

На правах рукописи

Федорченко Кристина Юрьевна

«МЕТОД НЕИНВАЗИВНОЙ ДИАГНОСТИКИ РАКА ЛЕГКОГО,
ОСНОВАННЫЙ НА АНАЛИЗЕ БЕЛКОВОГО И ПЕПТИДНОГО СОСТАВА
КОНДЕНСАТА ВЫДЫХАЕМОГО ВОЗДУХА ЧЕЛОВЕКА»

Специальность 03.01.02 – Биофизика

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата биологических наук

Москва – 2017

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте биохимической физики им. Н.М. Эмануэля Российской академии наук.

Научный руководитель:

Варфоломеев Сергей Дмитриевич,
член-корреспондент РАН, доктор химических наук,
профессор, научный руководитель Федерального
государственного бюджетного учреждения науки
Института биохимической физики им. Н.М.
Эмануэля Российской академии наук.

Официальные оппоненты:

Шишкин Сергей Сергеевич,
доктор биологических наук, профессор,
заведующий лабораторией биомедицинских
исследований Федерального государственного
учреждения «Федеральный исследовательский
центр «Фундаментальные основы биотехнологии»
Российской академии наук» (ФИЦ Биотехнологии
РАН);

Маевский Евгений Ильич,
доктор медицинских наук, профессор, заместитель
директора по науке, заведующий лабораторией
энергетики биологических систем Федерального
государственного бюджетного учреждения науки
Института теоретической и экспериментальной
биофизики Российской академии наук.

Ведущая организация:

Федеральное государственное бюджетное
учреждение науки Институт биоорганической
химии им. академиков М.М. Шемякина и Ю.А.
Овчинникова Российской академии наук.

Защита состоится "21" февраля 2018 г. в 12 часов 00 минут на заседании диссертационного совета Д 002.039.01 при Федеральном государственном бюджетном учреждении науки Институте биохимической физики им. Н.М. Эмануэля Российской академии наук по адресу: 119334, Москва, ул. Косыгина, 4.

С диссертацией можно ознакомиться в библиотеке Федерального Государственного бюджетного учреждения науки Института химической физики им. Н.Н. Семенова Российской академии наук по адресу: 119991, Москва, Ленинский проспект, 38 и на сайте <http://new.chph.ras.ru/diss-sovet/razmeshchennye-dissertatsii/213-dissertatsiya-fedorchenko-kristiny-yurevny>.

Автореферат разослан: " ____ " декабря 2017 г.

Ученый секретарь
диссертационного совета Д 002.039.01,
кандидат химических наук

Мазалецкая Лидия Ивановна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования.

Степень разработанности темы исследования.

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 Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов Министерства здравоохранения и социальных служб США (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). 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 инновационных технологий, таких как метаболомика, протеомика, масс-спектрометрия, обладает огромным потенциалом в области профилирования биомаркеров выдыхаемого воздуха.

Биомаркеры выдыхаемого воздуха могут оцениваться для понимания патомеханизма заболевания и также во вспомогательных целях при назначении соответствующей терапии.

Среди патологий дыхательной системы человека можно выделить рак легкого, для которого разработка новых подходов в дифференциальной диагностике на фоне других заболеваний легких и респираторного тракта является наиболее актуальной.

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) Оработать 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. Основным результатом исследования – информация о сравнительном белковом и пептидном составе КВВ больных раком легкого и пациентов с хронической обструктивной болезнью легких (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 – внедрения нового метода ранней диагностики онкологических заболеваний, – так и для фундаментальных исследований, связанных с изучением процессов развития патологических изменений на ранних стадиях онкологических заболеваний. Созданные в рамках исследования подходы к проведению масштабных исследований КВВ в условиях клиники, а так же подходы к сбору, хранению, подготовке и анализу образцов являются абсолютно новыми и станут основой 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 диссертанта.

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 (МС/МС) производились автором при участии ведущего 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.

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 полиакриламидном геле (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. В работе использовали современное оборудование ЦКП "Новые материалы и технологии" ИБХФ РАН.

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. Диссертационная работа состоит из введения, обзора литературы, описания материалов и методов исследования, результатов и их обсуждения, выводов, списка цитируемой литературы, приложения. Количество страниц 141, ссылок 238, рисунков 35, таблиц 21.

Сокращения, принятые в работе. MALDI-TOF – Времяпролетная масс-спектрометрия на основе матрично-активированной лазерной десорбции/ионизации, БАЛ – бронхо-альвеолярный лаваж, ВЭЖХ-МС/МС – высокоэффективная жидкостная хроматография в сочетании с тандемной масс-спектрометрией, ИМ – индуцированная мокрота, ИЦР – ионно-циклотронный резонанс, КВВ – конденсат выдыхаемого воздуха, МС – масс-спектрометрия, МС/МС – тандемная масс-спектрометрия, ПААГ – полиакриламидный гель, ТХУ – трихлорусусная кислота, ХОБЛ – хроническая обструктивная болезнь легких.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

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

В главе говорится об актуальности разработки методов ранней диагностики патологических состояний организма, о взаимосвязи состояния организма и респираторной системы, а также о существующих методах мониторинга показателей респираторной системы человека, области их применения и недостатках. Дыхательная система выполняет важнейшую функцию жизнеобеспечения и отражает образ жизни человека и состояние его здоровья. Однако ограничения имеющихся методов оценки состояния дыхательной системы не позволяют проводить раннюю диагностику таких социально значимых заболеваний, как ХОБЛ и рак легкого. Приведены данные многочисленных исследований, рассмотрены преимущества и недостатки каждого из существующих методов, критически оценены причины высокого уровня смертности от рака легкого и отсутствия скрининговых программ для этого заболевания.

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

Будучи неинвазивной, а, следовательно, легкой в проведении процедурой, такая проба дыхания, как КВВ, позволяет клиницистам и исследователям оценивать различные процессы, происходящие в организме человека. Сбор КВВ может быть осуществлен даже для очень тяжелых пациентов и повторен через короткие промежутки времени. Считается, что исследование КВВ 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.

