,36]. Экспрессия АФП отсутствует примерно в 80% случаев раннего развития ГЦР [37]. Уровень АФП может повышаться при онкологических заболеваниях другой локализации (гепатоидной аденокарциноме поджелудочной железы, герминогенных опухолях и др.), при хронических заболеваниях печени, острых и хронических гепатитах и даже при беременности [38–42]. Таким образом, уровни АФП в сыворотке не могут надежно отличить хроническое повреждение печени, например фиброз или цирроз от ГЦР. В ряде случаев применяется также второй маркер – ДГП. Но его чаще применяют для мониторинга терапии и при условии повышенных значений до начала лечения. Иногда эти маркеры дополняют другими не слишком специфичными анализами, включая измерение уровня аполипопротеина, альфа-2-макроглобулина и гамма-глутамиттрансферазы.

В международной клинической практике помимо АФП используются несколько биомаркеров: предполагаемый белок 7 теплового шока 70 кДа (putative heat shock 70 kDa protein 7, HSP70B (ген *HSPA7*), глипикан 3 (glypican 3, GPC3 (ген *GPC3*)) и глутаминсинтетаза (glutamine synthetase, GLNA (ген *GLUL*)) [43]. Тем не менее, данные биомаркеры имеют ряд диагностических ограничений и не учитываются во всех основных рекомендациях. В отличие от нормальных клеток многие опухолевые образования, включая ГЦР, синтезируют в повышенном количестве белки семейства теплового шока HSP70 (*HSPA1A*, *HSPA1B*, *HSPA4*, *HSPA5*, *HSPA8*, *HSPA13* и *HSPA14*) и секретируют их во внеклеточный матрикс [44,45]. Кроме того, было показано, что уровни белков семейства HSP70 в сыворотке позволяют различать хронический гепатит, цирроз печени и ГЦР. Но следует отметить, что это наблюдение было ограничено анализом относительно небольшой когорты (86 здоровых доноров, 50 доноров с хроническим гепатитом и

47 пациентов с ГЦР), и учитывать, что белки семейства HSP70 секретируется и другими опухолями [46].

1.6.2. Молекулярно-генетические маркеры на основе геномного секвенирования

Недавние достижения в области секвенирования нового поколения (Next-Generation Sequencing, NGS) позволили выявить генетические и молекулярные механизмы возникновения ГЦР [47]. Данные механизмы включают в себя гомозиготные делеции белок-кодирующего гена *CDKN2A* на хромосоме 9 и высокие уровни амплификации ДНК на хромосомах 6p21 (белок-кодирующего гена *VEGFA*) и 11q13 (белков, кодируемых генами *FGF19* и *CNND1*). Метаанализ 2018 года показал, что метилирование промотора гена *CDKN2A* связано с повышенным риском развития ГЦР [47]. В исследовании, выполненном в 2021 году, было показано, что увеличение экспрессии гена *CDKN2A* влияет на регуляцию ассоциированных с опухолью макрофагов [48]. Наиболее распространенными являются мутации в белок-кодирующих генах *TP53* и *CTNNB1*, выявляющиеся у 25-30% пациентов с ГЦР [49]. Исследования показали, что у 18-40% пациентов с ГЦР выявляются мутации в гене *CTNNB1*, белковый продукт которого участвует в метаболической регуляции сигнального пути Wnt/ β -катенин [50].

Как было показано во множестве исследований, генетические вариации достаточно сильно зависят от исследуемой популяции, следовательно, имеют ограничения применимости для массового скрининга и диагностики [51,52]. Исследование вариации числа копий генов в когорте из 39 пациентов, проведенное в 2022 году, показало, что трансверсии T>G/A>C и транзиции C>T/G>A были доминирующими вариантами мутаций. Мутация в гене *TP53* была наиболее распространенной (76,9%); также мутации обнаруживались в генах *MUC16*, *MUC12* и *MUC4*. Наиболее распространенными были p.Arg249Ser (23,5%) и миссенс-мутация (82,3%) в варианте TP53 NP_001119586.1. Кроме того, миссенс-мутации возникали в генах и кодируемых ими белках EGF, CTNNB1,

BRCA2, NTRK3, LRP1B, AXIN1 и IRS2. Также было обнаружено, что уровень амплификации генов *FGF3*, *TERT* и *SOX2* составлял соответственно 15,4%, 12,8% и 12,8%. Эти значения существенно превышали данные, представленные в базе данных cBioPortal³, что свидетельствует о более высокой степени амплификации этих генов в исследуемой когорте [53].

Развитие технологий NGS для персонализированного подхода лечения, безусловно, имеет свои преимущества; тем не менее, они ограничены ДЭГ среди различных популяций. Поэтому необходима разработка более чувствительного и специфичного метода поиска ДЭГ и дифференциально экспрессируемых белков (ДЭБ) для поиска биомаркеров. Такие панели могут быть созданы и при использовании данных о протеоформах.

1.7. Модели клеточных линий гепатоцеллюлярного рака и их сравнение

Профили экспрессии генов в клетках опухолей и в полученных из них клеточных линиях весьма схожи. Однако так как условия культивирования клеток *in vitro* отличаются от условий роста *in vivo* в опухолях, уровни экспрессии отдельных генов могут существенно изменяться. Химиотерапевтические препараты, для которых была доказана эффективность на клеточных линиях, не всегда дают положительные результаты при воздействии на опухоль [54]. Поэтому в работах, посвящённых, например, поиску потенциальных биомаркеров или терапевтических мишеней, выбор клеточных линий для использования в качестве моделей ГЦР становится актуальной задачей. В настоящее время насчитывается порядка 30 клеточных линий, используемых в качестве модели ГЦР, среди них наиболее часто используемой является HepG2. Данная клеточная линия была получена в 1979 году из образца опухоли печени 15-летнего юноши европейского происхождения с дифференцированной гепатоцеллюлярной карциномой [55]. Клетки HepG2 имеют свойства эпителиальных гепатоцитов и могут быть использованы для изучения различных биологических процессов,

³ <https://www.cbioportal.org> Геномная база данных о мутациях в РНК и ДНК при раковых заболеваниях

связанных с печенью, включая образование опухолей, метаболизм лекарственных препаратов и токсикологию. Культура клеток HepG2 широко применяется во множестве экспериментальных исследований, посвящённых обнаружению функциональных генов, связанных с патологическими процессами в печени, а также в разработке новых лекарственных препаратов для лечения ГЦР и иных заболеваний печени. Тем не менее, использование клеточной линии HepG2 остается довольно спорным, несколько исследований предполагают ее происхождение от эпителиальной гепатобластомной опухоли, а не от гепатоцеллюлярной карциномы [56,57].

В связи с этим особый интерес представляет работа 2015 года, авторы которой сравнили экспрессию 1386 генов в клетках 200 образцов опухолей ГЦР и более чем в 1 тыс. клеточных линий, включая 25 наиболее используемых в исследованиях, ассоциированных с ГЦР [58]. В работе использовали клинические данные пациентов из базы данных TCGA (<https://tcga-data.nci.nih.gov/tcga>) и данные экспрессии генов клеточных линий из энциклопедии линий раковых клеток (Cancer Cell Line Encyclopedia, CCLE). Из 25 клеточных линий только четыре, а именно HepG2, HuH7, Hep3B и PLC/PRF/5, тесно коррелировали с опухолями и показали адекватный ответ на применение противораковых лекарственных препаратов. В работе отмечается, что недостаточное описание клинических образцов, потенциальное повреждение генома химиотерапевтическими препаратами или недостаточная дифференциация могут быть причиной низкой корреляции для некоторых клеточных линий [58].

Немаловажным является тот факт, что использование модельных клеточных линий ограничено способностью сохранять и воспроизводить биологические особенности первичной опухоли из-за перекрестного загрязнения другими линиями, длительного культивирования и, как следствие, увеличению генетического дрейфа [59]. Исследования модельных клеточных линий в основном проводятся с помощью транскриптомных подходов, при этом в силу разобшенности транскриптома и протеома выбор «правильной» клеточной линии

для моделирования конкретного заболевания имеет фундаментальное значение [60,61].

1.8. Методы исследования белков для изучения разнообразия протеоформ

Развитие геномных технологий, среди которых прежде всего следует отметить полимеразную цепную реакцию (1985) [62] и методы масс-спектрометрического анализа крупных биомолекул (1987) [63], а также появление баз данных белковых и геномных последовательностей и работы по проекту «Геном человека» (1990-2022) [64], сделали возможным крупномасштабный анализ белкового состава объектов исследований.

Термин «протеом» был введен Майклом Уилкинсом в 1994 г и является аналогом термина «геном», где префикс «протео-» относится к белкам, а суффикс «-ом» обозначает целостность и совокупность, т. е. полный перечень белков, присутствующих в клетке, ткани, системе организма или в целом организме в определенный момент времени или в определенных условиях [65]. Используемые для такого рода исследований методы стали называть протеомными. Протеомы, или наборы, обнаруживаемых в образце белков, получают в результате сопоставления последовательностей белков или белок-кодирующих генов в базах данных с результатами масс-спектрометрического анализа.

Можно выделить две основные «дорожные карты» идентификации белков исследуемого образца. Первая предполагает разделение белков, гидролиз выделенных белков, регистрацию спектров смесей полученных пептидов и идентификация белков по набору зарегистрированных пептидных масс поиском в белковых или нуклеотидных базах данных. Для подавляющего числа белков последовательность из 10-15 аминокислот оказывается уникальной и позволяет однозначно идентифицировать белок (с точностью до гомолога). В таком варианте анализа используют разделение белков в геле, чаще всего используют 2DE. Второй вариант протеомного анализа состоит в гидролизе всех белков

образца в одной пробирке, определение молекулярных масс и регистрации спектров фрагментации (МС/МС) продуктов гидролиза. Так как производительность масс-спектрометров не бесконечна, смесь пептидов предварительно разделяется по гидрофобности, а в результате по времени, жидкостной хроматографией. Такой анализ часто обозначают как LC-MS/MS (Liquid Chromatography coupled with Tandem Mass Spectrometry). Для идентификации каждого белка достаточно получить спектры фрагментации для нескольких, иногда одного пептида. Для идентификации белка используют несколько пептидов, полученных в результате специфичного ферментативного гидролиза, чаще всего трипсином, и измеряют молекулярные массы или регистрируют спектры фрагментации этих пептидов. Тем не менее, из-за ПТМ и возможности существования белка в комплексах с другими белками, солями, липидами и другими соединениями, идентификация белка простым измерением его молекулярной массы и сравнением с расчетными данными в базах данных зачастую оказывается невозможной. Очевидно, что LC-MS/MS анализ ограниченно годен для поиска изоформ или протеоформ белков – различные формы белка будут иметь близкие наборы пептидов с одинаковыми временами выхода и спектрами фрагментации. В то же время модификации белка, изменяя состав молекул, меняет их массу, подвижность и изоэлектрическую точку. Поэтому для поиска протеоформ использование предварительного разделения в геле оказывается предпочтительным или даже необходимым.

1.8.1. Масс-спектрометрические методы

История развития МС демонстрирует значительный прогресс от первых простых приборов до современных высокоточных систем. Их применение привело к важным научным открытиям и разработке новых методов анализа. МС позволяет идентифицировать и характеризовать белки и их модификации с высокой точностью и чувствительностью [66]. Более того, МС имеет важное значение в медицине, включая диагностику инфекционных болезней, наследственные болезни обмена веществ (нарушения метаболизма аминокислот,

органических кислот и дефектов митохондриального бета-окисления жирных кислот), мониторинг лекарственной терапии и поиск новых биомаркеров [67]. В целом, развитие МС продолжается, и она остается одним из ключевых инструментов в современных исследованиях.

Масс-спектрометр работает на основе ионизации пептидов в смеси, сортировки и детектирования ионов и анализа полученного спектра при помощи компьютерных вычислительных методов. В процессе ионизации пептиды приобретают положительный заряд и превращаются в ионы. Затем ионы проходят через вакуумную систему, где они сортируются в зависимости от величины отношения массы к заряду (m/z) и регистрируются детектором. По полученному спектру определяется масса каждого иона. Ранее использование МС в биологии было ограничено из-за жесткой ионизации, которая могла приводить к разрушению молекул. Однако в 1980-х годах были разработаны мягкие методы ионизации, например, лазером после кристаллизации белков с матрицей (светочувствительным органическим веществом). Этот метод, известный как MALDI (Matrix-Assisted Laser Desorption / Ionization), стал революционным техническим прорывом в масс-спектрометрии белков. Он позволил более мягкую ионизацию, не повреждая молекулы, и стал широко применяемым методом для исследования биологических молекул. Метод MALDI основан на окружении молекул исследуемого вещества матрицей, которая под действием лазерного излучения ионизирует соседние молекулы без их разрушения. Этот метод был интегрирован с масс-спектрометром, использующим времяпролетный детектор (time of flight, TOF). В TOF-детекторе ионы перемещаются в вакуумной трубке и достигают чувствительной пластины (фотоэлектронного умножителя), которая служит детектором. Из-за разности потенциалов, которые постоянны по отношению ко всем ионам, ионы с меньшим значением отношения массы к заряду (m/z) и более заряженные ионы движутся быстрее. Следовательно, время пролета иона различается в зависимости от значения m/z иона. Таким образом, метод

MALDI-TOF позволяет получать информацию о массе ионов и проводить анализ молекулярных компонентов образца (Рисунок 3).

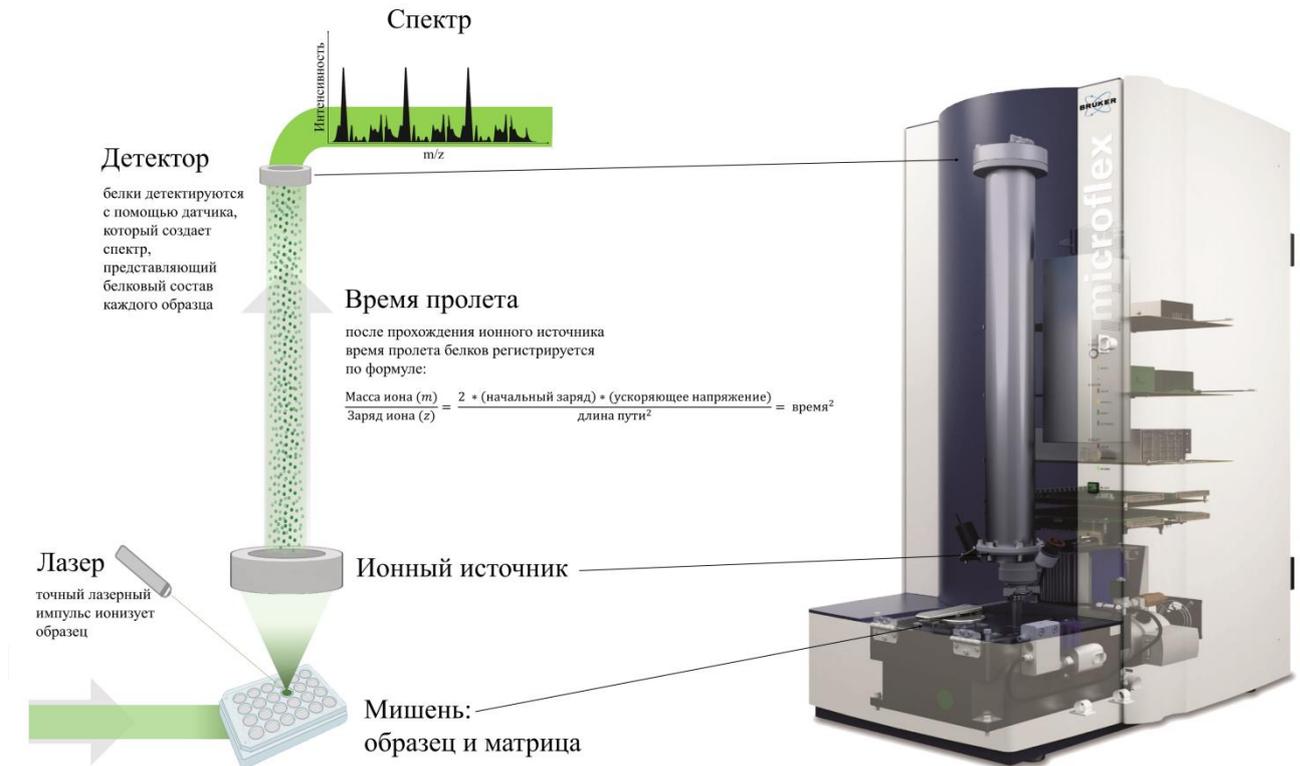


Рисунок 3. Схема работы времяпролетного масс-спектрометра.

Совершенствование методов МС достигается путем внедрения новых аналитических подходов. В настоящее время наиболее частым подходом для анализа сложных белковых смесей является сочетание жидкостной хроматографии и МС с использованием ионизации распылением в электрическом поле (Liquid Chromatography coupled with Tandem Mass Spectrometry with Electrospray Ionization, LC-ESI-MS/MS). Данный метод заключается в предварительном фракционировании жидкостной хроматографией на компоненты в зависимости от их физико-химических свойств, таких как размер, гидрофобность, заряд и др.. Затем проводится ионизация электрораспылением, в результате которой молекулы превращаются в ионы. В дальнейшем ионы отбираются и пропускаются через анализатор масс-спектрометра с множественным сканированием, где происходит их фрагментация и регистрация

результатов. Этот метод применяется для идентификации, структурного анализа и количественного определения белков и пептидов в сложных образцах.

Несмотря на свои многочисленные преимущества, он имеет некоторые недостатки и ограничения, такие как, сложность подготовки образцов для минимизации потерь белков и пептидов, ограниченную чувствительность к некоторым классам соединений из-за их физико-химических свойств и возможное перекрестное загрязнение образцов (кросс-контаминации), в особенности при анализе сложных смесей. Также он не способен в полной мере регистрировать единичные протеоформы и, следовательно, необходимо использовать дополнительные методы предварительного фракционирования образца. Такое ограничение решается путем применения комбинаций различных методов [68–71], что позволяет упростить процесс анализа сложных биологических объектов, и способствует получению более точных и надежных результатов идентификации белков [72]. Одним из наиболее ярких примеров комбинации является 2DE в сочетании с МС высокого разрешения [73].

1.8.2. Двумерный гель-электрофорез (2DE)

Первый вариант двумерного разделения белков электрофорезом был представлен Смитсом в 1956 году [74]. В его работе использовалась фильтровальная бумага для первой стадии разделения и крахмальный гель для второй. Однако этот метод имел низкое разрешение и был ограничен по информативности. Дальнейшее развитие и улучшение 2DE опиралось на постоянное совершенствование методов одномерного гель-электрофореза и использование новых сред для разделения. Особенно важным прорывом стало использование изоэлектрического фокусирования (ИЭФ) для первой стадии разделения белков по изоэлектрическим точкам или их зарядам, позволяющее достигать высокого разрешения разделения [75–77].

Главным достоинством метода 2DE является проведение разделения по двум независимым параметрам – изоэлектрическая точка (pI) и молекулярная

масса (M_w). Вначале выполняется ИЭФ в первом направлении для разделения по pI . Затем идет разделение во втором направлении с использованием электрофореза в полиакриламидном геле, обеспечивающим высокое разрешение по размеру или молекулярному весу белка (Рисунок 4).

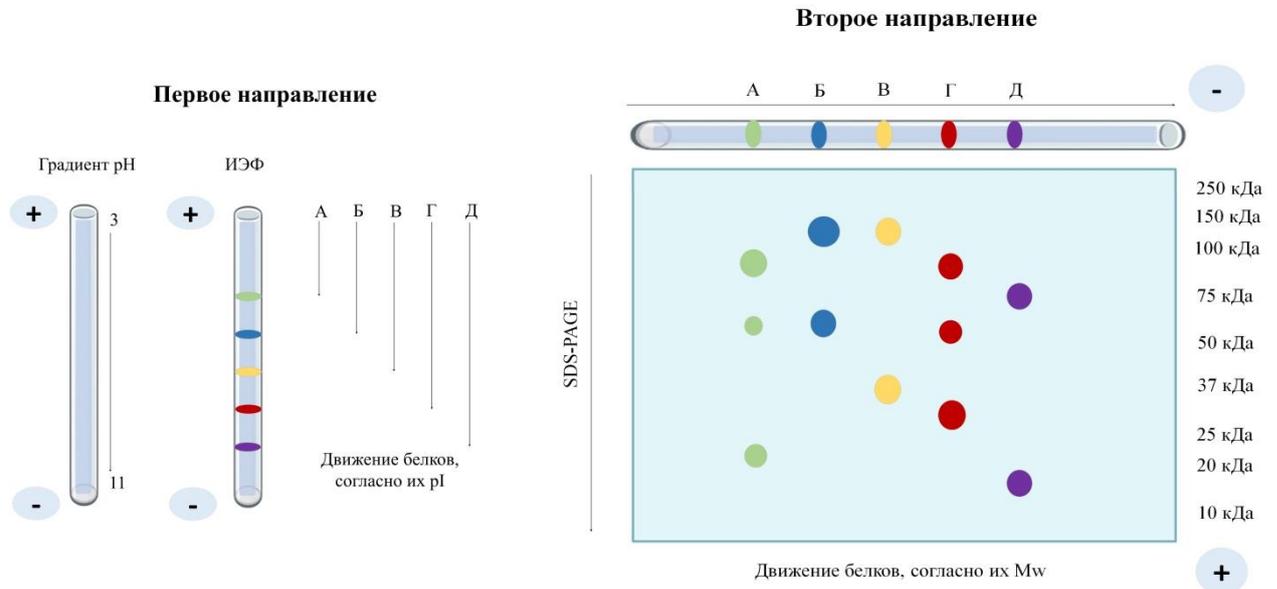


Рисунок 4. Схематическое представление 2DE. Первое направление – разделение по изоэлектрической точке (pI), второе направление – разделение в полиакриламидном геле (SDS-PAGE) по молекулярной массе (M_w).

Оба эти метода обладают превосходным разрешением, особенно ИЭФ, достигая разрешения до 0,001 единицы pH [78,79]. Уникальность метода 2DE заключается в том, что разрешение, полученное в процессе ИЭФ, не теряется при последующем разделении второго направления [80]. Благодаря этим свойствам, метод 2DE был признан наиболее мощным подходом для разделения сложных белковых смесей [81,82]. Хотя протокол этого подхода остается практически неизменным с 1975 года, за последние десятилетия были внесены некоторые улучшения. Одно из значительных улучшений заключалось в использовании коммерческих иммобилизованных градиентов pH (IPG-полоски), которые почти полностью заменили градиент pH , основанный на амфолитах-носителях в гелях, существенным образом упростив проведение экспериментов и их воспроизводимость [83]. Дополнительным улучшением было внедрение новых

хаотропных и детергентных веществ, улучшающих растворимость гидрофобных белков [84]. В результате этих усовершенствований метод 2DE стал еще более эффективным и мощным инструментом для разделения и анализа сложных белковых образцов [68–71].

Существует две основные характеристики, влияющие на количество пятен в 2DE – разрешение и чувствительность. Если мы рассматриваем 2DE только как метод разделения, то следует говорить только о разрешении. Улучшение метода достигается подбором параметров первого и второго направления 2DE. При этом обнаружение пятен в значительной степени зависит от используемого метода окрашивания. Выбор метода окрашивания зависит от различных условий, таких как чувствительность, линейность и совместимость с последующими методами исследования, например масс-спектрометрией. Основными используемыми красителями являются бриллиантовый синий Coomassie (R250, R350, G250) и нитрат серебра. Окрашивание Coomassie обладает умеренной чувствительностью, но обеспечивает хорошую линейность и точность. Краситель связывается с основными и ароматическими аминокислотами благодаря электростатическим и гидрофобным взаимодействиям, и совместим с последующим масс-спектрометрическим анализом.

Окрашивание нитратом серебра обладает более высокой чувствительностью по сравнению с Coomassie, но характеризуется меньшей линейностью и точностью [85]. Кроме того, оно не всегда подходит для анализа методом масс-спектрометрии из-за возможного сшивания белков при использовании формальдегида в качестве восстановителя. Также доступны методы окрашивания, основанные на флуоресцентных метках (например, Sypro Ruby или Flamingo). Эти методы обеспечивают высокую чувствительность, линейность и совместимость с МС [86].

1.8.3. Иммуноблоттинг и иммунохимические методы

Суть иммунохимических методов анализа заключается во взаимодействии между целевым белком (или другим антигеном) и специфическим антителом. Этот метод может использоваться для обнаружения белков с низкой концентрацией в сложных смесях, для контроля уровня экспрессии белков и для проверки их очистки. Благодаря высокой точности идентификации белков этот метод получил широкое применение как в научных исследованиях, так и в клинической практике. Такие методы активно развиваются с начала 1970-х годов.

В 1979 году был разработан метод иммунодетекции рибосомальных белков L7 и L12 *E.Coli*, а также рибосомальных белков куриной печени, используя полиакриламидный гель и нитроцеллюлозную мембрану (Вестерн блоттинг). Это позволило показать, что белки, закрепленные на поверхности мембраны, могут быть легко обнаружены при помощи специфических антител [87].

Процесс иммунодетекции белков с использованием предварительного разделения белков в полиакриламидном геле (ПААГ) состоит из трех основных этапов: разделение, перенос и детекция. На первом этапе сложные смеси белков и/или нуклеиновых кислот разделяются электрофоретическим методом в ПААГ. На втором этапе разделенные белки переносятся на различные типы мембран. Наконец, на третьем этапе происходит детекция белков при помощи иммунохимических методов, окрашивания или авторадиографии для их идентификации.

Связывание белков на мембране основано на гидрофобных и электростатических взаимодействиях. Для переноса белков из полиакриламидного геля на мембраны используются два типа мембран: нитроцеллюлозные и поливинилиденфторидные (polyvinylidene difluoride, PVDF). Нитроцеллюлозные мембраны имеют ряд недостатков, такие как хрупкость, низкая ёмкость связывания белка (80-100 мкг/см²), неустойчивость к растворителям и ограниченность повторного использования с другими

антителами. Поэтому чаще всего используют мембраны из PVDF, которые не имеют ограничений в использовании по сравнению с нитроцеллюлозными мембранами. Для оценки эффективности переноса белков с геля на мембрану можно использовать окрашивание с помощью различных красителей, например, Coomassie Blue или Ponceau S. Coomassie Blue является наиболее распространенным красителем, в то время как Ponceau S лучше растворяется в воде, что упрощает его последующую полную отмывку и последующий анализ антителами. Хотя, часто в такой отмывке и нет необходимости. Например, окраска белков Coomassie Blue R (не G) позволяет не только более точно оценить перенос белков на мембрану, но и провести более полный иммуноанализ [88].

Иммунохимические методы обладают высокой чувствительностью и широко используются в исследованиях, лабораторных и клинических анализах. Однако их использование имеет ряд ограничений, связанных в первую очередь с тем, что не для каждого целевого белка имеются специфические антитела, и для крупномасштабного исследования может потребоваться большое число различных антител. Напомним, геном человека содержит ~ 20 000 белок-кодирующих генов. Кроме того, выявление различных форм белка может потребовать получение отдельных антител для каждой формы. На результат эксперимента будет влиять локализация эпитопов, и, если эпитоп принадлежит константной области белка, различить возможные модификации последнего будет затруднительно. Иммунохимические методы могут использоваться для выявления протеоформ, но их использование имеет ряд ограничений, связанных с необходимостью проведения множества предварительных анализов и сложностью подбора антител. Это приводит не только к увеличению трудозатрат, но и к увеличению стоимости таких методов.

1.8.4. Биоинформатический анализ

В эпоху высокопроизводительных технологий, таких как NGS, MS и др., объем данных, генерируемых в исследованиях, возрастает экспоненциально. Все это приводит к тому, что биоинформатический анализ становится неотъемлемой

частью современных биологических исследований, выполняя центральную функцию в интерпретации и управлениями огромными массивами данных [89,90].

В зависимости от цели можно выделить основные направления биоинформатического анализа: анализ последовательностей, структурная биоинформатика, функциональная аннотация генов и белков, анализ геномов и белков, анализ метаболических и сигнальных путей и моделирование структур. Интерпретация биоинформатического анализа зависит от нескольких ключевых факторов, которые отражают правильность и полноту выводов на основе полученных данных. К ключевым факторам относятся: качество и объем данных, выбор и настройка алгоритмов поиска, базы данных и аннотированные ресурсы, статистическая обработка данных.

