

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«КУРЧАТОВСКИЙ ИНСТИТУТ»



На правах рукописи

Власов Иван Николаевич

**ТРАНСКРИПТОМНЫЙ АНАЛИЗ МОНОНУКЛЕАРНЫХ ЛЕЙКОЦИТОВ
ПЕРИФЕРИЧЕСКОЙ КРОВИ ПАЦИЕНТОВ С ЛЕТАЛЬНЫМИ И
НЕЛЕТАЛЬНЫМИ ИСХОДАМИ COVID-19 ПРИ ИНФИЦИРОВАНИИ
20В И ДЕЛЬТА ВАРИАНТАМИ SARS-COV-2**

1.5.3. – Молекулярная биология

Диссертация
на соискание ученой степени
кандидата биологических наук

Научный руководитель:
д-р биол. наук, проф.
Сломинский Петр Андреевич

Москва, 2025

ОГЛАВЛЕНИЕ

1. ВВЕДЕНИЕ.....	4
2. ОБЗОР ЛИТЕРАТУРЫ.....	11
2.1 Высокопроизводительное секвенирование.....	11
2.2 Обработка результатов секвенирования.....	14
2.3 Коронавирусы в патофизиологии человека.....	34
2.4 Полнотранскриптомные исследования при COVID-19.....	50
3. МАТЕРИАЛЫ И МЕТОДЫ.....	79
3.1 Описание выборок пациентов.....	79
3.2 Изоляция первичных мононуклеарных лейкоцитов периферической крови.....	80
3.3 Выделение и секвенирование РНК.....	81
3.4 Обработка и биоинформатический анализ данных.....	87
3.5 Описание подтверждения на независимой выборке.....	89
4. РЕЗУЛЬТАТЫ.....	92
4.1 Транскриптомное профилирование в группе зараженных вариантом 20В.....	92
4.2 Транскриптомное профилирование в группе зараженных вариантом Дельта.....	100
4.3 Исследование липидных профилей в крови пациентов ОРПТ с COVID-19, вызванным заражением вариантами 20В и Дельта SARS-CoV-2.....	108
5. ОБСУЖДЕНИЕ.....	109
5.1 Сравнение разработанных пайплайнов.....	109

5.2 Обсуждение результатов, полученных в процессе транскриптомного анализа в группе зараженных 20В вариантом SARS-COV-2.....	112
5.3 Обсуждение результатов, полученных в процессе транскриптомного анализа в группе зараженных Дельта вариантом SARS-COV-2.....	118
5.4 Сравнительный анализ ДЭГ, выявленных в группах инфицированных 20В вариантом и Дельта вариантом SARS-CoV-2.....	128
6. ЗАКЛЮЧЕНИЕ.....	133
7. ВЫВОДЫ.....	135
8. СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.....	137
9. СПИСОК ЛИТЕРАТУРЫ.....	139

1. ВВЕДЕНИЕ

Актуальность темы исследования

COVID-19 — это инфекционное заболевание, вызываемое заражением коронавирусом SARS-CoV-2. В декабре 2019 года в Ухане, провинции Hubei, Китай, произошло несколько случаев заболевания ранее неизвестной респираторной инфекцией. Инфекция быстро распространилась сначала по всей провинции, а затем и по всей стране [1]. Вирусом, вызывающим эту инфекцию, оказался новый коронавирус, имеющий последовательность, на 79.6 % совпадающую с выявленным ранее вирусом атипичной пневмонии, также известным как **Severe acute respiratory syndrome-related coronavirus (SARS-CoV)** [2]. К марту 2020 года новый коронавирус настолько широко распространился по всему миру, что Всемирная организация здравоохранения (ВОЗ) официально назвала ситуацию «пандемией» [3]. За 6 лет в мире произошло более 700 миллионов подтвержденных случаев заражения SARS-CoV-2, при этом более 7 миллионов случаев закончились гибелью зараженных пациентов. На 2025 год пик пандемии COVID-19 миновал, но тем не менее, по данным ВОЗ, в мире все еще происходит более тысячи подтвержденных случаев COVID-19 каждую неделю [4].

Заболевание COVID-19 может сопровождаться множеством различных симптомов, самыми распространенными из которых являются повышение температуры, кашель, интенсивное чувство усталости и потеря чувства обоняния и вкуса. Большая доля случаев заболевания также сопровождается головными болями, невралгиями, болью в горле, а также кожными проявлениями и проявлениями со стороны желудочно-кишечного тракта [5]. Одними из самых опасных клинических проявлений COVID-19 являются поражения органов дыхания, в тяжелых случаях приводящие к неспособности пациента дышать без искусственной вентиляции и острому респираторному дистресс-синдрому [6].

За время пандемии COVID-19 вирус претерпел значительное число мутаций, приведших к возникновению множества функционально различных вариантов вируса. Многие из возникших вариантов были более трансмиссивны, более

вирулентны и более способны к избеганию ответа иммунной системы, чем оригинальный вариант «Ухань» [7]–[9]. Кроме того, многие вновь возникшие варианты также проявляли устойчивость к нейтрализации антителами к ранним вариантам и распространенным вакцинам, в связи с чем возникновение новых вариантов сопровождалось новыми волнами заболевания [10]–[12].

Вариант SARS-CoV-2 Дельта был впервые выявлен в марте 2021 года в Индии, и в течении нескольких месяцев он распространился по нескольким другим странам [7]. В мае 2021 года его трансмиссивность была оценена как не меньшая, чем трансмиссивность варианта Альфа [13]. С другой стороны, в вышедших позже исследованиях его трансмиссивность была оценена как более высокая, чем у варианта Альфа [14], [15]. К сентябрю 2021 года этот вариант стал наиболее распространенным среди вариантов SARS-CoV-2 [7]. Этот вариант ассоциирован со значительно более высокой вероятностью попадания в отделение реанимации и интенсивной терапии (ОРИТ), и повышенной смертностью [16]. Эффективность моноклональных антител, входящих в рекомендованные протоколы лечения, а также антител из крови вакцинированных и переболевших другими вариантами вируса значительно ниже против варианта Дельта, чем против варианта Альфа [11]. Вакцины BNT162b2 и ChAdOx1 nCoV-19 также значительно менее эффективны против варианта Дельта [11], [12]. Случаи заболевания COVID-19 в 2025 г. вызываются подвариантами варианта Омикрон [17], [18], характеризующегося высокой трансмиссивностью и способностью к иммунному избеганию, но не высокой вирулентностью. То, что вирус постоянно эволюционирует, требует постоянного продолжения изучения случаев COVID-19.

Одним из методов такого исследования является масштабное параллельное транскриптомное секвенирование, позволяющее идентифицировать изменения в экспрессии генов между различными биологическими состояниями. Выявление дифференциально экспрессирующихся генов в тканях пациентов, заболевание у которых привело к различным исходам, позволяет пролить свет на факторы, определяющие течение заболевания. Помимо этого, дифференциально экспрессирующиеся гены также представляют собой потенциальные мишени для

терапевтического вмешательства.

Для данного исследования также важно отметить, что далеко не всем существующим вариантам SARS-CoV-2 были присвоены буквенные коды. Как правило, все предшествующие первым ключевым вариантам (например, альфа), в число которых входит в том числе исследуемый в нашей работе вариант 20B, рассматривались как вариант «Ухань» - то есть оригинальный вариант, положивший начало эпидемии, несмотря на отличия в последовательности. Например, в работе [19], рассматривающей эпидемиологию SARS-CoV-2 на территории РФ в тот же период, что и работа на основании которой мы идентифицируем исследуемый нами вариант вируса [20], описываются присутствующие в популяции не VOC варианты, как «Ухань».

В связи с этим **целью** настоящей работы было изучение изменения экспрессии генов на уровне транскриптома в тканях пациентов с летальными и нелетальными исходами COVID-19, вызванного двумя различными вариантами вируса SARS-CoV-2 – вариантами 20B и Дельта. Для достижения этой цели были поставлены следующие задачи:

Задачи исследования

1. Провести сравнение профилей экспрессии генов в периферической крови пациентов отделения реанимации и интенсивной терапии с летальными и нелетальными исходами COVID-19, вызванного заражением 20B вариантом SARS-CoV-2.
2. Провести сравнение профилей экспрессии генов в периферической крови между группами пациентов отделения реанимации и интенсивной терапии с летальными и нелетальными исходами COVID-19, вызванного заражением Дельта вариантом SARS-CoV-2.
3. Исследовать роль дислипидемии в определении исхода заболевания на независимых выборках пациентов с COVID-19, вызванного заражением вариантами 20B и Дельта SARS-CoV-2.

Научная новизна работы

В настоящей работе проведен полнотранскриптомный анализ экспрессии в мононуклеарных лейкоцитах периферической крови в двух группах пациентов с COVID-19, и выявлены гены, дифференциально экспрессирующиеся между группами пациентов с летальными и нелетальными исходами. В группе пациентов с 20B вариантом SARS-CoV-2 была показана дифференциальная экспрессия генов, ассоциированных с процессами активации рецепторов липопротеинов низкой плотности и метаболизмом липидов. В группе пациентов с Дельта вариантом SARS-CoV-2 была показана дифференциальная экспрессия генов, ассоциированных с негативной регуляцией иммунного ответа, системой комплемента и негативной регуляцией вирусных процессов. Исследование на независимых выборках подтвердило значимую роль обмена липидов для определения исходов заболевания COVID-19, вызванного заболеванием 20B вариантом, но не Дельта вариантом SARS-CoV-2. На основании анализа обогащения терминами Gene Ontology Biological Processes были выявлены группы ключевых дифференциально экспрессирующихся генов – *PPARG*, *CD36*, *STAB1*, *ITGAV* и *ANXA2* в случае 20B варианта, и *ISG15*, *SERPING1*, *C1QA*, *C1QB*, и *VSIG4* в случае Дельта варианта.

Теоретическая и практическая значимость работы

1. Разработан биоинформатический подход, позволяющий выявить ключевые процессы, определяющие исход инфекции SARS-CoV-2.
2. Полнотранскриптомный анализ мононуклеарных лейкоцитов периферической крови между группами пациентов ОРИТ с летальными и нелетальными исходами COVID-19, вызванного заражением 20B вариантом SARS-CoV-2 показал значимость метаболизма липидов для определения исхода заболевания, вызванного этим вариантом SARS-CoV-2.
3. Полнотранскриптомный анализ мононуклеарных лейкоцитов периферической крови между группами пациентов ОРИТ с летальными и нелетальными исходами COVID-19, вызванного заражением Дельта вариантом SARS-CoV-2, подчеркнул

важную роль баланса негативной и позитивной регуляции иммунной системы в определении исхода заболевания, вызванного этим вариантом SARS-CoV-2.

4. Различие между профилями экспрессии в случае двух различных вариантов SARS-CoV-2 подчеркнуло значимость исследований специфических вариантов вируса и продемонстрировало вариант-обусловленную природу процессов, влияющих на определение исхода заболевания в различных случаях.

Методология и методы исследования

В ходе выполнения работы были использованы стандартные молекулярно-биологические и биоинформатические методы исследований. Было проведено секвенирование РНК, отбор дифференциально экспрессирующихся генов с дальнейшим анализом обогащения. Были построены и проанализированы сети генов и терминов, и на их основании были отобраны ключевые гены. Также было проведено исследование уровней липидов на независимых выборках пациентов с COVID-19.

Положения, выносимые на защиту

1. Используемый в работе подход позволяет выявлять ключевые значимо изменённые биологические процессы при совместном применении разных алгоритмов выявления дифференциальной экспрессии.
2. У пациентов с тяжелой формой COVID-19, вызванной 20B вариантом вируса SARS-CoV-2, в определении исхода заболевания играют роль процессы, связанные с активностью рецепторов липопротеинов низкой плотности.
3. Увеличение экспрессии генов *PPARG*, *CD36*, *STAB1*, *ITGAV* и *ANXA2* характерно для больных с тяжелой формой COVID-19, вызванной 20B вариантом вируса SARS-CoV-2.
4. При летальном и нелетальном исходе заболевания COVID-19, вызванного Дельта вариантом вируса SARS-CoV-2, различаются механизмы, связанные с регуляцией активации системы комплемента и ряда иммунных процессов. Понижение экспрессии генов *ISG15*, *SERPING1*, *CIQA*, *CIQB*, и *VSIG4* имеет место при

летальных исходах.

5. Метаболизм липопротеинов низкой плотности может играть роль в определении исхода тяжелых случаев заболевания COVID-19, вызванным 20B вариантом, но не Дельта вариантом SARS-CoV-2.

Личный вклад соискателя

Основные результаты работы получены автором лично или при его непосредственном участии.

Степень достоверности и апробация результатов исследования

Работа выполнена в соответствии с общепринятыми этическими и научными принципами. Выводы и основные положения, выносимые на защиту, обоснованы фактическим материалом, полученным в результате проведения экспериментов с использованием надежных методов исследования, а также анализа и интерпретации данных. Материалы диссертации были представлены на четырех научных конференциях: Международный Конгресс «VIII Съезд Вавиловского общества генетиков и селекционеров, посвященный 300-летию российской науки и высшей школы» (г. Саратов, 14–19 июня 2024 года), Курчатowski геномный форум, Международный форум природоподобных технологий (г. Москва, 21 – 22 октября 2024 года), II Конгресс Международного общества клинической физиологии и патологии (г. Москва, 13-15 июня 2025 года), научная конференция «Молекулярная генетика и генетические технологии» НИЦ «Курчатowski института» (г. Москва, Россия, 19 мая 2025 года).

Публикации

Основные результаты диссертационной работы представлены в 7 публикациях, в том числе в 3 статьях в рецензируемых научных журналах, индексируемых базами данных Web of Science, Scopus, РИНЦ, и в 4 тезисах докладов научных мероприятий.

Структура и объем диссертации

Диссертация состоит из следующих разделов: «Введение», «Обзор

литературы», «Материалы и методы», «Результаты», «Обсуждение», «Заключение», «Выводы», «Список сокращений и условных обозначений» и «Список литературы», который включает 299 источников. Работа изложена на 175 страницах, содержит 15 рисунков, 8 таблиц.

2. ОБЗОР ЛИТЕРАТУРЫ

2.1 Высокопроизводительное секвенирование

2.1.1 Общая характеристика высокопроизводительного секвенирования

Термин массовое параллельное секвенирование (МПС), известный в англоязычной литературе как *next generation sequencing* (NGS), используется для обозначения нескольких различных технологий и коммерческих платформ для секвенирования, объединенных несколькими общими чертами. Во всех случаях при МПС параллельно секвенируются миллионы коротких фрагментов ДНК. Результатом такого секвенирования становится большое количество последовательностей нуклеотидов (ридов), длины которых для большинства секвенирующих платформ не превышают 600 пар нуклеотидов (п.н.). После этого эти короткие последовательности картируются на референсный геном или *de novo* собираются в непрерывные участки (контиги). Секвенируемый материал может представлять собой целый геном, определенные его участки (например, экзоны) или полученные другим образом последовательности ДНК – например, продукты обратной транскрипции всего транскриптома. При этом каждая буква секвенируемой последовательности оказывается прочитана несколько раз. Глубиной покрытия называют среднее число прочтений каждой буквы [21].

Непосредственное секвенирование ридов осуществляется различными способами в зависимости от платформы. Среди принципов этого секвенирования следует выделить три — секвенирование посредством достраивания цепи (Illumina, PacBio и многие другие) [22], секвенирование посредством лигирования цепи (SOLiD) [23] и секвенирование посредством измерения концентрации ионов (IonTorrent) [24].

На данный момент разработано множество различных применений МПС, таких как определение последовательности и вариаций генома, картирование регуляторных областей генома, определение трехмерной организации генома и характеристика транскриптома [25].

Исторически первым применением МПС было определение

последовательности генома. МПС было изначально создано как значительно более дешевая и менее трудоемкая альтернатива методам секвенирования, использованным в проекте «Геном человека». Ранним вариантам МПС удалось снизить стоимость секвенирования п.н. в 100 раз в сравнении с ценой в проекте «Геном человека» [25], и с тех пор был достигнут прогресс в цене приблизительно на 5 порядков. Секвенирование для определения последовательности подразделяется на *de novo* секвенирование и ресеквенирование. В процессе *de novo* секвенирования риды объединяются в контиги на основании их перекрывающихся участков [26]. В процессе ресеквенирования определяется наилучшее положение ридов в ранее отсеквенированном референсном геноме [27]. Оба этих подхода используются для того, чтобы выявить структурные вариации в геноме, такие как однонуклеотидные замены, инсерции, делеции, дубликации, инверсии и транспозиции.

Другим применением МПС является картирование регуляторных областей генома. Выявление прямого контакта между транскрипционным фактором и ДНК *in vivo* может демонстрировать роль транскрипционного фактора в регуляции экспрессии генов поблизости от области связывания. Используя обработку формальдегидом для фиксации связей между ДНК и ДНК-связывающим белком, с последующим обогащением библиотеки для секвенирования при помощи преципитации белка со специфичными антителами позволяет определить такие области связывания в геноме [28]. Такой метод получил название секвенирование с иммунопреципитацией хроматина (ChIPseq).

Сходным образом с ChIPseq, фиксация присутствующих *in vivo* связей и близких контактов позволяет также определить, какие части генома, не находящиеся поблизости с точки зрения последовательности, находятся поблизости физически в трехмерной структуре ядра клетки. Такой подход позволяет достичь двух различных целей. Первая – это определение дистальных регуляторных участков, влияющих на регуляцию экспрессии генов, находящихся на большом расстоянии в линейной последовательности, но близко расположенных в трехмерной структуре ядра [29]. Такой метод получил название анализ

иммунопреципитации хроматина с парно-концевыми метками. Второй подход использует информацию о физической близости различных участков генома для воссоздания трехмерной структуры ядра и получил название Hi-C [30].

В данной работе мы более подробно остановимся на секвенировании совокупности белок-кодирующих генов транскриптома при помощи платформ Illumina.

2.1.2 Масштабное параллельное секвенирование РНК (RNA-seq)

Регуляция экспрессии генов, в особенности белок-кодирующих, играет роль во всех процессах, происходящих в организме, включая патогенез различных заболеваний. Исходя из этого факта, выявление дифференциальной экспрессии – значимого различия между экспрессией одно и того же гена в двух различных биологических состояниях – представляет собой важный инструмент для исследования любых биологических процессов. RNA-seq позволяет протестировать дифференциальную экспрессию для тысяч генов одновременно, что делает его незаменимым инструментом при исследовании любых биологических процессов [31].

Первым этапом RNA-seq является получение интересующей исследователя фракции РНК. Известно, что рибосомальная РНК (рРНК) представлена в наибольшем количестве в клетке. Так как количественный анализ рРНК не представляет интереса в большинстве экспериментов, ее необходимо удалить из выделяемой фракции либо путем специфичного удаления [32], [33], либо путем специфичного выделения целевой фракции [34], [35].

Большинство белок-кодирующих генов млекопитающих несут поли-А концы, что позволяет относительно просто выделить их из тотальной РНК при помощи гибридизации с биотинилированными олиго-Т зондами или 3' концевое обогащения с использованием олиго-Т праймеров в процессе обратной транскрипции. В тех случаях, когда целевые РНК не несут поли-А хвостов (например, микроРНК), удаление рРНК может производиться на основании ее последовательности с использованием биотинилированных зондов или зондо-

направленного расщепления. Другим подходом может являться ПЦР-основанное обогащение других фракций при помощи так называемых not-so-random праймеров, которые включают в себя последовательности, встречающиеся только в целевой фракции РНК [36]. Помимо методов, основанных на последовательности, также разработаны методы, использующие другие свойства рРНК, такие как ее высокая представленность (метод C₀T гибридизации) или кэпирование 5' монофосфатом [32].

После выделения необходимой фракции РНК следующим шагом является измельчение РНК и проведение обратной транскрипции или проведение обратной транскрипции на целых РНК с последующим измельчением. В случае если измельчение происходит на уровне РНК, оно может осуществляться при помощи основных растворов, растворов дивалентных катионов или РНКазы-III. В случае если измельчение происходит на уровне кДНК, оно может осуществляться при помощи ультразвукового измельчения, транспозонного тагментирования или измельчения ДНКазы [37].

Полученные кДНК нужного размера лигируются с адаптерами, необходимыми для амплификации и последующего секвенирования. В некоторых протоколах в процессе получения кДНК к ним также добавляются случайные или имеющие predetermined последовательности молекулярные ярлыки (molecular labels), которые позволяют избежать ошибки, связанной с неравномерной амплификацией кДНК. После этого кДНК с адаптерами и, возможно, молекулярными ярлыками, амплифицируются при помощи 8-12 циклов ПЦР, что позволяет поднять количество кДНК до необходимого для секвенирования минимума. Полученные кДНК, называемые библиотеками, затем секвенируются, и обрабатываются дальше в виде ридов [31].

2.2 Обработка результатов секвенирования

2.2.1 Качество и удаление низкокачественных оснований

Отсеквенированные библиотеки выводятся в форматах fastq или fasta, в которых указывается последовательность каждого полученного рида, и, в случае формата fastq, «качество» каждого нуклеотида в рида. На данный момент

существует три различных способа кодировки качества в зависимости от платформы, однако все они так или иначе основаны на логарифмической трансформации оценки вероятности ошибки. Например, в случае современных программ для обработки сиквенсов Illumina качество (Q_{PHRED}) записывается как

$$Q_{PHRED} = -10 \times \log_{10}(P_e), \text{ где } P_e - \text{ оценка вероятности ошибки;}$$

с последующим добавлением натуральной величины (33 или 64) для простоты кодирования в формате ASCII [38].

$$PHRED_{33} = Q_{PHRED} + 33$$

$$PHRED_{64} = Q_{PHRED} + 64$$

Учитывая доступность информации о вероятности ошибки, следующим шагом может быть удаление оснований с высокими вероятностями ошибок. В обработке библиотек для выявления полиморфизмов или определения последовательности генома такой шаг является необходимым, однако в случае RNA-seq, где определение точной последовательности генов и транскриптов не является основной целью, существуют аргументы за и против применения этого шага. С одной стороны, на примере симулированных данных продемонстрировано, что удаление оснований низкого качества снижает общее количество неправильно определённых оснований в последовательностях транскриптов [39]. В тоже время в другом исследовании с использованием симулированных данных было показано, что удаление оснований низкого качества приводит к изменению итоговых оценок экспрессии генов [40]. Также, на примере данных SEQC [41] продемонстрировано, что удаление адаптеров и оснований низкого качества не приводит к улучшению результатов количественной оценки экспрессии генов [42], хотя последнее может, в том числе, зависеть от специфического алгоритма выравнивания. Таким образом, можно сделать вывод, что прежде, чем принять решение о включении этого шага в процесс обработки, необходимо проверять эффект удаления оснований низкого качества на каждый разработанный подход для квантификации экспрессионных данных.

2.2.2 Картирование

В случае хорошо изученных объектов исследования, таких как человек, мышь или крыса, для квантификации относительных уровней транскрипта полученные сиквенсы выравниваются на существующий референсный геном или транскриптом, чтобы определить их принадлежность к определенным генам и транскриптам.

Задача картирования заключается в поиске наилучшего выравнивания для каждого рида в геноме или транскриптом. Для того чтобы рассмотреть проблему картирования рида на транскриптом, необходимо сначала рассмотреть решение проблемы картирования рида в случае генома. Наиболее простым подходом к картированию рида будет просто протестировать рид в каждой возможной позиции его в геноме и оценить общее число совпадений в последовательности для каждой позиции. Проблема такого подхода заключается в том, что он крайне вычислительно сложный, т.е. его исполнение займет очень большое число операций. Число операций, нужное для выполнения такого алгоритма, пропорционально произведению размера генома, длины рида и числа ридов, что слишком много для любого практического применения [43]. Вследствие этого, в рамках любого практически применимого метода картирования необходимо сузить область поиска потенциальных позиций для ридов. Например, возможно однократно выявить все уникальные фрагменты, равные по длине риду, и имплементировать быстрый алгоритм поиска по всем позициям (например, хэш-поиск) для выявления совпадений. Однако, такой метод потребовал бы слишком много оперативной памяти/места на жестком диске, так как число уникальных последовательностей более 30 п.н. в геноме крайне высоко. Логичным следующим шагом в таком случае будет алгоритм, называемый «seed and extend» – в этом случае вместо того, чтобы картировать положение всех последовательностей, равных по длине риду, картируется положение всех более коротких последовательностей – например, длиной в 8 нуклеотидов (т. н. k-мер). Так как число возможных k-меров длиной 8 нуклеотидов равно 65536, в то время как число возможных k-меров длиной 30 нуклеотидов составляет приблизительно 10^{19} , такой подход позволяет

значительно уменьшить максимальный требуемый размер оперативной памяти. После картирования положения всех k -меров длиной 8 нуклеотидов в геноме при совпадении участка ридов с такой последовательностью остается только проверить участок контекста в геноме/транскриптоме, окружающий такую последовательность. Такой алгоритм картирования имплементирован, например, в инструментах SSAHA [44] и Stampy [45]. Однако такой подход чреват проблемами в случаях, когда возможны неполные совпадения между последовательностью ридов и генома (миссмапчи), так как учет ошибок в первой фазе картирования был бы затруднителен. В случае, если мы, например, позволим каждому k -меру содержать одну замену – это значительно увеличит встречаемость такого k -мера в геноме. Дальнейший рост числа допустимых миссмапчей постепенно сведет на нет вычислительное преимущество этого метода. Другой проблемой такого подхода является то, что представленность k -меров в геноме неравномерна – определенные последовательности встречаются чаще, чем другие, некоторые значительно чаще, чем другие [43].

Для решения этих проблем были разработаны методы, основанные на массивах и деревьях суффиксов. Суффикс – это подпоследовательность строки, конец которой совпадает с концом строки. Дерево суффиксов – это совокупность всех суффиксов, присутствующих в геноме, где корень дерева – это конец строки, и каждый путь от корня к листу дерева отражает один из суффиксов. Большим преимуществом такой формы данных является то, что часто встречающиеся последовательности в такой структуре данных представлены очень компактно, и их уникальный контекст, позволяющий отличить различные случаи их появления в геноме/транскриптоме, легко доступен при поиске. Также поиск по дереву суффиксов – это вычислительно простая операция. Недостатком такого подхода будет то, что включение любых миссмапчей приводит к экспоненциальному росту размера дерева суффиксов. Инструмент MPscan имплементирует такой подход к картированию, однако он работает только для картирования, не включающего миссмапчи [43].

Массив суффиксов – это совокупность всех суффиксов, присутствующих в

геноме/транскриптом, отсортированных в лексикографическом порядке. Очевидно, что для генома/транскриптома такой массив будет требовать исключительно много оперативной памяти/места на диске. Однако эта проблема может быть решена, если только первый нуклеотид и его позиция в геноме будет сохранена для каждого суффикса, а затем при необходимости этот суффикс будет получен исходя из этой информации и последовательности референсного генома. Другой проблемой для массива суффиксов является то, что поиск по массиву является отдельной задачей, требующей решения. Одним из самых распространенных среди инструментов для картирования методов поиска по массиву суффиксов является трансформация Буроуса-Вилера (Burrows–Wheeler transform, BWT) [46]. Этот метод включает запись наиболее длинного суффикса, начинающегося с каждой позиции генома в форме цикла, где после последнего нуклеотида идет символ \$, а затем первый нуклеотид. После этого сохраняется только последний нуклеотид для каждого суффикса, и такой набор букв для всех позиций в геноме называется BWT этого генома. Вместе с описанной выше позицией такая репрезентация генома становится очень компактной. При этом свойства этой трансформации таковы, что конкатенация BWT с лексикографически отсортированной версией BWT представляет собой массив, включающий все подстроки из 2 букв, встречающиеся в геноме. В свою очередь конкатенация BWT с лексикографически отсортированным массивом всех подстрок из двух букв представляет собой массив всех подстрок из 3 букв, встречающихся в геноме. Повторяя эту процедуру, возможно полностью воссоздать весь массив суффиксов из только двух этих значений – позиции начала суффикса в геноме и BWT. Это удобная и очень компактная репрезентация целого генома, однако воссоздавать весь массив суффиксов, чтобы найти в нем интересующий ряд вычислительно не эффективно, т. к. потребует множества циклов лексикографической сортировки. Для того, чтобы извлечь из BWT последовательность ряда буква за буквой, был предложен алгоритм Феррагина-Манзини [47]. Алгоритм также использует только BWT и лексикографически отсортированный BWT для поиска рядов. Последовательности рядов выявляются справа налево – то есть с последнего

нуклеотида рида. На первом шаге алгоритма определяются все позиции нуклеотидов, совпадающих с последним нуклеотидом рида в отсортированном BWT. Затем среди этих позиций отбираются те позиции, в которых в неотсортированном BWT находятся нуклеотиды, совпадающие со вторым с конца нуклеотидом в риде. Далее выявляется, какой позиции в отсортированном BWT соответствуют выявленные на предыдущем шаге нуклеотиды, и эта процедура повторяется до тех пор, пока не будет определена позиция начала рида в BWT. Эта процедура может быть реализована крайне компьютерно эффективно, что позволяет определять положение огромного числа различных ридов с большой скоростью. Данный подход используется в основе многих из самых популярных инструментов для геномного выравнивания, включая Bowtie-2 [48], BWA/BWA-SW [46] и SOAP2 [49]. Этот подход часто называется FM-индексом, сокращенно от Full-text index in Minute space.

Описанные выше подходы подходят для поиска положения рида, идеально совпадающего с участком последовательности референсного генома. Проблема картирования ридов с заменами слишком обширна сама по себе, чтобы осветить ее здесь, но вкратце этот подход требует вычисления «качества» картирования рида, обратно пропорциональному заменам и инсерциям/делециям. Как правило, это происходит за счет суммирования «весов» различных ошибок, отличающих рид от референсного генома в его различных положениях.

Следует отметить, что методы выравнивания, использующиеся для полногеномного выравнивания, как правило, не подходят для выравнивания транскриптома в силу специфики строения транскриптов. Например, риды могут попасть на точку соединения экзонов, что является ожидаемой ситуацией при картировании транскриптома и необычной ситуацией, указывающей на присутствие перестройки генома или ошибку выравнивания при картировании генома. [50]. Такие ситуации, как правило, называются «гэп»-ами. Качество выравнивания рида, находящегося на стыке экзонов, окружающих интрон длиной в 100 п.н. будет значительно снижено, несмотря на то что это потенциально наилучшее положение для этого рида в транскрипте. На данный момент

разработано несколько методов решения этой проблемы, которые имплементированы во множестве различных инструментов картирования транскриптома. Далее некоторые из них будут рассмотрены.

TopHat – это один из самых хорошо известных и часто используемых методов картирования транскриптома. TopHat дополняет работу Bowtie, разбивая риды, которые не выравниваются при помощи Bowtie, предположительно по причине их попадания в гэп, на меньшие фрагменты для дальнейшего выравнивания. Если эти фрагменты выравниваются на значительном геномном расстоянии друг от друга, TopHat идентифицирует такой рид как находящийся на соединении экзонов. Преимуществом этого подхода является возможность определения точек соединения экзонов без априорной информации о их расположении [50], [51]. Однако, независимые исследования показали, что TopHat и сопряженные с ним методы выявления дифференциальной экспрессии уступают другим современным методам картирования транскриптома и определения дифференциальной экспрессии [52], [53].

HISAT2 – более современный метод картирования транскриптома, основанный на модификации FM-index-а, называемой graph FM-index (GFM-index). Конструкция этого индекса начинается с построения линейного графа референсного генома, где каждый узел представляет собой один из нуклеотидов референсного генома. Затем к этому графу добавляются пути, представляющие собой замены, инсерции и делеции (до 20 п.н. длиной). Путь в таком графе будет представлять собой последовательность нуклеотидов референсного генома или одного из его вариантов. Затем этот граф сортируется по методу, называемому “индексной сортировкой” [54]. Таблица ребер такого отсортированного графа представляет собой чрезвычайно компактный индекс, при помощи которого возможно определить положение любой последовательности рида. Помимо этого глобального индекса, HISAT2 также создает множество малых индексов для областей ~ 57 т.п.н., облегчающих и ускоряющих поиск возможных гэпов и соединений экзонов, сужая возможную область поиска. Также преимущество такого малого индекса заключается в том, что он помещается в кэш памяти

процессора, обращение к которой быстрее, чем к оперативной памяти. Так как HISAT2 изначально был разработан для того, чтобы выравнивать в том числе транскриптомы, этот алгоритм не использует специфические процедуры для выравнивания ридов, попадающих на гэпы. Вместо этого HISAT2 использует подобранные веса, придаваемые определенным «ошибкам» выравнивания, таким как гэпы, так чтобы выравнивание все еще было возможно при попадании рида на стык экзонов. Влияние очень длинной делеции в риде в такой системе на оценку его качества выравнивания будет меньше, чем в любом алгоритме, описанном ранее [54].

STAR – другой широко используемый метод картирования транскриптома. Основная концепция STAR заключается в использовании Наибольшего Картируемого Префикса (НКП) – самой длинной подстроки картируемого рида, для которой находится идеальное совпадение по последовательности в геноме. После выявления такой подстроки и места ее картирования процесс повторяется для некартированных частей рида, что отличает STAR от алгоритмов, основанных на сходных принципах. Такой подход позволяет этому алгоритму автоматически выявлять точки стыка экзонов и миссмапчи. Непосредственное картирование подстрок осуществляется при помощи суффиксного массива [55]. После того, как все НКП выявлены, они сшиваются в контиги с использованием весов для совпадений, миссмапчей, инсерций, делеций и стыков экзонов. Этот шаг сходен с подходом, который используется в алгоритмах геномного картирования, с тем исключением, что вместо целых ридов используются их фрагменты [56].

Использование фрагментов ридов при картировании транскриптома является распространенной тенденцией, которую лучше всего иллюстрирует алгоритм Rsubread. Этот алгоритм разбивает каждый картируемый рид на несколько подридов, каждый из которых картируется отдельно. После картирования подриды «голосуют» за правильную позицию рида в картируемом транскриптом. Такой дизайн также позволяет автоматически определять инсерции, делеции и стыки экзонов [57].

На данный момент разработано множество инструментов для выравнивая

транскриптома, использующих множество различных подходов. К сожалению, на данный момент не существует по-настоящему всеобъемлющих сравнений между этими методами, что делает проблему выбора между ними затруднительной. Поэтому для того, чтобы сделать выбор, основанный на объективных метриках, необходимо проводить сравнения между интересующими методами самостоятельно. Первым подходом к такому сравнению может служить использование симулированных транскриптомных данных. Такой сет данных позволит сравнить информацию о каунтах для генов, полученных при помощи алгоритма картирования, с истинной информацией о относительном количестве каждого из транскриптов. Проблемой в работе с симулированными данными является то, что симулированный датасет может не отражать особенностей, присущих биологическим данным. Вторым подходом может быть использование данных ПЦР с обратной транскрипцией (Reverse transcription polymerase chain reaction, RT-PCR) для верификации данных, полученных при помощи метода секвенирования транскриптома с использованием различных инструментов для картирования. Проблемой этого подхода является высокая цена получения таких данных. На данный момент в общем доступе выложен датасет, включающий данные секвенирования транскриптома, а также более 10 тыс. генов, данные о дифференциальной экспрессии которых получены при помощи RT-PCR [41]. Несмотря на важность этого датасета как инструмента для сравнения, некоторые технические особенности делают интерпретацию результатов, полученных при помощи его использования, затруднительными. Другой проблемой такого подхода является то, что он позволяет протестировать только алгоритм обработки секвенирования транскриптома в целом, а не его индивидуальные шаги. Существует возможность сравнения различных алгоритмов картирования и их влияния на точность получаемых данных о дифференциальной экспрессии «при прочих равных» – то есть, когда остальные переменные шаги в процедуре анализа секвенирования транскриптома находятся в одинаковом состоянии, и только инструмент для картирования варьируется. Использование такого подхода на данных SEQC в сочетании с использованием симулированных данных

представляется наилучшим подходом к тестированию алгоритмов картирования, а также других шагов обработки данных секвенирования транскриптома.

2.2.3 Риды, картирующиеся в несколько возможных регионов

Одной из проблем картирования является картирование мультиридов – ридов, равновероятно картирующихся в несколько геномных регионов [58]. Основными причинами появления мультиридов, как правило, является то, что нуклеотидная последовательность рида происходит из области генома или транскриптома, для которых существует несколько других геномных областей с высокосходными последовательностями. Такими областями могут быть гены, имеющие несколько паралогов и/или области низкой сложности (богатые повторами) и/или различные изоформы одного и того же транскрипта [59]. Многие из таких последовательностей могут быть абсолютно идентичными или нести небольшие различия в последовательности. Однако зачастую даже эти малые различия недостаточны для того, чтобы определить истинное положение рида только на основании его последовательности из-за существования полиморфизмов, а также ошибок в референсном или секвенируемом геноме.

