

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ИНСТИТУТ БИОХИМИЧЕСКОЙ ФИЗИКИ ИМ. Н.М. ЭМАНУЭЛЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК

На правах рукописи

Федорченко Кристина Юрьевна

Метод неинвазивной диагностики рака легкого, основанный на анализе белкового и пептидного состава конденсата выдыхаемого воздуха человека

03.01.02 – биофизика

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата биологических наук

НАУЧНЫЙ РУКОВОДИТЕЛЬ:

член-корреспондент РАН, доктор химических наук,
профессор **С.Д. ВАРФОЛОМЕЕВ**

Москва 2017

ОГЛАВЛЕНИЕ

СПИСОК СОКРАЩЕНИЙ	5
ВВЕДЕНИЕ	7
ОБЗОР ЛИТЕРАТУРЫ	13
Глава 1. Обзор существующих подходов для диагностики состояния респираторной системы человека применительно к социально значимым заболеваниям. Проблемы ранней диагностики заболеваний органов дыхания	13
1.1 Взаимосвязь состояния организма человека и показателей респираторной системы человека и обзор существующих методов мониторинга этих показателей.....	14
1.2 Проблемы ранней диагностики социально значимых заболеваний органов дыхания..	20
Глава 2. Конденсат выдыхаемого воздуха человека: перспективы и проблемы для разработки метода ранней диагностики патологических изменений состояния организма	24
2.1 Состав конденсата выдыхаемого воздуха.....	25
2.2 Исследование газовой компоненты конденсата выдыхаемого воздуха и её диагностическая значимость.....	25
2.3 Исследование липидной компоненты конденсата выдыхаемого воздуха и её диагностическая значимость.....	26
2.4 Исследование белковой компоненты конденсата выдыхаемого воздуха и её диагностическая значимость.....	26
2.5 Протеом конденсата выдыхаемого воздуха в мультимодальных скрининговых программах.....	28
2.6 Методы сбора и определения белкового состава конденсата выдыхаемого воздуха.....	28
ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ	31
Глава 3. Материалы и методы	31
3.1 Материалы.....	31
3.2 Методы.....	31
3.2.1 Описание групп пациентов/доноров КВВ.....	32
3.2.2 Сбор и хранение проб КВВ.....	32
3.2.2.1 Система сбора R-Tube.....	32
3.2.2.2 Система сбора Eco-Screen.....	33
3.2.3 Пробоподготовка КВВ.....	34
3.2.3.1 Подготовка препарата КВВ для проведения хромато-масс-спектрометрического анализа (ВЭЖХ-МС/МС).....	34
3.2.3.2 Обезжиривание и осаждение белков КВВ.....	34
3.2.3.3 Осаждение белков с помощью дезоксихолата натрия и трихлоруксусной кислоты..	35
3.2.3.4 Концентрирование пептидов на колонке C18 (ZipTip C18).....	35

3.2.4 Хромато-масс-спектрометрический анализ (ВЭЖХ-МС/МС) проб КВВ.....	36
3.2.5 Электрофорез в полиакриламидном геле.....	37
3.2.6 Гидролиз белков в ПААГ.....	37
3.2.7 Получение масс-спектров белков, сконцентрированных в ходе электрофореза.....	37
3.2.8 Биоинформатический анализ результатов.....	38
3.2.8.1 Создание реляционной базы данных проб КВВ.....	38
3.2.8.2 Анализ масс-листов с помощью Mascot.....	38
3.2.8.3 Анализ белков, с наибольшей вероятностью обнаруживаемых в пробе.....	39
3.2.8.4 Сопоставление пептидной и белковой выдачи программы Mascot.....	39
3.2.8.5 Анализ пептидов, с наибольшей вероятностью обнаруживаемых в пробе.....	40
3.2.9 Построение прогностических моделей на основании статистического анализа пептидного состава проб КВВ.....	40
РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ.....	42
Глава 4. Особенности анализа масс-спектров белков КВВ.....	42
4.1 Принцип работы программы Mascot.....	42
4.2 Описание и выбор параметров программы Mascot в зависимости от эксперимента.....	44
4.2.1 Описание параметров программы Mascot.....	44
4.2.2 Выбор параметров программы Mascot.....	46
4.2.2.1 Параметры с неизменяемыми значениями.....	46
4.2.2.2 Параметры с изменяемыми значениями.....	47
Глава 5. Основные свойства КВВ и методика пробоподготовки КВВ для анализа белкового состава, основанные на исследовании 17 молодых здоровых некурящих доноров.....	49
5.1 Основные свойства КВВ	49
5.2 Подбор оптимальной методики концентрирования белковых смесей для масс-спектрометрического определения/измерения.....	50
5.2.1 Чувствительность ВЭЖХ-МС/МС анализа.....	51
5.2.2 Оценка концентрации белка в КВВ.....	51
5.2.3 Сравнительный анализ методик концентрирования при пробоподготовке КВВ для масс-спектрометрического анализа.....	53
5.2.3.1 Лиофилизация.....	53
5.2.3.2 Осаждение белков ТХУ и концентрирование на колонке С18.....	54
5.2.3.3 Обессоливание и обезжиривание смесью метанол/хлороформ.....	56
5.2.4 Ограничения ферментативного гидролиза трипсином.....	57
5.3 Выводы главы.....	58
Глава 6. Исследование белкового состава КВВ здоровых доноров. Анализ кератинового фона окружающего воздуха, а также КВВ, собранных с помощью защитного фильтра.....	59

6.1 Исследование белкового состава КВВ здоровых доноров.....	59
6.2 Интерпретация белкового состава КВВ здоровых доноров с помощью MetaCore.....	60
6.3 Анализ белкового состава объединенной пробы конденсатов выдыхаемого воздуха здоровых доноров.....	61
6.4 Анализ кератинового фона окружающего воздуха, а также КВВ, собранных с помощью защитного фильтра.....	64
6.5 Выводы главы.....	65
Глава 7. Исследование белкового состава КВВ пациентов с ХОБЛ, пневмонией и раком легкого.....	67
7.1 Результаты анализа белкового состава КВВ пациентов с такими основными диагнозами, как ХОБЛ и внебольничная пневмония.....	68
7.2 Результаты анализа белкового состава КВВ больных с диагнозом «рак легкого».....	78
7.3 Сравнительный анализ белкового состава КВВ больных с диагнозом «рак легкого» и белкового состава ткани, пораженной опухолевым процессом.....	84
7.4 Сравнительный анализ белкового состава КВВ больных с диагнозом «рак легкого» с уровнем экспрессии генов при данном заболевании.....	88
Глава 8. Новые подходы к разработке метода неинвазивной диагностики заболеваний легких на основе анализа состава КВВ	93
8.1 Создание базы данных проб и доноров КВВ.....	94
8.2 Анализ границ применимости программы Mascot при идентификации белкового и пептидного состава КВВ.....	95
8.3 Проблемы современного этапа развития метода и способы их решения.....	99
8.4 Метод классификации проб КВВ, основанный на пептидах.....	101
8.4.1 Формирование итоговых данных.....	102
8.4.2 Анализ масс-спектров промывочных заколов и проб на предмет загрязнения.....	102
8.4.3 Анализ идентифицированных пептидов в пробах КВВ, обработанных по предложенной в работе методике	105
8.5 Построение аналитических моделей, предсказывающих наличие фокусного заболевания у донора КВВ.....	112
8.6 Возврат на уровень белков при анализе пептидов по описанному методу.....	117
ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ.....	119
СПИСОК ЛИТЕРАТУРЫ.....	121
ПРИЛОЖЕНИЕ.....	142

СПИСОК СОКРАЩЕНИЙ

AUC – площадь под ROC-кривой, величина, часто используемая для оценивания качества упорядочивания алгоритмом объектов двух классов

CEA – Канцероэмбриональный антиген

CID – столкновительно-индуцированная фрагментация

CYFRA – опухолевый маркер немелкоклеточного рака легких и некоторых других раковых заболеваний (фрагмент Цитокератина 19)

ELISA – иммуноферментный анализ

ECoScreen – стационарное устройство для сбора конденсата выдыхаемого воздуха фирмы VIASYSHealthcare, Германия

FDA (USA) – Управление по санитарному надзору за качеством пищевых продуктов и медикаментов Агентства Министерства здравоохранения и социальных служб США

FiO₂ – фракционная концентрация кислорода во вдыхаемой газовой смеси

GO – GeneOnthology, биоинформатическая база данных

SDS-PAAG – электрофорез в полиакриламидном геле и денатурирующих условиях

HPLC – высокоэффективная жидкостная хроматография

MALDI-TOF – Времяпролетная масс-спектрометрия на основе матрично-активированной лазерной десорбции/ионизации

NSE – нейрон-специфическая енолаза

PaO₂ – парциальное давление кислорода в артериальной крови

RTube – портативное устройство для сбора конденсата выдыхаемого воздуха фирмы Respiratory Research, США

VOC – летучие органические вещества

БАЛ – бронхоальвеолярный лаваж

БСА – бычий сывороточный альбумин

ВЭЖХ-МС/МС – высокоэффективная жидкостная хроматография в сочетании с тандемной масс-спектрометрией

ИМ – индуцированная мокрота

ИЦР – ионно-циклотронный резонанс

КВВ – конденсат выдыхаемого воздуха

КТ – компьютерная томография

МС – масс-спектрометрия

МС/МС – тандемная масс-спектрометрия

ОРДС – острый респираторный дистресс-синдром

ПААГ – полиакриламидный гель

ПОЛ – перекисное окисление липидов

ПЦР – полимеразно-цепная реакция

ТХУ – трихлоруксусная кислота

УЗИ – ультразвуковое исследование

ХОБЛ – хроническая обструктивная болезнь легких

ВВЕДЕНИЕ

Актуальность работы

Дыхательная система выполняет важнейшую функцию жизнеобеспечения и отражает образ жизни человека и состояние его здоровья. Химические анализы дыхания имеют широкий спектр различных применений: от одобренного FDA (USA) измерения выдыхаемой фракции оксида азота для мониторинга эффективности противовоспалительной терапии при бронхиальной астме до определения летучих органических веществ и профилирования нелетучих биомаркеров в охлажденной дыхательной пробе, называемой конденсатом выдыхаемого воздуха (КВВ). [Silkoff et al., 2004; Phillips, 1997; Horvath et al., 2005; Buszewski et al., 2007] Будучи неинвазивной, а, следовательно, легкой в проведении процедурой, проба дыхания при этом позволяет клиницистам и исследователям оценивать различные процессы, происходящие в организме человека. Сбор такой пробы может быть осуществлен даже для очень тяжелых пациентов и повторен через короткие интервалы времени. Исходя из всего вышесказанного, считается, что исследование дыхания может быть идеальным кандидатом для скрининговых программ.

Кроме таких широко известных составляющих, как водород, кислород, углекислый газ, инертные газы и пары воды, выдох содержит также тысячи летучих и нелетучих компонентов, главным образом, в следовых количествах, что превращает их обнаружение в достаточно сложную задачу. Применение современных высокочувствительных технологий при анализе проб составляет основу правильного анализа этого типа биоматериала. Использование метаболомики, протеомики и масс-спектрометрии обладает огромным потенциалом в области профилирования биомаркеров выдыхаемого воздуха. [Czitrovsky et al., 2002] Биомаркеры выдыхаемого воздуха оцениваются для понимания патомеханизма заболевания, а также во вспомогательных целях при назначении соответствующей терапии [Horvath et al., 2009]. Тем не менее, необходимо заметить, что большинство исследователей (как следует из анализа опубликованных статей) не применяют протеомный подход, а концентрируются на поиске отдельных маркеров, что ограничивает возможности применения метода, особенно в случае комплексных заболеваний.

Хроническая обструктивная болезнь легких (ХОБЛ) и пневмония являются самыми распространенными респираторными заболеваниями и характеризуются тем, что ведущую роль в патогенезе данных заболеваний играет воспаление [Steel et al., 2013; Eltom et al., 2014]. Тем

не менее, работы по выявлению различий в протеомах КВВ здоровых и больных с диагнозом ХОБЛ и пневмония, а также оценка возможности использования КВВ для профилирования данных заболеваний на настоящий момент отсутствуют.

Рак легкого является одним из самых смертоносных видов рака и характеризуется самой высокой частотой и смертностью. [Ferlay et al., 2010; Jemal et al., 2008] Пятилетний процент выживаемости при раке легкого составляет всего лишь 15%. [Subramaniam et al., 2013] Рак легкого является одной из основных причин смерти в России, от которой умирает 27% онкологических больных [Чиссов и др., 2013]. Хирургическое лечение рака легкого доступно только для пациентов, диагностированных на ранней стадии. Плохой прогноз же связан с несколькими причинами:

– рак легкого – это сложный многоступенчатый процесс, и, хотя молекулярные стадии патогенеза успешно изучаются на уровне генов, транскрипционном уровне (мРНК) и уровне белков, канцерогенный механизм остается невыясненным [Petty et al., 2004].

– 50% пациентов, которым поставлен диагноз, уже имеют метастазы [Kumar et al., 2003], а в случае немелкоклеточного рака легкого метастазы при постановке диагноза имеются у более, чем 75% пациентов [Petty et al., 2004].

Причинами позднего диагностирования являются отсутствие ранних симптомов [Sone et al., 2007] или их неспецифичность (например, кашель или одышка, которые принимаются пациентами за последствия долгого курения или возрастные изменения) [Detterbeck et al., 2009], а также скрытое воспалением течение в тех случаях, когда рак легкого сопутствует пневмонии (параанкротная пневмония) или ХОБЛ.

Если диагноз поставлен тогда, когда рак еще локализован, и проводится хирургическое вмешательство, прогноз улучшается, а 5-летний уровень выживаемости достигает 52%. [Reed et al., 2004] Раннее диагностирование рака – условие успеха лечения и снижения смертности. Но, если для рака кишечника, груди, шейки матки скрининговые программы позволили уменьшить уровень смертности и улучшить прогноз [Cuzick, 1999], то для рака легкого такого рода программы пока не увенчались знаковым успехом [Gazdar et al., 1999].

В связи со всем вышесказанным, развитие нового потенциального метода ранней диагностики рака легкого, диагностики ХОБЛ и пневмонии является актуальной медицинской проблемой. Несмотря на очевидный потенциал использования протеомного подхода в анализе такой пробы дыхания, как КВВ, протеомные исследования этого объекта ограничены отдельными работами и до сих пор не имели системного характера.

Цель и задачи исследования

Целью данного исследования являлась разработка нового метода неинвазивной ранней диагностики рака легкого посредством анализа белкового и пептидного состава конденсата выдыхаемого воздуха человека с использованием масс-спектрометрии ультравысокого разрешения.

В **задачи** исследования входило:

- 1) Отработать оптимальную методику подготовки проб КВВ к масс-спектрометрическому анализу;
- 2) Определить белковый и пептидный состав проб КВВ для контрольной группы;
- 3) Определить белковый состав проб КВВ больных раком легкого, ХОБЛ, пневмонией и осуществить их системный анализ с учетом имеющихся данных по биохимии белков, анамнезу пациентов и клинической картины заболевания;
- 4) Определить пептидный состав конденсатов выдыхаемого воздуха больных раком легкого, ХОБЛ, пневмонией и осуществить их системный анализ с сопоставлением с полученными данными по белкам;
- 5) На основе анализа данных по пептидному составу КВВ предложить диагностическую модель для выявления рака легкого на фоне других респираторных заболеваний.

Научная новизна работы

Основной результат исследования – информация о сравнительном белковом и пептидном составе КВВ больных раком легкого и пациентов с ХОБЛ и пневмонией – является абсолютно новым. На основании полученных экспериментальных результатов показано, что анализ выдыхаемого воздуха – идеальный кандидат для скрининговых программ, открывающий, в сочетании с биоинформатическими подходами, новые возможности в области персонализированной медицинской диагностики. На основании проведенного исследования предложена панель белковых биомаркеров для диагностики рака легкого начальных стадий, а также создана аналитическая модель прогнозирования наличия рака легкого у донора на основе пептидного состава КВВ.

Теоретическая и практическая значимость работы

Полученные результаты могут послужить базой как для прикладных исследований – создания нового метода ранней диагностики онкологических заболеваний, – так и для фундаментальных исследований, связанных с изучением процессов развития патологических изменений на ранних стадиях онкологических заболеваний. Создаваемые в рамках проекта

подходы к проведению масштабных исследований КВВ в условиях клиники, а так же подходы к сбору, хранению, подготовке и анализу образцов будут абсолютно новыми и послужат базой для создания протоколов методов диагностики на стадии НИОКР.

Положения, выносимые на защиту

1. На основании исследования белкового состава КВВ пациентов с ХОБЛ, пневмонией и раком легкого, показано, что результаты анализа протеомов по группам различаются между собой и согласуются с клинической картиной рассматриваемых заболеваний.

2. На основании исследования белкового состава КВВ пациентов с диагностированным раком легкого 1-2 стадии выделены 19 белков, которые предложены в качестве диагностической панели для рака легкого.

3. На основании исследования пептидного состава КВВ всех исследуемых групп доноров, построена линейная аналитическая модель прогнозирования наличия у донора рака легкого и проверена с помощью группы доноров, не включенных в машинное обучение. Модель показала хорошую прогностическую способность ($AUC=0.99$), определив раковые образцы.

Личный вклад автора

Автор принимал активное участие в постановке задач исследования, самостоятельно проводил анализ литературных данных, участвовал в подборе методов исследования и сборе проб, осуществлял все этапы пробоподготовки, проведения экспериментов и обработки полученных результатов, а также подготовку материалов к публикациям. Измерения методом тандемной масс-спектрометрии производились автором при участии старшего научного сотрудника отдела масс-спектрометрии Института биохимической физики им. Н.М. Эмануэля РАН Кононихина А.С., биоинформатическая и статистическая обработка полученных результатов с использованием языков программирования осуществлялась при участии сотрудника лаборатории кинетики ферментативных реакций МБЦ МГУ имени М.В. Ломоносова Митрофанова С.И. Материалы диссертации в полном объеме доложены автором в устных докладах на ряде российских и международных конференций.

Достоверность полученных результатов

Достоверность экспериментальных результатов, полученных в работе, и обоснованность выводов обеспечивалась применением общепринятых физико-химических методов исследования. При проведении данной работы были использованы современные методы исследования белков и пептидов: одномерный и двумерный электрофорез в ПААГ, триптический гидролиз белков, тандемная хромато-масс-спектрометрия ультравысокого

разрешения, MALDI-TOF-масс-спектрометрия. Кроме того, в исследовании использовались биоинформатические методы работы с большими массивами данных, математические методы компьютерного анализа (корреляционный анализ, кластерный анализ, метод логистической регрессии). Достоверность результатов обеспечивалась инструментальной и статистической оценкой погрешности измерений, согласованием полученных результатов с литературными данными, а также согласованием данных, полученных различными методами исследования.

Апробация работы

Результаты работы были доложены на IV съезде Российского общества биохимиков и молекулярных биологов (Новосибирск, 2008 г.), III Троицкой конференции по медицинской физике и инновациям в медицине (Троицк, 2008 г.), Международной конференции From Promises to Practice. Applications of Science and Technology in Food, Healthcare, Energy and Environment (Греция, 2008 г.), Международном конгрессе Kongress der Deutschen gesellschaft fur pneumologie und beatmungsmedizin e.V. (Германия, 2011), Международном конгрессе Annual Congress European Respiratory Society (Нидерланды, 2011), Международном конгрессе 13th World Congress of the Human Proteome Organization, the 7th EuPA annual conference and the 6th Spanish Proteomics Society Congress (Испания, 2014), Международной конференции Biocatalysis-2015: Fundamentals and Applications (Россия, 2015), Международном конгрессе 14th World Congress of the Human Proteome Organization (Канада, 2015), XXV Национальном конгрессе по болезням органов дыхания (Россия, 2015), XI Международной научно-практической конференции "Пилотируемые полеты в космос" (Россия, 2015), Международном конгрессе Kongress der Deutschen gesellschaft fur pneumologie und beatmungsmedizin e.V. (Германия, 2016), V Съезде Физиологов СНГ/V Съезде Биохимиков России (Россия, 2016), Международном конгрессе Kongress der Deutschen gesellschaft fur pneumologie und beatmungsmedizin e.V. (Германия, 2017), Международном конгрессе Annual Congress European Respiratory Society (Италия, 2017), Международном конгрессе 16th World Congress of the Human Proteome Organization (Ирландия, 2017), XXVI Национальном конгрессе по болезням органов дыхания (Россия, 2017), Международной конференции по персонализированной онкологии (Россия, 2017). Работа также докладывалась в рамках семинаров ИБХФ РАН, семинара МБЦ МГУ и семинара в ФГБУ Научно-исследовательский институт пульмонологии ФМБА России на базе 57 городской клинической больницы.

Публикации

По материалам диссертации опубликовано 25 печатных работ, из них 8 – в рецензируемых научных российских и иностранных журналах по списку ВАК, 2 – главы в монографиях, 15 – в тезисах конференций.

Структура и объем диссертации

Диссертация написана в классической форме и содержит следующие разделы: введение; обзор литературы; материалы и методы; результаты, полученные в работе, и их обсуждение; выводы; список цитируемой литературы; приложение.

Диссертация содержит 141 страницу, 35 рисунков, 21 таблицу, 238 литературных ссылок.

ОБЗОР ЛИТЕРАТУРЫ

Глава 1. Обзор существующих подходов для диагностики состояния респираторной системы человека применительно к социально значимым заболеваниям. Проблемы ранней диагностики заболеваний органов дыхания

Большинство заболеваний легче предотвратить, чем лечить, а лечение на ранних стадиях обычно более эффективное и менее затратное по сравнению с запущенными стадиями. Профилактическая направленность медицины реализуется сегодня в разработке и внедрении программ скрининга социально значимых заболеваний. [Тюрин, 2011]

Эффективность лечения самых различных заболеваний напрямую зависит от верности и своевременности постановки диагноза, кроме того, в ходе проведения лечения возникает необходимость мониторинга ответных изменений состояния организма для внесения корректировок в назначенный курс терапии. Таким образом, можно сформулировать следующие основные задачи, которые в идеале должны решать методы диагностики патологических изменений состояния организма (в том числе методы ранней диагностики):

- Возможность постановки диагноза, как при наличии, так и в отсутствие симптомов у пациента;
- Получение информации о прогнозе течения болезни у пациентов с установленным диагнозом;
- Мониторинг эффективности проводимой терапии для конкретного пациента;
- Низкий процент ложноположительных результатов при проведении диагностики.

Скрининг (от англ. screening – массовое обследование) – комплекс мероприятий в системе здравоохранения, проводимых с целью выявления и предупреждения развития различных заболеваний у населения [Минздрав РФ №36ан, 2015]. Целью скрининга является раннее выявление заболеваний для обеспечения раннего начала лечения, что, в свою очередь, может иметь решающее значение для облегчения состояния пациентов и снижения смертности. Для обеспечения выполнения основной цели скрининговых исследований и во избежание как ложноположительных, так и ложноотрицательных ошибок раннего диагностирования, методы ранней диагностики должны обладать достаточной чувствительностью и специфичностью. Хотя скрининг и ранняя диагностика связаны с целым рядом недостатков, таких как денежные затраты на такого рода программы, случаи гипердиагностики и, напротив, случаи не выявления

заболевания по причине недостаточной точности используемых методов, в целом, массовые скрининговые обследования обеспечивают повышение уровня здоровья населения и неизменно обращают на себя внимание со стороны Всемирной организации здравоохранения (ВОЗ).

1.1 Взаимосвязь состояния организма человека и показателей респираторной системы человека и обзор существующих методов мониторинга этих показателей

Дыхательная система выполняет важнейшую функцию жизнеобеспечения и отражает образ жизни человека и состояние его здоровья. Установлено, что ранние, прогностически значимые признаки неблагоприятных средовых влияний проявляются, прежде всего, в изменении отношений между сердечно-сосудистой и респираторной системами [Глазачев, 1997]. Так, например, состояние респираторной системы при ожирении является одним из важных компонентов в плане определения и прогнозирования соматического здоровья, физического статуса, а также возможных ограничений тех или иных видов двигательной активности [Пшеннова и др., 2012]. Изменения функциональных показателей респираторной системы нередко проявляются раньше клинических симптомов и могут дать не только информацию о состоянии самой респираторной системы, но и более полное представление о состоянии человека в данный момент [Пшеннова и др., 2012].

Ниже перечислены основные методы диагностики состояния респираторной системы человека:

Пульсоксиметрия. Пульсоксиметрия широко используется в анестезиологии и интенсивной терапии пациентов. Среди внутренних ограничений пульсоксиметрии можно выделить то, что она не чувствительна к изменениям артериального парциального давления кислорода (PaO_2) на высоких уровнях и не позволяет различить нормальный гемоглобин и метгемоглобин или карбоксигемоглобин. На результаты измерения могут оказывать влияние такие факторы, как лак для ногтей [Hinkelbein et al., 2007], темный цвет кожи пациента [Feiner et al., 2007], измененная перфузия кожи, карбоксигемоглобин. Показания пульсоксиметрии следует использовать для обеспечения раннего предупреждения, снижая потребность измерения газов крови. В рандомизированном контролируемом исследовании в более чем 20000 хирургических больных [Moller et al., 1993], использование пульсоксиметрии не было связано с уменьшением послеоперационных осложнений и смертности, но 80% анестезиологов чувствовал себя более уверенно при использовании пульсоксиметра.

Объемная капнография. Капнограмма выдоха обеспечивает качественную информацию о волновой форме кривой, связанной с механической вентиляцией и количественной оценкой выдыхаемого CO_2 . Было показано, что вид графика, построенного на основании данных капнографии коррелирует с тяжестью обструкции дыхательных путей. [You et al., 1994] Данные о концентрации углекислого газа в выдыхаемом воздухе, позволяют врачу в реальном времени получать ценную информацию об адекватности вентиляции легких, а также ряд других важных данных, касающихся состояния кровообращения и метаболизма у больного. [Ершов, 2013] Объемная капнография особенно необходима тогда, когда изменения в концентрации CO_2 в выдохе следует избегать (в критическом состоянии, у неврологических больных с нормальными легкими). Также метод в сочетании с анализом на D-димер используется как высокочувствительный скрининг, чтобы исключить диагноз тромбоэмболии легочной артерии [Kline et al., 2001]. Было показано, что объемная капнография также может быть отличным инструментом для мониторинга тромболитической эффективности у больных с легочной эмболией [Verschuren et al., 2004]. Широкое применение объемной капнографии ограничивается необходимостью наличия сложного и дорогого оборудования.

Газы крови. Наиболее часто для оценки тяжести легочной недостаточности и в составе комплекса мер по определению тяжести повреждения легких используется индекс оксигенации $\text{PaO}_2 / \text{FiO}_2$ (где PaO_2 – парциальное давление кислорода в артериальной крови, FiO_2 – парциальное давление кислорода во вдыхаемом воздухе) [Bernard et al., 1994]

Экстравазкулярные воды в легких. Измерение экстравазкулярных вод в легких является количественной мерой отека легких и коррелирует со смертностью [Sakka et al., 2002]. Метод может использоваться для больных в критическом состоянии, позволяет измерять производительность сердечно-сосудистой системы. Измерение экстравазкулярных вод в легких может быть использовано в комбинации с другими сердечно-сосудистыми и легочными параметрами для диагностики отека легких. [Monnet et al., 2007]

Анализ механики дыхания. Измерения дыхательной механики просты для выполнения и предоставляют полезную и актуальную информацию для оценки степени тяжести заболевания и управления искусственной вентиляцией легких. Они действительно надежны только в пассивных условиях вентиляции. [Brochard et al., 2012].

Работа дыхания представляет собой площадь под кривой (интеграл) в координатах (объем легких)/(давление). [Cabello et al., 2006]. Мониторинг работы дыхания помогает в определении времени поддержания искусственной вентиляции легких. Он может быть использован для оценки влияния различных режимов вентиляции, понимания механизмов

заболевания (обострение астмы, обострение ХОБЛ), оценки влияния терапевтического вмешательства (например, бронходилататора [Mancebo et al., 1991]) и искусственной вентиляции легких. Данный метод используется, в основном, для клинических исследований.

Измерение объема легких (спирометрия). В условиях физиологического покоя достаточно надежным методом массовых исследований основных дыхательных объемов человека является стандартная (нефорсированная) спирометрия [Исупов, 2015]. Показатели, полученные с помощью указанного метода, позволяют судить не только о состоятельности респираторной системы, но и, отчасти, об общем физическом развитии конкретного лица, косвенно, весьма приблизительно оценивать его физическую тренированность [Агаджанян, 1986; Исупов и др., 2013; Исупов и др., 2014]. Появление и развитие различных модификаций форсированной («скоростной») спирометрии, существенно расширив диагностические возможности функционально-диагностических методик исследований легочных объемов, позволило осуществлять количественную оценку трахеобронхиальной проходимости, общего аэродинамического сопротивления воздухоносных путей с высокой точностью. Форсированная спирометрия (наряду с пневмотахометрией) позволяет более полно и объективно, нежели обычная спирометрия, исследовать общее физическое состояние человека на основе оценки скоростно-силовых качеств ин- и экспираторной мускулатуры [Исупов и др., 2013; Исупов, 2015].

С помощью спирометрии измеряют величины обычной жизненной емкости легких (ЖЕЛ) и экспираторной форсированной ЖЕЛ (ЭФЖЕЛ, л), секундной фракции ЭФЖЕЛ (ЭФЖЕЛ 1, л), двухсекундной фракции ЖЕЛ (ЭФЖЕЛ2), дыхательного объема (ДО, мл), резервного объема выдоха (РОВыд, л) и его форсированного варианта (ФРОВыд, л).

Прямое измерение объема легких в конце выдоха доступно только для пассивных пациентов (механическая вентиляция) и применяется в исследовательских целях при ведении пациентов с острым респираторным дистресс-синдромом [Brochard et al., 2012].

Ультразвуковое исследование (УЗИ) грудной клетки. В настоящее время УЗИ грудной клетки применяется для выявления плеврального выпота и мало используется в диагностике заболеваний легких из-за распространенного мнения о малой информативности метода, поскольку воздушная легочная ткань не проводит ультразвук. [Сафонов и др., 2014] Однако при потере воздушности он проникает в толщу легкого на всю глубину патологических изменений [Reiβig et al., 2009]. Проведенные исследования показали, что ультразвуковая визуализация при воспалительных заболеваниях легких по точности превышает рентгенографию грудной клетки [Kroegel et al., 2000]. УЗИ грудной клетки является

информативным радиологически безопасным методом динамического контроля за обратным развитием пневмонии, позволяющим своевременно оценить эффективность этиотропного лечения. [Сафонов и др., 2014] Ультразвуковое исследование грудной клетки может быть полезным в случае пассивных пациентов для раннего выявления отека а также других нарушений, таких как пневмоторакс или плеврит [Remerand et al., 2010; Reissig et al., 2011]. Также ультразвуковое исследование грудной клетки может применяться для диагностики при опухолях, крупных образованиях плевры и средостения, но только в сочетании с другими методами, такими как рентгенография, компьютерная и магнитно-резонансная томография.

Компьютерная томография (КТ) грудной клетки. В течение последних 20 лет рентгеновская КТ стала одним из важнейших методов диагностики заболеваний органов дыхания. Это обусловлено высокой точностью метода в выявлении патологических изменений органов и тканей грудной полости. [Тюрин, 2003] КТ используется для диагностики полостных образований в легких [Яковлев и др., 2012], которые часто выявляются при воспалительных заболеваниях (острый абсцесс легкого при пневмониях; воздействия патогенных микроорганизмов при туберкулезе легких, злокачественных опухолях, кистах, эхинококкозе, синдроме Вегенера) [Шейх и др., 2011]. КТ изображения могут также быть использованы для вычисления средней плотности легких и количественной оценки соответствующих количеств воздуха и ткани, но этот подход в настоящее время ограничен для исследования [Pelosi et al., 2001, Puybasset et al., 1998]. КТ грудной клетки предложили использовать как метод ранней диагностики рака легкого [Henschke et al., 2001]. Однако вопрос о принципиальной целесообразности скрининга рака легкого с помощью лучевых методов исследования до настоящего времени остается предметом дискуссии. Низкодозовая КТ показала обнадеживающие результаты [Henschke et al., 1999], но среди них было большое количество ложноположительных [Swensen et al., 2003], кроме того, этот метод является дорогостоящим [Amann et al., 2011] и имеет случаи негативного воздействия облучения на пациентов [Amann et al., 2011; Welch et al., 2007]. Не ясно, смогут ли эти программы реально снизить смертность больных раком легкого, а также окажутся ли такие дорогостоящие программы экономически оправданными [Тюрин, 2003].

Исследование индуцированной мокроты, бронхоальвеолярного лаважа, бронхобиопсий. Традиционными методами изучения степени воспаления в трахеобронхиальной системе являются оценка обычной и индуцированной мокроты, бронхоальвеолярного лаважа (БАЛ), бронхобиопсий [Lehmann et al., 2003; Van Beurden et al., 2003].

БАЛ может быть использован для оценки кровоизлияний и измерения нейтрофилов, эозинофилов, гиалиновых мембран, липидных включений и раковых клеток (хотя это требует тщательного цитологического исследования отобранной альвеолярной жидкости), что способствует правильной диагностике ряда заболеваний и позволяет оценить активность воспалительного процесса в бронхоальвеолярном пространстве. Более того, например, при альвеолярном протеинозе БАЛ является не только диагностической, но и терапевтической процедурой. [Самсонова, 2006] Многие исследования показали более высокие уровни воспалительных медиаторов (цитокинов и др.) и фиброцитов в БАЛ умерших пациентов с острым респираторным дистресс-синдромом по сравнению с оставшимися в живых [Quesnel et al., 2012]. Анализ жидкости БАЛ может также помочь выявить пациентов, которые могут реагировать на стероиды. Среди ограничений метода можно отметить необходимость эндоскопического оборудования и навыков работы с ним, отсутствие стандартизированной техники. Потенциальным осложнением является гипоксия. Можно заключить, что детальный анализ БАЛ подходит для бактериологических целей, но подробное цитологическое исследование или оценка биомаркеров часто остаются в рамках клинических испытаний [Brochard et al., 2012].

Индуцированная мокрота (ИМ) – субстрат, отхаркиваемый пациентом после ингаляции 3-5 % раствора хлорида натрия. Биоматериал можно получить в этих условиях в 75-100 % случаев даже у здоровых людей [Дулина, 2009; Profita et al., 2010; Wang et al., 1997]. Исследование клеточного состава ИМ может служить способом оценки типа и интенсивности воспалительного процесса при острых и хронических заболеваниях легких, в том числе и на их доклинических стадиях. Так, в клеточном составе ИМ курильщиков в сравнении с некурящими увеличено количество нейтрофилов и лимфоцитов [Brightling, 2006; Kunz et al., 2011]. Различия в цитограммах пациентов с бронхиальной астмой в фазе обострения и пациентов с ХОБЛ позволили использовать метод ИМ для дифференциальной диагностики данных заболеваний. Также метод может использоваться как первичная диагностическая процедура при подозрении на рак легкого [Чучалин и др., 1998; Добрых и др., 2013].

Исследование бронхобиоптатов играет важную роль в диагностике всех видов доброкачественных и злокачественных опухолей легких, также диагностическая значимость бронхобиопсий достаточно высока при ряде неспецифических воспалительных заболеваний и меньше – при саркоидозе и туберкулезе легких. В последние годы исследования бронхобиоптатов позволили описать состояние стенки бронхов у больных бронхиальной астмой в период ремиссии. [Черняев и др., 2011]

Несмотря на диагностическую значимость описанных выше методов и широкое применение их в клинической практике, следует однако отметить, что получение БАЛ и бронхобиоптатов являются инвазивными травматичными методами, вследствие чего не могут проводиться повторно через короткий промежуток времени, а также у тяжелых больных, детей и в амбулаторной практике. Индуцирование мокроты ингаляциями гипертонического раствора хлорида натрия также обладает рядом недостатков, поскольку может влиять на состав получаемых образцов и вызывать бронхоконстрикцию [Balint et al., 2001; Paredi et al., 2002].

Рентгенологическое исследование грудной клетки (флюорография). Проводится для диагностики и мониторинга различных заболеваний респираторной системы. Рентгенография грудной клетки считается самым распространенным диагностическим методом, который традиционно используется в повседневной клинической практике для диагностики пневмонии, особенно в критических условиях [Rubinowitz et al., 2007; Bowen et al., 2013]. Кроме пневмонии, рентгенологическое исследование используется для первичной диагностики обструктивных болезней легких, в том числе ХОБЛ, хотя в данном случае метод имеет ограниченное применение и имеет цель исключить другие патологические состояния, которые могут иметь сходные клинические проявления или сочетаться с ХОБЛ. К ним, прежде всего, относятся туберкулез и рак легкого. При обострении ХОБЛ лучевое исследование проводится для исключения пневмонии или абсцесса, а также венозного застоя и отека легких при левожелудочковой недостаточности [Тюрин, 2004]. Рентгенологическое исследование при туберкулезе дает возможность избежать неправильных оценок данных перкуссии и аускультации, с его помощью уточняется характер морфологических изменений, размер, фаза и динамика в процессе наблюдения и лечения [Калмыкова, 2009]. Основой обследования больных с гнойными заболеваниями легких и плевры являются методы лучевой диагностики. Многоосевая рентгеноскопия и рентгенография в стандартных проекциях выполняются для уточнения локализации абсцессов, дифференциальной диагностики с осумкованными эмпиемой плевры и пиопневмотораксом. [Петухов и др., 2010]. Также рентгенография грудной клетки выполняется при заболеваниях поджелудочной железы, например, при хроническом панкреатите. Рентгенологические признаки заболеваний поджелудочной железы зависят от локализации процесса, объема и фазы заболевания. На рентгенограммах грудной клетки при хроническом панкреатите следует исключить жидкость в левой плевральной полости, дисковидные ателектазы над левым куполом диафрагмы, ограничение или полное отсутствие ее подвижности, усиление легочного рисунка или очаговые воспалительные изменения в базальных отделах левого легкого. Эти изменения указывают на осложненное течение

панкреатита и требуют дифференциальной диагностики с другими заболеваниями. [Сиваш, 2008]

Рентгенологическое исследование грудной клетки может использоваться для первичной диагностики рака легкого [Лещук и др., 2010], однако, эффективность этого метода подверглась сомнению ввиду доказанной невозможности определения рака легкого на ранних стадиях и высокой смертности пациентов. В настоящее время данный метод не рекомендован для диагностики рака легкого. [Manser et al., 2013; Richard et al., 2017]

1.2 Проблемы ранней диагностики социально значимых заболеваний органов дыхания

ХОБЛ. Хроническая обструктивная болезнь легких (ХОБЛ) и астма являются хроническими воспалительными заболеваниями дыхательных путей, ставшими одними из основных проблем общественного здравоохранения [Abramson et al., 2014]. ХОБЛ в настоящее время является одной из ведущих причин заболеваемости и смертности в мире [Zong et al., 2015] и, по данным последнего отчета Всемирной организации здравоохранения, находится на 4-м месте в списке 10 ведущих причин смерти (рисунок 1). Улучшение методов ранней диагностики пациентов с ХОБЛ, которым еще может помочь немедикаментозное лечение, имеет первостепенную важность [Paone et al., 2008]. К сожалению, ранние симптомы ХОБЛ трудноразличимы и часто игнорируются пациентами, по этой причине очень часто диагноз ХОБЛ ставится на поздних стадиях болезни, когда пациенты испытывают существенное ухудшение качества жизни [Paone et al., 2014]. Спирометрия на сегодняшний день является зачастую единственным диагностическим методом, используемый для оценки статуса заболевания. Легочные функциональные тесты отражают тяжесть заболевания, но не его активность, результаты измерения строго зависят от правильного выполнения и интерпретации данных. Кроме того, клинические симптомы, на которые пульмонологи полагаются при выборе правильного диагностического подхода, являются субъективными и неспецифическими. Ограничения спирометрии и клинической истории заставляют искать новые показатели заболевания. [Paone et al., 2016]

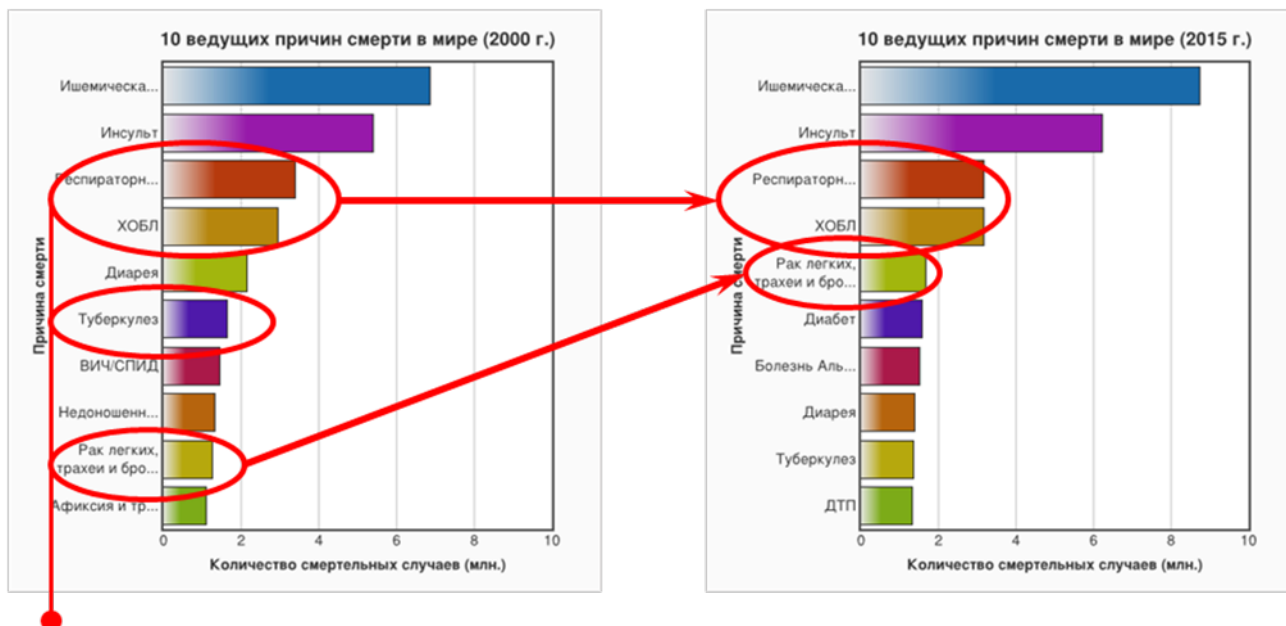


Рисунок 1. Ведущие причины смерти по данным ВОЗ.

Рак легкого. Рак легкого является одним из самых смертоносных видов рака и характеризуется наибольшим уровнем заболеваемости среди них [Ferlay et al., 2010; Jemal et al., 2008]. С течением заболевания выживаемость уменьшается, но для I-II стадий возможно излечение с удовлетворительными отдаленными результатами: пятилетняя выживаемость составляет примерно 60 и 46% соответственно [IARC, 2012]. Несмотря на развитие медицины, заболевание в большинстве случаев по-прежнему диагностируется на поздних стадиях развития из-за отсутствия эффективных методов ранней диагностики и неполного представления о процессах онкогенеза. 50% пациентов, которым поставлен диагноз «рак легкого», уже имеют метастазы [Kumar et al., 2003], а в случае немелкоклеточного рака легкого метастазы при постановке диагноза имеются более чем у 75% пациентов [Petty et al., 2004]. Раннее диагностирование - залог успешного лечения и снижения смертности, однако, при данном заболевании ранние симптомы могут отсутствовать [Sone et al., 2007] или быть неспецифичными (кашель, одышка), что позволяет принять их за последствия курения, простуду или сопутствующие заболевания.

На сегодняшний день среди методов диагностики рака легкого можно назвать следующие: рентгенологическое исследование, исследование индуцированной мокроты, низкодозовая компьютерная томография, флуоресцентная бронхоскопия, эндобронхиальная аспирация ультразвуковой трансbronхиальной иглой, исследование биомаркеров крови. Однако, несмотря на такое разнообразие, проблема скрининга населения с целью выявления рака легкого на ранних стадиях до сих пор не решена.

КТ для диагностики рака легкого. Вопрос о целесообразности использования КТ в качестве скринингового метода обсуждался в предыдущих разделах и он остается открытым [Shieh et al., 2017].

Бронхоскопия и эндобронхиальная аспирация ультразвуковой трансbronхиальной иглой являются инвазивными и потому не могут быть эффективно использованы для скрининга большого количества людей.

Рентгенография грудной клетки и исследование индуцированной мокроты в качестве методов ранней диагностики рака легкого результатов не дали [Coleman et al., 1999; Horva et al., 2009; Aberle et al., 2008]. Более того, данные методы в настоящее время не рекомендованы для диагностики рака легкого ввиду доказанной неэффективности. [Manser et al., 2013; Richard et al., 2017; Fontana et al., 1991]

Белковые биомаркеры крови для диагностики рака легкого. Доступность крови делает ее хорошим образцом для протеомных исследований. Тем не менее, метод оценки биомаркеров крови инвазивен, кроме того требует дальнейшей стандартизации и накопления и обработки статистических данных [Мирошниченко и др., 2009]. При анализе предполагается, что либо непосредственно опухоль, либо ответ организма на воспаление приводит к изменению белкового состава. Кровь циркулирует по всему телу, поэтому невозможно установить, из какого источника белки попали в пробу. Выделить из такой сложной смеси конкретные биомаркеры трудно. Некоторые белки находятся в крови в большом количестве, а концентрация маркеров часто изменяется незначительно. [Carter et al., 2012] Негосударственными медицинскими компаниями в России предлагаются около 20 видов анализов на онкомаркеры, при этом только 6 из них (CEA, Ca-125, Ca 72-4, Cyfra-21-1, NSE, SCCA) могут быть использованы при диагностике рака легкого, но не рекомендуются самими компаниями в качестве скрининговых методов по причине недостаточной чувствительности, специфичности и неоднозначности интерпретации результатов. По данным Гамбургской Группы по Стандартизации Опухолевых Маркеров чувствительность тестов с использованием карциноэмбрионального антигена (CEA), фрагментов цитокератина 19 (CYFRA) и нейронспецифической энолазы (NSE) составляла всего 58%, 66%, и 59%, соответственно [Maeda et al., 1996; Stieber et al., 1999].

Еще одной сложностью при диагностике рака легкого являются случаи наличия в анамнезе пациента других заболеваний (например, ХОБЛ) и случаи ассоциации рака легкого с вторичными параканкрозными изменениями, чья частота составляет до 30%. Поскольку в таких случаях проявляются клинические симптомы пневмонии или эмфиземы, большинство

амбулаторных пациентов с осложненным раком оказываются в общих палатах. Воспалительная паренхиматозная патология может способствовать ложноположительной или ложноотрицательной диагностике даже при использовании высокоточных лучевых методов [Bisenkov et al., 2006].

Таким образом, недостаток доступных неинвазивных методов исследования приводит к запоздалой диагностике, несвоевременному лечению и развитию осложнений у пациентов с заболеваниями органов дыхания.

Глава 2. Конденсат выдыхаемого воздуха человека: перспективы и проблемы для разработки метода ранней диагностики патологических изменений состояния организма

Выдох человека представляет собой аэрозоль, содержащий, помимо смеси газов, молекулы летучих органических соединений, а также микрокапли воды с растворенными в них нелетучими органическими соединениями. Летучие и нелетучие органические соединения являются продуктами метаболической активности в организме и, следовательно, отражают текущее состояние клеток, тканей и микробиоты. Состав соединений в выдохе является значительным источником информации о здоровье. В период с 1997 по 2017 год в научной литературе, по данным базы научных публикаций PubMedCentral, упоминается около 4,5 тысяч клинических исследований, связанных с анализом выдыхаемого воздуха. Из них примерно 1,5 тысячи посвящены диагностике рака. Метаболизмы раковых и нормальных клеток существенно различаются. В данный момент обнаруживается все больше механизмов, посредством которых раковые клетки перестраивают метаболические пути для обеспечения питания. Одним из наиболее хорошо описанных метаболических изменений является «эффект Варбурга» [Otto, 2016]. Этот эффект заключается в том, что в раковых клетках в 200 раз повышается анаэробный гликолиз даже при достаточном содержании кислорода. Такое изменение является ключевым для выживания опухолевых клеток, поскольку обеспечивает поступление структурных компонентов, в которых нуждается растущая опухоль. Эффект Варбурга дает основной вклад в характерный для рака профиль летучих органических соединений и влияет на профиль летучих органических соединений в выдыхаемом воздухе, что может использоваться при диагностике рака, в частности, рака легкого.