Учитывая, что в протеомных исследованиях используется множество различных методов и программных платформ, используемых научными лабораториями, и, как следствие, генерируется большое количество форматов данных, что приводит к трудностям при обмене данными между исследователями. К тому же переход от одного формата данных к другому может вызывать ряд проблем, таких как потеря или же искажение метаданных. Для исключения проблем, связанных с потерей или искажением метаданных, полученных в ходе протеомных исследований, в частности МС (как в наиболее используемом методе), организацией «Протеом человека» (The Human Proteome Organization, HUPO)⁴ были разработана стандартизация формата данных МС – mzML (Mass Spectrometry Markup Language). Он представляет собой иерархически структурированные данные одного прогона МС (включая метаданные о спектрах и сами спектры) в текстовом формате данных. Первый релиз данного формата (версия 1.0.0) был опубликован в 2008 году на конференции Американского общества масс-спектрометрии и заменил собой два отдельных формата данных XML для кодирования необработанных исходных данных. Первый – mzData, разработанный рабочей группой по стандартам масс-спектрометрии (PSI-MS), и

⁴ <https://hupo.org/HPP-Data-Interpretation-Guidelines>

второй – mzXML, разработанный в Институте системной биологии (Institute for Systems Biology (ISB), Сиэтл, Вашингтон). В ноябре 2022 года была выпущена версия 1.1.1 и на данный момент обновлений не планируется [91].

Достоверность результатов, полученных в ходе эксперимента, обеспечивается выбором правильных алгоритмов и их настройкой. Для этого в процессе анализа данных применяют следующий алгоритм – подход «мишень-приманка» («target-decoy») и последующая оценка ложноположительных результатов (False Discovery Rate, FDR). Такой алгоритм идентификации основан на том, что помимо реальных последовательностей белков и пептидов в MS-данных (мишень) создаются, случайно сгенерированные последовательности, которые не соответствуют реальным белкам или пептидам, чаще всего путем инвертирования реальных последовательностей белков и пептидов (приманка) [92]. FDR при проверке гипотезы совпадения исходного спектра с теоретическими последовательностями пептидов (peptide-spectrum match, PSM), происходит путем деления количества ложноположительных спектров на количество реальных спектров [93]. Рекомендуемым порогом уровня ложноположительных результатов является 1%, и как только доля PSM с ложными результатами достигает этого порога, поиск прекращается. Если применять порог более 1%, вероятность ложноположительных результатов возрастает многократно, что приводит к недостоверности полученных результатов.

Не менее важным фактором, помимо стандартизации форматов данных и выбора правильного алгоритма анализа, являются базы данных и аннотированные ресурсы, от достоверности которых зависит конечный результат исследований. Так как недостаточно аннотированные или ошибочные записи могут привести к неверной функциональной аннотации генов или белков. В Таблице 2 представлены наиболее часто используемые веб-ориентированные базы данных.

Таблица 2. Веб-ориентированные базы данных для интерпретации экспериментальных данных.

База данных	Описание и функции	Ссылка
UNIPROT	Крупнейшая база данных белков, представляющая подробную информацию о последовательностях белков, их функциях, структурах и взаимодействиях.	www.uniprot.org
NextProt	Является расширением базы данных Uniprot и объединяет информацию из различных источников, предоставляя более детализированные данные о белках.	www.nextprot.org/
STRING	Предсказание и анализ белковых взаимодействий, и функциональных аннотаций.	www.string-db.org
PANTHER	Классификация белков по семействам, предсказание их молекулярных функций, клеточной локализации, биологических процессов и путей.	www.pantherdb.org
KEGG PATHWAY	Содержит информацию о молекулярных сигнальных и метаболических путях в различных организмах. Представляет собой интерактивную карту, позволяющую визуализировать молекулярные взаимодействия и связи между различными компонентами пути, такими как гены, белки и метаболиты.	www.genome.jp/kegg/pathway.html
Human Protein Atlas	Представляет собой базу данных о картировании всех белков человека в клетках, тканях и органах различными омиксными методами – иммуногистохимическими, транскриптомными и протеомными.	www.proteinatlas.org
Gene Expression Omnibus (GEO2R)	Сравнение двух или более групп образцов, для идентификации генов, которые по-разному экспрессируются в разных экспериментальных условиях.	www.ncbi.nlm.nih.gov/geo/geo2r
cBioPortal	Содержит данные о мутациях, о количестве копий ДНК, об уровнях экспрессии мРНК и микроРНК, об уровнях белков и уровнях фосфобелков, данные о метилировании ДНК и обезличенные клинические данные.	https://www.cbioportal.org
The Cancer Genome Atlas (TCGA)	Содержит геномные, эпигеномные, транскриптомные и протеомные данные для 33 типов рака и более чем 86000 пациентов.	https://portal.gdc.cancer.gov

1.8.4.1. Базы данных протеоформ человека

Развитие и совершенствование методов исследования приводит к тому, что накопленные знания необходимо систематизировать в единую систему для их

успешного использования. Примером успешных белковых баз данных человека стали базы Uniprot и NextProt. Обе эти базы представляют собой данные о различных вариантах «существования белка», т. е. об его изоформах, предсказаниях о наличии ПТМ и других модификациях. Часть данных представлены в виде биоинформатических предсказаний, которые выявлены лишь на сходстве его последовательности или структуры с другими белками. Однако, многие белки, предсказанные с использованием геномного или транскриптомного анализа, не удастся обнаружить в силу того, что они были предсказаны неправильно из-за их недостаточного содержания, или же имеется ограничение экспрессии белок-кодирующего гена по времени. Поэтому получение данных о протеоформах, их систематизация и стандартизация в последние годы представляет собой большой научный интерес [94].

Для стандартизации данных о протеоформах консорциумом CTDP был предложен метод номенклатуры протеоформ «ProForma» [95], в дальнейшем она была расширена до версии «ProForma 2.0» [96]. Основные элементы номенклатуры включают в себя данные об аминокислотной последовательности в виде буквенного обозначения аминокислот, ПТМ в квадратных скобках сразу после модифицированной аминокислоты и локализацию модификаций в виде числовой нумерации [96]. Также консорциумом была создана первая база данных, содержащая информацию о протеоформах белков – Proteoform Atlas⁵. Она содержит в себе информацию о четырех клеточных линиях: HeLa S3, MCF7, H1299 и V16F10, а также данные о плазме, лейкоцитах и мононуклеарных клетках периферической крови, полученные различными методами анализа. В дальнейшем на основании базы данных Proteoform Atlas консорциум разработал базу данных исключительно о протеоформах белков человека – The Human Proteoform Atlas (HpfA), которая содержит 37 071 уникальную экспериментально подтвержденную протеоформу из 30 наборов данных 27 рецензируемых публикаций (дате доступа 18 сентября 2023 года) [97].

⁵ <http://atlas.topdownproteomics.org/>

Идентификация единичных протеоформ является достаточно сложной и важной задачей, но в то же время при таком подходе невозможно оценить изменения всего профиля протеоформ для конкретного биологического объекта. В 2020 году под руководством Нарыжного С.Н. была создана база «2DE-pattern», содержащая данные обо всех протеоформных паттернах для каждого отдельного белка и биологического объекта (<http://2de-pattern.pnpi.nrcki.ru>). На данный момент в базе содержатся данные о протеоформных паттернах следующих клеточных линий: HepG2, GliA-L (глиобластомные клетки) и LEH (эмбриональные фибробласты легких человека). Причем показаны не только протеоформные паттерны, полученные секционным методом 2DE, но и также полу-виртуальным 2DE [98], что обеспечивает более точную идентификацию протеоформ и позволяет расширить информацию о полученных профилях [99]. На Рисунке 5 представлена главная страница и в качестве примера, страница с информацией о белке EF2 (elongation factor 2), для которого приведены данные о его идентификации в вышеперечисленных клеточных линиях различными методами.

The image displays two screenshots of the '2DE-pattern' database website. The top screenshot shows the main search interface with a search bar containing 'P13639-1' and a 'Show' button. Below the search bar, there are links to 'Sectional 2DE' and 'Semi-virtual 2DE' for the selected protein. The bottom screenshot provides a detailed view for 'P13639-1 (EF2 Elongation factor 2 isoform 1)'. It features a grid of 2D electrophoresis images for three cell lines: Glioblastoma (GliA-L), Human embryonic lung fibroblasts (LEH), and Hepatocellular carcinoma cells (HepG2). For each cell line, there are two columns of images: 'Sectional 2DE' and 'Semi-virtual 2DE'. The 'Semi-virtual 2DE' images show a grid with colored spots representing protein spots. Below the images, there is a list of references related to the protein and the database.

References:

1. S. Naryzhny, E. Zorina, A. Kopylov, V. Zgoda, O. Kleyzi, A. Archakov, Next Steps on In-Silico 2DE Analyses of Chromosome 18 Proteoforms. *J Proteome Res.* 2018 17 (12): 4085-4096. doi: 10.1021/acs.jproteome.8b00286. PMID: 30228724
2. Naryzhny S. Inventory of proteoforms as a current challenge of proteomics: Some technical aspects. *J. Proteomics* 2018, 21874-21918(18): 30220-3
3. S. Petrov, A. I. Kopylov, I. A. Kopylov, S. G. Kopylov, A. V. Zhdanova, L. A. Pavlov, S. N. Naryzhny. SEARCHING FOR SPECIFIC MARKERS OF GLIOBLASTOMA: ANALYSIS OF PROTEOFORMS OF GLIOBLASTOMA CELLS. *Cell and Tissue Biology* 2018, 12(6): 455-459
4. S. Naryzhny. Book from Electrophoresis - 18e Science Practical Applications (ISBN: 978-953-51-6104-2) Chapter title: Two-dimensional gel electrophoresis (2DE) as an information base for human proteome. 2018 Chapter 06, 91-106. Publisher: InTech
5. S. Naryzhny, V. Zgoda, A. Kopylov, E. Petrov, S. Kibov, A. Archakov. Variety and dynamics of proteoforms in the human proteome: aspects of markers for hepatocellular carcinoma. *Proteomics* 2017, 17(4): 531-541. doi:10.1002/prote.201600003
6. Naryzhny S, Maryashova M, Zgoda V, Archakov A. Database of protein spectra from human liver. *Data in brief*. 2017 May 4; 12: 288-288
7. Naryzhny S.N. Low-resolution Full Realization of 2DE Power. *Review. Proteomics* 2016, 16(4): 335. doi:10.1002/prote.201600003 (registering DOI).
8. Naryzhny S, Zorina E, Kopylov A, Petrov S, Archakov A. Semi-virtual two dimensional gel electrophoresis: IF-ESI LC-MS/MS. *MethodsX*. 2017 Aug 31; 4: 260-264. doi: 10.1016/j.mex.2017.08.004. eCollection 2017. PMID:28912169
9. Stanislav Naryzhny, Mikhey Klopot, Natalia Borotina, Elena Zorina, Victor Zgoda, Olga Kleyzi, Natalia Belyakova, Olga Logina. A database for inventory of proteoform profiles: "2DE-pattern". *Electrophoresis* 2020, 41(10): 1-7

Рисунок 5. Главная страница и страница с информацией о белке EF2 базы данных «2DE-pattern».

Основным отличием базы данных «2DE-pattern» от базы «Proteoform Atlas» является то, что в ней содержатся данные о протеоформах согласно их координатам положения (pI и Mw) на двумерной электрофореграмме, в то время как в «Proteoform Atlas» для идентификации протеоформ используется четкая химическая формула согласно номенклатуре ProForma.

1.9. Заключение литературного обзора

По данным Всемирной организации здравоохранения рак печени является шестым по распространенности среди остальных видов рака и занимает 3 место при причине смертности. По статистике за 2020 год во всем мире было выявлено 905 677 случаев рака печени, при этом смертность составила более 91% (830 180 человек), значительно превосходя уровни смертности среди других видов рака. Ввиду многих ограничений ранней диагностики и лечения данного заболевания многие научные коллективы продолжают поиск как диагностических, так и прогностических биомаркеров, с целью улучшения общей выживаемости пациентов и качества их жизни.

Использование известных методов функциональной диагностики чаще всего позволяют определить лишь стадию заболевания, но при этом сложные молекулярные механизмы возникновения ГЦР остаются недостаточно изученными, а поиск потенциальных маркеров для ранней диагностики затруднен. Применение различных омикс-технологий, таких как транскриптомика и протеомика, позволяют расширить накопленные знания и открывают возможность поиска новых биомаркеров, в том числе более специфических маркеров (протеоформ), а также для идентификации потенциальных мишеней для терапевтического вмешательства. Интегративный анализ омиксных данных в контексте клинических параметров позволяет выявить связи между молекулярными профилями рака и его клиническими характеристиками, такими как прогноз выживаемости, чувствительности к терапии и вероятность рецидива, что особенно важно для разработки методов диагностики и выбора наиболее эффективных терапевтических стратегий. Высокотехнологические аналитические

платформы, создание большого количества аннотированных баз данных и применение современных биоинформатических методов для анализа омиксных данных обеспечивают статистическую валидацию и интерпретацию результатов, делая процесс исследования более эффективным и доступным.

Таким образом, комплексное профилирование обеспечивает глубокое молекулярное понимание ГЦР и открывает новые перспективы для разработки персонализированных подходов в диагностике и лечении.

2. МАТЕРИАЛЫ И МЕТОДЫ

Поиск протеомных профилей и протеоформных паттернов производили с использованием панорамного и секционного протеомного профилирования. Общая схема эксперимента представлена на Рисунке 6.

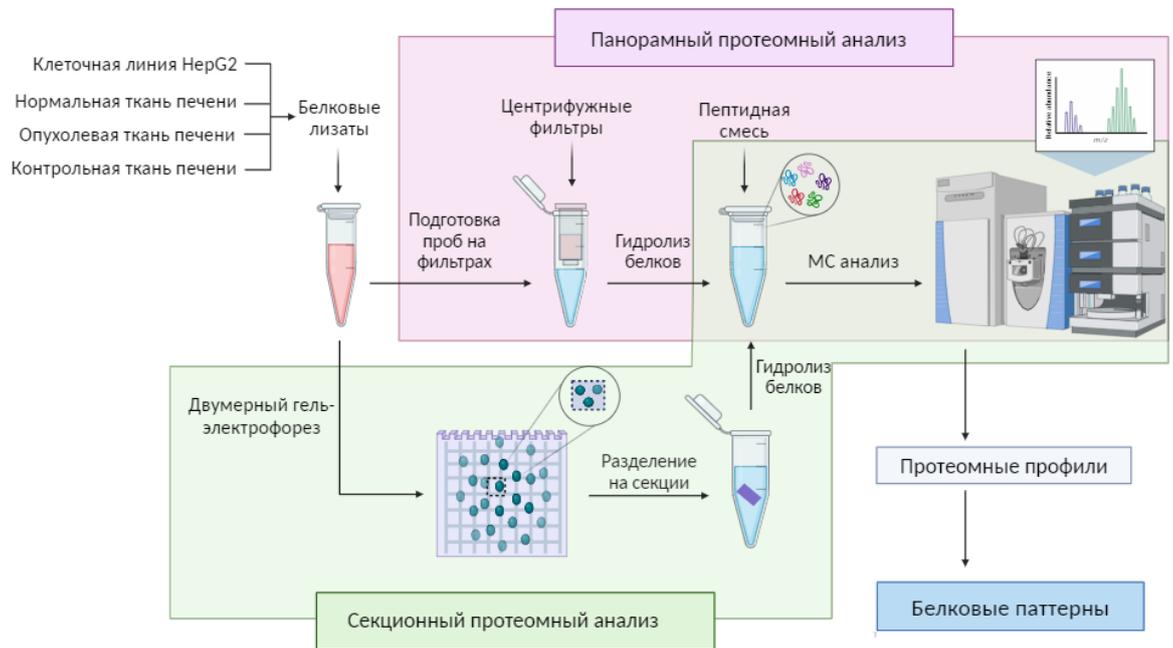


Рисунок 6. Общая схема эксперимента.

Следуя этой схеме для всех исследуемых нами образцов, а именно клеточной линии HepG2, ткани печени от здорового донора и от пациентов с гистологически подтвержденным диагнозом ГЦР, проводили предварительную подготовку проб с целью экстракции белков для последующего анализа. Подготовка проб для панорамного протеомного профилирования была выполнена с использованием центрифужных фильтров, содержащих мембрану, состоящую из ацетата целлюлозы с размерами пор – 30 кДа. Преимуществом такой подготовки является максимальная очистка белковой смеси за счет удаления излишков детергента из клеток, связанных с додецилсульфатом натрия (sodium dodecyl sulfate, SDS), тем самым повышая эффективность гидролиза, поскольку полученная смесь пептидов не содержит длинных фрагментов, получившихся в результате пропуска протеазой сайтов расщепления. Секционное протеомное

профилирование проводили путем предварительного фракционирования белковой смеси 2DE. В последующем гель разделялся на 96 секций, каждая из которых имела свой диапазон pI и Mw. После разделения белки каждой секции гидролизовали трипсином для получения пептидной смеси. Полученные пептидные смеси как при панорамном, так и при секционном профилировании, анализировали на приборе Orbitrap Q-Exactive (Thermo Scientific, США). Биоинформатическую обработку результатов проводили с использованием программного обеспечения SearchGui (v. 3.3.20), а визуализация полученных данных – на платформе PeptideShaker (v. 1.16.45). Поиск последовательностей белков проводили с использованием базы данных полного протеома человека Swiss-Prot. В ходе статистического анализа была сформирована сводная таблица, содержащая информацию по всем анализируемым пробам для каждого из методов профилирования. В дальнейшем на основе протеомных профилей, полученных в ходе секционного профилирования, формировались паттерны протеоформ, представленные в виде трехмерных графиков, отражающих протеомный профиль исследуемого образца.

Более подробная информация об анализируемых пробах и методах приведена далее.

2.1. Характеристика образцов

2.1.1. Клеточная культура НерG2

Клетки линии НерG2 были любезно предоставлены лабораторией клеточной биологии ИБМХ. Клетки культивировали в модифицированной по Дульбекко среде Игла с добавлением среды Хама F12 (DMEM / F12) или в среде RPMI-1640 (Биолот, Россия) с добавлением 10%-ной эмбриональной телячьей сыворотки и 100 ЕД/мл пенициллина при условиях 5% CO₂, 37°C.

2.1.2 Образцы ткани печени

Полная характеристика образцов ткани печени приведена в Приложении 1.

Образец нормальной ткани печени был получен от компании ILSbio, LLC (<http://www.ilsbio.com>). Ткань печени весом 1,5 г получена от здорового донора мужского пола, погибшего в автокатастрофе, через 2 ч. 45 мин. после вскрытия. Транспортировка ткани печени осуществлялась в сухом льду с предварительной заморозкой в жидком азоте.

Образцы ткани печени от пациентов с гистологически подтвержденным диагнозом «гепатоцеллюлярный рак» были предоставлены Первым отделением абдоминальной хирургии и онкологии РНЦХ им Б.В. Петровского. От каждого из пациентов были получены 2 образца – опухолевая ткань и контрольная ткань печени в пробирках, содержащих стабилизирующий раствор RNAlater (Thermo Fisher Scientific, США). Выписка № 01/14/21 из Протокола заседания лечебно-этического комитета Российского научного центра хирургии им. Б.В. Петровского № 14 от 16.12.2021.

До проведения дальнейших исследований образцы хранили при температуре -80°C .

2.2. Предварительная подготовка образцов

2.2.1. Экстракция белков из клеток HepG2

Для экстракции белков клетки (в логарифмической фазе роста) снимали с питательной среды 0,25%-ным раствором трипсин-ЭДТА (этилендиаминтетрауксусная кислота), далее промывали фосфатно-солевым буфером (ФСБ) с ингибиторами протеаз. К клеткам ($\sim 10^7$) добавляли 100 мкл лизирующего буфера (7М мочевины, 2М тиомочевины, 4% CHAPS, 1% ДТТ, 2% амфолиты, pH 3–10, смесь ингибиторов протеаз) и экстрагировали белки. Количество экстрагированного белка составляло ~ 2 мг.

2.2.2. Экстракция белков из ткани печени человека

Для экстракции белков фрагмент ткани печени, весом приблизительно 1 г, помещали в керамическую ступку, заливали жидким азотом и измельчали при

помощи керамического пестика. Полученный порошок переносили в 1,5-мл пробирки (Eppendorf, Германия), примерно по 200 мг на одну пробирку. Для исключения протеолитической и химической деградации белков дальнейшая подготовка проб проводилась на льду. Двукратно к каждой пробирке добавляли 500 мкл ФСБ с ингибиторами протеаз, пипетировали и затем центрифугировали (2 мин, 5000 g, температура 4°C). После центрифугирования надосадочную жидкость удаляли. Полученный осадок растворяли в 600 мкл лизирующего буфера и обрабатывали 6 раз на ультразвуковом гомогенизаторе SONOPULS HD 2070 (Bandelin Electronic, Германия) в режиме: 2 сек. импульс, 2 сек. перерыв. Затем центрифугировали 5 мин. при 10000 g, при температуре 4°C. Надосадочную жидкость отбирали, делили на аликвоты по 100 мкл и хранили при -80°C.

2.3. Подготовка проб для протеомного профилирования

2.3.1. Протокол подготовки проб с использованием центрифужных концентраторов

Подготовку образцов проводили с использованием центрифужных концентраторов с размерами пор 30 кДа, Microcon YM-30 (Merck, США). Для этого пробы белков, полученные в результате экстракции, содержащие необходимое количество общего белка (300 мкг), помещали в центрифужные концентраторы.

Дальнейшая подготовку проб осуществляли в несколько этапов: *восстановление дисульфидных связей* (100 мМ ДТТ в растворе 100 мМ Трис-НСl, рН 8,5), *алкилирование сульфгидрильных групп* (50 мМ йодацетамид в растворе 8 М мочевины, 100 мМ Трис-НСl, рН 8,5) и *гидролиз трипсином* (Trypsin Gold, Promega, США). Каждый из этапов сопровождался предварительным перемешиванием на шейкере Yellowline TTS 2 (ИКА, Германия) в течение 20-30 сек. инкубацией при различных условиях в термомиксере Comfort (Eppendorf, США) и центрифугированием при 9800 g в течение 15 мин. при температуре 20°C (центрифуга 5415R, Eppendorf, Германия). После каждого этапа удаление

последовательно применяемых растворов проводили добавлением 200 мкл промывочного буфера (8 М мочевины в 100 мМ Трис-НСl, рН 8,5).

Восстановление проводили инкубацией в течение 1 ч. при температуре 56°C. Алкилирование – инкубацией в течение 1 ч. в темноте при температуре 20°C. Перед гидролизом пробы дважды промывали добавлением 200 мкл буферного раствора (50 мМ аммоний бикарбонат, рН 8,5).

2.3.2. Двумерный гель-электрофорез

Изоэлектрофокусирование (ИЭФ) проводили с использованием полосок геля с иммобилизованным градиентом рН (IPG-гелевые полоски 7 см (рН 3–11) (GE Healthcare, США)) в соответствии с протоколом производителя. Для этого образцы, подготовленные согласно разделам 2.2.1 и 2.2.2, смешивали с регидрирующим буфером (7 М мочевины, 2 М тиомочевина, 2% CHAPS, 0.3% ДТТ, 2% IPG буфер, рН 3–10, 0,001% бромфеноловый синий) в конечном объеме 150 мкл. Пассивную регидратацию проводили при 4°C в кювете Immobiline DryStrip Reswelling Tray (Amersham Biosciences, США) в течение ночи для максимальной загрузки образца в гель. ИЭФ проводили при 20°C на приборе Hoefer™ IEF100 (Thermo Fisher Scientific, США) с предустановленными настройками производителя для 7-см полосок. ИЭФ завершали при достижении 9000 вольт-часов. Перед разделением по молекулярной массе полоски геля с разделенными белками по рI, помещали в уравнивающий раствор: 50 мМ Трис-НСl, рН 6,8, 6 М мочевины, 2% SDS, 30% глицерин и восстанавливающий агент 1% ДТТ. Затем ДТТ заменяли на алкилирующий агент 5% йодацетамид.

Электрофорез в полиакриламидном геле, в денатурирующих условиях (SDS-PAGE). Для приготовления гелей использовали коммерческий раствор ReadySol IEF 40 % T / 3% C (GE Healthcare, США) в конечной концентрации геля 12%. Раствор 1,5М Трис-НСl (рН 8,8) подготавливали заранее и хранили при 4°C. В качестве инициатора процесса полимеризации использовали 10%-ный персульфат аммония и катализатор тетраметилэтилендиамин (ТЕМЕД). Указанные выше

растворы доводили до конечного объема дистиллированной водой, который составлял для заливочной камеры (gel caster, Amersham Biosciences, США) с размером геля 80x80x1 мм – 30 мл. Для получения ровной поверхности геля при полимеризации после заливки растворов в заливочную камеру сверху осторожно наносили 20%-ный раствор этанола с небольшим количеством бромфенолового синего. Для контроля полимеризации геля небольшое количество заливаемого раствора оставляли в пробирке. Среднее время полимеризации гелей составляло 40 мин.

После полимеризации кассету с гелем помещали во вставку для системы электрофореза Hoefer miniVE (12%-ый гель размером 80x90x1 мм). Согласно положениям электродов, сверху на гель накладывали IPG-гелевую полоску. Прочный контакт между гелем и полоской обеспечивался заливкой 1 мл горячего раствора 0,5%-ной агарозы в электродном буфере (25 мМ Трис, рН 8,3, 200 мМ глицин, 0,1% SDS). Электрофорез проводили при комнатной температуре с переменным напряжением от 100 до 130 V. Время проведения электрофореза составляло 1,5 ч.

Окрашивание гелей. Для визуализации белковых пятен после электрофореза гель предварительно фиксировали в растворе 25%-ого изопропанола с 10%-ной уксусной кислотой в течение 15 мин. для удаления остатков SDS. Окрашивание гелей в кюветах проводили в течение 3 ч., используя краситель Coomassie R350 (0,1% R350, 30% метанол, 10% уксусная кислота). Улучшение визуализации белковых пятен в геле обеспечивалось удалением избыточно окрашенного фона многократной отмывкой водой (по 30 мин.) на орбитальном шейкере KS 260 basic (ИКА, Германия).

Секционный анализ гелей после двумерного гель-электрофореза. Окрашенный гель разделяли на 96 секций в соответствии ранее экспериментально установленными координатами (pI/Mw) позиций следующих белковых пятен, подтвержденных MALDI-TOF MS: актин цитоплазматический (АКТВ, pI 5,29/Mw 42052); белок78 кДа, регулируемый глюкозой (GRP78, pI 5,07/Mw 72333); альфа-3

цепь тропомиозина (TPM3, pI 4,68/Mw 32950); статмин 1 (STMN1, pI 5,76/Mw 17292); альфа-энолаза (ENOА, pI 7,01/Mw 47481) [88].

Каждую из полученных секций геля переносили в 1,5-мл чистые пробирки Eppendorf, разрезая на небольшие фрагменты размерами ~1 мм. Удаление остатков фиксирующего раствора проводили путем добавления 300 мкл 100 мМ бикарбоната аммония. Для исключения влияния красителя на результат, фрагменты геля с окрашенными пятнами, промывали 300 мкл раствора 50%-ного ацетонитрила с 50 мМ бикарбоната аммония. При сильном окрашивании фрагментов этот этап повторяли два раза. Затем раствор заменяли 100%-ным ацетонитрилом (150 мкл). Полученные фрагменты высушивали в вакуумной центрифуге SpeedVac (Eppendorf, Германия) в течение 15 мин.