Учитывая, что мультириды составляют значительную (от 17% до 52 % в зависимости от вида) часть всех ридов, различные подходы к картированию мультиридов могут иметь значимые эффекты на оценку дифференциальной экспрессии [59]. Для разрешения этой проблемы было разработано несколько различных подходов, которые будут рассмотрены далее.

Исторически первым и наиболее простым подходом является удаление всех мультиридов [60], [61]. Такой подход неизбежно приводит к тому, что экспрессия генов, имеющих паралоги или различные транскрипционные изоформы, будет занижена. Для того чтобы частично избежать этой проблемы, возможно также внести поправку на возможность уникального картирования последовательности генов и транскриптов исходя из уникальности их последовательности в референсном геноме [62].

Более современные подходы к решению этой проблемы включают в себя

распределение мультиридов при помощи информации о распределении уникально картирующихся ридов. В наиболее простой имплементации риды просто распределяются по возможным экзонам/транскриптам пропорционально тому, сколько такие экзоны/транскрипты несут уникально картирующихся ридов [58], [63]. Несмотря на простоту этого подхода, получаемые таким образом результаты лучше согласуются с результатами микроарреев, чем если бы мультириды были просто удалены [58]. Более сложная имплементация этой идеи представляет собой создание генеративной модели распределения ридов с параметрами, соответствующими уровням экспрессии, распределению ридов и ошибке секвенирования. Эти параметры определяются на основании распределения уникально картирующихся ридов методом Expectation-maximization [64]. После этого генеративная модель используется для того, чтобы распределить мультириды по методу максимального правдоподобия. Такой подход получил название RSEM. Согласно тестированию на симулированных данных, оценки экспрессии, использующие данные, получаемые при помощи RSEM, точнее отражают истинные изменения экспрессии чем оценки, использующие данные, получаемые при помощи удаления или пропорционального распределения мультиридов [64].

2.2.4 Нормализация

После картирования ридов на геном или транскриптом, следующим шагом в анализе данных RNA-seq является подсчет общего числа ридов, картированных на область, принадлежащую определенному транскрипту или гену. Это число называется «каунты» и считается мерой экспрессии гена. Однако, прежде чем использовать каунты для выявления дифференциальной экспрессии, необходимо внести поправки, чтобы избежать некоторых систематических ошибок.

В каждой библиотеке RNA-seq находится определенное число (как правило, от 10 до 50 млн.) ридов. Общее число ридов для каждого гена зависит от общего числа ридов в библиотеке (размера библиотеки), из-за чего вариация в размере библиотеки может приводить к вариации в каунтах генов [65]. Так, если в двух

библиотеках, полученных из одного и того же образца, было отсекавлено 15 млн. ридов в первой библиотеке, а во второй библиотеке было отсекавлено 30 млн. ридов – соотношение каунтов между ними будет в среднем составлять 1 к 2. Очевидно, с точки зрения анализа транскриптома это ошибка. Для того, чтобы избежать подобных ошибок, было разработано несколько различных методов нормализации каунтов: нормализация на общее число каунтов, на медиану или верхний квартиль числа каунтов, quantile, DESeq и TMM.

При нормализации на сумму всех каунтов (ТС – total counts) каунты для каждого гена или транскрипта делятся на общую сумму всех каунтов (размер) в библиотеке. Сами по себе, нормализованные таким образом каунты не используются в дальнейшем в расчётах, так как они всегда крайне малы. Как правило, для облегчения восприятия данных они умножаются на какое-то большое постоянное число. В двух самых распространенных методах нормализации на общее число каунтов – CPM (counts per million) и RPKM (reads per kilobase per million) – таким числом является миллион [58]. При использовании RPKM дополнительно производится нормализация на длину гена, то есть каунты для каждого гена, уже нормализованные на размер библиотеки, делятся на длину гена и умножаются на тысячу.

Основной проблемой нормализации на размер библиотеки является то, что наличие или отсутствие высоко экспрессируемых генов после нормализации на размер библиотеки может влиять на расчет экспрессии других генов и приводить к ложно положительным результатам при выявлении дифференциальной экспрессии [65]. Методы, описанные ниже, лишены этого недостатка, так как они либо не принимают самые высоко экспрессируемые гены в расчет фактора нормализации, либо имеют отдельные факторы нормализации для разных групп генов.

Методы median и upper quartile нормализуют каунты на медиану и верхний квартиль каунтов по каждому гену. В обоих случаях для подсчета этих величин используются только каунты по генам/транскриптам, отличные от нуля хотя бы в одной библиотеке [66].

Тогда как методы CPM, RPKM, median и upper quartile применяются для

каждой библиотеки ридов отдельно, все методы, описанные ниже, применимы только для группы библиотек. Такой группой могут являться, например, все библиотеки, полученные в рамках одного эксперимента.

Метод *quantile normalization* использует проекцию в измерениях квантилей для того, чтобы привести распределение каунтов во всех образцах в группе к единообразию. Изначально этот метод был разработан Volstad и коллегами [67] для микроарреев, и позже был адаптирован для RNA-seq в пакете программ «limma» [68].

Метод DESeq из одноименного пакета для обработки данных RNA-seq нормализует каждую библиотеку на специальный фактор, который рассчитывается исходя из предположения что большинство генов не дифференциально экспрессированы, и в среднем соотношение каунтов по ним в двух образцах должны составлять 1 к 1. Этот фактор рассчитывается как медиана отношений каунтов каждого гена к геометрическому среднему каунтов этого гена во всех образцах в группе [69].

Другой метод, основанный на этом же предположении, называется Trimmed Mean of M-values (TMM). Для понимания этого метода необходимо ввести понятие «кратное соотношение» (fold change, FC). Кратным соотношением называется отношение числа ридов в гене/транскрипте в экспериментальном образце к числу ридов в этом же гене/транскрипте в референсном образце. Для того чтобы рассчитать фактор нормализации в методе TMM, одна из библиотек берется в качестве референса, после чего рассчитывается FC для всех генов во всех остальных библиотеках. После этого для каждого из остальных образцов рассчитывается взвешенное среднее логарифмированных FC (logFC) по всем генам, которое и является фактором нормализации. Из этого расчета исключаются все наиболее экспрессируемые гены и все гены с наибольшими различиями между образцами [70].

При сравнении всех этих методов на симулированных данных было продемонстрировано, что выявление дифференциальной экспрессии с использованием всех этих методов обладает сходной мощностью, однако число

ложноположительных результатов при включении высокоэкспрессируемых генов в анализируемые данные наиболее низко при использовании методов DESeq и TMM [65].

2.2.5 Выявление дифференциальной экспрессии

Полученные на предыдущем этапе нормализованные каунты считаются мерой экспрессии генов/транскриптов, к которым они относятся. Наиболее простым и очевидным следующим шагом может служить вычисление FC по каждому гену и определение значимых различий между контрольной и экспериментальной группами при помощи статистических тестов, таких как T-Стьюдента или U-Манна-Уитни. Однако, такой подход чреват несколькими проблемами, специфичными для секвенирования транскриптома.

Первые две проблемы вытекают из специфических особенностей секвенирования транскриптома. Каунты, очевидно, представляют собой положительные целые числа, что само по себе исключает гипотезу о нормальности их распределения. Следовательно, все параметрические тесты, делающие предположение о нормальности распределения, не могут быть использованы.

Само по себе это не исключает возможности использования непараметрических тестов. Однако в силу другой проблемы – как правило, малого числа биологических и технических повторов в экспериментах по секвенированию транскриптома, мощность таких тестов может быть недостаточной, особенно учитывая большое число проверяемых независимых гипотез (равное числу тестируемых генов/транскриптов) и вытекающую из этого необходимость в поправках на множественное тестирование. Для того чтобы преодолеть эти проблемы, необходимо создать корректную модель распределения каунтов и использовать статистические тесты, подходящие для этой модели. Создание корректной модели требует выбора подходящего распределения и подбора хорошо описывающих распределение каунтов параметров – как правило, математическое ожидание и дисперсию. Малое число повторностей может привести к тому, что подобранные параметры будут далеки от того, чтобы точно описывать имеющиеся

данные. На данный момент разработаны подходы для того, чтобы частично скомпенсировать малое число повторов в экспериментах по секвенированию транскриптома. Идея таких подходов заключается в том, что возможно сделать некоторые предположения о распределении каунтов, в частности, о постоянности взаимоотношения математического ожидания и дисперсии, что позволит использовать информацию о распределении каунтов по другим генам для более корректного моделирования каунтов по каждому гену [71].

Наиболее подходящими распределениями для моделирования каунтов могут служить распределение Пуассона [72] и негативно-биномиальное (NB) распределение [73]. Необходимо отметить, что каунты подвержены двум типам вариации – технической и биологической. Биологическая вариация происходит от различий между исследуемыми образцами и особями, тогда как техническая возникает под воздействием случайных факторов в процессе пробоподготовки и сиквенса. Следует отметить, что, тогда как существует возможность свести биологическую вариацию в эксперименте на нет – например, сравнивая две группы состоящие исключительно из технических повторов, то техническая вариация присутствует абсолютно во всех экспериментах. В случае, если единственным источником вариации является техническая вариация, в таком эксперименте распределение Пуассона подходит для моделирования каунтов. Однако в большинстве экспериментов в той или иной степени используются биологические повторы. Было показано, что в экспериментах с биологическими повторами в группах наблюдается так называемая «овердисперсия» [74]. Овердисперсией называется ситуация, когда наблюдаемая вариабельность выходит за пределы ожидаемой по модели распределения Пуассона. Важно отметить, что распределение Пуассона предполагает жесткую взаимосвязь между значением математического ожидания и дисперсии. В таких случаях NB распределение является наиболее подходящим. Для того, чтобы понять почему NB распределение больше подходит для моделирования каунтов в экспериментах с «овердисперсией», требуется рассмотреть вариабельность в смешанной Гамма-Пуассон интерпретации NB распределения. В такой интерпретации общая вариабельность

будет представлять собой $var = \mu + \mu^2 \times \phi$, где « μ » и первое слагаемое представляет собой техническую вариабельность, аналогичную дисперсии в распределении Пуассона, а второе слагаемое – биологическую вариабельность, аналогичное дисперсии в Гамма распределении. Во втором слагаемом, « ϕ », аналогичное параметру Гамма распределения « k », задающему форму кривой, представляет собой отношение вариации экспрессии гена к среднему экспрессии гена [75]. Если, как упомянуто выше, сделать предположение о том, что соотношение среднего и дисперсии каунтов по всем генам является постоянным, то это дает возможность использовать информацию не только о нескольких повторностях каунтов по гену в группе, но и о всех других многочисленных исследуемых генах для того, чтобы оценить вариабельность каунтов. Кроме того, разложение вариабельности на эти два слагаемых позволяет более точно моделировать вариабельность каунтов в экспериментах с биологическими повторами, и, следовательно, овердисперсией. В силу этих причин NB распределение используется для моделирования каунтов в нескольких из самых широко используемых и эффективных методов выявления дифференциальной экспрессии. В рамках этого обзора мы рассмотрим два наиболее широко известных и часто используемых из таких методов – edgeR и DESeq2.

Пакет для программной среды R “edgeR”, изначально был разработан для серийного анализа экспрессии генов (SAGE) [76]. Тем не менее, этот подход и связанное с ним программное обеспечение подходит для работы широкого круга задач, связанных с анализом каунтов. К таким задачам относятся RNA-seq [60], CHIP-seq, подсчет видов при баркодировании и оценка количества белка в протеомных исследованиях [77]. Каунты по каждому гену/транскрипту в edgeR моделируются как негативно биномиально распределенные случайные величины. Дисперсия каунтов по каждому гену оценивается при помощи метода условного максимального правдоподобия, где регрессором (переменной, от которой происходит регрессия зависимой переменной – в данном случае дисперсии) являются каунты. Затем эмпирическая Байесова процедура применяется для того, чтобы свести дисперсию по всем генам к консенсусному значению, и именно на

этом шаге для оценки дисперсии используется информация, полученная от всех генов в образце. После этого достоверные межгрупповые различия по каунтам определяются при помощи модифицированной версии точного теста Фишера, адаптированной для того, чтобы подходить выборкам с овердисперсией [77].

Удобной особенностью этого метода является то, что в нем возможно выявлять различия в мультифакторных экспериментах. Мультифакторными экспериментами называются такие эксперименты, в которых существует более чем один фактор, действующий на экспрессию генов в исследуемых образцах. Факторы, воздействующие на экспрессию генов, могут быть как объектом исследования, так и не представляющими интереса для исследователя. Такие нежелательные факторы, как правило, называются запутывающими (confounding) факторами. В пакете программ используется регрессия дисперсии каунтов по генам к обобщенным линейным моделям. В такой регрессионной модели логарифм дисперсии по каждому гену раскладывается на логарифм суммы всех каунтов в библиотеке плюс вектор состояний по факторам с коэффициентами для каждого. После подбора коэффициентов в такой обобщенной линейной модели становится возможно определить, какие доли дисперсии возникают в результате действия каких факторов, что в конечном итоге позволяет провести несколько межгрупповых сравнений одновременно, при этом свести к минимуму влияние запутывающих факторов [70].

Сходным образом с edgeR, в пакете программ «DESeq2» для R каунты моделируются при помощи негативно-биномиального распределения. Опять же сходным образом DESeq2 использует предположение о постоянности соотношения среднего и дисперсии по всем генам для того, чтобы улучшить точность оценки дисперсии, и, как следствие, статическую мощность теста на дифференциальную экспрессию. Кроме того, сходным образом обобщенные линейные модели используются в DESeq2 для анализа мультифакторных экспериментов [71]. В отличие от edgeR, DESeq2 использует не условную, а стандартную форму максимального правдоподобия для оценки дисперсии каунтов по каждому гену. Вместо того, чтобы сводить дисперсию каунтов по всем генам к консенсусному

значению, эмпирический Байесовский подход используется для того, чтобы привести соотношение дисперсии и среднего к гладкой кривой. Вместо теста Фишера значимые различия между группами определяются при помощи теста Уолда (Wald) [71].

Последний подход, на котором хотелось бы остановиться, реализован в пакете программ `limma` для программной среды R. Интересной особенностью `limma` является то, что каунты в нем моделируются как нормально распределенные. Обоснование этого заключается в том, что правильный подбор параметров несет значительно большее влияние на точность итогового выявления дифференциальной экспрессии, чем выбор подходящего распределения [68].

Изначально «`limma`» разрабатывалась как метод выявления дифференциальной экспрессии в микроарреях, платформе, мерка экспрессии гена на которой общепринято считается нормально распределенной. Уже тогда в этом методе были имплементированы процедуры, увеличивающие мощность теста на дифференциальную экспрессию за счет использования дополнительной информации, содержащейся в каждом анализе образца. На микроарреях такая информация представляет собой дубликаты и трипликаты зондов на определенные гены. В «`limma`» использовались линейные модели для того, чтобы разделить показатели экспрессии для каждого индивидуального зонда для гена и общие для нескольких зондов на ген [78]. Современная имплементация «`limma`» использует линейные модели для анализа мультифакторных экспериментах по секвенированию транскриптома и микроаррей экспериментов.

С популяризацией RNA-seq, в «`limma`» были добавлены методы, использующие модификации созданных ранее подходов для выявления дифференциальной экспрессии в экспериментах по секвенированию транскриптома. Именно этим определяется тот факт, что каунты в «`limma`» моделируются как нормально распределенные. Однако несмотря на различие в распределениях, в «`limma`» также как в `edgeR` и `DESeq2` вариабельность каунтов по каждому гену представляет собой некую компромиссную величину, состоящую из вариабельности по индивидуальному гену в группе образцов и консенсусной

вариабельности по всем генам. Эмпирическая Байесова процедура также используется для того, чтобы прийти к описанному выше консенсусному значению вариабельности по всем генам [68]. Для выявления значимой дифференциальной экспрессии в «limma» используется модерированный t-test Стьюдента.

Уникальной особенностью «limma» представляются использование весов измерений на всех этапах анализа дифференциальной экспрессии – от имплементированных в “limma” методов нормализации до весов для каждого гена в группе или даже индивидуального гена в одном образце. Эти веса могут быть заданы априорно, а могут быть подобраны на основании экспрессионных данных. Основная идея использования этих весов заключается в том, что они репрезентативны для «качества» индивидуальных генов или измерений. Чем более «качественным», то есть надежным и стабильным является определенный ген или чем ближе определенное измерение к ожидаемым по модели величинам, тем больше его «вес» – тем большее влияние он оказывает на другие гены и измерения в шагах с использованием информации о других генах [68]. Эти «веса» были изначально разработаны для микроарреев и главным образом подбирались так, чтобы быть обратно-пропорциональными вариабельности в повторах одного гена в группе и на одном микроаррее [79]. В имплементации для секвенирования транскриптома эти веса определяются как то, насколько индивидуальные гены укладываются в оцениваемый на основании всех генов во всех образцах тренд соотношения среднего к дисперсии. Чем лучше индивидуальное измерение и ген укладываются в этот тренд, тем выше их вес [80].

Таким образом, можно сделать вывод, что несмотря на разнообразие индивидуальных подходов для выявления дифференциальной экспрессии, у самых часто используемых из них есть легко заметные общие черты. Все они тем или иным способом направлены на улучшение мощности теста за счет точной оценки параметров распределения при помощи использования информации от всех генов. Все они используют линейные модели в каком-то виде для обработки мультифакторных экспериментов. Все они используют модифицированные версии широко распространенных параметрических статистических тестов для выявления

дифференциальной экспрессии. К сожалению, так же, как и в случае инструментов для картирования, не существует всеобъемлющих исследований, сравнивающих различные подходы, однако использование симулированных данных и датасета SEQC для сравнения различных алгоритмов также может быть использовано для выбора наилучшего инструмента.

На данный момент в области для анализа экспериментов по секвенированию транскриптома разработано множество различных инструментов и подходов, однако общая структура и идея этого метода достаточно ясно установлена. Общий подход в подавляющем большинстве случаев будет заключаться в подрезке по качеству, картированию на референс, подсчету каунтов и использованию математических моделей для выявления дифференциальной экспрессии. Выбор наилучших подходов для осуществления каждого из этих этапов является наиболее важной проблемой в практическом применении секвенирования транскриптома.

2.3 Коронавирусы в патофизиологии человека

2.3.1 Краткая история пандемии COVID-19

Семейство вирусов *Coronaviridae* – таксономическая группа вирусов позвоночных. Они относятся к порядку *Nidovirales*, и включают в себя три рода – *Alphacoronavirus*, *Betacoronavirus* и *Gammacoronavirus*. Вирусы из семейства *Coronaviridae*, как правило, имеют сферическую или палочковидную форму и покрыты оболочкой. Их геном представляет собой одну смысловую одноцепочечную молекулу РНК около 30 т.п.н. длиной [81]. Геном *Coronaviridae* – один из самых больших среди РНК-содержащих вирусов.

Коронавирусы прикрепляются к специфическим рецепторам на поверхности клеток эпителия хозяина и вводят свой геном внутрь при помощи слияния своей оболочки либо с клеточной мембраной, либо с ограничивающей мембраной эндоцитарной везикулы. Прикрепление осуществляется при помощи spike-белка, одного из двух поверхностных белков коронавируса. Весь процесс репликации вируса протекает в цитоплазме. Смысловая цепь РНК вируса служит как матрицей для синтеза белков, так и матрицей для синтеза анти-смысловой цепи РНК [82].

Коронавирусы вызывают респираторные и кишечные инфекции у животных и человека. Болезнь у инфицированных, как правило, протекает легко или асимптоматически, однако встречаются и исключения. Коронавирусы не считались высокопатогенными для человека до вспышки «атипичной пневмонии» (SARS) в провинции Гуандон (Guandong) в Китае, произошедшей в 2002 году [83]. Заболевание начиналось с подъема высокой температуры и относительно мягких респираторных симптомов, быстро перерастающих в пневмонию в течение нескольких последующих дней [84]. К февралю 2003 года болезнь распространилась в соседние страны и регионы. К июлю 2003 года ВОЗ зарегистрировала 8437 случаев атипичной пневмонии по всему миру, из которых 813 закончились смертью [84].

15 мая 2003 года вышла статья Fouchier и др. [85] в которой было продемонстрировано выполнение постулатов Коха для вируса **SARS-associated**

coronavirus (SCV). Исполнение постулатов Коха, модифицированных для вирусов, считается золотым стандартом для доказательства инфекционной природы заболевания и демонстрацией того, что заболевание вызывается исследуемым вирусом [86]. В качестве близкого вида для экспериментального заражения была использована макака-крабоед.

Позже этот вирус был переименован интернациональным комитетом по таксономии вирусов сначала в **Severe acute respiratory syndrome coronavirus** в 2004 [87], а затем в 2009 году в **Severe acute respiratory syndrome-related coronavirus** (SARS-CoV) [88]. Последнее название сохранилось за ним и по сей день.

13-го июля 2012 года в Jeddah, Саудовской Аравии, из мокроты больного был выделен новый коронавирус, названный HCoV-EMC. Больной поступил в больницу с температурой, кашлем, отхаркиванием и одышкой. Через 11 дней больной погиб от прогрессирующих легочной и почечной недостаточности [89]. В 2013 году вирус был переименован в MERS (Middle East Respiratory Syndrome)-CoV [90], [91]. По данным ВСО, в общей сложности к июлю 2019 года было зарегистрировано 2442 лабораторно подтвержденных случая MERS и 843 связанных с ним смерти [92].

В декабре 2019 года в Ухане, провинции Hubei, Китай произошло несколько случаев заболевания ранее неизвестной респираторной инфекцией. Инфекция быстро распространилась сначала по всей провинции, а затем и по всей стране [1]. Вирусом, вызывающим эту инфекцию, оказался новый коронавирус, имеющий последовательность, на 79.6 % совпадающую с SARS-CoV. Помимо этого, вирус на 96.2 % совпадает по последовательности с коронавирусом летучей мыши *Rhinolophus affinis* BatCoV RaTG13 [2]. К марту 2020 года новый коронавирус настолько широко распространился по всему миру, что ВОЗ официально назвала ситуацию «пандемией» [3]. Болезнь получила название COVID-19, а вызывающий ее вирус- SARS-COV-2 [93]. По данным ВОЗ, на 2 августа 2022 года в мире произошло 596 млн. подтвержденных случаев заболевания COVID-19, включая 6 млн. случаев, закончившихся смертельным исходом [4]. На 28 мая 2025 года эта цифра выросла до 777 млн. случаев, более 7 млн. из которых закончились

смертельным исходом.

В настоящий момент наиболее вероятным источником происхождения SARS-COV-2 является зоонозный переход из естественного резервуара. Предположительно, основным резервуаром для всех известных коронавирусов, поражающих человека, являются летучие мыши [94]. Однако SARS-подобные коронавирусы не могут заражать человека напрямую. Для этого им требуется пройти рекомбинацию или мутацию в организме естественного или промежуточного хозяина. Такими хозяевами для MERS и SARS, возможно, служили верблюды и вивера [94]. Промежуточным хозяином для SARS-COV-2 мог служить панголин [95].

2.3.2 Классификация вариантов COVID-19

Однако в результате определенных замен или их сочетания могут возникать варианты, отличающиеся от оригинального вируса (т. н. вариант «Ухань») с точки зрения заразности, симптомов, тяжести, трудности обнаружения и эффективности мер противодействия против них (табл. 1). Несколько таких вариантов возникло с начала пандемии. WHO присваивает этим вариантам греческие буквы, но у них также присутствуют коды происхождения, вкратце обозначающие их позиции на филогенетическом дереве. Center for Disease control and prevention (CDC) классифицирует функциональные варианты SARS-CoV-2 как Variants of Interest (VOI), Variants of Concern (VOC), Variants Being Monitored (VBM) и потенциально Variants of High Consequence (VOHC) [96]. Стоит отметить, что ни один вариант на данный момент не классифицирован и не был классифицирован как VOHC.

Таблица 1. История классификации функциональных вариантов SARS-COV-2 по данным WHO и CDC [96], [97]. Красный цвет обозначает классификацию варианта как VOC, жёлтый - VOI, зелёный – VBM/VUM.

Функциональные варианты	29.12. 2020	26.02. 2021	19.03. 2021	07.05. 2021	15.06. 2021	29.06. 2021	21.09. 2021	26.11. 2021	21.09. 2022	14.04. 2022
Alpha	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Beta	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Gamma	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Delta	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Epsilon	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Eta	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Iota	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Kappa	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
B.1.617.3	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Zeta	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Mu	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX
Omicron	XXX	XXX	XX	XXX	XXX	XXX	XXX	XXX	XXX	XXX

К VOC относят варианты, замены в которых привели к

продемонстрированному повышению инфективности, более высокой опасности заболевания для населения (повышение госпитализаций или смертей), значительному снижению эффективности иммунного ответа после предыдущего заболевания другим вариантом или вакцинации, а также повышению вероятности ошибки в процессе диагностики. В начале пандемии (с 29 декабря 2020 года) к VOI были отнесены варианты Alpha, Beta и Gamma. 19 марта 2021 года к ним присоединился вариант Epsilon, а 15 июня 2021 года вариант Delta. В сентябре 2021 года Alpha, Beta, Gamma и Epsilon были категоризированы как VBM, оставив Delta единственным VOC до ноября 2021 года. В ноябре 2021 года был обнаружен новый вариант Omicron, отличающейся значительно повышенной инфективностью и по сей день являющийся VOC [97]. 14 апреля 2022 года Delta был переквалифицирован в VBM [96]. На данный 2023 год, вариант Omicron оказался единственным циркулирующим VOC [97], и в дальнейшем подварианты этого варианта стали доминирующими среди распространенных вариантов вируса. В связи с этим ВОЗ приняла решение изменить классификацию вариантов вируса, не присуждая более греческие буквы вариантам, не категоризированным как VOC [98]. Стоит отметить, что на 28 мая 2025 года ни один варианта SARS-CoV-2 не охарактеризован как VOC.

К VOI относятся варианты с генетическими маркерами, ассоциированным с пониженной нейтрализацией антителами, выработанными в процессе предыдущих инфекций или вакцинации, изменениями в связывании с рецептором, сниженной эффективностью лечения, возможным воздействием на методы диагностики, или предсказанной повышенной заразностью или тяжестью заболевания. Варианты Epsilon, Eta, Iota, Kappa, Zeta и вариант без присвоенной греческой буквы с кодом происхождения B.1.617.3 ранее были обозначены как VOI, но на уже в середине 2023 они были категоризированы как VBM [96]. На 28 мая 2025 года к VOI относят только вариант JN.1, впервые описанный в августе 2023 года и составляющий более 90 % случаев COVID-19 и по сей день [17].

Варианты, которые циркулируют на низком уровне, те, которые циркулировали ранее, но более не обнаружены, и замены, в которых при этом

обладают потенциальным или продемонстрированным эффектом на медицинские меры противодействия, относят к VBM/VUM. Считается, что на данный момент такие варианты не представляют высокой опасности. Ранее к VBM/ VUM относили варианты Alpha, Beta, Gamma, Epsilon, Delta, Eta, Iota, Kappa, Zeta, Mu и вариант без присвоенной греческой буквы с кодом происхождения В.1.617.3 [96]. На 28 мая 2025 года к VUM относят варианты KP.3, KP.3.1.1, LB.1, XEC, LP.8.1 и NB.1.8.1.

2.3.3 Проникновение вируса в клетку

SARS-COV-2 проникает в эпителиальные клетки, связываясь присутствующим на поверхности вируса шиповидным гликопротеином S и с присутствующим на поверхности клетки ангиотензин-конвертирующим ферментом (ACE2). Трансмембранный белок ACE2 присутствует на поверхности клеток в мозге, сердце, легких, кишечнике и почках, что делает эти органы потенциальными мишенями вируса [99]. Роль ACE2 была продемонстрирована путем трансфекции человеческого ACE2 в клетки линии ВНК (Baby Hamster Kidney), в норме не подверженные инфекции SARS-COV-2. Такая трансфекция делала клетки уязвимыми к инфекции [100]. Стоит отметить, что SARS также использует ACE2 для проникновения в клетку и имеет рецептор-связывающий мотив, очень сходный по аминокислотной последовательности с SARS-COV-2 [101], а антитела к S белку SARS-COV нейтрализуют инфективность SARS-COV-2 [100].

Помимо ACE2, SARS-COV-2 также может потенциально проникать в клетку, используя CD147 – гликопротеид из суперсемейства иммуноглобулинов. Wang и др. [102] продемонстрировали, что S белок связывается с CD147 *in vitro*. В клеточных культурах с нокдауном CD147 инфективность SARS-COV-2 значительно снижена, а в культурах с его оверэкспрессией повышена. Введение Меплазумаба, антитела к CD147, снижало инфективность вируса в клеточных культурах. При инфекции SARS-COV-2 в легких трансгенных мышей, несущих человеческий CD147, развиваются значительно более тяжелые гистопатологии, чем в легких мышей дикого типа [102].

Согласно результатам *in silico* моделирования, проведенного Ibrahim и др., еще одним рецептором, при помощи связывания с которым SARS-COV-2 может проникать в клетки, является GPR78, орфанный рецептор, связанный с G-белком [103].

Помимо присоединения к ACE2, SARS-COV-2 для успешного проникновения в клетку также требуется праймирование S белка сериновой протеазой TMPRSS2 и, в меньшей степени, цистеиновой протеазой CatB/L. Ингибирование обеих этих протеаз приводит к подавлению инфективности вируса, а ингибирование только TMPRSS2 к значительному ее снижению [101]. В эпителии кишечника праймирование S белка также может производиться TMPRSS4, другой сериновой протеазой [104]. В клеточных культурах Hela и Calu-3 было показано, что Furin, белок из семейства субтилизин-подобных конвертаз, также может способствовать праймированию S белка, внося разрез в сайт разрезания, отличный от разрезаемого TMPRSS2 [105].

После присоединения вирус эндоцитируется клеткой, после чего происходит слияние клеточной и вирусной мембран [1].

Как упоминалось выше, клетки мозга, сердца, легких, кишечника и почек экспрессируют ACE2, что делает эти органы уязвимыми для поражения вирусом и находит отражение в симптомах. Наиболее частые симптомы COVID-19 включают в себя повышение температуры, кашель и усталость [5]. Более тяжелые случаи включают такие симптомы, как пневмония, острый респираторный дистресс-синдром (ОРДС), высокая вирусная нагрузка в плазме крови, острое повреждение миокарда и вторичные инфекции [6].

Подавляющее большинство пациентов с COVID-19 имеют какие-то из описанных выше респираторных симптомов [6], что указывает на высокое сродство вируса к клеткам эпителия дыхательных путей. Поражение легких, вызываемое пролиферацией вируса и сопутствующим воспалением, видимо на рентгене как затемнение [106].

Миокардит, также являющийся одним из симптомов COVID-19, может быть вызван как системным воспалением и цитокиновым штормом, так и прямым

заражением кардиомиоцитов вирусом [107]. Само по себе вызванное связыванием с ним вируса истощение и недостаточность ACE2 рецептора, и, как следствие, его неспособность конвертировать ангиотензин II в ангиотензин₁₋₇ может приводить к вазоконстрикции, повышенному воспалению и тромбозу [108], которые также могут быть причинами повреждения миокарда.

Помимо описанных выше симптомов, среди пациентов с COVID-19 также часто встречаются неврологические симптомы, такие как энцефалит, энцефалопатия, головная боль, тошнота, припадки и anosmia (потеря обоняния) [109]. Существуют основания считать, что вирус способен проникать в ЦНС как через кровь [109], так и напрямую, как например при интраназальной инфекции [110]. Вирус способен пролиферировать в мозге, приводя к гибели нейронов, ремоделированию сосудов и локальной ишемии [110].

2.3.4 Различная инфективность и вирулентность вариантов

Как было упомянуто выше, функциональные варианты вируса различаются с точки зрения их инфективности, вирулентности и эффективности мер противодействия против них. Наиболее значимыми параметрами, изменения в которых относительно варианта «Ухань» необходимо оценить, чтобы составить понимание о различии вариантов вируса, являются следующие: трансмиссивность, тяжесть/летальность, сниженная эффективность обнаружения при помощи RT-PCR, устойчивость к нейтрализации пост-вакцинальной сывороткой крови и устойчивость к лечению моноклональными анти-телами, входящими в рекомендованные протоколы лечения, их смесями и/или противовирусными препаратами (табл. 2) [111].

Вариант SARS-COV-2 Альфа впервые был обнаружен в Соединенном Королевстве (СК), в сентябре 2020-го года. Вариант быстро стал доминантным вариантом в СК и распространился по всему миру [112], что указывает на его преимущество в приспособленности перед вариантом «Ухань» [7]. В общей сложности этот вариант несет 17 мутаций, 8 из которых приходится на S-белок. Этот вариант на 43-90% более инфективен, чем циркулировавшие ранее в СК

варианты [8]. Данные о тяжести и риске смерти противоречивы – существуют как данные, указывающие на отсутствие различий, так и данные, указывающие на увеличение тяжести и риска смерти в сравнении с другими вариантами [7].

Вариант SARS-COV-2 Бета был впервые обнаружен в декабре 2020 года в Южной Африке. Вскоре после этого этот вариант был выявлен в других странах [113]. В общей сложности этот вариант несет 9 мутаций в S белке, 3 из них в рецептор-связывающем домене. Этот вариант приблизительно на 50% более инфективен, чем ранее циркулировавшие в Южной Африке варианты [7]. Этот вариант также ассоциирован с повышенным риском госпитализации, но не попадания в ОРИТ или смерти [114]. Мутации в S белке этого варианта позволяют ему частично или полностью избегать нейтрализации плазмой переболевших другими вариантами и используемыми в рекомендованных протоколах лечения моноклональными антителами [115], [116]. Этот вариант также обладает повышенной устойчивостью к нейтрализации антителами вакцинированных несколькими различными вакцинами [117], [118] и повышенной устойчивостью относительно варианта Альфа к вакцинации Novavax [119].

Таблица 2. Свойства значимых вариантов SARS-COV-2 (по данным, собранным в обзоре Fernandes et al. 2022 [111] и Choi, Smith 2021 [7]). Стрелка вверх означает повышенную характеристику в сравнении с вариантом «Ухань». Прочерк означает, что изменение характеристики в сравнении с вариантом «Ухань» не показано.

Вариант	Трансмиссивность	Тяжесть/ летальность	Устойчивость к нейтрализации поствакцинально й сывороткой крови	Устойчивость к лечению рекомендованными протоколами
Alpha	↑	-	-	-
Beta	↑	↑	↑	↑
Gamma	-	↑	↑	↑
Delta	↑	↑	↑	↑
Epsilon	↑	-	↑	↑
Omicron	↑	-	↑	-

Вариант SARS-COV-2 Гамма впервые был выявлен в январе 2021 года в группе людей, путешествующих из Бразилии в Японию [120]. Гамма вариант несет 10 мутаций в S белке, три из которых располагаются в рецептор связывающем домене [7]. Вариант ассоциирован со значительно повышенным риском госпитализации и попадания в ОРИТ [8]. Вариант Гамма устойчив к нескольким видам моноклональных антител, входящих в рекомендованные протоколы лечения, и к нейтрализации сывороткой крови вакцинированных и переболевших другими вариантами вируса [9]. Помимо этого, вакцины BNT162b2 и mRNA-1273 менее эффективны против варианта Гамма, чем против вариантов Альфа и Бета [121].

Вариант SARS-COV-2 Дельта был впервые выявлен в марте 2021 года в Индии, и в течении нескольких месяцев он распространился по нескольким другим странам [7]. В мае 2011 года его трансмиссивность была оценена как не меньшая, чем трансмиссивность варианта Альфа [13], однако в вышедших позже исследованиях его трансмиссивность была оценена как более высокая, чем у варианта Альфа [14], [15]. К сентябрю 2021 года этот вариант стал наиболее распространенным среди вариантов SARS-COV-2 [7]. Этот вариант ассоциирован со значительно более высокой вероятностью попадания в ОРИТ, необходимости искусственной вентиляции и смерти (4.9 ОШ) [16]. Эффективность

моноклональных антител, входящих в рекомендованные протоколы лечения, а также антител из крови вакцинированных и переболевших другими вариантами вируса значительно ниже против варианта Дельта, чем против варианта Альфа [11]. Вакцины BNT162b2 и ChAdOx1 nCoV-19 также значительно менее эффективны против варианта Дельта [11], [12].