В последнее время несколько групп исследователей проводят работы по изучению потенциала КВВ как альтернативы традиционным методам исследования состояния легких [Silkoff et al., 2004; Phillips, 1997; Horvath et al., 2005; Buszewski et al., 2007; Kurova et al., 2009]. С целью улучшения ведения пациентов с раком легкого и повышения выживаемости, существует необходимость в разработке стратегий обследования и ранней диагностики, которые были бы чувствительными, специфичными и неинвазивными. С точки зрения диагностики и прогноза по заболеванию, КВВ представляется привлекательным методом, который имеет ряд преимуществ. Это простой, безопасный, недорогой и неинвазивный метод отбора проб нижних дыхательных путей у людей, с помощью которого осуществляется забор образцов непосредственно из местного окружения легких [Conrad et al., 2008]. Кроме того,

очень важно, что сбор КВВ не вторгается в процесс основного заболевания и может повторяться в течение короткого периода времени, даже у больных астмой и у пациентов после трансплантации легких [Kurova et al., 2010].

2.1 Состав конденсата выдыхаемого воздуха

КВВ представляет собой охлажденную дыхательную пробу. Происхождение молекул, определяемых в КВВ, различно. Они могут быть первоначально локализованы в ротовой полости и глотке, в трахеобронхиальной системе и в альвеолах. Считается, что основными нелетучими компонентами КВВ являются метаболические продукты клеток, выстилающих воздухопроводящие пути [Ferlay et al., 2010]. Биологические молекулы попадают в пробу из бронхиального и альвеолярного аэрозолей, формирующихся при испарении жидкости, прилежащей к эпителиальной выстилке. Кроме таких широко известных составляющих, как водород, кислород, углекислый газ, инертные газы и пары воды, выдох содержит также тысячи летучих и нелетучих компонентов, главным образом, в следовых количествах, что превращает их обнаружение в достаточно сложную задачу. Применение современных высокочувствительных технологий при анализе проб составляет основу правильного анализа этого типа биоматериала. Использование метаболомики, протеомики и масс-спектрометрии обладает огромным потенциалом в области профилирования биомаркеров выдыхаемого воздуха. [Czitrovszky et al., 2002]

2.2 Исследование газовой компоненты конденсата выдыхаемого воздуха и её диагностическая значимость

Оксид азота NO и аммиак являются наиболее изучаемыми маркерами в выдыхаемом воздухе, описаны их изменения при различных заболеваниях. Определение аммиака в КВВ может использоваться для дифференциальной диагностики вирусных и бактериальных заболеваний легких [Zetterquist et al., 2009]. Определение уровней NO и аммиака также используют для диагностики муковисцидоза и контроля проводимой терапии [Robroeks et al., 2008; Newport et al., 2009], диагностики астмы, тяжести и контроля развития заболевания [Ratnawati et al., 2006; Robroeks et al., 2007; Aldakheel et al., 2016].

Перекись водорода образуется из радикалов кислорода, выделяющихся активированными клетками воспаления: нейтрофилами, макрофагами и эозинофилами. Она

обнаруживается в КВВ и является потенциальным маркером окислительного стресса в легких. Уровень H_2O_2 используется в диагностике астмы [Teng et al., 2011]. Количество H_2O_2 в КВВ пациентов с ХОБЛ также значимо выше, чем в КВВ здоровых доноров, что было показано в нескольких исследованиях [Dekhuijzen et al., 1996; Kostikas et al., 2003; Nowak et al., 1999]. Lases и соавторы [Lases et al., 2000] обнаружили повышенный уровень H_2O_2 в КВВ пациентов с раком легкого, которым была сделана операция по резекции легкого.

2.3 Исследование липидной компоненты конденсата выдыхаемого воздуха и её диагностическая значимость

Все эйкозаноиды образуются из общего предшественника – арахидоновой кислоты, относящейся к полиненасыщенным жирным кислотам. К эйкозаноидам принадлежат четыре основных группы соединений: простагландины, простаглицлины, тромбоксаны и лейкотриены. Эти вещества обладают широким спектром биологической активности и являются потенциальными медиаторами воспаления, ответственными, в том числе, за сужение и расширение бронхов. Определение продуктов перекисного окисления липидов, также как и измерение перекиси водорода и азотосодержащих соединений, может дать важную информацию об окислительном статусе дыхательной системы.

Было показано, что уровень эйкозаноидов, в частности, лейкотриенов повышен у доноров с астмой [Montuschi et al., 2002; Ono et al., 2008; Kielbasa et al., 2008], ХОБЛ [Kostikas et al., 2005] и немелкоклеточным раком легкого [Carpagnano et al., 2011], в сравнении со здоровым контролем.

2.4 Исследование белковой компоненты конденсата выдыхаемого воздуха и её диагностическая значимость

Человек выдыхает белки экзогенного и эндогенного происхождения. Наличие первых объясняется их попаданием в дыхательную систему во время вдоха, а ко вторым относятся белки эпителиальных клеток дыхательных путей, а также продукты ферментативного расщепления белков [Фёдоров, 2013]. При развитии патологических процессов в дыхательных путях характерно изменение белкового состава их тканей, которое может иметь диагностическое значение; при этом белки и пептиды выдыхаемого воздуха выступают как специфические маркеры заболевания. С учетом особенностей рака легкого, анализ белкового и

пептидного состава КВВ имеет большой потенциал именно для диагностики данного заболевания.

В исследовании Bloemen и др. [Bloemen et al., 2011] в КВВ детей с астмой отмечен повышенный уровень цитокератина 1, в сравнении со здоровой группой, хотя никаких различий по другим белкам (цитокератины 2, 5, 6, 8, 9, 10, 14, 16, альбумин, актин, лизоцим, калгранулин В, десмин) найдено не было. В исследовании Zietkowski и др. [Zietkowski et al., 2009], в КВВ пациентов с астмой отмечен повышенный уровень с-реактивного белка. Протеомный подход к исследованию КВВ был продемонстрирован Fumagalli и др. [Fumagalli et al., 2012] при сравнении пациентов с ХОБЛ, эмфиземой легкого, а также курильщиков и группы здорового некурящего контроля. Согласно опубликованным данным, цитокератины 1, 5, 9, 14 и 26 преобладали в КВВ как здоровых курильщиков, так некурящих доноров, в группе ХОБЛ преобладающей белковой составляющей КВВ являлись цитокины (IL-1 α , IL-1 β , IL-2, IL-12 α , IL-12 β , IL-15, фактор некроза опухоли α), а также были обнаружены интерфероны α и γ и С3-компонент комплемента. При этом в КВВ группы с ХОБЛ отсутствовала значимая часть ферментов (8%), идентифицированных в КВВ группы здорового контроля. Также в КВВ группы курильщиков была обнаружена β -цепь гемоглобина, что, по мнению авторов, отражало гемолиз, вызванный связанным с курением окислительным стрессом. Различия в протеомах были также обнаружены между группами с ХОБЛ и эмфиземой легкого. Kullmann и соавторы [Kullmann et al., 2008] привели данные, что профиль цитокинов в КВВ пациентов с раком легкого отличался от КВВ здоровых доноров. Повышенный уровень цитокинов, в частности IL-6, IL-2 и фактора некроза опухоли α , а также пептидного гормона лептина и сосудосуживающего пептида эндотелина-1 в КВВ пациентов с немелкоклеточным раком легкого, в сравнении со здоровым контролем, показан в исследованиях Carpagnano и др. [Carpagnano et al., 2002; Carpagnano et al., 2007; Carpagnano et al., 2004] В исследовании Brussino et al. [Brussino et al., 2014] показана корреляция повышенного уровня цитокинов (IL-6, IL-17, фактор некроза опухоли α) в КВВ с повышением уровня фактора роста эндотелия сосудов и диаметром опухоли по данным КТ. Dalaveris и др. и Gessner и др. [Dalaveris et al., 2009; Gessner et al., 2010] также продемонстрировали повышение уровней фактора некроза опухоли α фактора роста эндотелия сосудов в КВВ пациентов с немелкоклеточным раком легкого. Однако повышенный уровень многих из вышеназванных маркеров демонстрировался также для КВВ пациентов с астмой и ХОБЛ, что может затруднить дифференциальную диагностику этих заболеваний. Среди интересных находок можно отметить исследование Cheng и др. [Cheng et al., 2011], в котором в КВВ пациентов с немелкоклеточным раком

легкого обнаружен регулируемый гормоном роста белок ТВС 1 (GRTP-1), а также исследование Zou и др. [Zou et al., 2013], в котором была показана корреляция СЕА (карциноэмбриональный антиген), SCC (антиген плоскоклеточного рака) и NSE (нейрон-специфическая енолаза) в КВВ с ранними стадиями немелкоклеточного рака легкого.

2.5 Протеом конденсата выдыхаемого воздуха в мультимодальных скрининговых программах

При рассмотрении вопроса о потенциальной роли исследований дыхания в клинической практике необходимо учитывать различные аспекты, включая эффективность, безопасность и стоимость. Основываясь на результатах своих исследований, Phillips и соавторы [Phillips et al., 2007] предположили, что исследование выдыхаемого воздуха вполне может занять свое место в мультимодальных скрининговых программах: пациенты из группы высокого риска могли бы сначала проходить такое исследование, а затем, в случае положительного результата, были бы направлены на КТ грудной клетки и далее, при необходимости, гистологическое подтверждение путем проведения бронхоскопии и/или биопсии легкого. Предположение авторов было основано на сравнительном анализе исследований выдыхаемого воздуха и других скрининговых методов [Horváth et al., 2009].

Несмотря на способность обнаруживать белки в КВВ и многочисленные исследования состава КВВ, было опубликовано лишь несколько исследований, документирующих использование протеомики для выявления различий между белковыми профилями КВВ больных раком легкого и здоровых людей (см. выше). Анализ белков КВВ зачастую ограничен специфическими белками [Carpagnano et al., 2004; Carpagnano et al., 2002] или измерением общей концентрации белка [Dwyer, 2004; Garey et al., 2004; McCafferty et al., 2004]. Необходимо заметить, что для того, чтобы считать скрининговый тест надежным, нужно убедиться в его способности отличить опухолевый процесс не только от здорового представителя, но и от заболеваний другого рода, каждое из которых скорее всего имеет свой уникальный белковый профиль [Dean et al., 2007].

2.6 Методы сбора и определения белкового состава конденсата выдыхаемого воздуха

Сбор КВВ. Сбор КВВ основан на конденсировании выдыхаемых газов и паров. Выдох направляется в устройство сбора и охлаждается, аэрозольные частицы конденсируются, капли собираются и отправляются на анализ или хранение. Сбор должен проводиться во время спокойного дыхания, с использованием носового зажима и уловителя слюны; поверхности, с которыми контактирует конденсат, должны быть инертными к компонентам. Носовой зажим используется для исключения носового дыхания и снижения вероятности попадания элементов слизистой носовой полости в дыхательный поток во время процедуры. Обычно сбор КВВ проводят в течение 10 минут: за это время можно получить 1–3 мл конденсата и процедура переносится пациентами без утомления [Dodig et al., 2013]. Количество КВВ зависит от количества воздуха, выдыхаемого в единицу времени (минутный объем) [Montuschi et al., 1999; Reinhold et al., 1999] температуры выдыхаемого воздуха и влажности.

Существует различное оборудование для сбора, но в настоящее время чаще всего используются системы R-Tube (Respiratory Research Inc., VA) и ECoScreen (VIASYS Healthcare, Hoechst, Germany). Согласно данным исследований по сравнению этих устройств [Hüttmann et al., 2011], pH и проводимость КВВ не зависят от устройства сбора. В одном исследовании были обнаружены различия в количествах некоторых компонентов КВВ, собранных этими устройствами, таких как эотаксин и цистеинил-лейкотриен [Soyer et al., 2006]. Czebe и соавторы также сравнили устройства сбора и разные температуры охлаждения для R-Tube, но не обнаружили существенных различий ни в объемах проб, ни в общем количестве белка в пробах, собранных при помощи R-Tube и ECoScreen [Czebe et al., 2008]. В другом исследовании [Hüttmann et al., 2011] объем КВВ оказался выше для проб, собранных с помощью R-Tube, но зато пробы, собранные при помощи ECoScreen, обнаруживали тенденцию к большим количествам общего белка и лучшей идентификации ряда белков. Тем не менее, авторы сделали вывод, что общее количество белка и значения для конкретных белков в КВВ зависят, скорее, от основного метода расчета и анализа [Hüttmann et al., 2011].

R-Tube представляет собой портативную одноразовую трубку, которая может быть использована в больнице, лаборатории и даже дома. Выдыхаемый воздух направляется через обратный клапан в охлажденную камеру для сбора, где происходит конденсация, далее клапан используется как поршень, который собирает капли со стенок. Слюна отделяется и скапливается в нижнем отсеке трубки, благодаря силе тяжести [Rosias et al., 2004]. В отличие от R-Tube, ECoScreen является стационарным устройством, которое имеет дополнительные функции, включая спирометр, отдельные камеры, позволяющие отбирать проб из различных

частей дыхательных путей, уловитель слюны и электрическую систему охлаждения. [Horvath et al., 2005]

Во время сбора образцов может происходить их загрязнение в том случае, если у пациентов повышенное слюноотделение. Для предотвращения таких загрязнений необходимо наличие очищающей системы или анализ образцов на предмет загрязнения слюной [Horvath et al., 2005; Corradi et al., 2008; Konstantinidi et al., 2015]. Загрязнение слюной может оказывать влияние на уровень различных маркеров в КВВ. Так, присутствие большого количества эйкозаноидов в слюне у детей при обострении астмы [Мозалевский и др., 1997] или нитритов/нитратов после принятия пищи с большим количеством этих соединений [Zetterquist et al., 1999] может оказывать влияние на уровень эйкозаноидов или маркеров метаболизма окиси азота в конденсате.

Анализ КВВ. КВВ является сложным объектом для протеомики из-за того, что это многокомпонентная смесь с высоким содержанием солей и липидов и с низкой концентрацией белка. Белковые биомаркеры часто присутствуют в небольших количествах, и обнаружение белка зависит от сложности белковой смеси, а также разрешающей способности и чувствительности используемого метода разделения и идентификации. Другая трудность состоит в том, что в выдыхаемом воздухе могут быть представлены как целые белки, так и пептиды различной длины. Основной причиной этого является присутствие гидролаз. Активность гидролаз усиливается при воспалении и обструкции эпителия дыхательных путей. Кроме того, пептиды могут проникать в легкие из системы кровообращения. Попытка идентифицировать белки методом иммунофлуоресценции (например, ELISA), в отличие от липидов в КВВ, может привести к потере результатов, поскольку моноклональное антитело может не взаимодействовать с фрагментом. Одним из подходов является прямой анализ сложной белковой смеси без разделения белков с использованием масс-спектрометрии [Cheng et al., 2011]. Этот подход позволяет идентифицировать пептиды, а использование масс-спектрометрии в сочетании с жидкостной хроматографией обеспечивает точный качественный анализ соединений и повышает селективность метода. Так, становится возможным распознать потенциальные маркеры в ничтожных концентрациях.

Другим подходом к анализу белковой компоненты в КВВ (тех белков, которые присутствуют в смеси целиком и могут быть разделены) является 2D-электрофорез в ПААГ, при котором происходит разделение белков по изоэлектрическим точкам и по молекулярным массам белков. [Hu et al., 2007; Лебедев, 2003] Визуализируют белки с помощью окрашивания кумасси, серебром, для идентификации используют MALDI-TOF масс-спектрометрию.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

ГЛАВА 3 Материалы и методы

3.1 Материалы

Реактивы для электрофореза, включая Precision Plus Protein™ standards и Silver Stain Plus Kit, были приобретены в Bio-Rad (США). Стандартные смеси белков MS standards UPS1 и MS-CAL1 были приобретены в Sigma-Aldrich (США), БСА приобретен в Serva Electrophoresis GmbH (Германия), модифицированный трипсин приобретен в Promega (США). Остальные реактивы, в том числе, ацетонитрил, гидрокарбонат аммония, дезоксихолат натрия, коричную кислоту, метанол, муравьиную кислоту, трихлоруксусную кислоту, хлороформ, высокой степени чистоты произведены в России.

3.2 Методы

3.2.1 Описание групп пациентов/доноров КВВ

В рамках настоящей работы обследовали 32 здоровых добровольца, 17 больных хронической обструктивной болезнью легких (ХОБЛ), 13 – внебольничной пневмонией, 46 больных с диагностированным раком легкого (таблица 1). Итого, за все время проведенной работы в исследовании КВВ приняли участие 108 человек.

Таблица 1. Характеристики основных групп доноров

Параметр	Группа доноров			
	Контроль	ХОБЛ	Пневмония	Рак легкого
Количество, чел.	32	17	13	46
Возраст, лет	От 20 до 45	От 53 до 76	От 19 до 60	От 32 до 81
Мужчины, чел. (%)	17 (53)	13 (76)	7 (54)	34 (74)
Женщины, чел. (%)	15 (47)	4 (24)	6 (46)	12 (26)
Курение: по наст. вр. / бывший / не курящий	8/0/24	12/5/0	4/2/7	23/5/16
Стадии заболевания	****	0/3/10/4*	2/4/2/3/2**	10/10/19/7***
Гистологический тип	–	–	–	21/10/1/1/1/1/10 [#]

* – стадии ХОБЛ: I/II/III/IV, ≥ 2 положительных критериев *Anthonisen*

** – PSI классы: I/II/III/IV/V

*** – стадии рака легкого: I/II/III/IV

**** – никаких симптомов аллергии, хронических респираторных заболеваний или острых респираторных симптомов на протяжении 2 месяцев перед сбором КВВ

[#] – плоскоклеточная карцинома / аденокарцинома / ангиокарцинома / мелкоклеточный / аденоплоскоклеточный / карциноид / крупноклеточный / другие виды рака с метастазами в лёгких

В качестве здоровых доноров приглашались добровольцы с неотягощенным анамнезом и отсутствием симптомов аллергии, хронических респираторных заболеваний или острых респираторных симптомов на протяжении 2 месяцев перед сбором КВВ.

Больные пневмонией и ХОБЛ в стадии обострения находились на стационарном лечении в пульмонологическом отделении ГКБ № 57 г. Москвы. Диагностику пневмонии и ХОБЛ осуществляли на основании общепринятых рекомендаций: WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) (<http://www.goldcopd.org>), Руководство по лечению инфекций нижних дыхательных путей у взрослых (Объединенная целевая группа Европейского респираторного общества и Европейского общества по клинической микробиологии и инфекционным болезням) [Woodhead et al., 2011]. Пациенты с диагностированным раком легкого находились на стационарном лечении в отделении торакальной хирургии МНИОИ имени П.А. Герцена, диагностика осуществлялась на основании данных компьютерной томографии органов грудной клетки и результатов исследования биопсии.

Исследование конденсата выдыхаемого воздуха было утверждено комитетами по этике ИБХФ РАН, НИИ пульмонологии РАН, МНИОИ имени П.А. Герцена, все доноры подписали информированное согласие на участие в исследовании.

3.2.2 Сбор и хранение проб КВВ

3.2.2.1 Система сбора R-Tube

Сбор КВВ пациентов с диагностированным раком легкого для облегчения процедуры и причинения как можно меньшего беспокойства пациентам осуществляли с помощью портативного устройства R-Tube (Respiratory Research, США) (рисунок 2). Образцы были собраны в первой половине дня после тщательного ополаскивания полости рта дистиллированной водой. Особое внимание уделялось исключению носового дыхания (с помощью носового зажима) и белкового загрязнения слюной и слизистой носоглотки. Образцы КВВ собирали в течение 10 мин с использованием предварительно охлажденной до -20°C R-Tube. Собранные пробы были помещены в полипропиленовые пробирки, заморожены и хранились в морозильнике при -85°C до проведения анализа.

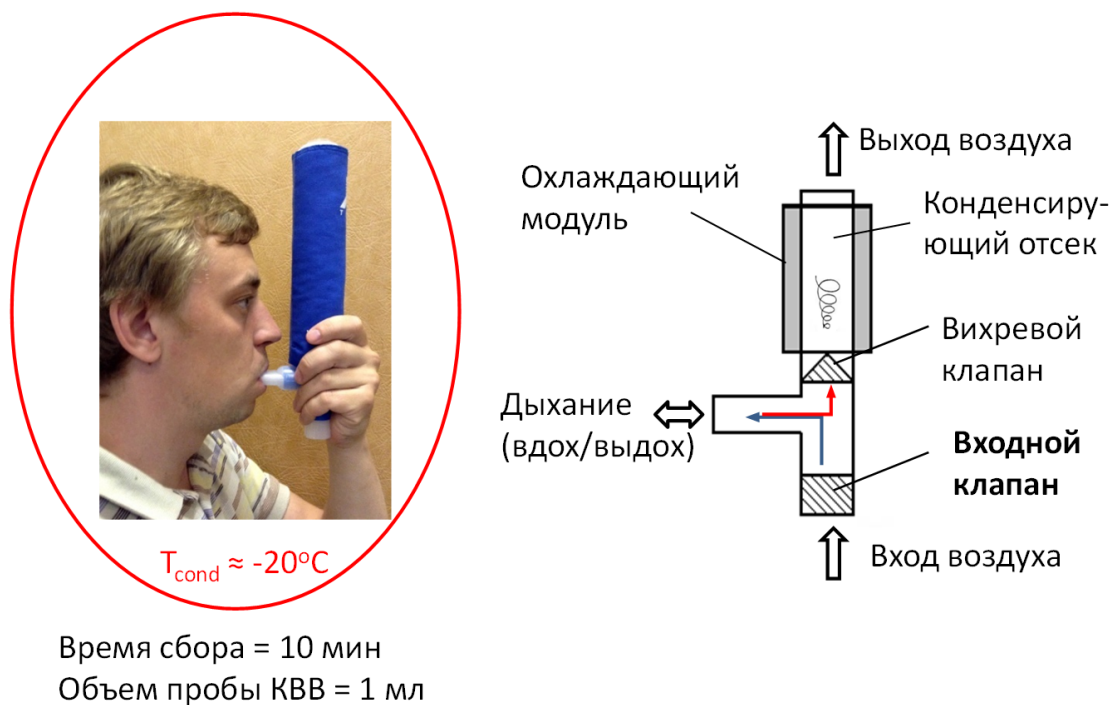


Рисунок 2. Конструкция R-Tube – портативного устройства для сбора КВВ.

3.2.2.2 Система сбора Eco-Screen

Конденсаты выдыхаемого воздуха пациентов с диагностированными ХОБЛ и внебольничной пневмонией собирали с помощью стационарного устройства ECoScreen. Каждый индивидуальный КВВ представлял собой конденсат воздуха, выдыхаемого человеком за 10 минут в режиме спокойного дыхания с носовой клипсой. Конструкция EcoScreen (VIASYSHealthcare, Германия) (рисунок 3) предотвращает попадание слюны в приемник. Конденсат собирался в тефлоновые приемники в модуле, охлаждаемом до -10°C .



Рисунок 3. Схема EcoScreen стационарного устройства для сбора КВВ.

3.2.3 Пробоподготовка КВВ

3.2.3.1 Подготовка препарата КВВ для проведения хромато-масс-спектрометрического анализа (ВЭЖХ-МС/МС)

Аликвоту КВВ (~объем 1 мл) переносили в полипропиленовые пробирки, устойчивые к низким температурам, с низкой белок-абсорбирующей поверхностью, и лиофилизировали (мягко сушили) в тех же пробирках до полного высыхания. Затем образцы КВВ гидролизовали модифицированным трипсином (Promega, США), добавленным в соотношении фермент/белок 1:100 (по массе) в 0,05 М NH_4HCO_3 буфер (pH 8,0) при 30°C. Реакцию останавливали после 20 часов инкубации добавлением 2 мкл 1% раствора муравьиной кислоты. Полученный раствор анализировали с помощью нанопоточной ВЭЖХ-МС/МС.

3.2.3.2 Обезжиривание и осаждение белков [Brannian et al., 1997]

В тонкостенную пробирку на 200 мкл поместили 10 мкл водной смеси белков (MS-CAL1, Sigma), к смеси добавили 7.5 объемов метанола и после интенсивного перемешивания на вортексе 7.5 объемов свежеперегнанного хлороформа. Смесь содержала 100 пмоль цитохрома С, 1 пмоль апомиоглобина, 1 пмоль альдолазы и 1 пмоль БСА. Белки осадили центрифугированием при 10000 об/мин в течение 5 мин на миницентрифуге (Eppendorf,

Германия). Раствор осторожно удалили микропипеткой, остаток высушили на воздухе под пористой пленкой, защищающей от пыли. В пробирку добавили 20 мкл 50 mM NH_4HCO_3 и тщательно обмыли стенки пробирки, затем в раствор добавили 0.02 мкг трипсина и оставили на ночь при 30 °C. Реакцию гидролиза остановили добавлением муравьиной кислоты до 0.1 %. Реакционную смесь анализировали с помощью нанопоточного ВЭЖХ-МС/МС в стандартных условиях. Вводимый объем пробы составлял 1 мкл, что соответствовало ≤ 50 фмоль каждого белка. Данные хромато-масс-спектрометрического анализа обрабатывали и анализировали, как описано в п. 3.2.4 и п. 3.2.8.2.

Лиофилизированные пробы КВВ были перерастворены в 10 мкл воды и экстрагированы, как описано выше.

3.2.3.3 Осаждение белков с помощью дезоксихолата натрия и трихлоруксусной кислоты [Bollag et al., 1991]

В пробирку на 1.5 мл поместили 1 мл смеси КВВ и белковых стандартов (MS standards UPS1, Sigma), взятых в эквимольном соотношении по 0,5 пикомоль. К смеси добавили 100 мкл 0,15 % дезоксихолата натрия, встряхнули на вортексе и оставили на 10 минут при комнатной температуре. Затем к пробе добавили 50 мкл 100% трихлоруксусной кислоты (ТХУ), перемешали на вортексе и оставили на 15 минут при -18 °C. Смесь центрифугировали 5 мин на максимальной скорости (10000 об/мин), после чего супернатант аккуратно удалили, а оставшийся осадок промыли дважды 150 мкл смеси метанол/хлороформ (1/1 по объему) и высушили под парафильмом при комнатной температуре. Осадок перерастворили в 20 мкл 50 mM NH_4HCO_3 и гидролизовали с 0,005 мкг/мкл трипсина при 30 °C 20 ч. Концентрация каждого белка составила 0.025 пмоль/мкл. Реакцию остановили с помощью 2 мкл 1 % муравьиной кислоты (до 0.1% кислоты в пробе). Полученный раствор (гидролизат) анализировали с помощью нанопоточного ВЭЖХ-МС/МС. Вводимый объем пробы составлял 4 мкл.

3.2.3.4 Концентрирование пептидов на колонке C18 (ZipTip C18)

Смеси белков с молекулярным весом от 12 до 67 кДа, взятых в эквимольном соотношении, были гидролизованы трипсином (0,02 мкг/мкл, 50 mM NH_4HCO_3 , 30 °C 16 часов). Были приготовлены две смеси: по 0,5 пмоль гидролизованных белков в 100 мкл воды и в 100 мкл гидролизата КВВ. Пептиды из данных смесей были сконцентрированы на картриджах ZipTip (Millipore) с обращенной фазой C18 согласно стандартному протоколу [Millipore]. Пептиды были смыты в 10 мкл 70%-ного метанола с 0,1 % муравьиной кислоты. Образцы были высушены и перерастворены в 10 мкл 0,1 % муравьиной кислоты. В качестве контроля 0,5 пмоль гидролизата стандартов растворили в 10 мкл 50%-ного метанола с 0,1 % муравьиной

кислоты (без обработки на ZipTip), высушили и перерастворили в 10 мкл 0,1 % муравьиной кислоты. Полученный гидролизат анализировали с помощью нанопоточного ВЭЖХ-МС/МС как описано выше. Вводимый объем пробы составлял 1 мкл, что соответствовало 50 фмоль каждого белка.

3.2.4 Хромато-масс-спектрометрический анализ (ВЭЖХ-МС/МС) проб КВВ

Хромато-масс-спектрометрический анализ (ВЭЖХ-МС/МС) проводили с использованием комбинированного измерительного устройства, состоящего из хроматографа Agilent 1100 (Agilent Technologies Inc., Санта-Клара, США) и масс-спектрометра LTQ FT Ultra (Thermo, Бремен, Германия) – масс-спектрометр ионного циклотронного резонанса (ИЦР), совмещенный с линейной квадрупольной ионной ловушкой, использующейся для накопления ионов и измерения спектров столкновительно индуцированной фрагментации (МС/МС) ионов (рисунок 4).

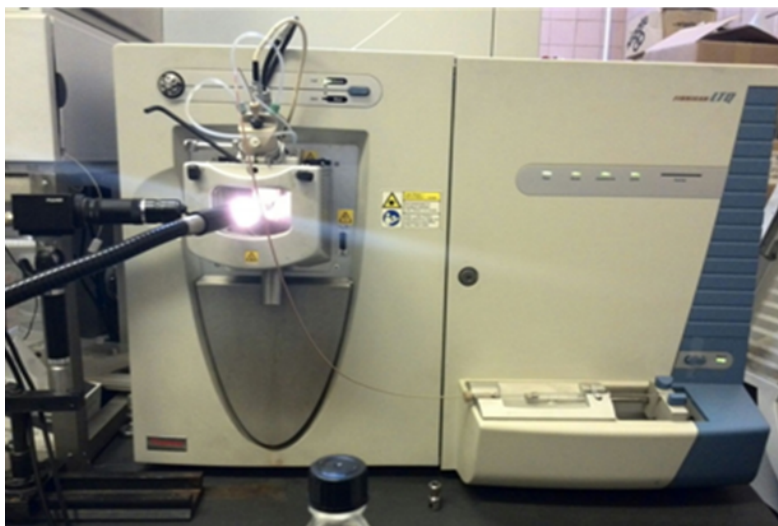


Рисунок 4. Гибридный масс-спектрометр LTQ-FT Ultra в тандеме с ВЭЖХ Agilent 1100

Объем вводимой на колонку пробы составлял 2 мкл, использовали колонку 75 мкм × 12 см с фазой Reprosil-Pur Basic C18, 3 мкм (Dr. Maisch HPLC GmbH, Аммербух-Энтринген, Германия), изготовленную в лаборатории по методике Манна и др. [Ishihama et al., 2002]. В качестве подвижной фазы использовали:

растворитель А – 0,1% раствор муравьиной кислоты в H₂O–НСООН (1000:1, по объему);

растворитель В – 0,1% раствор муравьиной кислоты в CH₃CN–НСООН (1000:1, по объему).

Проводили градиентную хроматографию с линейным увеличением относительного содержания растворителя В в потоке от 3% до 50% за 40 минут.

Масс-спектрометрический анализ фракций пептидов осуществляли при помощи программы Xcalibur (Thermo Finnigan, Сан Хосе, США) в 2-х стадийном режиме

автоматического измерения спектров. На первой стадии в масс-спектрометре ИЦР измеряли точные массы пептидов в диапазоне m/z 300-1600 с разрешением $R=50000$ для m/z 400 (число ионов в ячейке ИЦР 5×10^6). Было использовано динамическое исключение с длительностью исключения 30 миллисекунд. На второй стадии из ИЦР масс-спектра в автоматическом режиме отбирались три максимальных пика, для которых производилась столкновительно-индуцированная фрагментация (CID) (число аккумулялированных ионов 3×10^4).

Масс-спектры продуктов хроматографии находились в диапазоне от 300 до 1600 m/z . Каждый образец КВВ был проанализирован трижды и для последующего анализа отобраны лишь те белки/пептиды, которые были обнаружены в двух или трех повторах.

3.2.5 Электрофорез в полиакриламидном геле

Электрофорез в полиакриламидном геле проводили в денатурирующих условиях с добавлением детергента додецилсульфата натрия по методу Лэмбли [Laemmli, 1970]. Прокрашивание гелей осуществляли с помощью красителя Silver Stain Plus Kit (Bio-Rad, USA). В качестве маркеров молекулярной массы использовали смесь белков Precision Plus Protein™ standards (Bio-Rad, USA), включающую в себя: фосфоорилазу b (97 кДа), бычий сывороточный альбумин (66,2 кДа), яичный альбумин (45 кДа) и карбоангидразу (31 кДа).

3.2.6 Гидролиз белков в ПААГ

Использовали несиликонизированные пробирки для ПЦР (200 мкл). Куски геля диаметром 1,5 мм вырезали в ламинарном шкафу иглой шприца и вносили в пробирки. Добавляли в пробирки с гелем по 60 мкл отмывающего краску раствора ацетонитрил/ NH_4HCO_3 (1:1) и инкубировали 30 мин на шейкере. Процедуру промывки повторяли несколько раз. Затем меняли отмывающий раствор на ацетонитрил, инкубировали 20 минут, отбирали из пробирок жидкость, и ставили открытые пробирки с кусками геля в термостат на 45°C на 15 минут для высушивания. В каждую пробирку, содержащую кусочек геля, добавляли 5-10 мкл рабочего раствора трипсина в зависимости от размера кусочка. Пробирки в течение 1 часа инкубировали на льду. После чего для проведения реакции расщепления пробирки переносили в термостат на 37°C и инкубировали 3 часа. Трипсинолиз останавливали добавлением 20 мкл свежеприготовленного 0,5% (об/об) раствора муравьиной кислоты в воде. Выдерживали при комнатной температуре на шейкере 30 минут, затем экстракт отбирали в чистые пробирки и замораживали до использования.

3.2.7 Получение масс-спектров белков, сконцентрированных в ходе электрофореза

Масс-спектры белка и смеси пептидов, образующихся после обработки белка трипсином, получали на MALDI-TOF масс-спектрометре MicroFlex с использованием

стандартной мишени – MSP target polished steel (Bruker, Германия). Масс-спектры получены в рефлекторной моде, детектировались положительные ионы. В случае измерения масс-спектра белка на мишень наносили смесь измеряемого образца с синапиновой кислотой в соотношении 1:1. В случае измерения масс-спектра смеси пептидов на мишень наносили смесь измеряемого образца с коричневой кислотой в соотношении 1:4. Затем высушивали образцы на воздухе и проводили измерение. Внешнюю калибровку проводили с использованием точных значений масс известных белков. Формирование масс-листа проводили в ручном режиме. Идентификацию фрагментов проводили путем поиска совпадений значений экспериментальных масс с массами фрагментов, аннотированных в базах данных (SwissProt/TrEMBL).

3.2.8 Биоинформатический анализ результатов

3.2.8.1 Создание реляционной базы данных проб КВВ

На основании полученных в ходе сбора проб у пациентов был осуществлен отбор параметров, которые потенциально могли повлиять на результаты анализа состава КВВ. Все отобранные параметры были собраны в сводную таблицу в формате Excel, информацию в которую заносили на этапе сбора пробы и последующих этапах пробоподготовки, получения и анализа масс-спектра. Для каждого образца КВВ определяли два десятка параметров.

Для моделирования и проектирования базы данных использовали программу MySQLWorkbench 6.3 CE.

Заполнение базы данных осуществляли с помощью PHPMyAdmin software. PHPMyAdmin – это веб-приложение с открытым кодом, написанное на языке PHP и представляющее собой веб-интерфейс для администрирования системы управления базами данных MySQL. PHPMyAdmin позволяет через браузер осуществлять администрирование сервера MySQL, запускать команды SQL и просматривать содержимое таблиц и баз данных.

3.2.8.2 Анализ масс-листов с помощью Mascot

Список из точных масс пептидов и масс их фрагментов использовали для поиска и идентификации белков по базе данных при помощи программы Mascot (Matrix Science, Лондон, Великобритания; version 2.0.04). Для идентификации белков использовали базы данных NCBItr (National Center for Biotechnology Information, США) и IPI Human (The European Bioinformatics Institute, Великобритания) с выбранной таксономией Homo sapiens. Для идентификации белков использовали следующие параметры поиска: фермент – трипсин; точность масс для родительского иона – 5 ppm; точность масс для MS/MS фрагментов – 0,50 Da; модификации – окисление метионина. Поскольку концентрация большинства белков в КВВ крайне низкая, и

белки, в основном, летят не полностью, а в виде отдельных фрагментов [Cheng et al., 2011], что создает определенные трудности при их идентификации, считали, что белок достоверно идентифицирован, если для него нашлось не менее двух уникальных пептидов (Score > 70) у одного из доноров или если данный белок при наличии не менее одного уникального пептида (Score > 30) нашелся у нескольких доноров рассматриваемой группы.

3.2.8.3 Анализ белков, с наибольшей вероятностью обнаруживаемых в пробе

Для аннотирования и анализа результатов были использованы биоинформатические базы данных GeneCards (<http://www.genecards.org>), GeneOntology (GO) (<http://geneontology.org>), MOPED (<https://www.proteinspire.org/MOPED>), BioGPS (<http://biogps.org>), UniProt (www.uniprot.org), а также продукты компании Quagen (<http://www.qiagen.com>) для профилирования различных физиологических и патологических процессов в организме человека.

MetaCore (Clarivate Analytics, США) использовали для первичного анализа метаболических путей для белков, наиболее часто встречающихся в КВВ здоровых доноров. Программное обеспечение MetaCore основано на базе данных белок-белковых, ДНК-белковых и низкомолекулярное соединение-белковых взаимодействиях в организме человека. MetaCore устанавливает метаболические и сигнальные пути и определяет действие биоактивных молекул. Коды UniProt для различно экспрессирующихся белков загружаются в MetaCore для последующего анализа, основанного на процессах из Gene Ontology.

3.2.8.4 Сопоставление пептидной и белковой выдачи программы Mascot

Провели оценку количества пептидов, используемых программой для определения белков. Провели сравнение пептидных файлов Mascot, полученных при использовании разных баз данных, но выводящих одинаковые результаты белкового анализа. Результаты обработки одного масс-листа программой Mascot с различными параметрами поиска сравнивали следующим образом:

- 1) с помощью bash-скрипта из обработанного программой Mascot масс-листа извлекали все уникальные пептиды, белки, их score и e-value. В случае, если пептид встречался более одного раза, его извлекали с наибольшим score;

- 2) проводили сравнение пептидов с достаточно низким e-value (e-value < 1). Достоверно присутствующими в пробе считали только такие пептиды, т.к. высокий e-value служит индикатором того, что такой пептид присутствует в выдаче не случайно;

- 3) списки пептидов группировали по сравниваемому признаку. С помощью команды bash определяли, какие пептиды соответствуют выбранному белку в файлах, посчитанных по

базам данных NCBI nr и IPI-human. В таблицу записывался наиболее часто встречающийся набор пептидов, а также отличия от него.

4) С помощью bash-скрипта из файлов в формате .dat был получен список пептидов, которые соответствуют первому пептиду из каждого query, исходя из предположения, что именно пептиды из этой записи чаще всего попадают в файл формата .html. С помощью bash-скрипта был получен список пептидов из .html-файла, который соответствовал данному .dat-файлу. Две таблицы со списком пептидов затем сравнивались в Excel с целью выявить несоответствия в количестве и последовательности пептидов.

5) С помощью скрипта на python была составлена упорядоченная таблица всех пептидов из dat-файла, которая сравнивалась со списком пептидов, полученным из соответствующего html-файла с целью выявить несоответствия в последовательности пептидов.

6) *.dat и *.html-файлы также сравнивали вручную, чтобы определить иные различия между ними, не касающиеся последовательности пептидов.

3.2.8.5 Анализ пептидов, с наибольшей вероятностью обнаруживаемых в пробе

Манипуляции с данными проводили с помощью python пакетов numpy [van der Walt et al., 2011], pandas [McKinney, 2010].

Статистический анализ основывали на на тесте Манна-Уитни и методе главных компонент. Разложение на 10 принцепальных компонент произведено с помощью python-библиотеки sklearn [Pedregosa et al., 2011]. На графиках отображались первые 2 главные компоненты.

Принадлежность пептидов белкам анализировали с помощью сервиса Protein Information Resource [Wu et al., 2003], пептиды аннотировали по UniRef100 базе человеческих белков (сборка UniProtKB release 2017_06). Для определения семейств идентифицированных по пептидам белков, с помощью сервиса GeneMania [Warde-Farley et al., 2010] белки, подтвержденные минимум двумя пептидами, анализировали в виде сети кодирующих их генов. Для каждого белка получили аннотацию функции, гена, связи с другими белками, функциональную аннотацию исследуемых генов (выбирали не более 4 разных биологических процессов).

Результаты статистического анализа полученных данных визуализировали с помощью python-библиотеки matplotlib [Hunter, 2007].

3.2.9 Построение прогностических моделей на основании статистического анализа пептидного состава проб КВВ

При построении аналитических моделей по предсказанию диагноза на основании пептидного состава пробы КВВ использовали логистическую регрессию и алгоритм машинного обучения Random Forest [Breiman, 2001].

В качестве критерия качества модели использовался AUC (площадь под ROC-кривой), а также чувствительность и специфичность. Данные показатели рассчитывались с использованием метода скользящего контроля (cross-validation). Для того, чтобы повысить приоритет записей, в которых есть заболевания, использовался метод up-sampling (подробное описание приведено в главе 8).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Глава 4. Особенности анализа масс-спектров белков КВВ

Глава посвящена принципу работы программы Mascot – наиболее часто используемой программы для анализа масс-спектров в протеомных исследованиях. Описаны изменяемые параметры программы и выбор параметров в соответствии с типом прибора и характеристиками образца, используемые программой базы данных. Выводы этой главы были опубликованы в статьях [Kurova et al., 2009] и [Курова и др., 2010].

4.1 Принцип работы программы Mascot [Курова и др., 2010]

Программа Mascot является на сегодняшний день одной из самых используемых программ для интерпретации результатов масс-спектрометрии белков. Mascot сопоставляет масс-лист, принятый на вход, со сгенерированными *in silico* масс-листами всех возможных пептидов. Информация об этих пептидах (их аминокислотная последовательность) берётся из белковых баз данных, выбранных пользователем. Теоретические масс-листы пептидов генерируются с использованием дополнительной информации о пробоподготовке и пробообработке, введённой пользователем (параметров поиска). Таким образом, теоретические и практические масс-листы могут считаться полученными в одинаковых условиях. Из всех существующих в настоящее время программных продуктов поисковая машина Mascot обладает наиболее широкими возможностями для сравнения экспериментальных и теоретических данных при идентификации белков и пептидов. Принцип работы Mascot основан на алгоритме MOWSE (MOlecular Weight SEarch), разработанном в 1993 году Дэррилом Паппином (Darryl Pappin) [Pappin et al., 1993]. Программа позволяет проводить идентификацию по вычисленным специфическим для каждого белка наборам масс пептидов. Совпадение каждой вычисленной величины массы пептида с экспериментальной величиной характеризуется заданной погрешностью. Вместо простого счета числа совпавших пептидов MOWSE-алгоритм использует эмпирически полученные факторы для определения статистического веса каждого совпавшего пептида. После сравнения экспериментальных и теоретических величин для каждого элемента определяется рейтинг (величина, характеризующая уровень достоверности результатов идентификации, – Score).

Программа Mascot представляет результаты в виде таблиц, где белки выстраиваются согласно своему рейтингу. Для каждого белка указывается список пептидов, теоретические массы которых совпадают в пределах заданной погрешности с экспериментальными данными. [Курова и др., 2010]

Результаты Mascot-идентификации проиллюстрированы на примере анализа масс-спектра гидролизата образца КВВ. Детектированный в спектре сигнал двухзарядного иона с m/z 968,057198 соответствует массе незаряженной частицы 1934,099844. В базе данных NCBI был найден пептид Десмоплакина III с последовательностью TLVTQNSGV EALIHAILR, масса которого отличается от экспериментального значения на 0,01 Да. В таблице 2 представлены теоретические массы фрагментов пептида TLVTQNSGV EALIHAILR, полученные методом столкновительной диссоциации в ионной ловушке, которые в пределах погрешности 0,5 Да совпадают с фрагментами катиона с m/z 968,057198. В таблице 2 также приведены фрагменты так называемых b- и y-серий, которые образуются при расщеплении пептидной связи. Ионы b-серии представляют собой протонированные N-концевые части пептида (слева от расщепляемой пептидной связи), а ионы y-серии — их протонированные C-концевые (правые) части. В таблице даны также массы двухзарядных ионов этих фрагментов и массы фрагментов с выбросом молекулы аммиака (b^* и y^*) или воды (b^0 и y^0). По базе данных NCBI программа Mascot идентифицировала рассмотренный фрагмент согласно массе иона и набору масс его фрагментов как пептид TLVTQNSGV EALIHAILR, принадлежащий десмоплакину III, с рейтингом 75. Следует отметить, что в данном случае программа Mascot предложила единственный вариант последовательности аминокислот. Так бывает далеко не всегда, причем пептиды с похожими последовательностями могут принадлежать совершенно разным белкам. В таких случаях следует рассматривать варианты белков, предложенные программой Mascot, с учетом информации об исследуемом образце, о локализации предполагаемых белков в организме и т.д. [Курова и др., 2010]

Таблица 2. Теоретические массы фрагмента TLVTQNSGVEALIHAILR Десмоплакина III, совпадающие с погрешностью ≤ 0.5 Да с экспериментальными массами фрагментов, полученных методом столкновительной диссоциации двухзарядного катиона с m/z 968.057198

Номер фрагмента	b	b ⁺⁺	b*	b ⁰	AK	y	y ⁺⁺	y*	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	Номер фрагмента
1					T							18
2					L							17
3	314.21				V		860.99					16
4				397.24	T	1621.90	811.45	1604.88				15
5			526.29		Q	1520.85		1503.83	752.42		751.93	14
6					N	1392.80		1375.77		1374.79		13
7					S	1278.75						12
8		401.21			G	1191.72						11
9	900.48				V	1134.70						10
10					E	1035.63			509.81		509.31	9
11	1100.56		1083.53		A	906.59						8
12	1213.64		1196.62		L	835.55						7
13	1326.73	663.87		1308.72	I	722.47						6
14	1463.79		1446.76		H	609.38						5
15	1534.82				A	472.32						4
16	1647.91				I	401.29						3
17	1760.99				L	288.20						2
18					R							1

Примечание. Ионы b- и y-серий, образующиеся при расщеплении пептидной связи, соответствуют протонированным N-концевому (слева от расщепляемой связи) фрагменту и C-концевому (справа от расщепляемой связи) фрагменту пептида; b⁺⁺ и y⁺⁺ — массы двухзарядных ионов фрагментов b- и y-серий; b* и y* — массы фрагментов b- и y-серий, утративших молекулу аммиака; b⁰ и y⁰ — массы фрагментов b- и y-серий, утративших молекулу воды. AK — аминокислота в пептиде.

Ниже описаны изменяемые параметры программы Mascot, используемые при методе поиска MS/MS Ions search, так как только этот метод может использоваться для исследования белкового состава КВВ по описанной методике.

4.2 Описание и выбор параметров программы Mascot в зависимости от эксперимента

4.2.1 Описание параметров программы Mascot

Параметры, задающие область поиска

База данных (Database). Данный параметр позволяет выбрать базу данных белковых последовательностей, по которой будет производиться поиск результатов.

Таксономия (Taxonomy). Данный параметр позволяет сузить область поиска одной или несколькими таксономическими группами (вид, род и т.д.). Такой подход, во-первых, позволяет значительно ускорить анализ результатов, во-вторых, гарантирует, что белки в выдаче будут принадлежать только исследуемому организму.

Параметры, описывающие пробоподготовку и пробообработку

Фермент (Enzyme). Данный параметр задаёт фермент, с помощью которого в процессе пробоподготовки фрагментировался исследуемый белок или белковая смесь.

Максимальное допустимое количество пропущенных сайтов гидролиза трипсином (Max Missed Cleavages). С помощью этого параметра задаётся максимальное количество пептидных связей в белке, которые были «пропущены» заданным ферментом.

Инструмент (Instrument). С помощью этого параметра задаётся тип масс-спектрометра, с помощью которого были получены данные. Все масс-спектрометры могут быть отнесены к одной из следующих групп (по способу ионизации и типу анализатора): ESI-QUAD-TOF, MALDI-TOF-PSD, ESI-TRAP, ESI-QUAD, ESI-FTICR, MALDI-TOF-TOF, ESI-4SECTOR, FTMS-ECD, ETD-TRAP, MALDI-QUAD-TOF, MALDI-QIT-TOF, MALDI-ISD, CID+ETD. Соответственно, эти типы приборов генерируют ионы с отличающимся значением m/z , что и учитывает данный параметр.

Заряд пептида (Peptide charge). С помощью этого параметра можно задать заряд исходного иона, полученного на этапе ионизации пробы. Как правило, при ионизации методом электроспрей большинство ионов имеют заряд $2+$, а при ионизации MALDI-PSD большинство ионов имеют заряд $+1$ (www.matrixscience.com). Таким образом, данный параметр принимает соответствующие значения в зависимости от способа ионизации.

Погрешность масс-спектрометра (MS/MS tol. \pm). Параметр задаёт максимальную погрешность измерения массы иона в MS/MS. Измеряется в тех же единицах, что и погрешность измерения массы пептида.