2.4. Гидролитическое расщепление белков

2.4.1. Гидролиз с использованием центрифужных концентраторов

Для гидролиза использовали раствор трипсина (Trypsin Gold, Promega, США) в буферном растворе: 50 мМ аммоний бикарбонат, pH 8,5. Концентрацию трипсина, необходимую для гидролиза, принимали из расчета массового соотношения общая масса фермента / общая масса белка – 1/100. Инкубацию проводили в течение ночи при температуре 37°C. Для повышения качества гидролиза образцов на следующий день добавляли раствор трипсина в том же массовом соотношении и инкубировали дополнительно в течение 2 ч. при температуре 37°C. Для завершения гидролиза к гидролизатам на фильтрах добавляли 50 мкл 30%-ого раствора муравьиной кислоты. Полученный раствор пептидов переносили в чистые стеклянные вставки объемом 250 мкл (Agilent, США). До дальнейшего анализа раствор пептидов высушивали в вакуумном концентраторе (Concentrator 5301, Eppendorf, Германия) при 45°C.

Перед проведением МС-анализа высушенные пептиды растворяли в 20 мкл 5%-ого раствора муравьиной кислоты. Конечную концентрацию принимали

равной 1 мкг/мкл, исходя из расчетов, что масса общего белка, взятого для анализа равна массе пептидов в пробе.

2.4.2. Гидролиз белков из геля

Гидролиз белков в высушенных кусочках геля проводили путем их инкубации в течение ночи при 37°C в растворе трипсина (Trypsin Gold, 10 г/мл) в 25 мМ бикарбонате аммония. Для экстракции пептидов добавляли 200 мкл 60%-ного ацетонитрила, содержащего 0,1%-ную трифторуксусную кислоту. С целью повышения качества экстракции эту процедуру проводили два раза. Полученный раствор пептидов переносили в чистые виалы и концентрировали в вакуумной центрифуге, избегая полного высыхания. Восстановление пептидов для дальнейшего МС-анализа проводили путем добавления к каждому осадку 20 мкл 5%-ной муравьиной кислоты.

2.5. Масс-спектрометрический анализ (LC-ESI-MS/MS)

Пробы, подготовленные согласно разделу 2.4, предварительно фракционировали на хроматографической системе Agilent HPLC серии 1100 (Agilent Technologies, США). Для этого пептидную смесь (~ 4 мкг) загружали на колонку Zorbax 300SB-C18, 5x0,3 мм (Agilent Technologies, США), затем промывали 5%-ным ацетонитрилом, содержащим 0,1% муравьиную кислоту, при скорости потока 300 нл/мин.

Пептиды разделяли на обращено-фазовой колонке Zorbax 300SB-C18, 150x75 мм (Agilent Technologies, США), варьируя время потока градиента (5–60% ацетонитрил и 0,1% муравьиная кислота). Для проб, подготовленных согласно разделу 2.4.1 (на центрифужных концентраторах), использовался 90-мин градиент, а для проб, подготовленных согласно разделу 2.4.2 (гидролиз из геля) – 60-мин. градиент. Скорость потока составляла 300 нл/мин.

МС-анализ проводили на приборе Orbitrap Q-Exactive (Thermo Scientific, США). Регистрацию масс-спектров проводили в режиме положительных ионов с

разрешением 30000 (m/z 400) для масс-спектров родительских ионов (МС 1) и с разрешением 7500 (m/z 400) для дочерних ионов (МС 2). Анализ проводили в двух технических повторах для секционного протеомного профилирования, в трех технических повторах (образцов тканей печени пациентов с ГЦР) и двух технических повторах (клеток линии HepG2 и ткани нормальной печени) для панорамного протеомного профилирования.

2.6. Биоинформатический анализ данных

Биоинформатический анализ данных для идентификации белков и пептидов для всех файлов МС проводили с использованием программного обеспечения SearchGui (v.3.3.20) с поисковыми системами OMSSA (v. 2.1.9) и MS-GF+ (v. 2018.04.09) со следующими параметрами: фермент – трипсин; максимум пропущенных разрезов – 2; фиксированные модификации – карбамилрование цистеина; переменные модификации – окисление метионина, фосфорилирование серина, треонина, триптофана, ацетилирование лизина, дезамидирование аспарагина; диапазон ошибки массы родительского иона – 20 ppm; ошибка массы фрагментов – 0,01 Да. Поиск последовательностей проводили с использованием базы данных полного протеома человека Swiss-Prot (по состоянию на 22.08.2022 г.), содержащей 20 376 последовательностей канонических белков и 21 975 сплайс-вариантов, а также обращенные последовательности для применения подхода «мишень-приманка» [98]. Данный подход применяли для оценки доли ложноположительных результатов FDR (false discovery rate), в качестве отсечения принимали значение FDR 1% для регистрации белков. Однонуклеотидные замены не были включены в базу данных, по причине того, что при проведении двумерного гель-электрофореза такие замены не несут значительного изменения белкового заряда молекулы (pI), что позволило использовать стандартную базу данных для поиска белков и их пептидов.

Визуализация полученных данных была выполнена на платформе PeptideShaker (v.1.16.45). Для последующего анализа идентифицированными

считали белки, для которых было обнаружено не менее 2 уникальных пептидных последовательностей.

2.7. Нормализация данных секционного протеомного профилирования

Количественная оценка относительного содержания белка выражена в величине экспоненциально модифицированного индекса содержания белка (emPAI, exponential modified protein abundance index), которая рассчитывалась с помощью встроенного алгоритма в SearchGui.

Вычисление абсолютного количественного содержания белка при секционном профилировании проводили путем нормализации emPAI для протеоформных паттернов каждого отдельного белка к сумме значения emPAI всех идентифицированных протеоформ в образце. Вычисления были проведены в программе Microsoft Excel 2010 по следующей формуле:

$$\text{Нормализованный emPAI} = \frac{\sum_{\text{emPAI}} p_1 + p_2 + p_3 + \dots + p_n}{\sum_{\text{emPAI}}} \times 100 \% \quad (1)$$

где, $\sum_{\text{emPAI}} p_1 + p_2 + p_3 + p_n$ – сумма значений emPAI протеоформ, относящихся к определенному белок-кодирующему гену (БКГ), а \sum_{emPAI} сумма значений emPAI всех идентифицированных протеоформ в образце. Полученная величина нормализованного emPAI отражает абсолютное количественное содержание белка в молярной процентной концентрации для определенного БКГ в образце.

Например, белок катепсин В (Cathepsin B, CTSB) в опухолевом образце от первого пациента был представлен 8 протеоформами в секциях А6, А8, А12, В6, В7, В8, В9 и В12 (emPAI протеоформ соответственно – 1,03, 0,42, 0,56, 0,86, 1,89, 0,19 и 0,43). Суммарный emPAI для данного белка – 5,38, а суммарный emPAI образца – 5652,08. Таким образом мы получили оценку относительно содержания данного белка в образце – 0,09 %.

2.8. Анализ и статистическая обработка данных

Для дальнейшего анализа составляли сводную таблицу, содержащую результаты по всем анализируемым пробам для каждого из методов – секционного и панорамного протеомного профилирования (Приложение 2). Данные по техническим повторам усредняли, полный список составил 5753 идентифицированных белка. Для дальнейшего статистического анализа анализируемые образцы были разделены на две группы: клеточная линия HepG2/нормальная ткань печени, и опухолевая/контрольная ткань печени. Для предварительного анализа данных использовали описательную статистику, которая включала в себя следующие параметры: среднее значение, медиана, стандартное отклонение. Проверку данных на соответствие нормальному распределению проводили с использованием критерия Колмогорова-Смирнова. Статистическую оценку достоверности результатов проводили с использованием критерия Стьюдента с уровнем значимости $p < 0,05$. Для дальнейшего анализа в каждой из анализируемых групп были отобраны белки, соответствующие следующим критериям: кратное изменение (Fold change, FC) $\geq 1,5$ и/или присутствие только в злокачественных клетках.

Визуализация статистических данных была выполнена в программе Graph Pad Prism (версия 8.0.1) (<https://www.graphpad.com>).

Функциональная аннотация идентифицированных белков по трем категориям: биологические процессы, молекулярные функции и белковые классы, выполнена с использованием веб-ресурса PANTHER [100].

Рисунки выполнены с использованием веб-сервиса BioRender (<https://www.biorender.com>) и Microsoft Powerpoint.

Графическое представление протеоформных паттернов выполнено в виде трехмерных графиков в программе Microsoft Excel.

3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1. Протеомное профилирование белков нормальной ткани печени и белков клеточной линии HepG2

3.1.1. Панорамное протеомное профилирование

Ввиду того, что в данной работе использовали различные типы биоматериала (культура клеток и ткань печени), необходимым условием выбора протокола подготовки проб для МС-анализа была его применимость для различных типов биоматериала. В настоящее время существует много различных протоколов, которые оптимизированы для таких типов исследований. Одним из них является протокол подготовки проб с использованием центрифужных концентраторов (filter-aided sample preparation, FASP). Центрифужный концентратор Microcon YM-30 представляет собой мембрану из ацетата целлюлозы с диаметром пор 30 кДа. Такой диаметр пор является наиболее применимым, позволяя одновременно провести идентификацию и количественное измерение множества белков в образце в довольно широком диапазоне концентраций, от нескольких микрограммов до нескольких миллиграммов. Важное преимущество состоит в том, что использование мембран позволяет очищать белок от излишков детергента (например, SDS), повышает качество ферментативного гидролиза белков и позволяет концентрировать белки [101].

Образцы нормальной ткани печени и клеточной линии HepG2, подготовленные согласно разделу 2.4.1, разделяли на аликвоты, содержащие одинаковое количество белка (1 мг), и анализировали согласно разделу 2.5. Данный подход позволил нам получить информацию о протеомных профилях исследуемых образцов, а также провести их сравнительный анализ. В результате панорамного протеомного профилирования для клеточной линии HepG2 было идентифицировано 867 белков, а для нормальной ткани печени – 797. Совпадающее количество составило 401 белок. При этом 466 и 396 белков были

идентифицированы только в клетках линии HepG2 или нормальной ткани печени соответственно (Рисунок 7).

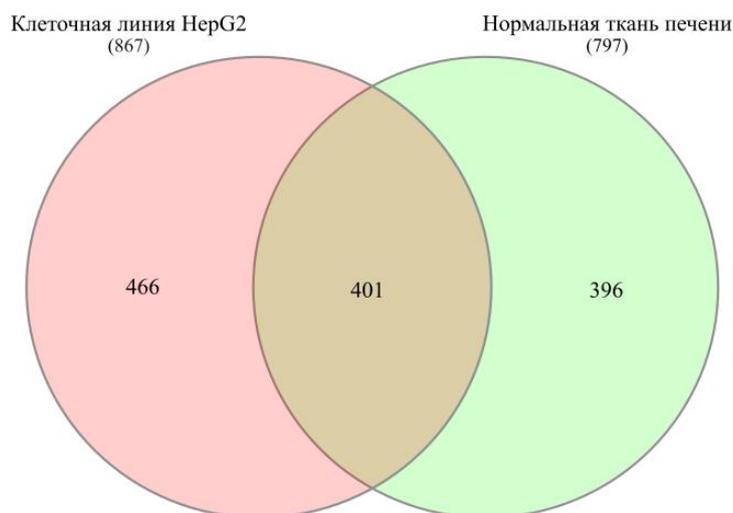


Рисунок 7. Количество белков, идентифицированных в клеточной линии HepG2 и нормальной ткани печени панорамным протеомным профилированием.

Такая разница в количестве идентифицированных белков может быть связана как с аналитическими ограничениями (недостаточное содержание белка для его идентификации), так и с тем, что клетки линии HepG2 содержат белки, характерные только для данной клеточной линии [102].

Количество белков, ассоциированных со злокачественными новообразованиями, значительно варьирует в зависимости от используемых критериев и базы данных. Однако, согласно различным источникам и базам данных, таких как Human Protein Atlas (HPA), The Cancer Genome Atlas (TCGA) можно выделить несколько тысяч белков, ассоциированных с различными видами рака. Ранее нами была проведена работа по анализу собственных и литературных данных для поиска белков в клетках HepG2, которые либо являются известными белковыми маркерами ГЦР, либо относятся к прогностическим белковым маркерам [103]. Основываясь на полученных результатах, мы отобрали 19 белков, которые в дальнейшем были проанализированы секционным протеомным профилированием для получения более детальной информации (Таблица 3).

Таблица 3. Список некоторых известных и потенциальных биомаркеров ГЦР.

№	Белок	Ген	Номер UniProt	Название	Ссылка
1	ANXA2	<i>ANXA2</i>	P07355	Annexin A2	[104]
2	CATB	<i>CTSB</i>	P07858	Cathepsin B	[9]
3	CYB5	<i>CYB5A</i>	P00167	Cytochrome b5	[9]
4	FETA	<i>AFP</i>	P02771	Alpha-fetoprotein Alpha-fetoprotein isoform Iso L3	[9,37,105–108]
5	FIBG	<i>FGG</i>	P02679	Fibrinogen gamma chain	[109]
6	FUCO2	<i>FUCA2</i>	Q9BTY2	Plasma alpha-L-fucosidase	[110]
7	GOLM1	<i>GOLM1</i>	Q8NBJ4	Golgi membrane protein 1	[111]
8	GPC3	<i>GPC3</i>	P51654	Glypican-3	[9,105,108]
9	HPT	<i>HP</i>	P00738	Haptoglobin	[9]
10	HSP74	<i>HSPA4</i>	P34932	Heat shock 70kDa protein 4	[108]
11	HSPB1	<i>HSPB1</i>	P04792	Heat shock protein beta-1	[94,97]
12	KITH	<i>TK1</i>	P04183	Thymidine kinase, cytosolic	[110]
13	OSTP	<i>SPP1</i>	P10451	Osteopontin	[116,117]
14	PGRC1	<i>PGRMC1</i>	O00264	Membrane-associated progesterone receptor component	[9]
15	SART3	<i>SART3</i>	Q15020	Squamous cell carcinoma antigen recognized by T- cells 3	[118]
16	SRC	<i>SRC</i>	P12931	Proto-oncogene tyrosine- protein kinase Src	[119]
17	SRPK1	<i>SRPK1</i>	Q96SB4	SRSF protein kinase 1	[120]
18	THRB	<i>F2</i>	P00734	Des-gamma- Carboxyprothrombin	[9,105]
19	ZA2G	<i>AZGP1</i>	P25311	Zinc-alpha-2-glycoprotein	[108]

Панорамным протеомным профилированием из данного списка нам удалось идентифицировать только шесть белков в клетках HepG2, а именно, PGRC1, CYB5, FIBG, FETA, CATB, HSP74. Из них FETA и HSP74 обнаружены только в клетках HepG2, причем для CATB было показано значительное увеличение содержания в клетках HepG2 (FC>7) по сравнению с нормальной тканью печени. Для PGRC1, CYB5 и FIBG наблюдали значительное повышение содержания в

ткани печени ($FC > 1,7$). Данные белки преимущественно экспрессируются в печени, и по различным данным, повышение их уровня экспрессии при ГЦР, свидетельствует о более агрессивном характере опухоли и возможном метастазировании [109,121,122].

3.3.2. Секционное протеомное профилирование

Модификации белка, посттрансляционные или генетические, могут приводить к изменению молекулярной массы или изоэлектрической точки белка. В этом случае будет меняться положение белка в геле при электрофоретическом разделении. Как было сказано в разделе 1, для идентификации белка достаточно обнаружение в масс-спектрах нескольких его пептидов, а белок идентифицируется с точностью до гомолога. Идентификация одного и того же белка в различных участках геля будет указывать на существование модификаций этого белка, его изоформ или протеоформ. Поэтому для детального анализа протеомных профилей мы применили секционное протеомное профилирование.

В процессе проведения 2DE для минимизации образования артефактов при подготовке проб использовался алкилирующий агент (йодацетамид), который предотвращает повторное окисление белка во время разделения по молекулярной массе и алкилирует остаточный ДТТ для минимизации образования вертикальных полос на геле. После разделения образцов двумерным электрофорезом и окрашивания гель разрезали на 96 секций согласно ранее определенным координатам мажорных белковых пятен (Рисунок 8). Далее каждую секцию геля обрабатывали согласно протоколу, описанному в разделе 2.4.2, и в дальнейшем анализировали с помощью LC-ESI-MS/MS. Такой подход за счет предварительного фракционирования перед проведением МС, позволяет идентифицировать одновременно десятки и сотни белков, а также получить протеоформный паттерн исследуемого образца.

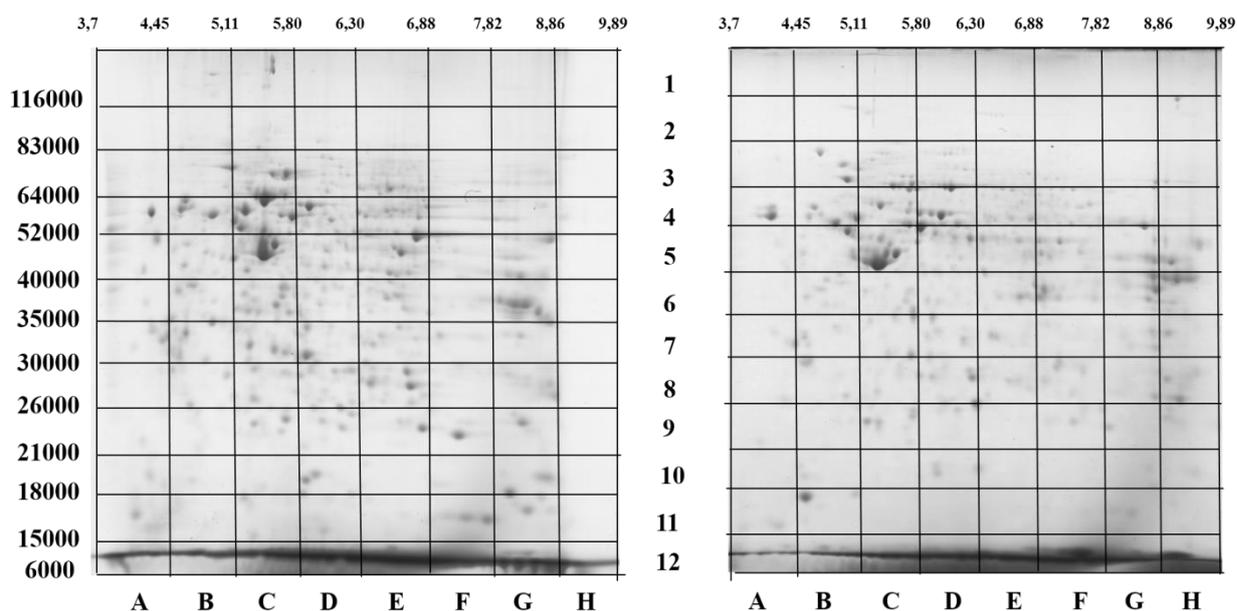


Рисунок 8. Двумерные электрофореграммы белков HepG2 (слева) и белков нормальной ткани печени (справа).

По оси абсцисс – изоэлектрическая точка (pI), ось ординат – молекулярная масса (kDa).

Суммарно во всех ячейках геля для клеточной линии HepG2 было идентифицировано 18992 протеоформы, кодируемые 3390 генами, для нормальной ткани печени – 20504 протеоформы, кодируемые 2482 генами. Количество идентифицированных протеоформ для клеток линии HepG2 согласуется с результатами транскриптомно-протеомного профилирования, полученными Киселевой О.И. и др. [123]. Небольшая разница в количестве идентифицированных протеоформ, по-видимому, обусловлена использованием в нашей работе стандартной библиотеки и комбинации других поисковых алгоритмов с целью оптимизации машинного времени обработки данных.

Графическое представление протеоформных паттернов было выполнено в виде трехмерных графиков в программе Microsoft Excel. На графиках, как и в случае исходной двумерной электрофореграммы, отображены положения протеоформ согласно параметрам ячейки, в которой они были детектированы (изоэлектрическая точка и молекулярная масса). Оценка их относительного количественного содержания выражена в значении emPAI и на трехмерных графиках представлена по оси аппликат (Рисунок 9).

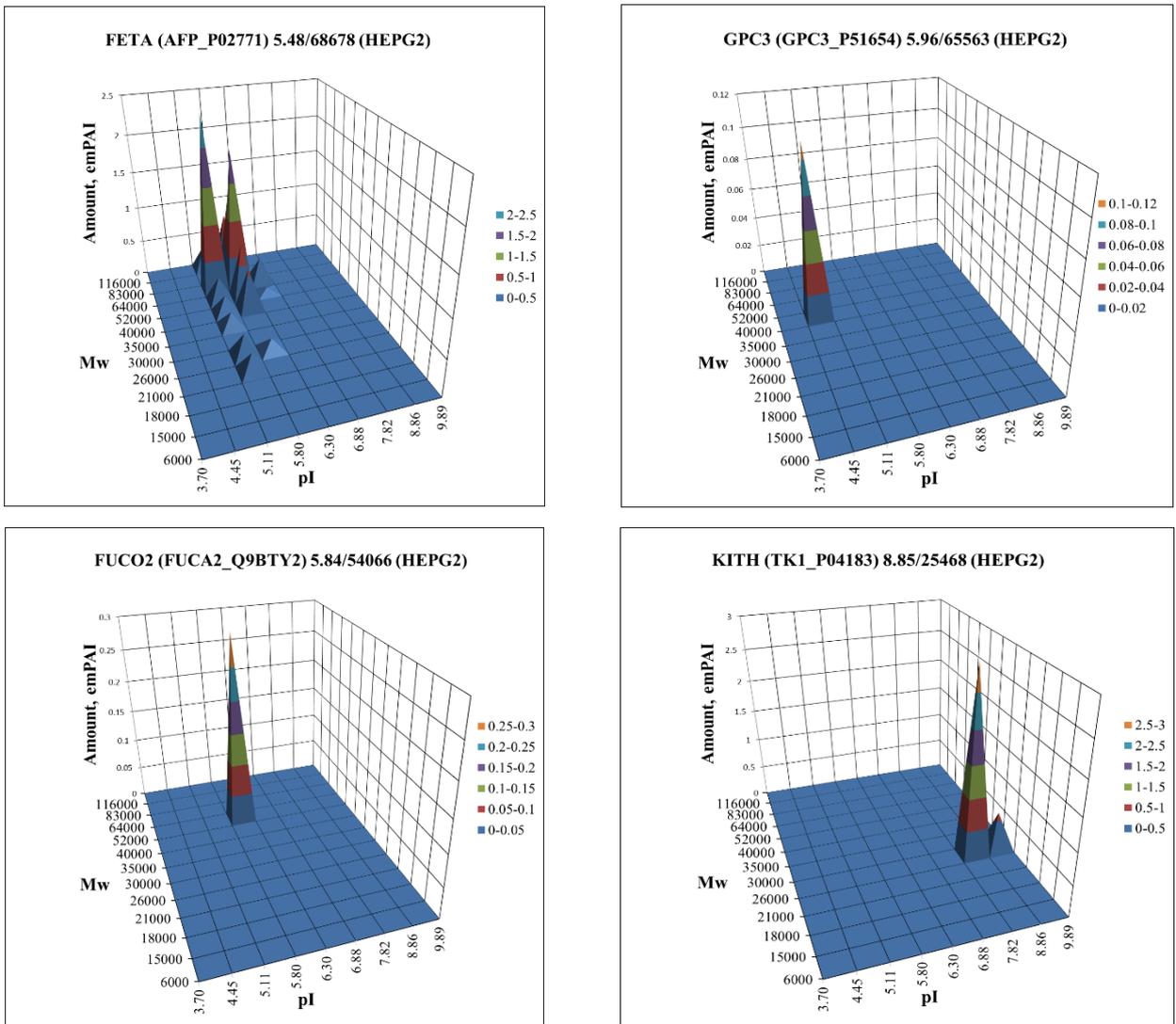


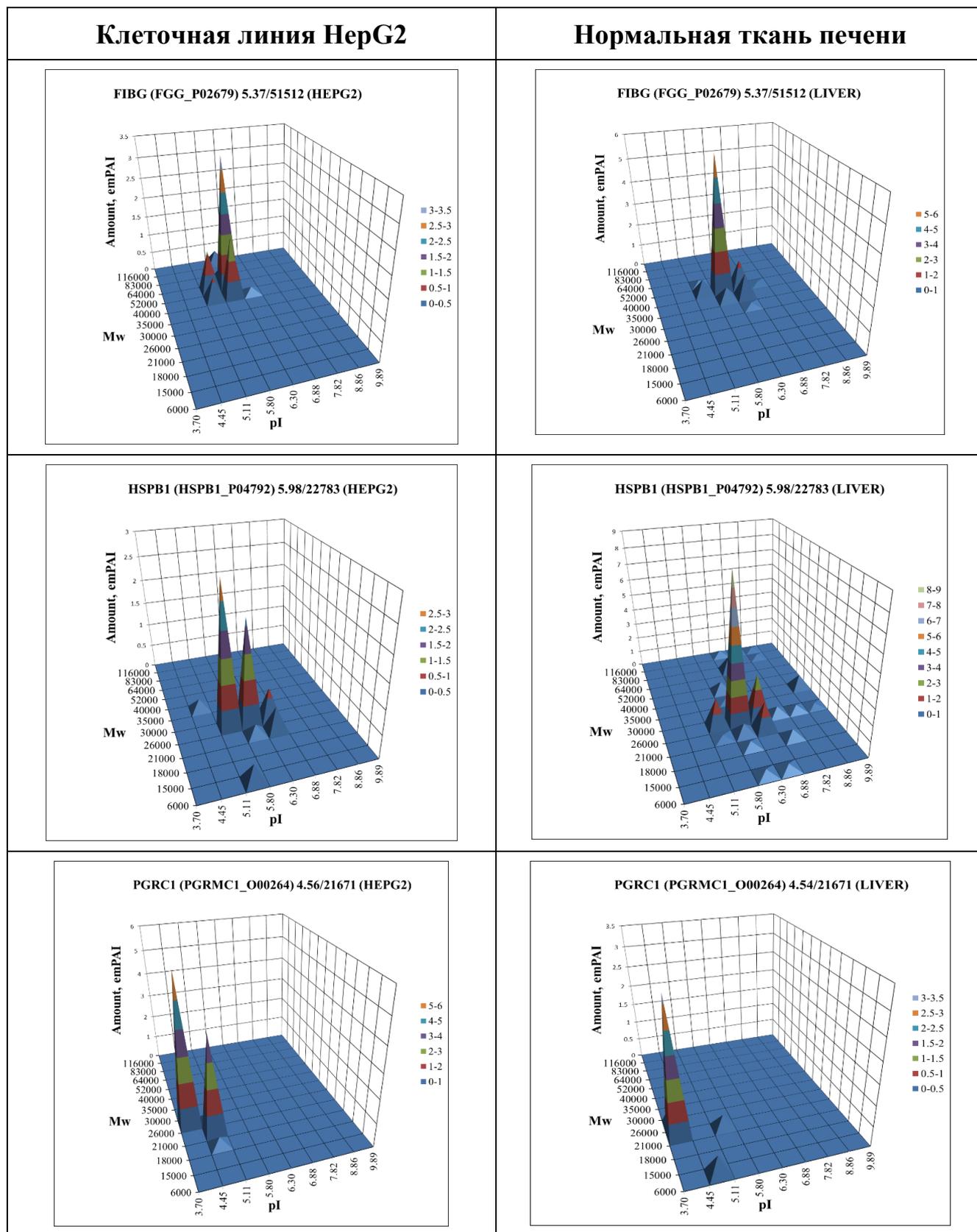
Рисунок 9. Протеоформные паттерны некоторых биомаркеров из Таблицы 1, детектируемые только в клетках HepG2. В верхней части каждого графика указано название белка, номер UniProt и теоретические параметры pI/Mw. По оси абсцисс – изоэлектрическая точка (pI), ось ординат – молекулярная масса (кДа), по оси аппликат – уровень белка (emPAI).

Практический интерес в визуализации таких профилей заключается в том, что мы получаем протеоформный паттерн для каждого белка, присутствующего в образце, что позволяет оценить вклад протеоформ в общий протеомный профиль. При идентификации протеоформы в ячейке, которая не соответствует своим каноническим параметрам исходного белка (молекулярной массы и изоэлектрической точки), возможно, спрогнозировать события, которые могли

привести к данным изменениям, в том числе потенциальные ПТМ. Визуализация такого «штрих-кода» значительно упрощает последующий направленный анализ протеоформ, представляющих практический интерес как в научных, так и клинических исследованиях.