Существует много разнообразного конденсирующего оборудования, но в настоящее время в большинстве исследований, посвященных КВВ, используются коммерчески доступные системы RTube (Respiratory Research Inc., VA) и ECoScreen (VIASYS Healthcare, Германия). Ввиду особенностей КВВ, использование иммуноферментного метода (например, ELISA) для идентификации белков может привести к потере части результатов. Альтернативный подход состоит в прямом анализе сложной белковой смеси с помощью ВЭЖХ-МС/МС без предварительного разделения.

На основании проведенного анализа литературных источников можно заключить, что:

– Развитие нового метода ранней диагностики рака легкого является актуальной медицинской проблемой;

– Несмотря на очевидный потенциал использования протеомного подхода в анализе КВВ, большинство исследователей (как следует из анализа опубликованных статей) не применяют протеомный подход, а концентрируются на поиске отдельных маркеров.

Глава 3. Материалы и методы

1) Доноры КВВ

В рамках настоящей работы обследовали 32 здоровых добровольца, 17 больных ХОБЛ, 13 – внебольничной пневмонией, 46 больных с диагностированным раком легкого (Таблица 1). Итого, за 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 (77)
Женщины, чел. (%)	15 (47)	4 (24)	6 (46)	12 (23)
Курение: по наст. вр. / бывший / не курящий	8/0/24	12/5/0	4/2/7	23/5/18
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. 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 КВВ было утверждено комитетами по этике ИБХФ РАН, НИИ пульмонологии РАН, МНИОИ имени П.А. Герцена и ИМБП РАН, все доноры подписали информированное согласие на участие в исследовании.

2) Сбор КВВ

Образцы были собраны в первой половине дня после тщательного ополаскивания полости рта дистиллированной водой. Особое внимание уделяли исключению носового дыхания (с помощью носового зажима) и белкового загрязнения слюной и слизистой носоглотки. Для исключения загрязнения образца КВВ слюной определяли присутствие в спектрах белка альфа-амилазы - специфического белка слюны.

Сбор КВВ пациентов с диагностированным раком легкого для облегчения процедуры и причинения как можно меньшего беспокойства пациентам осуществляли с помощью портативного устройства RTube (Respiratory Research, США). Образцы КВВ собирали в течение 10 мин с использованием предварительно охлажденной до -20°C Rtube.

Конденсаты выдыхаемого воздуха пациентов с диагностированными ХОБЛ и внебольничной пневмонией собирали с помощью стационарного устройства ECoScreen (VIASYS Healthcare, Германия). Образцы КВВ собирали в течение 10 мин с использованием тефлоновых приемников в модуле, охлаждаемом до -10°C .

Собранные пробы помещали в полипропиленовые пробирки, замораживали и хранили в морозильнике при -85°C до проведения анализа.

3) Обработка пробы КВВ перед масс-спектрометрическим анализом

Концентрирование белка для электрофореза в полиакриламидном геле осуществляли путем осаждения белка при помощи ТХУ или ацетона. Электрофорез в ПААГ проводили в денатурирующих условиях с добавлением детергента додецилсульфата натрия по методу Лэмли. Гидролиз белка в ПААГ проводили с использованием трипсина в течение 3 часов при 37°C , останавливали добавлением свежеприготовленного 0,5% (об/об) раствора муравьиной кислоты в воде. Осаждение белков на колонке C18 осуществляли с помощью ZipTip в соответствии с рекомендациями производителя (www.millipore.com). Обезжиривание пробы осуществляли обработкой смесью метанол/хлороформ.

Концентрирование белка с помощью лиофилизации проводили следующим образом: аликвоту КВВ (~объем 1 мл) переносили в полипропиленовые пробирки, устойчивые к низким температурам, с низкой белок-абсорбирующей поверхностью, и лиофилизовали в тех же пробирках до полного высыхания. Затем образцы КВВ подвергали гидролизу 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.

4) Масс-спектрометрический анализ образцов

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.

Эксперименты ВЭЖХ-МС/МС проводили на системе, состоящей из хроматографа Agilent 1100 (Agilent Technologies Inc., Санта-Клара, США) и гибридного масс-спектрометра LTQ-FT Ultra (Thermo, Бремен, Германия) — масс-спектрометр ионного циклотронного резонанса, совмещенный с линейной квадрупольной ионной ловушкой, использующейся для накопления ионов и измерения спектров столкновительно индуцированной фрагментации (МС/МС) ионов. Объем вводимой на колонку пробы составлял 1 мкл, использовалась колонка 75 мкм × 12 см с фазой Reprosil-Pur Basic C18, 3 мкм (Dr. Maisch HPLC GmbH, Аммербух-Энтринген, Германия). 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 Electron, Бремен, Германия) 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 5×10⁶). На второй стадии из ИЦР масс-спектра выбирали три максимальных пика, для которых производили столкновительно-индуцированную фрагментацию.

5) Биоинформационный анализ данных

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 NCBI nr и IPI Human 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, 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.

Для аннотирования и анализа результатов были использованы биоинформатические базы данных GeneCards, GeneOntology (GO), MOPED, BioGPS, UniProt, а также продукты компании Quagen 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.

Для моделирования и проектирования базы данных использовали программу MySQLWorkbench 6.3 CE. Заполнение базы данных осуществляли с помощью PHPMyAdmin software.

Для статистической обработки результатов использовали методы многопараметрической статистики.

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.

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

В главе приведены такие показатели проб КВВ контрольной группы, как рН, объем, концентрация белка, полученная на основании данных электрофореза в ПААГ. Разбавленность образцов является значительной проблемой при оценке биомаркеров из КВВ, поэтому необходимость концентрирования очевидна, но в то же время количество этапов пробоподготовки должно быть минимизировано, чтобы уменьшить потери белка. Показано, что лиофилизация является наиболее мягким и эффективным методом концентрирования проб КВВ перед протеомным анализом в сравнении с такими методами, как осаждение белков ТХУ, осаждение белков на колонке C18 (Рисунок 1). Кроме того, показано, что осаждение белков смесью метанол/хлороформ эффективно для идентификации белков, однако метод 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.