Прекурсор (Precursor). Mascot может принимать на вход данные различного типа. Входные данные в некоторых форматах (например, .PKS и .XML) не содержат информацию о соотношении массы к заряду исходного иона. Параметр Precursor позволяет исследователю указать эту информацию при работе с данными в указанных выше форматах. При этом значение заряда определяется значением параметра Peptide Charge.

Параметры, описывающие белки

Модификации (Modifications). На сегодняшний день известно более 200 белковых модификаций [Jensen, 2006]. Все эти модификации могут существенно изменять массу белка. Параметр “Modifications” позволяет учесть такие белковые модификации. При этом Mascot поддерживает два типа модификаций: fixed modifications (обязательно присутствующие модификации) и variable modifications (возможные модификации).

Первый тип модификаций (Fixed modifications) – это те модификации, которые будут применяться Mascot к каждому соответствующему аминокислотному остатку.

Возможные белковые модификации факультативны. Эти модификации тестируются программой Mascot во всех возможных сочетаниях.

Однако большое количество возможных модификаций существенно замедляет работу Mascot, так как каждая такая модификация влечёт за собой появление большого количества новых пептидов. Mascot позволяет включить в рассмотрение до девяти возможных модификаций. Тем не менее, рекомендуется выбирать как можно меньше модификаций, чтобы избежать описанных выше проблем.

Масса белка (Protein Mass). Параметр определяет максимально возможную массу белка в исследуемой пробе. Единицы измерения – kDa.

Погрешность измерения массы пептида (Peptide tolerance). Параметр задаёт максимально возможную погрешность измерения массы пептида. Может измеряться в следующих единицах: %, mmu (milli-mass units), ppm (parts per million), Da.

4.2.2 Выбор параметров программы Mascot

4.2.2.1 Параметры с неизменными значениями

Значения следующих параметров из групп «параметры, описывающие пробоподготовку и пробообработку» и «параметры, описывающие белки», определяли, исходя из способа пробоподготовки и пробообработки КВВ (таблица 3):

Таблица 3. Параметры с неизменными значениями

Параметр	Значение
Enzyme	Trypsin
Instrument	ESI-trap
MS/MS tol	0.5 Da
Precursor	-
Protein mass	30 kDa
Peptide tolerance	10.0 ppm
Peptide charge	2+

Значения параметров, представленные в таблице 3, были выбраны по следующим причинам:

- Значение параметра «Enzyme» определяется способом пробоподготовки (ферментативный гидролиз трипсином);
- Значения параметров «Instrument», «MS/MS tol», «Peptide tolerance», «Precursor», «Peptide charge» определялись масс-спектрометром, с помощью которого обрабатывали пробы. Масс-спектрометр LTQ-FT Ultra относится к типу приборов ESI-trap (приборы с методом ионизации ESI) с указанной погрешностью измерений.
- Значение параметра «Protein mass» выбрали, исходя из особенностей исследования [Kurova et al., 2009].

4.2.2.2 Параметры с изменяемыми значениями

Значения параметров из этой группы не определялись способом пробоподготовки и пробообработки КВВ.

- Модификации белков. Модификации белков для исследования отбирали следующим образом: для всех белковых модификаций, доступных для анализа в программе (84 белковых модификации), выяснили функции модификаций, частоту встречаемости модификаций и белки, в которых эти модификации происходят. Далее, из отобранных групп модификаций для анализа выбрали только часто встречающиеся модификации (более одной модификации данного типа на белок). При сопоставлении получившегося списка со спецификой пробы, в качестве учитываемой модификации выбрали окисление метионина.

- База данных. Программа Mascot работает только с базами данных, содержащими информацию о белках в Fasta-формате [Perkins et al., 1999]. В настройках Mascot по умолчанию доступно 4 варианта баз данных, содержащих информацию о белках человека: NCBIInr, cRAP, SwissProt, Human_EST (www.matrixscience.com). В протеомных исследованиях также может использоваться база данных IPI-human [Рябоконе и др., 2014]. База данных cRAP (The common Repository of Adventitious Proteins, <http://www.thegpm.org/crap/>) содержит в себе информацию о белках, которые не должны присутствовать в исследуемой пробе (выделяется три группы таких белков: группа белков, часто присутствующих в лаборатории – например, BSA; группа белков, оказавшихся в пробе вследствие физического контакта с окружающей средой; группа белков, используемых как внутренний количественный стандарт). В силу узконаправленной специализации, базу данных cRAP исключили из рассмотрения. Human_EST (Expressed Sequence Tag, <http://www.ncbi.nlm.nih.gov/dbEST/>) – база данных, представляющая часть базы данных GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). В базе данных EST содержится информация о коротких последовательностях кДНК разных организмов, в том числе и человека. База данных Human_EST не содержит аминокислотных последовательностей, поэтому была исключена из возможных вариаций параметра «Database(s)» программы Mascot. SwissProt (<http://www.uniprot.org/>) – курируемая база данных с высоким уровнем аннотации белков (все белки аннотированы вручную), в которой отсутствуют избыточные данные. Эта база данных очень часто используется в исследовании человеческого протеома, однако, в рамках данного исследования использовались более крупные базы данных IPI-human и NCBIInr. IPI-human – часть более крупной базы данных International Protein Index и построена на основе информации о белках из трёх крупных баз данных: UniProt Knowledgebase, Ensembl и RefSeq. База данных ежемесячно обновлялась до 2011 года. Сейчас проект закрыт, однако IPI всё ещё

используется в протеомных исследованиях. NCBIInr (NCBI non-redundant protein data base, <http://www.ncbi.nlm.nih.gov/protein>) была создана NCBI для поискового сервиса BLAST. Эта база данных объединяет в себе данные из GenBank, PDB, Swiss-Prot, PIR, PRF и в настоящее время активно используется при исследовании белков.

В данном исследовании работа велась с IPI-human и с NCBIInr. По причине закрытия проекта IPI-human во время проведения исследования, все данные, обработанные с помощью этой базы данных, были пересчитаны по базе данных NCBIInr.

- Таксономия. В данном исследовании при анализе проб выбирали таксон *Homo Sapiens*.
- Max Missed cleavages. Данный параметр характеризует максимальное количество трипсиновых сайтов, по которым не произошло расщепление белка при пробоподготовке. В среднем, 10% связей в белке являются сайтами трипсиновой рестрикции [Perkins et al., 1999], однако рестрикция происходит не по всем трипсиновым сайтам. Пропуск трипсиновых сайтов может происходить по стерическим причинам, либо из-за того, что трипсин хуже расщепляет белок, если с сайтом рестрикции соседствует основной аминокислотный остаток [Keil et al., 1992]. Второе обстоятельство не учитывается Mascot, так как оно существенно замедляет работу программы. Однако в Mascot включен параметр Max Missed Cleavages, который позволяет пользователю самому предполагать, сколько трипсиновых сайтов будет пропущено при пробообработке. Учитывая малую концентрацию пептидов в пробе и их небольшую длину, можно предполагать, что трипсиновые сайты не будут пропускаться по стерическим причинам. Однако второй фактор, тем не менее, может оказывать влияние на пропуск трипсином сайтов, что было показано на примере стандартной смеси белков человека с добавлением матрицы KBB (см. Главу 5). В данном исследовании задавался параметр Max Missed Cleavages, равный 2.

Глава 5. Основные свойства КВВ и методика пробоподготовки КВВ для анализа белкового состава, основанные на исследовании 17 молодых здоровых некурящих доноров

Определение белкового состава КВВ является непростой задачей, для решения которой необходимо подобрать условия сбора пробы и отработать основные этапы пробоподготовки. Кроме того, необходимо на модельных белковых смесях, содержащих сопоставимые с КВВ количества белка, показать принципиальную возможность идентификации белков. Данные вопросы рассмотрены в настоящей главе.

Выводы этой главы были опубликованы в статьях [Kurova et al., 2009] и [Kurova, Kurochkin et al., 2009].

5.1 Основные свойства КВВ

Для отработки методики было собрано КВВ у семнадцати доноров (7 мужчин возраста 21-36 лет и 10 женщин возраста 20-32 лет). Процедуру осуществляли в течение 10 минут, отведенных согласно медицинским рекомендациям по сбору КВВ, в отсутствие защитного фильтра перед воздухозаборником. Были исследованы тефлоновые и полипропиленовые виалы для сбора внутри конденсора, поскольку они совместимы с системой EcoScreen. Обнаружено, что оба вида материалов совместимы с масс-спектрометрическим анализом. Несмотря на данные литературы о более высокой степени извлечения альбумина *in vitro* из конденсирующих систем со стеклянным и силиконовым покрытием [Rosias et al., 2006], даже в той же работе извлечение альбумина кажется значительно более воспроизводимым для полипропиленового, алюминиевого и тефлонового покрытий, чем для стеклянного или силиконового. Кроме того, предполагалось, что сорбция белка не является существенной и вполне воспроизводима при температуре сбора -10 °С для всех внутренних частей EcoScreen, поскольку замороженная проба остается в сборнике на короткое время.

Для каждого донора из первой набранной группы 17 здоровых некурящих добровольцев контролировали общий объем выдыхаемого воздуха и pH КВВ. pH КВВ измерялся сразу после сбора (без дегазирования) с помощью откалиброванного pH-метра со стеклянным микроэлектродом (Beckman, USA) (таблица 4).

Таблица 4. Данные для проб 17 здоровых некурящих доноров

№	Пол	Возраст	pH	Общий объем выдыхаемого воздуха, л/10 мин	Объем КВВ, мл/10 мин
1	Ж	27	7,20	106,6	2,3
2	Ж	20	7,04	155,0	2,4
3	Ж	22	7,14	128,4	2,3
4	М	21	7,14	104,0	1,3
5	Ж	25	6,99	154,8	2,5
6	Ж	25	7,17	142,5	1,2
7	М	32	7,07	148,6	2,5
8	М	28	7,12	123,0	2,0
9	М	30	6,93	138,0	2,4
10	Ж	34	7,21	117,9	2,0
11	Ж	32	7,18	130,0	2,5
12	М	27	7,26	150,0	2,3
13	М	36	7,09	385,0	3,0
14	Ж	25	7,07	116,0	2,6
15	М	30	7,16	92,5	2,0
16	Ж	22	7,12	470,0	2,5
17	Ж	32	7,20	332,0	2,5

Объем конденсата варьировался в пределах 1,2–3,0 мл. Объем воздуха для здоровых доноров составил 129 ± 16 л/10мин. Среднее значение pH КВВ доноров составило $7,12 \pm 0,08$ без деаэрации, что является показателем здоровья дыхательных путей доноров, в соответствии с данными литературы [Horváth et al., 2005], где было показано, что измерение pH КВВ является достоверным методом оценки и наблюдения за воспалением дыхательных путей.

Концентрацию белка в КВВ оценили с помощью SDS-ПААГ электрофореза с окраской серебром [Kurova et al., 2009]. Принимая во внимание чувствительность окраски серебром (0,5 нг/полоса), можно заключить, что 0,25 мл КВВ содержит < 1 мкг белка, соответственно, концентрация белка в КВВ составляет < 4 мкг/мл.

5.2 Подбор оптимальной методики концентрирования белковых смесей для масс-спектрометрического определения/измерения [Kurova et al., 2009]

Низкие концентрации белка и сложная разнообразная матрица КВВ (наличие солей и липидов, затрудняющих МС-анализ) требуют оптимизации пробоподготовки, для уменьшения

потерь белка. Кроме того, за 10 мин, отведенные согласно медицинским рекомендациям по сбору КВВ, можно получить не более 3 мл образца. Чтобы улучшить извлечение белка и минимизировать количество этапов прободготовки, а также ограничить количество переносов пробы из пробирки в пробирку, для сбора КВВ использовался тефлоновый или полипропиленовый коллектор, а размороженный конденсат переносили из коллектора в специальные пробирки с низкосорбирующей поверхностью. Все прочие операции с образцами КВВ проводили в этих же пробирках вплоть до забора образца для хромато-масс-спектрометрического анализа. [Kurova et al., 2009]

Ввиду экстремально низкой концентрации белка в КВВ, было принято решение о дополнительной проверке возможности достоверной идентификации белков в пробах. Для этого были проведены модельные эксперименты со смесями белковых стандартов.

5.2.1 Чувствительность ВЭЖХ-МС/МС анализа

Для идентификации белков по данным масс-спектрометрии была использована поисковая программа Mascot, которая проводит поиск и идентификацию белков по базам данных первичных последовательностей белков и соответствующих генов. Из всех существующих на сегодняшний день программных продуктов, поисковая машина Mascot обладает наиболее широкими возможностями для сравнения экспериментальных и теоретических данных при идентификации белков и пептидов (более подробно см. Гл.4).

Анализ смеси белковых стандартов (4 белка) до высушивания позволил нам идентифицировать до 40 фемтомоль каждого представленного белка с высоким покрытием их первичных последовательностей (таблица 5). Это количество белка было принято за предел чувствительности данного метода.

5.2.2 Оценка концентрации белка в КВВ

Общая концентрация белка в КВВ не превышает 4 мкг/мл (см. выше), и при этом в одной пробе КВВ может содержаться около 100 различных белков, при этом объем образца для проведения измерения на масс-спектрометре составляет 1-2 мкл. Не сложно произвести оценку, что для идентификации белков в КВВ, исходно получаемую пробу необходимо сконцентрировать не менее, чем в 20-40 раз.

ОЦЕНКА:

1. Общая концентрация белка в КВВ $\text{менее } 4 \text{ мкг/мл} = 4 \times 10^{-3} \text{ г/л}$
2. В качестве оценки примем, что $\langle M_{\text{белка}} \rangle = (5 \div 50) \times 10^3 \text{ г/моль}$
3. Значит концентрация белка в КВВ $(4 \times 10^{-3} \text{ г/л}) / ((5 \div 50) \times 10^3 \text{ г/моль}) \sim 10^{-6} \div 10^{-7} \text{ М}$
4. Количество белка в пробе объемом 1 мл $10^{-9} \div 10^{-10} \text{ моль} = 10^6 \div 10^5 \text{ фемтомоль}$

5. Количество различных белков в пробе КВВ до 100 штук,
 6. Значит количество каждого белка в пробе около $10^4 \div 10^3$ фемтомоль
 7. Пусть в масс-спектрометр вводится 1 мкл пробы (1/1000 часть от исходного объема пробы КВВ в 1 мл)
 8. Тогда количество каждого белка вводимого в масс-спектрометр от 10 до 1 фемтомоль.
- Поскольку чувствительность метода оценена в 40 фемтомоль (но могут анализироваться смеси, содержащие и большее количество белка), то пробу КВВ необходимо сконцентрировать примерно в 40 раз.

Разбавленность образцов является значительной проблемой при оценке биомаркеров из КВВ: концентрации белков и так малы, а дополнительное разбавление, которое может быть вызвано избыточной влагой в выдыхаемом воздухе или слезью, может снизить ее до неопределяемой величины. [Carter et al., 2012] Таким образом, для анализа белкового состава КВВ возникает необходимость в подборе оптимального способа концентрирования пробы. (рисунок 5) Поскольку при этом можно ожидать увеличения концентрации и низкомолекулярной составляющей пробы (солей, липидов и т.д.), необходимо проверить, как на идентификацию белков повлияет применение методик обессоливания и обезжиривания.

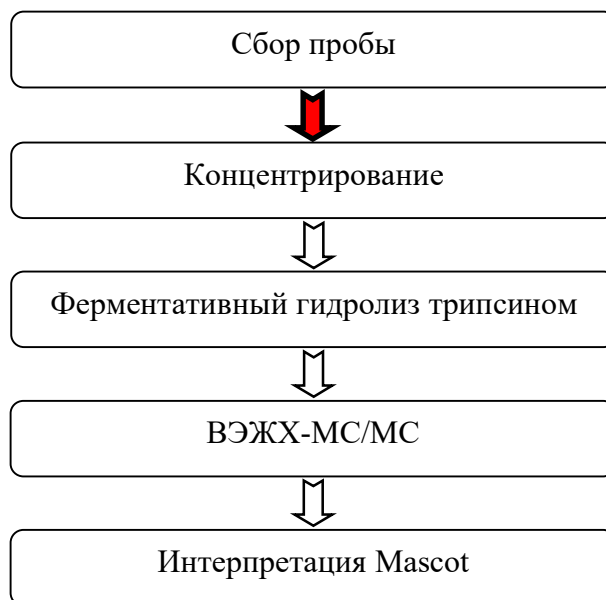


Рисунок 5. Схема анализа с учетом особенностей КВВ.

5.2.3 Сравнительный анализ методик концентрирования при пробоподготовке КВВ для масс-спектрометрического анализа [Kurova et al., 2009]

Использовались четыре методики, позволяющие сконцентрировать, обессолить или обезжирить белковые препараты (лиофилизацию, осаждение белков на колонке C18, осаждение смесью метанол/хлороформа и осаждение смесью ТХУ/дезоксихолата) на сильно разбавленных водных растворах стандартных смесей белков, содержащих от 5 до 48 белков, в отсутствии и присутствии раствора КВВ. После концентрирования белки подвергали обработке трипсином и анализировали при помощи нанопоточного ВЭЖХ-МС/МС с ионизацией электроспреем, как описано выше (см.Гл.3).

5.2.3.1 Лиофилизация

Анализ стандартных белковых смесей до лиофилизации позволил идентифицировать с высоким покрытием первичной последовательности 40 фмоль каждого белка (таблица 5), что приняли за предел чувствительности данного метода.

Таблица 5. Результаты идентификации белков в водном растворе с концентрацией 40 фмоль до и после лиофилизации.

Белки стандартной смеси	Молекулярный вес, Da	Mascot score		Количество идентифицированных пептидов	
		До высушивания	После высушивания	До высушивания	После высушивания
Цитохром С (лошадиный)	12360,96	174	67	12	2
Миоглобин (лошадиный)	16951,27	254	120	12	6
Альдолаза (из мышц кролика)	39211,28	355	233	23	12
Альбумин (бычий сывороточный)	66429,09	292	118	22	2

После лиофилизации количество идентифицированных пептидов уменьшилось на 50-90 %, тем не менее, были идентифицированы все белки стандартной смеси с высокими значениями Score (таблица 5). Схожие результаты были получены для модельных эквимольных смесей из 48 рекомбинантных белков человека. 43 белка идентифицировали в смеси до лиофилизации, после лиофилизации это количество снизилось до 33 белков (рисунок 6). Количество характеристических пептидов, позволяющих идентифицировать белок, и их индивидуальные скоры, рассчитанные программой Mascot, также уменьшились.

Для проверки влияния матрицы КВВ на идентификацию белков, к водному раствору модельной эквимольной смеси из 48 рекомбинантных белков человека добавили индивидуальный КВВ. После лиофилизации этой смеси идентифицировано 37 белков из

стандартного набора, в противовес 33 идентификациям из аналогичного эксперимента, но без добавления КВВ. Таким образом, показано, что матрица КВВ не снижала количество идентифицируемых белков.

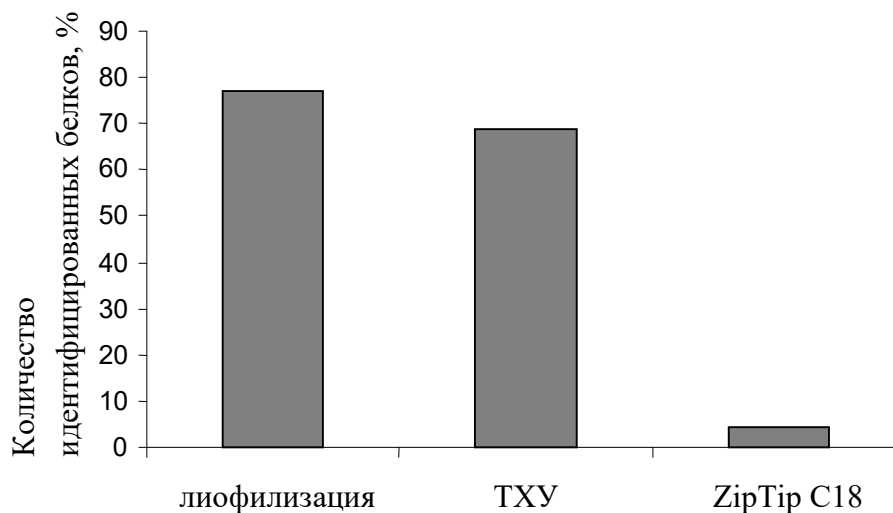


Рисунок 6. Количество белков стандартной смеси 48 рекомбинантных белков человека с добавлением КВВ, идентифицированных после использования различных концентрирующих методик.

5.2.3.2 Осаждение белков ТХУ и концентрирование на колонке C18

После осаждения ТХУ эквимольной модельной смеси 48 рекомбинантных белков человека с добавлением индивидуального КВВ идентифицировано 33 белка смеси. После концентрирования модельной смеси с добавлением КВВ при помощи ZipTip (колонка C18) идентифицировано только 2 белка модельной смеси. Твердофазная экстракция с помощью ZipTip C18 могла бы быть очень полезна при пробоподготовке для ВЭЖХ-МС/МС, особенно в случае КВВ, содержащего много солей и липидов, однако дополнительная стадия пробоподготовки приводит к снижению количества идентифицированных пептидов и, как следствие, к снижению достоверности (Mascot score) идентифицируемого белка. Количество идентифицированных пептидов проанализированных проб модельных смесей после ZipTip C18 было крайне мало, по сравнению с пробами без дополнительной обработки или сконцентрированных с помощью других методик. Идентификация и Mascot score модельных белков также существенно снизились. Результаты сравнения данных методик и лиофилизации представлены в таблице 6 и на рисунке 7.

Таблица 6. Результаты идентификации белков стандартной смеси для разных методов концентрирования

Название белка по базе данных UniProt (код белка в базе)	Количество идентифицированных пептидов				
	стандартная смесь белков	стандартная смесь белков после лиофилизации	стандартная смесь с белков добавлением КВВ		
			лиофилизация	осаждение ТХУ	осаждение на колонке C18
Альфа-лактальбумин (P00709)	4	1	–	–	–
Аннексин А5 (P08758)	17	7	8	3	–
Антитромбин-III (P01008)	20	6	9	15	–
Бета-2-микроглобулин (P61769)	3	2	3	2	–
ВН3 interacting domain death agonist (P55957)	9	1	1	4	–
Карбоангидраза 1 (P00915)	2	–	2	–	–
Карбоангидраза 2 (P00918)	9	–	3	1	–
Каталаза (P04040)	23	8	13	10	–
Катепсин D (P07339)	9	–	7	10	–
Катепсин G (P08311)	12	4	7	10	–
Компонент комплемента C5 (P01031)	4	1	4	3	–
С-реактивный белок (P02741)	5	2	1	–	–
Креатинкиназа М-типа (P06732)	25	11	15	21	–
Цитохром b ₅ (P00167)	4	2	2	–	–
Цитохром с (P99999)	10	1	5	6	–
Фактор роста эпидермиса (P01133)	–	–	–	–	–
Белок, связывающий жирные кислоты (P05413)	5	–	1	1	–
Гельсолин (P06396)	5	2	–	–	–
Глутатион S-трансфераза А1 (P08263)	16	8	8	10	–
Глутатион S-трансфераза Р (P09211)	15	1	8	8	–
ГТФаза Hras (P01112)	2	–	–	–	–
Гемоглобин альфа (P69905)	–	–	–	–	–
Гемоглобин бета (P68871)	2	1	1	–	1
Гистидил-тРНК синтетаза (P12081)	28	3	14	27	–
Инсулинподобный фактор роста II (P01344)	–	–	–	–	–
Интерлейкин-8 (P10145)	4	–	–	–	–
Лактоферрин (P02788)	6	2	5	2	–
Лептин (P41159)	4	–	–	–	–
Лизоцим С (P61626)	1	–	–	–	–
Тау-белок, ассоциированный с микротрубочками (P10636)	22	18	17	17	–
Миоглобин (P02144)	11	3	8	5	–
НАДФ Н дегидрогеназа 1 (P15559)	4	2	2	1	–
Неддин (Q15843)	6	–	2	1	–
пептидилпролил-транс изомеразы А (P62937)	13	2	6	11	–
Пероксиредоксин 1 (Q06830)	13	9	11	12	–
Тромбоцитарный фактор роста (P01127)	–	1	1	–	–
Ретинол-связывающий белок (P02753)	7	3	4	8	–
Рибозилгидроникотинамиддегидрогеназа (хинон) (P16083)	–	–	–	–	–
Серотрансферрин (P02787)	23	14	13	23	–
Сывороточный альбумин (P02768)	2	1	3	28	2
Малый ибиквитинзависимый модификатор 1 (P63165)	4	2	2	2	–
Супероксиддисмутаза (P00441)	4	2	4	3	–
Тиоредоксин (P10599)	4	4	4	4	–
Фактор некроза опухолей (P01375)	4	–	nd	7	–
Убиквитин (P62988)	9	4	5	3	–
Убиквитин-конъюгирующий фермент E2 I (P63279)	9	1	4	6	–
Убиквитин-конъюгирующий фермент E2 С (O00762)	10	3	5	9	–
Убиквитин-конъюгирующий фермент E2 E1 (P51965)	4	–	3	5	–

Показано, что наибольшее количество белков идентифицировалось после лиофилизации (рисунок 6). Осаждение ТХУ показало близкие результаты, однако эта методика является более трудоемкой и не безопасной для здоровья, чем лиофилизация. Осаждение белков на колонке С18 позволило идентифицировать только 2 белка модельной смеси, что говорит и неприменимости данного метода к нашей задаче. По результатам проведенного исследования, не рекомендовано использовать ZipTip С18 для пробоподготовки КВВ для ВЭЖХ-МС/МС анализа.

5.2.3.3 Обессоливание и обезжиривание смесью метанол/хлороформ

Исследовано осаждение белков смесью метанол/хлороформ. Данная методика могла бы быть полезна для данного вида проб, поскольку позволяет избавиться от излишков солей и липидов. Обработка модельной смеси белков смесью метанол/хлороформ/вода в соотношении 7.5/7.5/1 позволила получить наибольшее количество идентифицированных пептидов в сравнении с пробами без дополнительной обработки и пробами после лиофилизации, метода, показавшего наилучшие результаты в сравнении с другими концентрирующими методиками (таблица 7 и рисунок 7). Предполагается, что это результат денатурации белков в органических растворителях.

Таблица 7. Идентификация белков стандартной смеси после осаждения смесью этанол/хлороформ.

Стандартные белки	Молекулярный вес, Da	Mascot score	Количество идентифицированных пептидов
Цитохром С (лошадиный)	12360.96	424	17
Миоглобин (лошадиный)	16951.27	462	13
Альдолаза (из мышц кролика)	39211.28	1622	26
Альбумин (бычий сывороточный)	66429.09	1798	29

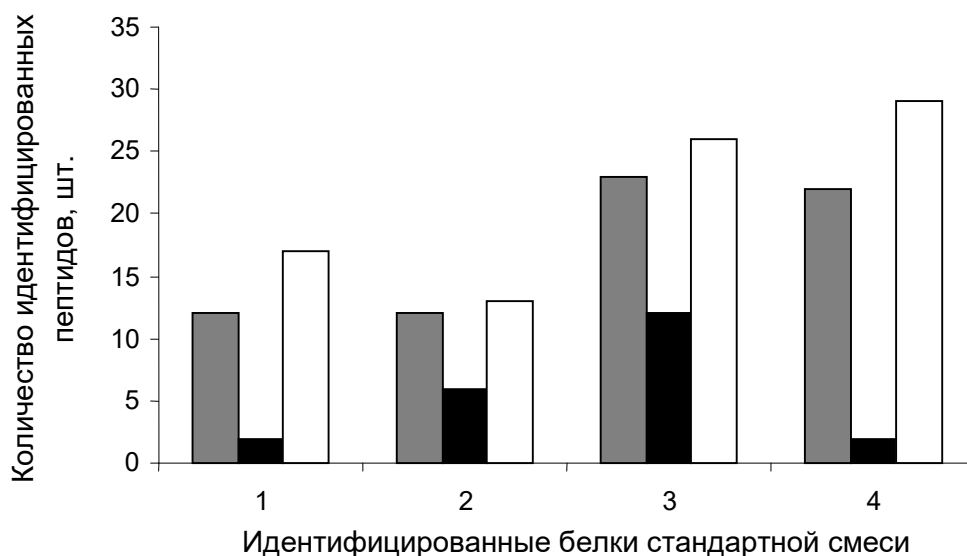


Рисунок 7. Количество идентифицированных пептидов в пробах без дополнительной обработки (серый), пробах после лиофилизации (черный), обработанных смесью метанол/хлороформ (белый). 1 – цитохром С (лошадиный), 2 - миоглобин (лошадиный), 3 - альдолаза (из мышц кролика), 4 - альбумин (бычий сывороточный).

К сожалению, несмотря на эффективность данной процедуры для идентификации белков, ее применение имеет ограничения, связанные с объемом пробы и концентрацией белка. В силу сильной разбавленности КВВ, данная методика подходит только в качестве дополнительного этапа пробоподготовки после концентрирования.

5.2.4 Ограничения ферментативного гидролиза трипсином

Некоторые белки не были идентифицированы вовсе: инсулин, эпидермальный фактор роста, альфа-цепь гемоглобина, инсулиноподобный фактор роста II, тромбоцитарный фактор роста (В-цепь), рибозилгидроникотинамиддегидрогеназа (NQO2). Данный результат можно объяснить двумя факторами: неполный гидролиз белков или осложненная ионизация пептидов во время МС-анализа трипсинового гидролизата. Хорошо известно, что инсулиноподобный фактор роста II, как и инсулин, устойчив к гидролизу трипсином при нормальных условиях по причине ковалентной димеризации [Zumstein et al., 1985]. Мы показали при помощи электрофореза и масс-спектрометрии, что инсулин устойчив к гидролизу трипсином при нормальных условиях и гидролизовался только после дополнительного восстановления/алкилирования. В случае NQO2, гидролизу могут препятствовать многочисленные карбоксикислотные группы и ароматические и гидрофобные основания, находящиеся вблизи сайтов гидролиза, что снижает специфичность трипсина к белковому

субстрату. Помимо белков с не подходящей для гидролиза трипсином структурой, некоторые белки могут быть потеряны в результате лиофилизации.

5.3 Выводы главы

Низкие концентрации белка и сложная разнообразная матрица КВВ требует оптимизации пробоподготовки, чтобы уменьшить потери белка. Кроме того, объем пробы КВВ не превышает 3 мл для здоровых доноров, при этом разбавленность образцов является значительной проблемой при оценке биомаркеров из КВВ: концентрации белков малы, а дополнительное разбавление, которое может быть вызвано избыточной влагой в выдыхаемом воздухе или слезью, может снизить ее до неопределяемой величины. Таким образом, необходимость концентрирования очевидна. Несмотря на снижение количества идентифицированных пептидов, по результатам данного исследования признано, что лиофилизация является наиболее мягким, эффективным и экологичным методом концентрирования проб КВВ перед протеомным анализом, в сравнении с такими методами, как осаждение белков ТХУ, осаждение белков на колонке С18 и осаждение белков смесью метанол/хлороформ.

Глава 6. Исследование белкового состава КВВ здоровых доноров. Анализ кератинового фона окружающего воздуха, а также КВВ, собранных с помощью защитного фильтра

Выводы этой главы были опубликованы в статьях [Kurova et al., 2009] и [Kurova, Kurochkin et al., 2009].

6.1 Исследование белкового состава КВВ здоровых доноров [Kurova et al., 2009]

Сбор КВВ здоровых доноров осуществляли с помощью обеих систем (КВВ 20 человек собрали с помощью EcoScreen, КВВ 5 человек собрали с помощью R-Tube, КВВ 6 человек собрали с помощью обоих устройств примерно в одно и то же время). Проводили сравнительный анализ списков белков, получаемых посредством сбора ECoScreen и R-Tube, и было показано, что устройство сбора не влияет на итоговый результат по идентифицируемым программой Mascot белкам.

Исследование КВВ здоровых доноров выявило, что основными белками проб являлись кератины, чей спектр, однако, отличался полиморфизмом для разных людей. Пары цитоскелетных кератинов 1/10 и 2/9 были инвариантными для всех проб. Не обнаружено никаких мутаций в последовательностях этих белков у здоровых доноров. Также к часто встречающимся в пробах можно отнести цитоскелетные кератины II типа (3, 4, 5, 6) и цитоскелетные кератины I типа (14, 15, 16). В то же время, набор кератинов и их изоформ достаточно ощутимо различался для индивидуальных проб. Особенно интересны изоформы кератинов различных типов. Например, в индивидуальных пробах нашлись пептидные фрагменты, идентифицирующие кератин 1В (кератин 77), кератин 28, кератин 2p, а также несколько изоформ кератина 6 (6 A, 6B, 6 C, 6 F, 6 irs 3). Таким образом, можно предположить, что разные люди отличаются разным уровнем экспрессии кератинов в дыхательных путях и, вероятно, «кератиновым полиморфизмом» эпителия.

Помимо кератинов, с высокой частотой в пробах идентифицировали дермцидин (известен как белок с антибактериальной активностью, производимый потовыми железами [Ghosh et al., 2012; Ghosh, Jana et al., 2012]), простагландин-Н₂ D-изомеразу (PGDS2), предшественник альфа-1-микроглобулина/бикунина (АМВР), убиквитин и цистатин А (эти белки были обнаружены в КВВ ≥ 50 % доноров).

6.2 Интерпретация белкового состава КВВ здоровых доноров с помощью MetaCore [Kurova et al., 2009]

Цистатин А был первоначально охарактеризован как ингибитор лизосомальных цистеиновых протеаз – катепсинов. Катепсины задействованы в процессинге и презентации антигенов, а также в ряде патологических состояний, таких как воспаление и рак. Тем не менее, в настоящее время предложена еще одна функция цистатинов: они индуцируют фактор некроза опухоли и синтез интерлейкина 10, а также стимулируют продуцирование оксида азота [Kopitar-Jerala et al., 2006].

Простагландин-Н₂ D-изомераза совместно с простагландин-Н-синтазой (циклооксигеназой) участвует в образовании простагландина D₂ из арахидоновой кислоты в тучных клетках [Tanaka et al., 2000] и Th2 клетках (Т-хелперах) [Matsuoka et al., 2000]. Простагландин D₂ является медиатором аллергических реакций и воспаления [Ichikawa et al., 2007; Brannian et al., 1997], активируя два разных типа рецепторов и вызывая сокращение гладкой мускулатуры дыхательных путей [Hirai et al., 2001]. Также он участвует в хемотаксисе эозинофилов и базофилов внутрь легких [Spik et al., 2005]. Таким образом, простагландин D₂ координационно, посредством двух рецепторов, регулирует аллергические реакции, особенно воспаление респираторного тракта [Urade et al., 2000]. Простагландин Н₂ D-изомераза участвует в аллергических и воспалительных реакциях [Psaty et al., 2005] и может быть хорошей мишенью для противоаллергической и противовоспалительной терапии.

Предшественник альфа-1-микроглобулина/бикунина представлен в крови двумя разными белками. Альфа-1-микроглобулин можно обнаружить в крови и соединительной ткани многих органов. Наиболее часто встречается там, где клетки граничат с окружающей средой, например, в легких, кишечнике, почках и плаценте. Альфа-1-микроглобулин ингибирует иммунологическую функцию белых клеток крови *in vitro*, и его распределение связывается с противовоспалительной и защитной функцией *in vivo* [Akerström et al., 2000]. Бикунин является ингибитором протеиназ Куница и ответственен за большую часть антитриптической активности мочи, поэтому также известен как ингибитор мочевого трипсина. Экскреция бикунина усиливается в условиях воспаления, поэтому его считают белком острой фазы воспаления. Тем не менее, экспрессия гена бикунина подавлена при воспалении. В плазме крови человека основная часть бикунина ковалентно связана с одним или двумя гомологичными тяжелыми цепями пептида, формируя таким образом высокомолекулярные ингибиторы протеиназ. Определение бикунина в моче дает информацию о тяжести системного протеолиза, происходящего во время воспаления. Было также показано, что при

воспалительных заболеваниях концентрация высокомолекулярных производных бикунина в плазме крови зависит как от их усиленной утилизации, так и от регуляции их биосинтеза [Mizon et al., 2002].

Дермцидин, известный как белок с антибактериальной активностью, производимый потовыми железами, вероятно, является «экзогенным» белком, поскольку он также обнаружен в конденсированном воздухе рабочего помещения (см. раздел 6.4), тем не менее, нельзя исключать, что часто встречающиеся в пробах КВВ здоровых доноров белки могли бы играть определенную диагностическую роль, особенно при совмещении качественного и количественного анализов.

Большинство белков выдыхаемого воздуха, согласно проведенному анализу с помощью программы MetaCore, были ассоциированы с ростом тканей и их регенерацией, например, долгосрочное укрепление нервно-мышечного соединения (p value $5,01 \times 10^{-6}$), регуляция синаптического роста к нервно-мышечным соединениям (p value $8,35 \times 10^{-6}$), развитие эпидермиса и эктодермы (p values 1×10^{-5}). Мы не обнаружили заболеваний, ассоциированных одновременно с значительным количеством белков, идентифицированных в КВВ здоровых доноров, большинство же «обнаруженных» заболеваний скорее всего отсутствовали у доноров на практике (к примеру, заболевания новорожденных). Таким образом, эти результаты подтверждают, что доноры могут быть признаны относительно здоровыми. Тем не менее, такие белки, как предшественник альфа-1-микроглобулина/бикунина, простагландин-Н2 D-изомеразы и цистатин А могут играть важную роль в диагностике респираторных заболеваний.

6.3 Анализ белкового состава объединенной пробы конденсатов выдыхаемого воздуха здоровых доноров [Kurova et al., 2009]

В ходе анализа индивидуальных проб КВВ здоровых доноров идентифицировали несколько десятков белков с высокими значениями Mascot Score, большую часть из которых составляют цитоскелетные кератины. Однако в списках выдачи Mascot присутствовали белки, чей низкий Score не позволил программе достоверно их идентифицировать, но это могло произойти по причине их малой концентрации в пробе. Чтобы повысить концентрацию минорных белков пробы, конденсаты выдыхаемого воздуха 18 здоровых некурящих доноров были смешаны (объединены). Одновременное увеличение количества мажорных белков (кератинов), солей и липидов в объединенной пробе КВВ может затруднить проведение трипсинолиза и масс-спектрометрического анализа, что следует отнести к недостаткам такого

подхода. Три различных методики (варианты А-В в разделе «Материалы и Методы») были испробованы с целью оптимизации анализа смесей КВВ.

В варианте А десять конденсатов после гидролиза трипсином были объединены в одну пробу и высушены. Пробу анализировали посредством закалывания в колонку 1 мкл и максимально возможного объема шприца, равного 7 мкл. Ожидали, что количество общих для всех проб компонентов должно возрасти в пять и тридцать пять раз соответственно. Путем анализа пептидных карт смеси с помощью программы Mascot и последовательностей в базе данных NCBI nr (human) смогли идентифицировать тридцать шесть белков. Программа Mascot идентифицирует белки, формируя наборы фрагментов характеристических пептидов. При получении общего Score белка принимается во внимание количество найденных характеристических пептидов и индивидуальный Score каждого пептида. Двадцать семь белков однозначно идентифицировали в соответствии с их уникальными наборами пептидных фрагментов. Несколько кератинов не смогли различить с их изоформами и гомологами, поскольку были с высокой вероятностью идентифицированы наборы пептидов, которые не содержали уникальных фрагментов.

Помимо кератинов, также достоверно идентифицировали альбумин, пролин-обогащенный белок 4 (слезный), иммуноглобулин IGHA1, цистатин А, белок-активатор тирозин-3/триптофан-5 – монооксигеназы, предшественник липокалина 1, S100 кальцийсвязывающий белок A9, десмоплакин. Этот набор белков был получен в ходе анализа максимально возможного в данных условиях объема пробы. В случае семикратного уменьшения вкалываемого объема пробы часть белков из этого списка не идентифицировалась вовсе. Это говорит о том, что пробы КВВ содержат много белков, чьи концентрации лежат ниже или равны чувствительности приведенного метода анализа.

В варианте Б половины проб индивидуальных КВВ пятнадцати доноров были смешаны и проанализированы. Кроме кератинов, идентифицировали дермцидин, цистатин А, простагландин H2 D-изомеразу человека, альфа-1-микроглобулин, гепарансульфатпротеогликан, убиквитин и супрабазин. Эти же белки, кроме супрабазина, идентифицировали в смеси после обезжиривания метанол/хлороформом (вариант В). Примечательно, что обезжиривание существенно увеличивает количество триптических фрагментов кератинов. Такое же улучшение было достигнуто со стандартными белками: цитохромом С, миоглобином, альдолазой, альбумином (таблица 8). Пятнадцать некератиновых белков КВВ, идентифицированных во всех объединенных пробах, представлены в таблице 9.

Таблица 8. Определение белковых стандартов после осаждения смесью метанол/хлороформ

Белок	Mascot score (количество идентифицированных пептидов)	
	без осаждения смесью метанол/хлороформ	после осаждения смесью метанол/хлороформ
цитохром С (лошадиный)	172 (12)	424 (17)
апомиоглобин (лошадиный)	254 (12)	462 (13)
альдолаза (из мышц кролика)	355 (23)	1622 (26)
альбумин (из плазмы быка)	292 (22)	1798 (29)

Некоторые белки, идентифицированные в объединенных пробах КВВ, наблюдались и в индивидуальных конденсатах, но со значительно меньшим значением Score (достоверность идентификации по алгоритму Mascot). Вероятно, увеличение значения Score было связано с увеличением концентрации белка в объединенной пробе. В то же время, если белок, присутствовавший в индивидуальном КВВ, не определялся ни в одной из объединенных проб, можно предположить, что его концентрация уменьшалась относительно других белков в смеси. В результате, спектры МС/МС таких пептидов становились беднее из-за ионного экранирования мажорными пептидами и давали меньшие значения Score для пептидов и белков. Тем не менее, исследовать белковый состав как индивидуальных проб КВВ, так и их смесей необходимо для получения исчерпывающей информации о них.

Таблица 9. Некератиновые белки, идентифицированные в объединенных пробах КВВ

Белок	количество идентифицированных пептидов	Белок	количество идентифицированных пептидов
Дермцидин	7	Кальцийсвязывающий белок А9	1
Альфа-тубулин	2	Десмоплакин	1
Сывороточный альбумин	4	Супрабазин	2
Пролинобогатенный (слезный) белок 4	2	ПростагландинН2-D-изомераза человека	1
Цистатин А	1	Гепарансульфатпротеогликан	1
Тирозин 3/триптофан 5 – монооксигеназа	1	Предшественник липокалина 1	2
Убиквитин	2	IGHA1	2
АМВР	3		

Сопутствующую проблему увеличения концентрации солей и липидов в пробе, мешающих масс-спектрометрическому анализу, можно решать стандартной методикой обессоливания и обезжиривания смесью метанол/хлороформ.

6.4 Анализ кератинового фона окружающего воздуха, а также КВВ, собранных с помощью защитного фильтра [Kurova et al., 2009]

Ряд кератинов, присутствующих в КВВ, относят к кожному эпителию и волосяным луковицам [Hoffmann et al., 2008], поэтому можно предположить, что они имеют экзогенное происхождение. Чтобы установить природу происхождения этих кератинов, использовано два подхода. Во-первых, были собраны КВВ двух доноров, которые дышали через защитный фильтр. Защитный фильтр представлял собой самодельное приспособление, которое должно было очистить вдыхаемый воздух от пыли – основного источника кератинов. Проведено сравнение полученных белковых составов КВВ при использовании анти-пылевого фильтра и без него между собой и не обнаружено различия. Белки, идентифицированные в КВВ, собранном с использованием защитного фильтра и без него, совпали (таблица 10.), если не считать небольшой разницы в значениях Score и количестве идентифицированных пептидов. Различия, более вероятно, следует отнести к особенностям ВЭЖХ-МС/МС-анализа, нежели к методу сбора конденсата. По этой причине, можно предположить, что экзогенные кератины накапливаются на выстилке дыхательных путей в течение всей жизни человека. Исходя из этого заключения, кератиновый состав КВВ должен зависеть от условий жизни и возраста человека. Кроме того, дыхание с фильтром оказалось несколько затруднено, поэтому такая процедура не может быть рекомендована пациентам с затрудненной вентиляцией легких или находящимся на искусственной вентиляции легких.

Таблица 10. Белки, идентифицированные в пробах КВВ, собранных с использованием защитного фильтра и без него.

Идентифицированный белок	Без фильтра	С фильтром
gi 17318569 Кератин 1	+	+
gi 40354192 Кератин 10	+	+
gi 55956899 Кератин 9	+	+
gi 47132620 Кератин 2	+	+
gi 908805 Кератин тип II	+	+
gi 4557701 Кератин 17	+	+
gi 181573 Цитокератин 8	+	+
gi 5031839 Кератин 6A	+	+
gi 15431310 Кератин 14	+	+
gi 18999435 Кератин 5	+	+
gi 1195531 Кератин 16	+	+
gi 24430190 Кератин 15	+	+
gi 31074629 Кератин 2p	+	+
gi 4502067 Альфа-1-микроглобулин/бикунин	+	+
gi 16751921 Дермцидин	+	+

Второй подход к установлению происхождения кератинов в КВВ заключался в исследовании белкового состава окружающего воздуха. С помощью установки, включающей помпу, прогоняющую воздух комнаты через емкость с водой, и конденсирующий блок, собран и сконденсирован воздух лабораторной комнаты. 1,5 мл и 5 мл конденсата были собраны за 5 часов и 20 часов, соответственно. Белковый состав полученных проб представлен в таблице 11. Идентификация ряда кератинов в окружающем воздухе позволила заключить, что они могут появляться в респираторном тракте извне и, вероятно, частично удерживаются респираторной системой. Выдыхаемые белки могут накапливаться в пыли помещения и вдыхаться людьми снова. К таким «свободноциркулирующим» белкам следует относиться с осторожностью при анализе белкового состава с точки зрения диагностической значимости и применения анализа КВВ для диагностики.

Таблица 11. Белки конденсированного воздуха лабораторной комнаты

Идентифицированный белок	Score	Найденные пептиды
А: 5 ч (1.5 мл)		
gi 114667176 Кератин, тип I цитоскелетный 14	244	10
gi 17318569 Кератин 1	176	8
gi 547754 Кератин, тип II цитоскелетный 2 эпидермальный	136	3
В: 20 ч (5 мл)		
gi 1346343 Кератин, тип II цитоскелетный 1	181	5
gi 122513 Гемоглобин субъединица бета-1	83	1

Были использованы пустые пробы (обработка реакционным буфером пустых пробирок) и БСА в качестве контроля чистоты и полноты трипсинолиза. Получены хорошие пептидные карты для БСА, ни одного пептида не было найдено в пустых пробах.

6.5 Выводы главы

Определено, что основными белками конденсата выдыхаемого воздуха являются цитоскелетные кератины. Кератины играют основную роль в поддержании целостности и механической стабильности как отдельных эпителиальных клеток, так и (посредством межклеточных контактов) эпителиальных тканей. Было идентифицировано 54 функциональных кератиновых гена. Также идентифицирован ряд некератиновых белков в КВВ здоровых доноров. Этот набор белков, включающих как кератины, так и некератиновые белки, может служить белковым фоном КВВ для дальнейших исследований. Знание фоновых белков поможет при поиске биомаркеров различных заболеваний в КВВ.

Исследование белков, циркулирующих в окружающем воздухе, также очень важно. Определение ряда белков (в основном, кератинов) в окружающем воздухе позволило предположить, что некоторые белки могут попадать в респираторный тракт экзогенным путем и как задерживаться в респираторной системе, так и возвращаться в окружающую среду и идентифицироваться с помощью использованной методики. Такие белки можно назвать «свободно циркулирующими». К таким «свободноциркулирующим» белкам следует относиться с осторожностью при анализе белкового состава с точки зрения диагностической значимости и применения анализа КВВ для диагностики.

Глава 7. Исследование белкового состава КВВ пациентов с ХОБЛ, пневмонией и раком легкого

Был проведен сравнительный анализ белкового состава КВВ здоровых доноров (курящих и некурящих) и белкового состава КВВ пациентов со следующими легочными заболеваниями: ХОБЛ, пневмония, рак легкого. Всего в данном исследовании приняло участие 108 человек, подробные характеристики доноров приведены в Главе 3.

Объем КВВ доноров с различными патологиями респираторной системы варьировался в пределах 0,2–1,5 мл.

В результате предварительного анализа образцов КВВ больных ХОБЛ методом двумерного электрофореза было выявлено, что белковый состав представлен изоформами кератина (рисунок 8), в то время как масс-спектрометрия позволила идентифицировать в образцах пептиды широкого ряда других белков.



Рисунок 8. Двумерный электрофорез смеси образцов КВВ больных ХОБЛ.