Для каждого из идентифицированных белков как для клеток линии HepG2, так и для нормальной ткани печени, мы получили протеоформные паттерны, сравнительный анализ которых позволил выделить наиболее изменяющиеся профили между тканью печени и клетками HepG2. Под изменением протеоформного паттерна следует рассматривать присутствие протеоформ в секциях геля, отличных от канонических параметров исходного белка. Из числа известных и прогностических маркеров ГЦР, представленных в Таблице 2, секционным профилированием нам удалось получить протеоформные паттерны для 17 белков, достоверно идентифицируемых по 2 и более пептидам (Таблица 4). Для двух белков, GPC3 и FUCO2, получены протеоформные паттерны, представленные 12 протеоформами и 1 протеоформой, соответственно, идентифицированные по 1 протеотипическому пептиду, т.е. принадлежащие только данному белку. Протеоформы для белков OSTP и GOLM1 не удалось идентифицировать в данном эксперименте. Особый интерес представляли белки из Таблицы 2, которые не были обнаружены в нормальной ткани печени. К ним относятся шесть белков, а именно альфа-фетапротеин (FETA), глипикан 3 (GPC3), альфа-L-фукозидаза плазмы (FUCO2), цитозольная тимидинкиназа (KITH), протоонкогенная тирозин-протеинкиназа SRC (SRC) и SRSF протеинкиназа 1 (SRPK1). В полученных нами данных, протеоформы перечисленных белков, кроме SRC и SRPK1, представлены в виде протеотипических пептидов. Протеоформы и соответствующие им пептиды для SRC и SRPK были представлены в виде белковых групп, и принадлежали всем трем изоформам для данных белков.

Таблица 4. Некоторые протеоформные паттерны, полученные для клеток HepG2 и нормальной ткани печени. В верхней части каждого графика указано название белка, номер UniProt и теоретические параметры pI/Mw. По оси абсцисс – изоэлектрическая точка (pI), по оси ординат – молекулярная масса (kDa), по оси аппликат – уровень белка (emPAI).



В результате нам не только удалось обнаружить большее количество белков, входящих в список известных и прогностических белковых маркеров ГЦР, но детализировать молекулярный профиль для клеток HepG2 и нормальной ткани печени для дальнейшего детального анализа.

3.3.3. Комбинирование результатов секционного и панорамного протеомного профилирования

Использование интегративного подхода для анализа сложных биологических объектов в протеомике позволяет проводить всесторонний анализ исследуемого объекта. Основываясь на экспериментальных данных, мы провели более детальный анализ полученных протеомных профилей не только для литературно известных белков и потенциальных биомаркеров, но и для наиболее измененных белков между клетками HepG2 и нормальной ткани печени. Основным критерием для анализа белков было изменение протеоформных паттернов между исследуемыми образцами и изменение относительного содержания белка более чем в 1,5 раза. Для этого данные, полученные в ходе секционного профилирования, нормализовали согласно разделу 2.7 и формировали таблицу, отражающую протеомный профиль, полученный двумя подходами. Среди более 3000 белков мы выделили 25 белков, соответствующих вышеуказанным критериям (Таблица 5). В данный список включен только 1 белок из проанализированных ранее, литературно известных биомаркеров, а именно АФП, который был обнаружен только в клетках HepG2, но не в печени.

Таблица 5. Экспериментально расширенный список известных и потенциальных биомаркеров ГЦР. FC – кратное изменение, СП – секционное профилирование, ПП – протеомное профилирование.

* – обнаружены только в клетках HepG2, но не в печени .

Номер Uniprot	Белок	Ген	Название	FC, (СП)	FC, (ПП)
P53396	ACLY	<i>ACLY</i>	ATP-citrate synthase	20,39	*
O60488	ACSL4	<i>ACSL4</i>	Long-chain-fatty-acid-CoA ligase 4	*	*
O60218	AK1BA	<i>AKR1B10</i>	Aldo-keto reductase family 1 member B10	56,04	*
P12429	ANXA3	<i>ANXA3</i>	Annexin A3	2,29	*
O15144	ARPC2	<i>ARPC2</i>	Actin-related protein 2/3 complex subunit 2	1,89	2,20
P40121	CAPG	<i>CAPG</i>	Macrophage-capping protein	5,42	*
P58107	EPIPL	<i>EPPK1</i>	Epiplakin	10,26	*
P02771	FETA	<i>AFP</i>	Alpha-fetoprotein	*	*
Q10471	GALT2	<i>GALNT2</i>	Polypeptide N-acetylgalactosaminyl transferase 2	*	*
P05546	HEP2	<i>SERPIND1</i>	Heparin cofactor 2	*	*
P14618	KPYM	<i>PKM</i>	Pyruvate kinase PKM	12,21	*
P17931	LEG3	<i>LGALS3</i>	Galectin-3	5,20	*
P23368	MAOM	<i>ME2</i>	NAD-dependent malic enzyme, mitochondrial	9,31	*
Q32P28	P3H1	<i>P3H1</i>	Prolyl 3-hydroxylase 1	11,67	*
Q8NC51	PAIRB	<i>SERBP1</i>	Plasminogen activator inhibitor 1 RNA-binding protein	132,38	*
P01833	PIGR	<i>PIGR</i>	Polymeric immunoglobulin receptor	21,62	*

Продолжение таблицы 5.

Номер Uniprot	Белок	Ген	Название	FC, (СП)	FC, (ПП)
Q96M27	PRRC1	<i>PRRC1</i>	Protein PRRC1	4,77	*
Q96A35	RM24	<i>MRPL24</i>	39S ribosomal protein L24, mitochondrial	16,35	*
Q96GC5	RM48	<i>MRPL48</i>	39S ribosomal protein L48, mitochondrial	3,04	*
P25815	S100P	<i>S100P</i>	Protein S100-P	*	*
Q13243	SRSF5	<i>SRSF5</i>	Serine/arginine-rich splicing factor 5	*	*
P42224	STAT1	<i>STAT1</i>	Signal transducer and activator of transcription 1-alpha/beta	2,04	*
Q9Y2Z4	SYYM	<i>YARS2</i>	Tyrosine-tRNA ligase, mitochondrial	9,44	*
P02786	TFR1	<i>TFRC</i>	Transferrin receptor protein 1	*	*
P02788	TRFL	<i>LTF</i>	Lactotransferrin	1,75	*

Использование предварительного фракционирования исследуемого образца методом 2DE не только позволяет получить более детальный молекулярный профиль исследуемого образца, но и также повысить количество обнаруженных белков. При сравнении результатов секционного и панорамного профилирования даже без учета идентифицированных протеоформ для каждого белка, было показано, что применение секционного протеомного профилирования позволило идентифицировать большее количество белков как для клеточной линии HepG2 (Рисунок 10), так и для нормальной ткани печени (Рисунок 11).

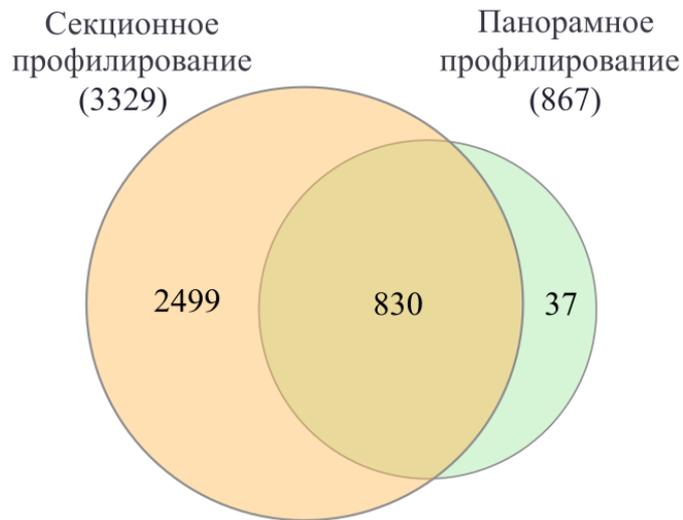


Рисунок 10. Количество белков, идентифицированных в клеточной линии HepG2 секционным и панорамным протеомным профилированием.

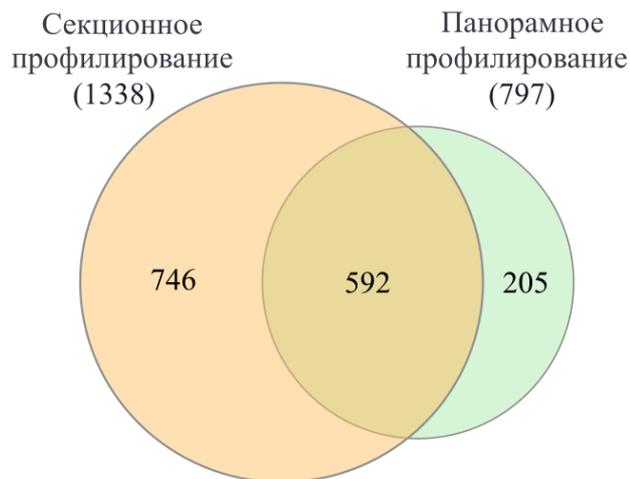


Рисунок 11. Количество белков, идентифицированных в нормальной ткани печени секционным и панорамным протеомным профилированием.

3.4. Протеомное профилирование белков в образцах пациентов с ГЦР

Согласно данным Американской Коллекции Типовых Культур (АТСС), клеточная линия HepG2 используется в качестве модели ГЦР⁶, что

⁶ <https://www.atcc.org/products/hb-8065> описание клеточной линии HepG2

подтверждается многочисленными публикациями различных научных коллективов. Поэтому для расширения и возможной применимости полученных результатов на белках клеток HepG2, мы провели такую же серию анализов (панорамным и секционным протеомным профилированием) образцов, полученных от пациентов с ГЦР.

3.4.2. Панорамное протеомное профилирование

Панорамное протеомное профилирование образцов, полученных от пациентов, позволило суммарно идентифицировать 1306 белков, из которых 788 белков были идентифицированы как в опухолевой, так и в контрольной ткани печени. При этом 324 и 194 белка были идентифицированы только в опухолевой и контрольной ткани печени, соответственно (Рисунок 12).

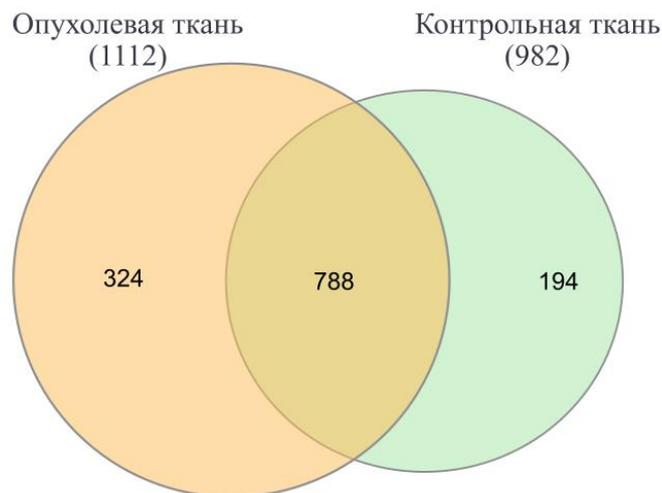


Рисунок 12. Количество белков, идентифицированных в опухолевой и контрольной ткани печени классическим протеомным профилированием.

При более детальном рассмотрении, для каждого из пациентов наблюдается примерно одинаковое количество идентифицированных белков (Рисунок 13).

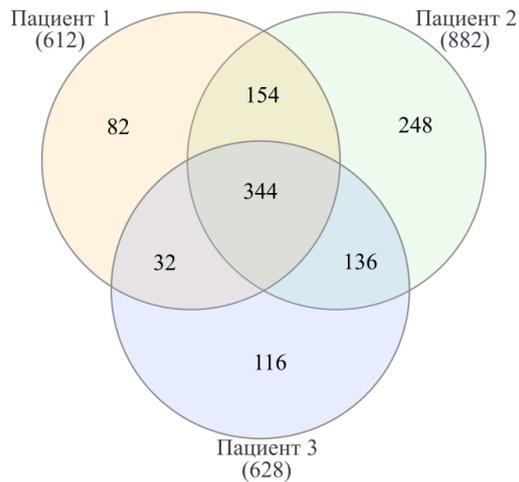
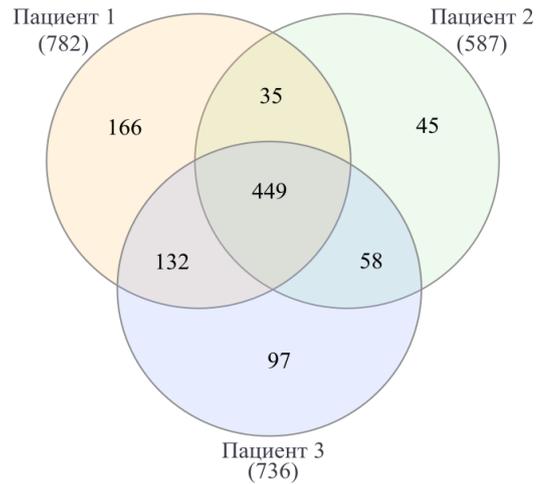
А**Б**

Рисунок 13. Количество идентифицированных белков классическим протеомным профилированием. **А:** опухолевая ткань печени, **Б:** контрольная ткань печени.

Из числа идентифицированных для каждого пациента белков, представленных на Рисунке 13, как в опухолевой ткани, так и в контрольной ткани печени, 260 белков были общими. Согласно классификации GO по белковым классам, эти белки в основном представлены ферментами взаимопревращения метаболитов (оксидоредуктазы, трансферазы и гидролазы), транспортными белками, шаперонами и белками цитоскелета.

Несмотря на гетерогенность опухолевой ткани, общее число копий клеток и баланс между белками с разным числом копий остается неизменным. Такая система белковой регуляции, называемая протеостазом, является важным фактором в регуляции различных процессов, связанных с биосинтезом, сборкой, транспортировкой и деградацией белка. В опухолевых клетках поддержание протеостаза особенно важно, поскольку они подвергаются повышенному уровню протеотоксического стресса из-за их быстрого деления и измененного метаболизма [124]. Такие белки играют значительную роль в ремоделировании протеома в зависимости от условий окружающей среды, и поэтому при злокачественных трансформациях важно представить уровни относительного

количественного содержания всех белков. На Рисунке 14 представлен график относительного количественного содержания белков (emPAI), которые были определены у всех пациентов (1306 белков), и, как уже ранее было сказано, мы видим, что баланс между белками с разным числом копий при различных состояниях между белками сохраняется.

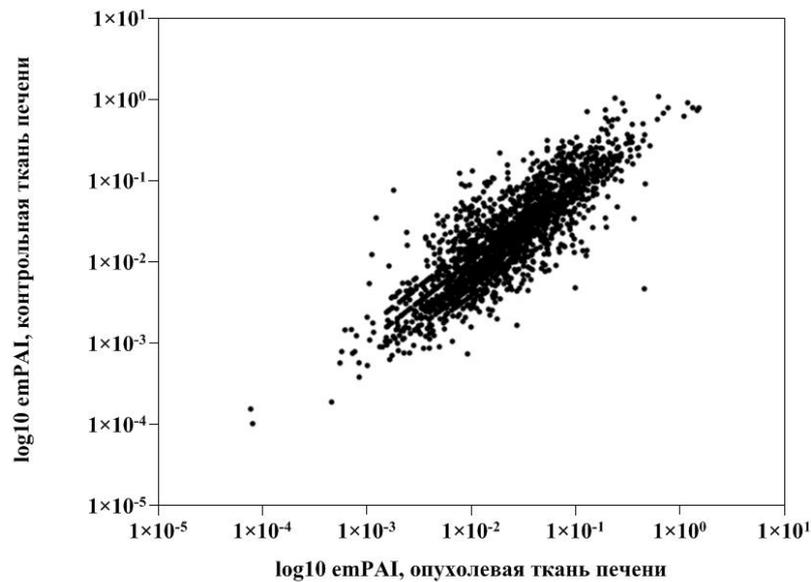


Рисунок 14. График относительного количественного содержания белка (emPAI) между образцами опухолевой и контрольной ткани печени.

Для обеспечения и воспроизводимости полученных результатов, были проведены биологические повторы проб от первого пациента (опухолевая и контрольная ткань печени). В результате в биологическом повторе нам удалось идентифицировать большее количество белков, как для опухолевой ткани, так и для контрольной ткани печени (Рисунок 15).

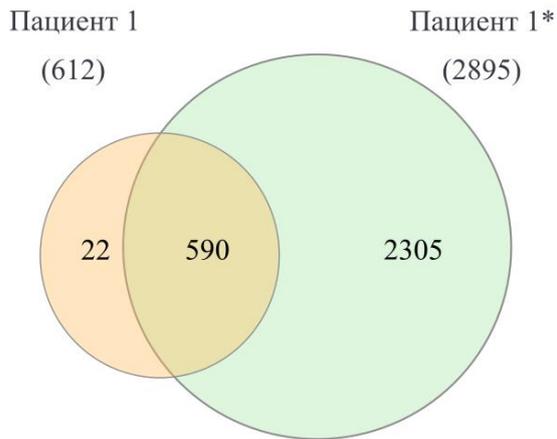
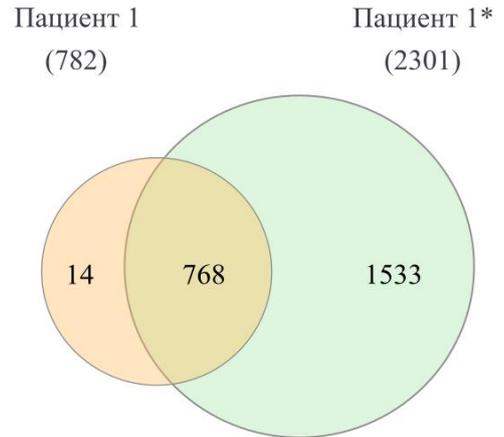
А**Б**

Рисунок 15. Количество белков, идентифицированных для первого пациента. **А:** опухолевая ткань печени, **Б:** контрольная ткань печени.

* биологический повтор проб.

При равных условиях эксперимента вероятной причиной улучшения качества анализа является замена предварительной колонки для фракционирования перед проведением МС. Тем не менее, практически все ранее идентифицированные белки были обнаружены. Исключение составили 22 белка в опухолевой ткани и 14 белков в контрольной ткани печени. Таким образом, подбор условий предварительного фракционирования может значительно улучшать качество анализа [125]. При этом полученные результаты практически не изменили уровни относительного количественного содержания всех белков, и следовательно, баланс между белками сохраняется.

Одной из важных задач нашей работы был поиск специфичных белковых сигнатур с целью выявления прогностических маркеров ГЦР. Для дальнейшего анализа полученные в ходе протеомного профилирования данные были объединены в единую сводную таблицу. Используя статистические методы анализа, были рассчитаны среднее значение, медиана, стандартное отклонение. Оценка достоверности результатов проводилась с использованием критерия Стьюдента с уровнем значимости $p < 0,05$. Для дальнейшего анализа были

отобраны белки, соответствующие следующим критериям: кратное изменение (Fold change, FC) $\geq 1,5$ и/или присутствие только в злокачественных клетках, а также $p < 0.05$). Таким образом было выявлено 109 ДЭБиз них 43 белка с повышенным уровнем содержания и 66 белков с пониженным уровнем содержания (Рисунок 16).

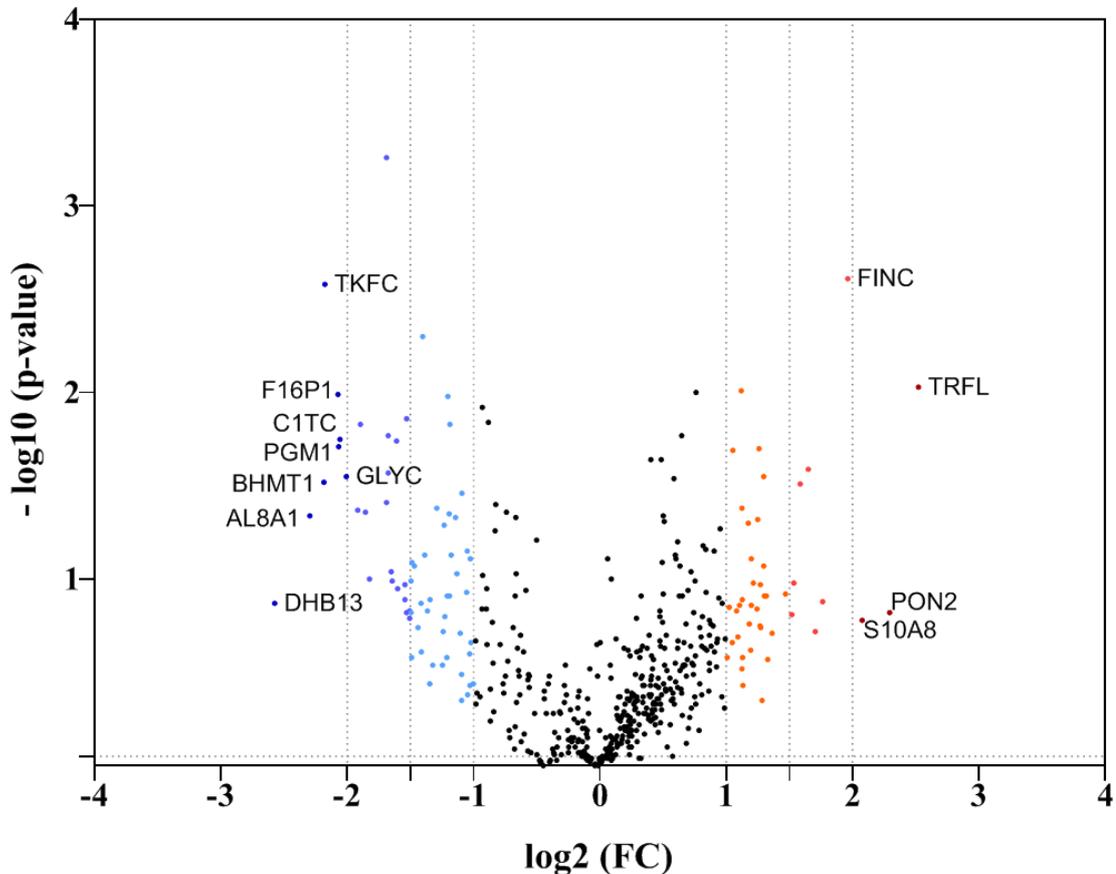


Рисунок 16. Дифференциально экспрессируемые белки в опухолевой и контрольной ткани печени.

Красным выделены белки с повышенным уровнем содержания, синим – с пониженным уровнем содержания.

Для более детального анализа был применен критерий $FC > 2$, и таким образом были отобраны 3 белка с повышенным уровнем содержания: TRFL (лактоферрин), S100A8 (белок S100-A8) и PON2 (параоксоназа/арилэстераза 2), и по литературным данным они связаны с неблагоприятным прогнозом общей выживаемости у пациентов [126–128]. Также было выявлено 8 белков с пониженным уровнем содержания ($FC > 2$) DHB13, AL8A1, BHMT1, TKFC, F16P1, PGM1, C1TC, GLYC и по литературным данным их сниженная экспрессия также

связана с неблагоприятным прогнозом выживаемости у пациентов [129–136]. Дополнительный анализ вышеперечисленных белок-кодирующих генов с использованием веб-ресурса cBioPortal в выборке ГЦР (TCGA, PanCancer Atlas) состоящей из 372 образцов / 372 пациентов, показал, что частота изменений в белок-кодирующих генах составляет 20,43 %: мутации – 2,17 % (8 случаев), амплификация – 14,63% (54 случая), глубокая делеция – 3,25% (12 случаев), множественные изменения – 0,54% (2 случая). Наиболее часто амплифицированным был белок-кодирующий ген S100A8 (38 случаев); при этом частота амплификации не зависела от пола и была выявлена у 20 мужчин и 18 женщин (средний возраст для обоих полов – 70 лет).

3.4.3. Секционное протеомное профилирование

Как и при анализе нормальной ткани печени и клеточной культуры HepG2, для расширения полученных данных панорамного протеомного профилирования опухолевой и контрольной ткани печени от пациентов с ГЦР был применен секционный подход. В качестве примера на Рисунке 17 представлены двумерные электрофореграммы для опухолевой и контрольной ткани печени от первого пациента, дальнейший анализ которых позволил визуализировать данные о протеоформах в виде трехмерных графиков паттернов протеоформ, которые отражают положение протеоформ на геле и относительное содержание белка. Двумерные электрофореграммы для второго и третьего пациента, а также биологического повтора для первого пациента представлены в Приложении 3.

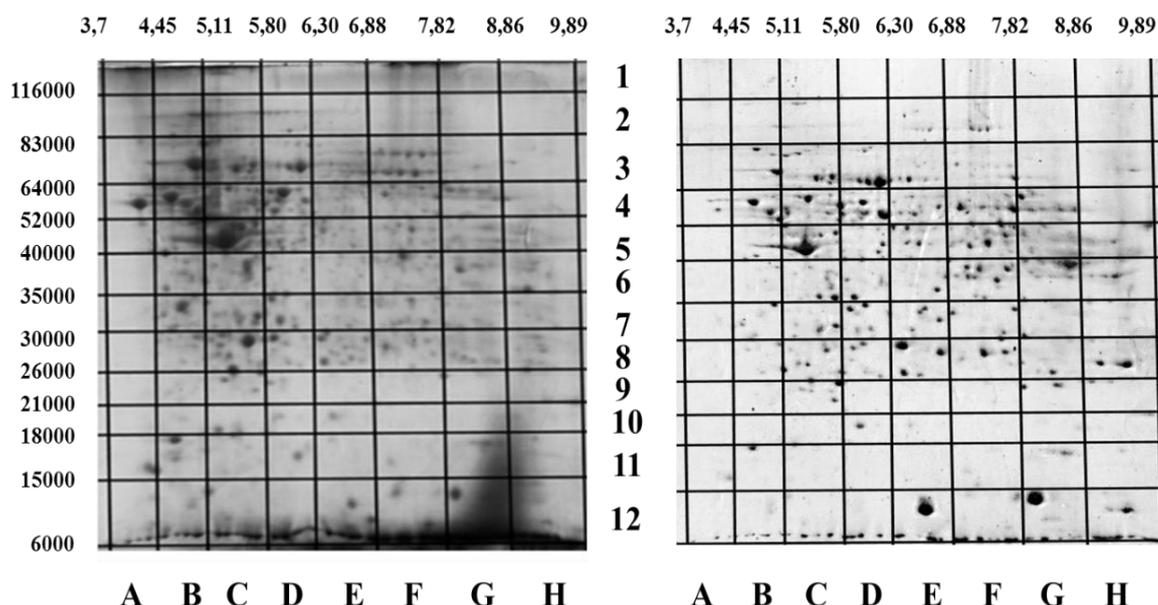


Рисунок 17. Двумерные электрофореграммы проб первого пациента. Слева – опухолевая ткань печени, справа – контрольная ткань печени.

По оси абсцисс – изоэлектрическая точка (pI), по оси ординат – молекулярная масса (кДа).

Таким образом, были получены данные не только о количестве идентифицированных белков, но и количестве протеоформ (Таблица 6).

Таблица 6. Количество идентифицированных белков и протеоформ для опухолевой и контрольной ткани печени.

	Количество белков	Количество протеоформ	Количество протеотипических пептидов	Количество белковых групп
Опухолевая ткань печени	4122	87086	48045	39041
Контрольная ткань печени	3510	57615	33410	24205

В дальнейшем данные, полученные в ходе секционного профилирования, нормировали согласно разделу 2.7. Как и в случае протеомного профилирования, формировали сводную таблицу, содержащую в себе данные обо всех идентифицированных белках и их нормированных величинах emPAI . Отбирали белки, соответствующие критериям отбора ($\text{FC} \geq 1,5$, и/или присутствие только в опухолевой ткани). Суммарно во всех пробах опухолевой и контрольной ткани

печени было идентифицировано 4526 белков. А при сравнении общего числа идентифицированных белков между опухолевой и контрольной тканью печени, общими оказались 3106 белков (Рисунок 18).