Вариант Эпсилон был впервые обнаружен в мае 2020 года в США в штате Калифорния, и к маю 2021 он распространился в 34 другие страны [122]. Мутации, которые несет этот вариант вируса в псевдовирусных частицах приводят к значительному увеличению инфективности [122]. Вирусная нагрузка в крови инфицированных вариантом Эпсилон превышает вирусную нагрузку в крови инфицированных другими вариантами вируса [122]. С момента обнаружения по тридцатое апреля 2021 общая доля заражений вариантом Эпсилон в штате Калифорния выросло до 50%, что может указывать на повышенную инфективность этого варианта вируса [10]. Вирус обладает значительной устойчивостью к нейтрализации антителами из плазмы выздоровевших после заражения вирусом анцестрального варианта, а также антителами вакцинированных вакцинами BNT162b2 и mRNA1273 и моноклональных антител, входящих в рекомендованные практики лечения [10].

Вариант Омикрон впервые был выявлен в ноябре 2021 года в Южной Африке, однако вскоре выяснилось, что он был занесен путешественниками из Голландии [123]. К декабрю 2021 года он был выявлен в 80 странах по всему миру [123]. На 21 октября 2022 года это единственный вариант, категоризированный ВОЗ как VOC [97]. Этот вариант несет наибольшее число мутаций в сравнении с другими вариантами — около 50 в общей сложности, при этом более 30 из них приходится на S белок [124]. Скорость роста процента инфицированных новым вариантом от общего числа больных с COVID-19 в южной Африке оказалась значительно выше, чем всех предыдущих вариантов [125]. Сходная ситуация наблюдалась в США и Великобритании в конце 2021-начале 2022 года [123]. *In silico* моделирование связывания рецептор-связывающего домена S белка варианта Омикрон с ACE2 указывает на более эффективное связывание, чем то, которое наблюдается в

варианте Дельта [126]. Другой проведенный *in silico* анализ также указывает на повышенную трансмиссивность варианта в сочетании со повышенной устойчивостью к нейтрализации моноклональными антителами [127]. Устойчивость варианта Омикрон к нейтрализации антителами из сыворотки крови переболевших COVID-19 и вакцинированных вакциной BNT также превосходит устойчивость других вариантов вируса, включая вариант Дельта [128].

Первые мутировавшие подварианты варианта Омикрон были выявлены в декабре 2021-январе 2022. К 16 марта 2023 года подварианты варианта Омикрон стали доминирующими вариантами COVID-19 и практически полностью вытеснили все другие варианты [129]. В связи с этим ВОЗ обновила свою классификацию вариантов, а также переклассифицировала все ранее циркулировавшие варианты и оригинальный вариант Омикрон как VBM/VUM [129]. Также было принято решение что впредь только VOC будут получать греческие наименования, а остальные варианты будут называться по номенклатуре их происхождения [98].

Первые образцы, в которых был выявлен подвариант ХВВ.1.5, были получены 5 января 2022 года. К ноябрю 2022 заражения этим вариантом (распространенность) составляли примерно 3% всех случаев COVID-19. К февралю 2023 года его распространенность достигла пика в 55% всех случаев COVID-19, после чего распространенность ХВВ.1.5 пошла на спад. К началу июня 2023 года его распространенность составляла 23% [130]. На данный момент (25.07.2023) этот подвариант является одним из 2 вариантов, классифицированных как VOI [97]. Согласно данным ВОЗ [130] и CDC [131], подвариант ХВВ.1.5 обладает более низкой инфективностью, чем другие циркулирующие на июнь 2023 года подварианты Омикрон, и его доля среди всех случаев COVID-19 будет продолжать падать. Однако стоит отметить, что все подварианты линии ХВВ.1.* [132], и специфически ХВВ.1.5 [133] обладают более высокой инфективностью и способностью к избеганию нейтрализации антителами к другим вариантам COVID-19, чем выявленные ранее подварианты Омикрон, такие как ВА.2.75. Полученные на данный момент данные указывают на то, что с точки

зрения тяжести протекания заболевания подвариант ХВВ.1.5 не отличается от других подвариантов Омикрон [134]. Помимо этого, различные противовирусные препараты, ранее показавшие эффективность против других вариантов SARS-COV-2, обладают различной эффективностью против ХВВ.1.5. Так, имдевимаб-казиривимаб, тиксагевимаб-цилгавимаб, сотровимаб и бектеловимаб не показали эффективности против этого подварианта в клинических условиях, в то время как для препаратов ремдезивир, молнупиравир, нитрматрелвир и энсителвир было продемонстрировано сохранившаяся эффективность в исследованиях *in vitro* [135].

Подвариант ХВВ.1.16 был впервые выявлен 9 января 2023 года. В марте 2023 года он был классифицирован ВОЗ как VUM, однако в силу неуклонного роста его распространенности 17 апреля 2023 он был переклассифицирован в VOI [136]. 17 апреля 2023 его распространенность составляла 10.2%. Через 4 недели его распространенность возросла до 16.8%. Этот подвариант – второй из двух, классифицированных на данный момент как VOI [97]. В отличие от ХВВ.1.5, по данным CDC [131] и ВОЗ [136], ХВВ.1.16 обладает повышенной инфективностью в сравнении с другими подвариантами варианта Омикрон, что, возможно, приведет к повышению его распространенности в будущем. Во многих географических регионах распространенность этого варианта растет быстрее, чем у всех остальных вариантов, что потенциально может привести к тому, что этот вариант станет наиболее распространенным в будущем [136]. С точки зрения способности к избеганию нейтрализации антителами [137] переболевших другими вариантами COVID-19, а также тяжести течения [138], [139] свойства ХВВ.1.16 крайне сходны со свойствами ХВВ.1.5, описанными выше. Показано, что противовирусный препарат сотровимаб эффективен в лечении инфекции ХВВ.1.16 [137].

Подвариант EG.5 был впервые выявлен 17 февраля 2023 года. 19 июня 2023 года он был классифицирован ВОЗ как VUM, однако в силу исключительно высокой скорости роста его распространенности 9 августа 2023 года он был переквалифицирован как VOI [97]. За 4 недели с 19 июня его распространенность выросла с 7% до 17%. На данный момент не опубликовано работ, посвященных этому варианту, однако такой быстрый рост распространенности в контексте

конкуренции с другими высокоинфективными подвариантами свидетельствует о более высокой инфективности EG.5 в сравнении с ХВВ.1.5 и ХВВ.1.16 [140]. Также существуют основания считать, что с точки зрения избегания нейтрализации антителами этот подвариант также обладает повышенной приспособленностью. Подвариант ХВВ.1.5 с внесенной мутацией F456L, обладающий идентичной последовательностью шиповидного гликопротеина с EG.5, устойчив к нейтрализации большинством антител, нейтрализующих ХВВ.1.5 дикого типа [141].

Помимо этих, 6 подвариантов варианта Омикрон на 2023 год были классифицированы как VUM. Это варианты ВА.2.75, СН.1.1, ХВВ, ХВВ.1.9.1, ХВВ.1.9.2 и ХВВ.2.3 (табл. 3).

Таблица 3. Подварианты варианта Омикрон по классификации ВОЗ [97]

Название	Дата выявления первого задокументированного образца, содержащего подвариант	Дата последней классификации	Классификация
XBB.1.5	05-01-2022	20-06-2023	VOI
XBB.1.16	09-01-2023	05-06-2023	VOI
BA.2.75	31-12-2021	06-07-2022	VUM
CH.1.1	27-07-2022	08-02-2023	VUM
XBB	13-08-2022	12-10-2022	VUM
XBB.1.9.1	05-12-2022	30-03-2023	VUM
XBB.1.9.2	05-12-2022	26-04-2023	VUM
EG.5	17-02-2023	09-08-2023	VOI
XBB.2.3	09-12-2022	17-05-2023	VUM

Вариант JN.1 был впервые выявлен 23 августа 2023 года. 19 декабря 2023 года он был классифицирован ВОЗ как VOI в силу его быстрого распространения и вытеснения случаев его предкового варианта. К этому времени его распространенность составляла от 5 до 20 % всех подтвержденных случаев COVID-19. Несмотря на то, что причины его повышенной заразности не были выявлены (в сравнении с другими подвариантами варианта Омикрон), его распространенность продолжала неуклонно расти [142]. К 9 февраля 2024 года этот вариант достиг распространенности в 89% всех подтвержденных случаев COVID-19 [18]. К апрелю 2024 года его распространенность достигла более чем 90% [17], и, вплоть до момента написания этих строк, (30 мая 2025 года), его распространенность не снижалась до менее чем 90 %, делая этот вариант наиболее значимым и доминантным подвариантом Омикрон. Этот вариант вытеснил все остальные VOI и является единственным классифицированным как VOI на данный момент. Н 30 мая 2025 года 6 других наиболее распространенных вариантов классифицировано как VUM. Это варианты KP.3, KP.3.1.1, XEC, LP.8.1, LP.8.1, NB.1.8.1 (табл. 4), все из которых происходят от JN.1 и несут дополнительные мутации по сравнению с этим вариантом.

Таблица 4. Подварианты варианта Омикрон по классификации ВОЗ по состоянию на 2025 год

Название	Дата выявления первого задокументированного образца, содержащего подвариант	Дата последней классификации	Классификация
KP.3	11-02-2024	03-05-2024	VUM
KP.3.1.1	27-03-2024	19-07-2024	VUM
LB.1	26-02-2024	28-06-2024	VUM
XEC	26-06-2024	03-02-2025	VUM
LP.8.1	01-07-2024	12-10-2022	VUM
NB.1.8.1	22-01-2025	23-05-2025	VUM

2.4 Полнотранскриптомные исследования при COVID-19

2.4.1 Процедура поиска и отбора исследований

С начала эпидемии было опубликовано множество исследований, посвященных различным молекулярным аспектам COVID-19. Одно из первых исследований, включающее связанный с COVID-19 полнотранскриптомный анализ, датируется апрелем 2020 года, и оно посвящено определению экспрессии *ACE2* в клетках из различных тканей и органов [143]. Для того чтобы получить представление о развитии таких исследований, на первом этапе был произведен поиск по ключевым словам “COVID-19” или “SARS-COV-2” и “RNA-seq”. Из полученных 604 статей были отобраны 42 статьи, в которых производится оригинальное исследование транскриптома ткани хозяина или культуры клеток, зараженных SARS-COV-2. Две статьи из этих 42 были исключены из дальнейшего анализа в силу глубоких методических проблем в статьях (отсутствие поправки на множественное тестирование, использование «внешнего» контроля) [144], [145]. Краткая информация о проанализированных статьях содержится в табл. 5.

В процессе анализа литературных данных было выделено несколько биологических процессов, выявленных сразу во многих статьях – ответ на интерфероны 1 и 2, цитокиновый сигналинг, дегрануляция нейтрофилов и активация системы комплемента.

2.4.2 Ответ на интерфероны 1 и 2

В 16 из 40 рассмотренных статьях авторы выделяют те или иные биологические процессы и сигнальные пути, связанные с ответом на интерфероны [146]–[161]. Как правило, эти пути либо значимо обогащены генами с пониженной экспрессией, либо специфически связаны с нарушением этих сигнальных путей. В этих статьях была выявлена дифференциальная экспрессия в следующих сравнениях:

- между больными с COVID-19 и выздоравливающими больными с COVID-19 или здоровыми контролями [146], [148], [150], [151], [153], [156], [160], [161];

- между больными с легкой тяжелой формой COVID-19 [148], [156]–[158], [161];
- между больными COVID-19 и больными другими респираторными и нереспираторными заболеваниями [147], [149], [154];
- между зараженными и незараженными клеточными культурами и тканями [149], [152], [154], [159].

Таким образом можно предположить, что ингибирование интерферонового сигналинга является характерной чертой тяжелого клинического течения COVID-19.

Интерфероны – это сигнальные белки из семейства цитокинов, активирующие иммунную систему для защиты от различных патогенов. Они считаются необходимой частью врожденного иммунного ответа [162]. Экспрессия интерферонов 1 и 3 индуцируется во всех типах иммунных клеток в ответ на взаимодействие *pattern recognition receptors* (PRR) с компонентами вирусов. Экспрессия интерферона 2 индуцируется в Т лимфоцитах и НК клетках в ответ на цитокины, такие как IL-12 [163]. Интерфероны, в свою очередь, стимулируют экспрессию интерферон-стимулируемых генов (ISG). Исходя из приведенных в литературе оценок, в общей сложности ответ на интерфероны модулируется индукцией около 2000 ISG. Большинство из них напрямую не реализуют функции ингибирования размножения вирусов, вместо этого индуцируя сигнальные молекулы, стимулирующие иммунный ответ в других клетках и распространяя сигналы из инфицированных тканей и клеток [164].

Интерфероны и индуцируемые ими гены участвуют в формировании antiviral факторов, переходе клеток в состояние антимикробной готовности, и модулируют множество иммунных процессов, таких как активность В-лимфоцитов и гуморальный иммунный ответ, активация цитотоксических клеток и системы комплемента, и экспрессия антигенов комплекса гистосовместимости I и II [163]. Эксперименты на клеточных культурах показали, что интерфероны 1 и 3 способны ингибировать размножение SARS-COV-2 в клетках эпителия человека [165]. Таким образом, ингибирование интерферонового ответа, в свою очередь,

может привести как значительному размножению вируса в зараженных клетках, так и к нарушению других иммунных функций, и, соответственно, к более тяжелому течению болезни.

2.4.3 Цитокиновый сигналинг

В 8 из 40 рассмотренных статей авторы выделяют биологические процессы и сигнальные пути, связанные с цитокиновым сигналингом [148], [153], [166]–[171].

В этих статьях дифференциальная экспрессия в таких путях и процессах была выявлена между:

- пациентами, больными COVID-19 и здоровыми контролями [153], [166]–[168], [170];
- тяжелыми и легкими случаями COVID 19 [148], [170], [171];
- зараженными COVID-19 и контрольными клеточными культурами [166]–[169].

Цитокины – это обширная категория белковых сигнальных молекул, играющих роль в коммуникации между клетками [172]. Как правило они включают хемокины, интерфероны, интерлейкины, лимфокины и факторы некроза опухоли, но не включают белковые гормоны и факторы роста. Цитокины играют крайне значимую роль в функционировании и регуляции иммунной системы [173].

В рассмотренных работах процессы и пути, связанные непосредственно с производством цитокинов и ответом на них, были значимо обогащены генами с повышенной экспрессией, что может свидетельствовать о индукции цитокиновых сигнальных путей в тяжелых случаях COVID-19. Роль самоподдерживающегося и неконтрольного повышения уровня провоспалительных цитокинов, известного как «цитокиновый шторм», в болезнях, вызываемых патогенными коронавирусами человека, была выявлена и до эпидемии COVID-19 [174]. Пациенты с таким фенотипом, как в случае COVID-19, так и других заболеваний, имеют значительно повышенный риск развития синдрома острой дыхательной недостаточности, необходимости искусственной вентиляции легких, и смерти [175].

Следует отметить, что синдром «цитокинового шторма» включает не только

повышение уровня циркулирующих цитокинов и его непосредственные последствия (так называемого «синдрома выделения цитокинов»), но и следующие признаки [175]:

- вторичный лимфоцитарный лимфогистиоцитоз;
- синдром активации макрофагов.

Кроме того, биологические процессы и сигнальные пути, связанные с цитокиновым сигналингом представляют собой более общие биологические процессы. Если же рассмотреть также процессы, связанные с пертурбацией перечисленных выше подвидов цитокинов – интерфероны, интерлейкины, лимфокины, хемокины и факторы некроза опухоли – то представленность статей, в которых выявлена дифференциальная экспрессия в связанных с цитокинами процессах, будет значительно выше.

Как упоминалось в предыдущей секции, дифференциальная экспрессия в процессах, связанных с ответом на интерфероны, встречаются в 16 из 40 [146]–[161]. Даже если не учитывать процессы, связанные с ответом на интерфероны, но учесть все статьи, в которых авторы отмечают дифференциальную экспрессию процессов и сигнальных путей, связанных с активацией макрофагов, интерлейкиновым сигналингом, хемокиновым сигналингом и сигналингом связанным с факторами некроза опухоли, то представленность процессов и сигнальных путей, связанных с цитокиновым сигналингом, возрастает до 14 из 40 [148], [150], [153], [160], [166]–[171], [176]–[179]. Учитывая, что различные цитокины могут иметь как про-, так и противовоспалительные эффекты, и то, что направление изменений экспрессии в связанных с цитокинами путях и процессах из рассмотренных статей менее однозначны, чем в случае интерферонов, нельзя однозначно сказать о том, какие из этих путей угнетены при инфекции SARS-COV-2, а какие активированы. Очевидно, что какая-то степень изменения экспрессии генов в биологических процессах и сигнальных путях, глубоко связанных с функционированием иммунной системы, будет выявляться при любом инфекционном заболевании. Однако совокупность рассмотренных данных позволяет предположить, что модуляция функционирования иммунной системы

при COVID-19 может играть значимую роль в прогрессе инфекции и ее тяжести, и что явление «цитокинового шторма» играет значимую роль в протекании болезни не только на уровне фенотипических проявлений, но и на уровне транскриптома.

2.4.4 Активация и дегрануляция нейтрофилов

В 9 из 40 рассмотренных статей авторы выделяют биологические процессы, связанные с активацией и дегрануляцией нейтрофилов [150], [161], [170], [171], [177], [180]–[183].

В этих статьях дифференциальная экспрессия в связанных с нейтрофилами путях и процессах была выявлена между:

- пациентами с COVID-19 и здоровыми контролями [150], [161], [170], [177], [181], [183];
- пациентами с тяжелой и легкой формой COVID-19 [170], [171], [180]–[182];
- пациентами с COVID-19 и другими респираторными заболеваниями [180].

В большинстве случаев эти процессы были обогащены генами с повышенной экспрессией и/или непосредственно связаны с активацией нейтрофилов, что может свидетельствовать о роли повышенной дегрануляции и активации нейтрофилов в тяжелом COVID-19.

Роль дегрануляции и активации нейтрофилов в патогенезе COVID-19 широко задокументирована. Высокое соотношение нейтрофилов к лимфоцитам может предсказывать повышенную вероятность летального исхода среди госпитализированных пациентов [184]. Повышенный уровень инфильтрации нейтрофилов [185] и повышенный уровень циркуляции внеклеточных ловушек нейтрофилов [186] часто наблюдаются в постмортальных тканях легкого пациентов, погибших при COVID-19.

Классический механизм работы нейтрофилов заключается в инфильтрации в зараженную ткань и выпуск внеклеточных ловушек, производящих активные формы кислорода (АФК). Одним из возможных объяснений иммунопатологической роли нейтрофилов в патогенезе тяжелой формы COVID-19 является обусловленная физическими факторами задержка миграции нейтрофилов из

микроциркуляторного русла легких, приводящая к аномально продолженному воздействию АФК, и, как следствие, повреждению клеток и тканей [187]. Сочетание активации нейтрофилов, системы комплемента и активации тромбоцитов может приводить к состоянию, называемому «иммунотромбозом», приводящему к дальнейшему повреждению микроциркуляторного русла легких из-за нарушения циркуляции. Иммунотромбоз является одним из предикторов тяжести исхода при COVID-19 [188]. Повышенная экспрессия генов в путях, связанных с активацией нейтрофилов, особенно в сравнении COVID-19 и других респираторных заболеваний, может свидетельствовать в пользу значительной роли нейтрофилов в повреждении тканей при COVID-19.

2.4.5 Активация системы комплемента

В 8 из 40 рассмотренных статей авторы выделяют биологические процессы, связанные с активацией системы комплемента [153], [160], [166], [180], [181], [189]–[191].

В этих статьях дифференциальная экспрессия в связанных с активацией системой комплемента путях и процессах была выявлена между:

- пациентами с тяжелой и легкой формой COVID-19 [180];
- пациентами с COVID-19 и пациентами с другими респираторными заболеваниями [180], [190], [191];
- пациентами с COVID-19 и здоровыми контролями [153], [160], [166], [181], [189].

В большинстве случаев эти процессы были обогащены генами с повышенной экспрессией и/или непосредственно связаны с активацией системы комплемента, что может свидетельствовать о роли активации системы комплемента в патогенезе тяжелой формы COVID-19.

Система комплемента играет значимую роль в антибактериальном и противовирусном врожденном и адаптивном иммунном ответе. Эта роль осуществляется как за счет прямой нейтрализации вирусных частиц, так и за счет активации других иммунных функций, таких как усиление хемотаксиса, регуляции

воспалительного и адаптивного иммунного ответа [192]. Однако, в случае COVID-19, активация системы комплемента может быть предиктором тяжелого течения заболевания даже в отсутствии значимого повышения уровня цитокинов [188]. Как упоминалось выше, продолжительная активации системы комплемента в сочетании с дегрануляцией нейтрофилов и активацией тромбоцитов может приводить к иммуотромбозу. Иммуотромбоз может приводить к повреждению тканей легкого, мочевыделительной системы, печени, сердца и, в итоге, к мультиорганной недостаточности [193]. Повышение экспрессии генов, связанных с активацией системы комплемента, подтверждает полученную другими методами информацию о значимости этого процесса для патогенеза тяжелого COVID-19.

Из обзора работ, посвященных RNA-seq при COVID-19 можно заключить, что дисрегуляция иммунного ответа играет значительную роль в патогенезе COVID-19 и определении тяжести исхода. В этом свете поиск целевых генов, через которые можно осуществить стимуляцию ингибированного интерферонового ответа [194], а также угнетение системы комплемента [188], [192], активации нейтрофилов [185], [187] и различных аспектов цитокинового шторма [175] представляется крайне важной задачей в разработке и улучшении терапии COVID-19.

Таблица 5. Сводная таблица статей с балк RNA-seq при COVID-19.

Название статьи	Дизайн эксперимента	Исследуемые ткани	Выделяемые авторами гены	Выделяемые авторами значимо обогащенные термины GO
Large-Scale Multi-omic Analysis of COVID-19 Severity	Пациенты с умеренными или тяжелыми респираторными симптомами, 102 COVID-19+ 26 COVID-19-	Лейкоциты крови	APOA1, APOA2, APOM, AGT, FBLN5, NID1, SERPINB1, NRP1, SERPINA4	«митотический клеточный цикл», «распознавание фагоцитоза», «позитивная регуляция активации В клеток», «активация системы комплемента (классический путь)», «антимикробный гуморальный ответ», «дегрануляция нейтрофилов», «ответ на фолиевую кислоту», «ремоделирование фосфотидил-холинов, плазменил-фосфотидил-холинов и липопротеинов высокой плотности»
Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19	13 пациентов с COVID-19 и 14 выздоравливающих контролей, в 5 временных точках	Периферическая кровь	Гены, вовлеченные в IL-1b вазодилаторный сигналинг, гены связанные с IFN 1 ответом	Обогащение по модулям - M2 – транскрипты, относящиеся к отказу ответа на интерферон первого типа, M4 – транскрипты, указывающие на присутствие мегакариоцитов в периферической крови, M7 – транскрипты, ассоциированные с эритроидной дифференцировкой и мегакариоцитами
Circuits between	88 пациентов с пневмонией от ковида,	Бронхоальвеолярный	Гены ответа на интерферон, гены,	Гены ответа на интерферон первого и второго типа

infected macrophages and T cells in SARS-CoV-2 pneumonia	211 пациентов с пневмонией от других патогенов и 42 интубированных пациентов без пневмонии.	лаваж	связанные с обогащением CD8+ Т клетками и уровнем Ц-реактивного белка.	
Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients	3 пациента с COVID-19 и 3 здоровых контроля	Бронхоальвеолярный лаваж и первичные мононуклеарные лейкоциты крови	(BALF и PBMC) IL6, IL6R, TP53 (BALF) IL10, CCL2/MCP-1, CXCL10/IP-10, CCL3/MIP-1A, CCL4/MIP1B, (PBMC) CTSL, CTSB, DDIT4, RRAS, CTSD, BIRC5, TNFSF10, CTSZ, NTRK1, IGFBP3, CCNB1, RRM2, CCNB2, GTSE1, CDK1, STEAP3, TP53I3.	Бронхоальвеолярный лаваж – «котрансляционные белки, расположенные в мембране», «белки, расположенные в эндоплазматическом ретикулуме», «вирусная транскрипция», «активация иммунных клеток» Первичные мононуклеарные лейкоциты крови – «активация системы комплемента», «гуморальный иммунный ответ, опосредованный циркулирующими иммуноглобулинами», «опосредованный В клетками иммунитет», «регуляция острого воспалительного ответа», «наведение аксонов», «наведение проекций нейронов», «биологические процессы, относящиеся к мРНК»
Disease severity-specific neutrophil signatures in	39 пациентов с COVID-19 и 19 здоровых контролей	Цельная кровь	CD177, RNASE2, HP, ELANE, OLFM4, MPO, RETN, MMP8, MMP9, S100A8,	«активация системы комплемента», «гуморальный иммунный ответ», «гуморальный иммунный ответ, опосредованный циркулирующими иммуноглобулинами», «активация

blood transcriptomes stratify COVID-19 patients			S100A9, S100A12, G0S2, CDC6, CDC25A, IFI27, IFITM3, SIGLEC1 CD169, IL10, SOCS3, ARG1, NELL2, RORC, KLRB1, TCF7 TCF1, RCAN3, BACH2, LEF1	нейтрофилов», «активация Т-клеток», «дифференцировка лимфоцитов»
HIF-1 α promotes SARS-CoV-2 infection and aggravates inflammatory responses to COVID-19	11 пациентов с COVID-19 и 9 здоровых контролей	Мононуклеарные лейкоциты крови	HIF- 1 α	«TNF сигналинг», «NF- κ B сигналинг», «HIF-1 сигналинг», «метаболизм»
DOCK2 is involved in the host genetics and biology of severe COVID-19	473 пациентов с COVID-19, eQTL, стратификация по тяжести болезни	Цельная кровь	DOCK2	-
Longitudinal transcriptome analyses show	Лонгитюдный анализ 18 пациентов с COVID-19 на этапе лечения,	Мононуклеарные лейкоциты	Транскрипционные факторы DDIT3, NR4A3, ZEB1,	«сигналинг интерферона 1-го типа», «гуморальный иммунный ответ, представляемый сигналингом рецепторов В

robust T cell immunity during recovery from COVID-19	восстановления и реабилитации со стратификацией по тяжести	крови	KLF10, JUNB и JUN), регуляция фосфорилирования AREG, DUSP8, OSM, MAP3K8, SOCS3, гены стимулируемые интерфероном TNF, IL1B, CXCR4, CD69, IER5	клеток», «аутофагия», «клеточный цикл», «продукция цитокинов», «регуляция дифференцировки Т клеток», «ответ на интерлейкин-1», «развитие иммунной системы», «регуляция коагуляции крови и активации тромбоцитов NIK/NF-κB»
Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions	215 пациентов с COVID-19, 517 пациентов, зараженных гриппоподобными заболеваниями с отрицательными тестами на COVID, 17 зараженных COVID-19 и 33 не зараженных Vero E6 cells	Клеточные культуры, назофарингитальные мазки	ACE2, IFI27, IFI6, CXCL10, CXCL11, CCL8	«ответ на интерферон альфа», «ответ на интерферон гамма»
Differential Co-Expression Network Analysis	17 здоровых контролей, 70 пациентов с COVID-19	Мононуклеарные лейкоциты крови	IL6, TNF, IL1B, CASP1	«элонгация митохондриальной трансляции», «терминация трансляции», «негативная регуляция сигналинга, опосредованная сигналингом интерферона-1-го типа»,

Reveals Key Hub-High Traffic Genes as Potential Therapeutic Targets for COVID-19 Pandemic				«сигналинг Toll-подобного рецептора», «сигналинг лектинового рецептора С-типа», «МАРК сигналинг», «NF-κB сигналинг», «сигналинг IL-17», «сигналинг NOD-подобного рецептора», «некроптоз», «сигналинг, опосредованный фактором некроза опухоли», «сигналинг, опосредованный интерфероном гамма», «сигналинг рецепторов Т-клеток», «нейтрофил опосредованный иммунитет», «активация нейтрофилов в иммунном ответе»
Transcriptional Profiling and Machine Learning Unveil a Concordant Biosignature of Type I Interferon-Inducible Host Response Across Nasal Swab and Pulmonary Tissue for COVID-19 Diagnosis	Анализ публично доступных датасетов GSE32155, GSE147507, GSE150316, GSE162835, GSE163151, GSE171668, GSE182569	Ткань легких и носовые мазки	BST2, IFIT1, IFIT2, IFIT3, IFITM1, ISG15, MX1, MX2, OAS1, OAS2, OAS3, OASL, RSAD2, STAT1.	«вирусный процесс», «Notch сигналинг», «движение ресничек», «болезнь Хантингтона», «развитие узора кровеносных сосудов», «переваривание и всасывание белков»

<p>Dual RNA-Seq analysis of SARS-CoV-2 correlates specific human transcriptional response pathways directly to viral expression</p>	<p>Анализ публично доступных датасетов GSE147507, CRA002390</p>	<p>ACE2 receptor-enhanced alveolar basal epithelial cell line (A549), bronchial epithelial cell line (NHBE), бронхоальвеолярный лаваж, мононуклеарные лейкоциты крови, ткань легких</p>	<p>CXCL5, CXCL8, CCL20, HIF1A, SPRR2A, SPRR2D, SPRR2E, PI3, KRT6B, ESF1, RHCG, MTSS1, OAS1, MX1, PARP9, DTX3L, ASS1</p>	<p>«активаность хемокинов», «сигналинг IL-17», «взаимодействие вирусных белков с цитокинами и цитокиновыми рецепторами», «клеточный ответ на липополисахариды», 'ASATC AAA G' TCF-3 мотив, «расположение котрансляционного белка в мембране», «расположение белка в эндоплазматическом ретикулуме», «транскрипция вируса», «активация системы комплемента», «ответ гуморальной иммунной системы, опосредованный циркулирующим иммуноглобулином», «иммунитет, опосредованный В клетками», «регуляция острого воспалительного ответа», «острый воспалительный ответ»</p>
<p>The whole blood transcriptional regulation landscape in 465 COVID-19 infected samples from Japan COVID-19 Task Force</p>	<p>106 асимптоматических больных COVID-19 и больных с легкой формой COVID-19 vs 359 больных с тяжелой и очень тяжелой формой COVID-19</p>	<p>Цельная кровь</p>	<p>TOR1AIP1, REST, STING1, CLEC4C, MYBL2</p>	<p>«дегрануляция нейтрофилов», «инициация иммунной системы», «антимикробные белки»</p>

<p>Integration and Reanalysis of Four RNA-Seq Datasets Including BALF, Nasopharyngeal Swabs, Lung Biopsy, and Mouse Models Reveals Common Immune Features of COVID-19</p>	<p>Анализ публично доступных датасетов PRJCA002273, GSE152075, GSE147507, PRJNA646535</p>	<p>Бронхоальвеолярный лаваж, назофарингеальные мазки, ткань легких, модели на мышах</p>	<p>CCL2, CXCL1, CXCL10, IL-1β, alarmin S100A8/S100A9</p>	<p>Бронхоальвеолярный лаваж – «рибосома», «сигналинг IL-17», «сигналинг TNF», «не-алкогольная жировая болезнь печени» Назофарингеальный мазок, ткани легких – «взаимодействие вирусных белков с цитокинами и цитокиновыми рецепторами», «взаимодействие цитокинов с цитокиновыми рецепторами», «хемокиновый сигналинг», «сигналинг NOD-подобного рецептора», «грипп А» Модели на мышах – «взаимодействие вирусных белков с цитокинами и цитокиновыми рецепторами», «сигналинг NOD-подобного рецептора», «хемокиновый сигналинг»</p>
<p>SARS-CoV-2 infection initiates interleukin-17-enriched transcriptional response in different cells from multiple organs</p>	<p>Анализ публично доступных датасетов GSE147507, GSE153970</p>	<p>Первичные клетки легкого, культура клеток A549</p>	<p>IL-6, CXCL8, CCL20, CXCL1, CXCL3</p>	<p>«цитокин-опосредованный сигналинг», «ответ на вирус», «сигналинг IL-17»</p>
<p>Immune</p>	<p>Кровь - 48 здоровых</p>	<p>Кровь,</p>	<p>JUN, POLRMT,</p>	<p>«белки окислительного фосфорилирования,</p>

system cells from COVID-19 patients display compromised mitochondrial-nuclear expression co-regulation and rewiring toward glycolysis	контролей, 119 пациентов с COVID-19, Эпителий респираторного тракта -157 здоровых контролей, 524 пациентов с COVID-19	эпителий респираторного тракта	JUND	кодируемые ядерной ДНК», «формирование реактивных форм кислорода», «гликолиз», «ответ на вирус связанный с митохондриями», «ответ на интерферон 1-го типа»
COVID-19-specific transcriptomic signature detectable in blood across multiple cohorts	Анализ публично доступных датасетов GSE161731, GSE149689, GSE163151, GSE163668, GSE157103, GSE152418, GSE161918, GSE150728, GSE160351	Первичные мононуклеарные лейкоциты крови, цельная кровь, лейкоциты, моноциты	IGHG1, IGHG3, IGHG4, IGLC2, PTX3, GP9, CMTM5, RUNX1, CDKN1A, CDKN1C	«активация системы комплемента (классический путь)», «производство иммуноглобулина», «гены, регулируемые RUNX1, участвующие в дифференцировке мегакариоцитов и функционировании тромбоцитов», «антимикробный гуморальный ответ», «активация тромбоцитов», «митотическое деление ядра», «гомеостаз многоклеточного организма», «позитивная регуляция пролиферации фибробластов»
SARS-CoV-2 early infection signature identified potential key	Анализ публично доступного датасета GSE147507	Клеточные линии человека, зараженные SARS-COV-	IL1A, CXCL2, TNFAIP3, MAFF, PPP1R15A, MFKBIA, PTX3, CXCL3, CCL20,	«интерфероновый сигналинг», «роль рецепторов в узнавании бактерий и вирусов»

infection mechanisms and drug targets		2	IFIT2, ARRDC3, EREG, ARSE, MAP2K6, DHCR7, UCP2, SLC25A10, VIL1, MCM5, DHCR24, SLC9A3R1, PFN1, TPPP3, DEGS2, RAB26	
The relationship between cytokine and neutrophil gene network distinguishes SARS-CoV-2-infected patients by sex and age	Анализ публично доступного датасета GSE152075	Назофарингальные мазки	CXCR1, CXCR2, IL-1 β , S100A9, ITGAM, DBNL, IL-18RAP, CCR1, GBP2, IRF7, CXCL8, LAMP1, CYBA, GBP2, IRF7, IFI30, CMSP, NMI	«дегрануляция нейтрофилов», «интерлейкиновый сигналинг», «связывание лигандов с рецепторами, сопряженными в G белками», «сигнальные каскады Toll-подобных рецепторов»
System-wide transcriptome damage and tissue identity loss in COVID-19 patients	39 пациентов, погибших от COVID-19 и 22 здоровых контроля	Ткани лимфоузлов, легкого, сердца, печени, почек	ACTA2, FLNA, CD163, CD68, COL1A1, COL1A2, COL6A3, MUC4, MUC5AC, MUC5B, CCL20, CXCL1, CXCL6, CXCL6, MMP1, IFIT3,	«активация макрофагов», «активация моноцитов», «фиброз», «сигналинг интерферона 1-го типа», «цитокиновый сигналинг», «активация системы комплемента»