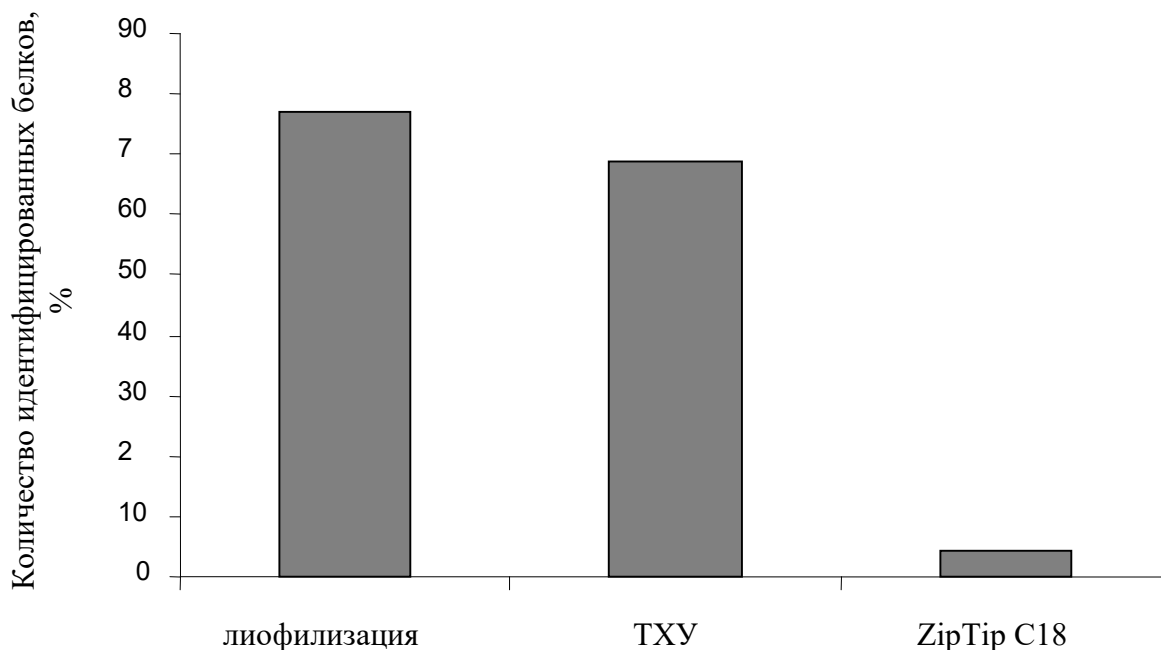


Рисунок 1. Результаты концентрирования смеси МС-стандартов (UPS1, Sigma) в КВВ различными методиками.

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

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, к часто встречающимся в пробах можно отнести цитоскелетные кератины II типа (3, 4, 5, 6) и цитоскелетные кератины I типа (14, 15, 16). Помимо кератинов, с высокой частотой в пробах идентифицировали дермцидин, простагландин-Н₂ D-изомеразу (PGDS2), предшественник альфа-1-микроглобулина/бикунина (AMBP), убиквитин, цистатин А, иммуноглобулин альфа (IGHA1) (эти белки были обнаружены в КВВ ≥ 50 % доноров). При интерпретации результатов в программе MetaCore (Clarivate Analytics, США) не было обнаружено заболеваний, ассоциированных с идентифицированным белковым составом КВВ, что косвенно подтверждает клинические данные.

Идентифицированный список, включающий как кератины, так и не кератиновые белки, может служить белковым фоном КВВ для дальнейших исследований.

Отдельный раздел главы посвящен исследованию белков, циркулирующих в окружающем воздухе, проведенному с помощью самодельной установки для конденсирования воздуха рабочих помещений. Определение ряда белков в окружающем воздухе (Таблица 2) 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. К таким «свободноциркулирующим» белкам следует относиться с осторожностью при анализе белкового состава с точки зрения диагностической значимости и применения анализа КВВ для диагностики.

Таблица 2. Белковый состав конденсата воздуха рабочих помещений.

Идентифицированный белок	Найденные пептиды
gi 114667176 Кератин, тип I цитоскелетный 14 Score: 244	10
gi 17318569 Кератин, тип II цитоскелетный 1 Score: 176	8
gi 547754 Кератин, тип II цитоскелетный 2 эпидермальный Score: 136	3
gi 122513 Гемоглобин субъединица бета-1 Score: 83	1

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

В главе приведены результаты анализа белкового состава проб КВВ пациентов с исследуемыми заболеваниями.

Показано, что КВВ группы с диагнозом ХОБЛ характеризовались наличием крови и обширной группой ядерных белков (47% от представленной группы белков ХОБЛ 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, который вырабатывается в ответ на активные формы кислорода и окислительный стресс, что соответствует клинической картине хронической обструктивной болезни легких.

В образцах КВВ больных пневмонией обнаружены белки острой фазы воспаления, фрагменты аннексина А1 и А2, цистатинов М и В, HSP90B1, принимающих участие в клеточном ответе на развивающуюся гипоксию. Кроме того, были идентифицированы белки альвеолярной базальной мембраны.

У 2 больных пневмонией идентифицирован белок POTEЕ (POTE ankyrin domain family, member E), который также был обнаружен у более 60% больных с немелкоклеточным раком легкого 1-2 стадии. Клинические обследования этих пациентов подтвердили подозрение на рак легкого, осложненный параканкрозной пневмонией, данные доноров КВВ перенесены в группу рака, для дальнейшего лечения пациенты были направлены в МНИОИ имени П.А. Герцена.

В ходе исследования всех групп доноров было идентифицировано более 350 белков, которые были взяты в дальнейший анализ (Рисунок 2).