Ввиду сложностей, связанных с малыми объемами проб, особенно для пациентов с раком легкого, а также малой репрезентативностью проведенных 2D-электрофорезов проб для целей исследования, основной анализ базировался на результатах ВЭЖХ-МС/МС проб КВВ. В ходе анализа, на основании предварительных экспериментов на пробах КВВ здоровых доноров, считали, что белок достоверно идентифицирован, если для него нашлось не менее двух уникальных пептидов ($\text{Score} > 70$) у одного из доноров или если данный белок при наличии не менее одного уникального пептида ($\text{Score} > 30$) нашелся у нескольких доноров рассматриваемой группы.

По результатам масс-спектрометрического анализа, инвариантными для всех проб являлись цитоскелетные кератины II типа (1, 2, 3, 4, 5, 6) и цитоскелетные кератины I типа (9, 10, 14, 15, 16). В предварительных экспериментах в данном исследовании и другими авторами было показано, что цитоскелетные кератины являются основными белковыми компонентами КВВ как курильщиков, так и некурящих здоровых людей [Hoffmann et al., 2008; Kurova et al., 2009]. Согласно каталогу белков человека, цитокератины СК 1, 2, 9 и 10 имеют эпидермальное происхождение, поэтому считается, что данные белки либо вносятся при пробоподготовке, либо имеют экзогенное происхождение, т.е. не относятся к белкам дыхательных путей. В исследовании [Kurova et al., 2009; Kurova et al., 2011] наша группа предположила, что эти экзогенные кератины в КВВ являются свободно циркулирующими белками воздуха. При анализе КВВ доноров с различными патологиями дыхательной системой другие авторы также отмечали присутствие кератинов, обозначенных выше [Cheng et al., 2011; Fumagalli et al., 2012], хотя при этом нельзя исключить, что уровень экспрессии ряда кератинов и, следовательно, количество белка для здоровых доноров и доноров с различными заболеваниями может различаться [Moll et al., 2008]. К сожалению, такой сложный объект для исследования, как КВВ, сильно ограничивает возможности использования количественных методов анализа [Cheng et al., 2011].

Выводы этой главы были опубликованы в статьях [Кононихин и др., 2015], [Федорченко и др., 2016] и [Анаев и др., 2017].

7.1 Результаты анализа КВВ пациентов с диагнозами ХОБЛ и внебольничная пневмония [Кононихин и др., 2015]

Анализ белкового состава КВВ всех представленных групп доноров выявил более 300 различных белков (рисунок 9).

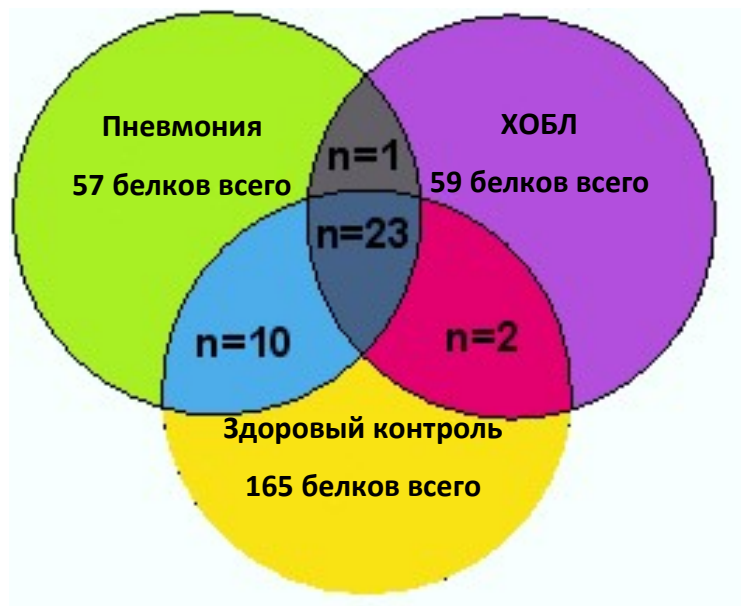


Рисунок 9. Диаграмма Венна для белков, идентифицированных в пробах КВВ контрольной группы и пациентов с диагнозами «ХОБЛ» и «внебольничная пневмония».

Далее приводятся результаты аннотирования и анализа групп здорового контроля и групп с диагнозами ХОБЛ и пневмония по белкам, не относящимся к кератинам, с использованием биоинформатических баз данных. Полный список не кератиновых белков, идентифицированных в пробах как контрольной группы, так и в группах с диагнозами ХОБЛ и пневмония, приведен в таблице 12 [Кононихин и др., 2015].

Таблица 12. «Общие» белки, идентифицированные в протеомах КВВ как контрольной группы, так и групп доноров с диагнозами ХОБЛ и пневмония.

Функция	Аннотация белка									
	Ген	Название белка ¹	M ¹ , Da	Семейство ¹	Кл. локализация ²		Экспрессия в тканях ^{3/4}			Вероятная связь с клин. картиной ⁶
					Вне-	Внутри-	Кровь	Кожа	Л&РТ ⁵	
Ферменты	PTGDS	Prostaglandin-H2 D-isomerase	21029	lipocalin	+	+	+/+	+/+	-/+	-
	PIP	Prolactin-inducible protein	16572	PIP	+	-	+/+	+/+	+/+	-
Регуляторные	AMBP	Alpha-1-Microglobulin/Bikunin Precursor	38999	lipocalin / -	+	+	+/+	+/+	-/+	-
	CSTA	Cystatin-A	11006	cystatin	-	+	+/+	+/+	+/+	-
	DMKN	Dermokine	47082	dermokine	+	-	+/+	+/+	-/+	-
	KNG1	Kininogen-1	71957	-	+	+	+/+	+/+	-/+	-
Транспортные	AGP2 (ORM2)	Alpha-1-Acid Glycoprotein 2	23603	lipocalin	+	-	+/+	+/+	-/+	-
	ALB	Serum albumin	69367	ALB / AFP / VDB	+	+	+/+	+/+	+/+	-
	LCN1	Lipocalin-1	19250	lipocalin	+	-	+/+	+/+	+/+	-
Структурные	DSP	Desmoplakin	331774	plakin	-	+	+/+	+/+	+/+	-
	FLG2	Filaggrin-2	248073	S100-fused protein	-	+	+/+	+/+	+/+	-
	HRNR	Hornerin	282390	S100-fused protein	+	+	+/+	+/+	+/+	-
	JUP	Junction plakoglobin	81745	beta-catenin	-	+	+/+	+/+	+/+	-
	SHROOM3	Protein Shroom3	216857	shroom	-	+	+/-	+/-	-/-	-
Защитные	AZGP1	Zinc-alpha-2-glycoprotein	34259	MHC class I	+	-	+/+	+/+	+/+	-
	CDSN	Corneodesmosin	51522	-	+	+	+/+	+/+	-/+	-
	DCD	Dermcidin	11284	-	+	-	+/+	+/+	+/+	-
	IGHA1	Immunoglobulin Heavy Constant Alpha 1	37655	-	+	-	+/+	+/+	+/+	-
	LYZ	Lysozyme C	16537	glycosyl hydrolase 22	+	+	+/+	+/+	+/+	-
	SPP1	Osteopontin	35423	osteopontin	+	-	+/+	-/+	-/+	-
Сократительные	ACTG1	Actin, cytoplasmic 2	41793	actin	+	+	+/+	+/+	+/+	-
	DNAH14	Dynein heavy chain 14, axonemal	399895	dynein heavy chain	-	+	+/+	+/+	-/+	-

¹ – Согласно базе данных UniProt

² – Согласно базе данных GeneCards

^{3/4} – Согласно базе данных MOPED/BioGPS

⁵ – Легкие&Респираторный Тракт

⁶ – Биологические/Патологические Процессы согласно GO/Quagen™. Расширенная информация представлена в таблице I-II в приложении

Как можно видеть из приведенных данных (таблица 12), 23 белка оказались характерны для всех рассматриваемых групп доноров. Среди них наибольшей частотой встречаемости в

пробах характеризовался дермцидин – белок-антибиотик, обладающий антибактериальной и протеолитической активностями (рисунок 10).

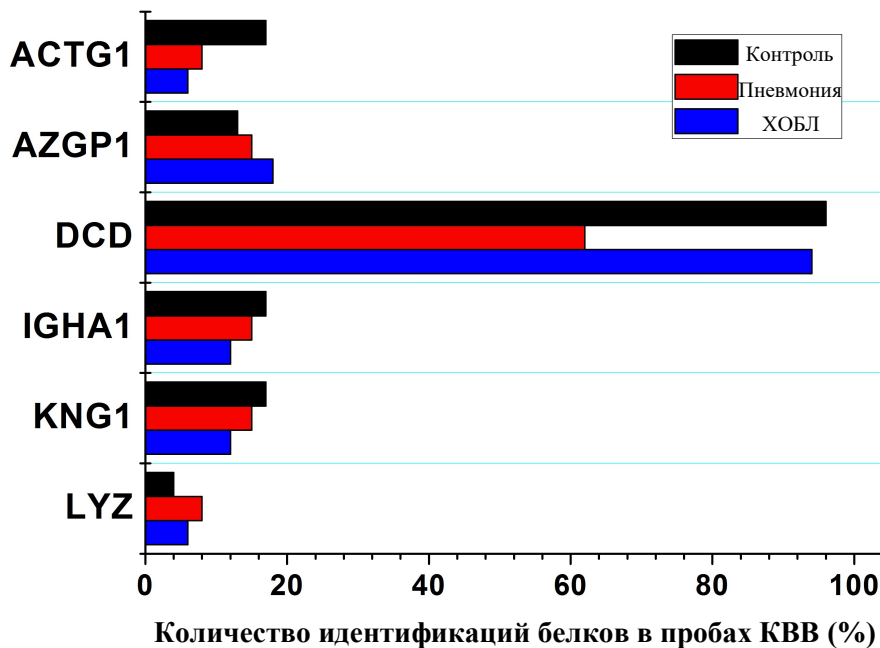


Рисунок 10. Процентное соотношение количества идентификаций белков, характерных как для протеомов КВВ здорового контроля, так и для групп доноров с диагнозами ХОБЛ и пневмония. Белки с приблизительно одинаковой частотой встречаемости и дермцидин.

Известно, что дермцидин секретируется потовыми железами человека [Ghosh et al., 2012; Ghosh, Jana et al., 2012], однако, в ряде исследований было показано, что дермцидин и его производные пептиды обнаруживались также в крови при острой ишемической болезни сердца [Mikhaylova et al., 2014], в плаценте [Cunningham et al., 1998], мозге и нейрональных клеточных линиях [Cunningham et al., 1998]. Согласно базе данных MOPED, экспрессия дермцидина также наблюдается во многих тканях, включая легкие и респираторный тракт. Более того, дермцидин был предложен в качестве вероятного онкогена в исследованиях, посвященных онкологическим заболеваниям [Stewart et al., 2008], также было показано, что он стимулирует пролиферацию опухолевых клеток мыши, крысы и человека [Stewart et al., 2008; Todorov et al., 1996; Park et al., 2010]. В этой связи интересно отметить, что, если в случае с группами здорового контроля и ХОБЛ дермцидин являлся практически инвариантным (частота встречаемости в пробах 96% и 94% соответственно), то для группы с диагнозом пневмония встречаемость этого белка резко упала до 60%. Эти данные безусловно требуют подтверждения дальнейшим набором статистики по заболеванию, возможно, с применением количественных методов определения уровня экспрессии данного белка, и последующего анализа причин снижения процента идентификации дермцидина в случае внебольничной пневмонии. [Кононихин и др., 2015]

Примерно с одинаковой частотой во всех рассматриваемых группах доноров идентифицировались такие белки, как иммуноглобулин альфа, кинниноген, цитоплазматический актин, сывороточный альбумин, цинк-альфа2-гликопротеин, лизоцим (рисунок 10).

На рисунке 11 представлены данные по частоте идентификации в пробах тех белков, которые были идентифицированы как в контрольной группе, так и в фокусных группах по рассматриваемым заболеваниям, но частота встречаемости этих белков в пробах здорового контроля значительно отличалась от групп с диагнозами ХОБЛ и пневмония. Как и в случае вышеописанного дермцидина, эти данные требуют дальнейшего анализа как с точки зрения статистической достоверности, так и с точки зрения анализа причин таких расхождений и оценки возможной диагностической значимости данных белков.

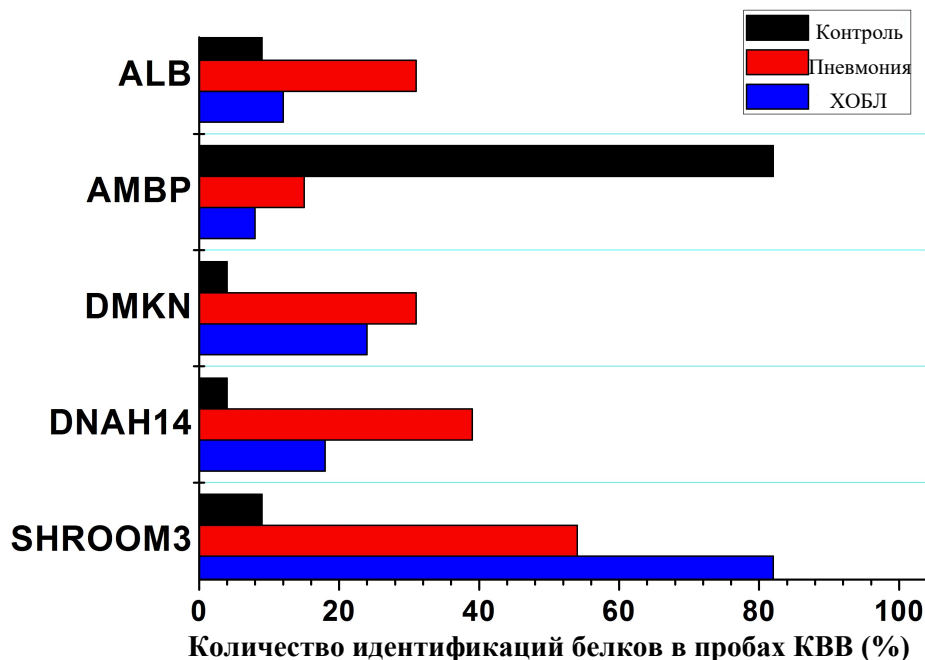


Рисунок 11. Процентное соотношение количества идентификаций белков, характерных как для протеомов КВВ здорового контроля, так и для групп доноров с диагнозами ХОБЛ и пневмония.

Среди представленных данных интересно отметить резкое снижение количества идентификаций предшественника альфа-1-микроглобулина / бикунина с более 80% идентификаций этого белка в пробах группы здорового контроля до порядка 10% идентификаций в пробах групп с диагнозами ХОБЛ и пневмония. Поскольку альфа-1-микроглобулин ингибирует иммунные функции лейкоцитов *in vitro*, а его распределение связывают с защитной и противовоспалительной ролью *in vivo* [Urade et al., 2000], а бикунин репрессируется при воспалении [Mizon et al., 2002], можно предположить, что острое

воспаление, сопровождающее как ХОБЛ, так и пневмонию, может вызывать снижение уровня экспрессии предшественника альфа-1-микроглобулина / бикунина. Уменьшение количества белка, в свою очередь, должно приводить к затруднению его идентификации в пробах КВВ. [Кононихин и др., 2015]

Для белка SHROOM3 наблюдается обратная динамика, характеризующаяся резким возрастанием количества идентификаций в пробах доноров с диагнозами ХОБЛ и пневмония по сравнению со здоровых контролем. Аналогичная обратная динамика (с менее выраженным ростом числа идентификаций) характерна также для динеина и дермокина. Интересен тот факт, что и SHROOM3, и динеин являются компонентами цитоскелета и связаны с микротрубочками. Про дермокин на настоящий момент известно мало. У этого белка открыто 4 изоформы, 3 из которых экспрессируются в коже и участвуют в дифференцировке кератиноцитов, а 4я распространена повсеместно [Leclerc et al., 2011]. Кроме того, есть данные, что уровень экспрессии дермокина повышается при воспалении [Hasegawa et al., 2013].

На рисунке 12 представлены данные по частоте идентификаций белков, характерных для контрольной группы и одной из групп по фокусным заболеваниям, но по каким-то причинам не идентифицированным для другой группы.

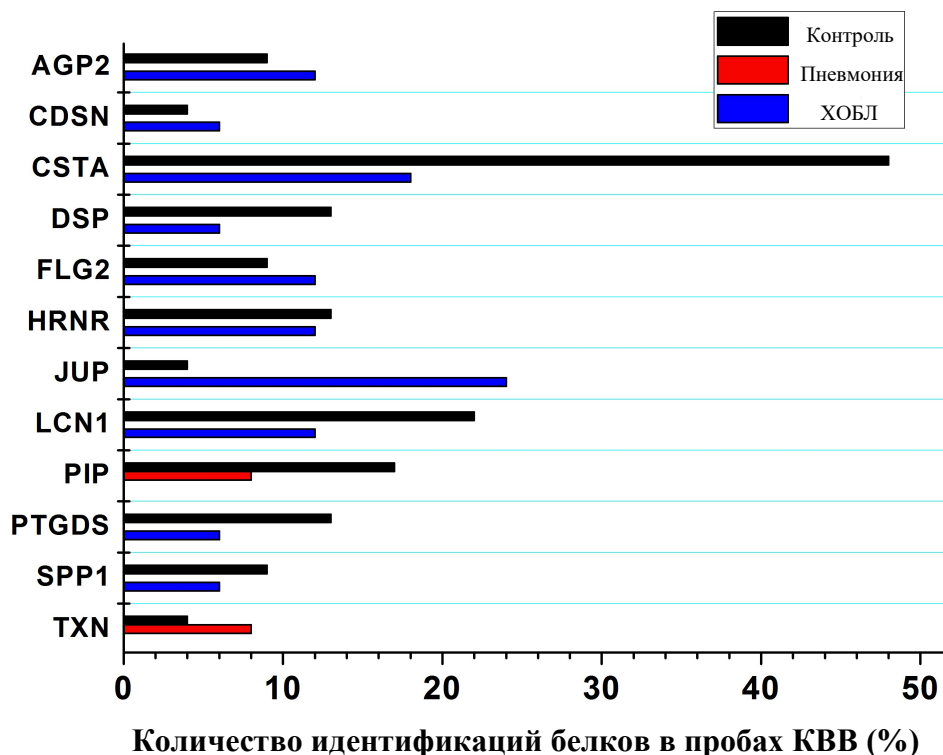


Рисунок 12. Процентное соотношение количества идентификаций белков, характерных для протеомов КВВ здорового контроля и одной из групп с фокусным заболеванием, но не нашедшихся ни у одного из доноров другой группы с фокусным заболеванием.

Как видно из рисунка 12, 10 белков были идентифицированы для контрольной группы и группы с диагнозом ХОБЛ, но отсутствовали в пробах группы с диагнозом пневмония, и 2 белка, напротив, идентифицировались только в контрольной группе и в группе с диагнозом пневмония, но не нашлись ни в одной пробе из группы с диагнозом ХОБЛ.

В таблице 13 приведен список белков, идентифицированных только в пробах группы с диагнозом ХОБЛ, и отсутствовавших в протеомах КВВ здорового контроля и группы с диагнозом пневмония.

Таблица 13. Белки, идентифицированные только в протеомах КВВ группы с диагнозом ХОБЛ.

Функция	Аннотация белка									
	Ген	Название белка ¹	M ¹ , Da	Семейство ¹	Кл. локализация ²		Экспрессия в тканях ^{3/4}			Вероятная связь с клин. картиной ⁶
					Вне-	Внутри-	Кровь	Кожа	Л&Р T ⁵	
Ферменты	ATP1A3	Sodium/potassium-transporting ATPase subunit alpha-3	111749	cation transport ATPase	-	+	+/+	+/+	-/+	-
	HTRA2	HtrA Serine Peptidase 2	48841	peptidase S1B	-	+	+/+	-/+	-/+	-
	NT5C1B	5'-Nucleotidase, Cytosolic IB	68804	5'-nucleotidase type 3	-	+	+/+	-/+	-/+	-
	PDE8B	High affinity cAMP-specific and IBMX-insensitive 3',5'-cyclic phosphodiesterase 8B	98979	cyclic nucleotide phosphodiesterase	-	+	+/+	-/+	-/+	+
	UBA6	Ubiquitin-Like Modifier Activating Enzyme 6	117970	ubiquitin-activating E1	n/a	n/a	+/+	-/+	-/+	-
Регуляторные	EIF2S1	Eukaryotic translation initiation factor 2 subunit 1	36112	eIF-2-alpha	-	+	+/+	+/+	+/+	+
	LACRT	Lacritin	14246	-	+	-	+/+	-/+	-/+	+
	NOC4L	Nucleolar complex protein 4 homolog	58468	CBF/MAK21	-	+	+/+	-/+	+/+	-
	TFAP2D	Transcription factor AP-2-delta	49578	AP-2	-	+	+/+	-/+	-/+	-
	TLE2	Transducin-like enhancer protein 2	79841	WD repeat Groucho/TLE	-	+	+/+	-/+	-/+	-
Рецепторные	CD3EAP	DNA-directed RNA polymerase I subunit RPA34	54986	eukaryotic RPA34 RNA polymerase subunit	-	+	+/+	-/+	+/+	-
	GPR179	G Protein-Coupled Receptor 179	257363	G-protein coupled receptor 3	-	+	+/-	-/-	-/-	-
	ZMYND8	Protein kinase C-binding protein 1	131692	-	-	+	+/+	-/+	-/+	-
Транспортные	HBB	Hemoglobin subunit beta	15998	globin	+	+	+/+	+/+	+/+	+
	HBG1	Hemoglobin subunit gamma-1	16140	globin	-	+	+/+	+/+	-/+	+
	OSBPL8	Oxysterol-binding protein-related protein 8	101196	OSBP	-	+	+/+	-/+	+/+	-
	SHBG	Sex hormone-binding globulin	43779	-	+	+	+/+	+/+	-/+	-
Структурные	COL1A1	Collagen alpha-1(I) chain	138941	fibrillar collagen	+	+	+/+	+/+	+/+	+
Защитные	PRDX1	Peroxiredoxin-1	22110	AhpC/TSA	-	+	+/+	+/+	+/+	+

¹ – Согласно базе данных UniProt² – Согласно базе данных GeneCards^{3/4} – Согласно базе данных MOPED/BioGPS⁵ – Легкие&Респираторный Тракт⁶ – Биологические/Патологические Процессы согласно GO/Quagen™. Расширенная информация представлена в таблице 2-II в приложении

Как видно из приведенных данных (таблица 13), КВВ группы с диагнозом ХОБЛ характеризовались наличием крови (в 2х пробах из 17 был обнаружен гемоглобин) и обширной группой ядерных белков (47% от представленной группы: перечислить против 15 % от группы белков пневмонии), появление которых свидетельствует о процессе разрушения ткани. Кроме того, интересно отметить наличие в КВВ пероксиредоксина в противовес тиоредоксину у контрольной группы и группы с диагнозом пневмония. Известно, что белок тиоредоксин вырабатывается в ответ на внутриклеточный NO и участвует в регуляции апоптоза [Sengupta et al., 2012]. В свою очередь, пероксиредоксин вырабатывается в ответ на активные формы кислорода и окислительный стресс [Cox et al., 2009], что соответствует клинической картине хронической обструктивной болезни легких [Pichavant et al., 2014].

В таблице 14 приведен список белков, идентифицированных только в пробах группы с диагнозом пневмония, и отсутствовавших в протеомах КВВ здорового контроля и группы с диагнозом ХОБЛ.

Таблица 14. Белки, идентифицированные только в протеомах КВВ группы с диагнозом пневмония.

Функция	Аннотация белка									
	Ген	Название белка ¹	M ¹ , Da	Семейство ¹	Кл. локализация ²		Экспрессия в тканях ^{3/4}			Вероятная связь с клин. картиной ⁶
					Вне-	Внутри-	Кровь	Кожа	Л&Р T ⁵	
Ферменты	KMT2D	Histone-lysine N-methyltransferase 2D	593389	Histone-lysine methyltransferase	-	+	+/+	-/+	-/+	-
	PDE4D	cAMP-specific 3',5'-cyclic phosphodiesterase 4D	91115	cyclic nucleotide phosphodiesterase	-	+	+/+	-/+	+/+	+
	PRSS3	Protease, Serine, 3	32529	peptidase S1	+	+	+/+	+/+	+/+	+
Регуляторные	BAG6	Large proline-rich protein BAG6	119409	-	-	+	+/+	+/+	-/+	+
	SMCHD1	Structural maintenance of chromosomes flexible hinge domain-containing protein 1	226374	-	-	+	+/+	-/+	-/+	-
	ZG16B	Zymogen granule protein 16 homolog B	22739	jacalin lectin	+	-	+/+	+/+	-/+	-
Рецепторные	EGFR	Epidermal Growth Factor Receptor	134277	Tyr protein kinase	+	+	+/+	+/+	+/+	+
Структурные	CRNN	Cornulin	53533	S100-fused protein	-	+	+/+	+/+	-/+	-
	*	*	-	histone H2B	-	+	+/+	+/+	+/+	+
	HSP90B1	Heat Shock Protein 90kDa Beta (Grp94), Member 1	92469	heat shock protein 90	-	+	+/+	+/+	+/+	+
	LAMA1	Laminin subunit alpha-1	337084	-	+	-	+/+	-/+	+/+	+
	SPRR3	Small proline-rich protein 3	18154	cornifin (SPRR)	-	+	+/+	-/+	-/+	-
Защитные	ANXA1	Annexin A1	38714	annexin	+	+	+/+	+/+	+/+	+
	ANXA2	Annexin A2	38604	annexin	+	+	+/+	+/+	+/+	+
	CST6	Cystatin-M	16511	cystatin	+	-	+/+	+/+	-/+	+
	CSTB	Cystatin-B	11140	cystatin	-	+	+/+	+/+	+/+	+
	S100A8	Protein S100-A8	10835	S-100	+	+	+/+	+/+	+/+	+
Сократительные	ACTA2	Actin, aortic smooth muscle	42009	actin	-	+	+/+	+/+	+/+	-
**	-	LOC100129958 similar to hCG1643231	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	-	LOC100290955 hypothetical protein XP_002342097	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

¹ – Согласно базе данных UniProt; ² – Согласно базе данных GeneCards

^{3/4} – Согласно базе данных MOPED/BioGPS

⁵ – Легкие&Респираторный Тракт

⁶ – Биологические/Патологические Процессы согласно GO/Quagen™. Расширенная информация представлена в таблице 3-П в приложении

* – Идентифицированы фрагменты белков семейства гистонов H2B

** – Гипотетические белки на основании данных секвенирования генома человека

Как видно из приведенных данных (таблица 14), КВВ группы с диагнозом пневмония характеризовались отсутствием цистатина А, характерного для всех остальных групп доноров, включая контрольную. Взамен этого «нормального» цистатина появляются два других: Цистатин-В и Цистатин-М. [Кононихин и др., 2015]

Известно, что при пневмонии существенно повышается уровень катепсина В [Czitrovszky et al., 2002]. Катепсин В – это цистеиновая протеаза, которую можно найти в лизосомах и которая играет роль в представлении молекул комплекса гистосовместимости второго класса [Zhang et al., 2000]. Цистатины, являясь важными ингибиторами протеаз во внеклеточных жидкостях, контролируют чрезмерную протеолитическую активность катепсинов. Цистатины второго типа (как Цистатин-М) имеют наибольшую аффинность к катепсинам В, Н, L, и S [Werle et al., 2003], также было показано, что возрастание количества цистатина В (цистатин 1 типа) связано с возрастанием количества катепсина-В [Gashenko et al., 2013].

Также для КВВ больных пневмонией было характерно наличие аннексина А1, аннексина А2 и HSP90В – белков, задействованных в процессах воспаления в организме. Экзогенный и эндогенный аннексин А1 является счетчиком-регулятором активности клеток врожденного иммунитета, чтобы достаточный уровень активации был достигнут, но не превышен [Perretti et al., 2009]. Аннексин А2 участвует в позитивной регуляции транскрипции гена COX II, что приводит к выработке простагландинов и воспалительной реакции [Derbyshire et al., 2005]. Было показано, что HSP90b1 является главным шапероном для многих Толл-подобных рецепторов, включая TLR1, TLR2, TLR4, TLR5, TLR6, TLR7, and TLR9 [Yang et al., 2007], многие из которых в свою очередь задействованы в регуляции функции В-клеток во время как физиологических [Pasare et al., 2005], так и патологических процессов [Groom et al., 2007].

Кроме роли при воспалительном процессе, Annexin A2 and HSP90B1 ассоциированы с гипоксией и входят в ПЦР-тест на гипоксию The Human Hypoxia Signaling Pathway RT² Profiler PCR Array (Qiagen). Гипоксия индуцирует быстрое временное увеличение экспрессии генов теплового шока. HSP90B1, являясь одним из них, способствует выживанию клеток путем стабилизации и сворачивания денатурированных в ходе стрессового воздействия белков [Grbovic et al., 2006]. Экспрессия аннексина А2 в клетках также увеличивается в условиях гипоксии [Genetos et al., 2010]. Исходя из вышесказанного, появление аннексина А2 и HSP90B1 в КВВ больных внебольничной пневмонией может свидетельствовать о клеточном ответе на развивающуюся гипоксию, что подтверждает клиническую картину заболевания [Chao et al., 2011].

Еще один интересный результат связан с появлением в КВВ больных внебольничной пневмонией белка ламинина. Ламинины являются гетеротримерными гликопротеинами внеклеточного матрикса, которые экскретируются эпителиальными клетками и которые имеют решающее значение для сохранения нормальной структуры и функции альвеолярной базальной мембраны. Ламинин в базальной мембране имеет поразительно долгий для белка период полураспада (от 30 до 60 дней), вследствие чего можно заключить, что ламинины и, возможно, другие белки базальной мембраны очень медленно меняются во взрослом легком [Urlich et al., 2011]. В этой связи можно предположить, что появление в КВВ фрагментов ламинина связано с его потерей/разрушением в альвеолярной базальной мембране. В свою очередь, потеря ламинина альвеолярным эпителием приводит к возрастанию количества коллагена в легких, который обеспечивает устойчивость к механическому повреждению, и к нарастанию количества клеток, участвующих в развитии воспаления [Urlich et al., 2011].

Из 40 белков, идентифицированных для фокусных групп по рассматриваемым заболеваниям и не характерных для контрольной группы, был выявлен только 1 общий белок: Collagen alpha-1(XVIII) chain, ингибитор пролиферации клеток эндотелия и ангиогенеза. Частота идентификаций этого белка в пробах обеих групп была примерно одинакова и составила, в среднем, 30% от общего числа КВВ для каждой группы. Еще одно общее свойство протеомов ХОБЛ и пневмонии – наличие в их составе цАМФ-специфичных 3',5'-циклических фосфодиэстераз. Эти фосфодиэстеразы (PDE) выступают критическим регулятором функции Т-клеток посредством своей способности гидролизовать внутриклеточный цАМФ, содействуя развитию воспаления [Vang et al., 2010]. Активированные Т-клетки продуцируют фактор некроза опухолей и другие цитокины, которые обычно идентифицируются в выдыхаемом воздухе пациентов с диагнозом ХОБЛ и другими легочными заболеваниями [Liang et al., 2012].

7.2 Результаты анализа белкового состава КВВ больных с диагнозом «рак легкого» [Федорченко и др., 2016]

Полный список не кератиновых белков, идентифицированных в пробах как контрольной группы, так и в группах с ХОБЛ, внебольничной пневмонией и раком легкого, приведен в таблице 15. Как можно видеть из приведенных данных (таблица 15), 21 белок оказался характерен для всех рассматриваемых групп доноров. Среди них наибольшей частотой встречаемости в пробах характеризовался дермцидин. Согласно базе данных MOPED, экспрессия дермцидина наблюдается во многих тканях, включая легкие и респираторный тракт. Более того, дермцидин был предложен в качестве вероятного онкогена в исследованиях,

посвященных онкологическим заболеваниям [Stewart et al., 2008], также было показано, что он стимулирует пролиферацию опухолевых клеток мыши, крысы и человека [Stewart et al., 2008; Todorov et al., 1996; Park et al., 2010].

Таблица 15. Белки, идентифицированные в протеомах КВВ всех групп доноров, включая контрольную

Функция	Описание белка									
	Название гена	Название белка	Mr ¹ , Да	Семейство ¹	Локализация в клетке ²		Экспрессия в тканях ^{3/4}			Вероятная связь с клинической картиной ⁶
					Вне.	Внутр.	Кровь	Кожа	Л&РТ ⁵	
Ферменты	PTGDS	Prostaglandin-H2 D-isomerase	21029	lipocalin	+	+	+/+	+/+	-/+	-
Регуляторные белки	AMBIP	Alpha-1-Microglobulin/Bikunin Precursor	38999	lipocalin / -	+	+	+/+	+/+	-/+	-
	CSTA	Cystatin-A	11006	cystatin	-	+	+/+	+/+	+/+	-
	DMKN	Dermokine	47082	dermokine	+	-	+/+	+/+	-/+	-
	KNG1	Kininogen-1	71957	-	+	+	+/+	+/+	-/+	-
Транспортные белки	AGP2 (ORM2)	Alpha-1-Acid Glycoprotein 2	23603	lipocalin	+	-	+/+	+/+	-/+	-
	ALB	Serum albumin	69367	ALB / AFP / VDB	+	+	+/+	+/+	+/+	-
	LCN1	Lipocalin-1	19250	lipocalin	+	-	+/+	+/+	+/+	-
Структурные белки	DSP	Desmoplakin	331774	plakin	-	+	+/+	+/+	+/+	-
	FLG2	Filaggrin-2	248073	S100-fused protein	-	+	+/+	+/+	+/+	-
	HRNR	Hornerin	282390	S100-fused protein	+	+	+/+	+/+	+/+	-
	JUP	Junction plakoglobin	81745	beta-catenin	-	+	+/+	+/+	+/+	-
	SHROOM3	Protein Shroom3	216857	shroom	-	+	+/-	+/-	-/-	-
Защитные белки	AZGP1	Zinc-alpha-2-glycoprotein	34259	MHC class I	+	-	+/+	+/+	+/+	-
	DCD	Dermcidin	11284	-	+	-	+/+	+/+	+/+	-
	IGHA1	Immunoglobulin Heavy Constant Alpha 1	37655	-	+	-	+/+	+/+	+/+	-
	LYZ	Lysozyme C	16537	glycosyl hydrolase 22	+	+	+/+	+/+	+/+	-
	SPP1	Osteopontin	35423	osteopontin	+	-	+/+	-/+	-/+	-
Сократительные белки	ACTG1	Actin, cytoplasmic 2	41793	actin	+	+	+/+	+/+	+/+	-
	DNAH14	Dynein heavy chain 14, axonemal	399895	dynein heavy chain	-	+	+/+	+/+	-/+	-

¹ – Молекулярная масса согласно базе данных UniProt

² – Согласно базе данных GeneCards

^{3/4} – Согласно базе данных MOPED/BioGPS

⁵ – Легкие и Респираторный Тракт

⁶ – Биологические/Патологические процессы согласно GO/Quagen™

Для дальнейшего анализа были использованы данные протеомов КВВ пациентов с 1 и 2 стадиями рака легкого, поскольку они представляют наибольший интерес с диагностической и прогностической точек зрения [**Reed et al., 2004**].

42 белка не кератиновой природы были идентифицированы только в КВВ доноров с диагностированным раком легкого 1-2 стадии и отсутствовали как в КВВ группы здорового контроля, так и в КВВ доноров с ХОБЛ и внебольничной пневмонией. На основании подробного аннотирования с помощью биоинформатических ресурсов, а также анализа частот встречаемости, были выделены 19 белков, которые можно было бы предложить в качестве диагностической панели для рака легкого при исследовании конденсата выдыхаемого воздуха (таблица 16). [**Федорченко и др., 2016**]

Таблица 16. Белки, идентифицированные только в протеомах КВВ доноров с диагностированным раком легкого 1-2 стадии и предлагаемые в качестве диагностической панели

Функция	Описание белка									
	Название гена	Название белка	Mr ¹ , Да	Семейство ¹	Локализация в клетке ²		Экспрессия в тканях ^{3/4}			Вероятная связь с клинической картиной ⁶
					Вне.	Внутр.	Кровь	Кожа	Л&Р T ⁵	
Ферменты	PANK2	Pantothenate kinase 2, mitochondrial	62681	type II pantothenate kinase	-	+	+/+	+/+	-/+	+
	DPYSL2	Dihydropyrimidinase-related protein 2	62294	DHOase	+	+	+/+	+/+	+/+	+
	DPYSL5	Dihydropyrimidinase-related protein 5	61421	DHOase	-	+	+/+	-/+	+/+	+
	DDX20	Probable ATP-dependent RNA helicase DDX20	92241	DEAD box helicase	-	+	+/+	-/+	+/+	+
	SEPT7	Septin-7	50680	Septin GTPase	+	+	+/+	+/+	+/+	+
Регуляторные белки	ATPIF1	ATPase inhibitor, mitochondrial	12249	ATPase inhibitor	-	+	+/+	+/+	-/+	+
	TBC1D1	TBC1 domain family member 1	133084	-	-	+	+/+	+/+	-/+	+
	TBC1D4	TBC1 domain family member 4	146563	-	-	+	+/+	+/+	-/+	+
	NUCKS1	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	27296	-	-	+	+/+	+/+	+/+	+
	POTEE	POTE ankyrin domain family member E	121363	In the N-terminal section belongs to the POTE family; in the C-terminal section belongs to the actin family	+	+	+/na	+/na	+/na	+
	SFRS1	Serine/arginine-rich splicing factor 1	27745	splicing factor SR	+	+	+/+	+/+	+/+	+
	SFRS3	Serine/arginine-rich splicing factor 3	19330	splicing factor SR	-	+	+/+	+/+	+/+	+
	SFRS4	Serine/arginine-rich splicing factor 4	56678	splicing factor SR	-	+	+/+	-/+	+/+	+
	SFRS6	Serine/arginine-rich splicing factor 6	39587	splicing factor SR	-	+	+/+	-/+	+/+	+
	WDR13	WD repeat-containing protein 13	53696	-	-	+	+/+	-/+	-/+	+
Транспортные белки	SYN1	Synapsin-1	74111	synapsin	-	+	-/+	-/+	-/+	+
Структурные белки	SPDL1	Protein Spindly	70172	Spindly	-	+	+/+	-/+	-/+	+
	BSDC1	BSD domain-containing protein 1	47163	-	-	+	+/+	-/+	-/+	+
	HSP90AA1	Heat shock protein HSP 90-alpha	84660	heat shock protein 90	+	+	+/+	+/+	+/+	+

¹ – Молекулярная масса согласно базе данных UniProt; ² – Согласно базе данных GeneCards; ^{3/4} – Согласно базе данных MOPED/BioGPS; ⁵ – Легкие и Респираторный Тракт; ⁶ – Биологические/Патологические процессы согласно GO/Quagen™

Как видно из приведенных данных (таблица 16), большинство представленных белков по своей функциональной характеристике относятся к категории регуляторных (в том числе, к этой же группе можно было бы отнести и категорию ферментов), имеют внутриклеточную локализацию, экспрессия всех представленных белков повышена при исследуемом заболевании. Более того, некоторые из представленных белков входят в коммерческие ПЦР-тесты на профилирование рака легкого по крови (Quagen™).[Федорченко и др., 2016]

В работе [Bera et al., 2006] было показано, что появление POTE ankyrin domain family member E характерно для многих видов рака (при этом он практически отсутствует в здоровых тканях). Последующие исследования экспрессии данного белка, в том числе для рака легкого [Wang et al., 2015], подтвердили его перспективность как маркера опухоли, и POTE ankyrin domain family member E был предложен в качестве мишени при разработке «противораковой вакцины» [Huang et al., 2013]. Серин/аргинин-обогащенный фактор сплайсинга 1 также часто попадет во внимание исследователей как маркер опухолевого процесса [Ezponda et al., 2010; Miguel et al., 2014].

Следует отметить, что среди белков, идентифицированных в КВВ онкопациентов, много участников митоза, а также процессов транскрипции, трансляции и альтернативного сплайсинга, что может отражать процесс неконтролируемого деления клеток опухоли (BSD domain-containing protein 1, Heat shock protein HSP 90-alpha, Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1, Probable ATP-dependent RNA helicase DDX20, Protein Spindly, Septin-7, Serine/arginine-rich splicing factor 1–6, WD repeat-containing protein 13). Большой интерес представляет идентификация группы сплайсинг-факторов (SR family), которые играют критическую роль в развитии опухоли [Da Silva et al., 2015].

Помимо представленной панели из 19 белков, можно отметить появление в КВВ доноров с раком легкого семейства HMG-I/Y белков и лактоферрина, что можно связать с индуцируемым опухолью разрастанием сети капилляров и иммунным ответом на активность раковых клеток [Рябокоть и др., 2014].

На основании полученных результатов анализа протеома конденсата выдыхаемого воздуха можно утверждать, что белковые профили различных групп доноров поддаются разделению и существует вероятность выделения специфической группы белков, присущей определенному состоянию/заболеванию респираторной системы. Следует отметить, что протеом онкологических больных, определяемый в КВВ, резко отличается от протеомов не только здорового молодого не курящего контроля, но и от протеомов больных ХОБЛ и пневмонией той же возрастной группы, что свидетельствует о потенциальной возможности

использования КВВ в качестве скринингового теста с последующей проверкой другими методами, например, компьютерной томографией. [Федорченко и др., 2016]

7.3 Сравнительный анализ белкового состава КВВ больных с диагнозом «рак легкого» и белкового состава ткани, пораженной опухолевым процессом

Следует отметить, что в КВВ идентифицировали 70 белков, которые встречаются чаще у пациентов с диагнозом «рак легкого», чем у здоровых доноров (данные не приведены) и 12 (Таблица 2 Приложение) белков, которые чаще встречаются у здоровых. На рисунке 13 представлено распределение белков КВВ пациентов с диагнозом «рак легкого» по частоте встречаемости в сравнении со здоровым контролем.

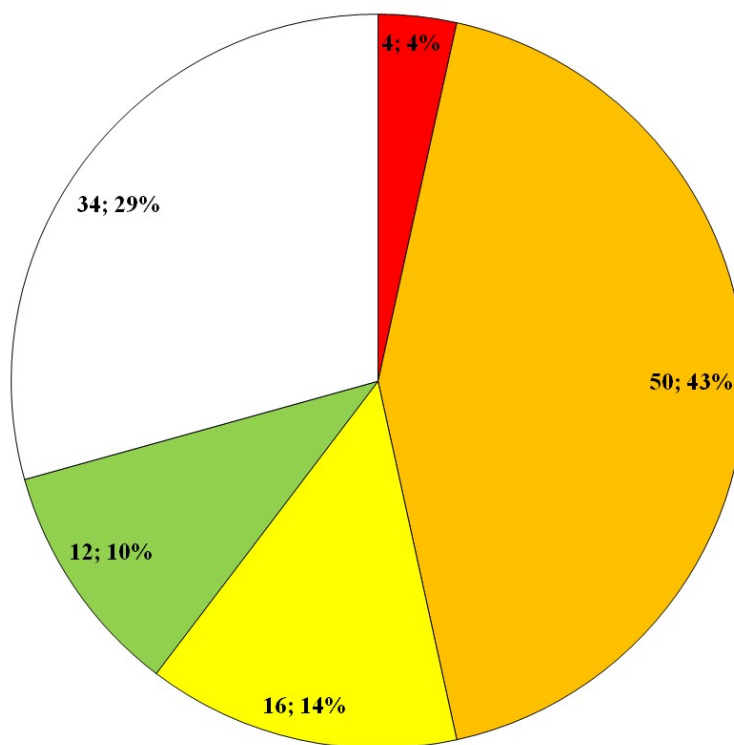


Рисунок 13. Распределение белков из КВВ по частоте встречаемости

- встречаются более, чем в 50% проб онкобольных, и не встречаются у здоровых
- встречаются менее, чем в 50% проб онкобольных, и не встречаются у здоровых
- встречаются у онкобольных чаще, чем у здоровых (у здоровых встречаются тоже)
- встречаются у онкобольных реже, чем у здоровых (у онкобольных встречаются тоже)
- встречаются примерно одинаково

Можно предположить, что различие по частоте встречаемости белков в КВВ могло бы быть связано с уровнем экспрессии этих белков в опухолевой ткани легкого. Для проверки этого предположения провели сравнение результатов анализа белкового состава КВВ пациентов с раком легкого с опубликованными данными по белковому составу опухолевой ткани легкого человека.

В качестве источника информации о белковом составе опухолевой ткани выбрали базу данных dbDEPC (a database of Differentially Expressed Proteins in Human Cancer). В ней предоставлена информация о белках с различным уровнем экспрессии в раке по сравнению с нормой или в различных типах рака, полученная, в основном, с помощью масс-спектрометрии. Всего в базе представлено 20 видов рака. Заполнение базы происходило следующим образом: вначале проводили автоматический поиск по ключевым словам (рак, масс-спектрометрия, протеомика) в базе данных PubMed, данные из найденных статей подвергали статистической проверке, идентификаторы белков сопоставляли с идентификаторами UniprotKB [Marimuthu et al., 2011; Xiao et al., 2012].

Список белков для дальнейшей работы находили по следующим параметрам поиска: норма против рака (Normal vs. Cancer), тип образца: ткань (tissue), организм: человек (Homo sapiens), название рака: рак легкого (в базе оказался представлен только один вид рака легкого: Lung Adenocarcinoma). По данному фильтру отобрали 8 статей, 2 из которых не подошли, поскольку в них рассматривали кровь, а не ткань легкого.

Несмотря на то, что в dbDEPC рак легкого представлен только аденокарциномой, это наиболее полная и удобная база данных из имеющихся на настоящий момент.

Всего тканевых белков из базы данных 105, 68 (таблица 5-П, см. приложение) повышено экспрессируются в опухолевых клетках аденокарциномы легкого и 37 (таблица 6-П, см. приложение) – пониженно экспрессируются. Среди пониженно экспрессирующихся белков КВВ и тканей пересечений не найдено. В группах повышено экспрессирующихся белков совпали три белка: дигидропиримидиназо-зависимый белок 5 (DPYL5), лактоферрин (LTF) и альфа-1-антихимотрипсин (ААСТ).

Дигидропиримидиназо-зависимый белок 5 изучен мало. DPYL4 – непосредственная мишень p54 при апоптозе, индуцированном повреждением ДНК. [Kimura et al., 2011] DPYL 2 – предлагается как маркер рака легкого. [Ocak et al., 2014] Дигидропиримидиназо-зависимые белки принимают участие в росте и дифференциации нейронов [He et al., 2012].

Дигидропиримидиназо-зависимые белки похожи по последовательностям: в выравнивании примерно 40% консервативных позиций, выделяются консервативные блоки.

(рисунок 14) Это позволяет предположить сходность функций этих белков. Возможно, DPYL5 также вовлечен в процесс развития опухоли.