			ISG15, STAT1, MX1, B2M, CD81, GNAS, HLA.B, HSPA1A, HSPB1, SERPING1	
Dynamic changes in gene-to-gene regulatory networks in response to SARS-CoV-2 infection	Анализ публично доступного датасета GSE147507	Клеточные линии человека, зараженные SARS-COV-2	IFIs, MXs, OASs, TRIMs, IFTMs, IRFs, STATs, CXCL1, CXCL2, CXCL3, CX3CL1, CCL20, IL6, IL1A, IL1B, IL32, CSF2, CSF3	«дифференциальная регуляция продукции цитокинов в макрофагах и Т-хелпер клетках IL-17A и IL-17F», «адгезия и диапедез гранулоцитов», «дифференциальная регуляция продукции цитокинов в кишечных эпителиальных клетках IL-17A и IL-17F», «сигналинг IL-17A и IL-17F в клетках желудочно кишечного тракта»
Detection of SARS-CoV-2 genome and whole transcriptome sequencing in frontal cortex of COVID-19 patients	9 пожилых пациентов с COVID-19 и 8 пожилых контролей, погибших от не связанных с COVID-19 причин	Фронтальные лобные доли	SLC14A1, HIF3A, PAPLN, RGS5, SERGEF, NCL, ZNF622, HBA2, HBB, HBA1, CTB-3601.7	«ответ на окислительный стресс», «транспорт кислорода», «катаболизм пероксида водорода», «активность пероксидазы», «активность антиоксидантов»
Genome-wide bioinformatic analyses predict key host and viral	Анализ публично доступного датасета GSE147507, GSE150316	Культура клеток NHBE, A549, Calu-3, зараженных	PABC1, HNRNP, EIF4B, PTEN, AKT1, AKT2, FOXO3, CREB3, IL32, PDGFB, CSF2,	«IL-1 опосредованный сигналинг», «активация нейтрофилов», «элонгация ДНК», «вирусный карциогенез»

factors in SARS-CoV-2 pathogenesis		кови́дом, зафиксированные в формалине образцы ткани легкого в парафине	IL-7. IL-18. IL-6, SERPINA3, KLF2	
Network perturbation analysis in human bronchial epithelial cells following SARS-CoV2 infection	Анализ публично доступного датасета GSE147507, GSE47962, GSE81909	Культура клеток NHBE, A549, HAE	<i>CSF3, DNAH7, CEACAM7, c17orf67, ANKAR, LOC401109, DBIL5P</i>	«процесс метаболизма коллагена», «метаболизм макромолекул многоклеточным организмом», «расщепление внеклеточного матрикса», «процесс метаболизма многоклеточного организма», «катаболизм коллагена», «катаболический процесс многоклеточного организма», «организация внеклеточного матрикса», «актиновый цитоскелет», «сокращение мышцы», «сократимая часть мышечного волокна», «процесс мышечной системы»
Immune transcriptomes of highly exposed SARS-CoV-2 asymptomatic seropositive versus seronegative	4 группы - 43 асимптоматических против 52 серонегативных индивидов, много взаимодействовавших с больными COVID-19, 4 больных COVID-19 с мягкими симптомами и	Мононуклеарные лейкоциты крови	IL10	«сигналинг TNF- α /NF κ B», «сигналинг mTORC1», «сигналинг IL2-STAT5», «сигналинг TGF β и воспалительный ответ»

individuals from the Ischgl community	4 серонегативных индивида без контактов с больными COVID-19 (Различия были выявлены между серопозитвными симптоматическими и остальными пациентами)			
A Distinct Dexamethasone-Dependent Gene Expression Profile in the Lungs of COVID-19 Patients	12 пациентов с COVID-19 против 8 пациентов с сепсисом или острым респираторным синдромом, вызванным другими причинами	Бронхоальвеолярный лаваж	ADAR, ATF3, IFITM3, IRF1, IRF7, ISG20, TRIM56, ABTB2, CNP, DDIT4, GALNT2, MAP3K14, PRKD2, SPSB1, SQLE, SSBP3, TNK2	«организация хроматина», «адаптивный иммунный ответ», «миграция лейкоцитов», «активация системы комплемента (классический путь)»
Cardiac SARS-CoV-2 infection is associated with pro-inflammatory transcriptomic alterations within the heart	10 погибших пациентов с сердечной инфекцией COVID-19 против 10 погибших пациентов без сердечной инфекции, но с COVID-19	Постмортальный забор ткани левого желудочка сердца	IFI44L, IFT3, TRIM25, NPPB, MB, MYPN, APOD	«ответ на интерферон 1-го типа», «клеточный ответа на интерферон 1-го типа», «сигналинг интерферона 1-го типа»

Blood gene expression predicts intensive care unit admission in hospitalised patients with COVID-19	51 пациент с COVID-19 попавший в ОРИТ против 27 пациентов, не попавших в ОРИТ	Цельная кровь	SNX2, ACAA1, FAM89B, RHETA1, KEAP1, BAIAP2, TRAPPC6A, AGXT, HES1, CDK5R1, SSBP4, UVRAG, LRIF1	«сигналинг рецепторов эпидермального фактора роста», «сигналинг рецептора, активируемого пероксисомным пролифератором альфа», «сигналинг TGFβ» «сигналинг фосфатидил-инозитола», «метаболизм дикарбоксилата»
Transcriptome Analysis Reveals Hub Genes Regulating Autophagy in Patients With Severe COVID-19	4 пациента с тяжелой формой COVID-19, 4 пациента с легкой формой COVID-19, 4 здоровых контроля	Мононуклеарные лейкоциты периферической крови	ATM, CHMP6, EP300, RIPK2, ATP6V0E1, VPS37B, ATP6V1E2, PLEKHM1, ZC3H12A, STAM, MVB12A, RALB	«аутофагия», «процесс, использующий механизм аутофагии»
Transcriptomic analysis of asymptomatic and symptomatic severe Turkish patients in SARS-CoV-2 infection	2 пациента с критически тяжелой формой COVID-19, 2 пациента с легкой формой COVID-19, 2 здоровых контроля	Мононуклеарные лейкоциты периферической крови	IL1B, CXCL8, IL10, JUN, VEGFA, CXCL10, SELL, TLR9, CCR5, IST1H4F	«TNF сигналинг», «сигналинг IL-17», «взаимодействие вирусного белка с цитокином и цитокиновым рецептором», «взаимодействие цитокинового рецептора с цитокином», «коклюш», «NF-κB сигналинг», «системная красная волчанка», «алкоголизм», «формирование внеклеточной ловушки нейтрофилов»
BNT162b2	4 пациента с COVID-	Мононуклеарные лейкоциты периферической крови	OAS1, OAS2, OAS3,	«ответ олиго-аденилат синтетазы на вирус»,

vaccination enhances interferon-JAKSTAT-regulated antiviral programs in COVID-19 patients infected with the SARS-CoV-2 Beta variant	19, получивших вакцину против 5 невакцинированных	рные лейкоциты периферической крови	BIRC5, CSF1, CTSG, CXCL9, CXCL10, GBP1, GBP4, GBP5, GBP6, HLA-H, IFI35, IFI 6, IFIT2, IFITM3, MX1, TRIM21, UBE2L6	«сигналинг интерферона-гамма второго типа», «интерфероновый сигналинг», «остановка клеточного цикла регулируемая p53»
Identifying novel host-based diagnostic biomarker panels for COVID-19: a whole-blood/nasopharyngeal transcriptome meta-analysis	Мета-анализ GSE163151, GSE151161, GSE152641, GSE169687, GSE172450, GSE152075, GSE156063, GSE188678	Цельная кровь и назофарингеальные мазки	-	«активация нейтрофилов», «клеточный цикл», «воспалительный ответ хозяина», «молекулярная сигнатура интерферона», «транспорт кислорода и газов», «связывания железа 2+», «связывание гема», «связывание гемоглобина альфа»
Comparative Upper Respiratory	5 пациентов с тяжелой формой COVID-19, 6 пациентов с легкой или	Назофарингеальные мазки	OAS1, OAS2, OAS3, IFIT5, RSAD2, IFITM1, ISG15,	«ранний ответ на интерферон 1-го типа», «ответ на интерферон 1-го типа», «противовирусный ответ»

<p>Tract Transcriptomic Profiling Reveals a Potential Role of Early Activation of Interferon Pathway in Severe COVID-19</p>	<p>асимптоматической формой COVID-19, 6 здоровых контролей</p>		<p>ZMYND10, SNORD14A, MTND6, MT-CYB</p>	
<p>Transcriptome Analysis of Peripheral Blood Mononuclear Cells in SARS-CoV-2 Naïve and Recovered Individuals Vaccinated With Inactivated Vaccine</p>	<p>5 выздоровевших пациентов с COVID-19 против 20 контролей, никогда не болевших COVID-19, обе группы после вакцинации инактивированной вакциной</p>	<p>Мононуклеарные лейкоциты периферической крови</p>	<p>CXCL8, IL1B, IL10, JUN, VEGFA, IL1A, CXCL1, CSCL2, CD80, CCL20</p>	<p>«сигналинг NF-kB », «TNF сигналинг», «сигналинг NOD-подобного рецептора», «сигналинг рецептора лектина типа C», «сигналинг IL-17»</p>
<p>Temporal transcriptomic analysis using</p>	<p>Анализ публично доступных датасетов GSE156701,</p>	<p>Цельная кровь и Мононуклеа</p>	<p>-</p>	<p>«защитный ответ на вирус», «сигналинг интерферона 1-го типа», «негативная регуляция починки ДНК», «интерферон-</p>

TrendCatcher identifies early and persistent neutrophil activation in severe COVID-19	CNP0001102, GSE161918, GSR171964, GSE169159	рные лейкоциты периферической крови		гамма опосредованный сигналинг»
Transcriptome Analysis of Peripheral Blood Mononuclear Cells Response in Patients with Severe COVID-19 Reveals Crucial Genes Regulating Protein Ubiquitination	4 пациента с тяжелой формой COVID-19 против 4 здоровых контролей	Мононуклеарные лейкоциты периферической крови	<i>ITCH, TRIM21, RNF130, FBXO11, UBE2J1, ASB16, FNIC, FOXA1, YY1, GATA2, MET2A, FOXC1</i>	«биосимметричный процесс изопентенил дисфосфата», «меволанатовый путь», «процесс метаболизма изопентенил дисфосфата», «процесс биосинтеза изопентенил дисфосфата», «транспорт соединений азота»
Identification of key molecules in COVID-19 patients significantly	Анализ публично доступных датасетов GSE157103, GSE156063, GSE152075	Мононуклеарные лейкоциты периферической крови	OAS1, OAS2, OAS3, OASL, HERC6, SERPING1, IFI6, IFI44, IFI44L, CMPK2, RSAD2, EPSTI1, and	«ответ на интерферон», «иммунная цитолитическая активность»

correlated with clinical outcomes by analyzing transcriptomic data			CXCL10	
Deregulation of ceRNA Networks in Frontal Cortex and Choroid Plexus of Brain during SARS-CoV-2 Infection Aggravates Neurological Manifestations : An Insight from Bulk and Single-Cell Transcriptomic Analyses	Анализ публично доступных датасетов GSE182297, GSE157852	назофарингеальные мазки	<i>CCL4, TNFA, IL6, IL1A, CCL2, CXCL2, IFN, CCR1, ITGAV, ITGA6, ITGB7, ITGB3, ITGA2B, ITGA5, ITGA6, ITGA9, ITGA4, ITGAE, ITGA8, CBP, CEBP, NFAT, ATF3, GATA6, HDAC2, TCF12</i>	«протеолиз», «процесс апоптоза», «вирусный процесс», «иммунный ответ», «врожденный иммунный ответ», «интерферон-гамма опосредованный сигналинг», «сигналинг интерферона 1-го типа»
Buffy Coat Transcriptomic Analysis Reveals Alterations in	5 пациентов с мягкой формой COVID-19 против 12 пациентов с тяжелой формой COVID-19	Белые клетки крови	LMNA, SDCBP, IFI16, PTEN, TRIM25, DDX58, PXN, ZYX, CBL, STAT3, RSP24,	«дегрануляция нейтрофилов», «воспалительный ответ», «цитокин опосредованный сигналинг», «врожденный иммунный ответ», «SRP-опосредованный котрансляционный транспорт белка,

Host Cell Protein Synthesis and Cell Cycle in Severe COVID-19 Patients			RSP6, XPO1, KIF14, KIF23, RPL5, RSP18, RSP25, RPL31, RSP2	располагающегося в мембране», «вирусная транскрипция», «nonsense-опосредованный катаболизм мРНК, транскрибируемой в ядре»
Multi-Design Differential Expression Profiling of COVID-19 Lung Autopsy Specimens Reveals Significantly Deregulated Inflammatory Pathways and SFTPC Impaired Transcription	14 пациентов с COVID-19 против 7 контролей без истории респираторных симптомов	Аутопсия тканей легкого	IFI44L, CD163, IFI6, STAT1, RSAD2, IFI44, MX1, IFIT3, CFB, ISG15, IFIT2, OAS2, EPSTI1, FUT1, MFAP4, LTBP4, LRP4, SPRED2, TRPC6, SOX7, IFI44L, IFI44, IFI6, STAT1, RSAD2, MX1, IFIT3, ISG15, IFIT2, OAS2, CD163, EPSTI1, FUT1, MFAP4, LTBP4, LRP4, SPRED2, TRPC6, SOX7	«сигналинг интерферона 1-го типа», «система комплемента», «активация макрофагов»
Long-term perturbation of the peripheral immune	Пациенты, восстанавливающиеся после тяжелой (13), средней (6) и легкой	Цельная кровь	S100A8, HMGB1, AZI2, GZMA, MX1, OAS3, OASL, PF4, GP9, MPL, F13A1,	«трансляция вирусной мРНК», «клеточный ответ на стресс», «процессинг рРНК», «передача сигнала», «гемостаз», «активация тромбоцитов»

system months after SARS- CoV-2 infection	(47) формы COVID-19 через 12, 16 и 24 недели после болезни		S100B	
--	--	--	-------	--

2.4.6 Генетические детерминанты тяжести течения и чувствительности к COVID-19

В то время как полнотранскриптомные исследования позволяют выявить биологические процессы и изменения экспрессии генов, определяющие различные свойства и особенности патогенеза и течения заболеваний, ассоциативные исследования дают информацию о генетических предпосылках тяжести течения заболевания и зачастую подчеркивают роль определенных генов в патогенезе, которая могла бы не быть выявлена при помощи полнотранскриптомных исследований.

На данный момент проведено множество ассоциативных исследований, направленных на выявление потенциальных генетических факторов, влияющих на вероятность заражения/чувствительность и тяжесть течения COVID-19. Однако, выявленные в этих работах однонуклеотидные замены зачастую не воспроизводятся в нескольких различных исследованиях [195].

Одними из первых выявленных в результате полногеномного поиска ассоциаций (GWAS) оказались локусы 3p21.31 и 9q34.2. Локус 3p21.31 содержит 6 генов – *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6* и *XCR1*. Стоит отметить, что ген *SLC6A20* кодирует транскрипционный фактор, позитивно регулирующий экспрессию *ACE2* [196]. Локус 9q34.2 содержит гены, кодирующие группу крови АВ0. Было показано, что группа крови 0 потенциально связана с пониженным риском инфекции COVID-19 [197]. Последующие исследования подтвердили значимость этих локусов для патогенеза COVID-19 [198], [199]. Также были выявлены новые локусы, связанные с повышенным риском инфекции – 6p21.32, содержащий ген *HLA-DBP1*, 21q22.11, содержащий ген *IFNAR2* и Xp22.2, содержащий ген *ACE2* [198]–[200]. Помимо этого, было продемонстрировано, что несинонимичные замены в генах *OAS1*, *DPP9*, *TYK2*, *CCHCR1* и *RAVER1* ассоциированы с более тяжелым течением COVID-19 [201].

Мета-анализ Эшети и др. [195] рассмотрел 46 работ, вышедших к 31 августа

2022 года. В этом мета-анализе были проанализированы ассоциации выявленных ранее полиморфизмов с риском более тяжелого течения COVID-19. Девять полиморфизмов (rs73064426, rs35044562, rs11385942, rs10490770, rs13078854, rs17763742, rs35731912, rs71325088 и rs72893671) в генах *LZTFL1* и *SLC6A20* из локуса 3p21.31, были ассоциированы со значительно (ОШ – 1.8) повышенным риском тяжелого течения COVID-19 [195]. Следует отметить, что все эти полиморфизмы находятся в значительном неравновесии по сцеплению друг с другом (>0.8). Также была подтверждена ассоциация полиморфизмов rs2109069-A и rs12610495-G в локусе 19p13.3, содержащем ген *DPP9*, с повышенным (ОШ 1.3) риском тяжелого течения COVID-19 [195]. Выявленные ранее ассоциации полиморфизмов в генах *IFNAR2*, *ABO*, *OAS1* и *TYK2* с тяжестью течения COVID-19 не были подтверждены в процессе мета-анализа.

Также в этом же мета-анализе были рассмотрены ассоциации выявленных ранее полиморфизмов с риском заражения COVID-19. Полиморфизм rs2531743-G в гене *SLC6A20* из локуса 3p21.31 ассоциирован с пониженным риском (ОШ 0.95) заражения COVID-19. В противоположность этому, полиморфизмы rs73062389-A и rs2271616-T и сцепленные полиморфизмы rs10490770-C/rs11385942-GA в том же локусе ассоциированы с повышенным риском заражения COVID-19. ОШ для rs73062389-A составляет 1.1–1.3, для rs2271616-T составляет 1.1–1.17, для пары сцепленных полиморфизмов ОШ не было оценено. Также были выявлены следующие значимо ассоциированные с повышенным риском заражения COVID-19 локусы: 2q32.3 (*CAVIN2*; *TMEFF2*), 3q12.3 (*CEP97*; *NXPE3*), 3q12.3 (*PDCL3P4*), 6p21.1 (*FOXP4-AS1*), 10q22.3 (*LRMDA* [*C10orf11*, *CDA017*]), 12q24.13 (*OAS1*), 19q13.33 (*PLEKHA4*), 21q22.11 (*IFNAR2*) и Xp22.2 (*ACE2*). ОШ для этих полиморфизмов не оценивался. [195]. Ранее выявленный эффект сцепленных полиморфизмов rs9411378-T и rs879055593-A в локусе 19q34.2, содержащем ген *ABO*, в мета-анализе подтвержден не был.

Исходя из приведенной выше информации, наиболее интересными генами с точки зрения их потенциальной роли в патогенезе COVID-19 являются *LZTFL1* и

SLC6A20, так как полиморфизмы в локусе, содержащим эти гены, имеют наибольшее OR из всех выявленных на данный момент полиморфизмов.

3. МАТЕРИАЛЫ И МЕТОДЫ

3.1 Описание выборок пациентов

3.1.1 Группа пациентов с COVID-19, вызванным заражением 20B вариантом SARS-CoV-2

С первого ноября 2020 года по 25 февраля 2021 200 пациентов были госпитализированы в отделение реанимации и интенсивной терапии (ОРИТ) первого Санкт-Петербургского государственного медицинского университета им. акад. И.П. Павлова с диагнозом COVID-19 по критерию NEWS и NEWS2 [202], [203]. Для включения пациентов в исследование использовались следующие критерии: русская национальность, возраст между 40 и 80 годами, отсутствие хронических коморбидностей, таких как рак, заболевания сердечно-сосудистой системы, сердечная или почечная недостаточность. Из 200 пациентов 44 соответствовали этим критериям. Все пациенты были COVID-19 положительными и им всем был поставлен диагноз «тяжелая пневмония». Инфекция SARS-CoV-2 была подтверждена при помощи RT-PCR. Пациенты наблюдались в течении 30 дней: 14 пациентов погибли и 30 выжили в течении этого периода. Пациенты для RNA-seq были отобраны по дополнительным критериям: возраст от 55 до 80 лет и мужской пол. В результате 8 (3 погибших, 5 выживших) пациентов с тяжелым COVID-19 были отобраны для исследования. Материал для секвенирования был отобран на момент госпитализации в ОРИТ. Клинические характеристики пациентов приведены в табл. 6. Все исследуемые лица дали информированное согласие в соответствии с рекомендациями Хельсинской декларации. Исследование было одобрено этическим комитетом первого Санкт-Петербургского государственного медицинского университета им. акад. И.П. Павлова.

Исходя из опубликованного анализа данных GISAID, на период сбора образцов основным вариантом, распространенным в северо-западном регионе России был вариант 20B (20B – Nextclade, B.1.1 – *Pango Lineage*) [19] – вариант,

являющийся непосредственным предком вариантов Альфа, Бета и Омикрон. Исходя из этого, в дальнейшем в статье эта группа пациентов будет рассматриваться как зараженная 20В вариантом SARS-CoV-2.

3.1.2 Группа пациентов с COVID-19, вызванным заражением Дельта вариантом SARS-CoV-2

С 25 июня 2021 года по 12 августа 2021 66 пациентов были госпитализированы в ОРИТ с диагнозом COVID-19, вызванным штаммом «Дельта» SARS-CoV-2. Диагноз и штамм вируса был определен при помощи RT-PCR. Для включения пациентов в исследование использовались следующие критерии: русская национальность, возраст между 50 и 70 годами, отсутствие хронических коморбидностей, таких как рак, заболевания сердечно-сосудистой системы, сердечная или почечная недостаточность. Из 66 пациентов 20 соответствовали этим критериям. Пациенты наблюдались в течении 30 дней, и в течении этого периода 10 пациентов погибли, а 10 выжили. В результате 10 пациентов (5 погибших и 5 выживших) были отобраны для исследования (табл. 7). Материал для секвенирования был отобран на момент госпитализации в ОРИТ. Все исследуемые лица дали информированное согласие в соответствии с рекомендациями Хельсинской декларации. Исследование было одобрено этическим комитетом первого Санкт-Петербургского государственного медицинского университета им. акад. И.П. Павлова.

3.2 Изоляция первичных мононуклеарных лейкоцитов периферической крови

Первичные мононуклеарные лейкоциты периферической крови (МПК) были изолированы из 8 мл венозной крови с добавлением антикоагулянта ЭДТА при помощи метода Ficoll-Paque (Ficoll-Paque PLUS, GE Healthcare, Чикаго, США) [204]. После центрифугирования МПК были собраны и дважды промыты при помощи фосфатного буфера (pH 7.4) для того, чтобы удалить богатую тромбоцитами фракцию плазмы.

3.3. Выделение и секвенирование РНК

Тотальная РНК была получена из МПК при помощи Тризола (TRIzol Reagent, Thermo Fisher Scientific, Волтам, США) в соответствии с протоколом изготовителя. Полученная РНК была немедленно заморожена при -80 °С.

Количество и качество РНК были оценены при помощи Nanodrop (Thermo Fisher Scientific, Волтам, США) и Agilent 2100 Bioanalyzer с набором RNA 6000 Nano Kit (Agilent Technologies, USA) в соответствии с протоколом изготовителя. Индекс целостности РНК (RIN) во всех образцах составлял более 8. Поли-а фракция была получена из тотальной РНК при помощи олиго-Т магнитных частиц Dynabeads® mRNA Purification Kit (Ambion, USA) в соответствии с протоколом изготовителя. Библиотеки для секвенирования были получены и поли-А фракции РНК при помощи кита NEBNext® Ultra™ II RNA Library Prep Kit (NEB, USA) и NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1) в соответствии с протоколом изготовителя. Концентрация библиотек для секвенирования была оценена при помощи кита Qubit dsDNA HS Assay Kit на флуорометре Qubit 2.0 (Thermo Fisher Scientific, USA). Распределение длинны полученных фрагментов было определено при помощи кита the Agilent High Sensitivity DNA Kit на Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Непарные риды длиной 50 п.н. были отсеквенированы из полученных библиотек при помощи платформы кита TruSeq SBS Kit V3 на платформе Illumina HiSeq1500 (Illumina, San Diego, CA, USA). Результаты секвенирования доступны для скачивания в Gene Expression Omnibus по идентификаторам GSE185863 (сет 1) и GSE272392 (сет 2).

Таблица 6. Данные о пациентах из группы пациентов с COVID-19, вызванным заражением вариантом 20B

Идентификатор пациента	Исход заболевания	Возраст	Пол	Т°С	Оксигенация	Частота дыхательных движений	Категория тяжести повреждения легких по компьютерной томографии (1-4)	Лечение	День смерти
38 34939	Погиб	66	м	37,1	65	26	4	<i>Tocilizumab 400 мг</i> CVVH 1470 мин	28
98 36367	Погиб	72	м	37,5	88	22	2	Ruxolitinib 10 мг в день	19
101 37339	Погиб	63	м	36,9	80	25	4	<i>Tocilizumab 400 мг</i> CVVH 1520 мин	30
114 37483	Выжил	79	м	36,7	92	22	3	<i>Baricitinib 4 мг</i> в день	-
51 35875	Выжил	72	м	38,0	92	22	4	<i>Tocilizumab 400 мг</i> CVVH 1440 мин	-
96 36891	Выжил	59	ж	37,6	86	27	4	<i>Tocilizumab 400 мг</i>	-
99 36444	Выжил	63	ж	36,6	88	21	4	<i>Tocilizumab 400 мг</i>	-

								CVVH мин 1490 мин	
126 37998	ВЫЖИЛ	74	ж	36,8	89	22	4	<i>Varicitinib</i> 4 мг в день	-

Таблица 7. Данные о пациентах из группы пациентов с COVID-19, вызванным заражением Дельта вариантом

Идентификатор пациента	Исход заболевания	Возраст	Пол	Т°С	Оксигенация	Частота дыхательных движений	Категория тяжести повреждения легких по компьютерной томографии (1-4)	Лечение	День смерти
222 25432	Погиб	45	м	37.3	97	16	2	Тocilizumab 400 мг Олокизумаб 64 мCVVH 1500 мин	26
184 23386	Погиб	43	м	36,7	89	22	3	Тocilizumab 400 мг CVVH 1440 мин	9
172 23580	Погиб	48	м	36.9	90	24	4	Тocilizumab 400 мг CVVH 1470 мин	9
163 23385/C2 021	Погиб	61	м	36.8	70	23	3	Тocilizumab 400 мг	4
213	Погиб	66	м	36.70	97	18	2	Олокизумаб	32

Идентификатор пациента	Исход заболевания	Возраст	Пол	Т°С	Оксигенация	Частота дыхательных движений	Категория тяжести повреждения легких по компьютерной томографии (1-4)	Лечение	День смерти
24802/C2 021								64 мг	
191 24043/C2 021	Погиб	73	м	36.7	89	19	3	Олокизумаб 64 мг Tocilizumab 400 мг	12
201 25023/C2 021	Выжил	43	м	38.6	90	17	2	Tocilizumab 400 мг CVVH min 1470 мин	17
171 23635/C2 021	Выжил	45	м	37.2	86	24	3	Tocilizumab 400 мг CVVH 1470 мин	16
207 25246/C2 021	Выжил	58	м	37.1	94	17	2	Tocilizumab 400 мг CVVH 1470 мин	10

Идентификатор пациента	Исход заболевания	Возраст	Пол	Т°С	Оксигенация	Частота дыхательных движений	Категория тяжести повреждения легких по компьютерной томографии (1-4)	Лечение	День смерти
223 26632/C2 021	Выжил	59	м	37.1	75	30	2	Tocilizumab 400 мг CVVH 1440 мин	18
166 23455/C2 021	Выжил	60	м	37.8	75	24	3	Tocilizumab 200 мг	10
183 24049/C2 021	Выжил	63	м	37.5	90	16	4	Tocilizumab 400 мг	34

3.4 Обработка и биоинформатический анализ данных

Для определения списка дифференциально экспрессирующихся генов (ДЭГ) полученные данные секвенирования были обработаны при помощи трех различных методов обработки данных секвенирования (Далее – метод обработки данных секвенирования – пайплайн). Сравнительный обзор пайплайнов представлен на рис. 1. Контроль качества полученных файлов в формате FastQ во всех трех случаях производился при помощи FastQC v0.11.9. Для дальнейшего анализа были отобраны дифференциально экспрессирующиеся гены, выявленные во всех трех пайплайнах одновременно.

3.4.1 Набор биоинформатических инструментов 1

В пайплайне 1 не производилось подрезание по качеству и удаление последовательностей праймеров. Выравнивание ридов на геном GRCh38 производилось при помощи HISAT2 (v2.6.1b) [54]. Подсчет числа ридов на ген производился при помощи featureCounts [205]. Выявление дифференциально экспрессирующихся генов производилось при помощи команды «DESeq» из пакета R (v.4.0.3) “DESeq2” [71]. В процессе анализа была использована модель экспрессия ~ возраст + выживание/не выживание + RIN (RNA integrity index) + время от сбора ткани до выделения РНК (ttr). В качестве дифференциально экспрессирующихся генов были выбраны те, для которых кратное изменение (FC) составляло более 1.5, и p.value теста “DESeq” с поправкой на множественное тестирование FDR <0.05.

3.4.2 Набор биоинформатических инструментов 2

Удаление последовательностей праймеров производилось при помощи Cutadapt (v3.4). Выравнивание ридов на геном GRCh38 производилось при помощи HISAT2 (v2.6.1b) [54]. Подсчет числа ридов на ген производился при помощи команды «htseq-count» из программы HTSeq (v.0.6.1). Выявление дифференциально экспрессирующихся генов производилось при помощи команды «DESeq» из пакета

R (v.4.0.3) “DESeq2” [71]. В процессе анализа была использована модель экспрессия ~ возраст + выживание/не выживание + RIN (RNA integrity index) + время от сбора ткани до выделения РНК (ttr). В качестве дифференциально экспрессирующихся генов были выбраны те, для которых кратное изменение (FC) составляло более 1.5, и p.value теста “DESeq” с поправкой на множественное тестирование FDR <0.05.

3.4.3 Набор биоинформатических инструментов 3

Подрезание по качеству и удаление последовательностей праймеров производилось при помощи AdapterRemovalV2 [206]. Для выравнивания ридов использовался референсный транскриптом, сгенерированный из референсного генома GRCH38 и аннотации генов GRCH38.92 при помощи команды «rsem-prepare-reference» из пакета “R” “RSEM” [64] с опцией –star для того, чтобы также сгенерировать аннотацию и индексы для STAR [56]. Выравнивание на референсный транскриптом (RSEM) и референсный геном (STAR) [56], а также подсчет симулированного числа ридов (псевдо-каунт) на ген при помощи команды rsem-calculate-expression из пакета “R” “RSEM”. Полученные псевдо-каунты были нормализованы при помощи метода TMM из пакета R “edgeR” [77] и метода CPM из команды “voom” пакета R “limma” [68]. Выявление дифференциально экспрессирующихся генов производилось при помощи команд “voom”, “lmer” и “eBayes” из пакета R “limma” [68]. В процессе моделирования «limma» была использована модель экспрессия ~ возраст + выживание/не выживание + RIN (RNA integrity index) + время от сбора ткани до выделения РНК (ttr). Недостающие значения характеристик RIN и «ttr» были импутированы при помощи метода «predictive mean matching» и пакета R “mice”. В качестве дифференциально экспрессирующихся генов были выбраны те, для которых кратное изменение (FC) составляло более 1.5, и p.value теста “limma” с поправкой на множественное тестирование FDR <0.05.

3.4.4 Обогащение терминами Gene Ontology

Обогащение терминами Биологические процессы из онтологии терминов Gene Ontology (GO enrichment) производился при помощи аппов ClueGO [207] v. 2.5.5 и Cluepedia [208] v. 1.5.5 для Cytoscape v. 3.6.1. Значимо обогащенные термины были выявлены при помощи правостороннего гипергеометрического теста с поправкой на множественное тестирование «FDR» ($p\text{-value} < 0.01$). Термины были объединены в группы на основании общего числа генов между терминами ($> 40\%$ общих генов). В процессе обогащения были отобраны исключительно термины от 3-го до 8-го уровня в онтологии Gene Ontology, для которых было выявлено не менее 3 ассоциированных ДЭГ и общее число ДЭГ составляет не менее 4% всех генов, ассоциированных с термином. В качестве бэкграунда для обогащения был использован список всех генов, для которых хотя бы в одном образце было выявлено не менее трех картированных ридов.

3.5 Описание подтверждения на независимой выборке

3.5.1 Сбор выборки пациентов для подтверждения полученных результатов

Для подтверждения полученных результатов были собраны дополнительные выборки пациентов ОРИТ первого Санкт-Петербургского государственного медицинского университета им. акад. И.П. Павлова с диагнозом COVID-19. Были собраны две группы пациентов – первая в период наибольшей распространенности варианта 20В (с 1 ноября 2020 по 25 февраля 2021), а вторая в период наибольшей распространенности варианта Дельта (с 1 июля 2021 по 31 августа 2021). Образцы (плазма крови и первичные мононуклеарные лейкоциты периферической крови) были собраны в день поступления пациентов в ОРИТ, а также через 7 дней после этого. В общей сложности, с учетом исходов – пациенты были подразделены на подгруппы – 28 пациентов, зараженных 20В вариантом SARS-CoV-2, из которых 13 погибли и 15 выжили по истечению 30 дней; а также 39 пациентов, зараженных Дельта вариантом SARS-CoV-2, из которых 21 погибли и 18 выжили по истечению 30 дней. Все пациенты удовлетворяли следующим критериям включения – возраст

между 40 и 80 годами, отсутствие хронических коморбидностей, таких как рак, цереброваскулярные заболевания, сердечная или почечная недостаточность. Однако пациенты с ожирением и гипертензией не исключались из выборки, так как это распространенные осложняющие факторы среди пациентов с тяжелым COVID-19. Не один из включенных пациентов не получал терапию, влияющую на уровень липидов в период до поступления в ОРИТ и в процессе терапии в ОРИТ.

3.5.2 Сбор плазмы крови

Плазма крови была изолирована из периферической венозной крови в вакуумные контейнеры с ЭДТА при помощи центрифугирования в течение 20 мин при 3000 g. Центрифугирование было проведено не более чем через 20 минут после забора крови. Образцы плазмы хранились при -80°C .

3.5.3 Липидный профиль плазмы

Концентрации общего холестерина (ТС), липопротеинов высокой плотности (ЛПВП), липопротеинов низкой плотности (ЛПНП) и триглицеридов в плазме крови были определены при помощи биохимического анализатора OLYMPUS AU400.

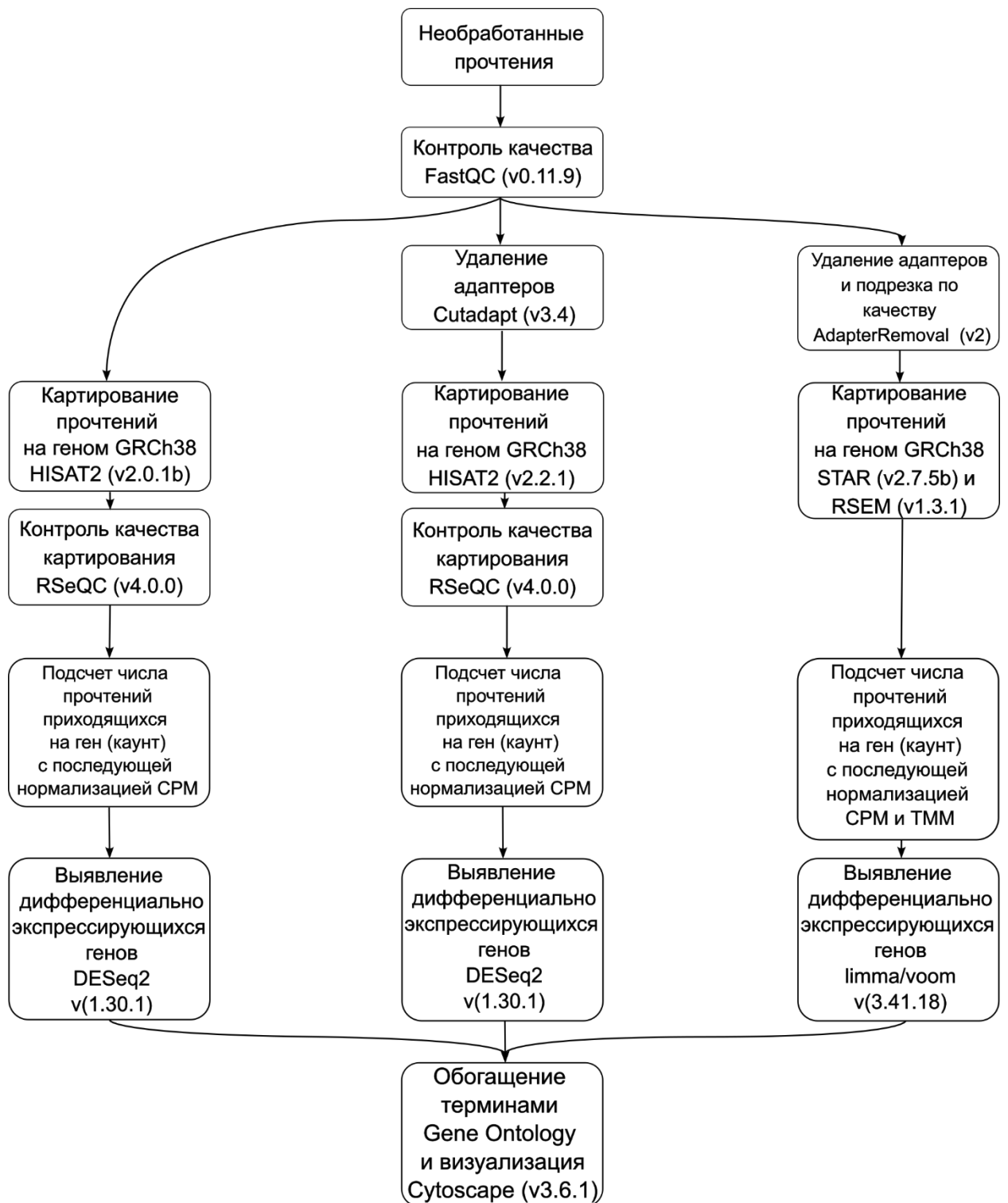


Рисунок 1. Схема трех использованных методов обработки данных секвенирования

4. РЕЗУЛЬТАТЫ

В рамках работы было проведено транскриптомное профилирование на двух группах пациентов – группа пациентов, госпитализированных в ОРИТ с первого ноября 2020 года по 25 февраля 2021 года, и группа пациентов, госпитализированных в ОРИТ с 25 июня 2021 года по 12 августа 2021 года. На основании данных о распространенности вариантов по территориям на момент госпитализации пациентов, мы можем сделать уверенное предположение, что первая группа пациентов была инфицирована вариантом 20B SARS-CoV-2. Заражение пациентов из группы два вариантом SARS-CoV-2 Дельта было подтверждено при помощи РТ-ПЦР. В дальнейшем группы будут называться «группа зараженных вариантом 20B» и «группа зараженных вариантом Дельта». Все пациенты находились в клинически тяжелом состоянии на момент госпитализации. Подробное описание клинических характеристик групп пациентов приведено в табл. 6 и 7 (в разделе «Материалы и Методы»).