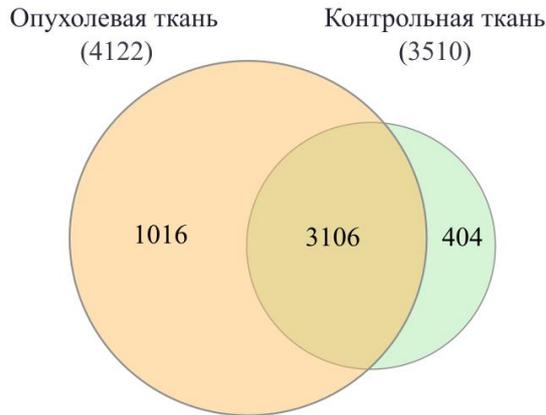


Рисунок 18. Количество белков, идентифицированных в опухолевой и контрольной ткани печени пациентов секционным профилированием.

Как и в случае анализа проб панорамным профилированием, секционное профилирование также демонстрирует поддержание протеостаза между общим числом копий клеток и сохранением баланса между белками с разным числом копий в опухолевой и контрольной ткани печени (Рисунок 19).

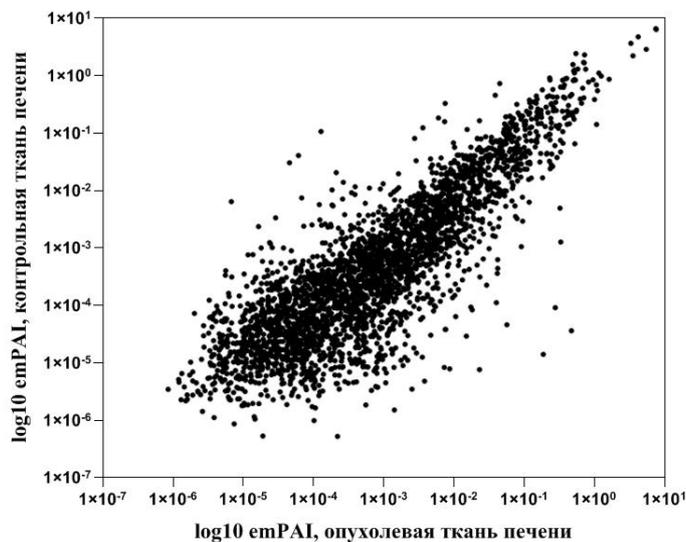


Рисунок 19. График зависимости относительного количественного содержания белка (emPAI) между опухолевой и контрольной тканью печени. По оси абсцисс – десятичный логарифм emPAI для опухолевой ткани печени, по оси ординат – десятичный логарифм emPAI для контрольной ткани печени.

3.2.3. Комбинирование результатов панорамного и секционного протеомного профилирования

Как и в случае сравнения данных, полученных двумя подходами с учетом предварительного фракционирования при помощи 2DE, для клеточной линии HepG2 и нормальной ткани печени, нам удалось идентифицировать значительно больше белков, чем при панорамном профилировании как в случае опухолевой ткани (Рисунок 20), так и в случае контрольной ткани печени (Рисунок 21).

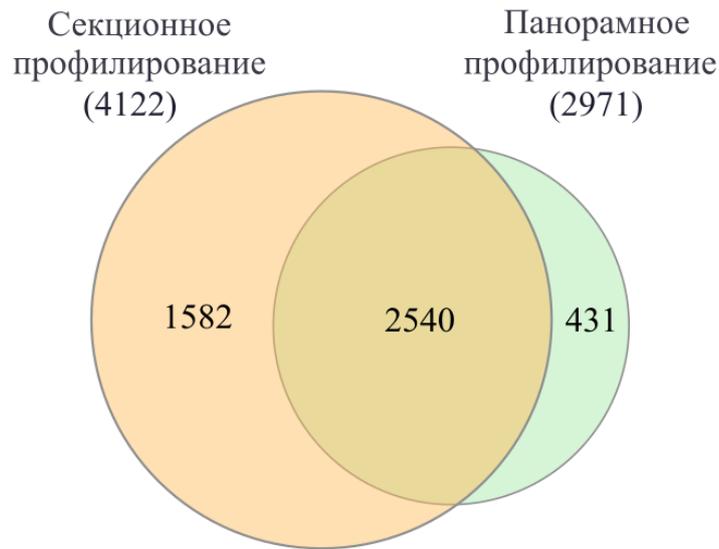


Рисунок 20. Количество белков, идентифицированных в опухолевых тканях печени секционным и панорамным протеомным профилированием.



Рисунок 21. Количество белков, идентифицированных в контрольной ткани печени секционным и панорамным протеомным профилированием.

Для дальнейшего анализа результаты, представленные в виде сводных таблиц для панорамного и секционного профилирования, были объединены в одну таблицу (Приложение 2). В таблице указаны расчетные данные об абсолютном количественном содержании белка в молярной процентной концентрации для определенного БКГ в опухолевой и контрольной ткани печени, полученного двумя подходами, FC и прогностическая ценность общей пятилетней выживаемости по данным базы Protein Atlas (по состоянию на апрель 2024 года).

Из числа идентифицированных белков для более детального анализа были отобраны белки с критерием отсечения $FC \geq 1,5$ как для белков, представленных в опухолевой ткани, так и для белков в контрольной ткани печени. Проанализировав полученные данные, не только о ДЭБ, но и о их протеоформных профилях, мы отобрали 52 белка, для которых изменение протеоформных профилей было наиболее выражено (Таблица 7).

Таблица 7. Список белков с измененными протеоформными профилями в образцах ГЦР.

* – обнаружены только в опухолевых образцах, но не в контроле.

Жирным выделены белки, для которых наблюдается изменение профилей в клетках НерG2.

Номер Uniprot	Белок	Ген	Название	ФС, (секционный профилирование)	ФС, (панорамное профилирование)
P53396	ACLY	<i>ACLY</i>	ATP-citrate synthase	17,24	*
O60488	ACSL4	<i>ACSL4</i>	Long-chain-fatty-acid-CoA ligase 4	375,38	*
P02771	FETA	<i>AFP</i>	Alpha-fetoprotein	*	*
O60218	AK1BA	<i>AKR1B10</i>	Aldo-keto reductase family 1 member B10	232,26	80,78
C9JRZ8	AK1BF	<i>AKR1B15</i>	Aldo-keto reductase family 1 member B15	81,9	*
P05186	PPBT	<i>ALPL</i>	Alkaline phosphatase, tissue-nonspecific isozyme	12,8	*
P12429	ANXA3	<i>ANXA3</i>	Annexin A3	7,5	5,33
P48444	COPD	<i>ARCNI</i>	Coatomer subunit delta	3,2	1,59
P40121	CAPG	<i>CAPG</i>	Macrophage-capping protein	31,22	*
P02741	CRP	<i>CRP</i>	C-reactive protein	35,27	*
P08311	CATG	<i>CTSG</i>	Cathepsin G	14,28	4,43
Q12805	FBLN3	<i>EFEMP1</i>	EGF-containing fibulin-like extracellular matrix protein 1	11,39	17,2
P58107	EPIPL	<i>EPPK1</i>	Epiplakin	29,61	*
P00734	THRB	<i>F2</i>	Prothrombin	2,21	2
Q10471	GALT2	<i>GALNT2</i>	Polypeptide N-acetylgalactosaminyltransferase 2	32,8	5,24
P04062	GLCM	<i>GBA</i>	Lysosomal acid glucosylceramidase	23,83	*
P30511	HLAF	<i>HLA-F</i>	HLA class I histocompatibility antigen, alpha chain F	8,09	*
P11215	ITAM	<i>ITGAM</i>	Integrin alpha-M	*	14,55
P05107	ITB2	<i>ITGB2</i>	Integrin beta-2	14,89	5,02
P80188	NGAL	<i>LCN2</i>	Neutrophil gelatinase-associated lipocalin	13,56	11,49
P17931	LEG3	<i>LGALS3</i>	Galectin-3	2,21	*

Продолжение таблицы 7.

Номер Uniprot	Белок	Ген	Название	FC, (СИ)	FC, (ПИ)
P02788	TRFL	<i>LTF</i>	Lactotransferrin	21,11	7,77
O00187	MASP2	<i>MASP2</i>	Mannan-binding lectin serine protease 2	2,51	*
P23368	MAOM	<i>ME2</i>	NAD-dependent malic enzyme, mitochondrial	5,58	1,88
P41218	MNDA	<i>MNDA</i>	Myeloid cell nuclear differentiation antigen	41,75	*
P05164	PERM	<i>MPO</i>	Myeloperoxidase	19,12	10,26
Q96A35	RM24	<i>MRPL24</i>	39S ribosomal protein L24, mitochondrial	1,77	*
Q96GC5	RM48	<i>MRPL48</i>	39S ribosomal protein L48, mitochondrial	2,16	*
P21589	5NTD	<i>NT5E</i>	5'-nucleotidase	1,59	1,51
Q32P28	P3H1	<i>P3H1</i>	Prolyl 3-hydroxylase 1	2,14	*
P01833	PIGR	<i>PIGR</i>	Polymeric immunoglobulin receptor	104,61	*
P07225	PROS	<i>PROS1</i>	Vitamin K-dependent protein S	9,53	3,7
Q96M27	PRRC1	<i>PRRC1</i>	Protein PRRC1	4,19	2,2
P08575	PTPRC	<i>PTPRC</i>	Receptor-type tyrosine-protein phosphatase C	14,94	8,3
P25815	S100P	<i>S100P</i>	Protein S100-P	13,43	*
Q8NC51	PAIRB	<i>SERBP1</i>	Plasminogen activator inhibitor 1 RNA-binding protein	18,9	*
P05546	HEP2	<i>SERPIND1</i>	Heparin cofactor 2	20,19	*
Q13243	SRSF5	<i>SRSF5</i>	Serine/arginine-rich splicing factor 5	2,92	*
P42224	STAT1	<i>STAT1</i>	Signal transducer and activator of transcription 1-alpha/beta	1,7	1,53
Q9NZ01	TECR	<i>TECR</i>	Very-long-chain enoyl-CoA reductase	1,94	1,8
P02786	TFR1	<i>TFRC</i>	Transferrin receptor protein 1	77,86	5,73
P07996	TSP1	<i>THBS1</i>	Thrombospondin-1	11,11	4,86
P49755	TMEDA	<i>TMED10</i>	Transmembrane emp24 domain-containing protein 10	2,42	1,65

Продолжение таблицы 7.

Номер Uniprot	Белок	Ген	Название	FC, (СП)	FC, (ПП)
Q9BUB7	TMM70	<i>TMEM70</i>	Transmembrane protein 70, mitochondrial	10,09	*
Q6PCB0	VWA1	<i>VWA1</i>	von Willebrand factor A domain-containing protein 1	39,24	*
Q9Y2Z4	SYYM	<i>YARS2</i>	Tyrosine-tRNA ligase, mitochondrial	*	2,02
P61981	1433G	<i>YWHAG</i>	14-3-3 protein gamma	1,59	1,54
O00602	FCN1	<i>FCN1</i>	Ficolin-1	11,97	*

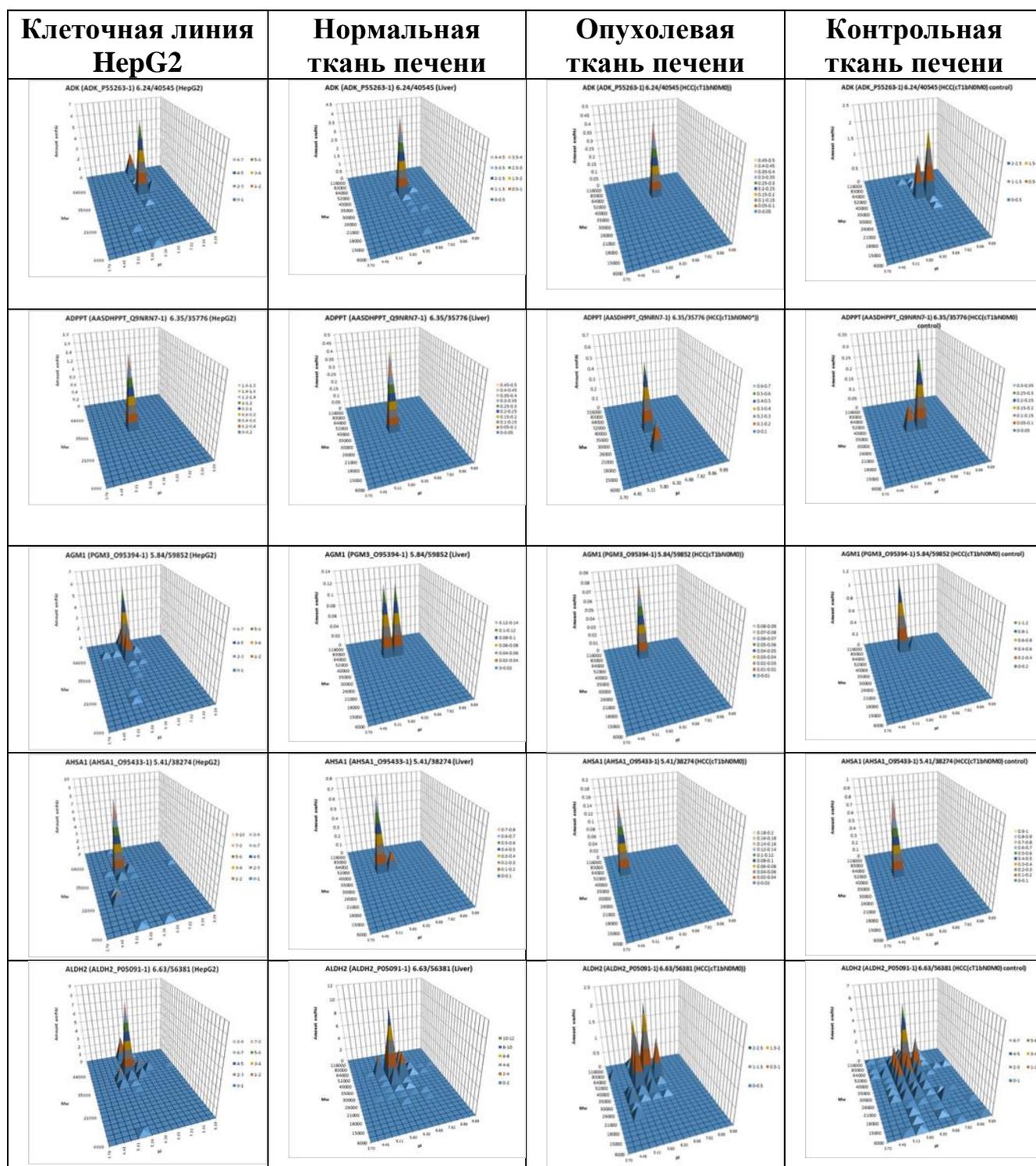
«Идеальный биомаркер» должен обладать высоким потенциалом секретиремости в кровь, что позволяет использовать неинвазивные методы для их выявления и количественного определения. Сигнальная последовательность (сигнальный пептид), нацеленная на секрецию белков, представляет собой короткую гидрофобную N-концевую последовательность, которая встраивается в мембрану эндоплазматического ретикулума и впоследствии отщепляется от белка [137]. Таким образом, секреторный белок можно определить как белок, который активно транспортируется из клетки. Для выявления потенциала секретиремости белков с измененными белковыми паттернами был проведен их поиск в базе данных «Протеома плазмы человека» (Human Plasma Proteome Database, HPPD) [138]. Данная база представляет собой наборы MS-данных о количественной оценке белков плазмы человека. Приблизительно 83% белков (43 / 52) были обнаружены в данной базе (зарегистрированы методом MS), при этом 23 из них имеют сигнальный пептид на N-конце. Это позволяет предположить, что они обладают относительно высоким потенциалом секреции в кровь и могут быть рассмотрены в качестве потенциальных кандидатов в биомаркеры для ГЦР (Приложение 4).

3.5. Сравнительный анализ белков клеточной линии HepG2, нормальной ткани печени и образцов ГЦР.

При сравнении протеоформных паттернов, полученных секционным протеомным профилированием опухолевой ткани печени и клеточной линии HepG2, оказалось, что в подавляющем числе случаев паттерны очень похожи, и мажорный пик соответствует теоретическим параметрам белка. Часть таких профилей представлена в Таблице 8. Более полный набор (1271 белок) представлен в Приложении 5. То есть, сильные изменения протеоформных паттернов являются довольно редкими случаями и в основном разница между паттернами связана с разным уровнем белка в анализируемом объекте.

Для белков, представляющих интерес в качестве потенциальных биомаркеров ГЦР, представленных в Таблице 7, только 25 паттернов, представленных в клеточной линии HepG2, согласуются с профилями ГЦР ($FC > 1.5$) по сравнению с контрольными образцами. Четыре белка не соответствуют критерию $FC > 1.5$, а 10 белков присутствуют только в нормальной ткани печени, остальные 13 белков не были обнаружены ни в клетках HepG2 ни в нормальной ткани печени. При этом протеоформные профили в нормальной ткани печени схожи с белковыми паттернами в контрольной ткани печени (Таблица 8).

Таблица 8. Примеры типичных белковых паттернов, полученных для всех типов анализируемых образцов (клеточная линия HepG2, нормальная ткань печени, опухолевая ткань печени, контрольная ткань печени).



Если рассматривать только белки, для которых обнаружено увеличение относительного количества содержания в анализируемом объекте или в контроле, то для опухолевой ткани печени и клеточной HepG2 общими были 700 белков, для контрольной и нормальной ткани печени 158 белков (Рисунок 22).

А



Б

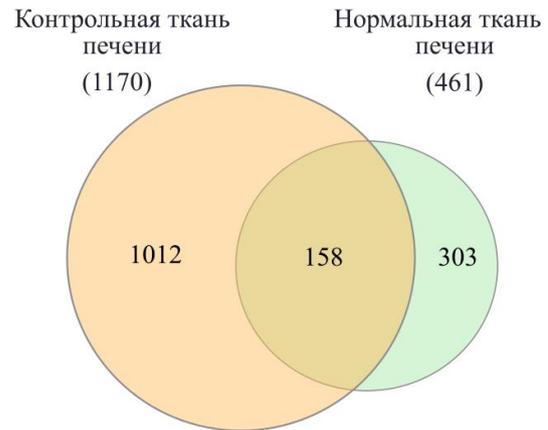


Рисунок 22. Количество белков с повышенным содержанием ($FC > 1.5$) в опытном или контрольном образце, полученное секционным и панорамным протеомным профилированием. **А:** опухолевая ткань печени и клеточная линия HepG2. **Б:** контрольная ткань печени и нормальная ткань печени.

Разница в количестве идентифицированных белков между контрольной и нормальной тканью печени, вероятно, связана с тем, что образцы, полученные от пациентов с ГЦР, были заражены вирусами гепатита В и С. Известно, что гепатит может вызывать различные изменения в печени; при этом обычно наблюдается портальное распространение воспаления с переходом на прилежащую паренхиму печени (фрагментарный некроз) [139]. Также два пациента получали предшествующую терапию, которая может оказывать влияние на белковый состав ткани печени. При лекарственном поражении печени вовлекается больше паренхима, чем портальные области и характеризуется апоптозом гепатоцитов, а также наличием многочисленных очаговых скоплений лимфоцитов и макрофагов [140].

Также были обнаружены белки, для которых FC не согласовывался между панорамным и секционным профилированием и, следовательно, они не были включены в выборку, представленную выше. При анализе этих данных оказалось, что только 2 белка из данного списка (204 и 101, соответственно) были обнаружены как в опухолевой и контрольной ткани печени, так и в клетках HepG2 и нормальной ткани печени, а именно, SRBS1 (Sorbin and SH3 domain-

containing protein 1) и HNRPK (Heterogeneous nuclear ribonucleoprotein K) (Рисунок 23).

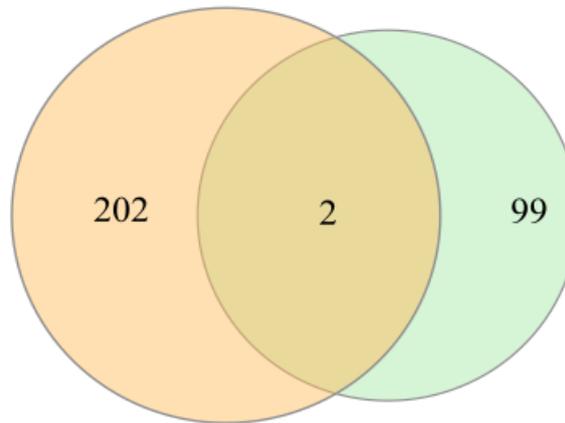


Рисунок 23. Количество белков с повышенным содержанием ($FC > 1.5$) в образцах ГЦР (опухолевая и контрольная ткань печени) (слева) и образцах HepG2 и нормальной ткани печени (справа).

Согласно литературным данным, SRBS1 является потенциальным маркером рака молочной железы и легких [141,142]. Было показано, что он имеет достаточно низкие уровни экспрессии в раковых клетках, и в значительной степени связан с низким уровнем общей выживаемости пациентов [143]. В полученных нами результатах, при секционном профилировании образцов, ГЦР наблюдается увеличение экспрессии данного белка, а при панорамном профилировании, наоборот увеличение экспрессии в контрольной ткани печени. Такая же ситуация наблюдается и при сравнении данных секционного и панорамного профилирования клеточной линии HepG2 и нормальной ткани печени. В случае белка HNRPK в работе Guo и соавторов с использованием 2DE и МАЛДИ было показано, что HNRPK высоко экспрессируется на ранних стадиях ГЦР по сравнению с циррозом печени [144]. Полученные нами результаты секционного профилирования согласуются с данными результатами; при этом в случае панорамного профилирования наблюдаются противоположные результаты: повышение в контроле, но не в злокачественных клетках.

3.5.2. ПТМ потенциальных биомаркеров ГЦР

Среди основных источников гетерогенности протеома особое внимание уделяется ПТМ ввиду их важной роли в регуляции различных клеточных процессов, обеспечении адаптивных механизмов, необходимых для клеточных функций и ответа на изменения окружающей среды. Идентификация и анализ ПТМ требует применения разнообразных методов и технологий, которые смогут точно обнаруживать и охарактеризовать их. Существует ряд биоинформатических программ, используемых для предсказания ПТМ, таких как NetPhos, ModPred, SignalP и др., также активно используется машинное обучение и структурное моделирование. Тем не менее, существуют и ряд ограничений, таких как сложность и разнообразие ПТМ, комплексные взаимодействия между ПТМ, специфичность исследуемого образца, точность предсказательных моделей и ограничение машинных алгоритмов поиска.

Вследствие того, что в нашем эксперименте процесс поиска возможных ПТМ ограничен машинными алгоритмами, существенным преимуществом является то, что мы можем предсказать возможные ПТМ исходя из физико-химических параметров протеоформы на двумерной электрофореграмме. Смещения положения протеоформ могут происходить как по первому направлению (изоэлектрическая точка), так и по второму направлению (молекулярная масса). На смещение протеоформ с одинаковой молекулярной массой в более кислую область рI, влияют ПТМ, добавляющие или удаляющие положительные заряды (фосфорилирование, ацетилирование, метилирование и т. Д.). Смещение в более щелочную область рI связано с модификациями карбоксильных групп (амидирование, этерификация). При этом смещение по двум параметрам, чаще всего вызваны убиквитинированием, сумоилированием или гликозилированием.

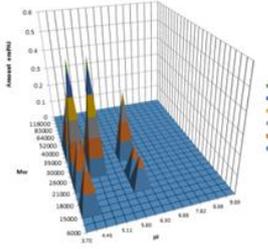
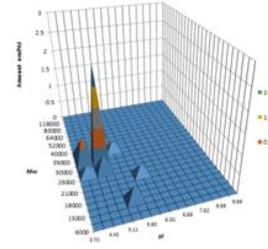
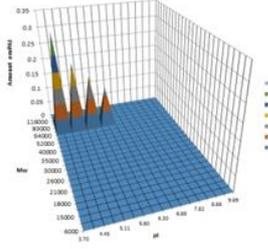
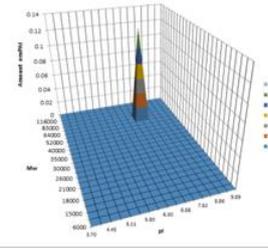
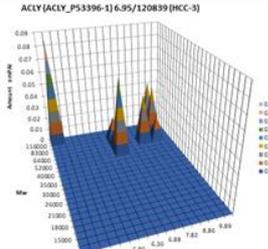
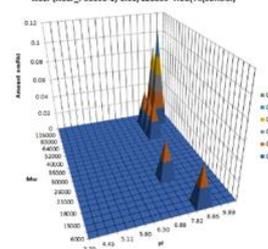
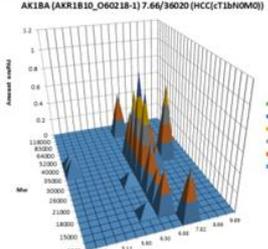
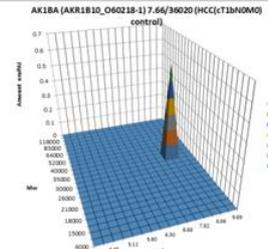
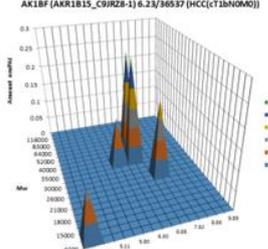
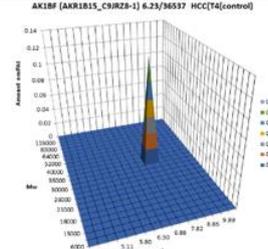
Поисковый запрос для анализа ПТМ в нашем эксперименте включал в себя наиболее часто представленные модификации: окисление метионина, фосфорилирование серина, треонина, триптофана, ацетилирование лизина,

дезамидирование аспарагина. Соответственно, данные модификации были обнаружены во всех типах исследуемых нами образцов. При этом окисление метионина и деамидирование аспарагина больше относятся к химической деградации белка и при комбинированном протеомном анализе могут быть причиной появления новых пиков (протеоформ). Они могут быть как артефактами в процессе подготовки проб, так и модификациями, влияющими на апоптоз, пролиферацию клеток или же возникать при воспалительных процессах [145–148]. Другой возможный источник артефактов – протеолиз, проявляющийся в виде большого количества фрагментов меньшего размера, чем исходная форма белка, что может быть визуализировано как по виду 2DE-геля, так и после МС-анализа [123]. В наших условиях такой протеолиз не наблюдался, так как для его предотвращения была использована максимально щадящая подготовка проб. Все этапы по подготовке проб проводили на льду, с добавлением ингибиторов протеаз – тиомочевина, протеазный коктейль. Кроме того, следует отметить, что протеолиз происходит в клетках и *in vivo*, и детектирование белковых фрагментов при анализе не обязательно является артефактом.

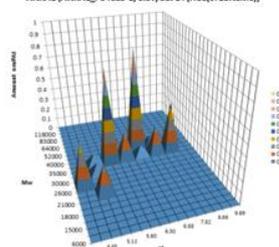
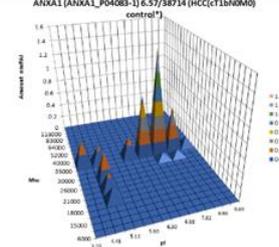
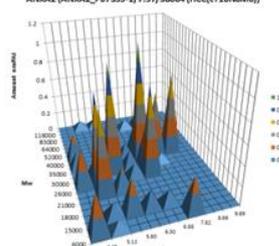
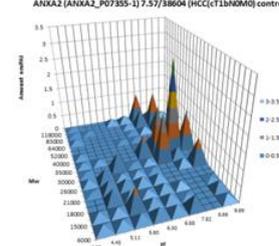
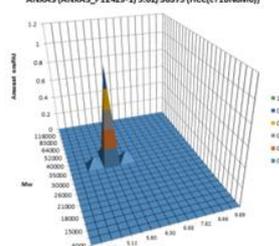
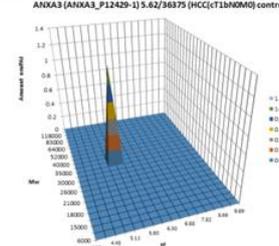
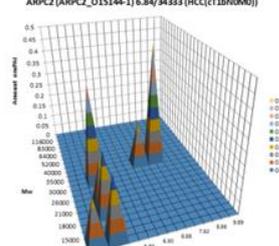
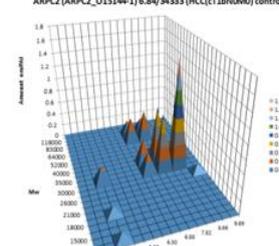
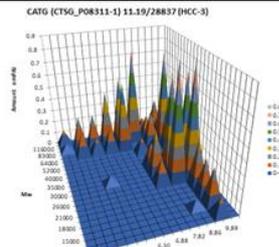
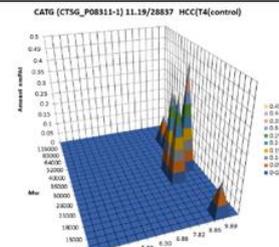
Основываясь на данных секционного протеомного профилирования для образцов опухолевой и контрольной ткани печени, особое внимание мы уделили 52 белкам из числа отобранных по протеоформным паттернам (Таблица 9). Также данные протеоформные паттерны дополнительно проанализированы в базе данных Proteoform Atlas для выявления возможных ПТМ. Важно отметить, что ПТМ, приведенные в Proteoform Atlas, идентифицированы для отдельных протеоформ и подразделяются на 3 категории достоверности, от 1 – автоматической валидации данных на комбинации пороговых значений на основании FDR, и до 3 – вручную валидированными данными СТДР. Для приведенных в Таблице 9 белков указанные ПТМ относятся к 1 категории достоверности. Для отобранных нами паттернов только для 10 протеоформ представлены модификации. Модификации для данных протеоформ были обнаружены исключительно в раковых клеточных линиях HeLa S3 и H1299.