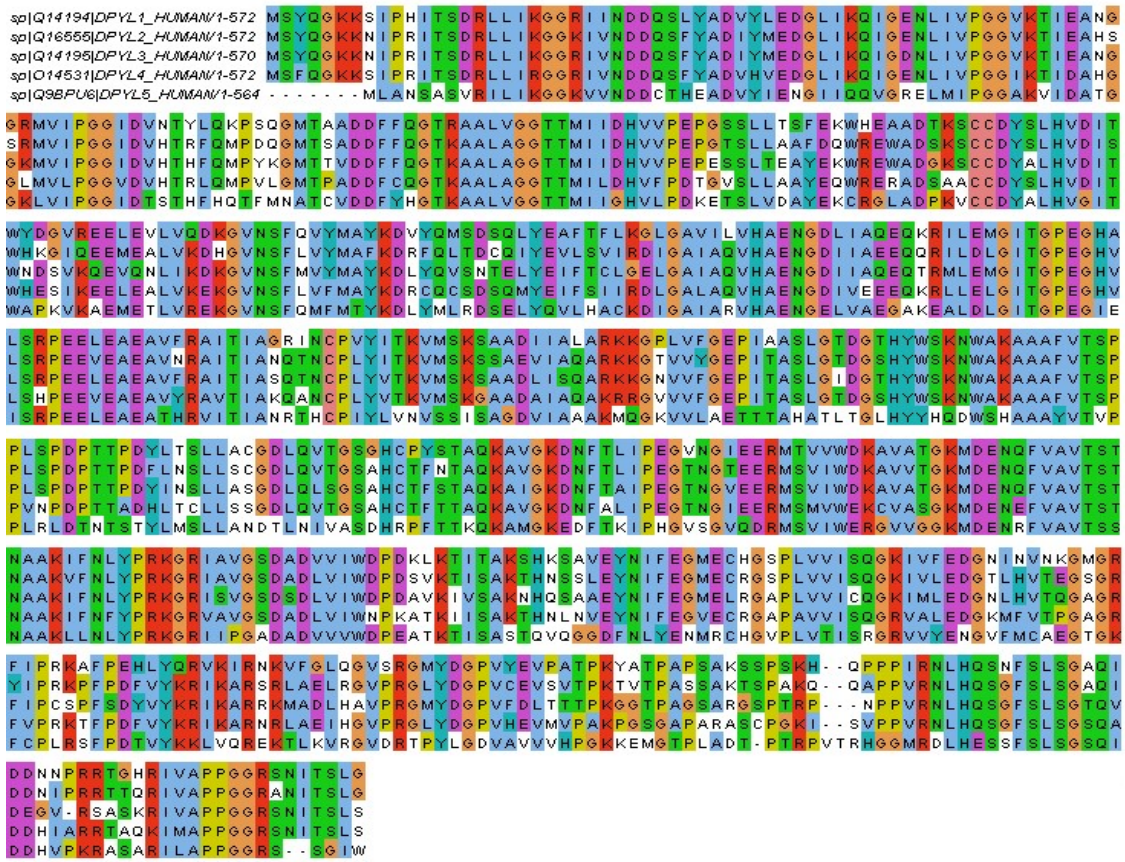


Рисунок 14. Выравнивание DPYL1, DPYL2, DPYL3, DPYL4 и DPYL5, полученное с помощью программы Jalview

Альфа-1-антихимотрипсин – один из наиболее распространенных ингибиторов сериновых протеиназ. Он является белком острой фазы воспаления. [Guergova-Kuras et al., 2011] При развитии опухоли, баланс между протеолизом и его ингибированием нарушается. Сериновые протеиназы, такие как эластаза, плазмин, тромбин, трипсин и катепсин G, наиболее тесно связаны с развитием опухоли. Эластаза – единственная протеиназа, которая способна разрушать нерастворимый эластин, что приводит к потере эластичности легких и разрушению альвеол. [Digumarti et al., 2011] Известно, что эластазы продуцируются и секретируются различными типами опухолевых клеток, в том числе при немелкоклеточном раке легкого. [Jin et al., 2016] Тем не менее, опухолевые клетки сами могут секретировать ингибиторы сериновых протеиназ, есть данные о том, что они положительно влияют на развитие опухоли. [Tung et al., 2013; Shau et al., 1992] Регулирование альфа-1-антихимотрипсина – необходимое условие для поддержания целостности тканей и ремонта повреждений. Альфа-1-антихимотрипсин предлагается в качестве маркера рака легкого. [Tsuda et al., 2002; Shimamura et al., 2004]

Лактоферрин – основной железо-связывающий белок, содержащийся в экзокринных жидкостях, таких как грудное молоко и выделения слизистых оболочек. Помимо своей основной функции связывания и переноса ионов железа, он имеет множество других функций, таких как антибактериальная, противовирусная, противогрибковая, противовоспалительная и противораковая. На ранних стадиях рака, нерегулируемая пролиферация клеток приводит к дефициту кислорода, вызывая гибель клеток, это приводит к воспалению и способствует экспрессии фактора роста эндотелия сосудов (VEGF). Лактоферрин уменьшает пролиферацию за счет снижения его экспрессии. [Koshikawa et al., 1996] Лактоферрин стимулирует производство естественных киллеров (NK) и лимфокинактивированных киллеров (LAK) [Kalsheker, 2009], интерферона-гамма (IFN-gamma), каспазы-1 [Северин, 2003], интерлейкина IL-18, с противометастатическими свойствами [Yamashita et al., 1997], кроме того он подавляет производство интерлейкинов IL-4, IL-6, и IL-10, способствующих пролиферации раковых клеток. [Koshikawa et al., 1996] Мутации на коротком плече третьей хромосомы, где расположен ген лактоферрина, является одним из наиболее распространенных событий в патогенезе рака легкого. Это значит, что лактоферрин играет важную роль в канцерогенезе. [Guergova-Kuras et al., 2011]

Для рекомбинантной формы лактоферрина – талактоферрина проводились клинические исследования по его применению в качестве терапии рака легкого [Grant et al., 1989].

На основании проведенного сравнения экспериментальных данных по белковому составу КВВ пациентов с диагнозом «рак легкого» с данными анализа белкового состава опухолевой ткани легкого, в КВВ найдены белки, встречающиеся в раковых тканях легкого: дигидропиримидиназо-зависимый белок 5, лактоферрин и альфа-1-антихимотрипсин. Все найденные белки относятся к повышенно экспрессирующимся белкам опухолевой ткани легкого по данным анализа литературы. По данным статей эти белки принимают участие в воспалительных и опухолевых процессах. Из этого можно заключить, что некоторые белки могут попадать в КВВ непосредственно из опухолевых тканей и служить маркерами заболевания. Это подтверждает возможность использования анализа КВВ в качестве метода ранней диагностики рака легкого.

7.4 Сравнительный анализ белкового состава КВВ больных с диагнозом «рак легкого» с уровнем экспрессии генов при данном заболевании

Помимо сравнения результатов анализа белкового состава КВВ пациентов с диагнозом «рак легкого» с опубликованными данными по белковому составу опухолевой ткани легкого, мы провели сравнение наших результатов с данными по экспрессии генов при раке легкого.

При проведении исследования была изучена информация о шести базах данных, содержащих данные, связанные с экспрессией генов в раковой ткани легкого. Выбранные БД оценивались по параметрам, представленным в таблице 17, и затем из всех баз данных была выбрана наиболее подходящая для работы.

Таблица 17. Параметры отбора БД

Название БД	Информация в БД	Источники информации	Удобство поиска	Комментарий
GeneCards	Энциклопедия генов, белков и болезней	Публикации, эксперименты	да	Достаточно удобна, но требуется знать, какие именно гены нужно искать
dbEST	Подразделение GenBank, содержащее данные и другую информацию по маркерным экспрессируемым последовательностям	Публикации	нет	Неудобна для работы
CGAP	Проект, содержащий информацию об экспрессии генов нормальных, предраковых и раковых клеток человека	Данные из других БД (RefSeq, UniProt)	нет	Не содержит достаточной информации для работы
UniGene	БД, содержащая совокупность транскрипционных последовательностей схожих транскрипционных локусов, информация о белковых сходствах, генной экспрессии, Кднк	Публикации, библиотеки данных	да	Аналогично GeneCards
MOPEP	БД белковой экспрессии модельных организмов	Эксперименты	да	Аналогично GeneCards
GEO DataSets	БД, содержащая проверенные сетки данных генной экспрессии, а также данные о сериях образцов и экспериментах	Публикации, эксперименты	да	Подходит для работы

На основе проведенного анализа для дальнейшей работы была выбрана база данных GEO DataSets. Анализ экспрессии генов проводили с использованием данных, представленных в базе данных GEO DataSets. Эта БД содержит информацию о каждом сете собранных образцов,

отдельно о каждом образце (включая данные об источнике образца, методе сбора и анализа), а также информацию об уровнях экспрессии генов в этих образцах.

В выбранной БД был проведен поиск по сетам, полученным из ткани легкого и содержащим более 50 образцов. Затем среди полученных сетов были отобраны те, которые содержали образцы и из здоровой ткани легкого, и из опухолевой. Этот фильтр был необходим для того, чтобы в дальнейшем можно было определить сверхэкспрессированные в раковой ткани гены.

Поскольку, как уже было упомянуто выше, сбор данных о генной экспрессии проводился при помощи микрочипирования на чипе Affymetrix GeneChip®, в БД GEO DataSets для каждого образца в выбранных сетях содержалась информация о фрагментах генов, исследованных на экспрессию. Сама информация содержалась в таблицах, состоящих из двух столбцов: в первом были указаны идентификаторы (ID) этих фрагментов в формате Affymetrix, а во втором - уровни экспрессии фрагментов. Используемый при сборе данных микрочип позволяет одновременно изучать экспрессию многих фрагментов генов (в данном случае - 54614) для каждого образца. С учетом количества образцов в одном сете (более 50), полученный объем информации очень трудно проанализировать, поэтому текущее исследование было ограничено одним сетом образцов.

Критерии для поиска использовали следующие: объект «Homo sapiens», диагноз «рак легкого», количество доноров не менее 50. По данным критериям для дальнейшего анализа отобрали статью [Lu et al., 2010]. В выбранном сете содержались образцы тканей легкого, собранные у пациентов с диагнозом "рак легкого". У каждого пациента было взято по два образца: из здоровой и опухолевой тканей. Благодаря этому стало возможным определить, какие из фрагментов генов были сверхэкспрессированы в раковой ткани легкого.

Для выполнения этой задачи был написан и запущен скрипт на языке программирования Python. Данный скрипт сначала усреднял значения экспрессии каждого фрагмента отдельно для групп "здоровых" и "опухолевых" образцов, а затем устанавливал отношения полученных значений. В качестве порога, после превышения которого фрагмент считался сверхэкспрессированным, было выбрано следующее значение: превышение уровня экспрессии фрагмента в опухолевой ткани более чем в два раза по сравнению с ее уровнем в здоровой.

По результатам работы скрипта было отобрано 646 фрагментов генов. Ясно, что если ген сверхэкспрессирован, то и все его фрагменты также сверхэкспрессированы, поэтому после работы скрипта не было потерь нужной информации.

Однако, идентификаторов фрагментов генов в формате Affymetrix не достаточно для того, чтобы выяснить, какие гены, а затем – белки, сверхэкспрессируются в опухолевой ткани. Для дальнейшей работы необходимо переконвертировать данные идентификаторы в формат какой-либо широко известной БД и затем выяснить, какие именно белки соответствуют фрагментам генов, сверхэкспрессированных в опухолевой ткани легкого. Описанную работу выполняют конвертеры ID. В данном исследовании было изучено три таких конвертера, и из них был выбран один, наиболее подходящий для работы. Параметры выбора представлены в таблице 18.

Таблица 18. Параметры отбора конвертера ID

Название конвертера	Возможность загрузки одновременно всех ID	Вид предоставляемой информации	Возможность сохранения результатов	Скорость работы	Комментарий
David	Нет	Информация о генах (названия, функции и т.д.)	Нет	Медленно	неудобен для работы
RefDic	Да	Идентификаторы, соответствующие загруженным, в интересующем формате	Да, но полученный файл неудобен для работы	Быстрее, чем David	удобен, но имеется лучший вариант
db2db	Да	Аналогично RefDic; также возможно выбирать одновременно несколько форматов ID для конвертации, есть фильтр по организмам	Да	Самый быстрый	Самый удобный для работы из изученных

Исходя из информации, полученной при изучении трех конвертеров, для дальнейшей работы был выбран конвертер db2db. Отобранные ранее идентификаторы были загружены в конвертер и переконвертированы в GenBank Protein GI. Таким образом, сразу стало возможным узнать последовательности белков, соответствующих фрагментам генов.

Сначала при помощи Python-скрипта из БД GenBank были скачаны последовательности всех белков, чьи GI были получены после конвертирования ID в формате Affymetrix. Затем из всех последовательностей была составлена белковая БД при помощи Bash-скрипта. Поскольку со стороны опухолевой ткани велась работа с белковыми последовательностями, названий белков, определенных в KBB программой Mascot, было недостаточно. Поэтому далее были проанализированы файлы данной программы, содержащие, помимо названий белков, их последовательности, аббревиатуры названий соответствующих генов и т.д.

После запуска Python-скрипта из файлов Mascot были получены последовательности всех белков, с наибольшей вероятностью соответствовавших белкам КВВ. Для этих последовательностей был проведен поиск blastp на полное сходство последовательностей. По результатам поиска была составлена сводная таблица, содержащая информацию о белках, общих для КВВ и раковой ткани легкого (таблица 19).

Таблица 19. Белки, общие для КВВ и раковой ткани легкого

Gene symbol	Название белка	Функции белка	Информация об ассоциации белка с раком легкого на основе данных из Pubmed
ALB	Serum albumin	Регуляция коллоидного осмотического давления крови, транспорт цинка в плазме	Ассоциирован
AZGP1	Zinc-alpha-2-glycoprotein	Стимуляция деградации липидов, вызывание экстенсивной потери жира	Ассоциирован с разросшимися опухолями
SPP1	Osteopontin	Гликофосфопротеин, содержащий N- и O-связанные олигосахариды; участвует в адгезии клеток	Ассоциирован
H2AFV HIST1H2AB HIST1H2AE HIST1H2AI HIST1H2AK HIST1H2AJ HIST1H2AG HIST1H2AL HIST1H2AM HIST1H2AD	Histone H2A	Формирование нуклеосомы, представляющей собой белковую глобулу, вокруг которой накручена нить ДНК	Ассоциирован
IGHA1 IGLV2-14 IGLV1-44 IGLC2 IGLC1 IGL@ IGLV3-21 IGLV2-11 IGLV1-40 IGLC3 IGHV4-31 IGHG1	Immunoglobulin (различные типы)	Антиген-связывающая и эффекторная (вызывают тот или иной иммунный ответ, например, запускают классическую схему активации комплемента) функции	Для IgA и IgG есть данные, что они ассоциированы с раковой опухолью легкого
KRT1 KRT10 KRT13 KRT14 KRT15 KRT16 KRT2 KRT5 KRT6B KRT7 KRT75	Keratin (различные типы)	Формирование твёрдых, но не минеральных структур, такие как волосы и ногти; образование наиболее прочных элементов внутри клеточного цитоскелета эпителиальных клеток	Для кератинов 7,10,14,15,16 есть данные, что они ассоциированы с раковой опухолью легкого. Кератин 14 является маркером в стволовых клетках и клетках-предшественницах в легочном эпителии, кератин 16 – маркером для плоскоклеточного рака легкого. Экспрессия кератина 7 – прогностический параметр для диагностики рака легкого на ранних стадиях.

Всего было найдено 36 общих белков, из них: 10 гистонов, 12 иммуноглобулинов и 11 кератинов. Как можно увидеть из таблицы 19, большинство найденных белков ассоциированы с раком легкого по данным из научных статей в базе данных PubMed.

Полученные данные также свидетельствуют о том, что часть белков и пептидов, определяемых в КВВ, с большой вероятностью попадают туда из опухолевой ткани легкого.

Выводы главы.

В проведенном исследовании было показано, что протеомный анализ конденсата выдыхаемого воздуха является информативным методом и обладает большим потенциалом в качестве метода диагностики респираторных заболеваний, поскольку позволяет установить различия не только между группами «больной-здоровый», но и между отдельными заболеваниями, такими как ХОБЛ, пневмония и рак легкого. Найдены пересечения белков, идентифицированных в КВВ доноров с раком легкого, с опубликованными данными по белкам, выделенным из опухолевой ткани легкого, и данными белкам, ассоциированными с генами, чья экспрессия повышена при данном заболевании.

Глава 8. Новые подходы к разработке метода неинвазивной диагностики заболеваний лёгких на основе анализа состава КВВ

8.1 Создание базы данных проб и доноров КВВ

Как следует из предыдущих глав, был накоплен большой экспериментальный материал и получены обнадеживающие результаты по анализу КВВ групп с различными заболеваниями. В настоящее время исследование продолжается, идёт постоянное накопление статистики по различным состояниям лёгких и респираторного тракта, в связи с чем поступающие данные должны быть сопоставлены с уже имеющимся. Для хранения и обработки накопленной информации была выбрана реляционная база данных, так как у неё простая и удобная схема данных, представленная в виде таблиц.

На основании полученных проб был осуществлен отбор параметров, которые потенциально могли повлиять на состав КВВ и с которыми в дальнейшем запланирована работа по сортировке получаемых данных. Все отобранные параметры были собраны в сводную таблицу (таблица в формате Excel), информацию в которую заносили на этапе сбора пробы и последующих этапах пробоподготовки, получения и анализа масс-спектра. Для образцов фиксировали такие параметры, как обозначение пробы; время сбора; pH образца; объем образца, полученного за 10 минут; указывали оператора (проводившего пробоподготовку и анализ образца); параметры программы, которые использовались для обработки масс-листов; масс-спектр в формате *.raw; масс-листы в формате *.msm, *.dat, *.htm. Информация по пациентам включала следующие строки: пол пациента; возраст пациента; объем выдыхаемого воздуха за 10 минут; статус курильщика; диагноз; результаты гистологического исследования; принимаемые лекарства; проводимое лечение; первые симптомы; сопутствующие заболевания; наследственность. Итого, два десятка параметров для каждого образца КВВ.

Создание базы данных начали с её моделирования: прописывались связи между полями и ключи. Главной особенностью реляционных баз данных является, то, что объекты внутри таких баз данных хранятся в виде набора двумерных таблиц, между которыми могут существовать связи по ключевым значениям. На этом этапе было важно общее понимание процесса работы с пробой, так как каждое поле отражает какую-то стадию, связанную с обработкой или изучением пробы. Следующим этапом в создании базы данных было проектирование её структуры. Для этого в созданных таблицах добавлялись столбцы (ячейки):

указывался тип данных и наименование столбцов. Эти действия производили с помощью программы MySQLWorkbench 6.3 CE.

На рисунке 15 показана блок-схема базы данных на этапе её моделирования.

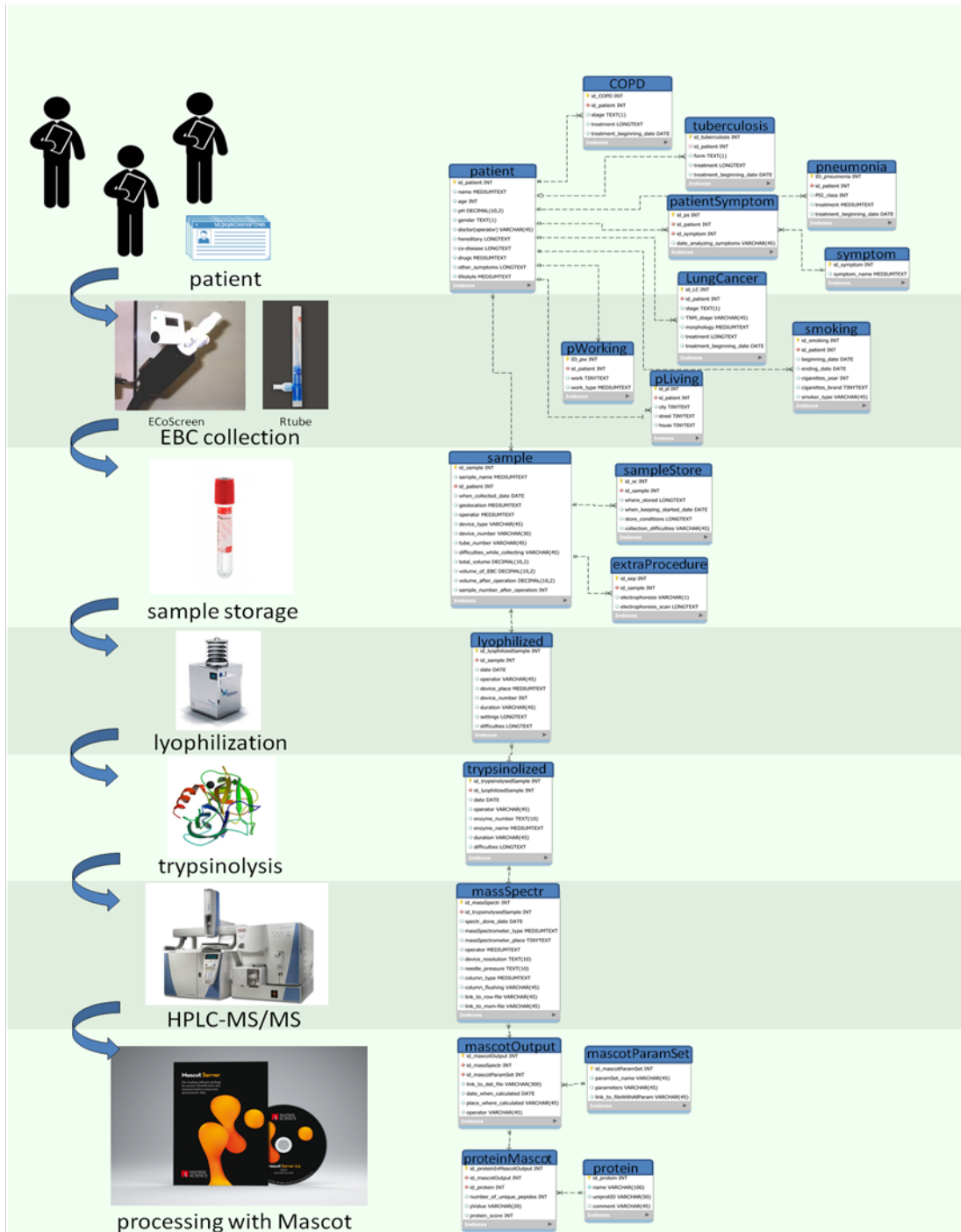


Рисунок 15. Блок-схема реляционной базы данных с привязкой к основным этапам работы с образцом КВВ. Картинка справа является результатом работы с MySQLWorkbench 6.3 CE. Каждое поле отражает какой-то этап работы с пробой, а внутри полей ячейки с нужными данными для исследования.

Далее в БД была загружена информация из сделанной ранее таблицы Excel, содержащей всю необходимую информацию о каждом пациенте, его пробе/пробах и файлах после пробообработки. Загрузка информации производилась с помощью сервера PHPMyAdmin (рисунок 16).

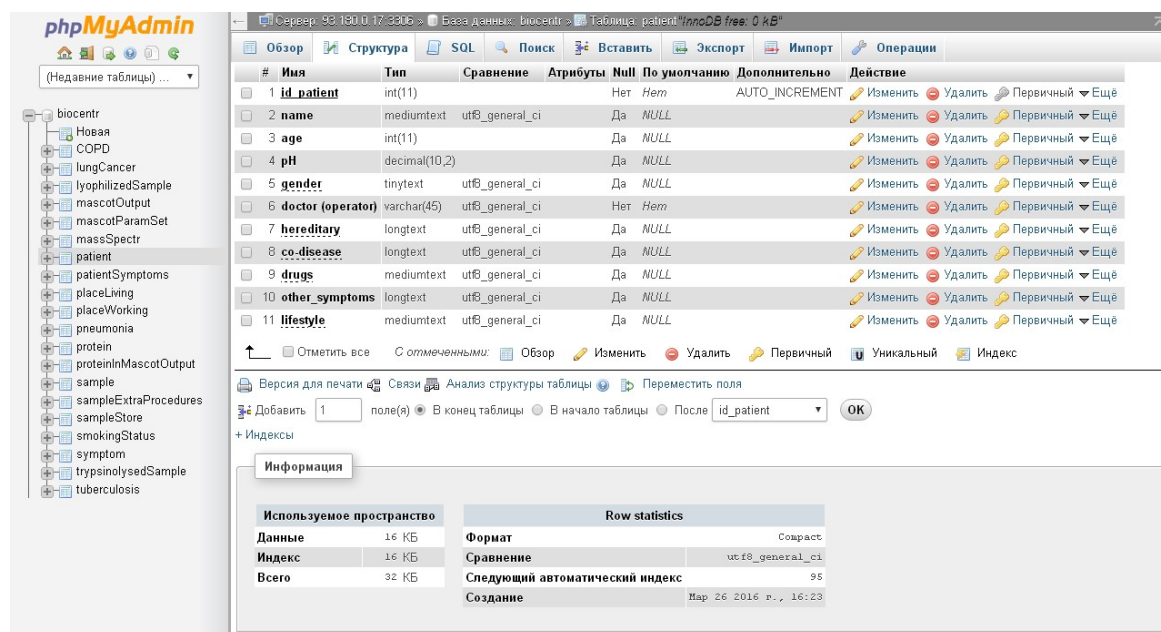


Рисунок 16. База данных, развёрнутая на сервере PHPMyAdmin, с загруженной информацией по образцам КВВ. Справа показаны ячейки поля patient.

Представленная база данных позволила производить поиск и сортировку пациентов по возрасту, полу, диагнозу и любому другому интересующему параметру. Более того, в базу данных загрузили файлы со спектрами и масс-листами.

Благодаря созданной базе данных, упорядочившей имеющуюся информацию, стало возможным провести статистическую обработку собранного массива данных о пробах КВВ доноров с различным состоянием респираторной системы и объединить их с новыми данными для расширения статистики по заболеваниям легочной системы.

8.2 Анализ границ применимости программы Mascot при идентификации белкового и пептидного состава КВВ

Несмотря на широкое применение программы Mascot в протеомных исследованиях (см. Главу 4), детали алгоритма Mascot засекречены [Hubbard et al., 2010]. Большинство MS/MS инструментов поиска выполняют одни и те же функции [Eng et al., 2011]. Все они читают собранные MS/MS-спектры, делают запрос к базе данных, чтобы выбрать пептид правильной

массы, оценивают вероятность этих пептидов против экспериментальных спектров и выдают предположительные пептидные идентификации. Схема процесса представлена на рисунке 17.

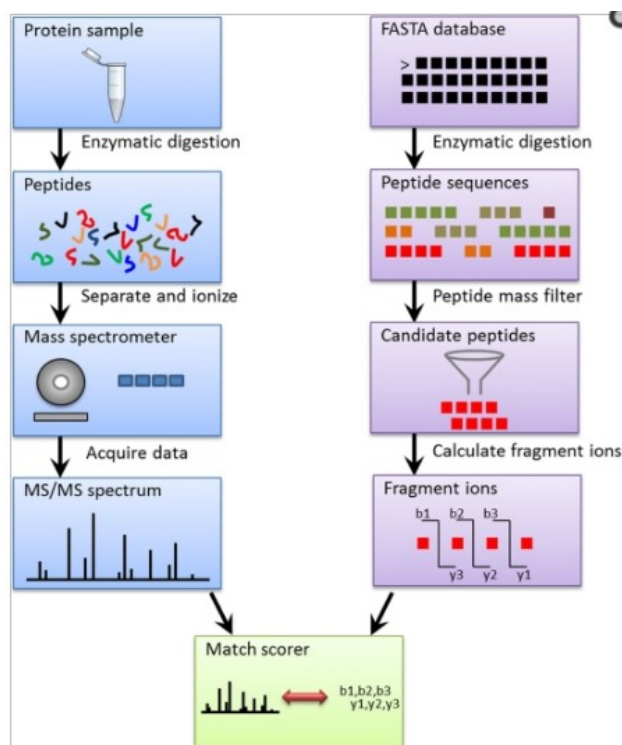


Рисунок 17. Схема работы программы, анализирующей MS/MS-спектр с целью идентификации пептидного состава пробы [Eng et al., 2011].

Для каждого MS/MS-спектра экспериментальная масса пептида может быть получена из отношения массы предшественника к его заряду (m/z) и предполагаемого или измеренного состояния заряда предшественника. Инструмент поиска в базе данных выбирает пептиды из базы данных последовательностей, которые имеют примерно такую же приблизительную массу, как экспериментальная масса пептида для спектра запросов. Каждому пептиду-кандидату присваивается рейтинг (Score), и показываются пептиды с самым высоким рейтингом для каждого спектра.

Белковые и пептидные последовательности, занесенные в базы данных (как правило, в формате FASTA) используются поисковыми инструментами при обработке MS/MS-спектра. Вместо того, чтобы рассмотреть все возможные последовательности (область действия *de novo* алгоритмов), эти инструменты ограничивают их поиск до последовательностей, присутствующих в базах данных. Обычно исследователи выбирают базу данных последовательностей, которая содержит все известные белки для интересующего организма на основании аннотации его генома [Eng et al., 2011]. Большинство поисковых систем используют

большие базы данных, состоящих из нескольких баз, такие, как NCBItr и Mascot MSDB [Hubbard et al., 2010]. Несмотря на это, нужно отметить, что в больших базах данных необязательно содержатся все возможные последовательности, которые были подтверждены недавним экспериментом по секвенированию генома. Кроме того, они чрезмерно избыточны и из-за этого могут приводить к ложноположительным идентификациям (в противовес ложноотрицательным в случае использования маленькой базы данных, так как какие-то изоформы или варианты могут не попасть в базу) (таблица 20).

Таблица 20. Статистика поиска по базам данных с помощью Mascot для коллекции спектров образцов белков. True positives – peptide-spectrum matches, где Mascot ion score превышает порог и пептиды принадлежат к представленным белкам. False positives – peptide-spectrum matches, где Mascot ion score превышает порог, но пептид не принадлежит представленному белку.

Protein	Species	Database	Taxonomic Filter	Mascot threshold	True positives	False positives
Molecular chaperone dnaK	<i>E. coli</i>	MSDB	None	48	11	15
		MSDB	<i>E. coli</i>	28	22	0
		E. coli K12	None	19	24	1
Ubiquitinprotein ligase	Human	MSDB	None	49	8	7
		MSDB	<i>Homo sapiens</i>	35	28	1
		IPI Human	None	33	32	1

В таблице 20 представлено, как размер базы данных может влиять на результаты поиска на примере человека и микроорганизма. Mascot был выбран, так как он считает scoring threshold, основанный на количестве совпадающих ионов-предшественников в базе данных. Ограничение поиска видоспецифичными наборами или использование видоспецифичной базы данных приводит к увеличению true positives.

Файл с данными Mascot – это файл в виде обычного текста (ASCII), содержащий список значений пиков, и, опционально, параметры поиска. Для Peptide Mass Fingerprint файл содержит список значений масс пептидов, по одному значению в каждой строке, опционально сопровождающихся пробелом и площадью пика или значению интенсивности. Для MS/MS Ions Search файл с данными должен содержать один или больше списков пиков MC/MC. Каждый MC/MC набор данных – список значений пар масс и интенсивности, разграниченных с помощью BEGIN IONS и END IONS. Файл с данными может включать встроенные параметры поиска. Большинство параметров поиска могут появляться только однажды в шапке файла с данными. Определенные параметры могут появиться внутри набора данных MC/MC (<http://www.matrixscience.com/>).

Выходные данные Mascot могут быть представлены в форматах XML и CSV. Информация, которая содержится в этих форматах, идентична. Формат XML идеален для

импорта в реляционную базу данных. Формат CSV может быть открыт в таких таблицах, как Microsoft Excel.

Mascot определяет peptide spectrum matches (PSMs), которые основаны на вероятности, что наблюдаемый результат представляет теоретическое значение в базе данных белковых последовательностей. Важно установить соответствующие параметры нормализации в алгоритме поиска, чтобы условия в каждом экспериментальном «проходе» были настолько точны, насколько это возможно. [Lin et al., 2014] Рейтинговые оценки идентифицированных пептидов (scores) сильно меняются в зависимости от различных поисковых параметров. В работе [Quandt et al., 2014] показано, что различные комбинации параметров поиска меняли уровень идентификации пептидов от 36% до 93%. Многие исследовательские группы приложили усилия по совершенствованию методов оценки данных, но использование этих методов не было широко принято и не стандартизуется [Balgley et al., 2007].

На основании анализа научной литературы из баз данных PubMed и PMC установлено, что для идентификации белков в КВВ с использованием масс-спектрометрии анализа разных инструментов и параметров для поиска и последующей стандартизации нет.

В ходе проведения анализа выданных html- и dat- файлов, полученных в результате работы программы Mascot, было обнаружено, что информация в html-формате файла представлена иначе, чем в dat-формате. Теряется часть информации, в частности, в protein hits попадают не все белки из dat-формата, теряется описание элементов и модификаций. Массы и Scores, как правило, округляются. Большое количество информации не отображается на html-странице, но представлено в ее коде. Кроме того, в html-формат файла не попадает большая часть пептидов из dat-формата. Почти во всех случаях, в html-формат файла попадают пептиды, встречавшиеся в dat-формате файла в верхних строчках ранжированного списка пептидов, которые соответствуют данному значению массы (чем выше строчка, тем лучше соответствие). Иногда последовательность аминокислот в пептидах может изменяться (например, запись XEGGGGGK в dat-файле преобразуется в запись SEGGGGGK в html-файле), что означает, что программа делает «подгонку» файла в формате .dat к теоретическим данным из базы, чтобы оформить белковую выдачу. Изменение последовательности – это источник ошибки при идентификации белков, которая может приводить как к неверной идентификации белка, так и к «потере» белка при анализе в случае, если фрагменты этого белка не были соотнесены с ним программой и оказались приписаны другим белкам или вообще не учитывались при формировании итогового файла.

В работе [Bera et al., 2006] было показано, что появление POTE ankyrin domain family member E характерно для многих видов рака, при этом он редко появляется в нормальных тканях. N-конец белка POTEЕ принадлежит семейству POTE (POTE ankyrin domain family member E), а С-конец принадлежит семейству актинов. Этот белок был ранее идентифицирован в пробах КВВ пациентов с диагностированным раком легкого начальных стадий и выделен в состав панели для ранней диагностики рака легкого [Федорченко и др., 2016]. Был дополнительно исследован набор файлов, полученных после анализа проб КВВ с помощью программы Mascot по базам данных NCBI и IPI. Показано, что наличие ранее определенного белка-маркера рака легкого POTEЕ в пробе и набор пептидов, по которым он определяется, напрямую зависели от того, по какой базе данных производился анализ (NCBI или IPI). В файлах IPI белку POTEЕ, как правило, соответствуют два пептида – AGFAGDDAPR и MSQELEINK. В файлах NCBI белку POTEЕ, как правило, соответствует один пептид AGFAGDDAPR, который встречается в белках других семейств (актина), т.е. этот пептид, встречающийся в белке-маркере, может быть не характерен для фокусного заболевания. В 3 из 25 файлов, полученных при анализе по базе данных IPI, в которых встречался POTEЕ, ему соответствует один пептид AGFAGDDAPR, а в 3 из 18 файлов, полученных при анализе по базе данных NCBI, в которых встречался POTEЕ, ему соответствуют два пептида AGFAGDDAPR и MSQELEINK. Таким образом, формируемый программой Mascot список белков зависит от используемой базы данных (NCBI или IPI HUMAN).

В настоящее время исследователи стараются использовать разные программы по анализу масс-спектров. Наиболее распространенными поисковыми движками являются Mascot (MatrixScience (Darryl Pappin)), Sequest (Jimmy Eng и John Yates), X!Tandem (проект – The Global Proteome Machine) и OMSSA (Национальный Институт Здоровья, США). [Автономов и др., 2009] Каждая из этих программ использует свой алгоритм поиска, который может быть эффективным в конкретной ситуации. Согласно данным работы по анализу биологических жидкостей человека, наиболее полная информация получается в результате объединения данных нескольких поисковых машин [Франкевич и др., 2006].

8.3 Проблемы современного этапа развития метода и способы их решения

Как можно видеть из графиков публикационной активности по теме, с 2013 года наблюдается «плато», т.е. тема в данный момент не развивается или развивается слабо. Количество статей по протеомному подходу к данному объекту ничтожно мало.

На основании анализа опубликованных данных и данных, представленных в данном исследовании, можно прийти к выводу, что «застой» в несомненно перспективном направлении обусловлен двумя причинами: особенностями объекта и не верным в силу этих особенностей подходом к исследованию.

Как можно заключить из Главы 2 и Главы 4, особенности КВВ как объекта для протеомного исследования следующие:

- КВВ – сложная многокомпонентная смесь, содержащая большое количество различных пептидов, а также липиды, соли;
- Концентрация белка находится на пределе чувствительности метода, при этом попытки концентрирования приводят к потерям, нивелирующим плюсы этой процедуры;
- Белки в респираторном тракте подвергаются действию различных протеаз и в КВВ попадают в виде фрагментов, при этом количество фрагментов и их набор постоянно варьируется и часто покрытие последовательности не превышает 5-10%;
- Низкая концентрация белка и сложность определения увеличивают ошибку, которую могут вносить загрязнение, различие методов пробоподготовки, сбора и анализа.

Основными инструментами для анализа КВВ являются ВЭЖХ-МС/МС и Mascot. В качестве конечного результата все опубликованные исследования использовали белковую выдачу программы. Однако,

- Если посмотреть опубликованные данные, большинство белков в отдельной пробе идентифицируется программой по пептидам, встретившимся в единичных случаях [Fumagalli et al., 2012; Audain et al., 2017];
- Mascot приписывает одни и те же пептиды разным белкам (см. выше), что, вкупе с малым покрытием последовательности, увеличивает ошибку идентификации;
- Не публикуются исходные масс-спектры и полный список идентифицированных пептидов, поэтому нет возможности сопоставить данные;
- Количество пептидов, которые программа Mascot идентифицирует и соотносит с белками, в среднем, составляет около 10% от общего количества идентифицируемых по данным МС/МС измерений пептидов (при выдаче данных в html-файле в формате белковой выдачи пептиды, которые не были поставлены в соответствие белку, не отображаются);
- Последовательность аминокислот в пептидах может изменяться (например, запись XEGGGGGK в dat-файле преобразуется в запись ESGGGGGK, см. выше), что означает, что программа делает подгонку файла в формате .dat к теоретическим данным из базы, чтобы оформить белковую выдачу.

Вышеперечисленные трудности и возможные ошибки белковых идентификаций дополняются тем, что белковые находки в рамках протеомного подхода невозможно проверить другими методами, например иммуноферментным анализом (в силу фрагментированности белков и малой представленности последовательностей, а также большого количества фрагментов, принадлежащих разным белкам). В результате, выводы исследований становятся слишком спекулятивными, а единственный однозначный вывод по теме делается о невозможности стандартизации методики.

На основании вышеизложенного, можно прийти к выводу, что для протеомного анализа КВВ в диагностических целях нужно использовать не белки, а пептиды. Таким образом, программа не будет подгонять результат под базу данных, результат не будет «множиться» по разным идентификациям, могут быть применены статистические подходы.

8.4 Метод классификации проб КВВ, основанный на пептидах

В ходе данного исследования был разработан метод классификации проб КВВ по рассматриваемому диагнозу, основанный на пептидах, идентифицируемых в пробе. При разработке метода была сохранена общая логика работы, отображенная на схеме (рисунок 18).

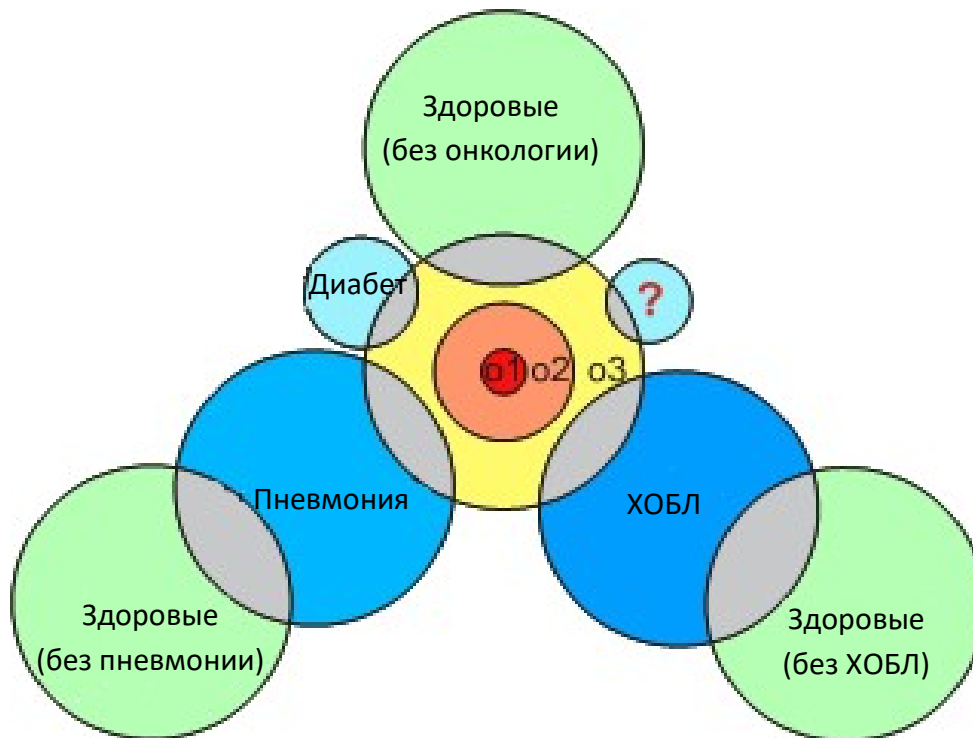


Рисунок 18. Схематичное представление полученных белковых профилей различных групп доноров, области пересечения – совокупность общих для групп белков. Область O1: маркеры рака, вошедшие в коммерческие наборы для ПЦР-профилирования рака легких по крови и не встречавшиеся в других группах; область O2: потенциальные маркеры рака, не встречавшиеся в других группах; область O3: белки, не связанные напрямую с фокусным заболеванием, либо встретившиеся в других группах доноров.

Рассматривались все исследуемые группы с учетом сопутствующих заболеваний, формировались группы пептидов, присущие тому или иному состоянию респираторной системы и организма человека. Однако, в отличие от анализа белков, идентифицируемых в КВВ, для пептидов стало возможным применить статистические подходы.

8.4.1 Формирование итоговых данных

Для проведения статистического анализа были сформированы итоговые таблицы.

Таблица с описанием доноров включала номер и название пробы, путь к файлам, содержащим данные по исходному масс-спектру и его анализу, время сбора, идентификаторы донора, основной диагноз, пол, возраст, статус курильщика, сопутствующий диагноз, наследственность, принимаемые лекарства, первые симптомы. Всего таким образом описано 108 человека.

Таблица с пептидами включала номер и название пробы, путь к файлам, содержащим данные по исходному масс-спектру и его анализу, параметры запуска Mascot, список полученных пептидов для каждой пробы. Пептиды для анализа отбирались в соответствии со значением Mascot Score.

Параметры запуска Mascot были следующие: TOL=10, TOLU=ppm, ITOL=0.5, ITOLU=Da, PFA=2, DB=NCBIInr, MODS=, MASS=Monoisotopic, CLE=Trypsin, SEARCH=MIS, CHARGE=2+ and 3+, REPORT=AUTO, FORMAT=Mascot generic, IT_MODS=Oxidation (M), TAXONOMY=Homo sapiens (human), INSTRUMENT=ESI-TRAP.

Общее количество файлов составило 291 (с учетом повторов для одного донора). Общее количество обнаруженных пептидов составило 74552.

8.4.2 Анализ масс-спектров промывочных заколов и проб на предмет загрязнения

Помимо частоты наличия пептидов в самих пробах не менее важно рассмотреть частоту наличия пептидов в промывках: если какой-то из пептидов встретился во всех промывках, то очевидно, что на такой показатель не стоит опираться при определении диагноза. В таблице 7-II (см. Приложение) приведены пептиды, которые чаще всего встречались в промывочных заколах. Как видно из таблицы 7-II, не всегда пептид, который часто встречался в промывочных заколах, также часто встречался и в пробах доноров КВВ. Например, DVFLGMFLYEYAR наблюдался во всех промывках, однако в пробах встретился всего в 18%. Соответственно, убирать из рассмотрения все пептиды, попавшие в этот список, не совсем верно, однако, в дальнейшем, при интерпретации результатов необходимо учитывать, что эти пептиды часто встречались при промывках.

Также для всех имеющихся экспериментов был сделан анализ по методу главных компонент (рисунок 19).

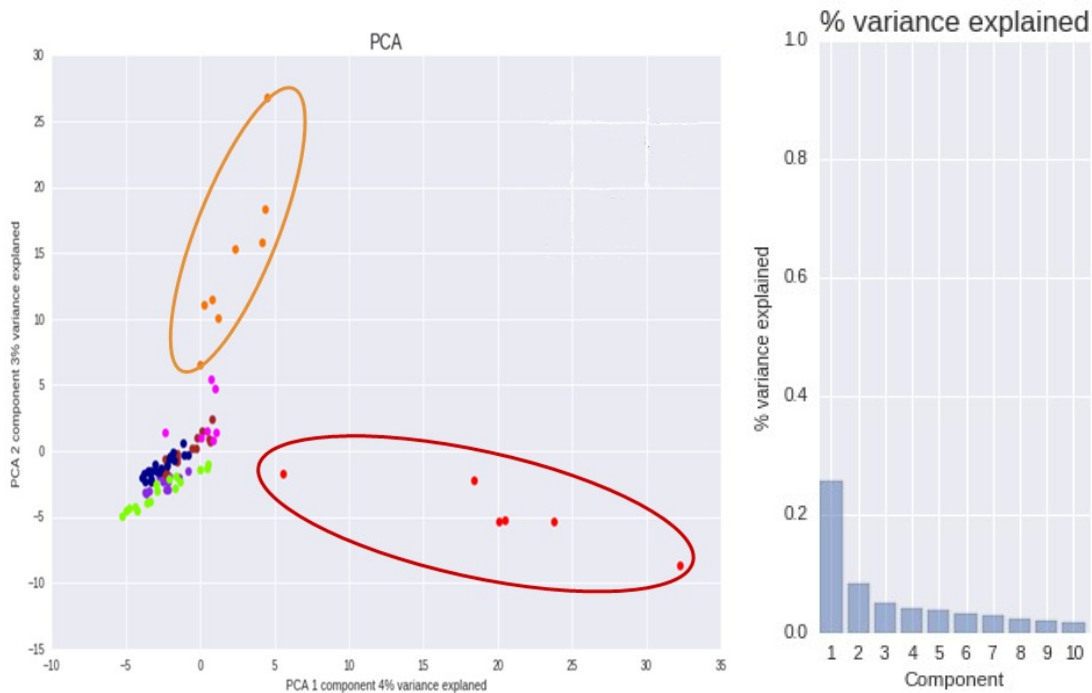


Рисунок 19. Анализ проведенных экспериментов по методу главных компонент (каждая точка на графике – это набор данных по одному эксперименту, т.е. анализу набора проб одной из целевых групп исследования, проведенному в одно время).

На рисунке 19 видно линейное скопление по центру точек и две выпавшие группы. Эти результаты означают, что большую часть экспериментов можно анализировать совместно, а два набора данных обладают иными характеристиками и должны быть удалены из рассмотрения по причине возможного внесения ошибки в итоговый результат анализа.

Такой же график был построен по времени проведения эксперимента (рисунок 20).

Видно, что цвета перемешиваются, что говорит о том, что большая часть результатов критически не загрязнена и может анализироваться совместно.

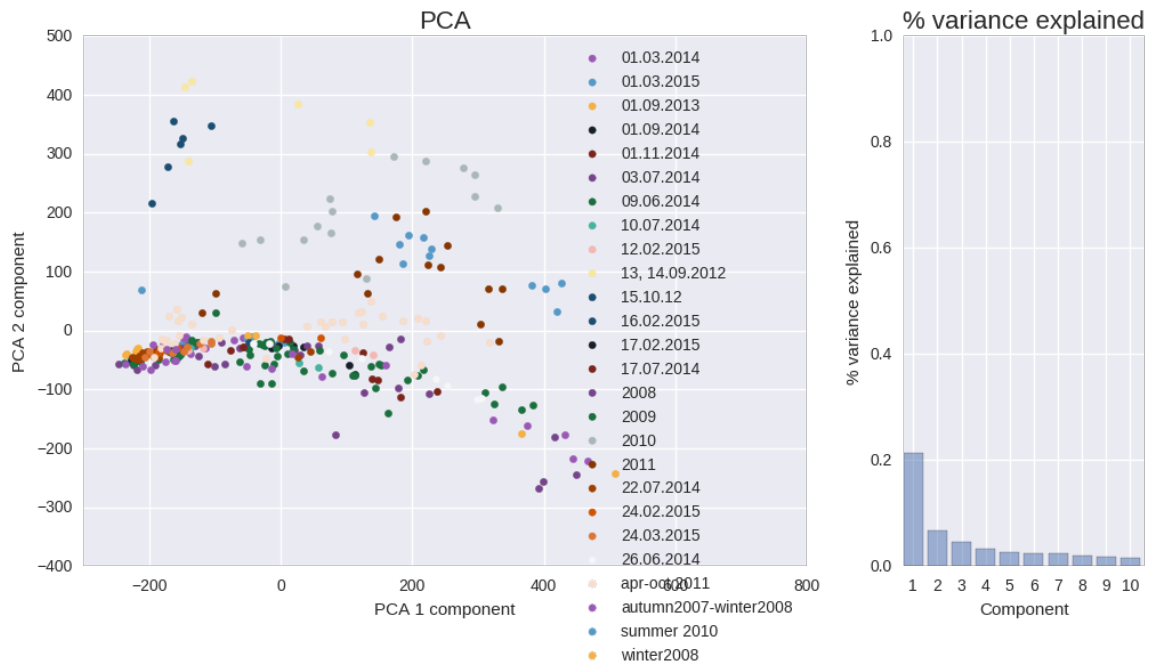


Рисунок 20. Анализ проведенных экспериментов по методу главных компонент (каждая точка на графике – это набор данных за указанный период времени).

8.4.3 Анализ идентифицированных пептидов в пробах КВВ, обработанных по предложенной в работе методике

После отсеивания экспериментов, выпавших на основании анализа по методу главных компонент (см. выше), в дальнейший анализ были взяты 52303 идентифицированных пептидов. Эти пептиды были ранжированы по длине (рисунок 21).