4.1 Транскриптомное профилирование в группе зараженных вариантом 20B

4.1.1 Секвенирование РНК

В процессе исследования были получены образцы крови от 200 пациентов ОРИТ с SARS-CoV-2. Из 200 образцов, для транскриптомного секвенирования согласно критериям, описанным в разделе Материалы и Методы, были отобраны 8. По исходу 30 дней 5 пациентов из отобранных выжили и 3 пациента погибли. Из отобранных образцов были получены первичные мононуклеарные лейкоциты периферической крови. Из первичных мононуклеарных лейкоцитов периферической крови была выделена мРНК, и проведено широкомасштабное параллельное секвенирование кДНК в двух технических повторах на пациента. Коэффициент корреляции Спирмана для по-генной экспрессии между техническими повторами составила от 0.86 до 0.89.

4.1.2 Выявление дифференциальной экспрессии между группами при помощи DESeq2 и limma/voom

Полученные данные секвенирования были обработаны при помощи трех различных пайплайнов, подробное описание которых находится в разделе “материалы и методы”. В процессе анализа описанные далее шаги были реализованы в каждом из пайплайнов с использованием различных биоинформатических инструментов. Полученные сиквенсы были выровнены на геномную сборку GRCH38. Затем числа ридов (каунты) по генам были использованы для создания линейной модели для выявления дифференциальной экспрессии между группами. Значимо дифференциально экспрессированными считались гены, для которых значение модуля кратного изменения (FC) составляло более 1.5 и значение $p.value$ с поправкой на множественное тестирование составляло < 0.05 . В общей сложности, в Пайплайне 1 было выявлено 1038 дифференциально экспрессирующихся генов (ДЭГ) (488 с повышенной и 550 с пониженной экспрессией в группе погибших по сравнению с выжившими); в Пайплайне 2 было выявлено 866 ДЭГ (399 с повышенной и 516 с пониженной экспрессией в группе погибших по сравнению с выжившими); в Пайплайне 3 было выявлено 516 ДЭГ (188 с повышенной и 328 с пониженной экспрессией в группе погибших по сравнению с выжившими). Стоит отметить, что в двух экспериментах были использованы разные критерии для отбора ДЭГ – в группе зараженных вариантом 20В использовалась поправка на множественное тестирование (FDR), а в группе зараженных вариантом Дельта нет.

4.1.3 Анализ обогащения терминов Gene Ontology

Для того, чтобы систематически выявить биологические процессы, влияющие на определение течения COVID-19, было проведено обогащение терминов Gene Ontology. Затем значимо обогащенные термины были определены в группы на основании общих ДЭГ. При обогащении ДЭГ, выявленных при помощи Пайплайна 1 были выявлены следующие значимо обогащенные термины: low-

density lipoprotein particle receptor activity (GO:0005041), leukocyte differentiation (GO:0002521), cargo receptor activity (GO:0038024). При обогащении ДЭГ, выявленных при помощи Пайплайна 2, были выявлены следующие значимо обогащенные термины: low-density lipoprotein particle receptor activity (GO:0005041), leukocyte differentiation (GO:0002521), cargo receptor activity (GO:0038024), lipoprotein particle receptor activity (GO:0030228), positive regulation of cell adhesion (GO:0045785), leukocyte activity (GO:0005041), leukocyte differentiation (GO:0002521), cargo receptor activity (GO:0038024). Стоит отметить, что в результате обогащения терминов ДЭГ, выявленных при помощи Пайплайна 3, термины были собраны в две группы, не связанные между собой общими генами. В группу А вошли следующие термины: low-density lipoprotein particle receptor activity (GO:0005041), leukocyte differentiation (GO:0002521), cargo receptor activity (GO:0038024), dendritic cell apoptotic process (GO:0097048), myeloid leukocyte differentiation (GO:0002573), mononuclear cell differentiation (GO:1903131), regulation of dendritic cell apoptotic process (GO:2000668), dendritic cell differentiation (GO:0097028)), тогда как в группа Б была представлена термином purinergic nucleotide receptor signaling pathway (GO:0035590). Сети генов и терминов для обогащений ДЭГ, полученных в каждом из пайплайнов представлены на рис. 2, 3 и 4.

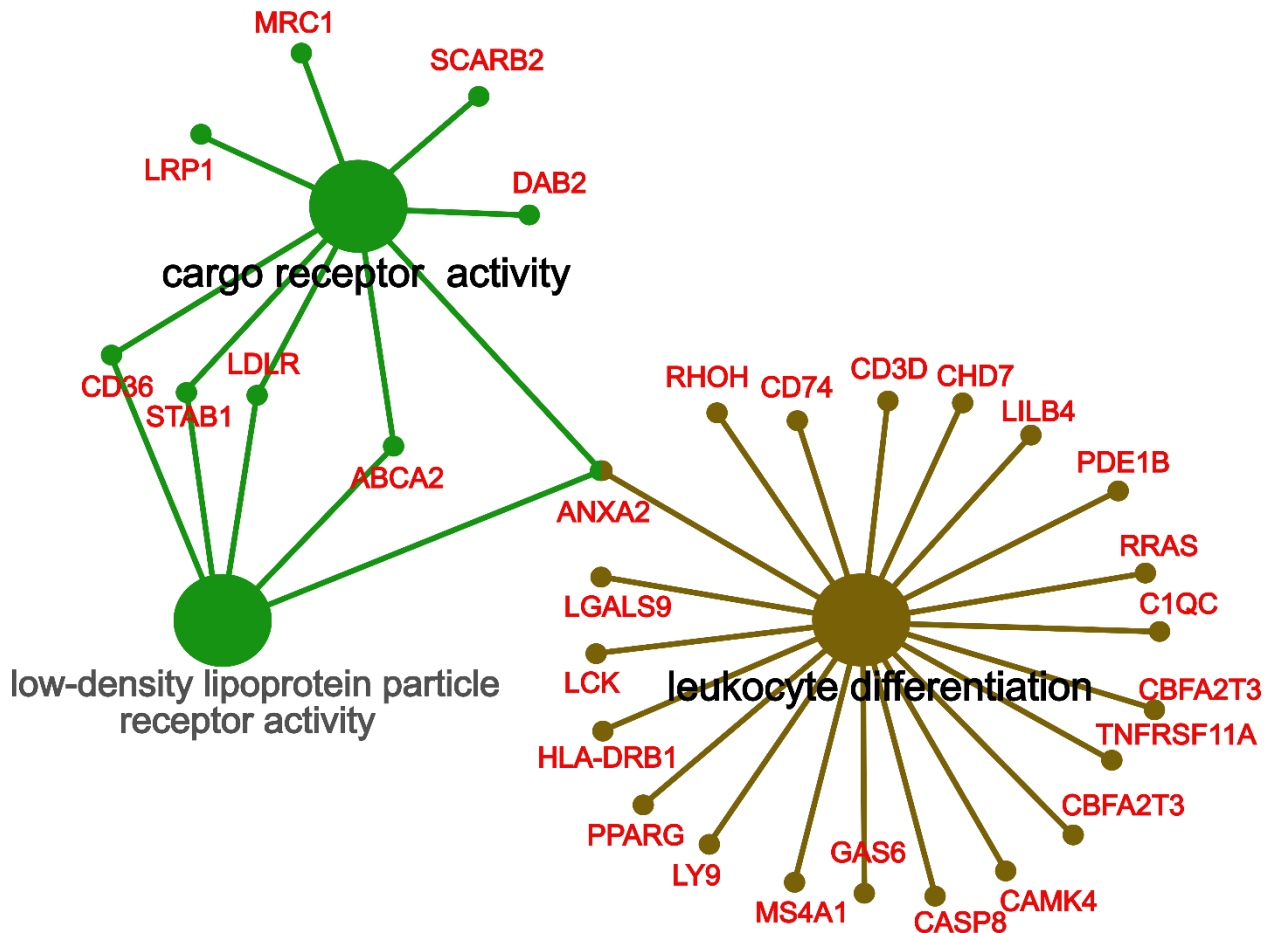


Рисунок 2. Сеть генов и терминов, построенная на основании обогащения терминов Gene Ontology в группе пациентов, зараженных вариантом 20B SARS-CoV-2, используя ДЭГ, полученные в Пайплайне 1. Цвет термина означает принадлежность группе, сформированной на основании общих ДЭГ. Термин с наибольшим *p*-value в каждой группе выделяется более темным черным шрифтом. Красный шрифт обозначает что узел является геном.

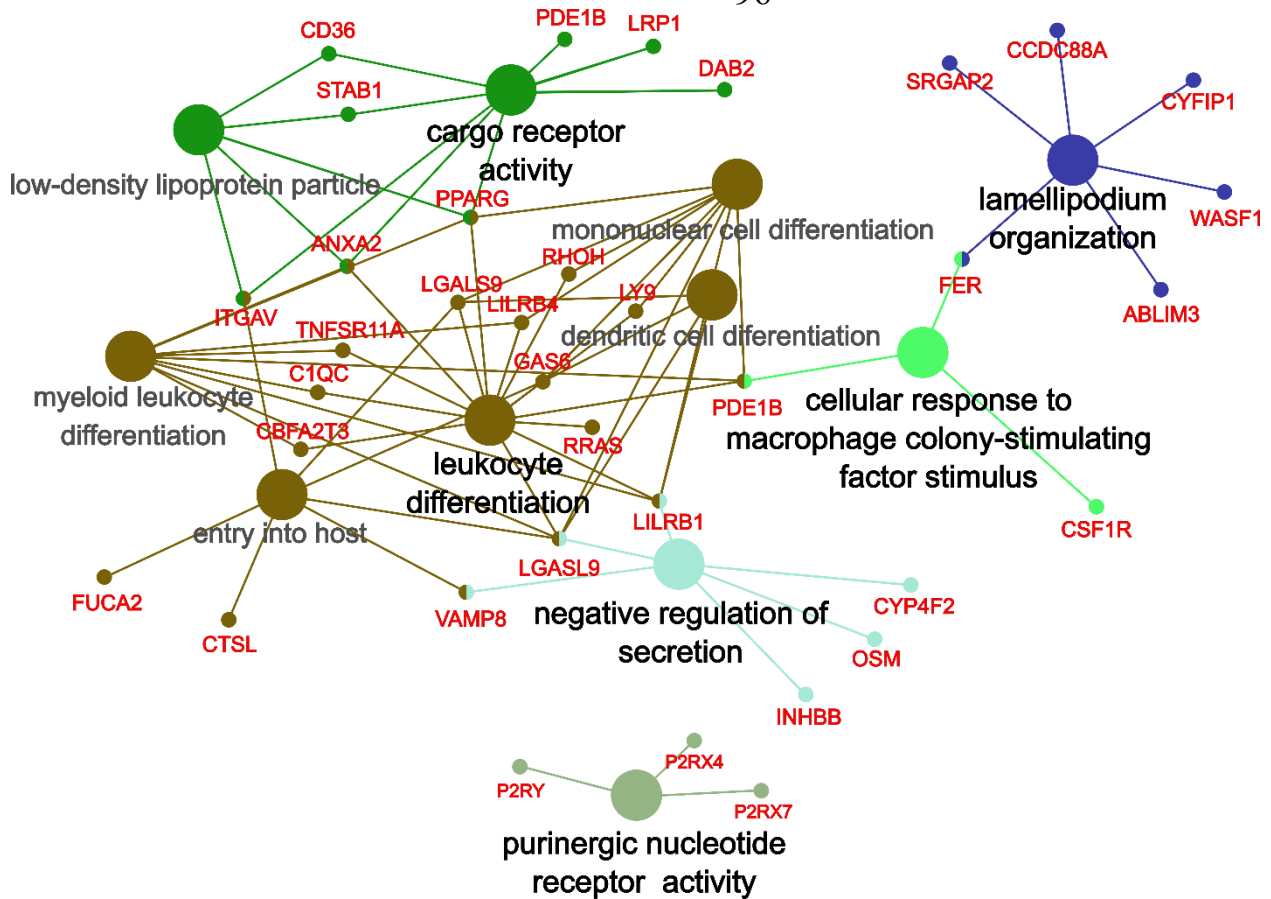


Рисунок 3. Сеть генов и терминов, построенная на основании обогащения терминов Gene Ontology в группе пациентов, зараженных 20B вариантом SARS-CoV-2, используя ДЭГ, полученные в Пайплайне 2. Цвет термина означает принадлежность группе, сформированной на основании общих ДЭГ. Термин с наибольшим *p*-value в каждой группе выделяется более темным черным шрифтом. Красный шрифт обозначает что узел является геном.

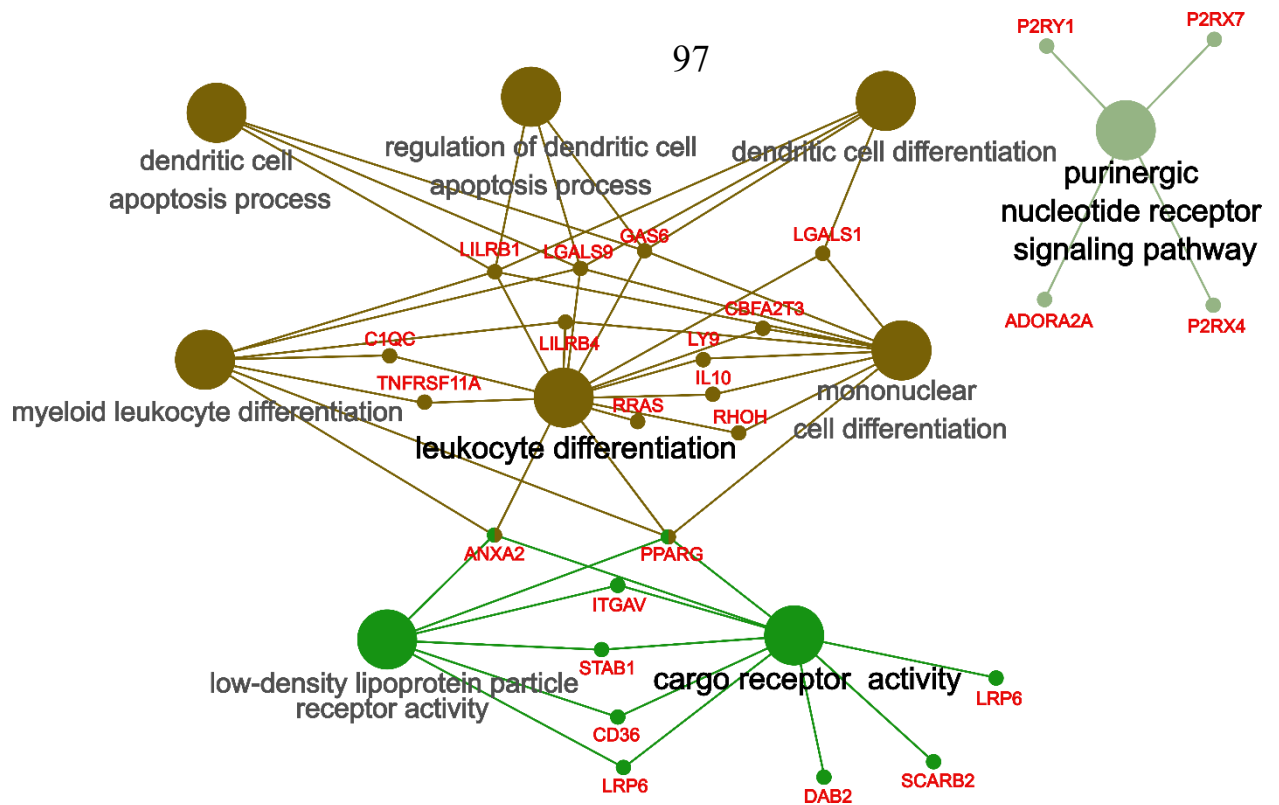


Рисунок 4. Сеть генов и терминов, построенная на основании обогащения терминов Gene Ontology в группе пациентов, зараженных 20B вариантом SARS-CoV-2, используя ДЭГ, полученные в Пайплайне 3. Цвет термина означает принадлежность группе, сформированной на основании общих ДЭГ. Термин с наибольшим *p*-value в каждой группе выделяется более темным черным шрифтом. Красный шрифт обозначает что узел является геном.

Сравнение терминов, выявленных при использовании разных пайплайнов показало, что термины *low-density lipoprotein particle receptor activity* (GO:0005041), *leukocyte differentiation* (GO:0002521), *cargo receptor activity* (GO:0038024) выявляются как обогащенные ДЭГ вне зависимости от используемых биоинформатических инструментов.

4.1.4 Выявление общих ДЭГ между тремя пайплайнами и обогащение терминами Gene Ontology

Используя гены, выявленные как дифференциально экспрессирующиеся во всех трех пайплайнах, была построена диаграмма Венна с учетом направления изменения экспрессии (рис. 5). В общей сложности был выявлен 361 общий ДЭГ, из которых 248 имели повышенную экспрессию, а 113 - пониженную у выживших пациентов по сравнению с погибшими. Далее было проведено обогащение терминами Gene Ontology с использованием этих генов (рис. 6). Три вышеупомянутых термина *density lipoprotein particle receptor activity* (GO:0005041),

leukocyte differentiation (GO:0002521), cargo receptor activity (GO:0038024) были также выявлены как значимо обогащенные в процессе этого анализа. Используя построенную сеть генов и терминов, были выявлены ключевые гены, определенные как гены, связанные с не менее чем тремя различными значимо обогащенными терминами. Такими генами оказались *STAB1*, *PPARG*, *CD36*, *ITGAV* и *ANXA2* (Таблица m). Стоит отметить, что экспрессия всех этих генов понижена в группе выживших по сравнению с группой погибших.

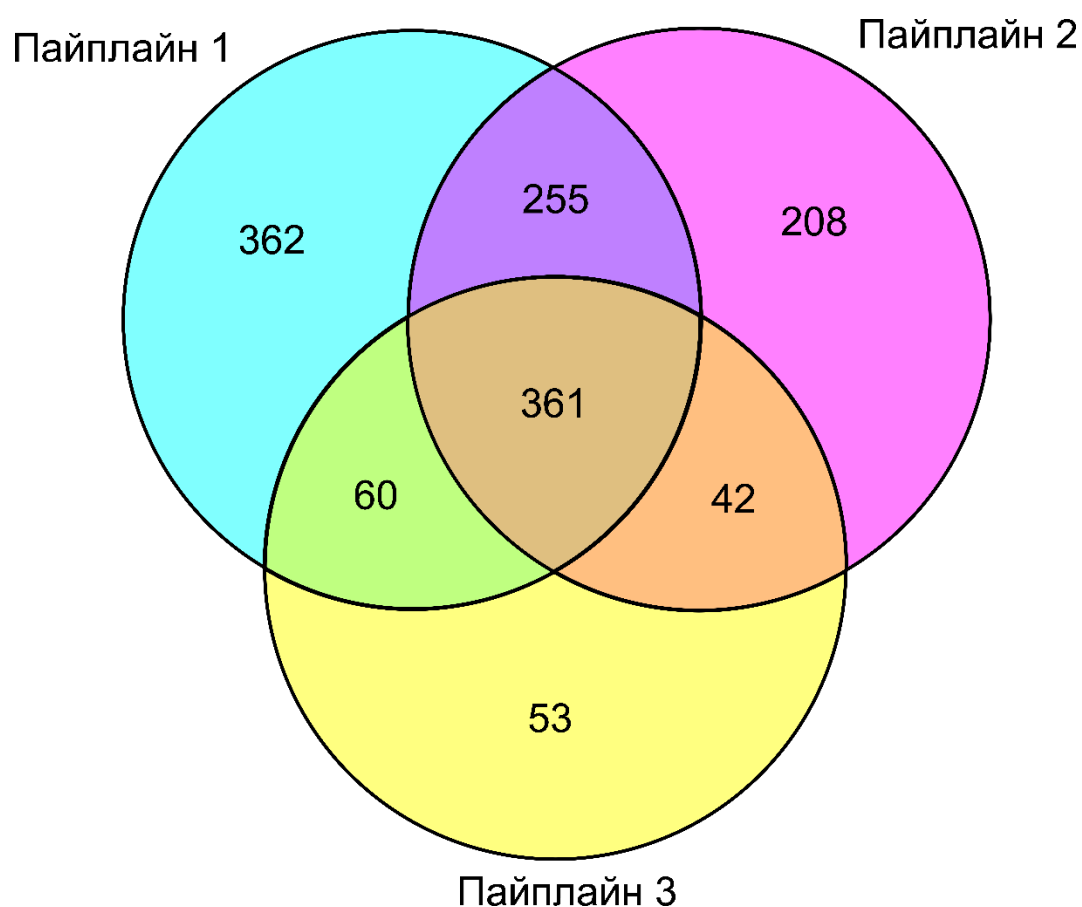


Рисунок 5. Диаграмма Венна для дифференциально экспрессирующихся генов между группой выживших пациентов и группой погибших пациентов, зараженных вариантом 20B SARS-CoV-2 для всех трех использованных пайплайнов.

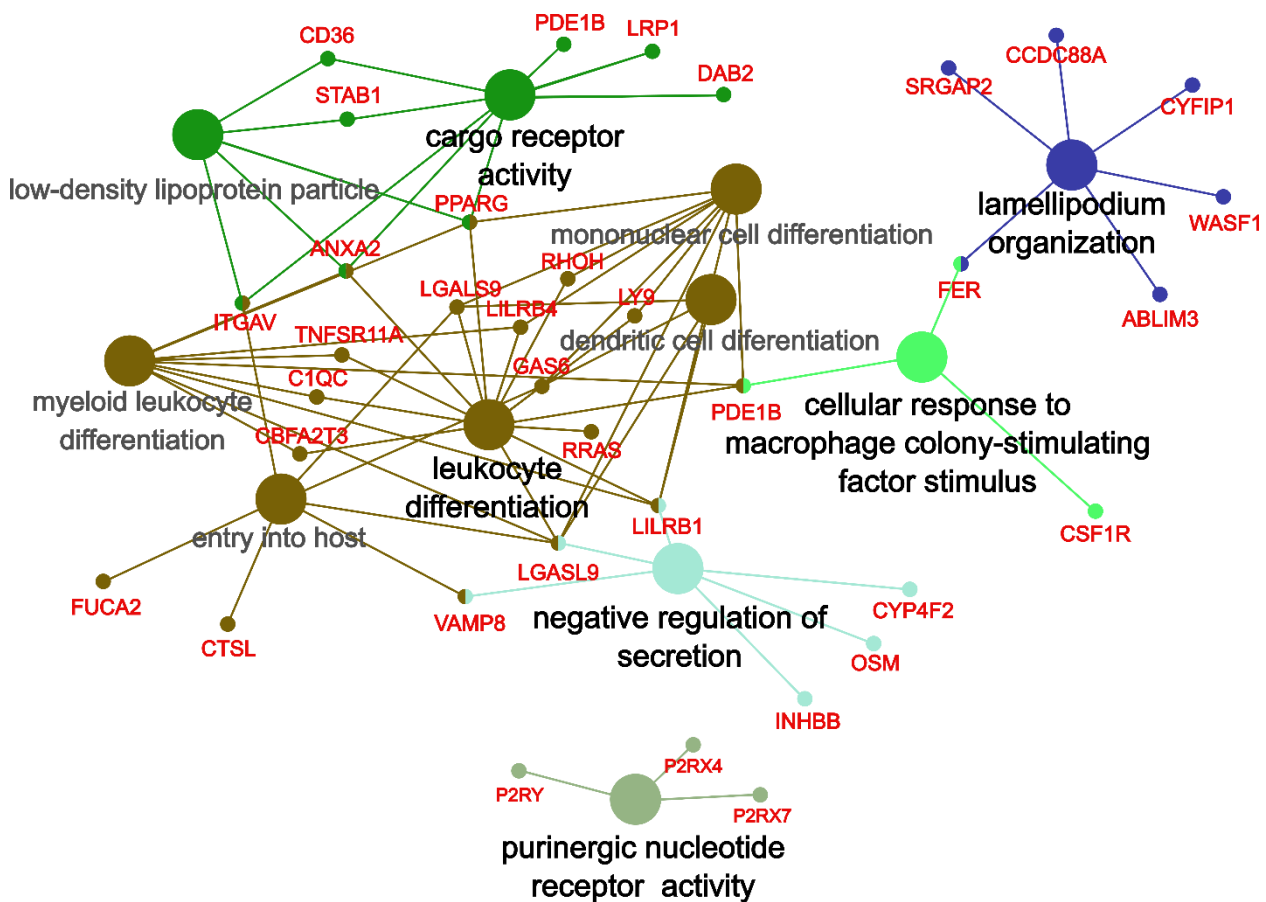


Рисунок 6. Сеть генов и терминов, построенная на основании обогащения терминов Gene Ontology ДЭГ, общих между всеми пайплайнами. Цвет обозначает принадлежность терминов к одной группе на основании общих ДЭГ. Цвет термина означает принадлежность группе, сформированной на основании общих ДЭГ. Термин с наибольшим p -value в каждой группе выделяется более темным черным шрифтом. Красный шрифт обозначает что узел является геном.

4.2 Транскриптомное профилирование в группе зараженных вариантом Дельта

4.2.1 Секвенирование РНК

В процессе исследования были получены образцы крови от 66 пациентов ОРИТ с SARS-CoV-2. У всех пациентов при помощи RT-PCR было подтверждено заражение штаммом коронавируса SARS-CoV-2 Дельта. Из 66 образцов для транскриптомного секвенирования, согласно критериям, описанным в разделе “материалы и методы”, были отобраны 10. По исходу 30 дней 5 пациентов из отобранных выжили и 5 пациентов погибли. Из отобранных образцов были получены первичные моноклеарные лейкоциты периферической крови. Из первичных моноклеарных лейкоцитов периферической крови была выделена мРНК, и проведено широкомасштабное параллельное секвенирование кДНК.

4.2.2 Выявление дифференциальной экспрессии между группами при помощи DESeq2 и limma/voom

Полученные данные секвенирования были обработаны при помощи трех различных пайплайнов, подробное описание которых приведено в разделе “материалы и методы”. В процессе анализа описанные далее шаги были реализованы в каждом из пайплайнов с использованием различных биоинформатических инструментов. Полученные сиквенсы были выровнены на геномную сборку GRCH38. Затем числа ридов (каунты) по генам были использованы для создания линейной модели для выявления дифференциальной экспрессии между группами. При использовании вышеупомянутых критериев отбора значимых генов, малое число генов было выявлено как дифференциально экспрессирующиеся, в связи с чем было принято решение на этом этапе определить дифференциально экспрессирующиеся гены как имеющие значение модуля кратного изменения (FC) составляло более 1.5 и значение $p.value$ без поправки на множественное тестирование составляло < 0.05 . Исходя из этих критериев, в

Пайплайне 1 было выявлено 429 ДЭГ (292 с повышенной и 156 с пониженной экспрессией в группе погибших по сравнению с выжившими). В Пайплайне 2 было выявлено 345 ДЭГ (266 с повышенной и 79 с пониженной экспрессией в группе погибших по сравнению с выжившими). В Пайплайне 3 было выявлено 306 ДЭГ (191 с повышенной и 115 с пониженной экспрессией в группе погибших по сравнению с выжившими).

4.2.3 Анализ обогащения терминов Gene Ontology

Было проведено обогащение терминами Gene Ontology, термины были сгруппированы на основании общих ДЭГ, и были построены сети генов и терминов (рис. 7, 8, 9). Термины считались значимо обогащенными по результатам одностороннего гипергеометрического теста с p -value с поправкой Бонферрони на мн. Тестирование < 0.05 . Большинство значимо обогащенных терминов во всех трех пайплайнах относились к процессам воспаления, врожденного иммунитета и вирусной репликации. В пайплайне таковые термины были представлены терминами GO:0070106 (interleukin-27-mediated signaling pathway), GO:0071639 (positive regulation of monocyte chemotactic protein-1 production), GO:0071605 (monocyte chemotactic protein-1 production), GO:0045071 (negative regulation of viral genome replication) и GO:0035456 (response to interferon-beta). В случае пайплайна 2, термины, относящиеся к системе врожденного иммунитета и регуляции вирусной репликации, были наиболее значимо обогащены (имели наименьшие p -value), включая GO:0002227 (innate immune response in mucosa), GO:0046597 (negative regulation of viral entry into host cell), GO:0006956 (complement activation), GO:0061844 (antimicrobial humoral immune response mediated by antimicrobial peptide), и GO:0045071 (negative regulation of viral genome replication). В случае пайплайна 3 большинство генов также относились к иммунному ответу: GO:0006958 (complement activation, classical pathway), GO:0039530 (MDA-5 signaling pathway), GO:0006956 (complement activation), и GO:0060339 (negative regulation of type I interferon-mediated signaling pathway).

4.2.4 Выявление общих ДЭГ между тремя пайплайнами и обогащение терминами Gene Ontology

Используя гены, выявленные как дифференциально экспрессирующиеся во всех трех пайплайнах, была построена диаграмма Венна с учетом направления изменения экспрессии (рис. 10). Был идентифицирован набор из 75 генов (59 с повышенной и 16 с пониженной экспрессией в группе погибших по сравнению с выжившими), выявленных как дифференциально экспрессированные во всех трех пайплайнах с учетом направления изменения экспрессии. Далее было проведено обогащение терминами Gene Ontology с использованием этих генов. В общей сложности было идентифицировано 14 значимо обогащенных терминов, 13 из которых связаны с иммунной системой.

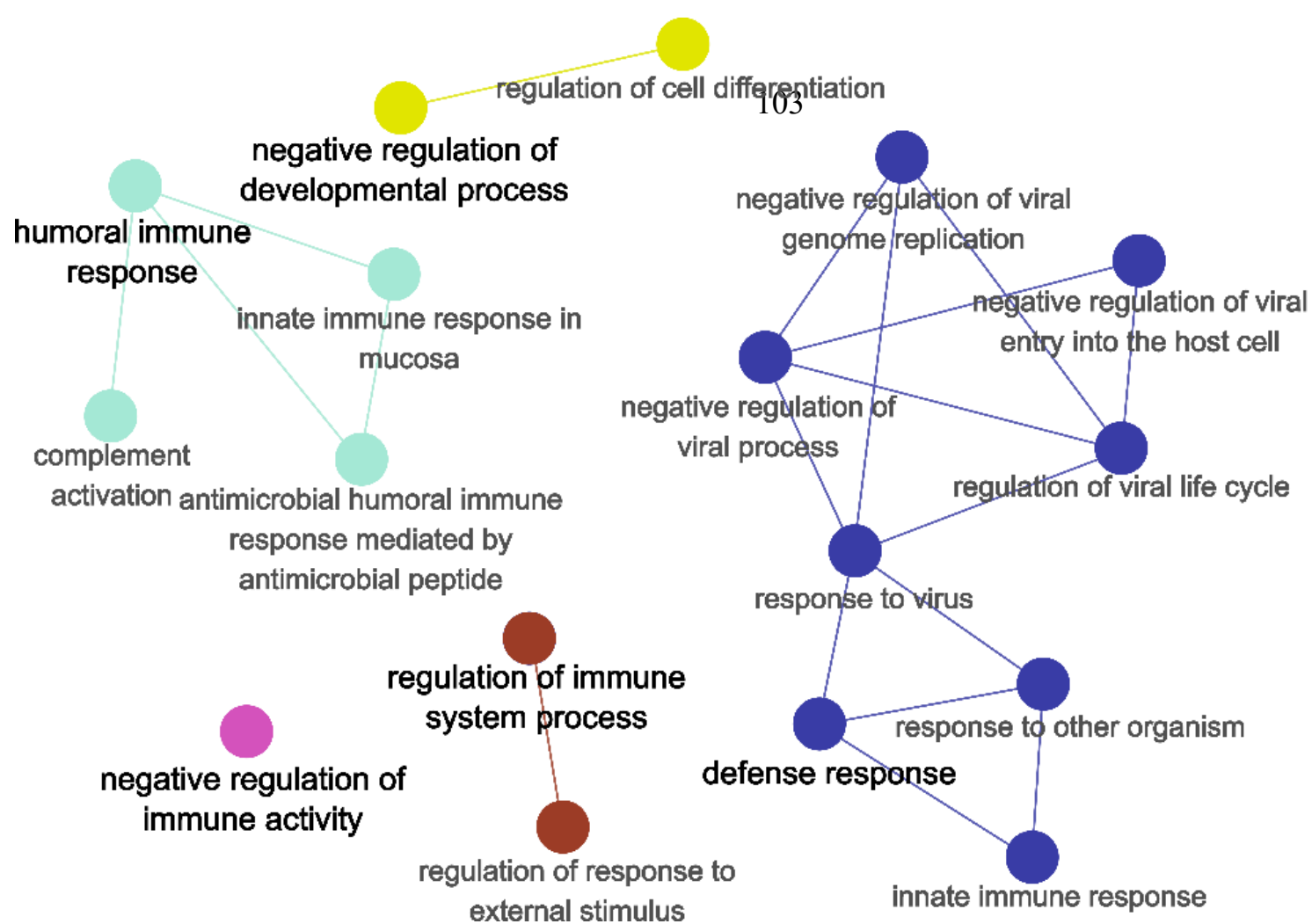


Рисунок 7. Сеть генов и терминов, построенная на основании обогащения терминов Gene Ontology в группе пациентов, зараженных вариантом SARS-CoV-2 Дельта, используя ДЭГ, полученные в Пайплайне 1. Цвет термина означает принадлежность группе, сформированной на основании общих ДЭГ. Термин с наибольшим p.value в каждой группе выделяется более темным черным шрифтом.