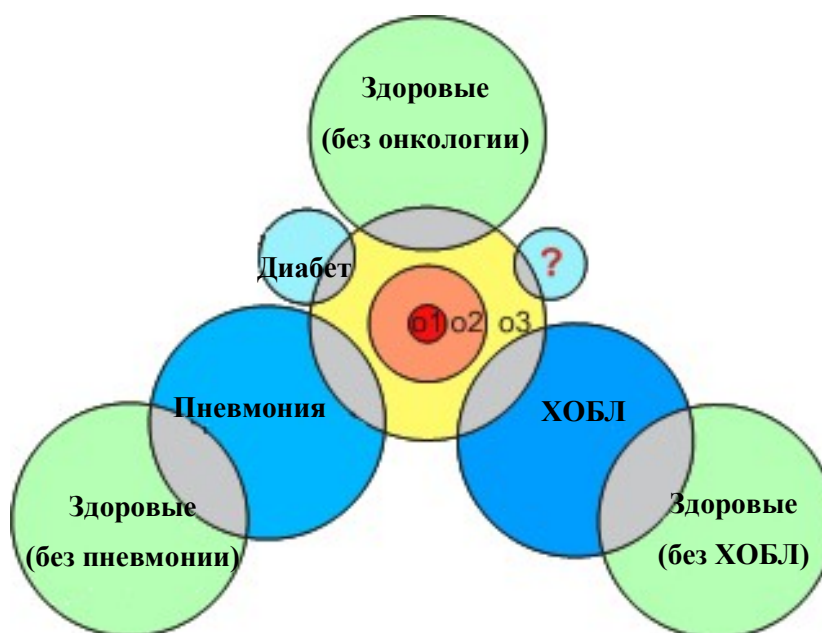


Рисунок 2. Схематичное представление полученных белковых профилей различных групп доноров, области пересечения – совокупность общих для групп белков. Область O1: маркеры рака, вошедшие в коммерческие наборы для ПЦР-профилирования рака легких по крови и не встречавшиеся в других группах; область O2: потенциальные маркеры рака, не встречавшиеся в других группах; область O3: белки, не связанные напрямую с фокусным заболеванием, либо встретившиеся в других группах доноров.

Помимо пересечений белкового состава образцов КВВ доноров с раком легкого с другими фокусными заболеваниями и здоровым контролем, выделилась группа доноров, имеющая в анамнезе диабет (Рисунок 1). В КВВ этих доноров, помимо белков, отражающих клиническую картину их основных заболеваний, присутствовали белки, связанные с углеводным транспортом, энергетическим обменом и метаболизмом инсулина. Еще одна выделившаяся группа - доноры, имеющие в анамнезе различные сосудистые заболевания (гипертоническая болезнь, ишемия аорты и т.д.) (область «?» на Рисунке 1). Так, например, для этой группы было характерно наличие в КВВ коллагена 1 типа. Возрастание количества этого белка часто ассоциируют с фиброзами.

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. Из них 2 белка вошли в область O1 (Рисунок 1), в том числе опухолевый антиген РOТEЕ, предложенный Национальным Институтом Здоровья (США) в качестве мишени для адресной доставки противораковых лекарств. 17 белков, участвующих в клеточном цикле и различных регуляторных процессах, например энергетическом обмене клетки, вошли в область потенциальных белковых маркеров рака легкого O2 (Рисунок 1). Итого, были выделены 19 белков, которые можно было бы предложить в качестве диагностической панели для рака легкого при исследовании белкового состава конденсата выдыхаемого воздуха (Таблица 3).

Таблица 3. 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}			Связь с клинической картиной ⁶
					Вне.	Внутр.	Кровь	Кожа	Л&РТ 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	-	-	+	+/+	-/+	-/+	+
	HSP90A1	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™.

88% пациентов с 1-2 стадией рака легкого были «идентифицированы» с помощью предложенной панели маркеров рака легких.

Также в главе приведен сравнительный анализ белкового состава конденсата выдыхаемого воздуха больных с диагнозом «рак легкого» и литературных данных по белковому составу ткани, пораженной опухолевым процессом. В качестве источника информации о белковом составе опухолевой ткани выбрали базу данных dbDEPC. Были найдены пересечения среди повышенно экспрессирующихся белков опухолевой ткани и белков КВВ (например, DPYSL5). Помимо сравнения с белками опухолевой ткани, проведен сравнительный анализ белкового состава конденсата выдыхаемого воздуха больных с диагнозом «рак легкого» с уровнем экспрессии генов при данном заболевании. Поиск наборов экспериментальных образцов велся в базе данных GEO Datasets, критерии для поиска использовали следующие: объект «Homo sapiens», диагноз «рак легкого», количество доноров не менее 50. В выбранном сете было найдено 36 пересечений с идентифицированными белками.

Глава 8. Новые подходы к разработке метода неинвазивной диагностики заболеваний легких на основе анализа состава конденсата выдыхаемого воздуха.

Глава посвящена созданию реляционной базы данных проб и доноров КВВ (Рисунок 3) и исследованию пептидного состава имеющихся образцов КВВ, а также построению линейной аналитической модели прогнозирования наличия у донора КВВ рака легкого на основе машинного обучения и логистической регрессии с кроссвалидацией. В базу данных была загружена информация из сделанной ранее таблицы Excel, содержащей всю информацию о каждом пациенте, его пробе/пробах и файлах после пробоподготовки.