Таблица 9. Белки с измененным белковым паттерном в опухолевой ткани печени.

* – приведены однобуквенные обозначения аминокислотных остатков, подвергающихся ПТМ.
– ПТМ не обнаружены.

Опухолевая ткань печени	Контрольная ткань печени	ПТМ (Proteoform Atlas)	ПТМ (экспериментальные данные)*
<p>1433G (YVNHAG_P01981-1) 4.8/28303 (HCC(T1bNOM))</p> 	<p>1433B (YVNHAB_P21946-1) 4.76/28082 (HCC(T1bNOM) control)</p> 	<p>N-acetylmethionine; Phosphotyrosine</p>	<p>Acetylation of K; Deamidation of N; Methylation of K;</p>
<p>5NTD (NTSE_P21589-1) 6.58/63368 (HCC(T1bNOM))</p> 	<p>5NTD (NTSE_P21589-1) 6.58/63368 (HCC(T1bNOM) control)</p> 	<p>–</p>	<p>–</p>
<p>ACY (ACY_P53396-1) 6.95/120839 (HCC-3)</p> 	<p>ACY (ACY_P53396-1) 6.95/120839 HCC(T4) control)</p> 	<p>–</p>	<p>Deamidation of N</p>
<p>AK18A (AKR1810_O60218-1) 7.66/36020 (HCC(T1bNOM))</p> 	<p>AK18A (AKR1810_O60218-1) 7.66/36020 (HCC(T1bNOM) control)</p> 	<p>–</p>	<p>Acetylation of K</p>
<p>AK18F (AKR1815_C9R28-1) 6.23/36337 (HCC(T1bNOM))</p> 	<p>AK18F (AKR1815_C9R28-1) 6.23/36337 HCC(T4) control)</p> 	<p>–</p>	<p>Deamidation of N; Phosphorylation of S</p>

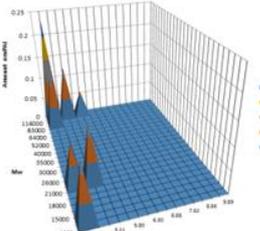
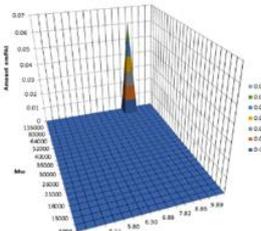
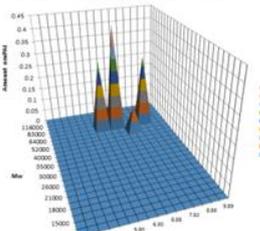
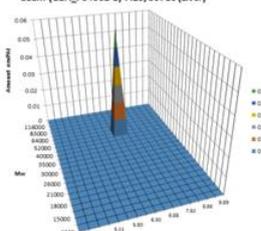
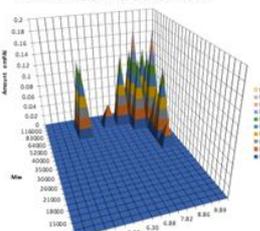
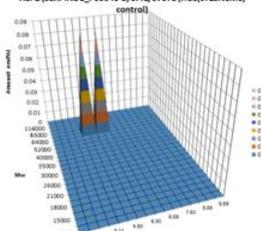
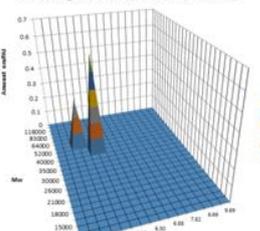
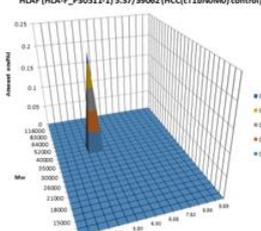
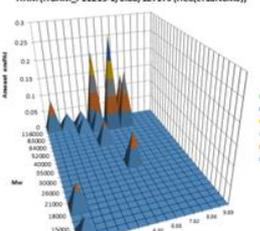
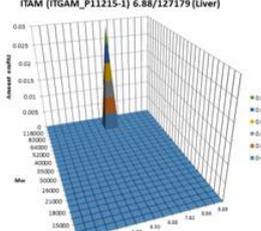
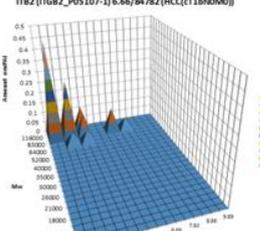
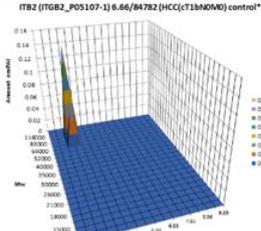
Продолжение таблицы 9.

Опухолевая ткань печени	Контрольная ткань печени	ПТМ (Proteoform Atlas)	ПТМ (экспериментальные данные)
<p>ANXA1 [ANXA1_P04083-1] 6.57/38714 (HCC(T1bN0M0))</p> 	<p>ANXA1 [ANXA1_P04083-1] 6.57/38714 (HCC(T1bN0M0) control*)</p> 	-	<p>Acetylation of K; Deamidation of N; Phosphorylation of S; Phosphorylation of T</p>
<p>ANXA2 [ANXA2_P07355-1] 7.57/38604 (HCC(T1bN0M0))</p> 	<p>ANXA2 [ANXA2_P07355-1] 7.57/38604 (HCC(T1bN0M0) control)</p> 	<p>N6-acetyllysine; N-acetylserine; Phosphotyrosine</p>	<p>Acetylation of K; Deamidation of N</p>
<p>ANXA3 [ANXA3_P12429-1] 5.62/36375 (HCC(T1bN0M0))</p> 	<p>ANXA3 [ANXA3_P12429-1] 5.62/36375 (HCC(T1bN0M0) control)</p> 	-	-
<p>ARPC2 [ARPC2_O15144-1] 6.84/34333 (HCC(T1bN0M0))</p> 	<p>ARPC2 [ARPC2_O15144-1] 6.84/34333 (HCC(T1bN0M0) control)</p> 	-	-
<p>CATG [CTSG_P08311-1] 11.19/28837 (HCC-3)</p> 	<p>CATG [CTSG_P08311-1] 11.19/28837 HCC(T4(control))</p> 	-	<p>Deamidation of N</p>

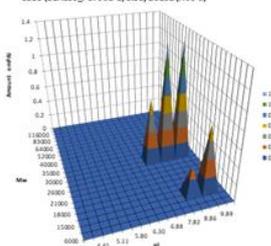
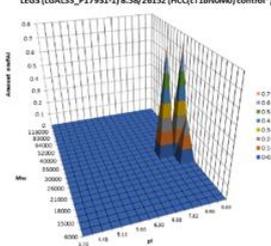
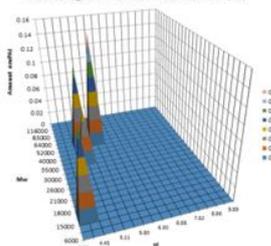
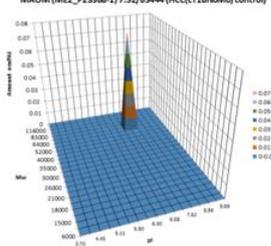
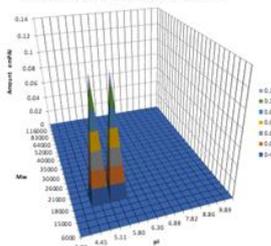
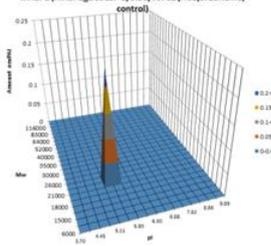
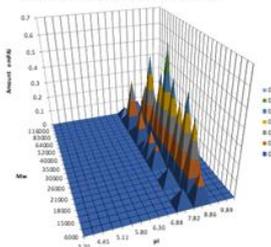
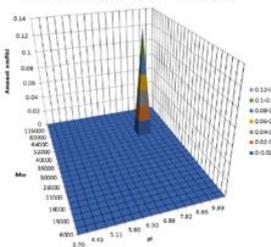
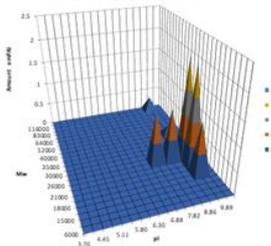
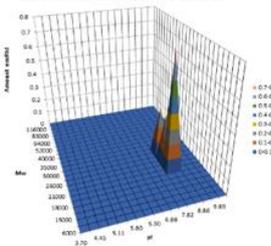
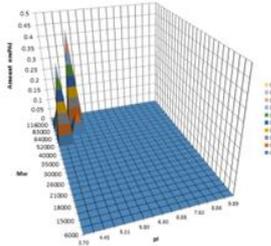
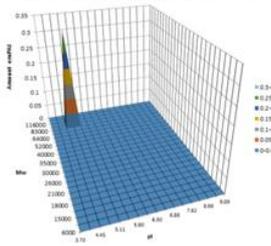
Продолжение таблицы 9.

Опухолевая ткань печени	Контрольная ткань печени	ПТМ (Proteoform Atlas)	ПТМ (экспериментальные данные)
		<p>—</p>	<p>Acetylation of K; Deamidation of N</p>
		<p>—</p>	<p>—</p>

Продолжение таблицы 9.

Опухолевая ткань печени	Контрольная ткань печени	ПТМ (Proteoform Atlas)	ПТМ (экспериментальные данные)
<p>GALT2 (GALNT2_Q10471-1) 8.63/64733 (HCC(Т1bN0M0))</p> 	<p>GALT2 (GALNT2_Q10471-1) 8.63/64733 HCC(T4(control))</p> 	—	—
<p>GLCM (GBA_P04062-1) 7.29/59716 (HCC(Т1bN0M0))</p> 	<p>GLCM (GBA_P04062-1) 7.29/59716 (Liver)</p> 	—	—
<p>HEP2 (SERPIND1_P05546-1) 6.41/57071 (HCC-3)</p> 	<p>HEP2 (SERPIND1_P05546-1) 6.41/57071 (HCC(Т1bN0M0) control)</p> 	—	—
<p>HLAF (HLA-F_P30511-1) 5.37/39062 (HCC(Т1bN0M0))</p> 	<p>HLAF (HLA-F_P30511-1) 5.37/39062 (HCC(Т1bN0M0) control)</p> 	—	Deamidation of N
<p>ITAM (ITGAM_P11215-1) 6.88/127179 (HCC(Т1bN0M0))</p> 	<p>ITAM (ITGAM_P11215-1) 6.88/127179 (Liver)</p> 	—	Phosphorylation of S; Phosphorylation of T
<p>ITB2 (ITGB2_P05107-1) 6.66/84782 (HCC(Т1bN0M0))</p> 	<p>ITB2 (ITGB2_P05107-1) 6.66/84782 (HCC(Т1bN0M0) control*)</p> 	—	—

Продолжение таблицы 9.

Опухолевая ткань печени	Контрольная ткань печени	ПТМ (Proteoform Atlas)	ПТМ (экспериментальные данные)
 <p>LEG3 (LGALS3_P17931-1) 8.58/26152 (HCC-3)</p>	 <p>LEG3 (LGALS3_P17931-1) 8.58/26152 (HCC(T1bN0M0) control*)</p>	-	Deamidation of N
 <p>MAOM (ME2_P23368-1) 7.52/65444 (HCC(T1bN0M0))</p>	 <p>MAOM (ME2_P23368-1) 7.52/65444 (HCC(T1bN0M0) control)</p>	-	-
 <p>MASP2 (MASP2_O00187-1) 5.39/75702 (HCC(T4))</p>	 <p>MASP2 (MASP2_O00187-1) 5.39/75702 (HCC(T1bN0M0) control)</p>	-	-
 <p>MNDA (MNDA_P41218-1) 9.76/45836 (HCC(T4))</p>	 <p>MNDA (MNDA_P41218-1) 9.76/45836 (HCC(T4) control)</p>	-	Deamidation of N
 <p>NGAL (LCN2_P0188-1) 9.02/22588 (HCC-3)</p>	 <p>NGAL (LCN2_P0188-1) 9.02/22588 (HCC(T4) control)</p>	-	Deamidation of N
 <p>P3H1 (P3H1_Q32P28-1) 5.05/83394 (HCC(T1bN0M0))</p>	 <p>P3H1 (P3H1_Q32P28-1) 5.05/83394 (HCC(T1bN0M0) control)</p>	-	-

Продолжение таблицы 9.

Опухолевая ткань печени	Контрольная ткань печени	ПТМ (Proteoform Atlas)	ПТМ (экспериментальные данные)
		-	Deamidation of N
		-	Deamidation of N; Methylation of K
		-	-
		-	-
		-	-
		-	-

Продолжение таблицы 9.

Опухолевая ткань печени	Контрольная ткань печени	ПТМ (Proteoform Atlas)	ПТМ (экспериментальные данные)
		N-acetylserine	—
		N-acetylthreonine	Methylation of K
		—	—
		—	—
		—	—
		—	—

Продолжение таблицы 9.

Опухолевая ткань печени	Контрольная ткань печени	ПТМ (Proteoform Atlas)	ПТМ (экспериментальные данные)
<p>THRB [F2_P00734-1] 5.63/70037 (HCC(T1bNOMO))</p>	<p>THRB [F2_P00734-1] 5.63/70037 (HCC(T1bNOMO) control)</p>	<p>—</p>	<p>Acetylation of K; Methylation of K; Phosphorylation of S</p>
<p>TMEDA (TMED10_P49755-1) 6.97/24976 (HCC(T4))</p>	<p>TMEDA (TMED10_P49755-1) 6.97/24976 HCC(T4) control)</p>	<p>Dimethylated arginine</p>	<p>Methylation of K</p>
<p>TMM70 (TMM70_Q08U87-1) 9.02/28969 (HCC(T4))</p>	<p>TMM70 (TMM70_Q08U87-1) 9.02/28969 (HCC(T1bNOMO) control)</p>	<p>N-acetylmethionine</p>	<p>—</p>
<p>TRFL (LTF_P02788-1) 8.5/78182 (HCC(T4))</p>	<p>TRFL (LTF_P02788-1) 8.5/78182 (HCC(T1bNOMO) control)</p>	<p>—</p>	<p>Acetylation of K; Deamidation of N</p>
<p>TSP1 (THBS1_P07996-1) 4.71/129383 (HCC(T1bNOMO))</p>	<p>TSP1 (THBS1_P07996-1) 4.71/129383 (HCC(T1bNOMO) control)</p>	<p>—</p>	<p>—</p>
<p>VWA1 (VWA1_Q6PC80-1) 7.18/46804 (HCC(T4))</p>	<p>VWA1 (VWA1_Q6PC80-1) 7.18/46804 (HCC(T1bNOMO) control)</p>	<p>—</p>	<p>—</p>

Согласно литературным данным, наиболее часто встречающимися ПТМ при ГЦР являются гликозилирование, фосфорилирование и убиквитинирование [149]. Показано, что использование некоторых препаратов (например, соранифеба) при лечении ГЦР может вызывать изменения паттернов гликозилирования, а, следовательно, изучаются возможности иммунотерапии, направленной на определенные ПТМ [150].

Как и в случае HepG2 и нормальной ткани печени, для ГЦР в экспериментально полученных нами данных, наиболее часто встречаемыми ПТМ являются окисление метионина (не представлено в Таблице 9) и деамидирование аспарагина. Но также присутствует ацетилирование лизина, фосфорилирование серина и треонина, и метилирование лизина (причем последнее обнаружено только в опухолевой ткани печени, но не в контрольной ткани печени). При рассмотрении протеоформных паттернов, полученных секционным протеомным профилированием, в опухолевой ткани печени для белка КРУМ (pyruvate kinase PKM), также обнаружено больше протеоформ по сравнению с контрольной тканью печени. При этом нам удалось обнаружить изоформу M1 белка КРУМ (P14618-2), представленную 5 пептидами, при этом 2 пептида (EAEEAAMFHR и KLFEEELVR) принадлежат только данной изоформе (Рисунок 24). По литературным данным, именно эта изоформа белка способствует более высокому росту опухолевых клеток [151]. Также она обладает высокой пируваткиназной активностью и не требует аллостерической активации D-фруктозо-1,6-бисфосфатом для своей активности в отличие от изоформы M2 [152].

гены потенциальных маркеров, представленные в Таблице 6. Для этого была использована система классификации PANTHER [100] (Рисунок 25). Она представляет собой аннотированную библиотеку, содержащую идентификаторы филогенетических связей между семействами генов и их белковых продуктов, таким образом, позволяя охарактеризовать белки с неизвестными функциями.

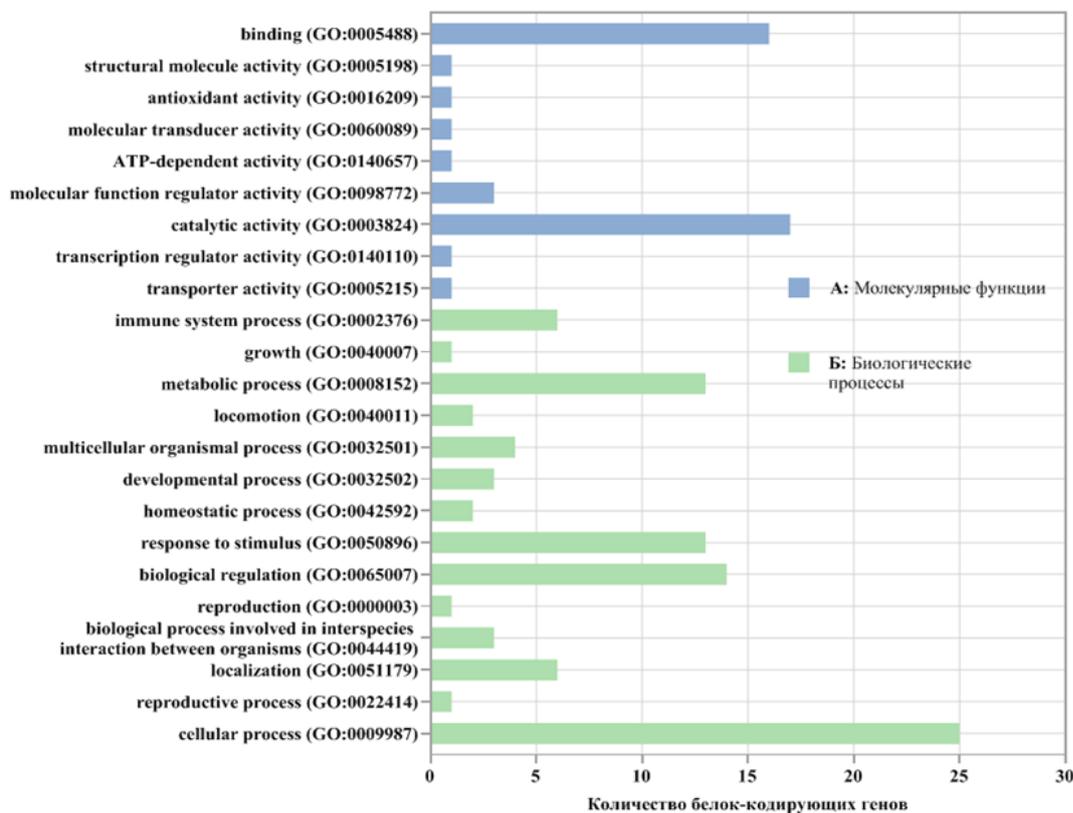


Рисунок 25. Классификация белок-кодирующих генов из списка потенциальных биомаркеров по молекулярным функциям и биологическим процессам.

Молекулярные функции представлены связыванием белков и различных соединений (ионов кальция, ионов металлов, гликопротеинов и др.), каталитической активностью эндопептидазы серинового типа и аллилспиртдегидрогеназы. При полном анализе молекулярных функций всех идентифицированных нами 4983 белок-кодирующих генов в образцах опухоли и неопухолевого контроля ГЦР наблюдается снижение количества белок-кодирующих генов, связанных с каталитической активностью. Известно, что уровень активности каталитических ферментов может заметно снижаться при

возникновении злокачественных трансформаций, что связано с изменениями в метаболических путях и дисфункцией самих ферментов [153].

Основными категориями, выявленными в биологических процессах, являются клеточные процессы (GO:0009987), биологическая и метаболическая регуляции (GO:0065007 и GO:0008152, соответственно) и ответ на стимул (GO:0050896). Наиболее представленными биологическими процессами являлась негативная регуляция апоптоза, биосинтез липидов, гомеостаз ионов кальция, ответ на механическую стимуляцию и фагоцитоз. Негативная регуляция апоптоза способствует развитию и прогрессированию рака, а также устойчивости опухолевых клеток к различным методам лечения [154].

При рассмотрении белковых классов для 52 белков, потенциальных маркеров, оказалось, что большинство из них является ферментами взаимодействия метаболитов (PC00262), представленных оксидоредуктазами, трансферазами, гидролазами и лигазами (Рисунок 26).

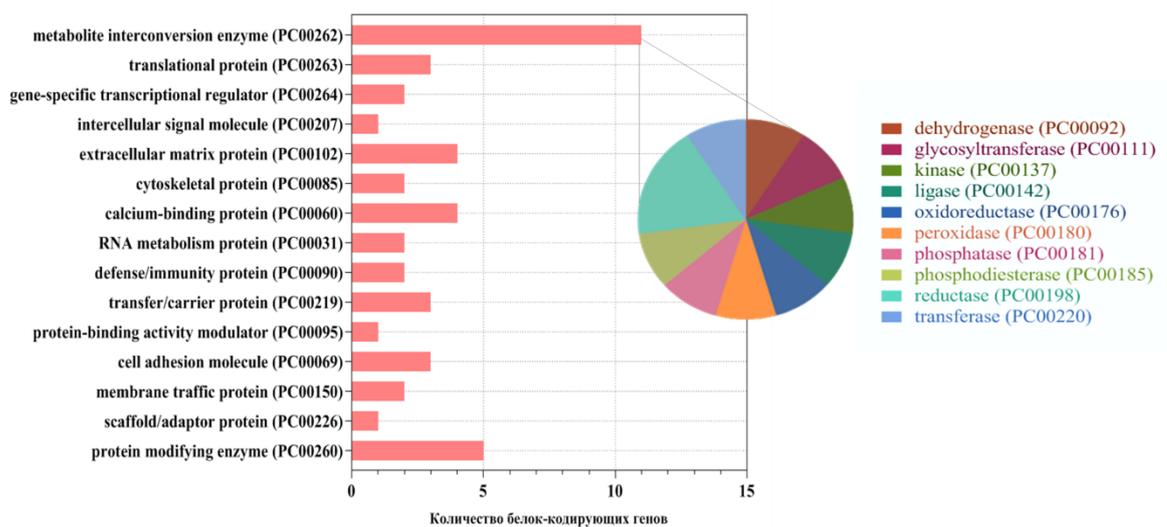


Рисунок 26. Классификация белок-кодирующих генов из списка потенциальных биомаркеров по белковым классам.

Известно, что метаболизм раковых клеток существенно отличается от метаболизма нормальных клеток, эти изменения позволяют раковым клеткам

адаптироваться к микросреде, что приводит к их быстрому росту и делению. В ряде исследований, показано, что белки, связанные с основными метаболическими путями, могут быть использованы в качестве потенциальных терапевтических мишеней [155–157].

ЗАКЛЮЧЕНИЕ

В ходе данной работы была показана высокая эффективность секционного протеомного профилирования образцов различного биологического происхождения, которая значительно превосходит панорамное протеомное профилирование без предварительного разделения 2DE. Данные, полученные панорамным протеомным профилированием, могут быть расширены и дополнены секционным протеомным профилированием. Комбинация полученных методов при исследовании клеточной линии HepG2 и нормальной ткани печени позволила составить «эталонные» протеомные профили; для клеточной линии HepG2 они включают 3419 белков, для нормальной ткани печени – 2570 белков. Использование секционного протеомного профилирования позволило не только составить расширенный протеомный профиль исследуемого образца, но и выявить протеоформные паттерны для каждого отдельного белка, присутствующего в образцах (для HepG2 – 3390 белков и 18992 протеоформы, для нормальной ткани печени – 2482 белка и 20504 протеоформы), что позволило перейти от модели клеточной линии непосредственно к образцам ГЦР. Таким образом, для всех пациентов были получены данные о протеомном профиле (опухолевая ткань печени – 4560 белков, контрольная ткань печени – 3865 белков) и протеоформные паттерны (для опухолевой ткани печени – 4122 белков, 87086 протеоформы и для контрольной ткани печени – 3510 белков и 57615 протеоформ).

При сравнительном анализе белков с повышенным и пониженным уровнем содержания между различными типами тканей и клеточной линией HepG2 были выявлены белки, которые присутствуют либо в контрольных образцах (нормальная и контрольная ткань печени), либо в образцах клеточной линии HepG2 и опухолевой ткани печени. Вероятные отличия в количестве идентифицированных белков могут быть связаны как с недостаточным содержанием белка для его детекции методом MS, так и с тем, что в HepG2 присутствуют белки характерные только для данной клеточной линии [102].

Также необходимо учитывать тот факт, что различия в количестве идентифицированных белков могут быть связаны с различным возрастом доноров (средний возраст – 56 лет), а клеточная линия HepG2 получена из ткани опухоли 15-летнего юноши. Возрастные изменения в печени включают в себя увеличение размера и плоидность гепатоцитов, а также изменения метаболизма [158].

Использование секционного протеомного профилирования позволило выделить наиболее измененные паттерны между опухолевой и контрольной тканью печени, а именно 52 белка, протеоформные паттерны которых могут являться прогностическими маркерами, представляя собой белковую сигнатуру ГЦР. Сравнительный анализ данных белков с белками клеточной линии HepG2 показал, что из 52 белков в клетках HepG2 наблюдается 25 белков с $FC > 1.5$, для 4 белков содержание белка практически не изменяется, а 10 белков имеют схожие белковые паттерны в нормальной ткани печени и в клетках HepG2. При этом общие белковые паттерны как в случае HepG2, так и ГЦР похожи. А вероятные расхождения между ГЦР и HepG2, по-видимому, связаны с небольшим объемом выборки пациентов (стадии, возраст и пол) и требуют расширения выборки. Тем не менее, полученные наборы данных могут в дальнейшем быть использованы для протеотипирования образцов в норме и при ГЦР, что позволяет сократить время и трудозатраты на разработку панелей биомаркеров, которые будут включать в себя не только одну конкретную протеоформу, но и комбинацию протеоформ. Также сравнение теоретических и экспериментальных координат протеоформ, полученных при секционном протеомном профилировании, может быть эффективно использовано в качестве дополнительного источника информации о ПТМ для каждого отдельного белка. Проведенный анализ ПТМ для 52 белков, которые могут являться потенциальными кандидатами в маркеры ГЦР показал, что наиболее представленными ПТМ являются ацетилирование, фосфорилирование и дезамидирование, поэтому в дальнейшем планируется не только расширение полученных данных о протеоформах, но и их углубленный анализ ПТМ. Для подтверждения характерных изменений полученных

протеоформных паттернов, потенциальных маркеров ГЦР, планируется расширить выборку пациентов с подтвержденным диагнозом ГЦР, а также проверить полученные результаты на других биологических объектах, например, клеточные линии, выделенные не из гепатоцитов или опухоли тканей, максимально удаленных от печени.