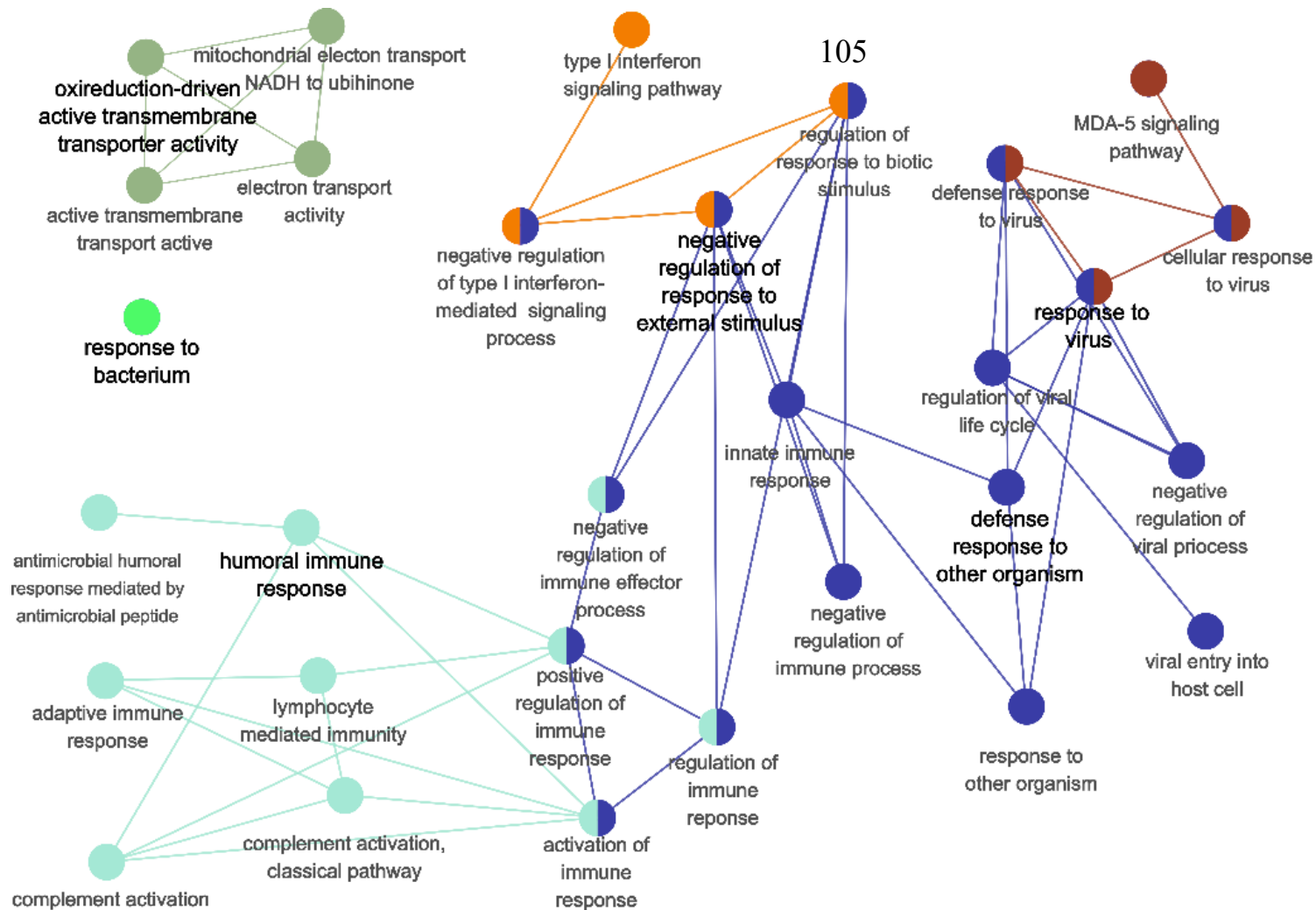


Рисунок 9. Сеть генов и терминов, построенная на основании обогащения терминов Gene Ontology в группе пациентов, зараженных вариантом SARS-CoV-2 Дельта, используя ДЭГ, полученные в Пайплайне 3. Цвет термина означает принадлежность группе, сформированной на основании общих ДЭГ. Термин с наибольшим p.value в каждой группе выделяется более темным черным шрифтом.

На основании результатов обогащения была построена сеть генов и терминов (рис. 11). На основании общих ДЭГ было выделено 5 групп терминов. Группа 1 состоит из 6 терминов – GO:0050777 (negative regulation of immune response, the “main term” with the highest significance/lowest p value), GO:0002709 (regulation of T cell mediated immunity), GO:0002697 (regulation of immune effector process), GO:0002698 (negative regulation of immune effector process), GO:0002449 (lymphocyte mediated immunity) и GO:0002460 (adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains). В дальнейшем, на основании термина с наименьшим p.value в этой группе и пересечением терминов по смыслу эта группа будет называться группой негативной регуляции иммунного ответа. Стоит отметить, что все ДЭГ, относящиеся к этой группе, имеют повышенную экспрессию у выживших пациентов по сравнению с погибшими. Группа 2 состоит из 3 терминов – GO:0048525 (negative regulation of viral process), GO:0051607 (defense response to virus), и GO:1903900 (regulation of viral life cycle). В дальнейшем эта группа будет называться группой негативной регуляции вирусных процессов.

Группа три состоит из 2 терминов – GO:0006956 (complement activation) and GO:0006959 (humoral immune response). Стоит отметить, что термин «complement activation» имеет наименьшее значение p.value среди всех значимо обогащенных терминов. Группы 4 и 5 состоят из 1 термина каждая – GO:0006936 (muscle contraction) и GO:0042742 (defense response to bacterium) соответственно.

На основании обогащения было выделено 8 генов, относящихся одновременно к 4 и более значимо обогащенным терминам – *CIQB*, *CIQA*, *ISG15*, *SERPING1*, *VSIG4*, *KLRD1*, *TRPM4* и *HFE*. Ген *ISG15* также примечателен тем, что это единственный ген, который относится к 3 разным группам терминов одновременно. Также стоит отметить, что все 8 генов имеют повышенную экспрессию у выживших пациентов по сравнению с погибшими.

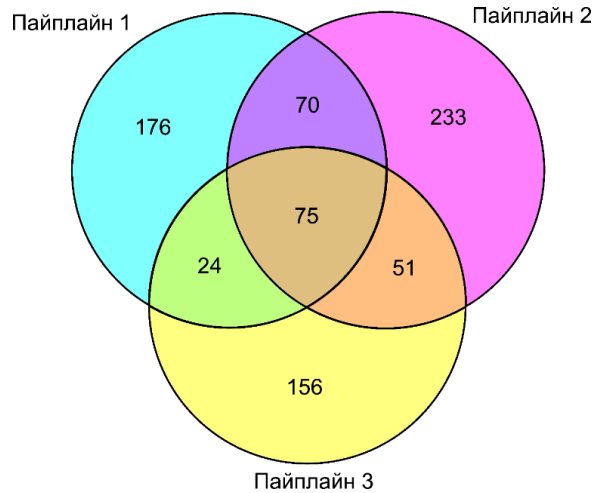


Рисунок 10. Диаграмма Венна для дифференциально экспрессирующихся генов между группой выживших пациентов и группой погибших пациентов, зараженных вариантом SARS-CoV-2 Дельта для всех трех использованных пайплайнов.

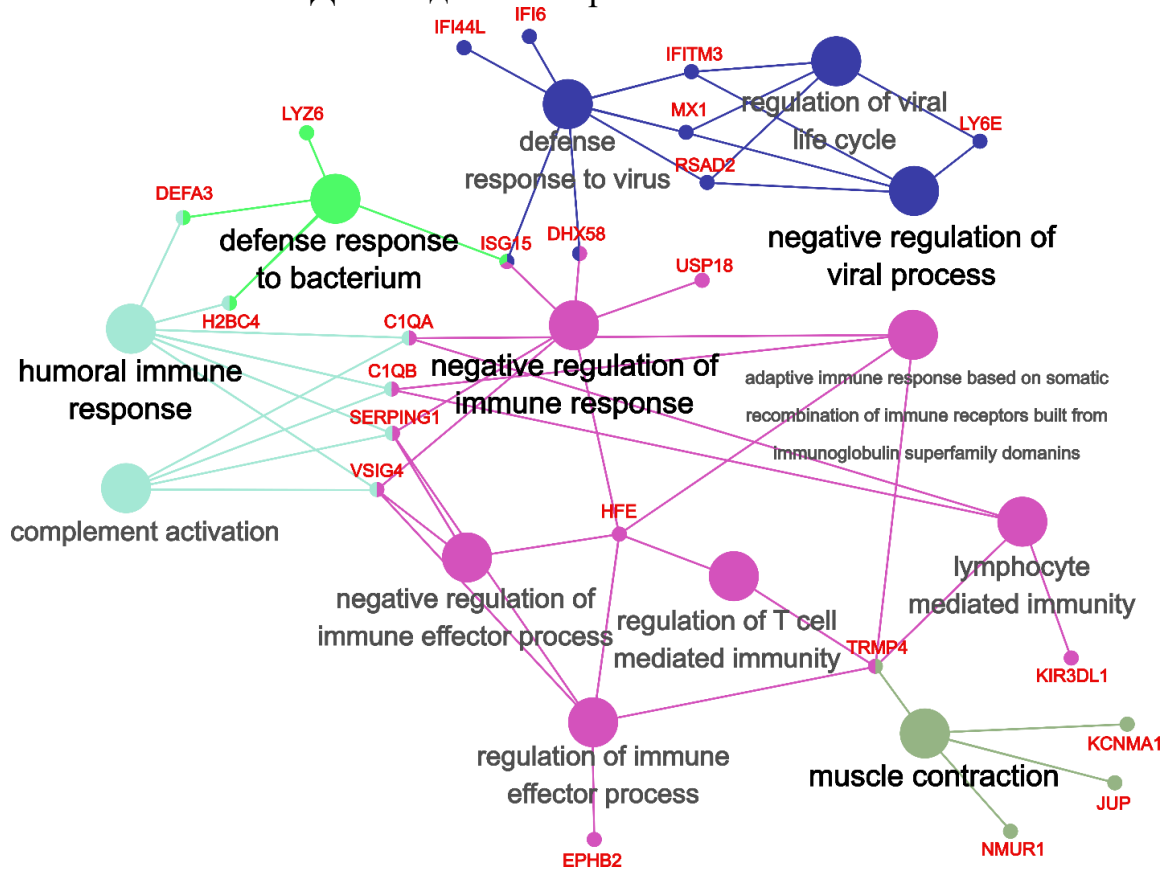


Рисунок 11. Сеть генов и терминов, построенных на основании обогащения терминов Gene Ontology ДЭГ, общих между всеми пайплайнами. Дифференциальная экспрессия определялась между группами выживших и погибших пациентов, зараженных вариантов SARS-CoV-2 Дельта. Цвет термина означает принадлежность группе, сформированной на основании общих ДЭГ. Термин с наибольшим p.value в каждой группе выделяется более темным черным шрифтом. Красный шрифт обозначает, что узел является геном.

4.3 Исследование липидных профилей в крови пациентов ОПИТ с COVID-19, вызванным заражением вариантами 20В и Дельта SARS-CoV-2

Для того чтобы выяснить, сопровождаются ли изменения в экспрессии рецепторов на уровне транскриптома изменениями на уровне липидома, нами было проведено исследование липидного профиля плазмы крови в группах пациентов ОПИТ с COVID-19, вызванным заражением вариантами SARS-CoV-2 20В и Дельта. Концентрации липидов между группами с летальными и нелетальными исходами были сравнены при помощи U-теста Манна-Уитни. Значимые различия между группами выживших и погибших не были выявлены, однако, в группе пациентов, инфицированных вариантом 20В SARS-CoV-2, была выявлена тенденция ($0.1 > p > 0.05$) к снижению уровня ЛПНП и общего холестерина в группе погибших по сравнению с выжившими (табл. 8). Подобные тенденции не были выявлены ни в каких других сравнениях, будь то в группе инфицированных Дельта вариантом SARS-CoV-2, или в концентрации липопротеинов высокой плотности и триглицеридов.

Таблица 8. Концентрации различных компонентов липидного профиля в группах пациентов ОПИТ с COVID-19 вызванным заражением 20В и Дельта вариантами SARS-CoV-2 с летальными и нелетальными исходами. ОХ – общий холестерол, ТГ – триглицериды, ЛПНП – липопротеины низкой плотности, ЛПВП – липопротеины высокой плотности.

Компонент липидного профиля	20В		Дельта	
	Летальный исход (N=21)	Нелетальный исход (N=18)	Летальный исход (N=13)	Нелетальный исход (N=15)
ОХ, mmol/l	3.4 (2.4-4.8) (p.val=0.054)	4.0 (2.6-5.9)	3.5 (2.1-5.8)	4.2 (2.5-6.7)
ЛПНП, mmol/l	1.8 (1.0-2.9) (p.val=0.087)	2.4 (1.5-3.7)	1.8 (0.1-3.5)	2.2 (0.9-4.3)
ЛПВП, mmol/l	0.8 (0.4-1.6)	1.0 (0.43-1.4)	0.8 (0.3-1.4)	0.9 (0.6-1.5)
ТГ, mmol/l	1.5 (1.0-5.4)	1.5 (1.9-3.0)	2.1 (0.9-4.3)	1.7 (0.9-4.5)

5. ОБСУЖДЕНИЕ

В этой работе было проведено два транскриптомных профилирования первичных мононуклеарных лейкоцитов периферической крови больных COVID-19 в разных временных периодах – исследование в группе пациентов, госпитализированных на момент, в который 20В был наиболее распространенным вариантом на северо-западе России (для которых мы будем предполагать, что заболевание было вызвано вариантом 20В), и исследование в группе пациентов, у которых при помощи RT-PCR было подтверждено заражение вариантом SARS-CoV-2 Дельта. В обоих исследованиях были сравнены уровни экспрессии в группах пациентов, выздоровевших и погибших в период 30 дней после поступления в ОРИТ. Экспрессия мРНК в первичных мононуклеарных лейкоцитах периферической крови пациентов была измерена при помощи полнотранскриптомного секвенирования. Материал для выделения РНК был получен в момент поступления больных в отделение ОРИТ. Такое сравнение позволяет нам выделить факторы, ассоциированные с летальными исходами тяжелых случаев заболевания. Далее будет проведено обсуждение результатов каждого из этих исследований отдельно, после чего результаты исследований будут сравнены между собой.

5.1 Сравнение разработанных пайплайнов

В этой работе для анализа транскриптомных данных было использовано 3 различных набора биоинформатических инструментов (пайплайнов). В этой части обсуждения будут рассмотрены специфические различия этих пайплайнов в 4 ключевых моментах обработки данных, а также будут рассмотрены возможные последствия этого решения для полученных результатов.

Ключевой этап 1 – удаление остаточных адаптеров и оснований низкого качества. Как было описано выше в разделе «Материалы и методы», пайплайн 1 не использует никаких методов для удаления адаптеров и оснований низкого качества,

пайплайн 2 использует Cutadapt [209] для удаления адаптеров и не проводит удаление оснований низкого качества, в то время как пайплайн 3 использует AdapterRemoval V2 [206] для обеих процедур. Как было упомянуто выше, данные о роли удаления адаптеров и удаления оснований противоречивы – существуют как данные о том, что это повышает точность оценки уровня экспрессии [39], так и о том, что это процедура не влияет на оценки экспрессии [42]. На данный момент не существует данных, показывающих, что эта процедура понижает точность оценок экспрессии, поэтому все три используемых подхода могут рассматриваться как эквивалентно допустимые.

Ключевой этап 2 – выравнивание ридов. Пайплайны 1 и 2 используют HISAT2 [54] для выравнивания ридов, при этом Пайплайн 2 использует самую последнюю доступную на момент проведения работы версию 2.2.1, в то время как Пайплайн 1 использует более устаревшую версию 2.0.1b. Однако, стоит отметить, что при подсчете числа ридов на ген корреляция между результатами выравнивания в Пайплайне 1 и 2 (R^2 Спирмена > 0.99 во всех образцах), что позволяет сделать вывод о том, что как минорная версия алгоритма для выравнивания, так и использование процедуры для подрезки остаточных адаптеров мало влияет на оценки экспрессии по генам. Пайплайн 3 в свою очередь использует STAR [56] для выравнивания ридов в интеграции с RSEM [64] для более точного выравнивания мультимапперов. К достоинствам STAR можно отнести как возможность встроенной интеграции с RSEM, так и то, что в недавно опубликованном исследовании было показано, что STAR как правило более точно картирует риды, особенно при большом количестве отличий секвенируемого от референсного генома [210]. Однако, в этой же работе было показано, что HISAT2 демонстрирует ненамного худшие результаты, чем STAR, делая его также допустимым выбором.

Ключевой этап 3 – нормализация. Пайплайны 1 и 2 используют встроенную в пакет “R” DESEQ2 [71] процедуру нормализации CPM, описанную в литературном обзоре ранее, в то время как Пайплайн 3 использует сочетание CPM из пакета “R” limma [68] (идентичную по сути с процедурой из DESEQ2) с

процедурой ТММ из пакета “R” edgeR [211]. Стоит отметить, что в обоих случаях такой выбор процедур диктуется в первую очередь рекомендациями разработчиков пакетов, используемых для следующего ключевого этапа, однако использование только процедуры СРМ рискует тем, что проблема композиции, описанная в секции введения про нормализацию, останется не решенной, что приведет к возможному влиянию высоко-представленных транскриптов на экспрессию всех остальных транскриптов в библиотеке [65].

Ключевой этап 4 – моделирование и выявление дифференциальной экспрессии. Пайплайны 1 и 2 используют DESEQ2 для построения модели экспрессии и проведения статистических тестов, в то время как Пайплайн 3 использует limma. На первый взгляд, DESEQ2 обладает намного большей «face validity» – натуральные числа, значения которых случайные величины из негативно биномиального распределения, значительно лучше представляют числа ридов, приходящихся на ген, чем числа из нормального распределения, которые могут быть дробными и отрицательным. Однако, в ранее опубликованных исследованиях было показано, что несмотря на это, результаты показываемые limma при прочих равных такие же или лучше, чем результаты методов, основанных на негативно-биномиальном распределении, включая DESEQ и edgeR [52]. Стоит отметить, что разница в точности оценки экспрессии между всеми тремя вышеупомянутыми методами невелика, поэтому выбор любого из них для пайплайнов был бы допустим.

Все используемые в различных пайплайнах подходы и инструменты относятся к «лучшим практикам», согласно независимым исследованиям [31]. Тем не менее, получаемые списки ДЭГ значительно различаются. Когда принимается решение о критериях, используемых при отборе ДЭГ (или любого другого примера решения отбора «значимых результатов» из большей группы при помощи статистического теста), важным аспектом является баланс между чувствительностью теста и его специфичностью. Общепринятый подход рекомендует в большинстве случаев придавать большее значение

ложноположительным результатам, и склонятся к стороне «специфичности», и выбранный в этой работе подход делает особенно сильный акцент на этом, так как в первую очередь мы рассматриваем гены, выявленные как ДЭГ во всех трех пайплайнах. Естественным результатом такого решения является повышенное число ложно отрицательных результатов, однако, с нашей точки зрения, это приемлемый выбор, учитывая, что зачастую проблемой в анализе транскриптома является трудность в интерпретации списков ДЭГ, а не недостаток их количества. Полученные нами результаты показывают, что выбранный подход пригоден для того, чтобы получать интерпретируемые и валидные результаты при анализе транскриптомных профилей.

5.2 Обсуждение результатов, полученных в процессе транскриптомного анализа в группе зараженных 20В вариантом SARS-COV-2

5.2.1 Роль метаболизма холестерина в патогенезе COVID-19

Дифференциальная экспрессия была выявлена между группами больных, выздоровевших и погибших в период 30 дней после поступления в ОРИТ при помощи трех отдельных пайплайнов. Выявленные ДЭГ были использованы для обогащения терминов Gene Ontology. В результате этого обогащения было показано, что термины «the low-density lipoprotein particle receptor activity pathway» (GO: 0005041) и «leukocyte differentiation» (GO:0002521) выявляются как значимо обогащенные вне зависимости от использованного пайплайна. Термин «the low-density lipoprotein particle receptor activity pathway» (GO: 0005041) был особенно значимо обогащен, учитывая, что 75% всех генов, аннотированных принадлежностью к этому термину, были дифференциально экспрессированы, с повышенной экспрессией в группе погибших по сравнению с выжившими. Этот термин связан с регуляцией поглощения клетками богатых холестерилами липопротеинов низкой плотности из кровотока [212].

К настоящему времени вышло несколько работ, связывающих метаболизм холестерина с чувствительностью к COVID-19 и тяжестью протекания

инфекции [213]–[215]. В нескольких работах было продемонстрировано, что заболевание COVID-19 сопровождается дислипидемией [216], [217]. В частности, уровни липидов в сыворотке крови понижаются после начала острого заболевания COVID-19, и продолжают понижаться одновременно с повышением концентрации С-реактивного белка вплоть до исхода болезни [217]. Более того, было показано, что концентрации липопротеинов низкой плотности (ЛПНП) являются предикторами негативных исходов болезни [216], [218], [219]. В выборке пациентов, чей исход COVID-19 завершился гибелью, уровень ЛПНП постоянно снижался в течение болезни вплоть до смерти. В одной из работ было показано, что уровень ЛПНП в группе погибших снижался вплоть до 60% от уровня при поступлении в ОРИТ. [218]. Таким образом, существует возможность, что наблюдаемый нами на уровне экспрессии пониженный уровень активности рецепторов ЛПНП у выживших пациентов мог приводить к их повышенной сопротивляемости к последствиям снижения уровня ЛПНП в сыворотке крови.

Другой фактор, играющий важную роль в процессе вирусной инфекции – это специфические клеточные белки, определяющие взаимодействия вирусов с заражаемыми клетками. Существуют гипотезы, предполагающие, что проникновение SARS-CoV-2 в клетку посредством связывания с ACE2 может зависеть от эндогенного синтеза холестерина [220], [221]. Богатые холестерином области мембраны клетки играют ключевую роль в шиповидно-гликопротеиновозависимом слиянии мембран, которое, в свою очередь, необходимо для проникновения SARS-CoV-2 в клетку и его репликации [220]. Более того, на примере нескольких других вирусных инфекций были показаны участие процессов, задействованных в клеточном гомеостазе холестерина, в проникновении вируса в клетку, и их роль в обеспечении процессов репликации вирусов [222], [223]. Также стоит отметить, что исследование способности заражения клеточных культур SARS-CoV-2 (изолят Ухань-1) при помощи полногеномных нокаутных CRISPR панелей выявило участие генов, связанных с метаболизмом холестерина, в репликации вируса [224]. В дополнение к тому, чтобы продемонстрировать роль

ACE2 как обособленного фактора в проникновении вируса в клетку, популяции клеток, резистентные к вирусу, также проявляют даунрегуляцию сигнального пути белков, связывающихся со стерольным регуляторным элементом (SREBP), а также гена *NPC1* (Niemann-Pick intracellular cholesterol transporter1) контролирующего экспорт холестерина из эндосом. Авторы рассуждают, что полученные ими данные указывают, что эти факторы могут быть необходимы для поддержки репликации SARS-CoV-2, и что фармакологическое ингибирование гомеостаза холестерина может быть методом борьбы с репликацией SARS-CoV-2 [224]. Помимо этого, вирус также может активировать семейство транскрипционных факторов SREBP, регулирующих биосинтез липидов и гомеостаз холестерина. Например, в первичных мононуклеарных лейкоцитах периферической крови пациентов с COVID-19 повышен уровень белка SREBP-2, который напрямую активирует несколько генов метаболизма холестерина [225]. Более того, было показано, что концентрация мРНК *SREBF2*, а также *SESNI* и *PCSK9* была повышена в первичных мононуклеарных лейкоцитах периферической крови пациентов с COVID-19 пропорционально тяжести протекания заболевания.

В общей сложности в этой работе было выявлено 5 генов (*PPARG*, *CD36*, *STAB1*, *ITGAV* и *ANXA2*), относящихся к метаболизму липидов, экспрессия которых снижена в группе выживших по сравнению с группой погибших во всех трех пайплайнах.

5.2.2 ANXA2

Аннексин А2 участвует в регуляции множества биологических процессов, включая гомеостаз сосудов, регуляцию воспаления и активации иммунной системы, повреждение и восстановление тканей и метаболизм холестерина [226]. В опубликованных ранее работах было показано, что Аннексин А2 играет роль эндогенного ингибитора активности пропротеиновой конвертазы субтилизин-кексинового типа 9 (PCSK9). Активность PCSK9 направлена на деградацию рецепторов липопротеинов низкой плотности, и посредством ингибирования его

активности *ANXA2* влияет на концентрации рецепторов ЛПНП и концентрацию холестерина в плазме крови [227], [228]. В то же время *AnxA2* играет роль в заражении несколькими различными вирусами (везивируса кроликов, вируса Репродуктивно-Респираторного синдрома свиней, вируса классической чумы свиней, вируса болезни Блютанга и вируса инфекционного бронхита кур) у нескольких различных видов животных (свиньи, кролики, крупный рогатый скот) [229]. На данный момент неизвестно, играет ли роль *AnxA2* непосредственно в инфекции SARS-CoV-2, или же его участие на процесс патогенеза COVID-19 происходит опосредованно через влияние на метаболизм холестерина и ЛПНП, но в полученных в этой работе данных понижение его экспрессии наблюдается в группе выживших пациентов с COVID-19 по сравнению с погибшими, что потенциально указывает на ассоциацию его повышенной экспрессии с отрицательными исходами COVID-19.

5.2.3 Скэвенджер-рецепторы CD36 и STAB1

Гены *CD36* and *STAB1* кодируют скэвенджер-рецепторы, распознающие модифицированные части ЛПНП. [230]–[232]. *CD36* был впервые описан как трансмембранный гликопротеин тромбоцитов. К данному моменту было установлено, что это член семейства скэвенджер-рецепторов класса Б [233] экспрессирующийся в том числе в макрофагах периферической крови и играющий значимую роль в функционировании системы врожденного иммунитета [234]–[237]. С другой стороны, этот гликопротеин также ассоциирован с развитием склеротических бляшек посредством своего участия в метаболизме окисленных липопротеинов низкой плотности, фосфолипидов и жирных кислот [238]–[240]. *CD36* – один из основных скэвенджер-рецепторов, ответственных за распознавание и интернализацию окисленных липопротеинов низкой плотности [232].

Рецептор-опосредованный эндоцитоз окисленных липопротеинов низкой плотности макрофагами моноцитарного происхождения активизирует репрограммирование врожденных иммунных ответов, что обычно называется

«тренированным» (отлично от «адаптивного») иммунитета [232].

Стоит упомянуть, что было показано обогащение репрезентативной сети транскрипционных модулей крови пациентов с COVID-19 молекулярными сигналами, характерными для моноцитов, в том числе повышенной экспрессией CD36 [241]. Возможным объяснением этого может служить то, что тяжелые случаи COVID-19 характеризуются повышением частоты групп классических моноцитов, экспрессирующих CD36 [242].

На основании полученных нами данных мы предполагаем, что повышенная экспрессия CD36 может быть ассоциирована с негативными исходами при заболеваниях COVID-19 и может играть роль в развитии негативного исхода посредством двух разных потенциальных механизмов – активации врожденного иммунитета и активацией процессов атерогенеза посредством стимуляции поглощения жирных кислот и окисленных липопротеинов.

Ген *STAB1* кодирует мультифункциональный скэвенджер-рецептор, активно экспрессируемый тканевыми макрофагами и синусоидальными эндотелиальными клетками. Известно, что экспрессия *STAB1* повышается при воспалении. Стабилин является необходимым компонентом эндоцитоза грам-положительных и грам-отрицательных бактерий, ЛПНП (включая окисленные и ацетилированные) и продуктов гликолизирования макрофагами [230], [231]. Мы предполагаем, что повышенная активность *PPARG* может привести к повышенной экспрессии *STAB1* и *CD36* [243], т.к. известно, что *PPARG* является ключевым регулятором транскрипции этих генов. В свою очередь, повышение количества рецепторов *STAB1* и *CD36* может привести к усилению эндоцитоза липидов и их локальному исчерпанию. Это согласуется с данными, указывающими на понижение уровня липидов в сыворотке больных с COVID-19, ассоциированное с плохим прогнозом исхода заболевания [244]. Было предложено несколько возможных механизмов, объясняющих эти факты – пониженный биосинтез ЛПНП из-за дисфункции печени, измененный метаболизм липидов из-за острого воспаления и/или более

интенсивная деградация липидов в связи с повышением концентрации свободных радикалов [244].

5.2.4 PPAR γ

В полученных нами данных было показано повышение экспрессии *PPARG* в группе погибших пациентов. PPAR γ может регулировать метаболизм липидов в иммунных клетках (посредством регуляции генов, участвующих в транспорте и метаболизме липидов, включая *CD36*). Помимо этого, активация PPAR γ может участвовать в презентации липидных антигенов Т лимфоцитам и модуляции иммунного ответа через дендритные клетки [245]. Таким образом, PPAR γ может служить связующим звеном между липидным микроокружением и дифференцировкой и функционированием иммунных клеток. PPAR γ обладает иммуномодуляторными функциями и играет роль в разрешении воспаления, и его повышенная экспрессия в группе погибших пациентов может служить признаком неразрешившегося воспаления в условиях истощения липидов, что характеризует в том числе тяжелые случаи COVID-19.

5.2.5 GAS6, LGALS9 и LILRB4

Наличие среди общих значимо обогащенных терминов термина «leukocyte differentiation» указывает на дисрегуляцию генов, связанных с врожденным иммунитетом, системой комплемента, и активацию Т и В лимфоцитов в группе погибших пациентов. Гены *GAS6* (growth arrest-specific protein 6), *LGALS9* (galectin 9), *LILRB4* (Leukocyte Immunoglobulin Like Receptor B4) имели повышенную экспрессию в группе погибших пациентов. В опубликованной ранее работе было показано, что концентрации продуктов этих генов были повышены в плазме крови больных COVID-19 (в сравнении со здоровыми контролями), а также что эти концентрации понижались со временем у пациентов, выживших в течении 30 дней после госпитализации в ОРИТ [246]. Концентрации Галектина 9 и LILRB4 были повышены в группе с легким течением болезни по сравнению с контрольной группой и были повышены в группах с тяжелым и критическим течением болезни

в еще большей степени [247]. Стоит также отметить, что *PPARG* регулирует экспрессию *LGALS9* таким образом, что повышение экспрессии *PPARG* приводит к повышению экспрессии *LGALS9* [248].

Таким образом, на основании полученных нами данных можно заключить, что индукция активности рецепторов ЛПНП в пациентах с COVID-19 ассоциирована с худшим прогнозом исхода заболевания. Наши результаты указывают на взаимосвязь гомеостаза липидов, иммунной системы и чувствительности к COVID-19.

5.3 Обсуждение результатов, полученных в процессе транскриптомного анализа в группе зараженных Дельта вариантом SARS-COV-2

5.3.1 Двойная роль иммунной системы в патогенезе COVID-19

Сходным образом с предыдущей частью исследования, дифференциальная экспрессия была выявлена между группами больных, выздоровевших и погибших в период 30 дней после поступления в ОРИТ при помощи трех отдельных пайплайнов. Заражение больных вариантом SARS-CoV-2 Дельта было подтверждено при помощи RT-PCR.

Было выявлено несколько значимо обогащенных терминов Gene Ontology: Biological Processes (GO:BP), подавляющее большинство которых относятся к различным иммунным и противовирусным процессам. Используя значимо обогащенные термины и ДЭГ, была построена сеть генов и терминов. Исходя из конструкции сети и общих ДЭГ между терминами, термины могут быть подразделены на 5 групп – негативная регуляция иммунного ответа (группа 1), негативная регуляция вирусного процесса (группа 2), активация гуморального иммунитета и системы комплемента (группа 3), защитный ответ на бактерии (группа 4) и сокращение мышцы (группа 5).

В современном понимании роль иммунного ответа в патогенезе COVID-19 имеет две стороны – с одной стороны, иммунный ответ необходим для того чтобы предотвратить репликацию вируса и остановить распространение инфекции, но, с

другой стороны, гиперактивация иммунной системы, характерная для тяжелых случаев COVID-19, вносит значимый вклад в повреждение тканей и летальность инфекции. Значимость быстрого ответа иммунной системы на заражение SARS-CoV-2 подчеркивается тем, что вирус имеет множество приспособлений, подавляющих ранний интерфероновый ответ и ответ врожденный иммунной систем, и введение экзогенного интерферона-1 пациентам с COVID-19 до и на ранних этапах инфекции приводит к лучшим прогнозам и понижению вирусной нагрузки [249]. С другой стороны, длительный иммунный ответ и самоподдерживающаяся гиперактивация иммунной системы вносят значимый вклад в морбидность и летальность тяжелых случаев COVID-19 [250]. Полученные нами данные хорошо укладываются и могут служить дополнительными подтверждением этого понимания. Значимое обогащение терминов, относящихся к негативной регуляции иммунного ответа и то, что большинство относящихся к этим терминам генов имеет повышенную экспрессию в группе выживших указывает на то, что негативная регуляция иммунного ответа важна для благоприятного исхода COVID-19. В то же время, присутствие терминов, относящихся к противовирусному ответу и негативной регуляции вирусных процессов, указывает на важность специфического иммунного ответа в благоприятных исходах COVID-19. Способность SARS-CoV-2 к избеганию иммунной системы и угнетению интерфероновому ответу хорошо задокументированы [249]. Значимо повышенная дифференциальная экспрессия генов, принадлежащих к связанным с противовирусным ответом терминам, указывает на то, что эффективное сопротивление механизмам подавления иммунного ответа при COVID-19 может играть важную роль развитию позитивных исходов COVID-19. Важность интерфероновому ответу также подчеркивается тем, что в группу 2 включены 3 гена, экспрессия которых индуцируется интерфероном – *IFI6*, *IFITM3* и *IFI44L*, экспрессия всех из которых повышена в группе выживших. Рассматривая сосуществование значимо обогащенных терминов, связанных с подавлением и индукцией иммунного ответа, стоит также отметить, что только два ДЭГ относятся

одновременно к группам 1 и 2, и что большинство ДЭГ относятся исключительно либо к группе 1, либо к группе 2.

То, что термины «гуморальный ответ» (часть адаптивного иммунитета) и «активация системы комплемента» (часть врожденного иммунитета) объединены в одну группу, скорее всего, связано с экспрессией 4 генов, относящихся одновременно к группам 1 и 3 – *SERPING1*, *VSIG4*, *CIQB* и *CIQ*. На основании нашего анализа эти гены представляются как особенно важные в ответе на COVID-19 и будут подробно рассмотрены позже. Однако, возвращаясь к рассмотрению сети генов и терминов, стоит сначала рассмотреть гены, относящиеся одновременно к группам 3 и 4, и принадлежащие термину «гуморальный ответ», но не термину «система комплемента». Группа 4, защитный ответ на бактерии, на первый взгляд представляется как не специфическая для заражения COVID-19, однако, если обратить внимание на гены, относящиеся к этой группе, и в особенности на их направление экспрессии, можно сделать некоторые выводы о потенциальной роли этих процессов в определении исхода COVID-19. Один из относящихся к группе 4 генов – *ISG15* также относится к группам 1 и 2, и является единственным геном, относящимся к трем группам терминов одновременно. В силу его ключевого положения в сети генов и терминов, далее он будет рассмотрен отдельно. Гены *LYZ* [251], *DEFA3* [252] и *H2BC4* [253] кодируют белки с задокументированными антимикробными свойствами. *DEFA3* and *H2BC4* относятся одновременно к термину «гуморальный иммунный ответ» и «ответ на бактерии», в то время как *LYZ* относится исключительно к термину «ответ на бактерии». Экспрессия всех трех этих генов повышена в группе погибших по сравнению с группой выживших. В целом, большинство генов, использованных для обогащения терминов Gene Ontology (60 из 74), в противоположность этому имеют пониженную экспрессию в группе погибших по сравнению с группой выживших. В силу этого маловероятно, что это случайное совпадение, и эта особенность в направлении экспрессии может указывать, что ослабление/отсутствие усиления анти-микробного и гуморального ответов может быть важным в определении положительного исхода COVID-19.

Таким образом, рассматривая значимое обогащение терминов и направление дифференциальной экспрессии связанных с ними генов, можно заключить, что своевременный и специфический противовирусный ответ, вкупе со своевременным ослаблением иммунного ответа, могут быть важны в определении позитивного исхода COVID-19. Далее будет рассмотрено, какую роль гены, относящиеся к этим иммунным процессам, могут играть в патогенезе COVID-19.

В общей сложности, на основании количества значимо обогащенных терминов, которым принадлежат гены, 8 ключевых ДЭГ могут быть выделены на основании нашего исследования – *CIQB*, *CIQA*, *ISG15*, *SERPING1*, *VSIG4*, *KLRD1*, *TRPM4* и *HFE*.