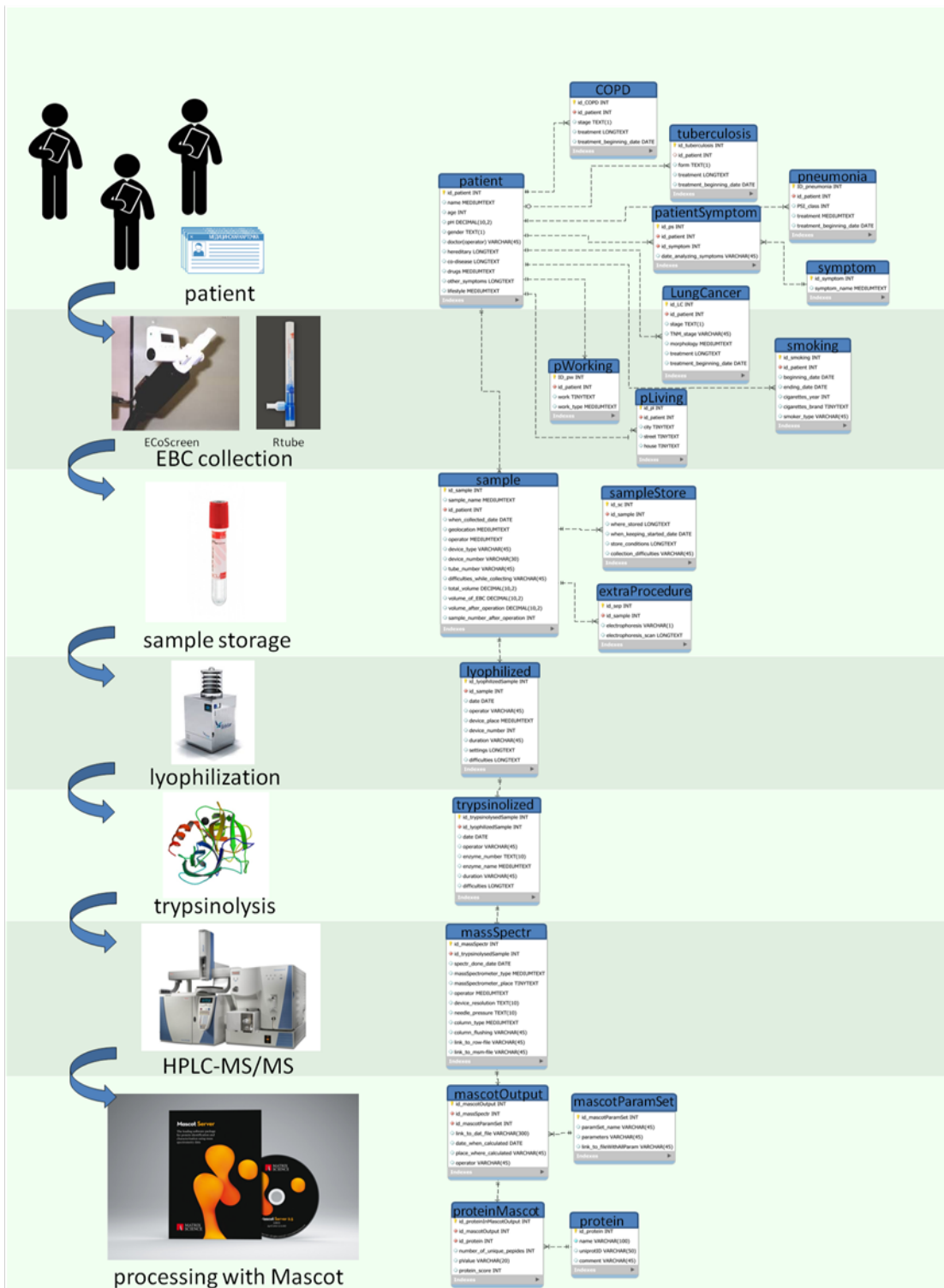


Рисунок 3. Блок-схема реляционной базы данных с привязкой к основным этапам.

Представленная база данных позволила производить поиск и сортировку пациентов по возрасту, полу, диагнозу и любому другому интересующему параметру (наследственность, принимаемые лекарства, первые симптомы и т.д.). Кроме того, в базу данных были загружены файлы со спектрами и масс-листами, что позволило автоматизировать и структурировать анализ.

Особенность анализа пептидного состава в данном исследовании состояла в том, что использовались триптические пептиды. В ходе анализа пептидного состава КВВ, идентифицировано порядка 55000 уникальных пептидов, которые были ранжированы по длине и аминокислотному составу (Рисунок 4, 5)

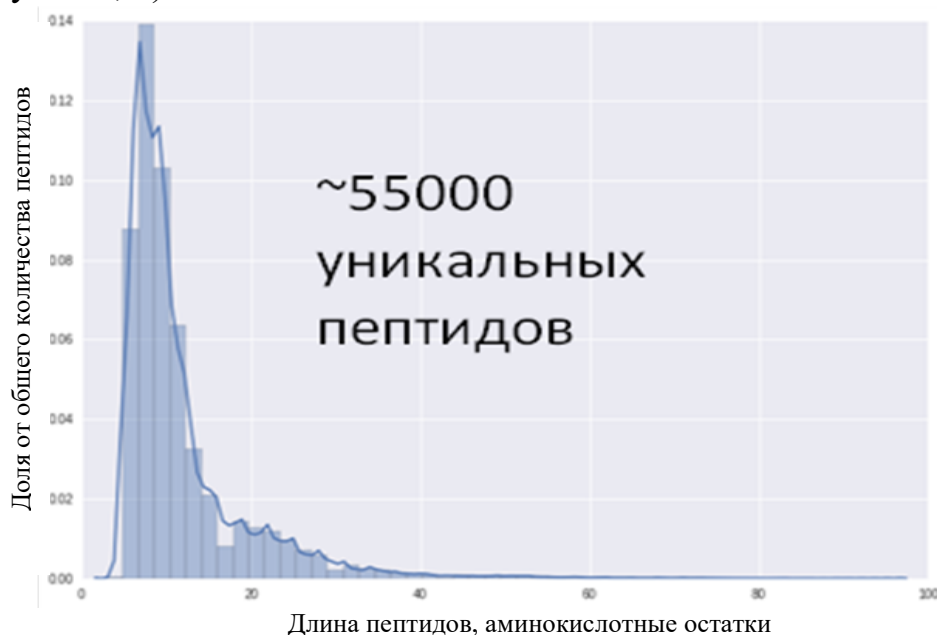


Рисунок 3. Распределение длин пептидов, идентифицированных в пробах КВВ и взятых в анализ после фильтрации редко встречающихся пептидов.