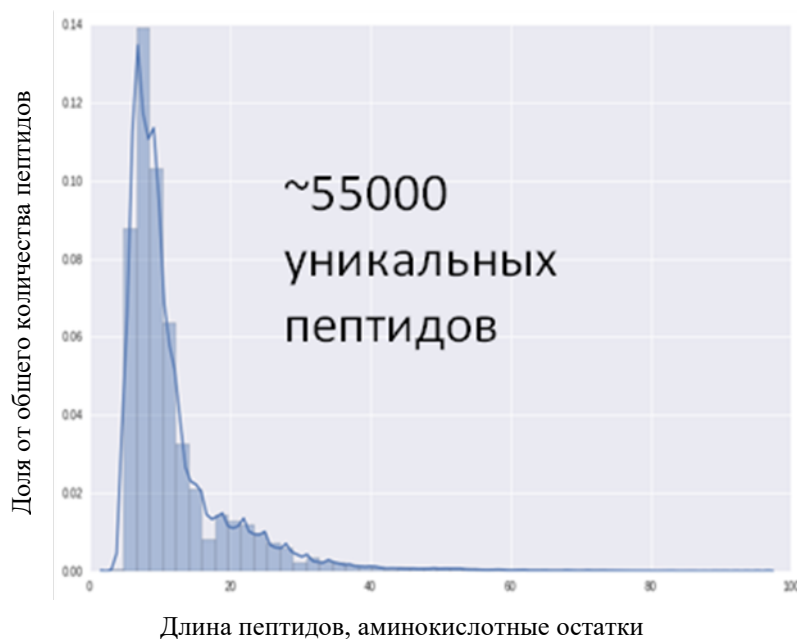


Рисунок 21. Распределение длин пептидов, идентифицированных в пробах КВВ и взятых в анализ после фильтрации редко встречающихся пептидов.

Как видно из представленных данных (рисунок 21), в пробах КВВ были представлены, в основном, короткие пептиды размером менее 20 аминокислотных остатков.

Помимо ранжирования пептидов, идентифицированных в пробах КВВ и включенных в анализ после удаления выпавших экспериментов, проведено сравнение пептидов по аминокислотному составу в сравнении с литературными данными по природному распределению аминокислот в мембранах и не мембранных структурах млекопитающих [Kumar, 2014] (рисунок 22).

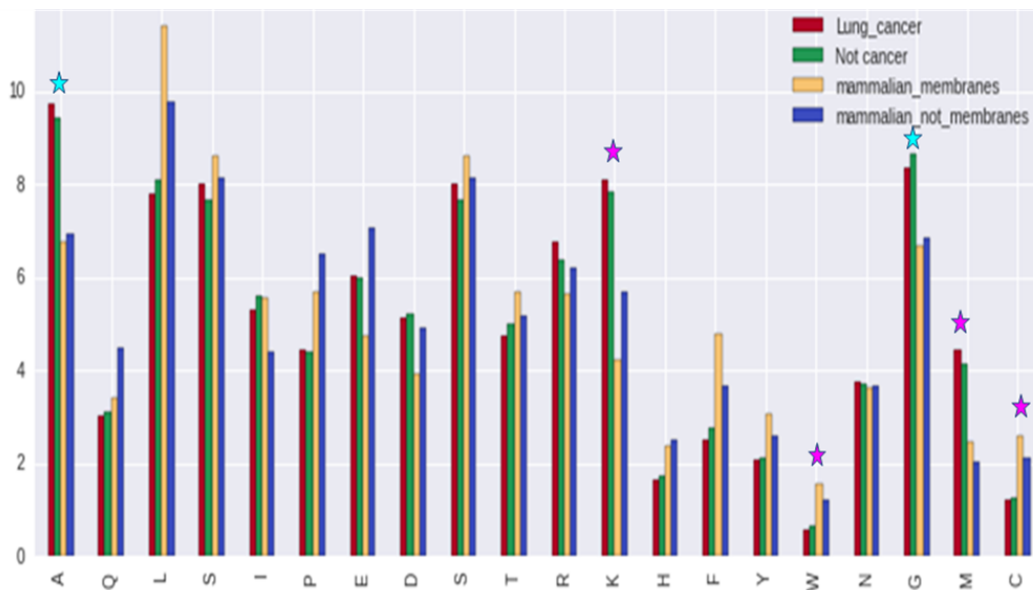


Рисунок 22. Аминокислотный состав пептидов, идентифицированных в пробах доноров КВВ, включенных в анализ, в сравнении с литературными данными по природному распределению аминокислот в мембранах и не мембранных структурах млекопитающих [Kumar, 2014]. Значимые отличия обозначены звёздочками.

Как видно из приведенных данных (рисунок 22), в пептидах КВВ наблюдалось существенно меньшее содержание триптофана и цитозина, существенно большее – аланина, глицина, метионина, аспарагина. Возможно, такое распределение связано с тем, что тяжелые аминокислоты реже попадают в выдыхаемый воздух, в отличие от легких аминокислот. Также можно заметить связь такого распределения с аминокислотным составом ферментов (консервативные и лабильные аминокислоты) и структурных компонентов клетки.

Также проанализировано среднее количество пептидов в одном образце (рисунок 23).

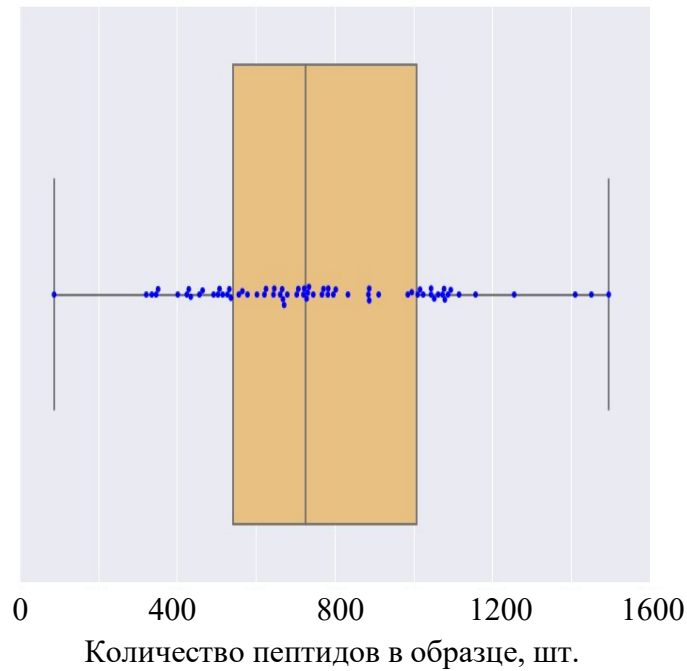


Рисунок 23. Среднее количество пептидов, идентифицированных в одном образце КВВ в рамках исследуемого подхода.

Как можно видеть из представленных данных (рисунок 23), среднее количество пептидов, идентифицированных в одном образце КВВ, составило примерно 700 штук.

На рисунке 24 представлено распределение количества идентифицированных пептидов в пробе КВВ в зависимости от наличия у донора диагноза «рак легкого».

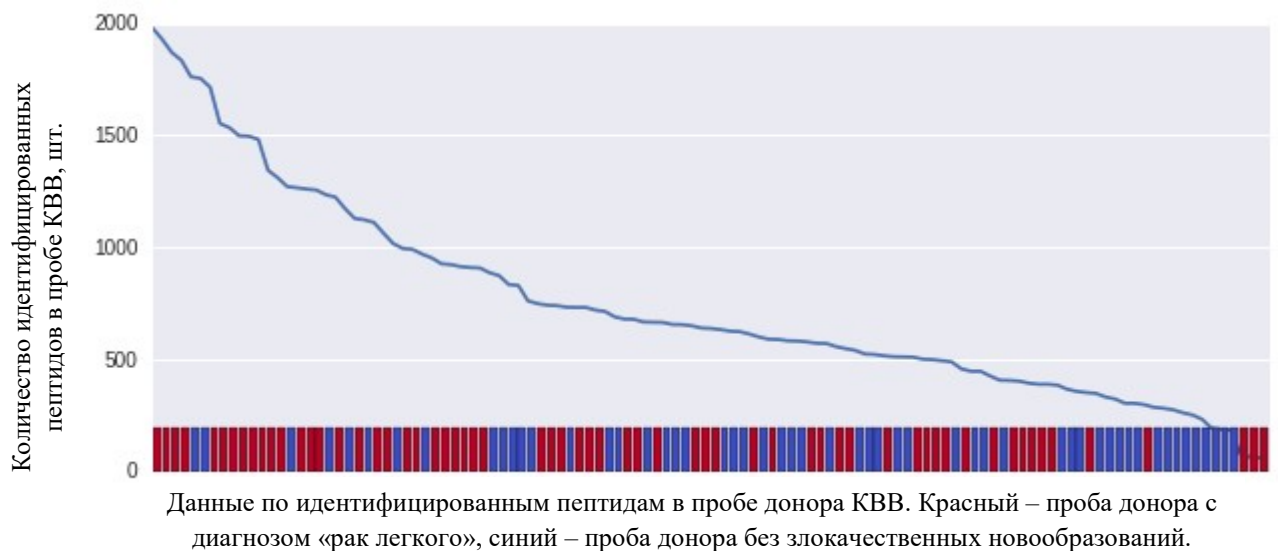


Рисунок 24. Количество пептидов, идентифицированных в пробе КВВ, в зависимости от наличия у донора диагноза «рак легкого» (отсортировано по убыванию).

Как видно из приведенных данных (рисунок 24), однозначного разделения по количеству идентифицируемых пептидов в пробе в зависимости от наличия рака легкого у донора не произошло, однако, можно заметить относительную «сконцентрированность» раковых проб у левого края графика, т.е. для них было характерно большее количество идентифицируемых пептидов в пробе. Этот вывод подтверждает тест Манна-Уитни, приведенный на рисунке 25.

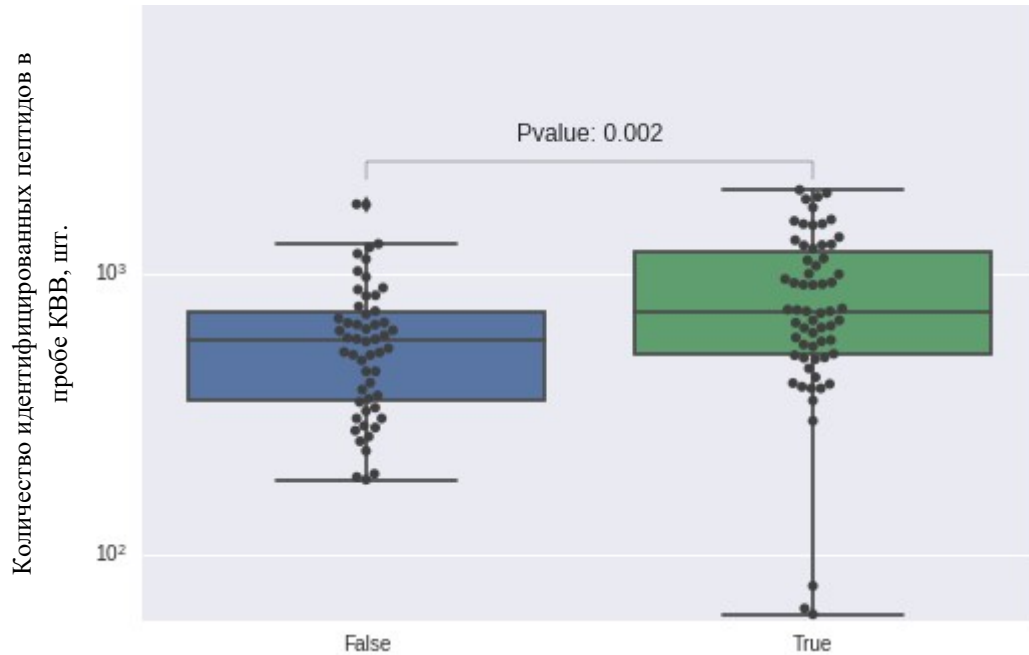


Рисунок 25. Воx-plot количества идентифицированных пептидов в пробах доноров КВВ. За нулевую гипотезу принимали наличие у донора рака легкого. Зеленое – гипотеза подтвердилась, синее – не подтвердилась.

По результатам теста Манна-Уитни, значение Р-критерия равно 0,002. У доноров с раком легкого количество идентифицированных пептидов в пробе, в среднем, значимо больше, чем количество идентифицированных пептидов в пробе донора без злокачественных новообразований. При этом нужно иметь в виду, что в тестировании принимали участие как здоровые доноры без патологий респираторной системы, так и пациенты с диагнозами «внебольничная пневмония» и «ХОБЛ».

После этого был проведен анализ частоты встречаемости идентифицированных пептидов в пробах КВВ. На рисунке 26 представлено распределение пептидов по количеству обнаружений.

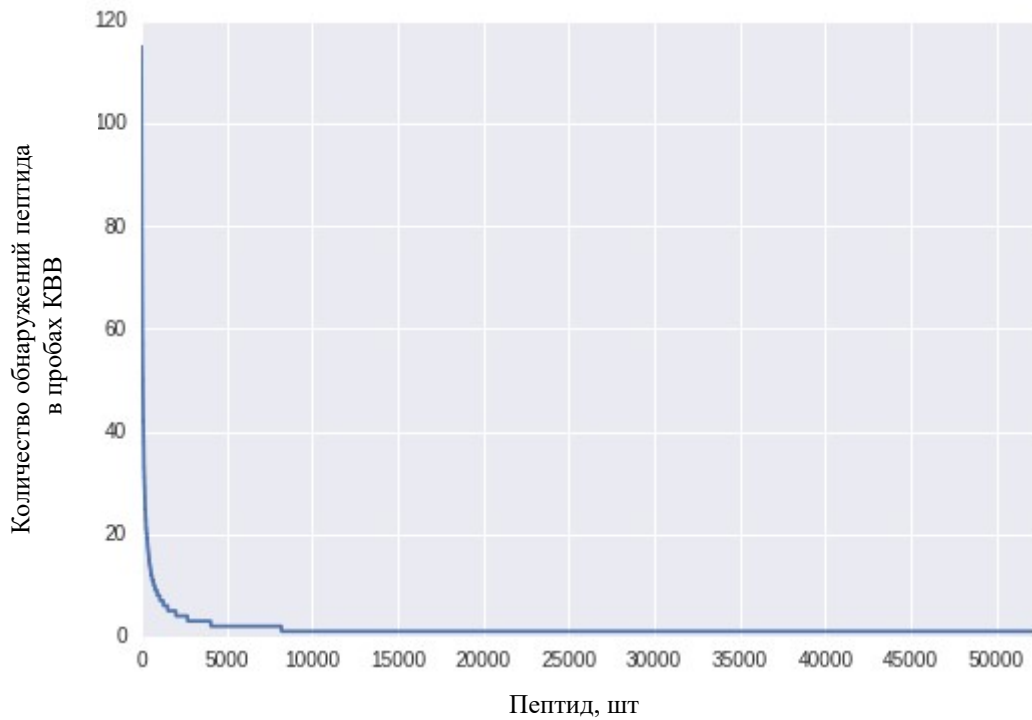


Рисунок 26. Распределение идентифицированных пептидов по количеству обнаружений в пробах КВВ.

Как видно из представленных данных (рисунок 26), основную массу идентифицированных пептидов составляли редко встречающиеся пептиды (встретившиеся менее, чем у 20 доноров КВВ). Из дальнейшего анализа исключались пептиды, которые наблюдались в менее, чем 5% проб. Всего осталось 787 пептидов (рисунок 27).

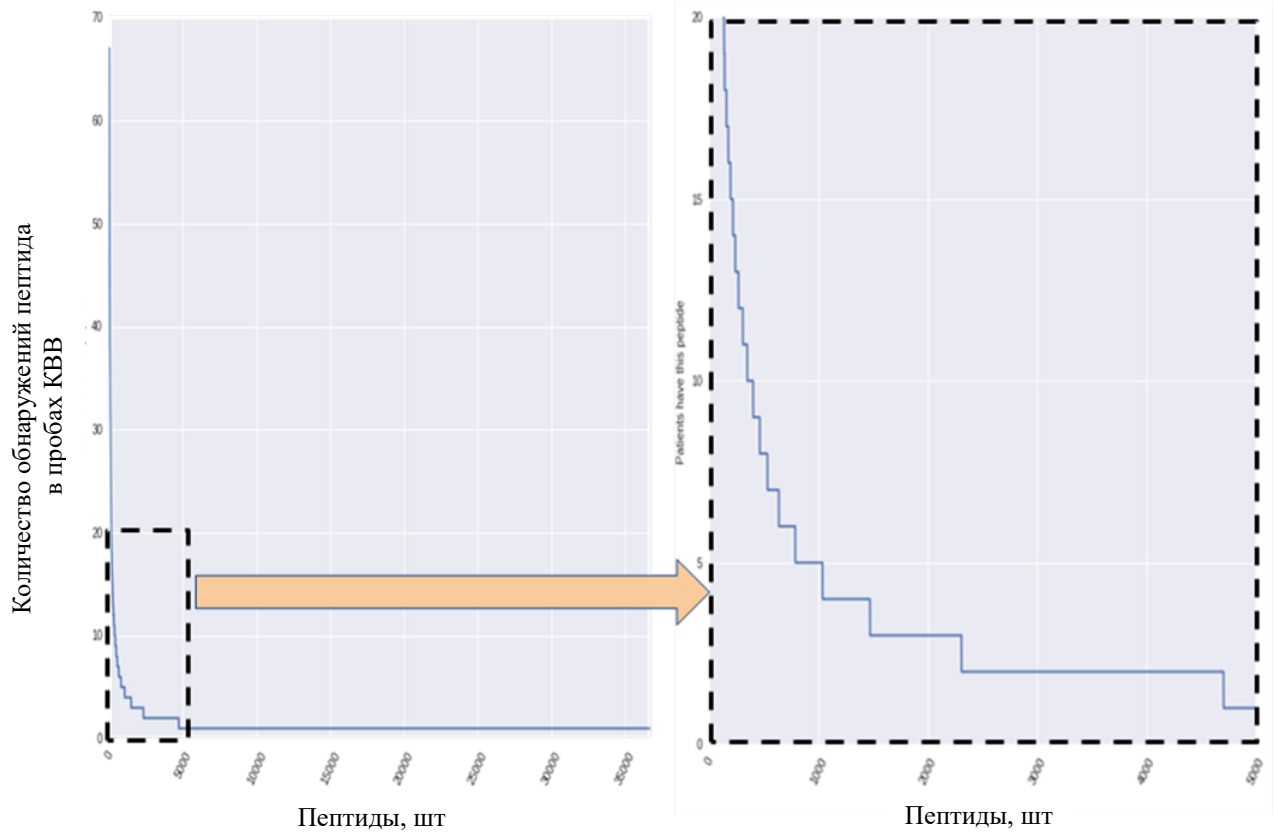
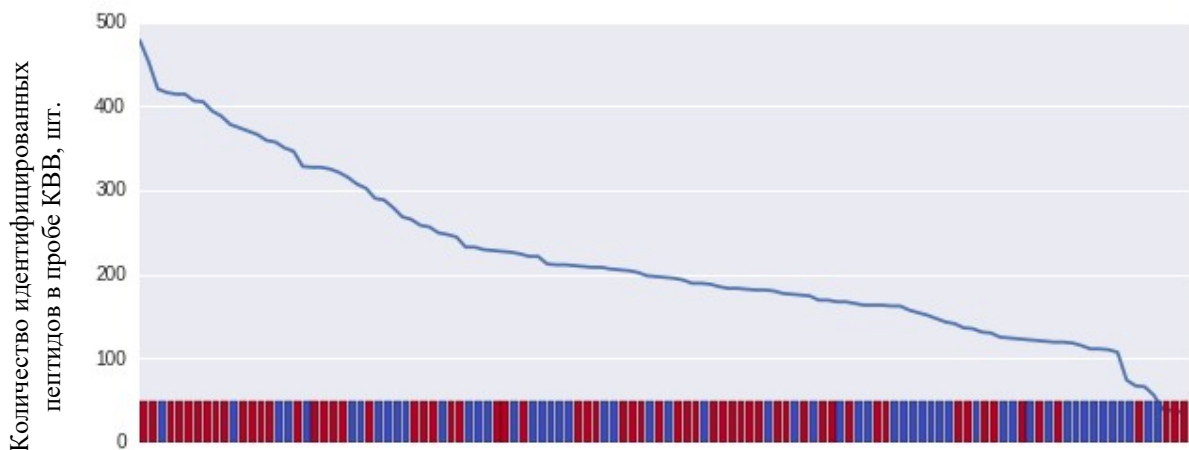


Рисунок 27. Результат фильтрации редко встречающихся пептидов среди идентифицированных пептидов в пробах КВВ (пептидов, встретившихся менее, чем в 5% проб).

После фильтрации редко встречающихся пептидов снова сделана проверка, сохранилась ли ассоциация количества идентифицированных пептидов в пробе КВВ с диагнозом «рак легкого». Результаты проверки приведены на рисунке 28.



Данные по идентифицированным пептидам в пробе донора КВВ. Красный – проба донора с диагнозом «рак легкого», синий – проба донора без злокачественных новообразований.

Рисунок 28. Количество пептидов, идентифицированных в пробе КВВ, в зависимости от наличия у донора диагноза «рак легкого» (после фильтрации редко встречающихся пептидов, отсортировано по убыванию).

Как видно из приведенных данных (рисунок 28), различия в количестве идентифицированных пептидов в пробе КВВ в зависимости от наличия диагноза «рак легкого» после фильтрации редко встречающихся пептидов не наблюдается. Этот вывод подтверждает тест Манна-Уитни, приведенный на рисунке 29.

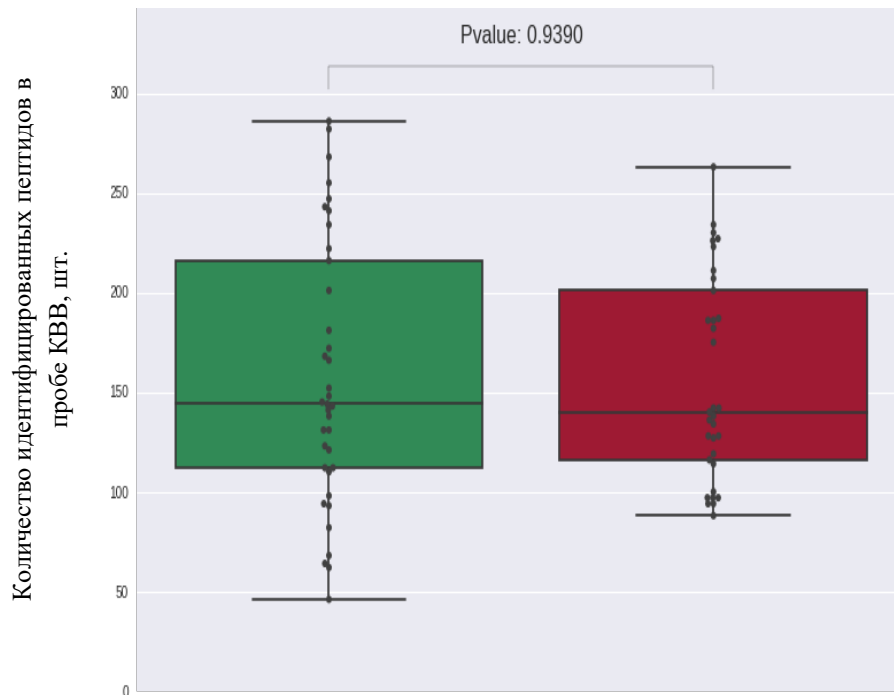


Рисунок 29. Вох-plot количества идентифицированных пептидов в пробах доноров КВВ после фильтрации редко встречающихся пептидов. За нулевую гипотезу принимали наличие у донора рака легкого. Красное – гипотеза подтвердилась, зеленое – не подтвердилась.

По результатам теста Манна-Уитни, значение Р-критерия равно 0.9, из чего следует, что после фильтрации пептидов, встречающихся менее, чем в 5% проб, количество идентифицированных пептидов в пробе КВВ, в среднем, не различается и не ассоциировано с наличием у донора злокачественных новообразований. При этом нужно иметь в виду, что в тестировании принимали участие как здоровые доноры без патологий респираторной системы, так и пациенты с диагнозами «внебольничная пневмония» и «ХОБЛ». Такое различие в результатах анализа до и после фильтрации редко встречающихся пептидов может говорить о наличии индивидуального полиморфизма проб КВВ доноров с диагнозом «рак легкого», что приводит к идентификации большого количества единично встречающихся пептидов.

Также была проведена иерархическая кластеризация образцов (рисунок 30).

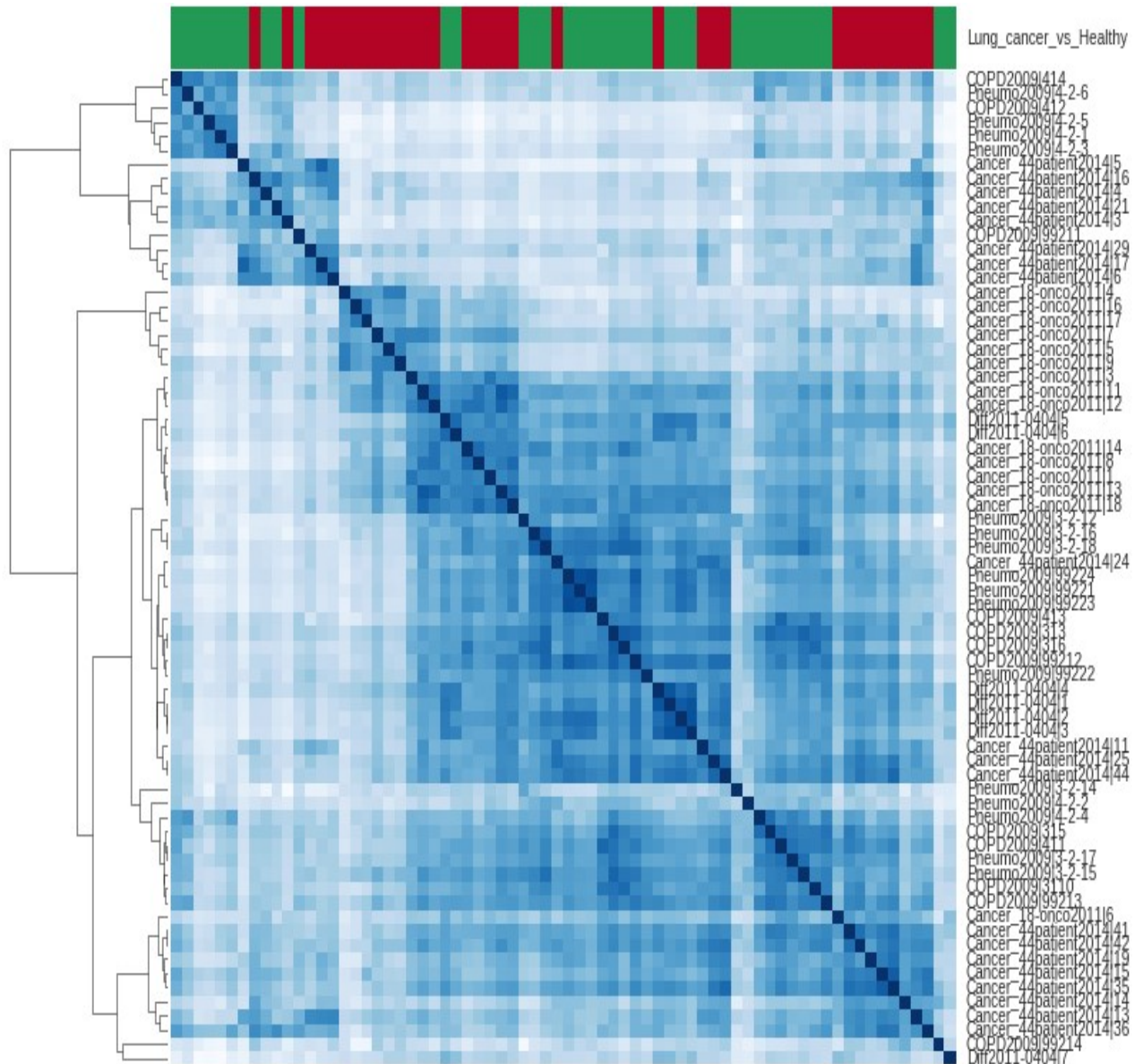


Рисунок 30. Иерархическая кластеризация проб доноров КВВ, включенных в анализ, на основании анализа идентифицированных пептидов. Красное – пробы доноров КВВ с диагнозом «рак легкого», зеленое – пробы доноров КВВ без злокачественных новообразований.

Как видно из приведенных данных (рисунок 30), иерархическая кластеризация образцов по данным анализа идентифицированных пептидов не позволила выделить группы «рак-не рак», однако, равномерное перемешивание проб подтвердило выводы анализа по методу главных компонент, что пробы, включенные в анализ, можно анализировать совместно.

8.5 Построение аналитических моделей, предсказывающих наличие фокусного заболевания у донора КВВ

В рамках решения задачи диагностики заболеваний по набору пептидов, обнаруживаемых в конденсате выдыхаемого воздуха, должна учитываться дополнительная информация, а именно: пол, возраст, наследственность, статус курильщика, сопутствующие заболевания донора. Классические статистические тесты (Chi-квадрат, Фишера и т.п.) для таких задач не применимы, т.к. они рассчитаны на небольшое количество признаков (в идеале 2). В таких случаях используются так называемые методы многопараметрической статистики. В частности, исходя из специфики данных, наиболее перспективным выглядит использование алгоритмов, основанных на построении "деревьев", т.е. алгоритмов, где результирующая модель представляет собой набор правил типа "если ((есть пептид1 и есть пептид2) или (есть пептид3 и нет пептида4)), то <диагноз>", а также комбинацию из большого количества таких правил. Качество таких моделей оценивается точностью предсказаний, также модели валидируются методом скользящего контроля (k-fold cross-validation), когда данные разбиваются на k блоков, затем один блок удаляется из рассмотрения и строится модель с использованием оставшихся k-1 блоков, после чего полученная модель применяется к удаленному блоку; данная процедура повторяется для каждого из k блоков. Метод скользящего контроля позволяет оценить точность модели на данных, которые не участвовали в ее построении. Таким образом, на выходе можно получить модель, которая "ставит диагноз", оценку ее точности (на имеющихся данных) и оценку ее точности на "реальных" данных.

Для построения аналитической модели, основанной на анализе идентифицированных в КВВ пептидов, использованы методы, обычно применяемые к данным по уровням экспрессии генов, т.е. количественным характеристикам, а не к вероятностям. За «количественную» характеристику принимался Mascot Score идентифицированного пептида, допуская, что эта численная характеристика в данном исследовании может быть приравнена интенсивности сигнала для гена в RNAseq.

Для всех пептидов был рассчитан fold change между значениями Mascot Score доноров КВВ с раком легкого и без по следующей формуле:

$$Peptide\,fold\,change = \frac{mean(peptidescore \in cancer)}{mean(peptidescore \in notcancer) + 0.0001}$$

+0.0001 добавлялось, чтобы не делить на 0.

На рисунке 31 приведен логарифм значений fold change для всех пептидов, участвующих в анализе, отсортированный по убыванию.

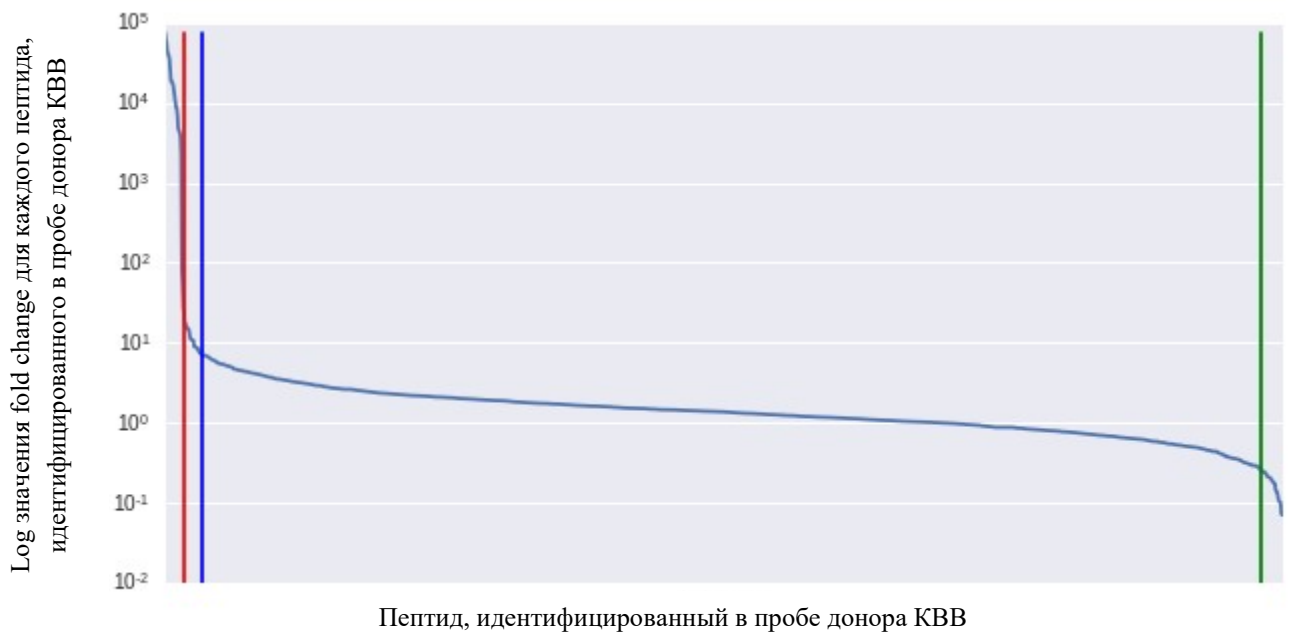


Рисунок 31. Log fold change для каждого пептида, идентифицированного в пробах КВВ и включенного в анализ после отсева выпавших экспериментов и фильтрации редко встречающихся пептидов.

Слева от красной линии (25 пептидов, $\text{foldchange} > 50$) находятся пептиды, которые практически не встречаются в не раковых пробах и встречаются в раке. Слева от синей линии (52 пептида, включая 25 левее красной) находятся пептиды, у которых, в среднем, Mascot Score был более, чем в 7 раз выше ($\text{foldchange} > 7$) у доноров КВВ с раком легкого, чем у доноров КВВ без рака легкого. Справа от зеленой линии (29 пептидов) находятся пептиды, у которых Mascot Score был понижен в пробах доноров КВВ с раком легкого относительно проб доноров КВВ без рака легкого ($\text{foldchange} < 0.25$).

Совмещенная таблица пептидов, идентифицированных в пробах доноров КВВ и включенных в анализ после отсева выпавших экспериментов и фильтрации редко встречающихся пептидов, отсортированная по значению fold change, представлена в Приложении.

В соответствии со значением fold change, были отобраны пептиды левее синей линии и правее зеленой (рисунок 31), всего 22 (таблица 21).

Таблица 21. Пептиды, выбранные для построения аналитической предсказательной модели.

Номер	Пептид	Score*	Коэффициент логистической регрессии
1	AFIQLWAFDAVK	0,549283	0,103
2	RSVMDSLTR	0,615143	0,392
3	MVANIDK	0,668160	0,247
4	ASAEKIMQIGK	0,708931	0,596
5	AEITDMR	0,747760	0,621
6	GTALAEVLK	0,785544	-0,252
7	LSGLLDLALGKDYVR	0,816607	-0,196
8	MFANEAAK	0,841249	-0,344
9	VNVDEVGGEALGR	0,859618	0,102
10	EEDGMIR	0,874701	0,796
11	DYYQGKYHLTHDEMDR	0,887246	0,679
12	FPGLPE	0,902479	0,424
13	DGKSGLIK	0,911888	0,252
14	FIGPNADVIR	0,921595	0,554
15	AEDVTRR	0,930257	-0,458
16	AASPSK	0,937724	-0,479
17	DVMMHVAS	0,941607	-0,334
18	VPLVPR	0,945639	-0,493
19	ILGMLIVS	0,949373	0,431
20	NNFCARSSE	0,954898	-0,311
21	TMRSKSLAK	0,958333	0,698
22	APAPNEAGR	0,961768	-0,402

*определен автоматически в соответствии со значением fold change и встречаемости в раковых или не раковых пробах на этапе отбора

Далее проведено машинное обучение на логистической регрессии с 10-кратной кроссвалидацией ($k\text{-folds}=10$). Оценка прогностической силы построенной модели проводилась с помощью ROC_AUC Score, оценка качества работы предложенного алгоритма проводилась с помощью F-меры (F_1).

На рисунке 32 показано, как возрастало значение ROC_AUC Score при добавлении каждого из 22 выбранных пептидов.

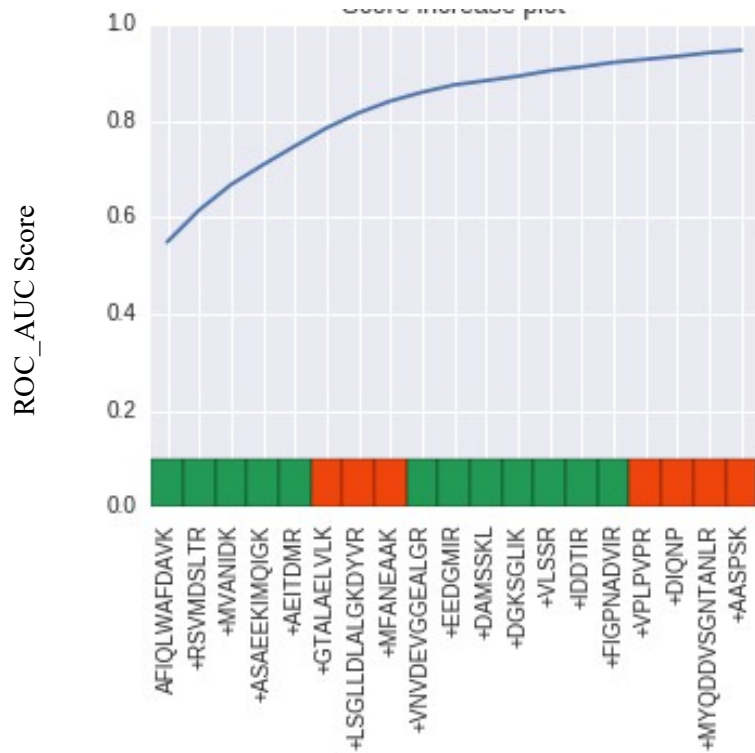


Рисунок 32. Зеленые – пептиды, встречавшиеся с высокой частотой в раковых пробах и практически не встречавшиеся в не раковых пробах, красные – пептиды, встречавшиеся с высокой частотой в не раковых пробах и практически не встречавшиеся в раковых пробах.

На рисунке 33 изображена построенная аналитическая модель по предсказанию наличия рака легкого у донора на основании анализа пептидного состава КВВ.

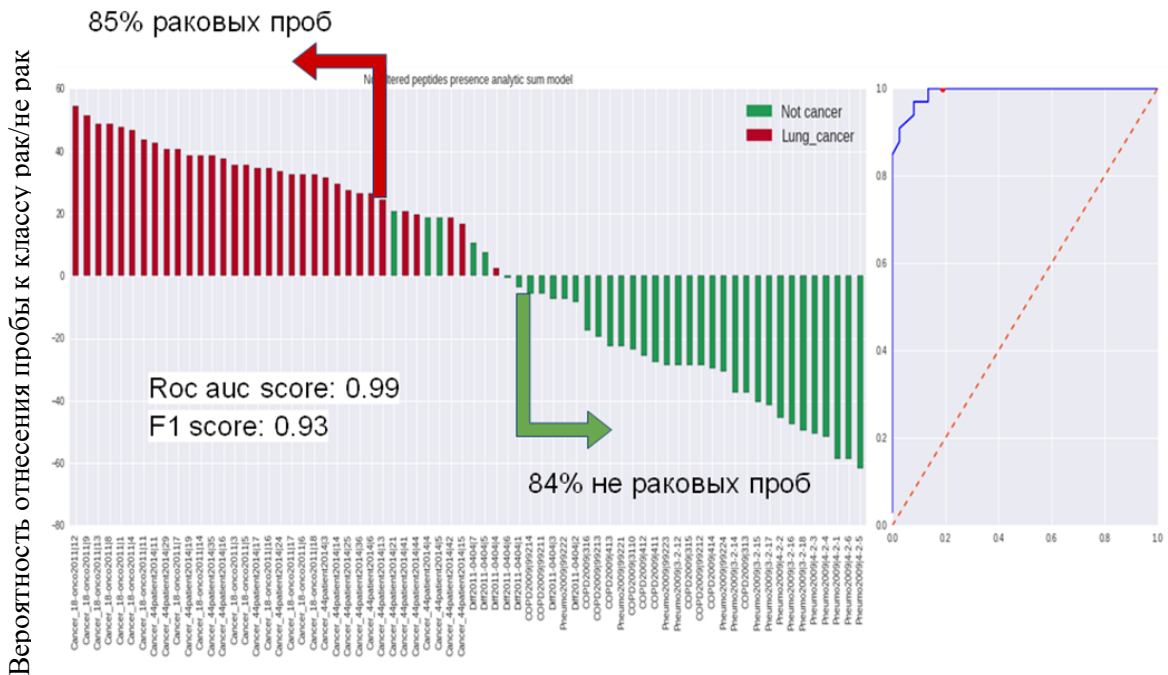


Рисунок 33. Красные пробы – раковые, зеленые пробы – не раковые. На графике приведены значения метрик машинного обучения предсказания: $F_1 = 0,93$; $Roc_auc = 0,99$.

Итоговые коэффициенты логистической регрессии для выделенных пептидов отображены на рисунке 34.

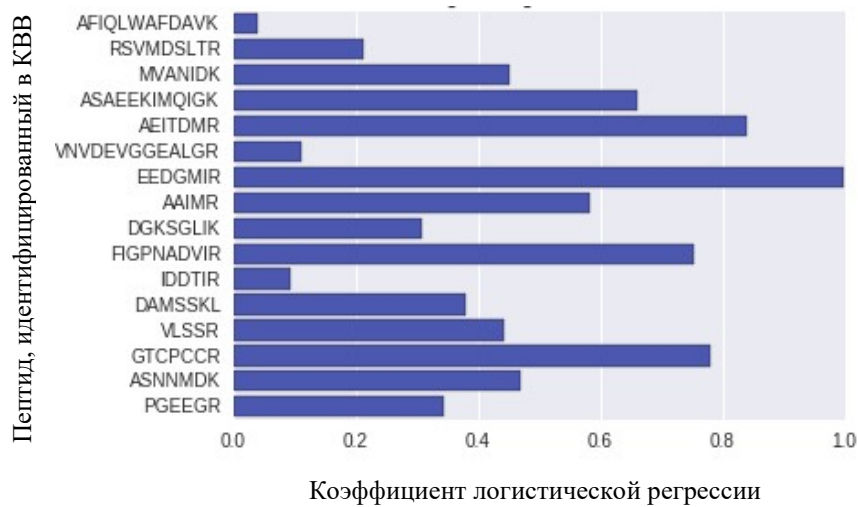


Рисунок 34. Коэффициенты логистической регрессии для выделенных пептидов

Предложенная аналитическая модель дает численные значения для проб по выделенным пептидам, которые относят пробу либо к раковой, либо к не раковой, и, по результатам проверки на группе проб, не включенных в машинное обучение, обладает хорошей прогностической силой ($F_1 = 0.93$; $Roc_auc = 0.99$). Поскольку в исследовании участвовали не только здоровые доноры, но также доноры с пневмонией, ХОБЛ и раком легкого, отягощенными сопутствующими заболеваниями, данная модель может использоваться для диагностики рака легкого по анализу пептидного состава КВВ даже для пациентов с отягощенным анамнезом.

8.6 Возврат на уровень белков при анализе пептидов по описанному методу

Принимая во внимание сложности работы с КВВ, описанные выше, а также полученные результаты, можно заключить, что пептиды КВВ больше подходят в качестве основы для разработки метода диагностики рака легкого, чем белки. Однако анализ идентифицируемых белков необходим для понимания происходящих в организме процессов и сопоставления результатов с опубликованными данными по белковым и пептидным маркерам рака легкого в других биологических жидкостях. Использованный в работе метод пробоподготовки КВВ привел к образованию триптических пептидов в исследуемых образцах, что сделало возможным возврат на белковый уровень анализа. Идентифицированные программой Mascot пептиды аннотировались по UniRef100 базе человеческих белков (сборка UniProtKB release 2017_06) с помощью сервиса Protein Information Resource [Wu et al., 2003]. Для каждого белка,

подтвержденного минимум двумя пептидами, получены аннотация функции, ген, связи с другими белками с помощью сервиса GeneMania [Warde-Farley et al., 2010]. На рисунке 35 представлен анализ белков, включающих пептиды, встречающиеся во всех группах доноров КВВ.

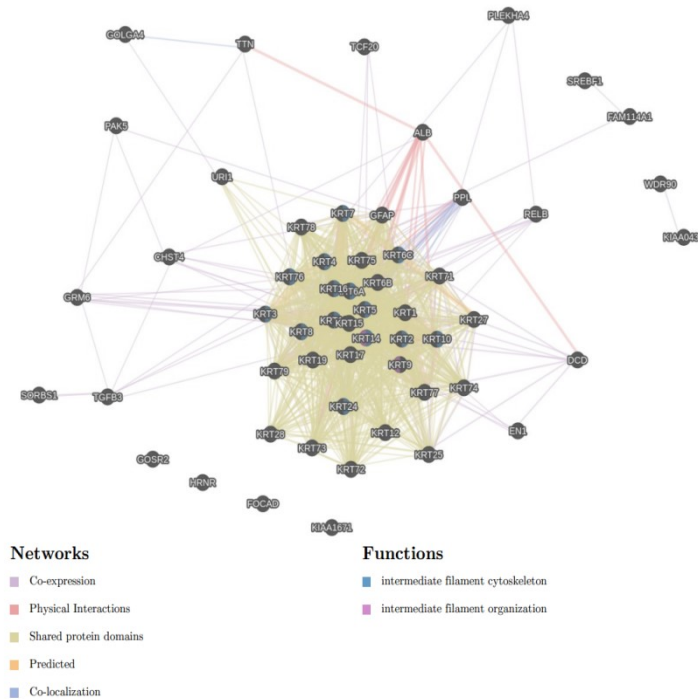


Рисунок 35. Карта взаимосвязей белков, имеющих в своем составе не менее двух идентифицированных пептидов из списка пептидов, определяемых во всех группах доноров, построенная в программе GeneMania.

Было обнаружено много совпадений с ранее идентифицированным белковым «фоном» КВВ (например, дермцидин, альбумин, хорнерин, цитоскелетные кератины I и II типа), что говорит о преемственности результатов двух видов анализа КВВ, использованных в работе.

Тем не менее, использование пептидов, идентифицируемых программой Mascot без соотнесения с белками на начальном этапе анализа, привело к расширению списка анализируемых пептидов. Так, при анализе группы пептидов, вносящих наиболее весомый вклад в отнесение пробы КВВ к раку легкого на основании работы аналитической модели, рассмотренной выше, обнаружено, что около половины всех «раковых» пептидов определились как не человеческие (не нашлись в базах данных по белкам человека).

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

Основной результат исследования – информация о сравнительном белковом и пептидном составе КВВ больных раком легкого и пациентов с ХОБЛ и пневмонией – является абсолютно новым. На основании полученных данных анализа протеома конденсата выдыхаемого воздуха можно утверждать, что белковые профили различных групп доноров поддаются разделению и существует вероятность выделения специфической группы белков, присущей определенному состоянию/заболеванию респираторной системы. Следует отметить, что протеом онкологических больных, определяемый в КВВ, отличается от протеомов не только здорового контроля, но и от протеомов больных ХОБЛ и пневмонией той же возрастной группы, что свидетельствует о потенциальной возможности использования КВВ в качестве скринингового теста с последующей проверкой другими методами, например, компьютерной томографией. Было показано, что анализ КВВ представляется перспективным неинвазивным методом диагностики состояния респираторного тракта человека, открывающим, в сочетании с биоинформатическими подходами, новые возможности в области персонализированной медицинской диагностики.

1. Разработаны методические основы определения белков и пептидов в конденсате выдыхаемого воздуха (КВВ) человека, сравнение методов пробоподготовки и концентрирования показало, что лиофилизация является наиболее мягким, эффективным и экологичным методом концентрирования проб КВВ перед протеомным анализом.