5.3.2 ISG15 – потенциальный ключевой регулятор ответа на COVID-19

Как упоминалось выше, ген *ISG15* – это единственный ген в сети генов терминов, который относится к трем разным группам и пяти различным значимо обогащенным терминам одновременно. *ISG15* кодирует интерферон-индуцируемый убиквитин подобный белок [254], ковалентно связывающийся как с веществами клетки-хозяина, так и с вирусными субстратами в процессе, получившем название «ИСДжилирование» («ISGylation»). ИСДжилирование играет важную роль в ингибировании репликации вируса. Точно неизвестно, какую роль играют ИСДжилированные белки и какой исход для белков определяет эта метка, однако существует несколько предполагаемых механизмов для ее действия. ИСДжилирование может влиять на локализацию белков клетки-хозяина, как было продемонстрировано на примере филамина Б [255], а также может нарушать агрегацию комплексов вирусных белков, как было показано на примере взаимодействия *ISG15* с вирусом гриппа Б и вирусом папилломы человека [256]. *ISG15* также может нарушать выход вируса из клетки-хозяина, как было показано на примере вируса человеческого иммунодефицита 1 (ВИЧ-1), в котором высвобождение вирионов ВИЧ-1, но не синтеза белков ВИЧ-1, было предотвращено посредством оверэкспрессии *ISG15* [257]. Также было показано,

что помимо взаимодействия с вирусными белками посредством ковалентной модификации, свободный ISG15 может действовать как цитокин. Внеклеточный ISG15 может служить лигандом интегрина ITGAL [258], связываясь с ним и стимулируя высвобождение of IFN γ и IL10 из лимфоцитов natural killer-ов (NK). Стоит отметить, что высвобождение IL10 может быть особенно значимым для прогноза COVID-19, так как ранее было показано, что уровень этого белка является независимым предиктором исхода COVID-19 [259].

В опубликованных ранее работах повышение экспрессии *ISG15* было выявлено в транскриптомном профилировании назофарингеальных мазков пациентов с COVID-19 [260]–[262]. С другой стороны, было показано, что *in vitro* заражение SARS-CoV-2 не приводило к повышению уровня мРНК *ISG15*, что, однако, может объясняться свойствами использованной клеточной культуры [263]. На важность роли *ISG15* также указывает то, что активность ISG15 в клетке регулируется SARS-CoV-2. Один из белков SARS-CoV-2 – папаин-подобная протеаза PLpro, участвует в процессинге вирусных белков и в регуляции посттрансляционной модификации белков клетки хозяина. Важно отметить, что у других коронавирусов человека PLpro используется для того, чтобы предотвратить убиквитинирование клеточных и вирусных белков, в то время как специфически у SARS-CoV-2 основным субстратом PLpro является ISG15 [264]–[266]. Основными целями PLpro являются комплексы ISG15 и IRF3, а также комплексы ISG15 и MDA5 [267], [268]. Эта адаптация SARS-CoV-2 указывает на значимость *ISG15* в противовирусном иммунитете и позволяет предположить, что повышенная экспрессия *ISG15*, и, как следствие, его повышенная активность играют важную роль в определении исходов COVID-19. Таким образом, *ISG15* представляется возможной целью для терапевтического вмешательства при COVID-19.

5.3.3 Система комплемента при COVID-19

Из 8 вышеупомянутых ключевых генов 4 относятся одновременно к терминам «активация системы комплемента» и «негативная регуляция иммунного

ответа», а также «гуморальный ответ» – *CIQA*, *CIQB*, *SERPING1* и *VSIG4*. Стоит отметить что все эти 4 гена имеют повышенную экспрессию в группе выживших по сравнению с группой погибших. Для того чтобы лучше понять роль этих генов в COVID-19, стоит сначала обсудить в целом полученные на данный момент данные о роли системы комплемента при COVID-19.

Система комплемента играет роль в вирусных инфекциях посредством опсонизации вирусов и клеток, зараженных вирусами, непосредственной нейтрализации вирусных частиц вне клеток, лизиса зараженных вирусами клеток и активации вирусоспецифических иммунных и воспалительных ответов [269]. Система комплемента может быть активирована посредством трех возможных путей – классического, неклассического и лектинового. Общепринятая точка зрения на активацию системы комплемента говорит о том, что активация любым из трех путей в конечном итоге приводит к одному и тому же результату, и приводит к тому же белковому каскаду, включающему одни и те же эффекторные молекулы. Однако на данный момент существуют данные, позволяющие предполагать, что путь активации системы комплемента может быть важен в определении исходов COVID-19. В наших собственных данных наблюдалось значительное изменение экспрессии генов *CIQA* и *CIQB*, кодирующих субъединицу первого компонента системы комплемента, специфически относящегося к активации системы комплемента по классическому пути. На данный момент нет ясной картины того, по какому именно пути активируется система комплемента при заражении SARS-CoV-2. Было показано, что все основные (S, N, M, E) белки SARS-CoV-2 взаимодействуют с первой субъединицей системы комплемента, активируя классический путь активации системы комплемента [270]. Однако, в другой работе была получена противоречащая этому информация, демонстрирующая то, что белки SARS-CoV-2 напрямую не связываются с C1Q, а активируют систему комплемента по лектиновому пути [271]. Показано, что отложения белков системы комплемента выявляются в различных тканях больных COVID-19, в том числе альвеолах, почках и капиллярах печени. Однако, большинство этих отложений

состоят из белков системы комплемента, относящихся к неклассическому пути активации системы комплемента [272]. Ретроспективное исследование концентрации нескольких компонентов системы комплемента (C1q, C3, C4 и C5) в когорте из 74 больных COVID-19 показало, что высокий уровень экспрессии компонентов, связанных с лектиновым и неклассическим путем активации системы комплемента ассоциирован с негативными исходами, в то время как для компонентов, ассоциированных с классическим путем активации, такая ассоциация выявлена не была [273]. Также на когорте из 71 пациента было показано, что уровень C1q в сыворотке крови был повышен у пациентов с легкой формой COVID-19 по сравнению с пациентами с тяжелой формой [274]. Стоит отметить, что дифференциальная экспрессия, наблюдаемая между пациентами с тяжелой и легкой формой, не обязательно будет выявлена в случае сравнения погибших и выживших пациентов. Тем не менее, если сделать предположение, что профили экспрессии выживших пациентов в сравнении с погибшими будут сходны с профилями экспрессии более легких случаев в сравнении с тяжелыми, полученные нами данные согласуются с картиной, описанной в ранее опубликованных работах. Наблюдаемая нами повышенная экспрессия *CIQA* и *CIQB* – субъединиц первого компонента системы комплемента, свидетельствует о активации комплемента посредством классического пути у выживших пациентов, что в полученных ранее данных ассоциировано с лучшими исходами COVID-19. Однако, при интерпретации этих данных нельзя сказать, является ли такое взаимоотношение причинно-следственным или же просто ассоциацией. В предыдущих работах было показано, что концентрации C1q, в отличие от концентраций IL6 и IL10, не являются независимыми предикторами клинических исходов, и их влияние зависит от концентраций других компонентов цитокинового шторма [259].

5.3.4 Двойная роль системы комплемента при COVID-19

В этом обсуждении ранее была представлена идея о двусторонней роли регуляции иммунной системы в патогенезе COVID-19. Полученные нами и в

литературе данные могут свидетельствовать о том, что система комплемента также играет двойную роль в определении исходов COVID-19. С одной стороны, гиперактивация системы комплемента предполагается как возможной участник развития дыхательной недостаточности при COVID-19 [275] и других заболеваниях [276], и ингибирование системы комплемента предполагается как возможная стратегия для улучшения исходов в тяжелых случаях COVID-19 [277], [278]. С другой стороны, система комплемента играет необходимую роль в иммунном ответе на инфекционные заболевания [279], и ее ингибирование приводит к значительному повышению вероятности возникновения ятрогенных побочных эффектов [277], [278]. Было предложено несколько возможных объяснений тому, почему система комплемента может играть такую роль. Первое – это тайминг, с предположением что ранняя часть воздействия активированной системы комплемента полезна, но при этом более отложенные эффекты вредны [269]. Второе объяснение предполагает, что решающим фактором является интенсивность, и к позитивным исходам ведет баланс в степени активации системы, а отсутствие активации и гиперактивация ведут к негативным исходам [279]. Третье объяснение это то, что специфический путь активации системы комплемента определяет, будет ли ее эффект вредным или полезным. Наши собственные данные поддерживают эту гипотезу, учитывая, что увеличение концентрации именно первого компонента системы комплемента, относящегося к классическому пути активации, но не других компонентов, наблюдается в группе выживших. В других работах также показано, что увеличение концентрации компонентов, относящихся к специфическому пути активации системы комплемента, а не повышение или понижение концентраций всех компонентов в целом ассоциировано с определенными исходами COVID-19 [272]–[274]. На основании этого можно сделать предположение, что повышение экспрессии *CIQA* и *CIQB* может обладать протективным эффектом при COVID-19 и считаться молекулярным маркером более легких случаев COVID-19. Однако, стоит отметить, что роль системы комплемента при COVID-19 сложна, не до конца изучена, и

требуется продолжать исследования в этой области.

5.3.5. VSIG4 и SERPING1 – потенциальные ключевые регуляторы системы комплемента при COVID-19

Два других гена с повышенной экспрессией в группе выживших, относящихся к терминам «активация системы комплемента», «негативная регуляция иммунного ответа» и «гуморальный ответ» — это *VSIG4* и *SERPING1*.

Ген *VSIG4* кодирует регуляторный белок, содержащий иммуноглобулиновый домен, также известный как «рецептор комплемента из иммуноглобулинового суперсемейства». Этот ген экспрессируется макрофагами и дендритными клетками и участвует в поглощении частиц, опсонизированных компонентами системы комплемента. *VSIG4* также ингибирует активацию неклассического пути активации системы комплемента посредством связывания с C3b [280], [281]. В силу последнего, гиперэкспрессия этого гена в группе выживших пациентов с COVID-19 по сравнению с погибшими в наших данных говорит в пользу идеи о значимости специфического пути активации системы комплемента для исходов COVID-19. Помимо регуляции системы комплемента и участия в фагоцитозе, взаимодействие *VSIG4* с его лигандами в макрофагах также вызывает иммуносупрессивные и противовоспалительные эффекты [282].

Роль *VSIG4* в патогенезе COVID-19 на данный момент не до конца изучена, и существуют данные, которые указывают на то, что повышение его экспрессии может быть как полезным, так и вредным фактором. Понижение экспрессии *VSIG4*, приводящее к ингибированию противовоспалительной активности, было ранее описано при хронической обструктивной болезни легких как значимый фактор в его патогенезе [283]. С другой стороны, ранее было показано, что гиперэкспрессия этого гена ассоциирована с негативными исходами при ОРДС [284] и раке легких [285], [286], что по аналогии может означать, что его повышенная экспрессия при COVID-19 также может быть связана с негативными исходами. Как было упомянуто ранее, на основании полученных нами данных мы рассматриваем

негативную регуляцию иммунной системы как важный фактор, определяющий позитивные исходы при COVID-19, и иммуносупрессивная и противовоспалительная активность *VSIG4*, его способность подавлять неклассический путь активации системы комплемента, вкупе с наблюдаемым повышением его экспрессии в группе выживших, позволяет предположить, что повышение его экспрессии скорее полезно при заболевании COVID-19.

Ген *SERPING1* кодирует ингибитор сериновой протеазы C1-INH, играющий роль в ингибировании врожденного иммунного ответа и активации системы комплемента [287]. В отличие от *VSIG4*, *SERPING1* способен ингибировать все три пути активации системы комплемента, а также способен облегчать комплемент-зависимое и воспалительное повреждение в нескольких моделях травмы [288], [289]. C1-INH ингибирует эти пути активации посредством формирования ковалентных связей с C1r и C1s (классический путь), MASP-1 (лектиновый путь), и C3b (неклассический путь) [287]. Стоит отметить, что, поскольку это механизм посттрансляционной регуляции, тот факт, что мы наблюдаем увеличение концентрации мРНК *CIQA* и *CIQB*, может не быть связан с повышением активности C1q, если она в свою очередь будет ингибирована на посттрансляционном этапе. *SERPING1* также является интерферон-индуцируемым геном, что вкупе с его способностью ингибировать иммунный ответ и активацию системы комплемента представляет его в роли негативной обратной связи, которая может ограничивать избыточную активацию иммунной системы и таким образом улучшать исходы заболеваний [287]. Это особенно актуально в случае тяжелого COVID-19, который характеризуется склонностью вызывать иммуноопосредованное повреждение различных тканей. В ранее опубликованной работе, ре-анализирующей три случая заболевания COVID-19 2020-го года, было показано повышение экспрессии *SERPING1* у больных по сравнению со здоровыми индивидами [290]. Также было показано, что T-аллель варианта rs78958998 расположенного в гене *SERPING1*, ассоциирована с повышенным риском виремии в случаях COVID-19 [291].

Таким образом, в процессе анализа дифференциальной экспрессии нами выявлены 5 ключевых генов, *ISG15*, *SERPING1*, *CIQA*, *CIQB*, и *VSIG4*, повышение экспрессии которых наблюдается в группе выживших по сравнению с погибшими. Проведенный анализ обогащения, в совокупности с литературными данными, свидетельствуют о важности этих генов в регуляции системы комплемента, противовирусного ответа и ингибировании избыточного иммунного ответа.

5.4 Сравнительный анализ ДЭГ, выявленных в группах инфицированных 20В вариантом и Дельта вариантом SARS-CoV-2

Две части этой работы были проведены сходным образом, со следующими двумя важными отклонениями –

В первой части исследования были сравнены группы из 3 погибших и 5 выживших, с двумя техническими повторами на образец, в то время как во второй были сравнены группы из 5 погибших и 5 выживших в одном повторе каждая

В первой части исследования ДЭГ были выявлены с поправкой FDR на множественное тестирование, в то время как во второй с этой поправкой гены выявлены не были, в силу чего было принято решение выявить ДЭГ без поправки на множественное тестирование

Вероятно, что второе различие является следствием первого. Помимо этих различий, между двумя этапами сбора образцов для исследований прошло около полутора лет (ноябрь 2020 – июнь 2021), за которые изменились многие обстоятельства, неподконтрольные исследовательской группе. Оба этих источника различий затрудняют интерпретацию данных. Тем не менее, даже принимая во внимание возможные запутывающие (confounding) факторы, наблюдается значимое изменение в наборе ДЭГ. Доля общих ДЭГ в группах пациентов с вариантами 20В и Дельта не достигает даже 5% ни в одном из случаев (рис. 12-15). Учитывая, что это наблюдается стабильно во всех пайплайнах и, в случае общих генов, предположение, что различия вызываются полностью различиями в методах обработки между эксперимента, можно исключить, так как различия такого

характера должны были бы иметь случайную природу, и наблюдаемая в этом случае картина скорее всего сильно различалась бы между пайплайнами. Таким образом, было бы логично предположить, что два основных потенциальных источника различий между группами – это различия, связанные с изменениями в протоколах лечения и статусе поступающих больных, и различия, связанные с особенностями вариантов вируса.

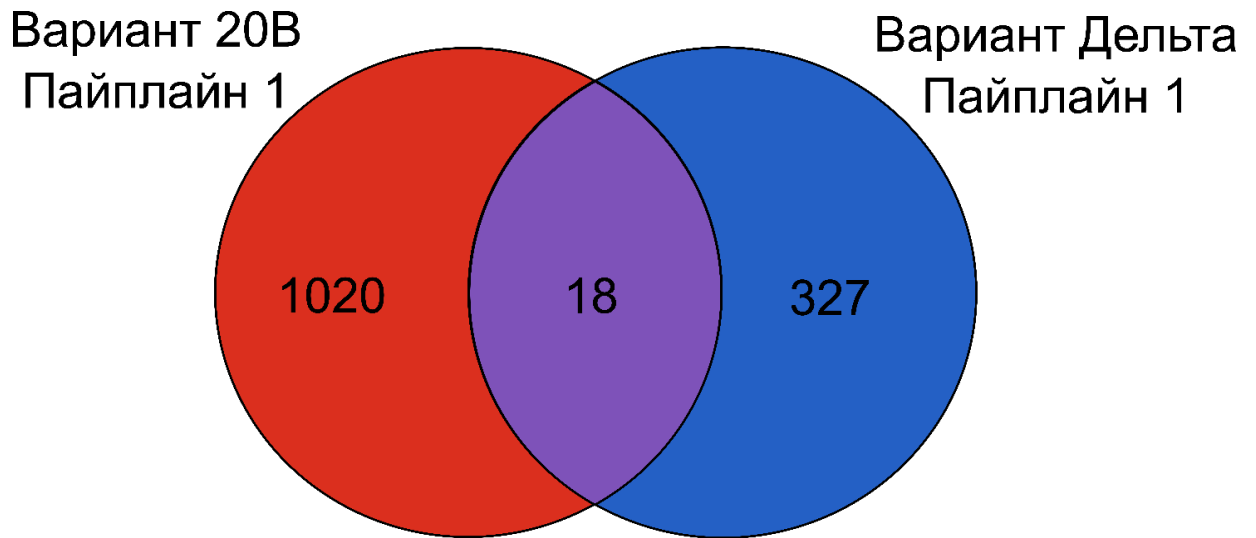


Рисунок 12. Диаграмма Венна по дифференциально экспрессирующимся генам между группами пациентов, зараженных вариантом SARS-CoV-2 20B и Дельта полученных в пайплайне 1.

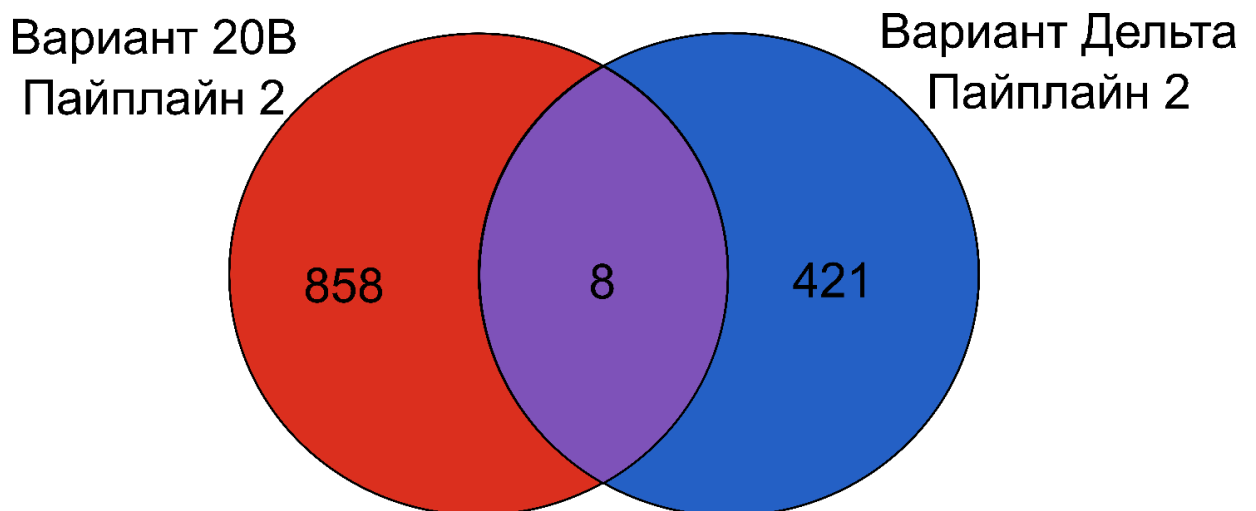


Рисунок 13. Диаграмма Венна по дифференциально экспрессирующимся генам между группами пациентов, зараженных вариантом SARS-CoV-2 20B и Дельта полученных в пайплайне 2.

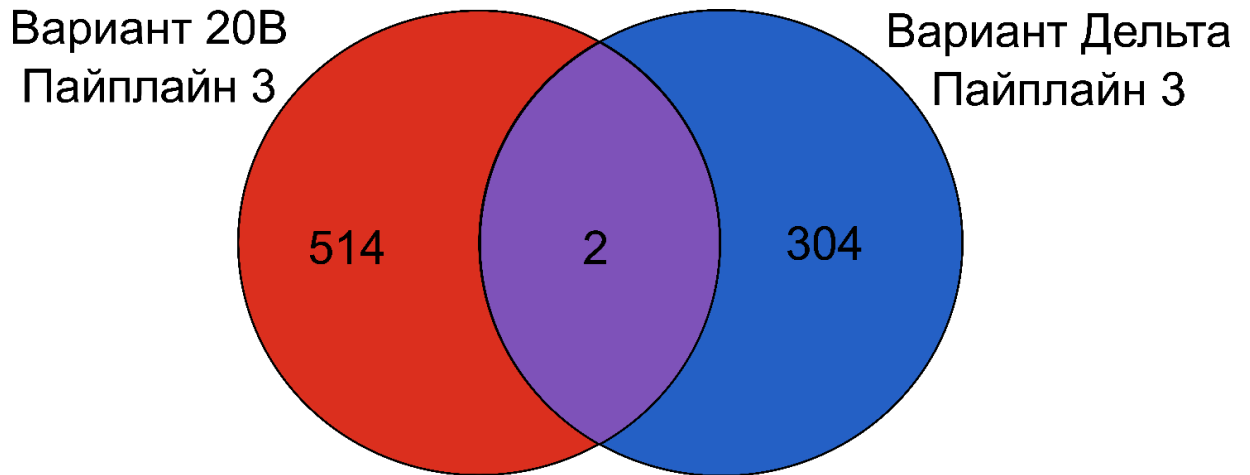


Рисунок 14. Диаграмма Венна по дифференциально экспрессирующимся генам между группами пациентов, зараженных вариантом SARS-CoV-2 20B и Дельта полученных в пайплайне 3.

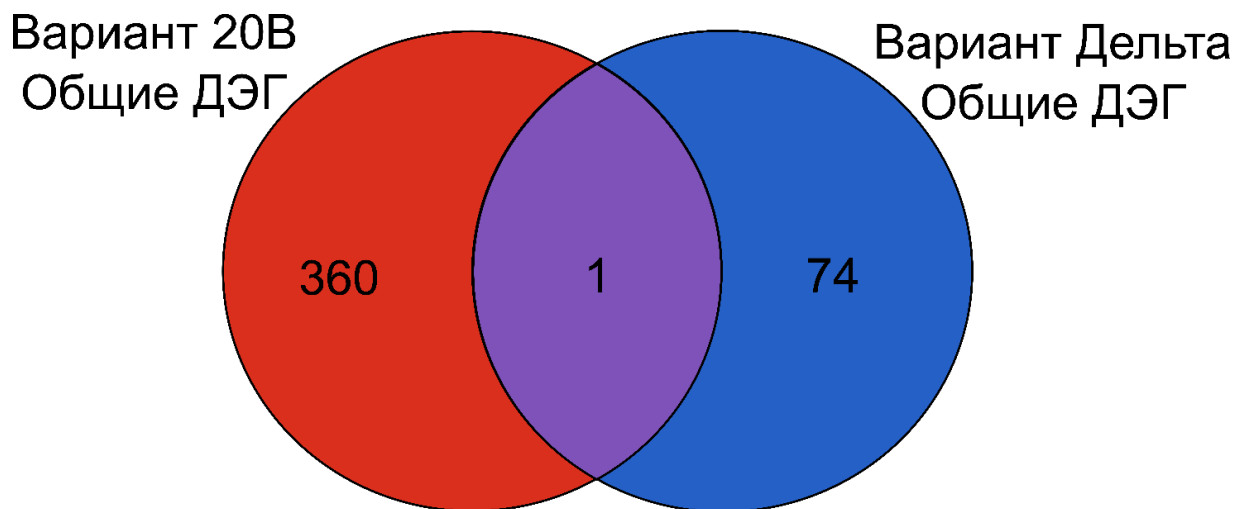


Рисунок 15. Диаграмма Венна по дифференциально экспрессирующимся генам между группами пациентов, зараженных вариантом SARS-CoV-2 20B и Дельта по генам, идентифицированным как дифференциально экспрессирующиеся во всех 3 пайплайнах.

К сожалению, поскольку оба этих источника различий полностью перекрываются в наших группах и имеют не случайную природу, с уверенностью сказать, который из них является причиной, невозможно. Однако, в рамках дальнейшей интерпретации полученных результатов, мы будем исходить из

предположения, что различия в экспрессии между группами как минимум частично объясняются особенностями различных вариантов вируса.

Варианты 20В и Дельта различаются различными наборами мутаций, в особенности в шиповидном гликопротеине S. Вариант B20 характеризуется двумя не синонимичными заменами в белке N SARS-CoV-2, R203K, G204R, и не несет никаких не синонимичных замен в шиповидном гликопротеине. по сравнению с оригинальным вариантом «Ухань» [292]. Дельта вариант характеризуется другим набором 7 миссенс мутаций (R158G, T478K, D950N в белке шиповидного гликопротеина, I82T в белке M, D63G в белке N, T120I в ORF7a, T60A в белке ORF9b) и делецией (156-157дел в белке S), и несколько из этих мутаций могут влиять на взаимодействие шиповидного гликопротеина с ACE2 и белками иммунной системы [7], [293], [294]. Однако роль этих мутаций не до конца понятна на данный момент. Например, не выявлено отличающих варианты Дельта и 20В мутаций, влияющих на структуру PLpro [295], [296]. Однако несмотря на то, что основным субстратом PLpro SARS-CoV-2 является ISG15, в группе зараженных 20В вариантом SARS-CoV-2 не выявляется дифференциальная экспрессия ISG15 между группами выживших и погибших.

6. ЗАКЛЮЧЕНИЕ

Кратко подводя итоги исследования, можно обобщить, что различия между погибшими и выжившими в группе пациентов с вариантом 20B SARS-CoV-2 преимущественно связаны с обменом липидов, а в группе пациентов с вариантом Дельта – с регуляцией иммунных процессов, и в особенности негативной регуляции иммунных процессов и регуляции активации системы комплемента. Идея о значимости обмена липидов, и в частности, о даунрегуляции сигнального пути рецепторов липопротеинов низкой плотности для исхода COVID-19 была многократно подтверждена в независимых исследованиях, включая метаанализ 22 опубликованных статей [219], [297], [298]. Полученные нами результаты исследования профиля липидов также говорят в пользу значимости обмена липидов, и, в частности, обмена липопротеинов низкой плотности для определения исходов заболевания в случае варианта 20B, но не варианта Дельта.

Одной из ключевых особенностей варианта Дельта, которые могут объяснить смещение ДЭГ от связанных с метаболизмом липидов генов к генам, связанным с иммунной системой, представляется улучшенное (по сравнению с вариантом 20B и Уханьским вариантом) иммунное избегание, характеризующее более поздние варианты SARS-CoV-2, в том числе вариант Дельта [299]. Логично предположить, что с улучшением способности к иммунному избеганию у вируса, влияние различий в тонкой регуляции иммунной системы хозяина, тоже будет расти. Для того, чтобы лучше понять, какую роль мутации SARS-CoV-2 играют в патогенезе и определении исхода COVID-19, необходимо проводить серийные исследования постоянно эволюционирующего вируса в реплицируемых обстоятельствах, таких как характеристики пациента (пол, возраст, вес и др.), получаемое лечение, статус по вакцинации. Безусловно, всегда будут существовать переменные, находящиеся вне контроля исследователей (в особенности учитывая, что новые варианты сменяют друг друга, и зачастую достаточно недолго сосуществуют в один и тот же временной период), однако, тем не менее, необходимо прилагать усилия для воспроизведения дизайнов эксперимента так, чтобы минимизировать влияние этих

запутывающих факторов. В современном мире, где доминируют более заразные и менее вирулентные подварианты варианта Омикрон, и с улучшением стандартов и возможностей систем здравоохранения, использованный в данной работе дизайн (выживших против погибших) становится все труднее и труднее воспроизвести. Однако, учитывая, что COVID-19 стал неотъемлемой частью современной жизни, другие дизайны, как, например, сравнение тяжелых случаев с легкими, все еще доступны, и поэтому необходимо продолжать подобные исследования, чтобы уровень понимания механизмов патогенеза COVID-19 успевал за бесконечно эволюционирующим вирусом.

7. ВЫВОДЫ

1. Использование трех различных наборов биоинформатических инструментов для транскриптомного анализа позволяет выявлять высоко сходные между различными наборами инструментов значимо обогащенные термины Gene Ontology: Biological Processes и общие для разных алгоритмов дифференциально экспрессирующиеся гены для дальнейшего анализа.
2. Сравнительный полнотранскриптомный анализ экспрессии генов в мононуклеарных лейкоцитах периферической крови пациентов с COVID-19, вызванным 20B вариантом вируса SARS-Cov-2, показал, что у пациентов с летальными и нелетальными исходами отличается экспрессия генов, связанных с активностью рецепторов липопротеинов низкой плотности по аннотации Gene Ontology.
3. Детальный анализ обогащения группы генов, связанных с активностью рецепторов липопротеинов низкой плотности по аннотации Gene Ontology, выявил следующие ключевые гены *PPARG*, *CD36*, *STAB1*, *ITGAV* и *ANXA2*, ассоциированные не менее, чем с 3 различными значимо обогащенными терминами Gene Ontology.
4. Сравнительный полнотранскриптомный анализ экспрессии генов в мононуклеарных лейкоцитах периферической крови пациентов с COVID-19, вызванным Дельта вариантом вируса SARS-Cov-2, показал, что у выживших и умерших пациентов отличается экспрессия генов, связанных с регуляцией иммунных процессов и системы комплемента по аннотации Gene Ontology.
5. Детальный анализ обогащения группы генов, связанных с регуляцией иммунных процессов и системы комплемента по аннотации Gene Ontology выявил следующие ключевые гены *ISG15*, *SERPING1*, *C1QA*, *C1QB*, и *VSIG4*, ассоциированные не менее чем с 4 различными значимо обогащенными терминами Gene Ontology.
6. Анализ концентрации липидов в плазме крови пациентов с летальными и нелетальными исходами COVID-19, вызванного вариантами 20B и Дельта

SARS-CoV-2, выявил тенденцию в различии концентрации липопротеинов низкой плотности в случае варианта 20B, но не варианта Дельта.

8. СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

Балк – не одноклеточный RNA-seq (“Bulk”)

ВОЗ – всемирная организация здравоохранения

ДЭГ – дифференциально экспрессирующийся ген

ЛПВП – липопротеины высокой плотности

ЛПНП - липопротеины низкой плотности

МПК – первичные мононуклеарные лейкоциты крови

МПС – масштабное/массовое параллельное секвенирование

ОРИТ – отделение реанимации и интенсивной терапии

ОХ – Общий холестерин

ОШ – отношение шансов

П.н. – пара нуклеотидов

Пайплайн – набор биоинформатических инструментов

ТГ - триглицериды

BWT – Burrows–Wheeler transform, (трансформация Барроуса-Уиллера)

ChIPseq – chromatin immunoprecipitation sequencing (секвенирование с иммунопреципитацией хроматина).

GEO – Gene Expression Omnibus (Библиотека экспрессии генов)

GO – Gene Ontology (Онтология генов)

GO:BP – Gene Ontology:Biological Processes (Онтология генов, раздел «Биологические процессы»)

MERS – Middle East Respiratory Syndrome (Ближневосточный респираторный синдром)

NGS – Next generation sequencing (секвенирование следующего поколения)

RNA-seq – RNA sequencing (масштабное параллельное секвенирование РНК)

SARS – Severe acute respiratory syndrome (тяжёлый острый респираторный синдром)

SARS-CoV – Severe acute respiratory syndrome-related coronavirus (коронавирус,

вызывающий тяжёлый острый респираторный синдром)

VBM – Variant being monitored (отслеживаемые варианты)

VOC – Variant of concern (варианты, вызывающие опасения)

VOI – Variant of interest (варианты, представляющие интерес)

VOHC – Variants of High Consequence (варианты, приводящие к значительным последствиям)

VUM – Variant under monitoring (отслеживаемые варианты)

9. СПИСОК ЛИТЕРАТУРЫ

1. C. L. Atzrodt, I. Maknojia, R. D. P. McCarthy, T. M. Oldfield, *et al.*, “A Guide to COVID-19: a global pandemic caused by the novel coronavirus SARS-CoV-2.,” *The FEBS journal*, vol. 287, no. 17, pp. 3633–3650, Sep. 2020, doi: 10.1111/febs.15375.
2. P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, *et al.*, “A pneumonia outbreak associated with a new coronavirus of probable bat origin.,” *Nature*, vol. 579, no. 7798, pp. 270–273, Mar. 2020, doi: 10.1038/s41586-020-2012-7.
3. WHO, “WHO Director-General’s opening remarks at the media briefing on COVID-19 - 11 March 2020,” 2020. [Online]. Available: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed Aug. 26, 2022).
4. WHO, “WHO Dashboard for COVID-19,” 2023. [Online]. Available: <https://covid19.who.int/> (accessed Aug. 26, 2022).
5. H. A. Rothan and S. N. Byrareddy, “The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak.,” *Journal of autoimmunity*, vol. 109, no. 2, p. 102433, Apr. 2020, doi: 10.1016/j.jaut.2020.102433.
6. C. Huang, Y. Wang, X. Li, L. Ren, *et al.*, “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.,” *Lancet (London, England)*, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.
7. J. Y. Choi and D. M. Smith, “SARS-CoV-2 Variants of Concern.,” *Yonsei medical journal*, vol. 62, no. 11, pp. 961–968, Nov. 2021, doi: 10.3349/ymj.2021.62.11.961.
8. N. G. Davies, S. Abbott, R. C. Barnard, C. I. Jarvis, *et al.*, “Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England.,” *Science (New York, N.Y.)*, vol. 372, no. 6538, Apr. 2021, doi: 10.1126/science.abg3055.
9. P. Wang, R. G. Casner, M. S. Nair, M. Wang, *et al.*, “Increased resistance of SARS-

- CoV-2 variant P.1 to antibody neutralization.,” *Cell host & microbe*, vol. 29, no. 5, pp. 747-751.e4, May 2021, doi: 10.1016/j.chom.2021.04.007.
10. M. McCallum, J. Bassi, A. De Marco, A. Chen, *et al.*, “SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern,” *Science (New York, N.Y.)*, vol. 373, no. 6555, pp. 648–654, Aug. 2021, doi: 10.1126/science.abi7994.
 11. D. Planas, D. Veyer, A. Baidaliuk, I. Staropoli, *et al.*, “Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization.,” *Nature*, vol. 596, no. 7871, pp. 276–280, Aug. 2021, doi: 10.1038/s41586-021-03777-9.
 12. A. Sheikh, J. McMEnamin, B. Taylor, C. Robertson, *et al.*, “SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness.,” *Lancet (London, England)*, vol. 397, no. 10293, pp. 2461–2462, Jun. 2021, doi: 10.1016/S0140-6736(21)01358-1.
 13. Public Health England, “SARS-CoV-2 variants of concern and variants under investigation in England,” *Sage*, no. April, pp. 1–50, 2021.
 14. A. Kucharski, N. Davies, R. Eggo, and S. Funk, “Modelling importations and local transmission of B.1.617.2 in the UK,” *CMMID COVID-19 Working Group*, no. May, pp. 1–7, 2021.
 15. H. Allen, A. Vusirikala, J. Flannagan, K. A. Twohig, *et al.*, “Increased household transmission of COVID-19 cases associated with SARS-CoV-2 Variant of Concern B.1.617.2: a national case-control study,” *Public Health England (PHE)*, pp. 1–21, 2021.
 16. S. W. X. Ong, C. J. Chiew, L. W. Ang, T. M. Mak, *et al.*, “Clinical and Virological Features of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variants of Concern: A Retrospective Cohort Study Comparing B.1.1.7 (Alpha), B.1.351 (Beta), and B.1.617.2 (Delta),” *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, vol. 75, no. 1, pp. e1128–e1136, 2022, doi: 10.1093/cid/ciab721.
 17. WHO, “Updated Risk Evaluation of JN.1, 15 April 2024,” [Online]. Available: <https://www.who.int/docs/default->

- source/coronaviruse/15042024_jn1_ure.pdf?sfvrsn=8bd19a5c_7. (accessed May 28, 2025)
18. WHO, “Updated Risk Evaluation of JN.1, 09 February 2024,” Feb. 2024. [Online]. Available: https://cdn.who.int/media/docs/default-source/documents/health-topics/sars/jn.1-9-february-2024.pdf?sfvrsn=9a39d825_3. (accessed May 30, 2025)
 19. V. A. Gushchin, A. A. Pochtovyi, D. D. Kustova, D. A. Ogarkova, *et al.*, “Dynamics of SARS-CoV-2 Major Genetic Lineages in Moscow in the Context of Vaccine Prophylaxis,” *International journal of molecular sciences*, vol. 23, no. 23, Nov. 2022, doi: 10.3390/ijms232314670.
 20. A. Gladkikh, V. Dedkov, A. Sharova, E. Klyuchnikova, *et al.*, “Epidemiological Features of COVID-19 in Northwest Russia in 2021.,” *Viruses*, vol. 14, no. 5, Apr. 2022, doi: 10.3390/v14050931.
 21. S. Behjati and P. S. Tarpey, “What is next generation sequencing?,” *Archives of disease in childhood. Education and practice edition*, vol. 98, no. 6, pp. 236–8, Dec. 2013, doi: 10.1136/archdischild-2013-304340.
 22. M. Ronaghi, M. Uhlén, and P. Nyren, “A sequencing method based on real-time pyrophosphate.,” *Science (New York, N.Y.)*, vol. 281, no. 5375, pp. 363, 365, Jul. 1998, doi: 10.1126/science.281.5375.363.
 23. J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, *et al.*, “Accurate multiplex polony sequencing of an evolved bacterial genome.,” *Science (New York, N.Y.)*, vol. 309, no. 5741, pp. 1728–32, Sep. 2005, doi: 10.1126/science.1117389.
 24. J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, *et al.*, “An integrated semiconductor device enabling non-optical genome sequencing.,” *Nature*, vol. 475, no. 7356, pp. 348–52, Jul. 2011, doi: 10.1038/nature10242.
 25. J. A. Reuter, D. V Spacek, and M. P. Snyder, “High-Throughput Sequencing Technologies,” *Molecular Cell*, vol. 58, no. 4, pp. 586–597, May 2015, doi: 10.1016/j.molcel.2015.05.004.
 26. M. G. Smith, T. A. Gianoulis, S. Pukatzki, J. J. Mekalanos, *et al.*, “New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density

- pyrosequencing and transposon mutagenesis.,” *Genes & development*, vol. 21, no. 5, pp. 601–14, Mar. 2007, doi: 10.1101/gad.1510307.
27. D. A. Wheeler, M. Srinivasan, M. Egholm, Y. Shen, *et al.*, “The complete genome of an individual by massively parallel DNA sequencing.,” *Nature*, vol. 452, no. 7189, pp. 872–6, Apr. 2008, doi: 10.1038/nature06884.
 28. T. H. Kim, Z. K. Abdullaev, A. D. Smith, K. A. Ching, *et al.*, “Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome.,” *Cell*, vol. 128, no. 6, pp. 1231–45, Mar. 2007, doi: 10.1016/j.cell.2006.12.048.
 29. M. J. Fullwood, M. H. Liu, Y. F. Pan, J. Liu, *et al.*, “An oestrogen-receptor- α -bound human chromatin interactome.,” *Nature*, vol. 462, no. 7269, pp. 58–64, Nov. 2009, doi: 10.1038/nature08497.
 30. E. Lieberman-Aiden, N. L. Van Berkum, L. Williams, M. Imakaev, *et al.*, “Comprehensive mapping of long-range interactions reveals folding principles of the human genome.,” *Science*, vol. 326, no. 5950, pp. 289–293, Oct. 2009, doi: 10.1126/science.1181369.
 31. R. Hrdlickova, M. Toloue, and B. Tian, “RNA-Seq methods for transcriptome analysis.,” *Wiley interdisciplinary reviews. RNA*, vol. 8, no. 1, p. 13, Jan. 2017, doi: 10.1002/wrna.1364.
 32. C. M. Sharma, S. Hoffmann, F. Darfeuille, J. Reignier, *et al.*, “The primary transcriptome of the major human pathogen *Helicobacter pylori*.,” *Nature*, vol. 464, no. 7286, pp. 250–5, Mar. 2010, doi: 10.1038/nature08756.
 33. M. S. Ko, “An ‘equalized cDNA library’ by the reassociation of short double-stranded cDNAs.,” *Nucleic acids research*, vol. 18, no. 19, pp. 5705–11, Oct. 1990, doi: 10.1093/nar/18.19.5705.
 34. D. K. Nam, S. Lee, G. Zhou, X. Cao, *et al.*, “Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 9, pp. 6152–6, Apr. 2002, doi: 10.1073/pnas.092140899.