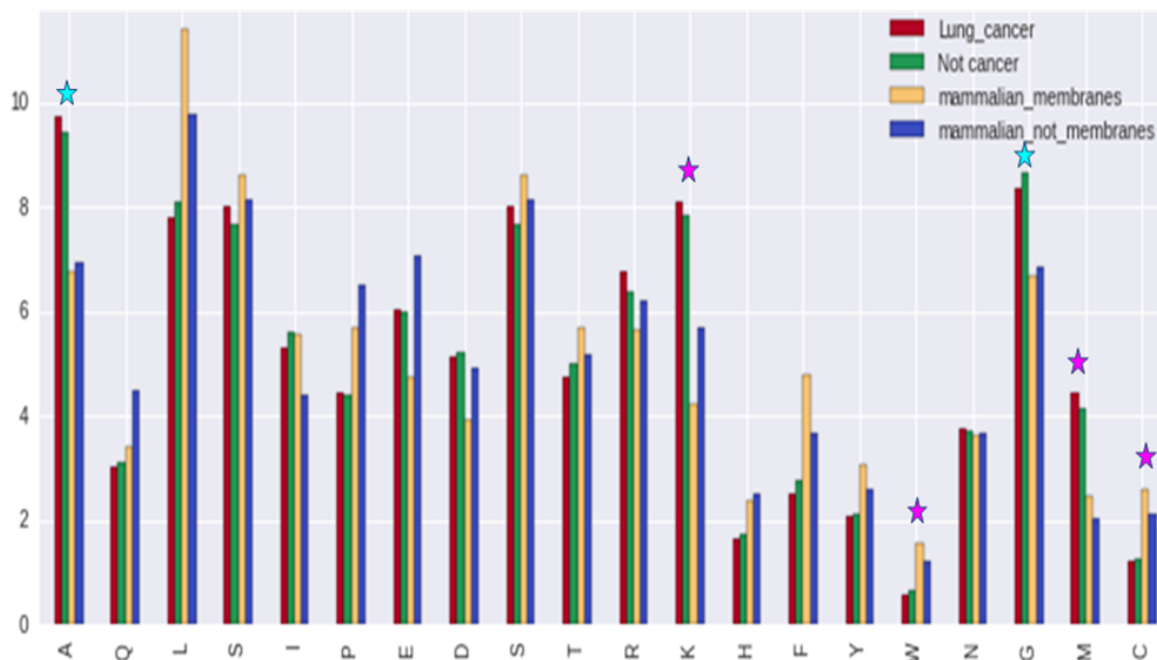


Рисунок 4. Аминокислотный состав пептидов, идентифицированных в пробах доноров КВВ, включенных в анализ, в сравнении с литературными данными по природному распределению аминокислот в мембранах и не мембранных структурах млекопитающих. Значимые отличия обозначены звёздочками.

Проведено сравнение когорт проб больных раком легкого и остальных доноров по пептидной нагрузке и показано, что значимой разницы не наблюдалось. Проведена иерархическая кластеризация образцов. (Рисунок 6)

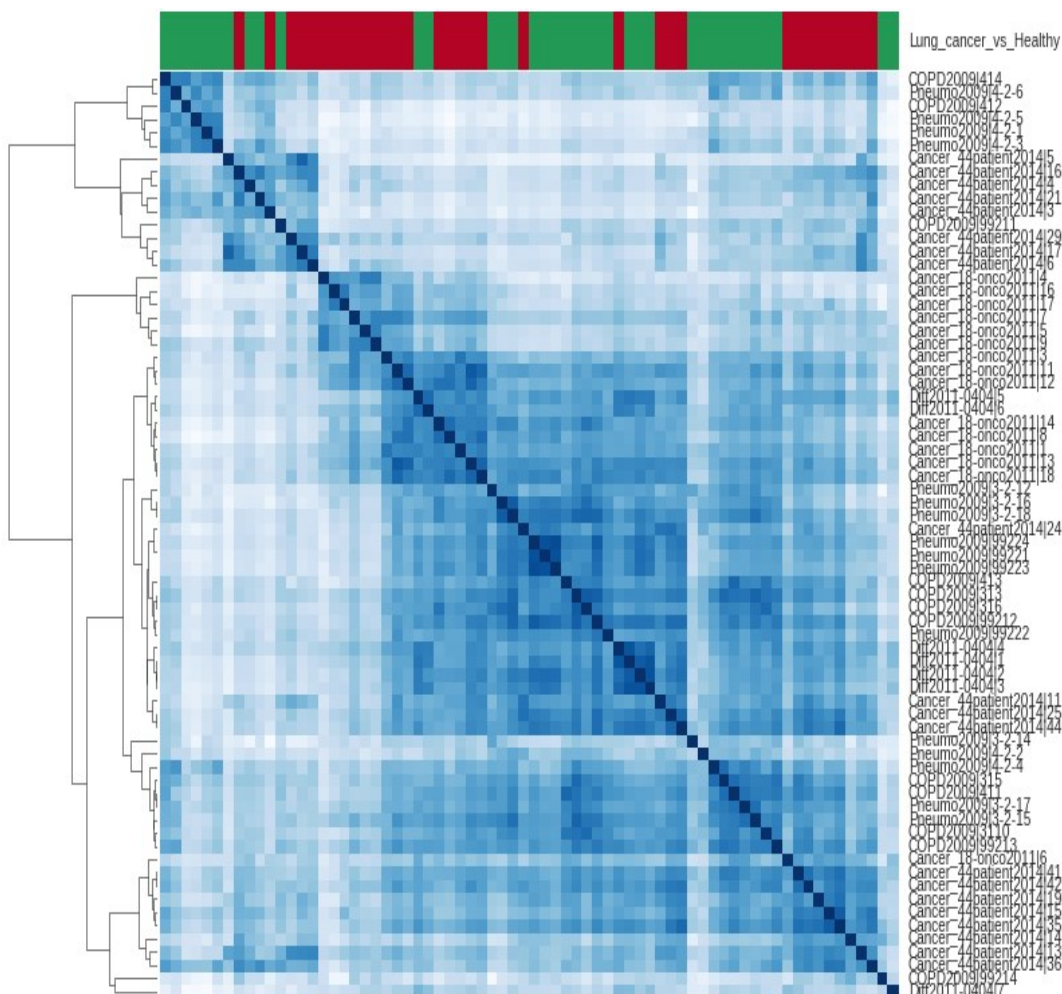


Рисунок 6. Иерархическая кластеризация проб доноров КВВ, включенных в анализ, на основании анализа идентифицированных пептидов. Красное – пробы доноров КВВ с диагнозом «рак легкого», зеленое – пробы доноров КВВ без злокачественных новообразований.

Также все пептиды были ранжированы на основании их встречаемости в пробах. Отдельный раздел главы посвящен анализу границ применимости программы Mascot при идентификации белкового и пептидного состава КВВ и приведено сравнение программ Mascot, X!Tandem и SEQUEST.