2. На основании исследования белкового состава КВВ здоровых доноров, определено, что основными белками конденсата выдыхаемого воздуха являются цитоскелетные кератины (было идентифицировано 54 функциональных кератиновых гена). Идентифицированный в пробах здоровых доноров набор белков, включающий как кератины, так и некератиновые белки, служил белковым фоном КВВ для дальнейших исследований. Определение ряда белков в окружающем воздухе показало, что некоторые белки могут попадать в респираторный тракт экзогенным путем, задерживаться в респираторной системе и возвращаться в окружающую среду.

3. На основании исследования белкового состава КВВ пациентов с ХОБЛ, пневмонией и раком легкого, показано, что результаты анализа протеомов по группам согласуются с клинической картиной рассматриваемых заболеваний. 42 белка не кератиновой природы были

идентифицированы только в КВВ доноров с диагностированным раком легкого 1-2 стадии и отсутствовали как в КВВ группы здорового контроля, так и в КВВ доноров с ХОБЛ и внебольничной пневмонией. На основании подробного аннотирования с помощью биоинформатических ресурсов, а также анализа частот встречаемости, были выделены 19 белков, которые предложены в качестве диагностической панели для рака легкого.

4. В ходе анализа пептидного состава КВВ, идентифицировано 55000 уникальных пептидов, которые были ранжированы по длине и аминокислотному составу, проведено сравнение когорт проб больных раком легкого и остальных доноров по пептидной нагрузке, проведена иерархическая кластеризация образцов, а также ранжированы все пептиды на основании их встречаемости в пробах и Mascot Score, допуская, что эта численная характеристика в эксперименте может быть аналогична интенсивности сигнала для гена в RNAseq.

5. На основании исследования пептидного состава КВВ всех исследуемых групп доноров, построена линейная аналитическая модель прогнозирования наличия у донора рака легкого и проверена с помощью группы доноров, не включенных в машинное обучение. Модель показала хорошую прогностическую способность ($AUC=0.99$), определив раковые образцы. Для некоторых из пептидов, имеющих наибольший вес при построении модели, была проанализирована принадлежность к белкам и найдены соответствия с данными по белковому составу КВВ.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Приказ Министерства здравоохранения РФ "Об утверждении порядка проведения диспансеризации определенных групп взрослого населения" от 3 февраля 2015 г. № 36ан // "Официальный интернет-портал правовой информации" (www.pravo.gov.ru) . - 4 марта 2015 г.
2. Автономов, Д. М. Создание базы данных точных массово-временных меток для качественного и количественного подхода в исследовании протеома мочи человека с использованием изотопного мечения / Д. М. Автономов, И. А. Агрон, А. С. Кононихин, Е. Н. Николаев // ТРУДЫ МФТИ. - 2009. - Т. 1, № 1. - С. 24-30.
3. Агаджанян, Н. А. Функции организма в условиях гипоксии и гиперкапнии / Н. А. Агаджанян. - М. : Медицина, 1986. - 269 с.
4. Анаев, Э. Х. Диагностика заболеваний легких на основе протеомного анализа конденсата выдыхаемого воздуха / Э. Х. Анаев, К. Ю. Федорченко, М. Э. Кушаева, А. М. Рябоконт, А. С. Кононихин, В. В. Бармин, О. В. Пикин, И. А. Попов, Е. Н. Николаев, С. Д. Варфоломеев, А. Г. Чучалин // Пульмонология. - 2017. - Т. 27, № 2. - С. 187–197.
5. Глазачев, О. С. Закономерности мультипараметрического взаимодействия функциональных систем у детей в радиоэкологически неблагоприятной среде: автореф. дис. ... д-ра мед. наук / О. С. Глазачев. — М., 1997. — 43 с.
6. Добрых, В. А. Диагностическое значение цитологического исследования секрета нижних дыхательных путей / В. А. Добрых, И. Е. Мун, О. А. Ковалева, А. А. Дигор, И. В. Уварова, А. М. Макаревич // Дальневосточный медицинский журнал. - 2013. - № 1. - С. 125-129.
7. Дулина Т. Р. Биомаркеры воспаления у больных бронхиальной астмой в индуцированной мокроте и выдыхаемом воздухе: автореф.дис. канд.мед.наук. – Москва, 2009. – 24 с.
8. Ершов, А. Л. Капнография и возможности её применения в условиях скорой медицинской помощи. Учебно-методическое пособие / А. Л. Ершов. - СПб.: Издательство «Авеню-АРТ», 2013.- 31 с.
9. Исупов, И. Б. Возрастные особенности показателей форсированного дыхания / И. Б. Исупов, Е. Ю. Надежкина, О. С. Филимонова // Клиническая геронтология. - 2013. - Т. 19, № 7-8. - С. 31-33.
10. Исупов, И. Б. Способ определения биологического возраста человека по показателям форсированной спирометрии / И. Б. Исупов, Е. Ю. Надежкина, О. С. Филимонова // Клиническая геронтология. 2014. - Т. 20, № 5-6. - С. 41-43.

11. Исупов, И. Б. Создание и совершенствование учебных приборов междисциплинарного назначения как составная часть модернизации образовательного процесса (Опыт конструирования. Обобщение) / И. Б. Исупов - Волгоград : Изд-во ВГАПО, 2015. - 224 с.
12. Калмыкова, Ю. С. Влияния комплексной программы физической реабилитации для больных инфильтративным туберкулезом легких без бактериовыделения на показатели функционального состояния дыхательной системы и на динамику морфологических изменений в легких / Ю. С. Калмыкова // ППМБПФВС. - 2009. - № 4. - С.43-47.
13. Кононихин, А. С. Протеомный анализ конденсата выдыхаемого воздуха в целях диагностики патологий дыхательной системы / А. С. Кононихин, К. Ю. Федорченко, А. М. Рябоконт, Н. Л. Стародубцева, И. А. Попов, М. Г. Завьялова, Э. Х. Анаев, А. Г. Чучалин, С. Д. Варфоломеев, Е. Н. Николаев // Биомедицинская химия. - 2015. - Т. 61, № 6. - С. 777–780.
14. Курова В. С. Масс-спектрометрический мониторинг белкового состава конденсата выдыхаемого воздуха больного, перенесшего трансплантацию легких / В. С. Курова, Э. Х. Анаев, А. С. Кононихин, И. А. Попов, К. Ю. Федорченко, Е. Н. Николаев, С. Д. Варфоломеев, А. Г. Чучалин // Известия Академии наук. Серия химическая. - 2010. - Т. 10, № 1. - С. 284–288.
15. Лебедев, А.Т. Масс-спектрометрия в органической химии / А.Т. Лебедев. - М.: БИНОМ. Лаборатория знаний, 2003. - 493 с.
16. Лещук, Т. Ю. Ранняя рентгенологическая диагностика рака легкого в Гродненской области / Т. Ю. Лещук, С. Э. Савицкий, Ф. С. Миклашевич // Журнал ГрГМУ. - 2010. - Т. 2, № 30. - С.138-140.
17. Мирошниченко, И. И. Биомаркеры в современной медико-биологической практике / И. И. Мирошниченко, С. Н. Птицина // Биомедицинская химия. - 2009. - Т. 55, № 4. - С. 425-440.
18. Мозалевский, А. Ф. Состав метаболитов арахидоновой кислоты в крови и слюне детей с бронхиальной астмой / А. Ф. Мозалевский, Т. Д. Травянка, А. А. Яковлев и др. // Укр.биохим. ж. - 1997. - Т. 69. - С. 162-168.
19. Петухов, В. И. Роль классических рентгенологических методов исследования и компьютерной томографии в диагностике гнойных заболеваний легких и плевры / В. И. Петухов, М. О. Русецкая, С. В. Шаврова // Проблемы здоровья и экологии. - 2010. - Т. 1, № 23. - С. 59-64.
20. Пшеннова, В. С. Состояние респираторной системы при ожирении / В. С. Пшеннова, И. С. Ежова, М. Кхир Бек, О. В. Александров // Рос мед ж. - 2012. - № 4. - С. 6-11.
21. Рябоконт, А. М. Сравнительный протеомный анализ конденсата выдыхаемого воздуха у пациентов с раком легкого методом масс-спектрометрии высокого разрешения / А. М.

- Рябокоть, Э. Х. Анаев, А. С. Кононихин, Стародубцева Н.Л., Киреева Г.Х., Попов И.А., Бармин В.В., Багров В.А., Пикин О.В., Николаев Е.Н., Варфоломеев С.Д. // Пульмонология. — 2014. — № 1. — С. 5–11.
22. Самсонова, М. В. Диагностические возможности бронхоальвеолярного лаважа / М. В. Самсонова // Практическая пульмонология. - 2006. - № 4. - С. 8-12.
23. Сафонов, Д. В. Динамическое ультразвуковое исследование грудной клетки при пневмониях у детей / Д. В. Сафонов, Т. И. Дианова, О. И. Голубцова, О. Н. Иванова // Фундаментальные исследования. - 2014. - № 10-9. - С. 1799-1804.
24. Северин, Е.С. Биохимия: Учебник для вузов / Е.С. Северин. - М.: ГЭОТАР-МЕД. - 2003. - 703 с.
25. Сиваш, Э. С. Рентгенологическое исследование при хроническом панкреатите / Э. С. Сиваш // ЭКГ. - 2008. - № 7. - С. 82-86.
26. Тюрин, И. Е. Компьютерная томография органов дыхания / И. Е. Тюрин // Практическая пульмонология. - 2003. - № 3. - С. 11-15.
27. Тюрин, И. Е. Рентгенологическое исследование при обструктивных заболеваниях легких / И. Е. Тюрин // Практическая пульмонология. - 2004. - № 4. - С.29-35.
28. Тюрин И. Е. Скрининг заболеваний органов дыхания: современные тенденции / И. Е. Тюрин // Практическая пульмонология. - 2011. - № 2. - С. 12-16.
29. Фёдоров, В. И. Белки, пептиды и аминокислоты выдыхаемого воздуха как маркеры бронхо-легочных заболеваний / В. И. Фёдоров, А. А. Карапузиков, М. К. Старикова // Бюллетень сибирской медицины. - 2013. - № 6 - С. 167-174.
30. Федорченко, К. Ю. Ранняя диагностика рака легкого на основе анализа протеома конденсата выдыхаемого воздуха / К. Ю. Федорченко, А. М. Рябокоть, А. С. Кононихин, С. И. Митрофанов, В. В. Бармин, О. В. Пикин, Э. Х. Анаев, И. В. Гачок, И. А. Попов, Е. Н. Николаев, А. Г. Чучалин, С. Д. Варфоломеев // Вестник Московского университета. Серия 2: Химия. - 2016. - № 2. - Р. 112–120.
31. Франкевич, В. Е. Высокоэффективный анализ и идентификация пептидов и белков с использованием масс-спектрометрии с квадрупольной ловушкой / В. Е. Франкевич, Р. Х. Зиганшин, Н. Б. Поляков, В. А. Гринкевич, А. С. Арсеньев // Масс-спектрометрия. - 2006. - Т. 3, № 3. - С. 161–168.
32. Черняев, А. Л. Гистологическая диагностика заболеваний легких / А. Л. Черняев, М. В. Самсонова // Практическая пульмонология. - 2011. - № 1. - С.34-38.

33. Чиссов, В. И. Злокачественные новообразования в России в 2011 году (заболеваемость и смертность) / Под ред. В. И. Чиссова, В. В. Старинского, Г. В. Петровой. - М., 2013. - 125 с.
34. Чучалин А. Г. Применение метода индуцированной мокроты для оценки интенсивности воспаления дыхательных путей / А. Г. Чучалин, С. Н. Авдеев, Э. Х. Анаев // Пульмонология. - 1998. - № 3. - С.81-87.
35. Шейх, Ж. В. Компьютерная томография в диагностике абсцедирующей пневмонии / Ж. В. Шейх, А. П. Дунаев, Н. С. Дребушевский // Вестн.РАР. - 2011. - Т. 1. - С. 38.
36. Яковлев, В. Н. Диагностика полостных образований в легких / В. Н. Яковлев, Ж. В. Шейх, А. В. Араблинский, В. Г. Алексеев, А. И. Синопальников, А. П. Дунаев, Н. С. Дребушевский // Клиническая медицина. - 2012. - № 7. - С. 59-62.
37. Aberle, D. R. Lung cancer screening with CT / D. R. Aberle, K. Brown // Clin Chest Med. - 2008. - V. 29. - P. 1-14.
38. Abramson, M. J. Distinguishing adult-onset asthma from COPD: a review and a new approach / M. J. Abramson, V. Perret, S. C. Dharmage, V. M. McDonald, C. F. McDonald // Int J Chron Obstruct Pulmon Dis. - 2014. - Vol. 9. - P. 945-962.
39. Akerström, B. Alpha-1-microglobulin: a yellow-brown lipocalin / B. Akerström, L. Lögdberg, T. Berggard, P. Osmark, A. Lindqvist // Biochim Biophys Acta. - 2000. - Vol. 1482. - P. 172-184.
40. Aldakheel, F. M. Relationships between adult asthma and oxidative stress markers and pH in exhaled breath condensate: a systematic review / F. M. Aldakheel, P. S. Thomas, J. E. Bourke, M. C. Matheson, S. C. Dharmage, A. J. Lowe // Allergy. - 2016. - Vol. 71, № 6. - P. 741-757.
41. Amann, A. Lung cancer biomarkers in exhaled breath / A. Amann, M. Corradi, P. Mazzone, A. Mutti // Expert Rev Mol Diagn. - 2011. - Vol. 11, № 2. - P. 207-217.
42. Audain, E. In-depth analysis of protein inference algorithms using multiple search engines and well-defined metrics / E. Audain, J. Uszkoreit, T. Sachsenberg, J. Pfeuffer, X. Liang, H. Hermjakob, A. Sanchez, M. Eisenacher, K. Reinert, D. L. Tabb, O. Kohlbacher, Y. Perez-Riverol // J Proteomics. - 2017. - Vol. 150. - P. 170-182.
43. Balgley, B. M. Comparative Evaluation of Tandem MS Search Algorithms Using a Target-Decoy Search Strategy / B. M. Balgley, T. Laudeman, L. Yang L., T. Song, C. S. Lee // Mol Cell Prot. - 2007. - Vol. 6. - P. 1599-1608.
44. Balint, B. Ionic composition of exhaled breath condensate in cystic fibrosis / B. Balint, S. A. Kharitonov, M. Kemp, L. E. Donnelly, P. L. Shah, M. E. Hodson, P.J. Barnes // Eur Respir J. - 2001. - Vol.18, Suppl. 33. - P. A158.

45. Bera, T.K. POTE paralogs are induced and differentially expressed in many cancers / T. K. Bera, A. Saint Fleur, Y. Lee, A. Kydd, Y. Hahn, N. C. Popescu, D. B. Zimonjic, B. Lee, I. Pastan // *Cancer Res.* - 2006. - Vol. 66, № 1. - P. 52–56.
46. Bernard, G. R. The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination / G. R. Bernard, A. Artigas, K. L. Brigham, J. Carlet, K. Falke, L. Hudson, M. Lamy, J. R. Legall, A. Morris, R. Spragg // *Am J Respir Crit Care Med.* - 1994. - Vol. 149. - P. 818-824.
47. Bisenkov, L. N. Diagnosis and treatment of lung cancer associated with paracancerous changes in the thoracic cavity / L. N. Bisenkov, A. S. Barchuk, S. A. Shalaev, I. M. Kuznetsov, G. S. Zubov, O. B. Il'ina // *Vopr Onkol.* - 2006. - Vol. 52, № 4. - P. 408-413.
48. Bloemen, K. A new approach to study exhaled proteins as potential biomarkers for asthma / K. Bloemen, R. van den Heuvel, E. Govarts, J. Hooyberghs, V. Nelen, E. Witters, K. Desager, G. Schoeters // *Clin Exp Allergy.* - 2011. - Vol. 41, № 3. - P. 346–356.
49. Bollag, D. *Protein Methods* / D. Bollag, S. Edelstein. - NY.: Wiley-Liss. Inc., 1991. - P. 72-73.
50. Bowen, S. J. British thoracic society Paediatric pneumonia audit: a review of 3 years of data / S. J. Bowen, A. H. Thomson // *Thorax.* - 2013. - Vol. 68. - P. 682–683.
51. Brannian, J. Localization of Lipid Peroxidation-Derived Protein Epitopes in the Porcine Corpus Luteum / J. Brannian, Y. Zhao, J. Burbach // *Biology of reproduction.* - 1997. - Vol. 57. - P. 1461-1466.
52. Breiman, L. *Random Forests* / L. Breiman // *Machine Learning.* - 2001. - Vol. 45, № 1. - P. 5–32.
53. Brightling, C. E. Clinical applications of induced sputum / C. E. Brightling // *Chest.* 2006. – Vol. 129. – P. 1344–1348.
54. Brochard, L. Clinical review: Respiratory monitoring in the ICU - a consensus of 16 / L. Brochard, G. S. Martin, L. Blanch, P. Pelosi, F. J. Belda, A. Jubran, L. Gattinoni, J. Mancebo, V. M. Ranieri, J.-C. M. Richard, D. Gommers, A. Vieillard-Baron, A. Pesenti, S. Jaber, O. Stenqvist, J.-L. Vincent // *Critical Care.* - 2012. - Vol. 16. - P. 219.
55. Brussino, L. Inflammatory cytokines and VEGF measured in exhaled breath condensate are correlated with tumor mass in non-small cell lung cancer / L. Brussino, B. Culla, C. Bucca, R. Giobbe, M. Boita, G. Isaia, E. Heffler, A. Oliaro, P. Filosso, G. Rolla // *J Breath Res.* - 2014. - Vol. 8, № 2. - P. 027110.

56. Buszewski, B. Human exhaled air analytics: biomarkers of diseases / B. Buszewski, M. Keszy, T. Ligor, A. Amann // *Biomed Chromatogr.* - 2007. - Vol. 21. - P. 553–566.
57. Cabello, B. Work of breathing / B. Cabello, J. Mancebo // *Intensive Care Med.* - 2006. - Vol. 32. - P. 1311-1314.
58. Carpagnano, G. E. Interleukin-6 is increased in breath condensate of patients with non-small cell lung cancer / G. E. Carpagnano, O. Resta, M. P. Foschino-Barbaro, E. Gramiccioni, F. Carpagnano // *Int J Biol Mark.* - 2002. - Vol. 17, № 2. - P. 141–145.
59. Carpagnano, G. E. Endothelin-1 is increased in the breath condensate of patients with non-small-cell lung cancer / G. E. Carpagnano, M. P. Foschino-Barbaro, O. Restaa, E. Gramiccioni, F. Carpagnano // *Oncology.* - 2004. - Vol. 66, №3. - P. 180–184.
60. Carpagnano, G. E. IL-2, TNF- α , and leptin: local versus systemic concentrations in NSCLC patients / G. E. Carpagnano, A. Spanevello, C. Curci, F. Salerno, G. P. Palladino, O. Resta, G. Di Gioia, F. Carpagnano, M. P. Foschino Barbaro // *Oncology Res.* - 2007. - Vol. 16, № 8. - P. 375–381.
61. Carpagnano, G. E. Neutrophilic airways inflammation in lung cancer: the role of exhaled LTB-4 and IL-8 / G. E. Carpagnano, G. P. Palladino, D. Lacedonia, A. Koutelou, S. Orlando, M. P. Foschino-Barbaro // *BMC Cancer.* - 2011. - Vol. 11. - article 226.
62. Carter, S. R. Exhaled Breath Condensate Collection in the Mechanically Ventilated Patient / S. R. Carter, C. S. Davis, E. J. Kovacs // *Respir Med.* - 2012. - Vol. 106, № 5. - P. 601-613.
63. Chao, J. Alveolar macrophages initiate the systemic microvascular inflammatory response to alveolar hypoxia / J. Chao, J. G. Wood, N. C. Gonzalez // *Respir Physiol Neurobiol.* - 2011. - Vol. 178, № 3. - P. 439–448.
64. Cheng, Z. Comparative proteomics analysis of exhaled breath condensate in lung cancer patients / Z. Cheng, A. K. Chan, C. R. Lewis, P. S. Thomas, M. J. Raftery // *J Cancer Therapy.* - 2011. - Vol. 2, № 1. - P. 1–8.
65. Coleman, M. P. Cancer Survival Trends in England and Wales 1971–1995: Deprivation and NHS Region. Studies on Medical and Population Subjects No. 61 / M. P. Coleman, P. Babb, P. Damiecki, P. C. Grosclaude, S. Honjo, J. Jones, G. Knerer, A. Pitard, M. J. Quinn, A. Sloggett, B. L. De Stavola. Great Britain. - Office for National Statistics The Stationery Office: London, 1999. - 695 p.
66. Conrad, D. H. Proteomics as a Method for Early Detection of Cancer: A Review of Proteomics, Exhaled Breath Condensate, and Lung Cancer Screening / D. H. Conrad, J. Goyette, P. S. Thomas // *J Gen Int Med.* - 2008. - Vol. 23, Suppl. 1. - P. 78-84.

67. Corradi, M. Collecting Exhaled Breath Condensate (EBC) with Two Condensers in Series: A Promising Technique for Studying the Mechanisms of EBC Formation, and the Volatility of Selected Biomarkers / M. Corradi, M. Goldoni, A. Caglieri, G. Folesani, D. Poli, M. Corti, A. Mutti // *J Aerosol Med Pulm Drug Deliv.* - 2008. - Vol. 21, № 1. - P. 35-44.
68. Cox, A. G. Mitochondrial peroxiredoxin involvement in antioxidant defence and redox signalling / A. G. Cox, C. C. Winterbourn, M. B. Hampton // *Biochem J.* - 2009. - Vol. 425, № 2. - P. 313-325.
69. Cunningham, T. J. Identification of a survival-promoting peptide in medium conditioned by oxidatively stressed cell lines of nervous system origin / T. J. Cunningham, L. Hodge, D. Speicher, D. Reim, C. Tyler-Polsz, P. Levitt, K. Eagleson, S. Kennedy, Y. Wang // *J Neurosci.* - 1998. - Vol. 18. - P. 7047-7060.
70. Cuzick J. Screening for cancer: future potential / J. Cuzick // *Eur J Cancer.* - 1999. - Vol. 35, № 5. - P. 685-692.
71. Czebe, K. Influence of condensing equipment and temperature on exhaled breath condensate pH, total protein and leukotriene concentrations / K. Czebe, I. Barta, B. Antus, M. Valyon, I. Horvath, T. Kullmann // *Respir Med.* - 2008. - Vol. 102. - P. 720-725.
72. Czitrovsky, A. A new method for the simultaneous measurement of particle size, complex refractive index and particle density / A. Czitrovsky, W. Szymanski, A. Nagy, P. Jani // *Meas Sci Technol.* - 2002. - Vol. 13. - P. 303-308.
73. Da Silva, M.R. Splicing Regulators and Their Roles in Cancer Biology and Therapy / M. R. da Silva, G. A. Moreira, R. A. Gonçalves da Silva, É. de Almeida Alves Barbosa, R. Pais Siqueira, R. R. Teixeira, M. R. Almeida, A. Silva Júnior, J. L. Fietto, G. C. Bressan // *Biomed Res Int.* - 2015. - Vol. 2015. - P. 150514.
74. Dalaveris, E. VEGF, TNF- α and 8-isoprostane levels in exhaled breath condensate and serum of patients with lung cancer / E. Dalaveris, T. Kerenidi, A. Katsabeki-Katsafli, T. Kiropoulos, K. Tanou, K. I. Gourgoulisanis, K. Kostikas // *Lung Cancer.* - 2009. - Vol. 64, № 2. - P. 219-225.
75. Dekhuijzen, P. N. R. Increased exhalation of hydrogen peroxide in patients with stable and unstable chronic obstructive pulmonary disease / P. N. R. Dekhuijzen, K. K. H. Aben, I. Dekker, L. P. Aarts, P. L. Wielders, C. L. van Herwaarden, A. Bast // *Am J Respir Crit Care Med.* - 1996. - Vol. 154, № 3. - P. 813-816.
76. Derbyshire, Z. E. Angiotensin II stimulated transcription of cyclooxygenase II is regulated by a novel kinase cascade involving Pyk2, MEKK4 and annexin II / Z. E. Derbyshire, U. M. Halfter, R. L. Heimark, T. H. Sy, R. R. Vaillancourt // *Mol Cell Biochem.* - 2005. - Vol. 271, № 1-2. - P. 77-90.

77. Detterbeck, F.C. The new lung cancer staging system / F.C. Detterbeck, D.J. Boffa, L.T. Tanoue // *Chest*. - 2009. - Vol. 136, № 1. - P. 260–271.
78. Digumarti, R. A Randomized, Double-Blind, Placebo-Controlled, Phase II Study of Oral Talactoferrin in Combination with Carboplatin and Paclitaxel in Previously Untreated Locally Advanced or Metastatic Non-small Cell Lung Cancer / R. Digumarti, Y. Wang, G. Raman, D.C. Doval, S.H. Advani, P.K. Julka, P.M. Parikh, S. Patil, S. Nag, J. Madhavan, A. Bapna, A.A. Ranade, A. Varadhachary, R. Malik // *J Thorac Oncol*. - 2011. - Vol. 6, № 6. - P. 1098-1103.
79. Dodig, S. Exhaled breath condensate – from an analytical point of view / S. Dodig, I. Cepelak // *Biochem Med (Zagreb)*. - 2013. - Vol. 23, № 3. - P. 281-295.
80. Dwyer, T. M. Sampling airway surface liquid: non-volatiles in the exhaled breath condensate / T. M. Dwyer // *Lung*. - 2004. - Vol. 182, № 4. - P. 241–250.
81. Eltom, S. Role of the Inflammasome-Caspase1/11-IL-1/18 Axis in Cigarette Smoke Driven Airway Inflammation: An Insight into the Pathogenesis of COPD / S. Eltom, M.G. Belvisi, C.S. Stevenson, S.A. Maher, E. Dubuis, K.A. Fitzgerald, M.A. Birrell // *PLoS One*. - 2014. - Vol. 11, №9. - P. e112829.
82. Eng, J. K. A Face in the Crowd: Recognizing Peptides Through Database Search / J. K. Eng, B. C. Searle, K. R. Clauser, D. L. Tabb // *Mol Cell Proteomics*. - 2011. - Vol. 10, № 11. - P. R111.009522.
83. Ezponda, T. The Oncoprotein SF2/ASF Promotes Non-Small Cell Lung Cancer Survival by Enhancing Survivin Expression / T. Ezponda, M. J. Pajares, J. Agorreta, J. I. Echeveste, J. M. López-Picazo, W. Torre, R. Pio, L. M. Montuenga // *Clin. Cancer Res*. - 2010. - Vol. 16, № 16. - P. 4113–4125.
84. Feiner, J. R. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender / J. R. Feiner, J. W. Severinghaus, P. E. Bickler // *Anesth Analg*. - 2007. - Vol. 105. - P. S18-23.
85. Ferlay, J. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008 / J. Ferlay, H.R. Shin, F. Bray, D. Forman, C. Mathers, D.M. Parkin // *Int J Cancer*. - 2010. - № 127. - P. 2893–2917.
86. Fontana, R. S. Screening for lung cancer. A critique of the Mayo Lung Project / R. S. Fontana, D. R. Sanderson, L. B. Woolner, W. F. Taylor, W. E. Miller, J. R. Muhm, P. E. Bernatz, W. S. Payne, P. C. Pairolero, E. J. Bergstralh // *Cancer*. - 1991. - Vol. 67, 4 Suppl. - P. 1155–1164.
87. Fumagalli, M. Profiling the Proteome of Exhaled Breath Condensate in Healthy Smokers and COPD Patients by LC-MS/MS / M. Fumagalli, F. Ferrari, M. Luisetti, J. Stolk, P. S. Hiemstra, D.

- Capuano, S. Viglio, L. Fregonese, I. Cerveri, F. Corana, C. Tinelli, P. Iadarola // *Int J Mol Sci.* - 2012. - Vol. 13, № 11. - P. 13894-13910.
88. Garey, K. W. Markers of inflammation in exhaled breath condensate of young healthy smokers / K. W. Garey, M. M. Neuhauser, R. A. Robbins, L. H. Danziger, I. Rubinstein // *Chest.* - 2004. - Vol. 125, № 1. - P. 22–26.
89. Gashenko, E. A. Evaluation of serum procathepsin B, cystatin B and cystatin C as possible biomarkers of ovarian cancer / E. A. Gashenko, V. A. Lebedeva, I. V. Brak, E. A. Tsykalenko, G. V. Vinokurova, T. A. Korolenko // *Int J Circumpolar Health.* - 2013. - Vol. 72. - P. 10.
90. Gazdar, A.F. Molecular detection of early lung cancer / A.F. Gazdar, J.D. Minna // *J Natl Cancer Inst.* - 1999. - Vol. 91, № 4. - P. 299–301.
91. Genetos, D. C. Hypoxia increases Annexin A2 expression in osteoblastic cells via VEGF and ERK / D. C. Genetos, A. Wong, S. Watari, C. E. Yellowley // *Bone.* - 2010. - Vol. 47. - P. 1013–1019.
92. Gessner, C. Angiogenic markers in breath condensate identify non-small cell lung cancer / C. Gessner, B. Rechner, S. Hammerschmidt, H. Kuhn, G. Hoheisel, U. Sack, P. Ruschpler, H. Wirtz // *Lung Cancer.* - 2010. - Vol. 68, № 2. - P. 177–184.
93. Ghosh, R. The role of dermcidin isoform 2: a two-faceted atherosclerotic risk factor for coronary artery disease and the effect of acetyl salicylic acid on it / R. Ghosh, U. K. Maji, R. Bhattacharya, A. K. Sinha // *Thrombosis.* - 2012. - Vol. 2012. - P. 987932.
94. Ghosh, R. The control of hyperglycemia in alloxan treated diabetic mice through the stimulation of hepatic insulin synthesis due to the production of nitric oxide / R. Ghosh, P. Jana, A. K. Sinha // *Exp Clin Endocrinol Diabetes.* - 2012. - Vol. 120. - P. 145-151.
95. Grant, A. J. Elastase activities of human bladder cancer cell lines derived from high grade invasive tumours / A. J. Grant, P. J. Russell, D. Raghavan // *Biochem Biophys Res Commun.* - 1989. - Vol. 162, № 1. - P. 308-315.
96. Grbovic, O. M. V600E B-Raf requires the Hsp90 chaperone for stability and is degraded in response to Hsp90 inhibitors / O. M. Grbovic, A. D. Basso, A. Ye. Q. Sawai, P. Friedlander, D. Solit, N. Rosen // *Proc Natl Acad Sci U S A.* - 2006. - Vol. 103. - P. 57–62.
97. Groom, J. R. BAFF and MyD88 signals promote a lupus like disease independent of T cells / J. R. Groom, C. A. Fletcher, S. N. Walters, S. T. Grey, S. V. Watt, M. J. Sweet, M. J. Smyth, C. R. Mackay, F. Mackay // *J Exp Med.* - 2007. - Vol. 204. - P. 1959–1971.
98. Guergova-Kuras, M. Discovery of Lung Cancer Biomarkers by Profiling the Plasma Proteome with Monoclonal Antibody Libraries / M. Guergova-Kuras, I. Kurucz, W. Hempel, N. Tardieu, J.

- Kádas, C. Malderez-Bloes, A. Jullien, Y. Kieffer, M. Hincapie, A. Guttman, E. Csánky, B. Dezsó, B.L. Karger, L. Takács // *Mol Cell Proteomics*. - 2011. - Vol. 10, № 12. - P. M111.010298.
99. Hasegawa, M. Altered expression of dermokine in skin disorders / M. Hasegawa, K. Higashi, C. Yokoyama, F. Yamamoto, T. Tachibana, T. Matsushita, Y. Hamaguchi, K. Saito, M. Fujimoto, K. Takehara // *J Eur Acad Dermatol Venereol*. - 2013. - Vol. 27, № 7. - P. 867-875.
100. He, Y. dbDEPC 2.0: updated database of differentially expressed proteins in human cancers / Y. He, M. Zhang, Y. Ju, Z. Yu, D. Lv, H. Sun, W. Yuan, F. He, J. Zhang, H. Li, J. Li, R. Wang-Sattler, Y. Li, G. Zhang, L. Xie // *Nucleic Acids Res*. - 2012. - Database issue. - P. D964-971.
101. Henschke, C. I. Early Lung Cancer Action Project: overall design and findings from baseline screening / C. I. Henschke, D. I. McCauley, D. F. Yankelevitz, D. P. Naidich, G. McGuinness, O. S. Miettinen // *Lancet*. - 1999. - Vol. 354. - P. 99–105.
102. Henschke, C. I. Early Lung Cancer Action Project: a summary of the findings on baseline screening / C. I. Henschke, D. I. McCauley, D. F. Yankelevitz // *Oncologist*. - 2001. - Vol. 6, № 2. - P. 147-152.
103. Hinkelbein J. Effect of nail polish on oxygen saturation determined by pulse oximetry in critically ill patients / J. Hinkelbein, H. V. Genzwuerker, R. Sogl, F. Fiedler // *Resuscitation*. - 2007. - Vol. 72. - P. 82-91.
104. Hirai, H., Tanaka, K., Yoshie, O., Ogawa, K. et al, Prostaglandin D2 selectively induces chemotaxis in T helper type 2 cells, eosinophils, and basophils via seven-transmembrane receptor CRTH2. *J. Exp. Med.* 2001, 193, 255–261.
105. Hoffman, R. M. Lung Cancer Screening / R. M. Hoffman, R. Sanchez // *Med Clin NA*. - 2017. - Vol. 101, № 4. - P. 769-785.
106. Hoffmann, H. J. Human skin keratins are the major proteins in exhaled breath condensate / H. J. Hoffmann, L. M. Tabaksblat, J. J. Enghild, R. Dahl // *Eur Respir J*. - 2008. - Vol. 31. - P. 380-384.
107. Horvath, I. Exhaled breath condensate: methodological recommendations and unresolved questions / I. Horvath, J. Hunt, P. J. Barnes, on behalf of the ERS/ATS TaskForce // *Eur Respir J*. - 2005. - Vol. 26. - P. 523–548.
108. Horvath, I. Exhaled biomarkers in lung cancer / I. Horvath, Z. Lazar, N. Gyulai, M. Kollai, G. Losonczy // *Eur Respir J*. - 2009. - Vol. 34. - P. 261–275.
109. Hu, S. Human saliva proteome analysis and disease biomarker discovery / S. Hu, J. A. Loo, D. T. Wong // *Exp Rev Proteomics*. - 2007. - Vol. 4, № 4. - P. 531-538.
110. Huang, Y.-H. Identification and enhancement of HLA-A2.1-restricted CTL epitopes in a new human cancer antigen-POTE / Y. H. Huang, M. Terabe, C. D. Pendleton, D. Stewart Khursigara, T. K.

- Bera, I. Pastan, J. A. Berzofsky // PLoS One Public Library of Science. - 2013. -Vol. 8, № 6. - P. e64365.
111. Hubbard, S. J. Proteome Bioinformatics / S. J. Hubbard, A. R. Jones (eds). - UK: Humana Press. - 2010. - 403 p.
112. Hunter, J. D. Matplotlib: A 2D Graphics Environment / J. D. Hunter // Computing in Science & Engineering. - 2007. - Vol. 9. - P. 90-95.
113. Hüttmann, E. M. Comparison of two devices and two breathing patterns for exhaled breath condensate sampling / E. M. Hüttmann, T. Greulich, A. Hattesoehl, S. Schmid, S. Noeske, C. Herr, G. John, R. A. Jörres, B. Müller, Vogelmeier C., Koczulla A. R. // PLoS One. -2011. - Vol. 6, № 11. - P. e27467.
114. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Personal habits and indoor combustions. Volume 100 E. A review of human carcinogens // IARC Monogr Eval Carcinog Risks Hum. 2012. - V. 100(Pt E). - P. 1-538.
115. Ichikawa, T. Possible impact of salivary influence on cytokine analysis in exhaled breath condensate / T. Ichikawa, K. Matsunaga, Y. Minakata, S. Yanagisawa, et al // Anal Chem Insights. - 2007. - Vol. 2. - P. 85–92.
116. Ishihama, Y. Microcolumns with self-assembled particle frits for proteomics / Y. Ishihama, J. Rappsilber, J. S. Andersen, M. Mann // J Chrom A. - 2002. - Vol. 979. - P. 233-239.
117. Jemal, A. Cancer Statistics / A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, M. J. Thun // CA Cancer J Clin. - 2008. - Vol. 58, №2. - P. 71–96.
118. Jensen, O. N. Interpreting the protein language using proteomics / O. N. Jensen // Nat Rev Mol Cell Biol. - 2006. - Vol. 7. - P. 391—403.
119. Jin, Y. Identification of GlcNAcylated alpha-1-antichymotrypsin as an early biomarker in human non-small-cell lung cancer by quantitative proteomic analysis with two lectins / Y. Jin, J. Wang, X. Ye, Y. Su, G. Yu, Q. Yang, W. Liu, W. Yu, J. Cai, X. Chen, Y. Liang, Y. Chen, B.H. Wong, X. Fu, H. Sun // Br J Cancer. - 2016. - Vol. 114, № 5. - P. 532-544.
120. Kalsheker N.A. Alpha 1-antichymotrypsin / N.A. Kalsheker // J Clin Pathol. - 2009. - Vol. 62, № 10. - P. 865-869.
121. Kielbasa, B. Eicosanoids in exhaled breath condensates in the assessment of childhood asthma / B. Kielbasa, A. Moeller, M. Sanak, J. Hamacher, M. Hutterli, A. Cmiel, A. Szczeklik, J. H. Wildhaber // Pediatric Allergy Immunol. - 2008. - Vol. 19, № 7. - P. 660–669.

122. Kimura, J. Identification of dihydropyrimidinase-related protein 4 as a novel target of the p53 tumor suppressor in the apoptotic response to DNA damage / J. Kimura, T. Kudoh, Y. Miki, K. Yoshida // *Int J Cancer*. - 2011. - Vol. 128, № 7. - P. 1524-1531.
123. Kline, J. A. Diagnostic accuracy of a bedside D-dimer assay and alveolar dead-space measurement for rapid exclusion of pulmonary embolism: a multicenter study / J. A. Kline, E. G. Israel, E. A. Michelson, B. J. O'Neil, M. C. Plewa, D. C. Portelli // *JAMA*. - 2001. - Vol. 285. - P. 761-768.
124. Konstantinidi, E. M. Exhaled Breath Condensate: Technical and Diagnostic Aspects / E. M. Konstantinidi, A. S. Lappas, A. S. Tzortzi, P. K. Behrakis // *SWJ*. - 2015. - Vol. 2015. - P. 435160.
125. Kopitar-Jerala, N. The role of cystatins in cells of the immune system / N. Kopitar-Jerala // *FEBS Lett*. - 2006. - Vol. 580. - P. 6295-6301.
126. Koshikawa, N. Purification and Identification of a Novel and Four Known Serine Proteinase Inhibitors Secreted by Human Glioblastoma Cells / N. Koshikawa, T. Nakamura, N. Tsuchiya, M. Isaji, H. Yasumitsu, M. Umeda, K. Miyazaki // *J Biochem*. - 1996. - Vol. 119, № 2. - P. 334-339.
127. Kostikas, K. Oxidative stress in expired breath condensate of patients with COPD / K. Kostikas, G. Papatheodorou, K. Psathakis, P. Panagou, S. Loukides // *Chest*. - 2003. - Vol. 124, № 4. - P. 1373–1380.
128. Kostikas, K. Leukotriene B4 in exhaled breath condensate and sputum supernatant in patients with COPD and asthma / K. Kostikas, M. Gaga, G. Papatheodorou, T. Karamanis, D. Orphanidou, S. Loukides // *Chest*. - 2005. - Vol. 127, № 5. - P. 1553–1559.
129. Kroegel, C. Transthorakale Sonographie: Grundlagen und Anwendung; ein Leitfaden für die Praxis. Stuttgart. / C. Kroegel, A. Reißig. – NY.: Thieme Verlag, 2000. – 116 p.
130. Kullmann, T. Differential cytokine pattern in the exhaled breath of patients with lung cancer / T. Kullmann, I. Barta, E. Csiszer, B. Antus, I. Horvath // *Pathology and Oncology Res*. - 2008. - Vol. 14, № 4. - P. 481–483.
131. Kumar, V. Robbins Basic Pathology, 7th Ed. / V. Kumar, R.S. Cotran, S.L. Robbins. - Philadelphia: Saunders, 2003. - 873 p.
132. Kumar, R. Amino acid frequency distribution among eukaryotic proteins / Kumar R., Gaur G. // *OABJ*. - 2014. - Vol. 5, № 2. - P. 6–11.
133. Kunz, L. I. Z. Smoking status and anti-inflammatory macrophages in bronchoalveolar lavage and induced sputum in COPD / L. I. Z. Kunz, T. S. Lapperre, J. B. Snoeck-Stroband, S. E. Budulac // *Respir Res*. - 2011. - Vol. 12, № 1. – P. 34.

134. Kurova V. S. Proteomics of exhaled breath: methodological nuances and pitfalls / V. S. Kurova, E. C. Anaev, A. S. Kononikhin, K. Y. Fedorchenko, I. A. Popov, T. L. Kalupov, D. O. Bratanov, E. N. Nikolaev, S. D. Varfolomeev // *Clin Chem Lab Med.* - 2009. - Vol. 47, № 6. - P. 706–712.
135. Kurova, V. S. Mass spectrometric monitoring of exhaled breath condensate proteome of a patient after lung transplantation / V. S. Kurova, E. C. Anaev, A. S. Kononikhin, I. A. Popov, K. Yu Fedorchenko, E. N. Nikolaev, S. D. Varfolomeev, A. G. Chuchalin // *Rus Chem Bull.* — 2010. — Vol. 59, no. 1. — P. 292–296.
136. Kurova, V. S. Exogenic proteins in the human exhaled breath condensate / V. S. Kurova, A. S. Kononikhin, D. A. Sakharov, I. A. Popov, I. M. Larina, A. G. Tonevitskiĭ, S. D. Varfolomeev, E. N. Nikolaev // *Bioorg Khim.* - 2011. - Vol. 37, № 1. - P. 55-60.
137. Kurova, V. S. Structural and catalytic polymorphism of human enzymes: Novel potential platforms for biomedical diagnostics / V. S. Kurova, I. N. Kurochkin, G. R. Kalamkarov, A. E. Bugrova, K. Yu Fedortchenko, S. D. Varfolomeev // *Biotechnol Adv.* — 2009. — Vol. 27, no. 6. — P. 945–959.
138. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4 / U. K. Laemmli // *Nature.* - 1970. — Vol. 227, №5259. — P. 680–685.
139. Lases, E. C. Oxidative stress after lung resection therapy / E. C. Lases, V. A. M. Duurkens, W. B. M. Gerritsen, F. J. L. M. Hass // *Chest.* - 2000. - Vol. 117. - P. 99–103.
140. Leclerc, E. A. The ubiquitous dermokine delta activates Rab5 function in the early endocytic pathway / E. A. Leclerc, L. Gazeilles, G. Serre, M. Guerrin, N. Jonca // *PLoS One.* - 2011. - Vol. 6, № 3. - P. e17816.
141. Lehmann, C. Breathing pattern and exhaled breath condensate sampling / C. Lehmann, M. Rothe, G. Becher // *Am J Respir Crit Care Med.* - 2003. - Vol. 167, № 9. - P. 50.
142. Liang, Y. Exhaled Breath Condensate: A Promising Source for Biomarkers of Lung Disease / Y. Liang, S. M. Yeligar, L. A. S. Brown // *ScientificWorldJournal.* - 2012. - Vol. 2012. - P. 217518.
143. Lin, M. S. Assessment of MS/MS Search Algorithms with Parent-Protein Profiling / M. S. Lin, J. J. Cherny, C. T. Fournier, S. J. Roth, D. Krizanc, M. P. Weir // *J Proteome Res.* - 2014. - Vol. 13 № 4. - P. 1823-1832.
144. Lu, T. P. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women / T. P. Lu, M. H. Tsai, J. M. Lee, C. P. Hsu, P. C. Chen, C. W. Lin, J. Y. Shih, P. C. Yang, C. K. Hsiao, L. C. Lai, E. Y. Chuang // *Cancer Epid Biomark Prev.* - 2010. - Vol. 19, № 10. - P. 2590-2597.

145. Maeda, K. Prognostic value of vascular endothelial growth factor expression in gastric carcinoma / K. Maeda, Y. Chung, Y. Ogawa, S. Takatsuka, S. M. Kang, M. Ogawa, T. Sawada, M. Sow // *Cancer*. - 1996. - Vol. 77. - P. 858–863.
146. Mancebo, J. Effects of albuterol inhalation on the work of breathing during weaning from mechanical ventilation / J. Mancebo, P. Amaro, H. Lorino, F. Lemaire, A. Harf, L. Brochard // *Am Rev Respir Dis*. - 1991. - Vol. 144. - P. 95-100.
147. Manser, R. Screening for lung cancer. Cochrane Data-base / R. Manser, A. Lethaby, L. B. Irving, C. Stone, G. Byrnes, M. J. Abramson, D. Campbell // *Syst Rev*. - 2013. - № 6. - P. CD001991.
148. Marimuthu, A. A Comprehensive Map of the Human Urinary Proteome / A. Marimuthu, R. N. O'Meally, R. Chaerkady, Y. Subbannayya, V. Nanjappa, P. Kumar, D. S. Kelkar, S. M. Pinto, R. Sharma, S. Renuse, R. Goel, R. Christopher, B. Delanghe, R. N. Cole, H. C. Harsha, A. Pandey // *J Proteome Res*. - 2011. - Vol. 10, № 6. - P. 2734-2743.
149. Matsuoka, T. Prostaglandin D2 as a Mediator of Allergic Asthma / T. Matsuoka, M. Hirata, H. Tanaka, Y. Takahashi, T. Murata, K. Kabashima, Y. Sugimoto, T. Kobayashi, F. Ushikubi, Y. Aze, N. Eguchi, Y. Urade, N. Yoshida, K. Kimura, A. Mizoguchi, Y. Honda, H. Nagai, S. Narumiya // *Science*. - 2000. - Vol. 287. - P. 2013–2017.
150. McCafferty, J. B. Effects of breathing pattern and inspired air conditions on breath condensate volume, pH, nitrite, and protein concentrations / J. B. McCafferty, T. A. Bradshaw, S. Tate, A. P. Greening, J. A. Innes // *Thorax*. - 2004. - Vol. 59, № 8. - P. 694–698.
151. McKinney, W. Data Structures for Statistical Computing in Python / W. McKinney // *Proceedings of the 9th Python in Science Conference*. - 2010. - P. 51-56.
152. Miguel, F.J. Identification of alternative splicing events regulated by the oncogenic factor SRSF1 in lung cancer / F. J. Miguel, R. D. Sharma, M. J. Pajares, L. M. Montuenga, A. Rubio, R. Pio // *Cancer Res*. - 2014. - Vol. 74, № 4. - P. 1105–1115.
153. Mikhaylova, M. Analysis of Y-P30/Dermcidin expression and properties of the Y-P30 peptide / M. Mikhaylova, A. Schumacher, C. Borutzki, J. R. Neumann, T. Macharadze, T. El-Mousleh, P. Wahle, A. C. Zenclussen, M. R. Kreutz // *BMC Res Notes*. - 2014. - Vol. 7. - P. 400.
154. Mizon, C. Urinary bikunin determination provides insight into proteinase/proteinase inhibitor imbalance in patients with inflammatory diseases / C. Mizon, F. Piva, V. Queyrel, M. Balduyck, E. Hachulla, J. Mizon // *Clin. Chem. Lab. Med*. - 2002. - Vol. 40. - P. 579-586.
155. Moll, R. The human keratins: biology and pathology / R. Moll, M. Divo, L. Langbein // *Histochem Cell Biol*. - 2008. - Vol. 129. - P. 705–733.