ВЫВОДЫ

1. Секционное протеомное профилирование позволяет обнаружить большее количество белков по сравнению с панорамным протеомным профилированием. Для клеточной линии HepG2 секционным и панорамным протеомными профилированием получено: 3391 / 867 белков, для нормальной ткани печени: 2482 / 797 белков, для опухолевой ткани печени: 4122 / 1306 белков, и для контрольной ткани печени: 3510 / 982 белков, соответственно.

2. На основании данных секционного протеомного профилирования для каждого из исследуемых типов образцов получены и визуализированы более 2500 протеоформных паттернов, что позволило определить специфические белковые сигнатуры и общие паттерны для злокачественных клеток печени человека. Показано, что в подавляющем числе случаев протеоформные паттерны очень похожи, и мажорный пик соответствует теоретическим параметрам белка, то есть не модифицирован. Отобран 1271 белковый паттерн, воспроизводимо визуализированный для всех 4 типов образцов.

3. По результатам панорамного протеомного профилирования опухолевой и контрольной ткани печени выявлены 109 дифференциально экспрессируемых белков, из них 43 белка с повышенным уровнем содержания и 66 белков с пониженным уровнем содержания. Из них 11 белков могут рассматриваться как потенциальные биомаркеры ГЦР.

4. Комбинация секционного и панорамного протеомного профилирования позволила выделить 52 белка со специфичными протеоформными паттернами в опухолевой ткани печени.

5. Сопоставление специфичных 52 протеоформных паттернов в опухолевой ткани печени с данными анализа клеток линии HepG2 и нормальной ткани печени показало, что в клетках линии HepG2 25 протеоформных паттернов отличаются от нормальной ткани печени схожим образом, что позволяет рассматривать HepG2 в качестве модельной клеточной линии при исследовании ГЦР.

ФИНАНСИРОВАНИЕ

Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 годы) (№ 122030100168-2).

БЛАГОДАРНОСТИ

Автор выражает искреннюю благодарность научному руководителю, д.б.н. Нарыжному Станиславу Николаевичу, за возможность выполнения исследования и помощь на всех этапах работы. Автор благодарит коллег из лаборатории системной биологии ИБМХ Згуду В.Г., Копылова А.Т. и Завьялову М.Г. за проведение масс-спектрометрического анализа, а также сотрудника лаборатории клеточной биологии ИБМХ Вахрушева И.В. за предоставление клеточной линии HepG2. Коллег из лаборатории протеомики ПИЯФ (Гатчина) Ронжину Н.Л., Клейст О.К, Белякову Н.В., Легину О.К., Клопова Н.В. за помощь в визуализации трехмерных графиков протеоформ. Багмета Н.Н. из ГНЦ РФ ФГБНУ «Российский научный центр хирургии имени академика Б. В. Петровского» за предоставление образцов тканей, полученных от пациентов с диагностированным гепатоцеллюлярным раком. Автор выражает искреннюю признательность Пономаренко Е.А., Медведеву А.Е. и Скворцову В.С. за плодотворное обсуждение и комментарии по работе.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

2DE	Two-dimensional electrophoresis (двумерный гель-электрофорез)
emPAI	Exponentially modified protein abundance index (экспоненциально измененный индекс содержания белка)
FASP	Filter-aided sample preparation (подготовка проб с использованием центрифужных концентраторов)
FC	Fold change (кратное изменение)
HBV	Вирус гепатита Б
HCV	Вирус гепатита С
LC-ESI-MS/MS	Liquid Chromatography coupled with Tandem Mass Spectrometry with Electrospray Ionization (жидкостная хроматография в сочетании с тандемной масс-спектрометрией с ионизацией электрораспылением)
SDS	Sodium dodecyl sulfate (додецилсульфат натрия)
SDS-PAGE	SDS-polyacrylamide gel electrophoresis (электрофорез в полиариламидном геле в присутствии додецилсульфата натрия)
АФП	Альфа-фетопротейн
ГЦР	Гепатоцеллюлярный рак (гепатоцеллюлярная карцинома)
ДКП	Дез-гамма-карбокси протромбин
ДЭБ	Дифференциально экспрессируемые белки
ДЭГ	Дифференциально экспрессируемые гены
ИЭФ	Изоэлектрическое фокусирование
МС	Масс-спектрометрия
НАЖБП	Неалкогольная жировая болезнь печени
НАСГ	Неалкогольный стеатогепатит
ОНП	Однонуклеотидный полиморфизм
ПТМ	Посттрансляционные модификации

СПИСОК ТЕРМИНОВ

Белковая сигнатура: фрагмент протеомного профиля, специфичный для конкретного типа образца.

Контрольная ткань печени: фрагмент ткани печени, полученной от пациентов с гистологически подтвержденным диагнозом ГЦР, без видимого поражения ГЦР.

Нормальная ткань печени: ткань печени, полученная от здоровых доноров без видимых патологий

Опухолевая ткань печени: фрагмент ткани печени пораженный ГЦР, полученной от пациентов с гистологически подтвержденным диагнозом ГЦР

Панорамное протеомное профилирование: хромато-масс-спектрометрическая идентификация LC-ESI-MS/MS с предварительной подготовкой проб по протоколу FASP

Паттерн протеоформ: наборов продуктов одного гена (протеоформ) с разными физико-химическими параметрами (изоэлектрическая точка и молекулярная масса)

Протеомный профиль: совокупность протеоформ, детектируемых в образце.

Протеоформа: совокупность белковых продуктов одного гена, включающие, в том числе белки, различающиеся по физико-химическим параметрам

Секционное протеомное профилирование: секционный двумерный гель-электрофорез (2DE) с последующей хромато-масс-спектрометрической идентификацией LC-ESI-MS/MS (Liquid Chromatography coupled with Tandem Mass Spectrometry with Electrospray Ionization).

Хаб-ген: ген, имеющий множество взаимодействий с другими генами (более 10), и, как правило, выполняющий значительную функцию в регуляции генов и биологических процессов.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Bray F. et al. Erratum: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries // *CA. Cancer J. Clin. American Cancer Society*, 2020. Vol. 70, № 4. P. 313.
2. Каприн А.Д., Старинский В.В., Шахзадова А.О. Злокачественные новообразования в России в 2021 году (заболеваемость и смертность) // М.: МНИОИ им. П.А. Герцена. 2022. P. 1–252.
3. Schlageter M. et al. Histopathology of hepatocellular carcinoma // *World J Gastroenterol*. 2014. Vol. 20, № 43. P. 15955–15964.
4. Непомнящая Е.М., Шапошников А.В., Юрьева Е.А. Гепатоцеллюлярная карцинома — новые положения классификации ВОЗ 2019, 5-е издание // *Архив патологии*. 2020. Vol. 82, № 6. P. 36–40.
5. Forner A., Reig M., Bruix J. Hepatocellular carcinoma // *Lancet. Elsevier*, 2018. Vol. 391, № 10127. P. 1301–1314.
6. Татаринцов Ю.С. Обнаружение эмбриоспецифического аглобулина в сыворотке крови больного первичным раком печени // *Вопр. мед. химии*. 1964. Vol. 10, № 1. P. 90–91.
7. Dunbar C., Kushnir M.M., Yang Y.K. Glycosylation profiling of the neoplastic biomarker alpha fetoprotein through intact mass protein analysis // *J. Proteome Res. American Chemical Society*, 2023. Vol. 22, № 1. P. 226–234.
8. Liebman H.A. et al. Des- γ -Carboxy (abnormal) prothrombin as a serum marker of primary hepatocellular carcinoma // *N. Engl. J. Med. Massachusetts Medical Society*, 1984. Vol. 310, № 22. P. 1427–1431.
9. Lee N.P. et al. Proteomic expression signature distinguishes cancerous and nonmalignant tissues in hepatocellular carcinoma // *J. Proteome Res.* 2009. Vol. 8, № 3. P. 1293–1303.
10. Behne T., Copur M.S. Biomarkers for Hepatocellular Carcinoma // *Int. J. Hepatol.* 2012. Vol. 2012. P. 859076.
11. Ponomarenko E.A. et al. The size of the human proteome: the width and depth //

- Int. J. Anal. Chem. 2016. Vol. 2016. P. 7436849.
12. Naryzhny S.N. et al. Combination of virtual and experimental 2DE together with ESI LC-MS/MS gives a clearer view about proteomes of human cells and plasma // *Electrophoresis*. 2016. Vol. 37, № 2. P. 302–309.
 13. Naryzhny S.N. et al. 2DE-based approach for estimation of number of protein species in a cell // *Electrophoresis*. Wiley Online Library, 2014. Vol. 35, № 6. P. 895–900.
 14. Smith L.M. et al. Proteoform: a single term describing protein complexity // *Nat. Methods*. 2013. Vol. 10, № 3. P. 186–187.
 15. Crick F.H. On protein synthesis. // *Symp. Soc. Exp. Biol.* 1958. Vol. 12. P. 138–163.
 16. Aebersold R. et al. How many human proteoforms are there? // *Nat. Chem. Biol.* 2018. Vol. 14, № 3. P. 206–214.
 17. Zhao Y., Jensen O.N. Modification-specific proteomics: Strategies for characterization of post-translational modifications using enrichment techniques // *Proteomics*. 2009. Vol. 9, № 20. P. 4632–4641.
 18. He F. Human Liver Proteome Project // *Mol. Cell. Proteomics*. 2005. Vol. 4, № 12. P. 1841–1848.
 19. Niu L. et al. Dynamic human liver proteome atlas reveals functional insights into disease pathways // *Mol. Syst. Biol.* 2022. Vol. 18, № 5. P. 10947.
 20. Hiotis S.P. et al. Hepatitis B vs. hepatitis C infection on viral hepatitis-associated hepatocellular carcinoma // *BMC Gastroenterol.* 2012. Vol. 12, № 1. P. 64.
 21. Kim W. et al. Comparison of Proteome between Hepatitis B Virus- and Hepatitis C Virus-Associated Hepatocellular Carcinoma // *Clin. Cancer Res.* 2003. Vol. 9, № 15. P. 5493–5500.
 22. Li Y. et al. Integrated bioinformatics analysis reveals key candidate genes and pathways associated with clinical outcome in hepatocellular carcinoma // *Front. Genet.* 2020. Vol. 11. P. 814.
 23. Jiang C.H. et al. Bioinformatics-based screening of key genes for transformation of liver cirrhosis to hepatocellular carcinoma // *J. Transl. Med.* 2020. Vol. 18, №

1. P. 40.
24. Kakar M.U. et al. Identification of differentially expressed genes associated with the prognosis and diagnosis of hepatocellular carcinoma by integrated bioinformatics analysis // *Biomed Res. Int.* 2022. Vol. 2022. P. 4237633.
25. Loo E. et al. BRAF V600E Mutation Across Multiple Tumor Types: Correlation Between DNA-based Sequencing and Mutation-specific Immunohistochemistry // *Appl. Immunohistochem. Mol. Morphol.* 2018. Vol. 26, № 10. P. 709–713.
26. Álvarez-García V. et al. Mechanisms of PTEN loss in cancer: It's all about diversity // *Semin. Cancer Biol.* 2019. Vol. 59. P. 66–79.
27. Contini C. et al. A top-down proteomic approach reveals a salivary protein profile able to classify Parkinson's disease with respect to Alzheimer's disease patients and to healthy controls // *Proteomics.* 2023. Vol. 24. P. e2300202.
28. Noor A. et al. Molecular Profiles of Amyloid- β Proteoforms in Typical and Rapidly Progressive Alzheimer's Disease // *Mol. Neurobiol.* 2022. Vol. 59, № 1. P. 17–34.
29. Chapman E.A. et al. Defining the Sarcomeric Proteoform Landscape in Ischemic Cardiomyopathy by Top-Down Proteomics // *J. Proteome Res.* 2023. Vol. 22, № 3. P. 931–941.
30. Garcia-Garcia A. et al. FUT8-Directed core fucosylation of N-glycans is regulated by the glycan structure and protein environment // *ACS Catal.* American Chemical Society, 2021. Vol. 11, № 15. P. 9052–9065.
31. Mao C. et al. Beyond antibody fucosylation: α -(1,6)-fucosyltransferase (Fut8) as a potential new therapeutic target for cancer immunotherapy // *Antib. Ther.* 2023. Vol. 6, № 2. P. 87–96.
32. Bastian K. et al. FUT8 Alpha-(1,6)-Fucosyltransferase in Cancer // *Int. J. Mol. Sci.* 2021. Vol. 22, № 1. P. 455.
33. Чекмазов И.А. et al. Рак печени: этиология, патогенез, итоги длительного клинико-эпидемиологического наблюдения // *Доказательная гастроэнтерология.* 2019. Vol. 8, № 1. P. 5–15.
34. Клинические рекомендации “Рак Печени (Гепатоцеллюлярный)” //

- Министерство здравоохранения Российской Федерации. 2020. P. 1–57.
35. Гусейнов А.З., Гусейнов Т.А. Современная диагностика опухолей печени (Обзор литературы) // Вестник Новых Медицинских Технологий. 2016. Vol. 10, № 4. P. 359–377.
 36. Кушлинский Н.Е., Любимова Н.В. Опухолевые маркеры. Общая характеристика, клиническое значение и рекомендации по использованию // Поликлиника. 2016. Vol. 1–3. P. 62–77.
 37. Saffroy R. et al. New perspectives and strategy research biomarkers for hepatocellular carcinoma. 2007. Vol. 45, № 9. P. 1169–1179.
 38. Kuo P.-C. et al. Hepatoid carcinoma of the pancreas // World J. Surg. Oncol. 2015. Vol. 13, № 1. P. 185.
 39. Pedrazzoli P. et al. Serum tumour markers in germ cell tumours: From diagnosis to cure // Crit. Rev. Oncol. Hematol. 2021. Vol. 159. P. 103224.
 40. O'Neill A.F. et al. α -Fetoprotein as a predictor of outcome for children with germ cell tumors: A report from the Malignant Germ Cell International Consortium // Cancer. John Wiley & Sons, Ltd, 2019. Vol. 125, № 20. P. 3649–3656.
 41. Głowska-Ciemny J. et al. Fetal and placental causes of elevated serum alpha-fetoprotein levels in pregnant women // Journal of Clinical Medicine. 2024. Vol. 13, № 2.
 42. Wang X. et al. Alpha-Fetoprotein as a predictive marker for patients with hepatitis B-related acute-on-chronic liver failure // Can. J. Gastroenterol. Hepatol. John Wiley & Sons, Ltd, 2018. Vol. 2018, № 1. P. 1232785.
 43. Galle P.R. et al. EASL clinical practice guidelines: management of hepatocellular carcinoma // J. Hepatol. Elsevier, 2018. Vol. 69, № 1. P. 182–236.
 44. Chuma M. et al. Expression profiling in multistage hepatocarcinogenesis: Identification of HSP70 as a molecular marker of early hepatocellular carcinoma // Hepatology. 2003. Vol. 37, № 1. P. 198–207.
 45. Wang B. et al. The expression profiles and prognostic values of HSP70s in hepatocellular carcinoma // Cancer Cell Int. BioMed Central, 2021. Vol. 21, № 1. P. 1–17.

46. Gehrman M. et al. Heat shock protein 70 serum levels differ significantly in patients with chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma // *Front. Immunol.* 2014. Vol. 5, № JUL. P. 1–7.
47. Zhou Y. et al. CDKN2A promoter methylation and hepatocellular carcinoma risk: A meta-analysis // *Clin. Res. Hepatol. Gastroenterol.* 2018. Vol. 42, № 6. P. 529–541.
48. Luo J., Wang J., Huang J. CDKN2A is a prognostic biomarker and correlated with immune infiltrates in hepatocellular carcinoma // *Biosci. Rep.* 2021. Vol. 41, № 10. P. BSR20211103.
49. Zucman-Rossi J. et al. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma // *Gastroenterology.* 2015. Vol. 149, № 5. P. 1226–1239.
50. Huo J., Wu L., Zang Y. Development and validation of a CTNNB1-associated metabolic prognostic model for hepatocellular carcinoma // *J. Cell. Mol. Med.* 2021. Vol. 25, № 2. P. 1151–1165.
51. Penha Mesquita A. et al. Gene variations related to the hepatocellular carcinoma: Results from a field synopsis and Bayesian reevaluation // *Gene.* 2023. Vol. 869. P. 147392.
52. Chen L. et al. Genetic factors in the clinical predictive model for hepatocellular carcinoma: Evidence from genetic association analyses // *J. Hepatol.* 2023. Vol. 79, № 1. P. e33–e35.
53. Ke L. et al. Somatic Mutation Profiles Revealed by Next Generation Sequencing (NGS) in 39 Chinese Hepatocellular Carcinoma Patients // *Frontiers in Molecular Biosciences.* 2022. Vol. 8.
54. Marin J.J.G. et al. Models for Understanding Resistance to Chemotherapy in Liver Cancer // *Cancers.* 2019. Vol. 11, № 11.
55. Aden D.P. et al. Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line // *Nature.* 1979. Vol. 282, № 5739. P. 615–616.
56. López-Terrada D. et al. Hep G2 is a hepatoblastoma-derived cell line // *Hum. Pathol. Elsevier Inc.,* 2009. Vol. 40, № 10. P. 1512–1515.
57. Qiu G.-H. et al. Distinctive pharmacological differences between liver cancer cell

- lines HepG2 and Hep3B // *Cytotechnology*. 2015. Vol. 67, № 1. P. 1–12.
58. Bagi C.M., Andresen C.J. Models of Hepatocellular Carcinoma and Biomarker Strategy // *Cancers*. 2010. Vol. 2, № 3. P. 1441–1452.
59. Blidisel A. et al. Experimental Models of Hepatocellular Carcinoma—A Preclinical Perspective // *Cancers*. 2021. Vol. 13, № 15.
60. Haider S., Pal R. Integrated Analysis of Transcriptomic and Proteomic Data // *Current Genomics*. 2013. Vol. 14, № 2. P. 91–110.
61. Wang D. et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues // *Mol. Syst. Biol.* 2019. Vol. 15, № 2. P. e8503.
62. Saiki R.K. et al. Enzymatic Amplification of β -Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia // *Science* (80-.). 1985. Vol. 230, № 4732. P. 1350–1354.
63. Karas M. et al. Matrix-assisted ultraviolet laser desorption of non-volatile compounds // *Int. J. Mass Spectrom. Ion Process.* 1987. Vol. 78. P. 53–68.
64. Hood L., Rowen L. The Human Genome Project: big science transforms biology and medicine // *Genome Med.* 2013. Vol. 5, № 9. P. 79.
65. Wilkins M.R. et al. Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it // *Biotechnol. Genet. Eng. Rev.* 1996. Vol. 13, № 1. P. 19–50.
66. Yates J.R., Ruse C.I., Nakorchevsky A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications // *Annu. Rev. Biomed. Eng.* 2009. Vol. 11, № 1. P. 49–79.
67. Banerjee S. Empowering Clinical Diagnostics with Mass Spectrometry // *ACS Omega*. American Chemical Society, 2020. Vol. 5, № 5. P. 2041–2048.
68. Lisitsa A. V et al. Application of Slicing of One-Dimensional Gels with Subsequent Slice-by-Slice Mass Spectrometry for the Proteomic Profiling of Human Liver Cytochromes P450 // *J. Proteome Res.* 2010. Vol. 9, № 1. P. 95–103.
69. Jafari M. et al. Comparison of in-gel protein separation techniques commonly used for fractionation in mass spectrometry-based proteomic profiling //

- Electrophoresis. 2012. Vol. 33, № 16. P. 2516–2526.
70. Maccarrone G., Turck C.W., Martins-de-Souza D. Shotgun Mass Spectrometry Workflow Combining IEF and LC-MALDI-TOF/TOF // *Protein J.* 2010. Vol. 29, № 2. P. 99–102.
 71. Spicer V. et al. 3D HPLC-MS with Reversed-Phase Separation Functionality in All Three Dimensions for Large-Scale Bottom-Up Proteomics and Peptide Retention Data Collection // *Anal. Chem.* 2016. Vol. 88, № 5. P. 2847–2855.
 72. Yates J.R. Mass Spectral Analysis in Proteomics // *Annu. Rev. Biophys. Biomol. Struct.* 2004. Vol. 33, № 1. P. 297–316.
 73. Wilkins M. Proteomics data mining // *Expert Rev. Proteomics.* 2009. Vol. 6, № 6. P. 599–603.
 74. Smithies O., Poulik M.D. Two-Dimensional Electrophoresis of Serum Proteins // *Nature.* 1956. Vol. 177, № 4518. P. 1033.
 75. Awdeh Z.L., Williamson A.R., Askonas B.A. Isoelectric focusing in polyacrylamide gel and its application to immunoglobulins // *Nature.* 1968. Vol. 219, № 5149. P. 66–67.
 76. Righetti P.G., Drysdale J.W. Isoelectric focusing in gels // *J. Chromatogr. A.* 1974. Vol. 98, № 2. P. 271–321.
 77. Dale G., Latner A.L. Isoelectric focusing in polyacrylamide gels // *Lancet.* 1968. Vol. 291, № 7547. P. 847–848.
 78. Klose J. From 2- D electrophoresis to proteomics // *Electrophoresis. Wiley Online Library*, 2009. Vol. 30, № S1. P. S142–S149.
 79. Hamdan M.H., Righetti P.G. Proteomics today: protein assessment and biomarkers using mass spectrometry, 2D electrophoresis, and microarray technology. John Wiley & Sons, 2005. Vol. 18.
 80. O'Farrell P.H. High resolution two-dimensional electrophoresis of proteins // *J. Biol. Chem.* 1975. Vol. 250, № 10. P. 4007–4021.
 81. Anderson N.G., Matheson A., Anderson N.L. Back to the future: The human protein index (HPI) and the agenda for post- proteomic biology // *Proteomics.* 2001. Vol. 1, № 1. P. 3–12.

82. Anderson N.G., Anderson L. The human protein index // *Clin Chem.* 1982. Vol. 28, № 4 pt 2. P. 739–748.
83. Görg A. et al. The current state of two- dimensional electrophoresis with immobilized pH gradients // *Electrophor. An Int. J.* 2000. Vol. 21, № 6. P. 1037–1053.
84. Seddon A.M., Curnow P., Booth P.J. Membrane proteins, lipids and detergents: not just a soap opera // *Biochim. Biophys. Acta (BBA)-Biomembranes.* 2004. Vol. 1666, № 1–2. P. 105–117.
85. Rabilloud T. et al. Silver-staining of proteins in polyacrylamide gels: a general overview // *Cell Mol Biol.* 1994. Vol. 40, № 1. P. 57–75.
86. Chevalier F. et al. Different impact of staining procedures using visible stains and fluorescent dyes for large-scale investigation of proteomes by MALDI-TOF mass spectrometry // *J. Proteome Res.* 2006. Vol. 5, № 3. P. 512–520.
87. Towbin H., Staehelin T., Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications // *Proc. Natl. Acad. Sci.* 1979. Vol. 76, № 9. P. 4350–4354.
88. Naryzhny S.N. Blue Dry Western: Simple, economic, informative, and fast way of immunodetection // *Anal. Biochem.* 2009. Vol. 392, № 1. P. 90–95.
89. Jungblut P.R., Thiede B., Schlüter H. Towards deciphering proteomes via the proteoform, protein speciation, moonlighting and protein code concepts // *J. Proteomics.* 2016. Vol. 134. P. 1–4.
90. Fey S.J., Larsen P.M. 2D or not 2D // *Curr. Opin. Chem. Biol.* 2001. Vol. 5, № 1. P. 26–33.
91. mzML 1.1.0 Specification [Electronic resource]. URL: <https://www.psdev.info/mzML>.
92. Elias J.E., Gygi S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry // *Nat. Methods.* 2007. Vol. 4, № 3. P. 207–214.
93. Benjamini Y., Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing // *Journal of the Royal Statistical Society.*

- Series B (Methodological). 1995. Vol. 57, № 1. P. 289–300.
94. Naryzhny S. Inventory of proteoforms as a current challenge of proteomics: Some technical aspects // *J. Proteomics*. 2019. Vol. 191. P. 22–28.
 95. LeDuc R.D. et al. ProForma: A Standard Proteoform Notation // *J. Proteome Res.* 2018. Vol. 17, № 3. P. 1321–1325.
 96. LeDuc R.D. et al. Proteomics Standards Initiative's ProForma 2.0: Unifying the Encoding of Proteoforms and Peptidoforms // *J. Proteome Res.* 2022. Vol. 21, № 4. P. 1189–1195.
 97. Hollas M.A.R. et al. The Human Proteoform Atlas: a FAIR community resource for experimentally derived proteoforms // *Nucleic Acids Res.* 2022. Vol. 50, № D1. P. D526–D533.
 98. Naryzhny S. et al. A semi-virtual two dimensional gel electrophoresis: IF–ESI LC-MS/MS // *MethodsX*. 2017. Vol. 4. P. 260–264.
 99. Naryzhny S. et al. A database for inventory of proteoform profiles: “2DE-pattern” // *Electrophoresis*. 2020. Vol. 41, № 12. P. 1118–1124.
 100. Thomas P.D. et al. PANTHER: Making genome-scale phylogenetics accessible to all // *Protein Sci.* 2022. Vol. 31, № 1. P. 8–22.
 101. Wiśniewski J.R. Filter-aided sample preparation for proteome analysis // *Microb. Proteomics. Methods Mol. Biol.* 2018. Vol. 1841. P. 3–10.
 102. Wiśniewski J.R. et al. In-depth quantitative analysis and comparison of the human hepatocyte and hepatoma cell line HepG2 proteomes // *J. Proteomics*. 2016. Vol. 136. P. 234–247.
 103. Naryzhny S. et al. Variety and Dynamics of Proteoforms in the Human Proteome: Aspects of Markers for Hepatocellular Carcinoma // *Proteomes*. 2017. Vol. 5, № 4. P. 33.
 104. Jiatao Lou, LingFei Zhang, Shaogang Lv, Jiang C.Z. and S. Biomarkers for Hepatocellular Carcinoma // *Biomark. Cancer*. 2017. P. 1–9.
 105. Wright L.M., Kreikemeier J.T., Fimmel C.J. A concise review of serum markers for hepatocellular cancer // *Cancer Detection and Prevention*. 2007. Vol. 31, № 1. P. 35–44.