Для построения аналитической модели, основанной на анализе идентифицированных в КВВ пептидов, были использованы методы, обычно применяемые к данным по уровням экспрессии генов. 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:

$$\text{Peptidefoldchange} = \frac{\text{mean}(\text{peptidescore} \in \text{cancer})}{\text{mean}(\text{peptidescore} \in \text{notcancer}) + 0.0001}$$

+0.0001 добавлялось, чтобы не делить на 0.

Созданная в исследовании предсказательная модель дает численные значения для проб по выделенным пептидам, которые относят пробу либо к раковой, либо к не раковой. (Рисунок 7)

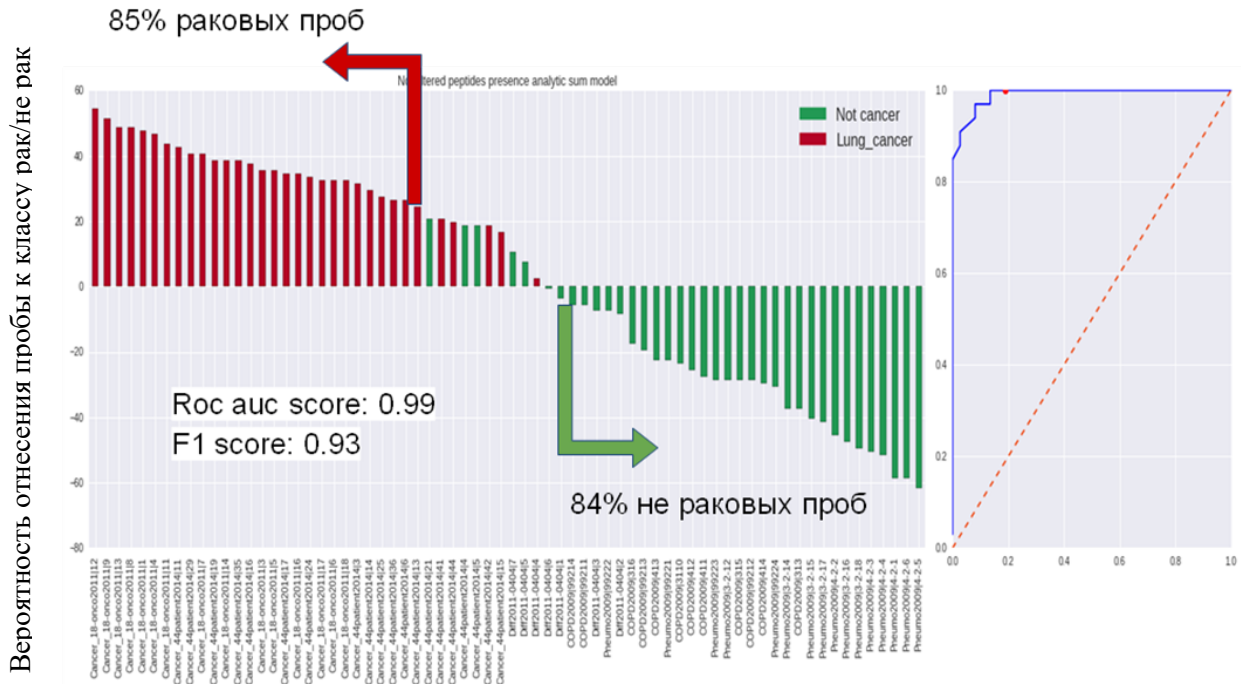
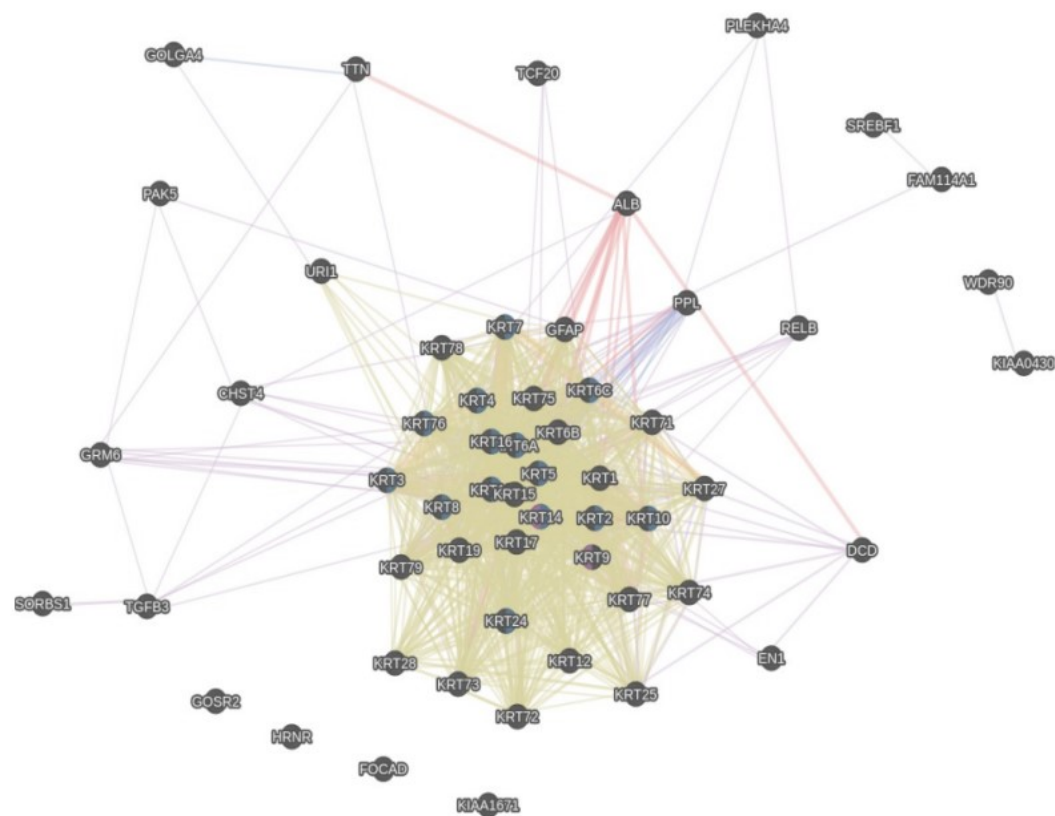


Рисунок 7. Красные пробы – раковые, зеленые пробы – не раковые. На графике приведены значения метрик машинного обучения предсказания: $F_1 = 0,93$; $Roc_auc = 0,99$.