156. Moller, J. T. Randomized evaluation of pulse oximetry in 20,802 patients: II. Perioperative events and postoperative complications / J. T. Moller, N. W. Johannessen, K. Espersen, O. Ravlo, B. D. Pedersen, P. F. Jensen, N. H. Rasmussen, L. S. Rasmussen, T. Pedersen, J. B. Cooper // *Anesthesiology*. - 1993. - Vol. 78. - P. 445-453.
157. Monnet, X. Assessing pulmonary permeability by transpulmonary thermodilution allows differentiation of hydrostatic pulmonary edema from ALI/ARDS / X. Monnet, N. Anguel, D. Osman, O. Hamzaoui, C. Richard, J. L. Teboul // *Intensive Care Med*. - 2007. - Vol. 33. - P. 448-453.
158. Montuschi, P. Increased 8-Isoprostane, a Marker of Oxidative Stress, in Exhaled Condensate of Asthma Patients / P. Montuschi, M. Corradi, G. Ciabattini, J. Nightingale, S. A. Kharitonov, P. J. Barnes // *Am J Respir Crit Care Med*. - 1999. - Vol. 160, № 1. - P. 216-220.
159. Montuschi, P. Exhaled leukotrienes and prostaglandins in asthma / P. Montuschi, P. J. Barnes // *J Allergy Clin Immunol*. - 2002. - Vol. 109, № 4. - P. 615-620.
160. Newport, S. Exhaled breath condensate pH and ammonia in cystic fibrosis and response to treatment of acute pulmonary exacerbations / S. Newport, N. Amin, A. J. Dozor // *Pediatr Pulmonol*. - 2009. - Vol. 44, № 9. - P. 866-872.
161. Nowak, D. Increased content of thiobarbituric acid-reactive substances and hydrogen peroxide in the expired breath condensate of patients with stable chronic obstructive pulmonary disease: no significant effect of cigarette smoking / D. Nowak, M. Kasielski, A. Antczak, T. Pietras, P. Bialasiewicz // *Respir Med*. - 1999. - Vol. 93, № 6. - P. 389-396.
162. Ocak, S. Discovery of New Membrane-Associated Proteins Overexpressed in Small-Cell Lung Cancer / S. Ocak, D.B. Friedman, H. Chen, J.A. Ausborn, M. Hassanein, B. Detry, B. Weynand, F. Aboubakar, C. Pilette, Y. Sibille, P.P. Massion // *J Thorac Oncol*. - 2014. - Vol. 9, № 3. - P. 324-336.
163. Ono, E. Increase in inflammatory mediator concentrations in exhaled breath condensate after allergen inhalation / E. Ono, H. Mita, M. Taniguchi, N. Higashi, T. Tsuburai, M. Hasegawa, E. Miyazaki, T. Kumamoto, K. Akiyama // *J Allergy Clin Immunol*. - 2008. - Vol. 122, № 4. - P. 768-773.
164. Otto, A. M. Warburg effect(s)—a biographical sketch of Otto Warburg and his impacts on tumor metabolism / A. M. Otto // *Cancer & Metabolism*. - 2016. - Vol. 4. - P. 5.
165. Paone, G. The combination of a smoking cessation programme with a routine rehabilitation centre increases stop-smoking rate / G. Paone, M. Serpilli, V. Conti, E. Girardi, R. S. Principe, G. Puglisi, L. DE Marchis, G. Schmid // *J Rehabil Med*. - 2008. - V. 40. - P. 672-677.
166. Paone, G. Long-term home non invasive mechanical ventilation increases systemic inflammatory response in chronic obstructive pulmonary disease: a prospective observational study /

- G. Paone, V. Conti, G. Biondi-Zoccai, E. De Falco, I. Chimenti, M. Peruzzi, C. Mollica, G. Monaco, G. Giannunzio, G. Brunetti, G. Schmid, V. M. Ranieri, G. Frati // *Mediators Inflamm.* - 2014. - V. 2014. - P. 503145.
167. Paone, G. Blood and sputum biomarkers in COPD and asthma: a review / G. Paone, V. Leone, V. Conti, L. De Marchis, E. Ialleni, C. Graziani, M. Salducci, M. Ramaccia, G. Munafò // *Eur Rev Med Pharmacol Sci.* - 2016. - V. 20, № 4. - P. 698-708.
168. Pappin, D. J. Rapid identification of proteins by peptide-mass fingerprinting / D. J. Pappin, P. N. Hojrup, A. J. Bleasby // *Curr. Biol.* - 1993. - Vol. 3. - P. 327.
169. Paredi, P. Analysis of expired air for oxidation products. / P. Paredi, S. A. Kharitonov, P. J. Barnes // *Am J Respir Crit Care Med.* - 2002. - Vol. 166, № 12. - P. S31-37.
170. Park, S. Y. Evaluation of plasma carcinogenic markers in rat hepatic tumors models induced by rat hepatoma N1-S1 cells and benzo[a]pyrene / S. Y. Park, S. Phark, M. Lee, Z. Zheng, S. Choi, N. H. Won, W. W. Jung, D. Sul // *Arch Pharm Res.* - 2010. - Vol. 33. - P. 247-255.
171. Pasare, C. Control of B-cell responses by Toll-like receptors / C. Pasare, R. Medzhitov // *Nature.* - 2005. - Vol. 438. - P. 364–368.
172. Pedregosa, F. Scikit-learn: Machine Learning in Python / F. Pedregosa, V. Varoquaux., A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay // *JMLR* 12. - 2011. - P. 2825–2830.
173. Pelosi, P. Recruitment and derecruitment during acute respiratory failure: an experimental study / P. Pelosi, M. Goldner, A. McKibben, A. Adams, G. Eccher, P. Caironi, S. Losappio, L. Gattinoni, J. J. Marini // *Am J Respir Crit Care Med.* - 2001. - Vol. 164. - P. 122-130.
174. Perkins, D. N. Probability-based protein identification by searching sequence databases using mass spectrometry data / D. N. Perkins, D. J. Pappin, D. M. Creasy, J. S. Cottrell // *Electrophoresis.* - 1999. - Vol. 20, № 18. - P. 3551-3567.
175. Perretti, M. Annexin A1 and glucocorticoids as effectors of the resolution of inflammation / M. Perretti, F. D'Acquisto // *Nat Rev Immunol.* - 2009. - Vol. 9, № 1. - P. 62-70.
176. Petty, R.D. Gene Expression Profiling in Non-small Cell Lung Cancer: From Molecular Mechanisms to Clinical Application / R.D. Petty, M.C. Nicolson, K.M. Kerr, E. Collie-Duguid, G.I. Murray // *Clin Cancer Res.* - 2004. - Vol. 10, № 10. - P. 3237-3248.
177. Phillips, M. Prediction of lung cancer using volatile biomarkers in breath / M. Phillips, N. Altorki, J. H. Austin, R. B. Cameron, R. N. Cataneo, J. Greenberg, R. Kloss, R. A. Maxfield, M. I.

- Munawar, H. I. Pass, A. Rashid, W. N. Rom, P. Schmitt // *Cancer Biomark.* - 2007. - Vol. 3. - P. 95–109.
178. Phillips, M. Method for the collection and assay of volatile organic compounds in breath / M. Phillips // *Anal Biochem.* - 1997. - Vol. 247. - P. 272–278.
179. Pichavant, M. Oxidative stress-mediated iNKT-cell activation is involved in COPD pathogenesis / M. Pichavant, G. Rémy, S. Bekaert, O. Le Rouzic, G. Kervoaze, E. Vilain, N. Just, I. Tillie-Leblond, F. Trottein, D. Cataldo, P. Gosset // *Mucosal Immunol.* - 2014. - Vol. 7, № 3. - P. 568–578.
180. Profita, M. Chronic obstructive pulmonary disease and neutrophil infiltration: role of cigarette smoke and cyclooxygenase products / M. Profita, A. Sala, A. Bonanno // *Lung Physiol.* - 2010. – Vol. 298, № 2. – P. 261–269.
181. Psaty, B. M. COX-2 inhibitors - lessons in drug safety / B.M. Psaty, C.D. Furberg // *N Engl J Med.* - 2005. - Vol. 352. - P. 1133–1135.
182. Puybasset, L. A computed tomography scan assessment of regional lung volume in acute lung injury. The CT Scan ARDS Study Group / L. Puybasset, P. Cluzel, N. Chao, A. S. Slutsky, P. Coriat, J. J. Rouby // *Am J Respir Crit Care Med.* - 1998. - Vol. 158. - P. 1644-1655.
183. Quandt, A. Using synthetic peptides to benchmark peptide identification software and search parameters for MS/MS data analysis / A. Quandt, L. Esponaa, A. Balaskoc, H. Weisser, M-Y. Brusniak, P. Kunszt, R. Aebersold, L. Malmström // *EuPA Open Proteomics.* - 2014. - Vol. 5. - P. 21–31.
184. Quesnel, C. Alveolar fibrocyte percentage is an independent predictor of poor outcome inpatients with acute lung injury / C. Quesnel, P. Piednoir, J. Gelly, L. Nardelli, V. Lecon, S. Lasocki, L. Bouadma, I. Philip, C. Elbim, F. Mentre, B. Crestani, M. Dehoux // *Crit Care Med.* - 2012. - Vol. 40. - P. 21-28.
185. Ratnawati. Exhaled breath condensate nitrite/nitrate and pH in relation to pediatric asthma control and exhaled nitric oxide / Ratnawati, J. Morton, R. L. Henry, P. S. Thomas // *Pediatr Pulmonol.* - 2006. - Vol. 41, № 10. - P. 929-936.
186. Reed, M. F. Survival after resection for lung cancer is the outcome that matters / M. F. Reed, M. Molloy, E. L. Dalton, J. A. Howington // *Am J Surg.* - 2004. - Vol. 188. - P. 598–602.
187. Reinhold, P. Breath condensate—a medium obtained by a noninvasive method for the detection of inflammation mediators of the lung / P. Reinhold, A. Langenberg, G. Becher, M. Rothe // *Berl Munch Tierarztl Wochenschr.* - 1999. - Vol. 112. - P. 254-259.

188. Reißig, A. Transthorakale Sonographie bei der Diagnostik pulmonaler Erkrankungen: ein systematischer Zugang / A. Reißig, C. Görg, G. Mathis // *Ultraschall in Med.* - 2009. - Vol. 30. - P. 438–458.
189. Reissig, A. Current role of emergency ultrasound of the chest / A. Reissig, R. Copetti, C. Kroegel // *Crit Care Med.* - 2011. - V. 39. - P. 839-845.
190. Remerand, F. Multiplane ultrasound approach to quantify pleural effusion at the bedside / F. Remerand, J. Dellamonica, Z. Mao, F. Ferrari, B. Bouhemad, Y. Jianxin, C. Arbelot, Q. Lu, C. Ichai, J. J. Rouby // *Intensive Care Med.* - 2010. - Vol. 36. - P. 656-664.
191. Robroeks, C. M. H. H. T. Exhaled nitric oxide and biomarkers in exhaled breath condensate indicate the presence, severity and control of childhood asthma / C. M. H. H. T. Robroeks, K. D. G. Van De Kant, Q. Jobsis, H. J. E. Hendriks, R. Van Gent, E. F. M. Wouters, J. G. M. C. Damoiseaux, A. Bast, W. K. W. H. Wodzig, E. Dompeling // *Clin Exp Allergy.* - 2007. - Vol. 37, № 9. - P. 1303-1311.
192. Robroeks, C. M. H. H. T. Biomarkers in exhaled breath condensate indicate presence and severity of cystic fibrosis in children / C. M. H. H. T. Robroeks, P. P. R. Rosias, D. Van Vliet, Q. Jobsis, J. B. L. Yntema, H. J. L. Brackel, J. G. M. C. Damoiseaux, G. M. Den Hartog, W. K. W. H. Wodzig, E. Dompeling // *Pediatr Allergy Immunol.* - 2008. - Vol. 19, № 7. - P. 652-659.
193. Rosias, P. P. Exhaled breath condensate in children: pearls and pitfalls / P. P. Rosias, E. Dompeling, H. J. Hendriks, J. W. Heijnsens, R. A. Donckerwolcke, Q. Jöbsis // *Pediatr Allergy Immunol.* - 2004. - Vol. 15, № 1. - P. 4-19.
194. Rosias, P. P. Breath condenser coatings affect measurement of biomarkers in exhaled breath condensate / P. P. Rosias, C. M. Robroeks, H. J. Niemarkt, A. D. Kester, J. H. Vernooij, J. Suykerbuyk, J. Teunissen, J. Heynens, H. J. Hendriks, Q. Jöbsis, E. Dompeling // *Eur. Respir. J.* - 2006. - Vol. 28. - P. 1036–1041.
195. Rubinowitz, A. N. Thoracic imaging in the ICU / A. N. Rubinowitz, M. D. Siegel, I. Tocino // *Crit Care Clin.* - 2007. - Vol. 23, № 3. - P. 539–573.
196. Sakka, S. G. Prognostic value of extravascular lung water in critically ill patients / S. G. Sakka, M. Klein, K. Reinhart, A. Meier-Hellmann // *Chest.* - 2002. - Vol. 122. - P. 2080-2086.
197. Sengupta, R. The role of thioredoxin in the regulation of cellular processes by S-nitrosylation / R. Sengupta, A. Holmgren // *Biochim Biophys Acta.* - 2012. - Vol. 1820, № 6. - P. 689-700.
198. Shau, H. Modulation of natural killer and lymphokine-activated killer cell cytotoxicity by lactoferrin / H. Shau, A. Kim, S.H. Golub // *J Leukoc Biol.* - 1992. - Vol. 51, № 4. - P. 343-349.

199. Shieh, Y. Low-Dose CT Scan for Lung Cancer Screening: Clinical and Coding Considerations / Y. Shieh, M. Bohnenkamp // *Chest*. - 2017. - V. 152, № 1. - P. 204-209.
200. Shimamura, M. Bovine lactoferrin inhibits tumor-induced angiogenesis / M. Shimamura, Y. Yamamoto, H. Ashino, T. Oikawa, T. Hazato, H. Tsuda, M. Iigo // *Int J Cancer*. - 2004. - Vol. 111, № 1. - P. 111-106.
201. Silkoff, P.E. The Aerocrine exhaled nitric oxide monitoring system NIOX is cleared by the US Food and Drug Administration for monitoring therapy in asthma / P. E. Silkoff, M. Carlson, T. Bourke, R. Katial, E. Ogren, S. J. Szeffler // *J Allergy Clin Immunol*. - 2004. - Vol. 114. - P. 1241–1256.
202. Sone, S. Long-term follow-up study of a population-based 1996–1998 mass screening programme for lung cancer using mobile low-dose spiral computed tomography / S. Sone, T. Nakayama, T. Honda, K. Tsushima, F. Li, M. Haniuda, Y. Takahashi, T. Suzuki, T. Yamanda, R. Kondo, T. Hanaoka, F. Takayama, K. Kubo, H. Fushimi // *Lung Cancer*. - 2007. - Vol. 58. - P. 329–341.
203. Soyer, O. U. Comparison of two methods for exhaled breath condensate collection / O. U. Soyer, E. A. Dizdar, O. Keskin, C. Lilly, O. Kalayci // *Allergy*. - 2006. - Vol. 61. - P. 1016–1018.
204. Spik, I. Activation of the Prostaglandin D2 Receptor DP2/CRT2 Increases Allergic Inflammation in Mouse / I. Spik, C. Brenuchon, V. Angeli, D. Staumont, S. Fleury, M. Capron, F. Trottein, D. Dombrowicz // *J. Immunol*. - 2005. - Vol. 174. - P. 3703–3708.
205. Steel, H.C. Overview of community-acquired pneumonia and the role of inflammatory mechanisms in the immunopathogenesis of severe pneumococcal disease / H.C. Steel, R. Cockeran, R. Anderson, C. Feldman // *Mediators Inflamm*. - 2013. - P. 490346.
206. Stewart, G. D. Variation in dermcidin expression in a range of primary human tumours and in hypoxic/oxidatively stressed human cell lines / G. D. Stewart, R. J. Skipworth, C. J. Pennington, A. G. Lowrie, D. A. Deans, D. R. Edwards, F. K. Habib, A. C. Riddick, K. C. Fearon, J. A. Ross // *Br J Cancer*. - 2008. - Vol. 99. - P. 126-132.
207. Stieber, P. Pro-gastrin-releasing peptide (ProGRP)--a useful marker in small cell lung carcinomas / P. Stieber, H. Dienemann, A. Schalhorn, U. M. Schmitt, J. Reinmiedl, K. Hofmann, K. Yamaguchi // *Anticancer Res*. - 1999. - Vol. 19, № 4A. - P. 2673-2678.
208. Subramaniam, S. Lung cancer biomarkers: State of the art / S. Subramaniam, R. K. Thakur, V. K. Yadav, R. Nanda, S. Chowdhury, A. Agrawal // *J Carcinog*. - 2013. - Vol. 12. - P. 3.

209. Swensen, S. J. Lung cancer screening with CT: Mayo Clinic experience / S. J. Swensen, J. R. Jett, T. E. Hartman, D. E. Midthun, J. A. Sloan, A. M. Sykes, G. L. Aughenbaugh, M. A. Clemens // *Radiology*. - 2003. - Vol. 226, № 3. - P. 756–761.
210. Tanaka, K. Cutting Edge: Differential Production of Prostaglandin D2 by Human Helper T Cell Subsets / K. Tanaka, K. Ogawa, K. Sugamura, M. Nakamura, S. Takano, K. Nagata // *Immunol*. - 2000. - Vol. 164. - P. 2277–2280.
211. Teng, Y. Hydrogen peroxide in exhaled breath condensate in patients with asthma: a promising biomarker? / Y. Teng, P. Sun, J. Zhang, R. Yu, J. Bai, X. Yao, M. Huang, I. M. Adcock, P. J. Barnes // *Chest*. - 2011. - Vol. 140, № 1. - P. 108-116.
212. Todorov, P. Characterization of a cancer cachectic factor / P. Todorov, P. Cariuk, T. McDevitt, B. Coles, K. Fearon, M. Tisdale // *Nature*. - 1996. - Vol. 379. - P. 739-742.
213. Tsuda, H. Cancer prevention by bovine lactoferrin and underlying mechanisms a review of experimental and clinical studies/ H. Tsuda, K. Sekine, K. Fujita, M. Ligo // *Biochem Cell Biol*. - 2002. - Vol. 80, № 1. - P. 131-136.
214. Tung, Y.T. Bovine lactoferrin inhibits lung cancer growth through suppression of both inflammation and expression of vascular endothelial growth factor / Y. T. Tung, H. L. Chen, C. C. Yen, P. Y. Lee, H. C. Tsai, M. F. Lin, C. M. Chen // *J Dairy Sci*. - 2013. - Vol. 96, № 4. - P. 2095-2106.
215. Urade, Y. Prostaglandin D synthase: structure and function / Y. Urade, O. Hayaishi // *Vitam. Horm.* - 2000. - Vol. 58. - P. 89–120.
216. Urich, D. Lung-specific loss of the laminin $\alpha 3$ subunit confers resistance to mechanical injury / D. Urich, J.L. Eisenberg, K. J. Hamill, D. Takawira, S. E. Chiarella, S. Soberanes, A. Gonzalez, F. Koentgen, T. Manghi, S. B. Hopkinson, A. V. Misharin, H. Perlman, G. M. Mutlu, G. R. Budinger, J. C. Jones // *J Cell Sci*. - 2011. - Vol. 124, № 17. - P. 2927–2937.
217. Van Beurden, W. J. Markers of inflammation and oxidative stress during lower respiratory tract infections in COPD patients / W. J. Van Beurden, F. W. Smeenk, G. A. Harff, P. N. Dekhuijzen // *Monaldi Arch Chest Dis*. - 2003. - Vol. 59, № 4. - P. 27380.
218. van der Walt, S. The NumPy Array: A Structure for Efficient Numerical Computation / S. van der Walt, S. C. Colbert, G. Varoquaux // *Computing in Science & Engineering*. - 2011. - Vol. 13. - P. 22-30.
219. Vang, A. G. PDE8 Regulates Rapid Teff Cell Adhesion and Proliferation Independent of ICER / A. G. Vang, S. Z. Ben-Sasson, H. Dong, B. Kream, M. P. DeNinno, M. M. Claffey, W. Housley, R. B. Clark, P. M. Epstein, S. Brocke // *PLoS One*. - 2010. - Vol. 5, № 8. - P. e12011.

220. Verschuren, F. Volumetric capnography as a bedside monitoring of thrombolysis in major pulmonary embolism / F. Verschuren, E. Heinonen, D. Clause, J. Roeseler, F. Thys, P. Meert, E. Marion, A. El Gariani, J. Col, M. Reynaert, G. Liistro // *Intensive Care Med.* - 2004. - Vol. 30. - P. 2129-2132.
221. Wang, B. M. Diagnosing pulmonary alveolar proteinosis / B. M. Wang, E. G. Stern, R. A. Schmidt, D. J. Pierson // *Chest.* - 1997. - Vol. 12. - P. 460-466.
222. Wang, Q. Serum levels of the cancer-testis antigen POTEE and its clinical significance in non-small-cell lung cancer / Q. Wang, X. Li, S. Ren, N. Cheng, M. Zhao, Y. Zhang, J. Li, W. Cai, C. Zhao, W. Cao, C. Zhou // *PLoS One.* - 2015. - Vol. 10, № 4. - P. e0122792.
223. Warde-Farley, D. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function / Warde-Farley D1, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. // *Nucleic Acids Res.* - 2010. - Vol. 38, (Web Server issue). - P. W214-220.
224. Welch, H. G. Overstating the evidence for lung cancer screening: the International Early Lung Cancer Action Program (I-ELCAP) study / H. G. Welch, S. Woloshin, L. M. Schwartz, L. Gordis, P. C. Gøtzsche, R. Harris, B. S. Kramer, D. F. Ransohoff // *Arch Intern Med.* - 2007. - Vol. 167. - P. 2289-2295.
225. Werle, B. Cystatins C, E/M and F in human pleural fluids of patients with neoplastic and inflammatory lung disorders / B. Werle, K. Sauckel, C. M. Nathanson, M. Bjarnadottir, E. Spiess, W. Ebert, M. Abrahamson // *Biol Chem.* - 2003. - Vol. 384, № 2. - P. 281-287.
226. Woodhead, M. Guidelines for the management of adult lower respiratory tract infections - Full version / M. Woodhead, F. Blasi, S. Ewig, J. Garau, G. Huchon, M. Ieven, A. Ortqvist, T. Schaberg, A. Torres, G. van der Heijden, R. Read, T. J. M. Verheij, Joint Taskforce of the European Respiratory Society and European Society for Clinical Microbiology and Infectious Diseases // *Clin Microbiology and Infection.* - 2011. - Vol. 17. - P. E1-E59.
227. Wu, C. H. The Protein Information Resource / C. H. Wu, L. S. Yeh, H. Huang, L. Arminski, J. Castro-Alvear, Y. Chen, Z. Hu, P. Kourtesis, R. S. Ledley, B. E. Suzek, C. R. Vinayaka, J. Zhang, W. C. Barker // *Nucleic Acids Res.* - 2003. - Vol. 31, № 1. - P. 345-347.
228. Xiao, H. Proteomic Analysis of Human Saliva From Lung Cancer Patients Using Two-Dimensional Difference Gel Electrophoresis and Mass Spectrometry / H. Xiao, L. Zhang, H. Zhou, J. M. Lee, E. B. Garon, D. T. Wong // *Mol Cell Proteomics.* - 2012. - Vol. 11, № 2. - P. M111.012112.

229. Yamashita, J. Tumor neutrophil elastase is closely associated with the direct extension of non-small cell lung cancer into the aorta / J. Yamashita, M. Ogawa, M. Abe, N. Hayashi, Y. Kurusu, K. Kawahara, T. Shirakusa // *Chest*. - 1997. - Vol. 111, №4. - P. 885-890.
230. Yang, Y. Heat shock protein gp96 is a master chaperone for toll-like receptors and is important in the innate function of macrophages / Y. Yang, B. Liu, J. Dai, P. K. Srivastava, D. J. Zammit, L. LeFrançois, Z. Li // *Immunity*. - 2007. - Vol. 26. - P. 215–226.
231. You, B. Expiratory capnography in asthma: evaluation of various shape indices / B. You, R. Peslin, C. Duvivier, V. D. Vu, J. P. Grilliat // *Eur Respir J*. - 1994. - Vol. 7. - P. 318-323.
232. Zetterquist, W. Salivary contribution to exhaled nitric oxide / W. Zetterquist, C. Pedroletti, J. O. N. Lundberg, K. Alving // *Eur Respir J*. -1999. - Vol. 13. - P. 327-333.
233. Zetterquist, W. Oral bacteria--the missing link to ambiguous findings of exhaled nitrogen oxides in cystic fibrosis / W. Zetterquist, H. Marteus, P. Kalm-Stephens, E. Nas, L. Nordvall, M. Johannesson, K. Alving // *Respir Med*. - 2009. - Vol. 103, № 2. - P. 187-193.
234. Zhang, T. Lysosomal cathepsin B plays an important role in antigen processing, while cathepsin D is involved in degradation of the invariant chain in ovalbumin-immunized mice / T. Zhang, Y. Maekawa, J. Hanba, T. Dainichi, B. F. Nashed, H. Hisaeda, T. Sakai, T. Asao, K. Himeno, R. A. Good, N. Katunuma // *Immunology*. - 2000. - Vol. 100, № 1. - P. 13–20.
235. Zietkowski, Z. High-sensitivity C-reactive protein in the exhaled breath condensate and serum in stable and unstable asthma / Z. Zietkowski, M. M. Tomasiak-Lozowska, R. Skiepkowski, B. Mroczko, M. Szmitkowski, A. Bodzenta-Lukaszyk // *Respir Med*. - 2009. - Vol. 103, № 3. - P. 379–385.
236. Zong, D. D. Epigenetic mechanisms in chronic obstructive pulmonary disease / D. D. Zong, R. Y. Ouyang, P. Chen // *Eur Rev Med Pharmacol Sci*. - 2015. - Vol. 19. - P. 844-856.
237. Zou, Y. CEA, SCC and NSE levels in exhaled breath condensate—possible markers for early detection of lung cancer / Y. Zou, L. Wang, C. Zhao, Y. Hu, S. Xu, K. Ying, P. Wang, X. Chen // *J Breath Res*. - 2013. - Vol. 7, № 4. - P. 047101.
238. Zumstein, P. P. Amino acid sequence of a variant pro-form of insulin-like growth factor II / P. P. Zumstein, C. Luthi, R. E. Humbel // *PNAS*. - 1985. - Vol. 82. - P. 3169-3172.

ПРИЛОЖЕНИЕ

Таблица 1-П. «Общие» белки, идентифицированные в протеомах КВВ как контрольной группы, так и групп доноров с диагнозами ХОБЛ и пневмония.

Gene	Protein Cell Localization	Biological/Pathological Processes
ACTG1	Cytoskeleton, Extracellular, Cytosol, Nucleus	innate immune response
AGP2 (ORM2)	Extracellular	regulation of immune system process; acute-phase response
ALB	Extracellular, Cytoskeleton, Cytosol, Endoplasmic reticulum, Mitochondrion, Nucleus, Plasma membrane	blood coagulation; platelet activation
AMBP	Extracellular, Plasma membrane	negative regulation of immune response; negative regulation of endopeptidase activity
AZGP1	Extracellular	antigen processing and presentation of peptide antigen via MHC class I; immune response; positive regulation of T cell mediated cytotoxicity
CDSN	Cytoskeleton, Extracellular	cell-cell adhesion; epidermal barrier integrity. The gene is located in the major histocompatibility complex (MHC) class I region on chromosome 6
CSTA	Cytoskeleton, Nucleus, Cytosol	negative regulation of peptidase activity; negative regulation of endopeptidase activity; cell-cell adhesion
DCD	Extracellular	defense response to fungus; defense response to bacterium; proteolysis
DMKN	Extracellular	–
DNAH14	Cytosol	microtubule-based movement
DSP	Cytoskeleton, Plasma membrane, Nucleus	adherens junction organization; desmosome organization; intermediate filament organization
FLG2	Cytosol, Nucleus	structural molecule activity
HRNR	Extracellular, Cytosol	establishment of skin barrier
IGHA1	Extracellular	immune response; antibacterial humoral response; positive regulation of respiratory burst
JUP	Cytoskeleton, Cytosol, Plasma membrane, Nucleus	desmosome assembly; cell morphogenesis; cell-cell adhesion; positive regulation of cell growth; regulation of cell proliferation; detection of mechanical stimulus
KNG1	Extracellular, Plasma membrane	negative regulation of blood coagulation; inflammatory response
LCN1	Extracellular	transport; response to stimulus; negative regulation of endopeptidase activity
LYZ	Extracellular, Cytoskeleton, Lysosome, Plasma membrane, Vacuole	inflammatory response; defense response to bacterium
PIP	Extracellular	positive regulation of gene expression; negative regulation of T cell apoptotic process/tumor suppressor signaling and anti-proliferative cascade
PTGDS	Endoplasmic reticulum, Extracellular space, Golgi apparatus	prostaglandin biosynthetic process; arachidonic acid metabolic process; transport
SHROOM3	Cytoskeleton, Cytosol, Plasma membrane, Nucleus	actin cytoskeleton organization; pattern specification process; cell morphogenesis
SPP1	Extracellular	extracellular matrix organization; inflammatory response/ Antiviral Response; Lung Cancer

TXN	Extracellular, Mitochondrion, Nucleus, Cytosol	oxidation-reduction process; cellular response to glucose stimulus; innate immune response; transcription, DNA-templated; signal transduction
------------	---	---

Таблица 2-П. Белки, идентифицированные только в протеомах КВВ группы с диагнозом ХОБЛ.

Gene	Protein Cell Localization	Biological/Pathological Processes
ATP1A3	Endoplasmic reticulum, Golgi apparatus, Plasma membrane	ATP biosynthetic process; transmembrane transport; cation transport
CD3EAP	Nucleus, Mitochondrion	transmembrane receptor protein tyrosine kinase signaling pathway; rRNA transcription
COL1A1	Endoplasmic reticulum, Extracellular	blood vessel development; extracellular matrix organization; platelet activation; protein transport; cellular response to various stimulus and hormones
EIF2S1	Cytosol, Endoplasmic reticulum	activation of signaling protein activity involved in unfolded protein response; gene expression; regulation of translational initiation in response to stress
GPR179	Plasma membrane	G-protein coupled receptor signaling pathway
HBB	Cytosol, Extracellular	blood coagulation; positive regulation of cell death; oxygen transport; response to hydrogen peroxide
HBG1	Cytosol	blood coagulation/ IL-4 Pathway
HTRA2	Cytosol, Endoplasmic reticulum, Mitochondrion, Nucleus, Plasma membrane	intrinsic apoptotic signaling pathway; positive regulation of apoptotic process
LACRT	Extracellular	positive regulation of cell proliferation; calcium-mediated signaling; positive regulation of NFAT protein import into nucleus
NOC4L	Nucleus	rRNA processing; ribosome biogenesis
NT5C1B	Cytosol	nucleotide metabolic process; small molecule metabolic process
OSBPL8	Cytosol	lipid transport; ; positive regulation of insulin receptor signaling pathway; positive regulation of protein kinase B signaling
PDE8B	Cytosol	cAMP catabolic process; regulation of transcription, DNA-templated; signal transduction; negative regulation of insulin secretion
PRDX1	Nucleus, Mitochondrion	response to reactive oxygen species; regulation of stress-activated MAPK cascade
SHBG	Extracellular, Cytosol	hormone transport
TFAP2D	Nucleus	negative regulation of apoptotic process; positive regulation of transcription, DNA-templated; positive regulation of transcription from RNA polymerase II promoter
TLE2	Nucleus	transcription, DNA-templated
UBA6	—	cellular protein modification process; ubiquitin-dependent protein catabolic process
ZMYND8	Nucleus	negative regulation of transcription from RNA polymerase II promoter

Таблица 3-П. Белки, идентифицированные только в протеомах КВВ группы с диагнозом ПНЕВМОНИЯ.

Gene	Protein Cell Localization	Biological/Pathological Processes
ACTA2	Cytoskeleton, Cytosol	regulation of blood pressure; response to virus
ANXA1	Cytoskeleton, Extracellular, Cytosol, Plasma membrane	neutrophil homeostasis; inflammatory response; positive regulation of prostaglandin biosynthetic process; alpha-beta T cell differentiation
ANXA2	Extracellular, Cytosol, Plasma membrane	angiogenesis; body fluid secretion; collagen fibril organization/ Hypoxia Signaling Pathway
BAG6	Cytosol, Nucleus	immune system process; ubiquitin-dependent protein catabolic process; transport
CRNN	Cytosol	cell-cell adhesion
CST6	Extracellular	anatomical structure morphogenesis; negative regulation of endopeptidase activity
CSTB	Cytosol, Nucleus	negative regulation of peptidase activity; negative regulation of endopeptidase activity; regulation of apoptotic process
EGFR	Endoplasmic reticulum, Endosome, Extracellular, Golgi apparatus, Nucleus, Plasma membrane, Cytosol, Cytoskeleton	response to stress; response to osmotic stress; cell-cell adhesion; innate immune response/Osmotic Stress
Histone H2B	Nucleus	nucleosome assembly; defense response to bacterium
HSP90B1	Cytosol, Endoplasmic reticulum	response to hypoxia; protein folding; cellular protein metabolic process; innate immune response
KMT2D	Nucleus	transcription, DNA-templated
LAMA1	Extracellular	cell adhesion; extracellular matrix organization
PDE4D	Cytosol	positive regulation of interferon-gamma production; positive regulation of interleukin-2 production; positive regulation of interleukin-5 production
PRSS3	Extracellular, Cytosol	small molecule metabolic process; innate immune response
S100A8	Cytoskeleton, Extracellular, Cytosol, Plasma membrane	cytokine production; leukocyte migration involved in inflammatory response; acute inflammatory response; chemokine production; defense response to bacterium; innate immune response
SMCHD1	Nucleus	chromosome organization
SPRR3	Cytosol	epidermis development; peptide cross-linking
ZG16B	Extracellular	–

Таблица 4-П. Белки КВВ доноров с диагнозом «рак легкого», которые чаще встречаются в здоровых тканях легкого

№	Белок (код в базе UniProt)	Семейство (домен)	№	Белок (код в базе UniProt)	Семейство (домен)
1	Цистатин-А (P01040)	цистатины	7	Кератин 9 (P35527)	промежуточные филаменты
2	Кератин 13 изоформа b (A1A4E9)	промежуточные филаменты	8	Кератин 17 (Q04695)	промежуточные филаменты
3	Кератин 6A (P02538)	промежуточные филаменты	9	предшественник альфа1-микроглобулина/бикунина (P02760)	липокалины
4	Кератин 16 (P08779)	промежуточные филаменты	10	Простагландин-Н2 D-изомераза (P41222)	липокалины
5	Кератин 15 (P19012)	промежуточные филаменты	11	Хорнерин (Q86YZ3)	S100-белки
6	Кератин 4 (P19013)	промежуточные филаменты	12	Убиквитин-40S рибосомальный белок S27a (P62979)	убиквитины

Таблица 5-П. Белки КВВ доноров с диагнозом «рак легкого», которые чаще встречаются в тканях аденокарциномы легкого

№	Белок (код в базе UniProt)	Семейство (домен)	№	Белок (код в базе UniProt)	Семейство (домен)
1	Zinc finger protein (O14887)	-	35	Four and a half LIM domains protein 2 (Q14192)	LIM zinc-binding domain
2	Zinc finger protein NFX.1 (O43831)	-	36	Four and a half LIM domains protein 5 (Q5TD97)	LIM zinc-binding domain
3	DNA-связывающий белок (O14887)	-	37	LIM and cysteine-rich domains protein 1 (Q9NZU5)	LIM zinc-binding domain
4	CDK5 regulatory subunit-associated protein 2 (Q96SN8)	-	38	Alpha-1-acid glycoprotein 1 (P02763)	липокалины
5	Белок 9, ассоциированный со сперматогенезом (Q9BWV2)	-	39	Лизофосфолипид ацилтрансфераза 5 (Q6P1A2)	мембран-связанные ацилтрансферазы
6	Iron-responsive element-binding protein 2 (P48200)	аконитазы/IPM изомеразы	40	HLA class I histocompatibility antigen, A-25 alpha chain (P18462)	МНС class I family
7	АДФ-рибозилциклаза 2 (Q10588)	ADP-рибозилциклазы	41	Ceruloplasmin (P00450)	multicopper oxidase family
8	Aldo-keto reductase family 1 member C1 (Q04828)	альдо/кеторедуктазы	42	Никотинамид фосфорибозилтрансфераза (P43490)	NAPRTase family
9	E3 ubiquitin-protein ligase MIB1 (Q86YT6)	ANK repeat	43	Nesprin-3 (Q6ZMZ3)	nesprin family
10	Ankyrin repeat domain-containing protein 36B (Q8N2N9)	ANKRD36 family	44	Паксиллин (P49023)	паксиллины
11	АТФ синтаза субъединица дельта, митохондриальная (P30049)	АТФазы, цепь epsilon	45	Ubiquitin carboxyl-terminal hydrolase isozyme L1 (P09936)	пептидазы C12
12	BMP and activin membrane-bound inhibitor homolog (Q13145)	BAMBI family	46	Гаптоглобин (P00738)	пептидазы S1
13	Регулятор апоптоза BAX (Q07812)	Bcl-2 family	47	Фактор коагуляции VII (P08709)	пептидазы S1
14	Фруктозобифосфат альдолаза А (P04075)	фруктозобифосфат альдолаза класс I	48	Дипептидиламинопептидаза-подобный белок 6 (P42658)	пептидазы S9B
15	Компонент комплемента C9 (P02748)	комплемент C6/C7/C8/C9	49	Фосфоглицератмутаза 1 (P18669)	фосфоглицерат-мутаза
16	Компонент комплемента C7 (P10643)	комплемент C6/C7/C8/C9	50	E3 ubiquitin-protein ligase ARIH2 (O95376)	RBR family
17	Protein cornichon homolog 4 (Q9P003)	cornichon family	51	Рибозофосфат-пирофосфокиназа 3 (P21108)	рибозофосфат-пирофосфокиназы
18	G1/S-специфичный циклин-D2 (P30279)	циклины	52	Serum amyloid A protein (P02735)	SAA family
19	Dihydropyrimidinase-related protein 5 (Q9BPU6)	DHOase family	53	Альфа-1-антихимотрипсин (P01011)	серпины
20	Protein Downstream neighbor of Son (Q9QXP4)	DONSON family	54	Ангиотензиноген (P01019)	серпины

21	Scavenger receptor class F member 1 (Q14162)	EGF-like domain	55	Гепарин-кофактор 2 (P05546)	серпины
22	Protein FAM189A2 (Q15884)	FAM189 family	56	Pigment epithelium-derived factor (P36955)	серпины
23	Ферритин (Q96CU0)	ферритины	57	Protein Z-dependent protease inhibitor (Q9UK55)	серпины
24	Фибриноген, альфа цепь (P02671)	fibrinogen C-terminal domain	58	C4b-binding protein alpha chain (P04003)	Sushi (CCP/SCR) domain
25	Фибронектин (P02751)	fibronectin type-I,II,III domain	59	Complement factor H-related protein 3 (Q02985)	Sushi (CCP/SCR) domain
26	EGF-containing fibulin-like extracellular matrix protein 1 (Q12805)	fibulin family	60	T-complex protein 1 subunit epsilon (P48643)	ТСП-1 шаперонины
27	C1GALT1-specific chaperone 1 (Q96EU7)	glycosyltransferase 31 family	61	Тромбоспондин-1 (P07996)	тромбоспондины
28	Муцин-5В (Q9HC84)	HAT repeat	62	Tumor necrosis factor receptor superfamily member 27 (Q9HAV5)	TNFR-Cys repeat
29	Бетаин--гомоцистеин S-метилтрансфераза 2 (Q9H2M3)	Hcy-binding domain	63	E3 ubiquitin-protein ligase CHIP (Q9UNE7)	TPR repeat
30	Low affinity immunoglobulin gamma Fc region receptor III-A (P08637)	Ig-like C2-type (immunoglobulin-like) domain	64	Лактотрансферрин (P02788)	трансферрины
31	DNA-binding protein Ikaros (Q13422)	Ikaros C2H2-type zinc-finger protein family	65	Триозефосфатизомераза (P60174)	триозефосфатизомеразы
32	Супероксиддисмутаза [Mn], митохондриальная (P04179)	Fe/Mn супероксиддисмутаза	66	Ephrin type-A receptor 3 (P29320)	Тур protein kinase family
33	Zinc finger protein 37A (P17032)	krueppel C2H2-type zinc-finger protein family	67	Tyrosine-protein kinase transmembrane receptor ROR2 (Q01974)	Тур protein kinase family
34	Zinc finger protein 138 (P52744)	krueppel C2H2-type zinc-finger protein family	68	Аполипопротеин В-100 (P04114)	vitellogenin domain

Таблица 6-П. Белки, которые чаще встречаются в здоровых тканях по сравнению с аденокарциномой легкого

№	Белок (код в базе UniProt)	Семейство (домен)	№	Белок (код в базе UniProt)	Семейство (домен)
1	UMP-СМР киназа (P30085)	аденилат киназы	20	Zinc finger and SCAN domain-containing protein 22 (P10073)	krueppel C2H2-type zinc-finger protein family
2	Аннексин А2 (P07355)	аннексины	21	Тромбоцитарный гликопротеин 1b, альфа-цепь (P07359)	LRR (лейцин-обогащенные) повторы
3	Аполипопротеин А-1 (P02647)	аполипопротеи-ны А1/А4/Е	22	mRNA-capping enzyme (O60942)	non-receptor class of the protein-tyrosine phosphatase family
4	АТФ синтаза, субъединица бета, митохондриальная (P06576)	АТФазы, альфа/бета цепи	23	Кальпаин-10 (Q9HC96)	пептидазы С2
5	Ингибитор протеазы Куница 4 (Q6UDR6)	ВРТИ/Куниц ингибиторный домен	24	Фактор коагуляции VII (P08709)	пептидазы S1
6	Предшественник Ovo-подобного фактора транскрипции 1 (O14753)	C2H2-type zinc finger	25	Субкомпонент комплемента C1 (P09871)	пептидазы S1
7	Внеклеточная супероксиддисмутаза [Cu-Zn] (P08294)	Cu-Zn-супероксиддисмутаза	26	GroUp IID секреторная фосфолипаза А2 (Q9UNK4)	фосфолипазы А2
8	Nidogen-2 (Q14112)	EGF-подобный домен	27	Ras-related protein Rab-6A (P20340)	Rab family
9	Альфа-енолаза (P06733)	енолазы	28	Ras-related protein Rab-39A (Q14964)	Rab family
10	Ретинол-связывающий белок 5 (P82980)	белки, связывающие жирные кислоты (FABP)	29	Гиппокальцин-подобный белок 1 (P37235)	рековерины
11	Белок, связывающий жирные кислоты, эпидермальный (Q01469)	белки, связывающие жирные кислоты (FABP)	30	Pigment epithelium-derived factor (P36955)	серпины
12	Фиколин-3 (O75636)	фиколинлектины	31	Карбонилредуктаза [НАДФ Н] 1 (P16152)	короткоцепочечные дегидрогеназы/редуктазы
13	Гемоглобин, субъединица бета (P68871)	глобины	32	Complement factor H-related protein 2 (P36980)	Sushi (CCP/SCR) domain
14	Лутропин, субъединица бета (P01229)	гликопротеиновые гормоны, бета-субъединицы	33	Тромбоспондин-1 (P07996)	тромбоспондины
15	Рецептор релаксина-3 2 (Q8TDU9)	G-белок сдвоенные рецепторы 1	34	Тромбоспондин-2 (P35442)	тромбоспондины
16	HAUS аугмин-подобный комплекс, субъединица 7 (Q99871)	HAUS7	35	CD27 антиген (P26842)	TNFR-Cys повторы
17	Белок, связывающий инсулин-подобный фактор роста, 1 (P08833)	IGFBP N-терминальный домен	36	Серотрансферрин (P02787)	трансферрины
18	Интерлейкин-20 (Q9NYY1)	интерлейкины-10	37	Аполипопротеин В-100 (P04114)	vitellogenin domain
19	Преламин-А/С (Q6UYC3)	промежуточные филаменты			

Таблица 7-II. Пептиды, наиболее часто встречающиеся в промывочных заколах.

№	Пептид	Содержание в промывочных пробах (%)	Содержание в пробах доноров (%)
1	DVFLGMFLYEYAR	100,0	18,4
2	FQNALLVR	88,9	23,3
3	LVAASQAALGL	88,9	17,5
4	ETLLQDFR	83,3	15,5
5	KVPQVSTPTLVEVSR	83,3	36,9
6	YTKKVPQVSTPTLVEVSR	83,3	19,4
7	AFIQLWAFDAVK	77,8	16,5
8	DDNPNLPR	77,8	13,6
9	LVNEVTEFAK	77,8	22,3
10	YEELQITAGR	77,8	71,8
11	AEFAEVSK	72,2	33,0
12	GVPLTS	72,2	0,0
13	SLVNLGGSKSISISVAR	72,2	2,9
14	LAADDFR	66,7	77,7
15	QPSSTLPS	66,7	0,0
16	RHPDYSVVLRLR	66,7	15,5
17	SDLELPGS	66,7	1,9
18	SLDLDSIIAEVK	66,7	78,6
19	TLLDIDNTR	66,7	69,9
20	TLLEGEESR	66,7	98,1
21	AEFAEVSKLVTDLTK	61,1	26,2
22	ANALLANGVELR	61,1	6,8
23	AVTLVSPKPDSTLINGLGR	61,1	11,7
24	EKPVV	61,1	0,0
25	FLEQQNQVLQTK	61,1	75,7
26	FSSSSGYGGGSSR	61,1	74,8
27	GGSGGSYGGGGSGGGYGGGSGSR	61,1	56,3
28	GPGNEAFTDWVHGMPEVSTGR	61,1	5,8
29	KQTALVELVK	61,1	10,7
30	KVPQVSTPTLVEVSRNLGK	61,1	4,9
31	LSQRFPKAEFAEVSK	61,1	11,7
32	LVVQSP	61,1	5,8
33	NKYEDEINKR	61,1	59,2
34	NLDLDSIIAEVK	61,1	58,3
35	SAPPL	61,1	0,0
36	TVAAPSVF	61,1	1,9
37	TYETTLEK	61,1	11,7
38	VTMQNLNDR	61,1	83,5

Таблица 8-II. Значения foldchange пептидов, идентифицированных в пробах доноров КВВ и включенных в анализ после отсева выпавших экспериментов и фильтрации редко встречающихся пептидов.

Peptide	Fold Change	Peptide	Fold Change
VNVDEVGGEALGR	75161.2903226	VISSR	8.87428673806
AADDTWEPFASGK	61290.3225806	ILGMLIVS	8.69793520683
RHPYFYAPELLFFAK	55322.5806452	ITRFIK	8.6862246129
RSVMDSLTR	42580.6451613	EIPLLYLLGEQYIKAMTDMSK	8.62626382239
TSESGELHGLTTEEEFV	40967.7419355	VCPSQTR	8.4042272297
ATVTPSPVKGK	38548.3870968	NKTIKACDK	7.82462535179
ELDRDTVFALVNYIFFK	34516.1290323	ALAGGKRR	7.79660799425
DAMSSKL	24677.4193548	LVAGLTNGYK	7.36002326131
DGKSGLIK	18548.3870968	EKEMK	7.27599538795
IDDTIR	18387.0967742	FPGLPE	7.08132169193
VESAIGK	17741.9354839	MFGPK	7.05103115285
QTALVELVKHKPK	16612.9032258	MFANEAAK	0.248799943421
MVANIDK	14516.1290323	DVMMHVAS	0.248783953251
TVADHK	12258.0645161	VPLPVR	0.243818302988
ASAEKIMQIGK	10161.2903226	ATLVSLPR	0.242165139505
AEITDMR	8709.67741935	DACPEKK	0.235860562461
IVSAK	8548.38709677	PYAGCR	0.234442623207
FIGPNADVIR	6935.48387097	DIQNP	0.232216265588
FPGLAK	4677.41935484	YWPGNTTLLLK	0.214899047298
ELPVNVK	4516.12903226	IIGMR	0.210194453224
EEDGMIR	4193.5483871	APDRKYK	0.207319961042
GKKGFIK	3870.96774194	MYQDDVSGNTANLR	0.206252524047
KGGAV	2258.06451613	MVLAGLIR	0.204868510458
DRLML	81.4312390511	NNFCARSSE	0.204287456105
GTCPCCR	51.9773866283	GIVGY	0.1934323277
KIKHSQFAR	25.9886933142	AEDVTRR	0.190492048015
SVVTVIDVIFYK	24.4794441172	ERTNDR	0.181410795474
ASNNMDK	17.3880563373	AASPSK	0.17690039661
YTIAALLSPYSYSTTA	17.2482787074	LKLVDNTHISTISASDI	0.174166676957
APHDPGR	15.9390516425	APAPNEAGR	0.174155930706
DYYQGYHLTHDEMDR	15.6492507036	EIPPIAK	0.145117754996
VEPGSIHK	14.7665818419	LSGLLDLALGKDYVR	0.136620324366
TMRSKSLAK	14.7269262114	SSEDPNEDIVER	0.135783983746
AFIQLWAFDAVK	14.1334719156	GHLTKH	0.124407967823
RTMIR	12.0999176574	IALICSNK	0.108858721136
VLSSR	11.1363598282	AGTPSK	0.104493558424
EPKAVMK	10.8207110898	TGSGDIENYNDATQVR	0.100159938163
DGPMLSQR	10.8165406042	LDNPMGIAGDSDFMEMAVSNIK	0.099523815404
AAIMR	10.7226347413	GTALAEVLK	0.0725781328413
PGEEGR	10.6143429612	IGIMR	0.0696623722824
LEKTTEKVK	8.915383587		