106. Oka H. et al. Multicenter prospective analysis of newly diagnosed hepatocellular carcinoma with respect to the percentage of Lens culinaris agglutinin-reactive α -fetoprotein // *Journal of Gastroenterology and Hepatology (Australia)*. 2001. Vol. 16, № 12. P. 1378–1383.
107. Shiota G., Miura N. Biomarkers for hepatocellular carcinoma // *Clinical Journal of Gastroenterology*. 2012. Vol. 5, № 3. P. 177–182.
108. Zhao Y.-J., Ju Q., Li G.-C. Tumor markers for hepatocellular carcinoma. // *Mol. Clin. Oncol.* 2013. Vol. 1, № 4. P. 593–598.
109. Zhu W.L. et al. Abnormal expression of fibrinogen gamma (FGG) and plasma level of fibrinogen in patients with hepatocellular carcinoma // *Anticancer Res.* 2009. Vol. 29, № 7. P. 2531–2534.
110. Zhang S.Y., Lin B.D., Li B.R. Evaluation of the diagnostic value of alpha-l-fucosidase, alpha-fetoprotein and thymidine kinase 1 with ROC and logistic regression for hepatocellular carcinoma // *FEBS Open Bio.* 2015. Vol. 5. P. 240–244.
111. Fimmel C.J., Wright L. Golgi protein 73 as a biomarker of hepatocellular cancer: Development of a quantitative serum assay and expression studies in hepatic and extrahepatic malignancies // *Hepatology*. 2009. Vol. 49, № 5. P. 1421–1423.
112. Wang C. et al. Heat shock proteins in hepatocellular carcinoma: Molecular mechanism and therapeutic potential // *Int. J. Cancer*. 2016. Vol. 138, № 8. P. 1824–1834.
113. Pan F. et al. The critical role of ferroptosis in hepatocellular carcinoma // *Front. Cell Dev. Biol.* 2022. Vol. 10. P. 882571.
114. Matsushima-Nishiwaki R. et al. Correlation between the complex of small heat shock proteins (HSPBs) and the progression in patients with hepatocellular carcinoma // *Arch. Biochem. Biophys.* 2022. Vol. 732. P. 109461.
115. Sun J. et al. Engineered small extracellular vesicles loaded with miR-654-5p promote ferroptosis by targeting HSPB1 to alleviate sorafenib resistance in hepatocellular carcinoma // *Cell Death Discov.* 2023. Vol. 9, № 1. P. 362.
116. Chaiteerakij R., Addissie B.D., Roberts L.R. Update on Biomarkers of

- Hepatocellular Carcinoma // Clin. Gastroenterol. Hepatol. 2015. Vol. 13, № 2. P. 237–245.
117. Shang S. et al. Identification of osteopontin as a novel marker for early hepatocellular carcinoma // Hepatology. 2012. Vol. 55, № 2. P. 483–490.
118. Schmilovitz-Weiss H. et al. Tissue expression of squamous cellular carcinoma antigen and Ki67 in hepatocellular carcinoma-correlation with prognosis: a historical prospective study. // Diagn. Pathol. 2011. Vol. 6. P. 121.
119. Dehm S., Senger M. a, Bonham K. SRC transcriptional activation in a subset of human colon cancer cell lines. // FEBS Lett. 2001. Vol. 487, № 3. P. 367–371.
120. Sigala I. et al. Expression of SRPK1 in gliomas and its role in glioma cell lines viability // Tumor Biol. 2016. Vol. 37, № 7. P. 8699–8707.
121. Tsai H.-W. et al. Progesterone receptor membrane component 1 as a potential prognostic biomarker for hepatocellular carcinoma // World J Gastroenterol. 2018. Vol. 24, № 10. P. 1152–1166.
122. Guo H. et al. Cytochrome B5 type A alleviates HCC metastasis via regulating STOML2 related autophagy and promoting sensitivity to ruxolitinib // Cell Death Dis. 2022. Vol. 13, № 7. P. 623.
123. Kiseleva O. et al. Empowering Shotgun Mass Spectrometry with 2DE: A HepG2 Study // International Journal of Molecular Sciences. 2020. Vol. 21, № 11.
124. Beck M. et al. The quantitative proteome of a human cell line // Mol. Syst. Biol. 2011. Vol. 7, № 1. P. 549.
125. Вавилов Н.Э. Детекция низкопредставленных белков 18 хромосомы методами таргетной протеомики: дис. канд. биол. наук: 1.5.4. М., 2023. P. 112.
126. Xu B.-H. et al. Aberrant amino acid signaling promotes growth and metastasis of hepatocellular carcinomas through Rab1A-dependent activation of mTORC1 by Rab1A // Oncotarget. 2015. Vol. 6, № 25. P. 20813–20828.
127. Zhang C. et al. S100 family members: potential therapeutic target in patients with hepatocellular carcinoma: A STROBE study // Medicine (Baltimore). 2021. Vol. 100, № 3.

128. Lu K. et al. Sustainable inflammation transforms hepatic cells by causing oxidative stress injury and potential epithelial-mesenchymal transition // *Int J Oncol.* 2016. Vol. 49, № 3. P. 971–980.
129. Chen J. et al. 17-beta-hydroxysteroid dehydrogenase 13 inhibits the progression and recurrence of hepatocellular carcinoma // *Hepatobiliary Pancreat. Dis. Int.* 2018. Vol. 17, № 3. P. 220–226.
130. Yao S. et al. Comprehensive Analysis of Aldehyde Dehydrogenases (ALDHs) and Its Significant Role in Hepatocellular Carcinoma // *Biochem. Genet.* 2022. Vol. 60, № 4. P. 1274–1297.
131. Jin B. et al. Downregulation of betaine homocysteine methyltransferase (BHMT) in hepatocellular carcinoma associates with poor prognosis // *Tumor Biol.* 2016. Vol. 37, № 5. P. 5911–5917.
132. Walakira A. et al. Integrative computational modeling to unravel novel potential biomarkers in hepatocellular carcinoma // *Comput. Biol. Med.* 2023. Vol. 159. P. 106957.
133. Yang J. et al. Loss of FBP1 facilitates aggressive features of hepatocellular carcinoma cells through the Warburg effect // *Carcinogenesis.* 2017. Vol. 38, № 2. P. 134–143.
134. Jin G.-Z. et al. Phosphoglucomutase 1 inhibits hepatocellular carcinoma progression by regulating glucose trafficking // *PLOS Biol. Public Library of Science,* 2018. Vol. 16, № 10. P. e2006483.
135. Matos J.M. et al. A Pilot Study of Proteomic Profiles of Human Hepatocellular Carcinoma in the United States // *J. Surg. Res.* 2009. Vol. 155, № 2. P. 237–243.
136. Dou C. et al. SHMT1 inhibits the metastasis of HCC by repressing NOX1-mediated ROS production // *J. Exp. Clin. Cancer Res.* 2019. Vol. 38, № 1. P. 70.
137. Owji H. et al. A comprehensive review of signal peptides: structure, roles, and applications // *Eur. J. Cell Biol.* 2018. Vol. 97, № 6. P. 422–441.
138. Nanjappa V. et al. Plasma Proteome Database as a resource for proteomics research: 2014 update // *Nucleic Acids Res.* 2014. Vol. 42, № D1. P. D959–D965.
139. Boyd A. et al. Medical liver biopsy: background, indications, procedure and

- histopathology // *Frontline Gastroenterol.* 2020. Vol. 11, № 1. P. 40–47.
140. Gasmi B., Kleiner D.E. Liver Histology: Diagnostic and Prognostic Features // *Clin. Liver Dis.* 2020. Vol. 24, № 1. P. 61–74.
141. Feng K. et al. SORBS1 inhibits epithelial to mesenchymal transition (EMT) of breast cancer cells by regulating PI3K/AKT signaling and macrophage phenotypic polarization // *Aging (Albany, NY)*. 2024. Vol. 16. P. 4789–4810.
142. Xu H. et al. circSORBS1 inhibits lung cancer progression by sponging miR-6779-5p and directly binding RUFY3 mRNA // *J. Transl. Med.* 2024. Vol. 22, № 1. P. 590.
143. Song L. et al. SORBS1 suppresses tumor metastasis and improves the sensitivity of cancer to chemotherapy drug // *Oncotarget*. 2016. Vol. 8, № 6. P. 9108–9122.
144. Guo Y. et al. Heterogeneous nuclear ribonucleoprotein K (hnRNP K) is a tissue biomarker for detection of early Hepatocellular carcinoma in patients with cirrhosis // *J. Hematol. Oncol.* 2012. Vol. 5, № 1. P. 37.
145. Gao J., Xu D. Correlation between posttranslational modification and intrinsic disorder in protein // *Biocomputing 2012*. 2011. P. 94–103.
146. Souza R.W.A. et al. Differential regulation of cysteine oxidative post-translational modifications in high and low aerobic capacity // *Sci. Rep.* 2018. Vol. 8, № 1. P. 17772.
147. Kim K. et al. Reducing protein oxidation in low-flow electrospray enables deeper investigation of proteoforms by top down proteomics // *EuPA Open Proteomics*. 2015. Vol. 8. P. 40–47.
148. Verrastro I. et al. Mass Spectrometry-Based Methods for Identifying Oxidized Proteins in Disease: Advances and Challenges // *Biomolecules*. 2015. Vol. 5, № 2. P. 378–411.
149. Wang Y.-W. et al. Post-translational modifications and immune responses in liver cancer // *Frontiers in Immunology*. 2023. Vol. 14. P. 1230465.
150. Wang Y., Chen H. Protein glycosylation alterations in hepatocellular carcinoma: function and clinical implications // *Oncogene*. 2023. Vol. 42, № 24. P. 1970–1979.

151. Morita M. et al. PKM1 Confers Metabolic Advantages and Promotes Cell-Autonomous Tumor Cell Growth // *Cancer Cell*. 2018. Vol. 33, № 3. P. 355-367.e7.
152. Christofk H.R. et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth // *Nature*. 2008. Vol. 452, № 7184. P. 230–233.
153. Bellisola G. et al. Catalase activity in human hepatocellular carcinoma (HCC) // *Clin. Biochem*. 1987. Vol. 20, № 6. P. 415–417.
154. Razaghi A. et al. Negative regulators of cell death pathways in cancer: perspective on biomarkers and targeted therapies // *Apoptosis*. 2018. Vol. 23, № 2. P. 93–112.
155. Shi Q. et al. A novel prognostic model for hepatocellular carcinoma based on pyruvate metabolism-related genes // *Sci. Rep*. 2023. Vol. 13, № 1. P. 9780.
156. Wu J. et al. Characterization of metabolic landscape in hepatocellular carcinoma // *World J. Gastrointest. Oncol*. 2021. Vol. 13, № 9. P. 1144.
157. Berndt N. et al. Metabolic heterogeneity of human hepatocellular carcinoma: implications for personalized pharmacological treatment // *FEBS J*. 2021. Vol. 288, № 7. P. 2332–2346.
158. Schmucker D.L. Aging and the Liver: An Update // *Journals Gerontol. Ser. A*. 1998. Vol. 53A, № 5. P. B315–B321.

ПРИЛОЖЕНИЕ 1. Описание образцов ткани печени

Характеристика образцов ткани печени от доноров.

№ 1 – здоровый донор.

№ 2-4 – пациенты с гистологически подтвержденным диагнозом ГЦР.

№	Пол	Возраст, лет	Основной диагноз, стадия	Сопутствующие заболевания	Предшествующее лечение
1	М	–	–	–	–
2	М	66	ГЦР, сТ1bN0M0	HBV, Гипертоническая болезнь 1 ст.; Открытоугольная глаукома левого глаза	–
3	М	64	ГЦР, сТ3NхM0	HBV, Гипертоническая болезнь 2 ст., Варикозная болезнь нижних конечностей	Трансартериальная химиоэмболизация (07.03.2018); Бисегментэктомия С5-С6 (07.04.2018); Резекция С7 (19.11.2018) – прогрессия (очаг в С4 печени)
4	М	50	ГЦР, сТ2NхM0	НСV; гипертоническая болезнь 3 ст.; Ожирение 3 ст.; Атеросклероз аорты и коронарных артерий; Мочекаменная болезнь	Противовирусная терапия (Викейра пак) от 2017 года

ПРИЛОЖЕНИЕ 2. Полный список белков идентифицированных двумя методами – секционным и панорамным протеомным профилированием

Данные размещены в Yandex Disk и доступны по ссылке:

<https://disk.yandex.ru/d/4ihxaTXOXMcumA>

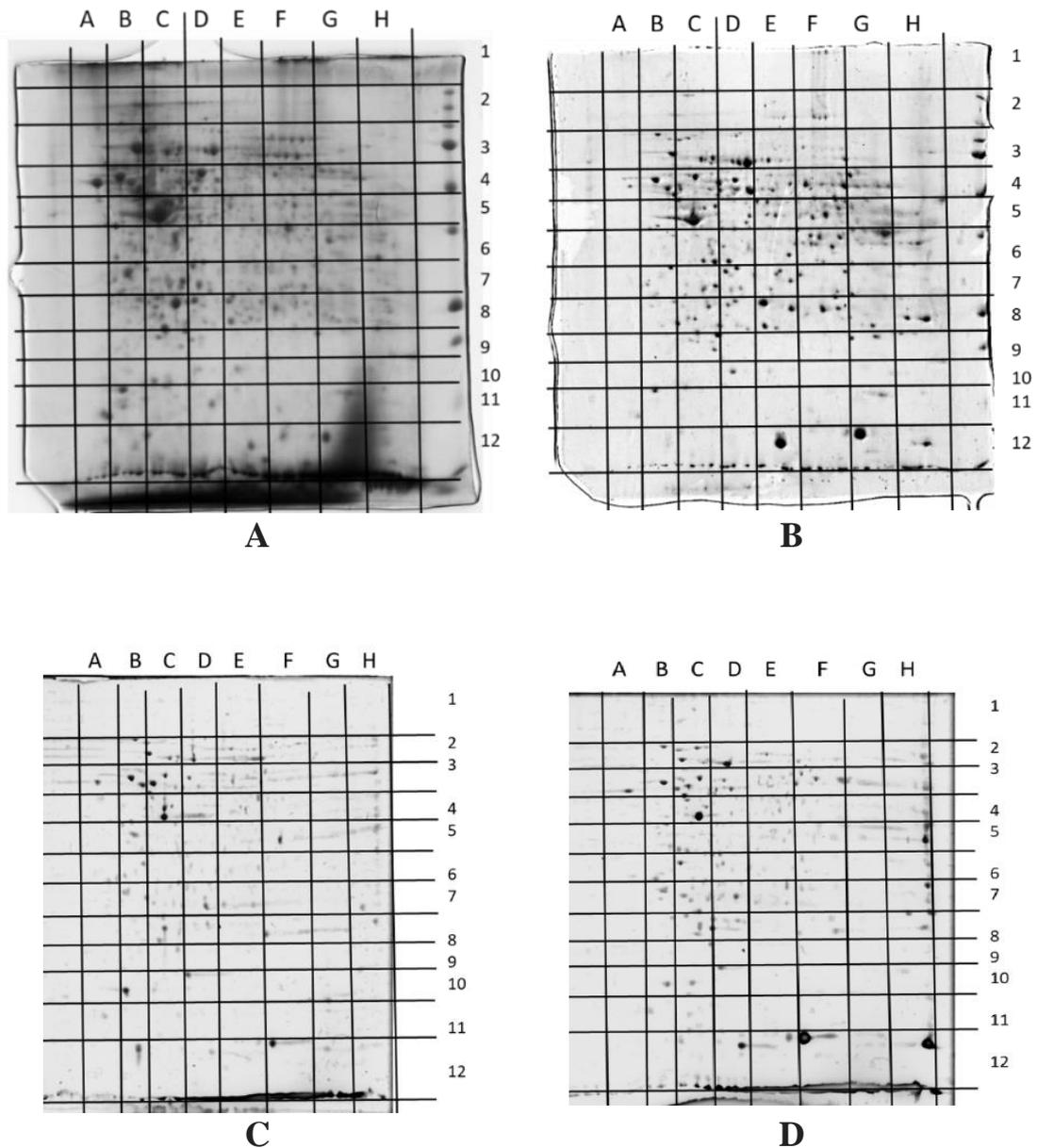


ПРИЛОЖЕНИЕ 3. Двумерные электрофореграммы образцов от пациентов с ГЦР

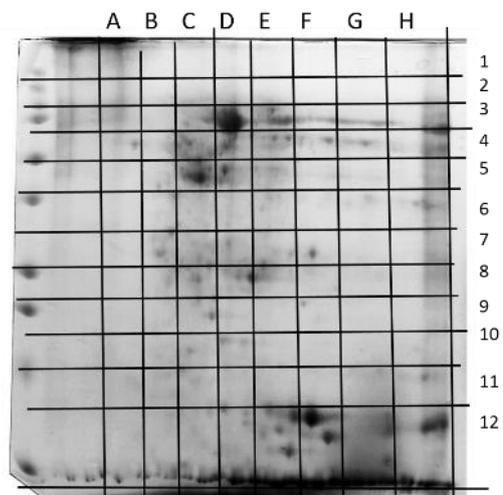
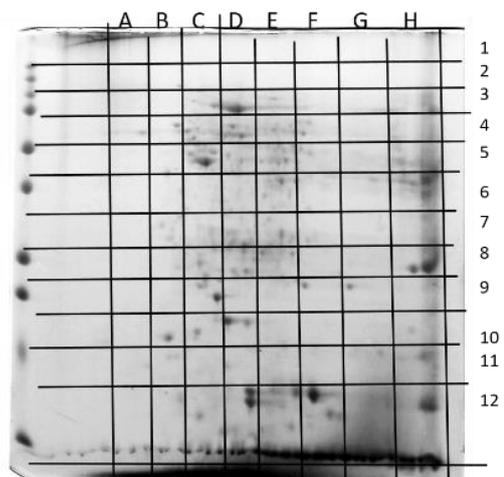
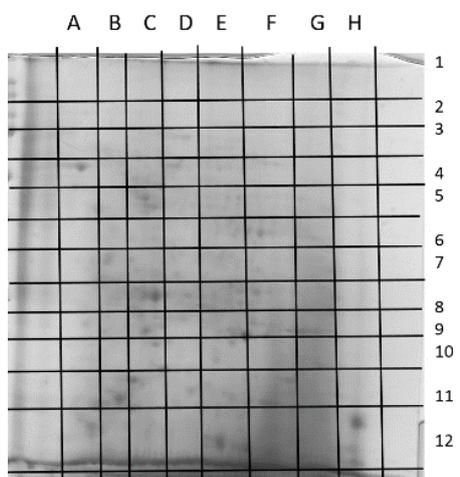
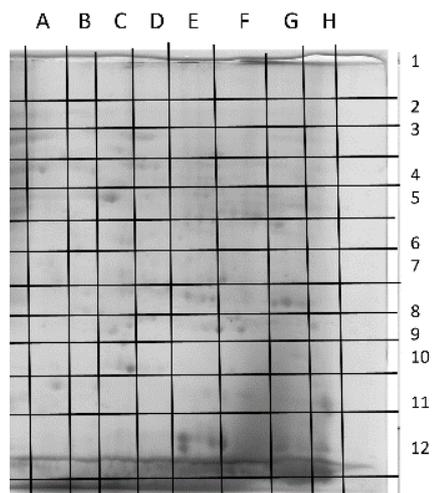
Опухолевая ткань печени: **A** – HCC1 (первый пациент), **C** – HCC2 (второй пациент), **E** – HCC3 (третий пациент), **G** – HCC1*(биологический повтор первого пациента).

Контрольная ткань печени: **B** – HCC1 (первый пациент), **D** – HCC2, (второй пациент), **F** – HCC3 (третий пациент), **H** – HCC1*(биологический повтор первого пациента).

Показаны секции в 2DE гелях, выбранные для последующего анализа ESI LC-MS/MS.



Продолжение Приложения 3.

**E****F****G****H**

ПРИЛОЖЕНИЕ 4. Сравнение данных, полученных для 52 белков с наиболее измененными паттернами протеоформ с базой данных Plasma Proteome Database (PPD)

Белки, представленные в базе данных Plasma Proteome Database помечены +.

* – обнаружены только в опухолевых образцах, но не в контроле.

Жирным шрифтом выделены 23 белка с сигнальным пептидом на N-конце.

Название	Номер Uniprot	Белок	Ген	Повышенная регуляция, определенная методом 2DE, в ГЦР	Повышенная регуляция, определенная методом FASP, в ГЦР	Plasma Proteome Database(http://www.plasmaproteomedatabase.org/)			
						Общие 42	PPD ID	Ген	Идентификатор гена
ATP-citrate synthase	P53396	ACLY	<i>ACLY</i>	17,24	1,19	+	HPRD_00155	<i>ACLY</i>	47
Long-chain-fatty-acid--CoA ligase 4	O60488	ACSL4	<i>ACSL4</i>	375,38	*	+	HPRD_02152	<i>ACSL4</i>	2182
Alpha-fetoprotein	P02771	FETA	<i>AFP</i>	*	*	+	HPRD_00074	<i>AFP</i>	174
Aldo-keto reductase family 1 member B10	O60218	AK1BA	<i>AKR1B10</i>	232,26	80,78				
Aldo-keto reductase family 1 member B15	C9JRZ8	AK1BF	<i>AKR1B15</i>	81,9	*				
Alkaline phosphatase, tissue-nonspecific isozyme	P05186	PPBT	<i>ALPL</i>	12,8	*	+	HPRD_01377	<i>ALPL</i>	249
Annexin A1	P04083	ANXA1	<i>ANXA1</i>	1,62	2,24	+	HPRD_01060	<i>ANXA1</i>	301
Annexin A2	P07355	ANXA2	<i>ANXA2</i>	1,85	1,52	+	HPRD_01061	<i>ANXA2</i>	302
Annexin A3	P12429	ANXA3	<i>ANXA3</i>	7,5	5,33	+	HPRD_00111	<i>ANXA3</i>	306
Coatomer subunit delta	P48444	COPD	<i>ARCNI</i>	3,2	1,59	+	HPRD_02893	<i>ARCNI</i>	372
Actin-related protein 2/3 complex subunit 2	O15144	ARPC2	<i>ARPC2</i>	2,02	2,54	+	HPRD_10367	<i>ARPC2</i>	10109

Продолжение Приложения 4.

Название	Номер Uniprot	Белок	Ген	Повышенная регуляция, определенная методом 2DE, в ГЦР	Повышенная регуляция, определенная методом FASP, в ГЦР	Plasma Proteome Database(http://www.plasmaproteomedatabase.org/)			
						Общие 42	PPD ID	Ген	Идентификатор гена
Macrophage-capping protein	P40121	CAPG	<i>CAPG</i>	31,22	*	+	HPRD_01088	<i>CAPG</i>	822
C-reactive protein	P02741	CRP	<i>CRP</i>	35,27	*	+	HPRD_00422	<i>CRP</i>	1401
Cathepsin G	P08311	CATG	<i>CTSG</i>	14,28	4,43	+	HPRD_00289	<i>CTSG</i>	1511
EGF-containing fibulin-like extracellular matrix protein 1	Q12805	FBLN3	<i>EFEMP1</i>	11,39	17,2	+	HPRD_03331	<i>EFEMP1</i>	2202
Epiplakin	P58107	EPIPL	<i>EPPK1</i>	29,61	*	+	HPRD_06500	<i>EPPK1</i>	83481
Prothrombin	P00734	THRB	<i>F2</i>	2,21	2	+	HPRD_01488	<i>F2</i>	2147
Polypeptide N-acetylgalactosaminyltransferase 2	Q10471	GALT2	<i>GALNT2</i>	32,8	5,24	+	HPRD_03782	<i>GALNT2</i>	2590
Lysosomal acid glucosylceramidase	P04062	GLCM	<i>GBA</i>	23,83	*	+	HPRD_06973	<i>GBA</i>	2629
HLA class I histocompatibility antigen, alpha chain F	P30511	HLAF	<i>HLA-F</i>	8,09	*	+	HPRD_00884	<i>HLA-F</i>	3134
Integrin alpha-M	P11215	ITAM	<i>ITGAM</i>	*	14,55	+	HPRD_00411	<i>ITGAM</i>	3684
Integrin beta-2	P05107	ITB2	<i>ITGB2</i>	14,89	5,02	+	HPRD_02506	<i>ITGB2</i>	3689
Neutrophil gelatinase-associated lipocalin	P80188	NGAL	<i>LCN2</i>	13,56	11,49	+	HPRD_02551	<i>LCN2</i>	3934
Galectin-3	P17931	LEG3	<i>LGALS3</i>	2,21	*	+	HPRD_01090	<i>LGALS3</i>	3958
Lactotransferrin	P02788	TRFL	<i>LTF</i>	21,11	7,77	+	HPRD_01028	<i>LTF</i>	4057
Mannan-binding lectin serine protease 2	O00187	MASP2	<i>MASP2</i>	2,51	*	+	HPRD_05484	<i>MASP2</i>	10747

Продолжение Приложения 4.

Название	Номер Uniprot	Белок	Ген	Повышенная регуляция, определенная методом 2DE, в ГЦР	Повышенная регуляция, определенная методом FASP, в ГЦР	Plasma Proteome Database(http://www.plasmaproteomedatabase.org/) ¹			
						Общие 42	PPD ID	Ген	Идентификатор гена
NAD-dependent malic enzyme, mitochondrial	P23368	MAOM	<i>ME2</i>	5,58	1,88	+	HPRD_01103	<i>ME2</i>	4200
Myeloid cell nuclear differentiation antigen	P41218	MNDA	<i>MNDA</i>	41,75	*	+	HPRD_15932	<i>MNDA</i>	4332
Myeloperoxidase	P05164	PERM	<i>MPO</i>	19,12	10,26	+	HPRD_06102	<i>MPO</i>	4353
39S ribosomal protein L24, mitochondrial	Q96A35	RM24	<i>MRPL24</i>	1,77	*				
39S ribosomal protein L48, mitochondrial	Q96GC5	RM48	<i>MRPL48</i>	2,16	*				
5'-nucleotidase	P21589	5NTD	<i>NT5E</i>	1,59	1,51	+	HPRD_00552	<i>NT5E</i>	4907
Prolyl 3-hydroxylase 1	Q32P28	P3H1	<i>P3H1</i>	2,14	*				
Polymeric immunoglobulin receptor	P01833	PIGR	<i>PIGR</i>	104,61	*	+	HPRD_01436	<i>PIGR</i>	5284
Pyruvate kinase PKM	P14618	KPYM	<i>PKM</i>	3,38	1,85	+	HPRD_01529	<i>PKM</i>	5315
Vitamin K-dependent protein S	P07225	PROS	<i>PROS1</i>	9,53	3,7	+	HPRD_01473	<i>PROS1</i>	5627
Protein PRRC1	Q96M27	PRRC1	<i>PRRC1</i>	4,19	2,2				
Receptor-type tyrosine-protein phosphatase C	P08575	PTPRC	<i>PTPRC</i>	14,94	8,3	+	HPRD_01050	<i>PTPRC</i>	5788
Protein S100-P	P25815	S100P	<i>S100P</i>	13,43	*	+	HPRD_02792	<i>S100P</i>	6286
Plasminogen activator inhibitor 1 RNA-binding protein	Q8NC51	PAIRB	<i>SERBP1</i>	18,9	*	+	HPRD_09570	<i>SERBP1</i>	26135
Heparin cofactor 2	P05546	HEP2	<i>SERPIND1</i>	20,19	*	+	HPRD_00795	<i>SERPIND1</i>	3053

Продолжение Приложения 4.

Название	Номер Uniprot	Белок	Ген	Повышенная регуляция, определенная методом 2DE, в ГЦР	Повышенная регуляция, определенная методом FASP, в ГЦР	Plasma Proteome Database(http://www.plasmaproteomedatabase.org/)			
						Общие 42	PPD ID	Ген	Идентификатор гена
Serine/arginine-rich splicing factor 5	Q13243	SRSF5	<i>SRSF5</i>	2,92	*	+	HPRD_02948	<i>SRSF5</i>	6430
Signal transducer and activator of transcription 1-alpha/beta	P42224	STAT1	<i>STAT1</i>	1,7	1,53	+	HPRD_02777	<i>STAT1</i>	6772
Very-long-chain enoyl-CoA reductase	Q9NZ01	TECR	<i>TECR</i>	1,94	1,8				
Transferrin receptor protein 1	P02786	TFR1	<i>TFRC</i>	77,86	5,73	+	HPRD_01812	<i>TFRC</i>	7037
Thrombospondin-1	P07996	TSP1	<i>THBS1</i>	11,11	4,86	+	HPRD_01765	<i>THBS1</i>	7057
Transmembrane emp24 domain-containing protein 10	P49755	TMEDA	<i>TMED10</i>	2,42	1,65	+	HPRD_12013	<i>TMED10</i>	10972
Transmembrane protein 70, mitochondrial	Q9BUB7	TMM70	<i>TMEM70</i>	10,09	*				
von Willebrand factor A domain-containing protein 1	Q6PCB0	VWA1	<i>VWA1</i>	39,24	*				
Tyrosine--tRNA ligase, mitochondrial	Q9Y2Z4	SYYM	<i>YARS2</i>	*	2,02	+	HPRD_13025	<i>YARS2</i>	51067
14-3-3 protein gamma	P61981	1433G	<i>YWHAG</i>	1,59	1,54	+	HPRD_05639	<i>YWHAG</i>	7532

ПРИЛОЖЕНИЕ 5. Тысяча типичных паттернов протеоформ, полученных для всех типов анализированных образцов (клеточная линия HepG2, нормальная ткань печени, опухолевая ткань печени, контрольная ткань печени)

Данные размещены в Yandex Disk и доступны по ссылке:

<https://disk.yandex.ru/d/4ihxaTXOXMcumA>