Модель показала хорошую прогностическую способность ($AUC=0.99$), определив раковые образцы в группе КВВ, не включенных в машинное обучение.

Помимо описания построенной модели, в главе затрагивается важный вопрос возврата от пептидов к белковому уровню, что может быть критичным для внедрения метода в клиническую практику, а также в свете сопоставления полученных результатов белкового анализа проб. Благодаря тому, что пептиды триптические, такой возврат становится возможным. Принадлежность пептидов белкам анализировали с помощью сервиса Protein Information Resource (PIR), пептиды аннотировали по UniRef100 базе человеческих белков (сборка UniProtKB release 2017_06). Для определения семейств идентифицированных по пептидам белков с помощью сервиса GeneMania 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). (Рисунок 8) Анализ пептидов, встречающихся во всех группах, дал много совпадений с опубликованным белковым «фоном» КВВ.



Networks

- Co-expression
- Physical Interactions
- Shared protein domains
- Predicted
- Co-localization

Functions

- intermediate filament cytoskeleton
- intermediate filament organization

Рисунок 8. Карта взаимосвязей белков, имеющих в своем составе не менее двух идентифицированных пептидов из списка пептидов, определяемых во всех группах доноров, построенная в программе GeneMania.

ВЫВОДЫ

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 (были идентифицированы продукты 54 функциональных кератиновых генов). 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 РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых журналах и главы монографий

1. 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 // *Clinical Chemistry and Laboratory Medicine*. — 2009. — Vol. 47, no. 6. — P. 706–712.

2. 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 // *Biotechnology Advances*. — 2009. — Vol. 27, no. 6. — P. 945–959.

3. 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. — № 1. — С. 284–288.

4. Varfolomeyev, S. The molecular pathway to personalized medicine / S. Varfolomeyev, Kurova V., **Fedorchenko K.** // *From Promises to Practice. Applications of Science and Technology in Food, Healthcare, Energy and Environment* – Edited by Serageldin I., Masood E. — *Bibliotheca Alexandrina Alexandria, Egypt*, 2010. — P. 255–262.

5. Курова, В. С. Структурный и каталитический полиморфизм ферментов человека. Современные потенциальные платформы биомедицинской диагностики / В. С. Курова, И. Н. Курочкин, Г. Р. Каламкарров, А. Е. Бугрова, **К. Ю. Федорченко**, С. Д. Варфоломеев // *Физическая химия биопроцессов*. — Издательская группа URSS Москва, 2014. — С. 673–724.

6. 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.

7. **Федорченко, К. Ю.** Ранняя диагностика рака легкого на основе анализа протеома конденсата выдыхаемого воздуха / **К. Ю. Федорченко**, А. М. Рябоконт, 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. — С. 112–120.

8. **Федорченко, К. Ю.** Влияние космического полёта на белковый состав конденсата выдыхаемого воздуха космонавтов / **К. Ю. Федорченко**, А. М. Рябоконт, А. С. Кононихин, С. И. Митрофанов, Е. А. Михантьева, А. И. Спасский, И. Р. Суходолов, И. А. Попов, А. В. Поляков, И. М. Ларина, Е. Н.

Николаев, С. Д. Варфоломеев // Известия Академии наук. Серия химическая. — 2016. — № 11. — С. 2745-2750

9. Kononikhin, A. S. Spaceflight induced changes in human proteome / A. S. Kononikhin, N. L. Starodubtseva, L. K. Pastushkova, D. N. Kashirina, **K. Yu. Fedorchenko**, A. G. Brhozovsky, I. A. Popov, I. M. Larina, E. N. Nikolaev // Expert Review of Proteomics. — 2017. — Vol. 14, no. 1. — P. 15–29.

10. 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.

Тезисы конференций:

Результаты работы были опубликованы в сборнике материалов IV съезда Российского общества биохимиков и молекулярных биологов (Новосибирск, 2008 г.); Альманахе клинической медицины по материалам III Троицкой конференции по медицинской физике и инновациям в медицине (Троицк, 2008 г.); спецвыпуске журнала *Pneumologie*, посвященном Международному конгрессу *Kongress der Deutschen gesellschaft fur pneumologie und beatmungsmedizin e.V.* (Германия, 2011); спецвыпуске журнала *European Respiratory Journal*, посвященном Международному конгрессу *Annual Congress European Respiratory Society* (Нидерланды, 2011); сборнике материалов Международного конгресса *13th World Congress of the Human Proteome Organization, the 7th EuPA annual conference and the 6th Spanish Proteomics Society Congress* (Испания, 2014); сборнике материалов Международной конференции *Biocatalysis-2015: Fundamentals and Applications* (Россия, 2015); сборнике трудов XXV Национального конгресса по болезням органов дыхания (Россия, 2015); сборнике материалов XI Международной научно-практической конференции "Пилотируемые полеты в космос" (Россия, 2015); спецвыпуске журнала *Pneumologie*, посвященном Международному конгрессу *Kongress der Deutschen gesellschaft fur pneumologie und beatmungsmedizin e.V.* (Германия, 2016); спецвыпуске журнала *Acta Naturae*, посвященном V Съезду Физиологов СНГ/V Съезду Биохимиков России (Россия, 2016); спецвыпуске журнала *Pneumologie*, посвященном Международному конгрессу *Kongress der Deutschen gesellschaft fur pneumologie und beatmungsmedizin e.V.* (Германия, 2017); спецвыпуске журнала *European Respiratory Journal*, посвященном Международному конгрессу *Annual Congress European Respiratory Society* (Италия, 2017); сборнике материалов Международного конгресса *16th World Congress of the Human Proteome Organization* (Ирландия, 2017); сборнике трудов XXVI Национального конгресса по болезням органов дыхания (Россия, 2017); сборнике материалов Международной конференции по персонализированной онкологии (Россия, 2017).