35. V. Bhargava, P. Ko, E. Willems, M. Mercola, *et al.*, “Quantitative transcriptomics using designed primer-based amplification,” *Scientific reports*, vol. 3, p. 1740, 2013, doi: 10.1038/srep01740.
36. C. D. Armour, J. C. Castle, R. Chen, T. Babak, *et al.*, “Digital transcriptome profiling using selective hexamer priming for cDNA synthesis,” *Nature methods*, vol. 6, no. 9, pp. 647–9, Sep. 2009, doi: 10.1038/nmeth.1360.
37. S. Picelli, A. K. Björklund, B. Reinius, S. Sagasser, *et al.*, “Tn5 transposase and tagmentation procedures for massively scaled sequencing projects,” *Genome research*, vol. 24, no. 12, pp. 2033–40, Dec. 2014, doi: 10.1101/gr.177881.114.
38. P. J. A. Cock, C. J. Fields, N. Goto, M. L. Heuer, *et al.*, “The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants,” *Nucleic acids research*, vol. 38, no. 6, pp. 1767–71, Apr. 2010, doi: 10.1093/nar/gkp1137.
39. M. D. Macmanes, “On the optimal trimming of high-throughput mRNA sequence data,” *Frontiers in genetics*, vol. 5, no. JAN, p. 13, 2014, doi: 10.3389/fgene.2014.00013.
40. C. R. Williams, A. Baccarella, J. Z. Parrish, and C. C. Kim, “Trimming of sequence reads alters RNA-Seq gene expression estimates,” *BMC bioinformatics*, vol. 17, no. 1, p. 103, Feb. 2016, doi: 10.1186/s12859-016-0956-2.
41. Seqc/Maqc-Iii Consortium, “A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium,” *Nature Biotechnology*, vol. 32, no. 9, pp. 903–914, 2014, doi: 10.1038/nbt.2957.
42. Y. Liao and W. Shi, “Read trimming is not required for mapping and quantification of RNA-seq reads at the gene level,” *NAR genomics and bioinformatics*, vol. 2, no. 3, p. lqaa068, Sep. 2020, doi: 10.1093/nargab/lqaa068.
43. S. Schbath, V. Martin, M. Zytnecki, J. Fayolle, *et al.*, “Mapping reads on a genomic sequence: An algorithmic overview and a practical comparative analysis,” *Journal of Computational Biology*, vol. 19, no. 6, pp. 796–813, Jun. 2012, doi:

- 10.1089/cmb.2012.0022.
44. Z. Ning, A. J. Cox, and J. C. Mullikin, "SSAHA: A fast search method for large DNA databases," *Genome Research*, vol. 11, no. 10, pp. 1725–1729, 2001, doi: 10.1101/gr.194201.
 45. G. Lunter and M. Goodson, "Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads," *Genome Research*, vol. 21, no. 6, pp. 936–939, Jun. 2011, doi: 10.1101/gr.111120.110.
 46. H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform.," *Bioinformatics (Oxford, England)*, vol. 25, no. 14, pp. 1754–60, Jul. 2009, doi: 10.1093/bioinformatics/btp324.
 47. P. Ferragina and G. Manzini, "Opportunistic data structures with applications," in *Proceedings 41st Annual Symposium on Foundations of Computer Science*, 2000, pp. 390–398, doi: 10.1109/SFCS.2000.892127.
 48. B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with Bowtie 2.," *Nature methods*, vol. 9, no. 4, pp. 357–9, Mar. 2012, doi: 10.1038/nmeth.1923.
 49. R. Li, C. Yu, Y. Li, T. W. Lam, *et al.*, "SOAP2: An improved ultrafast tool for short read alignment," *Bioinformatics*, vol. 25, no. 15, pp. 1966–1967, Aug. 2009, doi: 10.1093/bioinformatics/btp336.
 50. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, *et al.*, "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions," *Genome Biology*, vol. 14, no. 4, p. R36, 2013, doi: 10.1186/gb-2013-14-4-r36.
 51. C. Trapnell, A. Roberts, L. Goff, G. Pertea, *et al.*, "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.," *Nature protocols*, vol. 7, no. 3, pp. 562–78, Mar. 2012, doi: 10.1038/nprot.2012.016.
 52. F. Rapaport, R. Khanin, Y. Liang, M. Pirun, *et al.*, "Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data.," *Genome biology*, vol. 14, no. 9, p. R95, 2013, doi: 10.1186/gb-2013-14-9-r95.
 53. M. Teng, M. I. Love, C. A. Davis, S. Djebali, *et al.*, "A benchmark for RNA-seq quantification pipelines.," *Genome biology*, vol. 17, no. 1, p. 74, Apr. 2016, doi:

- 10.1186/s13059-016-0940-1.
54. D. Kim, J. M. Paggi, C. Park, C. Bennett, *et al.*, “Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.,” *Nature biotechnology*, vol. 37, no. 8, pp. 907–915, Aug. 2019, doi: 10.1038/s41587-019-0201-4.
 55. U. Manber and G. Myers, “Suffix Arrays: A New Method for On-Line String Searches,” *SIAM Journal on Computing*, vol. 22, no. 5, pp. 935–948, Oct. 1993, doi: 10.1137/0222058.
 56. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, *et al.*, “STAR: ultrafast universal RNA-seq aligner.,” *Bioinformatics (Oxford, England)*, vol. 29, no. 1, pp. 15–21, Jan. 2013, doi: 10.1093/bioinformatics/bts635.
 57. Y. Liao, G. K. Smyth, and W. Shi, “The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads.,” *Nucleic acids research*, vol. 47, no. 8, p. e47, May 2019, doi: 10.1093/nar/gkz114.
 58. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, *et al.*, “Mapping and quantifying mammalian transcriptomes by RNA-Seq.,” *Nature methods*, vol. 5, no. 7, pp. 621–8, Jul. 2008, doi: 10.1038/nmeth.1226.
 59. B. Li, V. Ruotti, R. M. Stewart, J. A. Thomson, *et al.*, “RNA-Seq gene expression estimation with read mapping uncertainty.,” *Bioinformatics (Oxford, England)*, vol. 26, no. 4, pp. 493–500, Feb. 2010, doi: 10.1093/bioinformatics/btp692.
 60. J. C. Marioni, C. E. Mason, S. M. Mane, M. Stephens, *et al.*, “RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays,” *Genome Research*, vol. 18, no. 9, pp. 1509–1517, 2008, doi: 10.1101/gr.079558.108.
 61. U. Nagalakshmi, Z. Wang, K. Waern, C. Shou, *et al.*, “The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing,” *Science*, vol. 320, no. 5881, pp. 1344–1349, 2008, doi: 10.1126/science.1158441.
 62. R. Morin, M. Bainbridge, A. Fejes, M. Hirst, *et al.*, “Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing.,” *BioTechniques*, vol. 45, no. 1, pp. 81–94, Jul. 2008, doi:

- 10.2144/000112900.
63. G. J. Faulkner, A. R. R. Forrest, A. M. Chalk, K. Schroder, *et al.*, “A rescue strategy for multimapping short sequence tags refines surveys of transcriptional activity by CAGE,” *Genomics*, vol. 91, no. 3, pp. 281–288, 2008, doi: 10.1016/j.ygeno.2007.11.003.
 64. B. Li and C. N. Dewey, “RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.,” *BMC bioinformatics*, vol. 12, no. 1, p. 323, Aug. 2011, doi: 10.1186/1471-2105-12-323.
 65. M.-A. Dillies, A. Rau, J. Aubert, C. Hennequet-Antier, *et al.*, “A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis.,” *Briefings in bioinformatics*, vol. 14, no. 6, pp. 671–83, Nov. 2013, doi: 10.1093/bib/bbs046.
 66. J. H. Bullard, E. Purdom, K. D. Hansen, and S. Dudoit, “Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments.,” *BMC bioinformatics*, vol. 11, no. 1, p. 94, Feb. 2010, doi: 10.1186/1471-2105-11-94.
 67. B. M. Bolstad, R. A. Irizarry, M. Astrand, and T. P. Speed, “A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.,” *Bioinformatics (Oxford, England)*, vol. 19, no. 2, pp. 185–93, Jan. 2003, doi: 10.1093/bioinformatics/19.2.185.
 68. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, *et al.*, “limma powers differential expression analyses for RNA-sequencing and microarray studies.,” *Nucleic acids research*, vol. 43, no. 7, p. e47, Apr. 2015, doi: 10.1093/nar/gkv007.
 69. S. Anders and W. Huber, “Differential expression analysis for sequence count data,” *Genome Biology*, vol. 11, p. R106, 2010, doi: 10.1186/gb-2010-11-10-r106.
 70. D. J. McCarthy, Y. Chen, and G. K. Smyth, “Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation.,” *Nucleic acids research*, vol. 40, no. 10, pp. 4288–97, May 2012, doi: 10.1093/nar/gks042.
 71. M. I. Love, W. Huber, and S. Anders, “Moderated estimation of fold change and

- dispersion for RNA-seq data with DESeq2.,” *Genome biology*, vol. 15, no. 12, p. 550, Dec. 2014, doi: 10.1186/s13059-014-0550-8.
72. H. H. Thygesen and A. H. Zwinderman, “Modeling Sage data with a truncated gamma-Poisson model.,” *BMC bioinformatics*, vol. 7, no. 1, p. 157, Mar. 2006, doi: 10.1186/1471-2105-7-157.
73. M. D. Robinson and G. K. Smyth, “Small-sample estimation of negative binomial dispersion, with applications to SAGE data.,” *Biostatistics (Oxford, England)*, vol. 9, no. 2, pp. 321–32, Apr. 2008, doi: 10.1093/biostatistics/kxm030.
74. A. L. Oberg, B. M. Bot, D. E. Grill, G. A. Poland, *et al.*, “Technical and biological variance structure in mRNA-Seq data: life in the real world.,” *BMC genomics*, vol. 13, no. 1, p. 304, Jul. 2012, doi: 10.1186/1471-2164-13-304.
75. H. Wu, C. Wang, and Z. Wu, “A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data.,” *Biostatistics (Oxford, England)*, vol. 14, no. 2, pp. 232–43, Apr. 2013, doi: 10.1093/biostatistics/kxs033.
76. M. D. Robinson and G. K. Smyth, “Moderated statistical tests for assessing differences in tag abundance.,” *Bioinformatics (Oxford, England)*, vol. 23, no. 21, pp. 2881–7, Nov. 2007, doi: 10.1093/bioinformatics/btm453.
77. M. D. Robinson, D. J. McCarthy, and G. K. Smyth, “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.,” *Bioinformatics (Oxford, England)*, vol. 26, no. 1, pp. 139–40, Jan. 2010, doi: 10.1093/bioinformatics/btp616.
78. G. K. Smyth, J. Michaud, and H. S. Scott, “Use of within-array replicate spots for assessing differential expression in microarray experiments.,” *Bioinformatics (Oxford, England)*, vol. 21, no. 9, pp. 2067–75, May 2005, doi: 10.1093/bioinformatics/bti270.
79. M. E. Ritchie, D. Diyagama, J. Neilson, R. van Laar, *et al.*, “Empirical array quality weights in the analysis of microarray data.,” *BMC bioinformatics*, vol. 7, p. 261, May 2006, doi: 10.1186/1471-2105-7-261.
80. G. K. Smyth, “limma: Linear Models for Microarray Data,” in *Bioinformatics and*

- Computational Biology Solutions Using R and Bioconductor*, New York: Springer-Verlag, 2005, pp. 397–420.
81. E. Domingo, “Part I. Introduction,” in *Virus Taxonomy*, Elsevier, 2012, pp. 1–20.
 82. P. Li, T. Positive, S. Single, and S. Rna, “The Positive Sense Single Stranded RNA Viruses,” in *Virus Taxonomy*, no. Figure 1, Elsevier, 2005, pp. 739–1128.
 83. J. Cui, F. Li, and Z.-L. Shi, “Origin and evolution of pathogenic coronaviruses.,” *Nature reviews. Microbiology*, vol. 17, no. 3, pp. 181–192, Mar. 2019, doi: 10.1038/s41579-018-0118-9.
 84. N. Zhong, B. Zheng, Y. Li, L. Poon, *et al.*, “Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People’s Republic of China, in February, 2003,” *The Lancet*, vol. 362, no. 9393, pp. 1353–1358, Oct. 2003, doi: 10.1016/S0140-6736(03)14630-2.
 85. R. A. M. Fouchier, T. Kuiken, M. Schutten, G. van Amerongen, *et al.*, “Aetiology: Koch’s postulates fulfilled for SARS virus.,” *Nature*, vol. 423, no. 6937, p. 240, May 2003, doi: 10.1038/423240a.
 86. T. M. Rivers, “Viruses and Koch’s Postulates.,” *Journal of bacteriology*, vol. 33, no. 1, pp. 1–12, Jan. 1937, doi: 10.1128/jb.33.1.1-12.1937.
 87. ICTV, “Proposal FT 2003.150V,” 2003. [Online]. Available: <https://ictv.global/ictv/proposals/2003.V150.Coronaviridae.pdf>. (accessed Aug. 24, 2022)
 88. R. J. de Groot, J. Ziebuhr, L. L. Poon, P. C. Woo, *et al.*, “Proposal 2008.085- 126V,” 2008. [Online]. Available: <https://ictv.global/ictv/proposals/2008.085-122V.v4.Coronaviridae.pdf>. (accessed Aug. 24, 2022)
 89. A. M. Zaki, S. van Boheemen, T. M. Bestebroer, A. D. M. E. Osterhaus, *et al.*, “Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia.,” *The New England journal of medicine*, vol. 367, no. 19, pp. 1814–20, Nov. 2012, doi: 10.1056/NEJMoa1211721.
 90. R. J. de Groot, S. C. Baker, R. S. Baric, C. S. Brown, *et al.*, “Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study

- Group.,” *Journal of virology*, vol. 87, no. 14, pp. 7790–2, Jul. 2013, doi: 10.1128/JVI.01244-13.
91. J. Ziebuhr, R. S. Baric, S. Baker, R. J. de Groot, *et al.*, “Proposal 2015.003a-eS,” no. June, pp. 1–12, 2015, [Online]. Available: https://ictv.global/ictv/proposals/2015.003a-eS.A.v2.Coronaviridae_12sp.pdf. (accessed Aug. 24, 2022)
 92. “Middle East respiratory syndrome coronavirus (MERS-CoV) - Saudi Arabia,” 2019. <https://www.who.int/emergencies/disease-outbreak-news/item/2019-DON173> (accessed Aug. 23, 2022).
 93. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, “The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2.,” *Nature microbiology*, vol. 5, no. 4, pp. 536–544, Apr. 2020, doi: 10.1038/s41564-020-0695-z.
 94. Y. Jin, H. Yang, W. Ji, W. Wu, *et al.*, “Virology, Epidemiology, Pathogenesis, and Control of COVID-19.,” *Viruses*, vol. 12, no. 4, p. 372, Mar. 2020, doi: 10.3390/v12040372.
 95. T. T.-Y. Lam, N. Jia, Y.-W. Zhang, M. H.-H. Shum, *et al.*, “Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins.,” *Nature*, vol. 583, no. 7815, pp. 282–285, Jul. 2020, doi: 10.1038/s41586-020-2169-0.
 96. CDC, “SARS-CoV-2 Variant Classifications and Definitions,” 2023. [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html>. (accessed Aug. 24, 2022)
 97. WHO, “Tracking SARS-CoV-2 variants,” 2023. [Online]. Available: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> (accessed Aug. 26, 2023).
 98. WHO, “Statement on the update of WHO’s working definitions and tracking system for SARS-CoV-2 variants of concern and variants of interest,” 2023 [Online]. Available: <https://www.who.int/news/item/16-03-2023-statement-on-the-update-of-who-s-working-definitions-and-tracking-system-for-sars-cov-2-variants-of->

concern-and-variants-of-interest, (accessed Mar. 23, 2023)

99. P. Nitin, R. Nandhakumar, B. Vidhya, S. Rajesh, *et al.*, “COVID-19: Invasion, pathogenesis and possible cure – A review,” *Journal of Virological Methods*, vol. 300, no. January, p. 114434, Feb. 2022, doi: 10.1016/j.jviromet.2021.114434.
100. A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, *et al.*, “Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.,” *Cell*, vol. 181, no. 2, pp. 281-292.e6, Apr. 2020, doi: 10.1016/j.cell.2020.02.058.
101. M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, *et al.*, “SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.,” *Cell*, vol. 181, no. 2, pp. 271-280.e8, Apr. 2020, doi: 10.1016/j.cell.2020.02.052.
102. K. Wang, W. Chen, Z. Zhang, Y. Deng, *et al.*, “CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells.,” *Signal transduction and targeted therapy*, vol. 5, no. 1, p. 283, Dec. 2020, doi: 10.1038/s41392-020-00426-x.
103. I. M. Ibrahim, D. H. Abdelmalek, M. E. Elshahat, and A. A. Elfiky, “COVID-19 spike-host cell receptor GRP78 binding site prediction.,” *The Journal of infection*, vol. 80, no. 5, pp. 554–562, May 2020, doi: 10.1016/j.jinf.2020.02.026.
104. R. Zang, M. F. G. Castro, B. T. McCune, Q. Zeng, *et al.*, “TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes,” *Science Immunology*, vol. 5, no. 47, pp. 1–14, 2020, doi: 10.1126/sciimmunol.abc3582.
105. R. Essalmani, J. Jain, D. Susan-Resiga, U. Andréo, *et al.*, “Distinctive Roles of Furin and TMPRSS2 in SARS-CoV-2 Infectivity.,” *Journal of virology*, vol. 96, no. 8, p. e0012822, Apr. 2022, doi: 10.1128/jvi.00128-22.
106. N. Zhu, D. Zhang, W. Wang, X. Li, *et al.*, “A Novel Coronavirus from Patients with Pneumonia in China, 2019.,” *The New England journal of medicine*, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.
107. J. S. Ho, C.-H. Sia, M. Y. Chan, W. Lin, *et al.*, “Coronavirus-induced myocarditis: A meta-summary of cases.,” *Heart & lung : the journal of critical care*, vol. 49, no. 6, pp. 681–685, Nov. 2020, doi: 10.1016/j.hrtlng.2020.08.013.

108. P. Verdecchia, C. Cavallini, A. Spanevello, and F. Angeli, “The pivotal link between ACE2 deficiency and SARS-CoV-2 infection,” *European journal of internal medicine*, vol. 76, no. January, pp. 14–20, Jun. 2020, doi: 10.1016/j.ejim.2020.04.037.
109. M. Pennisi, G. Lanza, L. Falzone, F. Fisicaro, *et al.*, “SARS-CoV-2 and the Nervous System: From Clinical Features to Molecular Mechanisms,” *International journal of molecular sciences*, vol. 21, no. 15, p. 5475, Jul. 2020, doi: 10.3390/ijms21155475.
110. E. Song, C. Zhang, B. Israelow, A. Lu-Culligan, *et al.*, “Neuroinvasion of SARS-CoV-2 in human and mouse brain,” *The Journal of experimental medicine*, vol. 218, no. 3, Mar. 2021, doi: 10.1084/jem.20202135.
111. Q. Fernandes, V. P. Inchakalody, M. Merhi, S. Mestiri, *et al.*, “Emerging COVID-19 variants and their impact on SARS-CoV-2 diagnosis, therapeutics and vaccines,” *Annals of medicine*, vol. 54, no. 1, pp. 524–540, Dec. 2022, doi: 10.1080/07853890.2022.2031274.
112. A. Chaillon and D. M. Smith, “Phylogenetic Analyses of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) B.1.1.7 Lineage Suggest a Single Origin Followed by Multiple Exportation Events Versus Convergent Evolution,” *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, vol. 73, no. 12, pp. 2314–2317, Dec. 2021, doi: 10.1093/cid/ciab265.
113. J. W. Tang, O. T. R. Toovey, K. N. Harvey, and D. D. S. Hui, “Introduction of the South African SARS-CoV-2 variant 501Y.V2 into the UK,” *The Journal of infection*, vol. 82, no. 4, pp. e8–e10, Apr. 2021, doi: 10.1016/j.jinf.2021.01.007.
114. T. Funk, A. Pharris, G. Spiteri, N. Bundle, *et al.*, “Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021,” *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, vol. 26, no. 16, pp. 1–10, Apr. 2021, doi: 10.2807/1560-7917.ES.2021.26.16.2100348.
115. S. Cele, I. Gazy, L. Jackson, S.-H. Hwa, *et al.*, “Escape of SARS-CoV-2 501Y.V2

- from neutralization by convalescent plasma.,” *Nature*, vol. 593, no. 7857, pp. 142–146, May 2021, doi: 10.1038/s41586-021-03471-w.
116. C. K. Wibmer, F. Ayres, T. Hermanus, M. Madzivhandila, *et al.*, “SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma.,” *Nature medicine*, vol. 27, no. 4, pp. 622–625, Apr. 2021, doi: 10.1038/s41591-021-01285-x.
117. Z. Wang, F. Schmidt, Y. Weisblum, F. Muecksch, *et al.*, “mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants.,” *Nature*, vol. 592, no. 7855, pp. 616–622, Apr. 2021, doi: 10.1038/s41586-021-03324-6.
118. X. Xie, Y. Liu, J. Liu, X. Zhang, *et al.*, “Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera.,” *Nature medicine*, vol. 27, no. 4, pp. 620–621, Apr. 2021, doi: 10.1038/s41591-021-01270-4.
119. E. Mahase, “Covid-19: Novavax vaccine efficacy is 86% against UK variant and 60% against South African variant.,” *BMJ (Clinical research ed.)*, vol. 372, no. February, p. n296, Feb. 2021, doi: 10.1136/bmj.n296.
120. T. Fujino, H. Nomoto, S. Kutsuna, M. Ujiie, *et al.*, “Novel SARS-CoV-2 Variant in Travelers from Brazil to Japan.,” *Emerging infectious diseases*, vol. 27, no. 4, pp. 1243–1245, Apr. 2021, doi: 10.3201/eid2704.210138.
121. T. Charmet, L. Schaeffer, R. Grant, S. Galmiche, *et al.*, “Impact of original, B.1.1.7, and B.1.351/P.1 SARS-CoV-2 lineages on vaccine effectiveness of two doses of COVID-19 mRNA vaccines: Results from a nationwide case-control study in France.,” *The Lancet regional health. Europe*, vol. 8, p. 100171, Sep. 2021, doi: 10.1016/j.lanepe.2021.100171.
122. X. Deng, M. A. Garcia-Knight, M. M. Khalid, V. Servellita, *et al.*, “Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant.,” *Cell*, vol. 184, no. 13, pp. 3426–3437.e8, Jun. 2021, doi: 10.1016/j.cell.2021.04.025.
123. Y. Araf, F. Akter, Y. Tang, R. Fatemi, *et al.*, “Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines.,” *Journal*

- of medical virology*, vol. 94, no. 5, pp. 1825–1832, May 2022, doi: 10.1002/jmv.27588.
124. L. Wang and G. Cheng, “Sequence analysis of the emerging SARS-CoV-2 variant Omicron in South Africa.,” *Journal of medical virology*, vol. 94, no. 4, pp. 1728–1733, Apr. 2022, doi: 10.1002/jmv.27516.
 125. B. Shanmugaraj, A. Malla, N. Khorattanakulchai, and W. Phoolcharoen, “SARS-CoV-2 omicron variant: Could it be another threat?,” *Journal of medical virology*, vol. 94, no. 4, pp. 1284–1288, Apr. 2022, doi: 10.1002/jmv.27532.
 126. S. Pascarella, M. Ciccozzi, M. Bianchi, D. Benvenuto, *et al.*, “The electrostatic potential of the Omicron variant spike is higher than in Delta and Delta-plus variants: A hint to higher transmissibility?,” *Journal of medical virology*, vol. 94, no. 4, pp. 1277–1280, Apr. 2022, doi: 10.1002/jmv.27528.
 127. M. Shah and H. G. Woo, “Omicron: A Heavily Mutated SARS-CoV-2 Variant Exhibits Stronger Binding to ACE2 and Potently Escapes Approved COVID-19 Therapeutic Antibodies.,” *Frontiers in immunology*, vol. 12, no. January, p. 830527, Jan. 2021, doi: 10.3389/fimmu.2021.830527.
 128. M. Hoffmann, N. Krüger, S. Schulz, A. Cossmann, *et al.*, “The Omicron variant is highly resistant against antibody-mediated neutralization: Implications for control of the COVID-19 pandemic.,” *Cell*, vol. 185, no. 3, pp. 447–456.e11, 2022, doi: 10.1016/j.cell.2021.12.032.
 129. WHO, “Statement on the update of WHO’s working definitions and tracking system for SARS-CoV-2 variants of concern and variants of interest,” 2023. [Online]. Available: <https://www.who.int/news/item/16-03-2023-statement-on-the-update-of-who-s-working-definitions-and-tracking-system-for-sars-cov-2-variants-of-concern-and-variants-of-interest> (accessed Mar. 16, 2023).
 130. WHO, “XBB.1.5 Updated Risk Assessment, 20 June 2023,” 2023. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/20230620xbb.1.5.pdf?sfvrsn=fff6f686_3. (accessed Jul. 25, 2023).

131. CDC, “COVID Data Tracker,” 2023. [Online]. Available: <https://covid.cdc.gov/covid-data-tracker/> (accessed Aug. 03, 2023).
132. Q. Wang, S. Iketani, Z. Li, L. Liu, *et al.*, “Alarming antibody evasion properties of rising SARS-CoV-2 BQ and XBB subvariants,” *Cell*, vol. 186, no. 2, pp. 279–286.e8, Jan. 2023, doi: 10.1016/j.cell.2022.12.018.
133. C. Yue, W. Song, L. Wang, F. Jian, *et al.*, “ACE2 binding and antibody evasion in enhanced transmissibility of XBB.1.5,” *The Lancet. Infectious diseases*, vol. 23, no. 3, pp. 278–280, Mar. 2023, doi: 10.1016/S1473-3099(23)00010-5.
134. R. P. Karyakarte, R. Das, S. Dudhate, J. Agarasen, *et al.*, “Clinical Characteristics and Outcomes of Laboratory-Confirmed SARS-CoV-2 Cases Infected With Omicron Subvariants and the XBB Recombinant Variant,” *Cureus*, vol. 15, no. 2, p. e35261, Feb. 2023, doi: 10.7759/cureus.35261.
135. R. Uraki, M. Ito, M. Kiso, S. Yamayoshi, *et al.*, “Antiviral and bivalent vaccine efficacy against an omicron XBB.1.5 isolate,” *The Lancet. Infectious diseases*, vol. 23, no. 4, pp. 402–403, Apr. 2023, doi: 10.1016/S1473-3099(23)00070-1.
136. WHO, “XBB.1.16 Updated Risk Assessment, 05 June 2023,” Jun. 2023.
137. D. Yamasoba, K. Uriu, A. Plianpaisuk, Y. Kosugi, *et al.*, “Virological characteristics of the SARS-CoV-2 omicron XBB.1.16 variant,” *The Lancet. Infectious diseases*, vol. 23, no. 6, pp. 655–656, Jun. 2023, doi: 10.1016/S1473-3099(23)00278-5.
138. R. P. Karyakarte, R. Das, M. V Rajmane, S. Dudhate, *et al.*, “Chasing SARS-CoV-2 XBB.1.16 Recombinant Lineage in India and the Clinical Profile of XBB.1.16 Cases in Maharashtra, India,” *Cureus*, Jun. 2023, doi: 10.7759/cureus.39816.
139. R. Pung, X. P. Kong, L. Cui, S.-R. Chae, *et al.*, “Severity of SARS-CoV-2 Omicron XBB subvariants in Singapore,” *The Lancet Regional Health - Western Pacific*, vol. 37, p. 100849, Aug. 2023, doi: 10.1016/j.lanwpc.2023.100849.
140. WHO, “EG.5 Initial Risk Evaluation, 9 August 2023,” Aug. 2023, Accessed: Aug. 18, 2023. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/09082023eg.5_ire_final.pdf?sfvrsn=2aa2daee_3. (accessed

Aug. 18, 2023)

141. A. Yisimayi, W. Song, J. Wang, F. Jian, *et al.*, “Repeated Omicron infection alleviates SARS-CoV-2 immune imprinting,” *bioRxiv*, p. 2023.05.01.538516, Jan. 2023, doi: 10.1101/2023.05.01.538516.
142. WHO, “Initial Risk Evaluation of JN.1, 19 December 2023,” 2023. Accessed: May 30, 2025. [Online]. Available: https://cdn.who.int/media/docs/default-source/documents/health-topics/sars/jn.1-9-february-2024.pdf?sfvrsn=9a39d825_3. (accessed May 30, 2025)
143. X. Zou, K. Chen, J. Zou, P. Han, *et al.*, “Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection.,” *Frontiers of medicine*, vol. 14, no. 2, pp. 185–192, Apr. 2020, doi: 10.1007/s11684-020-0754-0.
144. S. M. Hammoudeh, A. M. Hammoudeh, P. M. Bhamidimarri, B. Mahboub, *et al.*, “Insight into molecular mechanisms underlying hepatic dysfunction in severe COVID-19 patients using systems biology.,” *World journal of gastroenterology*, vol. 27, no. 21, pp. 2850–2870, Jun. 2021, doi: 10.3748/wjg.v27.i21.2850.
145. A. B. M. M. K. Islam, M. A.-A.-K. Khan, R. Ahmed, M. S. Hossain, *et al.*, “Transcriptome of nasopharyngeal samples from COVID-19 patients and a comparative analysis with other SARS-CoV-2 infection models reveal disparate host responses against SARS-CoV-2.,” *Journal of translational medicine*, vol. 19, no. 1, p. 32, Jan. 2021, doi: 10.1186/s12967-020-02695-0.
146. J. P. Bernardes, N. Mishra, F. Tran, T. Bahmer, *et al.*, “Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19.,” *Immunity*, vol. 53, no. 6, pp. 1296-1314.e9, Dec. 2020, doi: 10.1016/j.immuni.2020.11.017.
147. R. A. Grant, L. Morales-Nebreda, N. S. Markov, S. Swaminathan, *et al.*, “Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia.,” *Nature*, vol. 590, no. 7847, pp. 635–641, Feb. 2021, doi: 10.1038/s41586-020-03148-w.
148. H.-Y. Zheng, M. Xu, C.-X. Yang, R.-R. Tian, *et al.*, “Longitudinal transcriptome

- analyses show robust T cell immunity during recovery from COVID-19.,” *Signal transduction and targeted therapy*, vol. 5, no. 1, p. 294, Dec. 2020, doi: 10.1038/s41392-020-00457-4.
149. D. Butler, C. Mozsary, C. Meydan, J. Foox, *et al.*, “Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions.,” *Nature communications*, vol. 12, no. 1, p. 1660, Mar. 2021, doi: 10.1038/s41467-021-21361-7.
150. A. Hasankhani, A. Bahrami, N. Sheybani, B. Aria, *et al.*, “Differential Co-Expression Network Analysis Reveals Key Hub-High Traffic Genes as Potential Therapeutic Targets for COVID-19 Pandemic.,” *Frontiers in immunology*, vol. 12, no. December, p. 789317, 2021, doi: 10.3389/fimmu.2021.789317.
151. H. Medini, A. Zirman, and D. Mishmar, “Immune system cells from COVID-19 patients display compromised mitochondrial-nuclear expression co-regulation and rewiring toward glycolysis.,” *iScience*, vol. 24, no. 12, p. 103471, Dec. 2021, doi: 10.1016/j.isci.2021.103471.
152. Y. Li, A. Duche, M. R. Sayer, D. Roosan, *et al.*, “SARS-CoV-2 early infection signature identified potential key infection mechanisms and drug targets.,” *BMC genomics*, vol. 22, no. 1, p. 125, Feb. 2021, doi: 10.1186/s12864-021-07433-4.
153. J. Park, J. Foox, T. Hether, D. C. Danko, *et al.*, “System-wide transcriptome damage and tissue identity loss in COVID-19 patients.,” *Cell reports. Medicine*, vol. 3, no. 2, p. 100522, Feb. 2022, doi: 10.1016/j.xcrm.2022.100522.
154. H. Bräuninger, B. Stoffers, A. D. E. Fitzek, K. Meißner, *et al.*, “Cardiac SARS-CoV-2 infection is associated with pro-inflammatory transcriptomic alterations within the heart.,” *Cardiovascular research*, vol. 118, no. 2, pp. 542–555, Jan. 2022, doi: 10.1093/cvr/cvab322.
155. L. Knabl, H. K. Lee, M. Wieser, A. Mur, *et al.*, “BNT162b2 vaccination enhances interferon-JAK-STAT-regulated antiviral programs in COVID-19 patients infected with the SARS-CoV-2 Beta variant.,” *Communications medicine*, vol. 2, no. 1, pp. 1–10, 2022, doi: 10.1038/s43856-022-00083-x.

156. S. A. Bhat, T. Shibata, M. Leong, J. Plummer, *et al.*, “Comparative Upper Respiratory Tract Transcriptomic Profiling Reveals a Potential Role of Early Activation of Interferon Pathway in Severe COVID-19.,” *Viruses*, vol. 14, no. 10, p. 2182, Oct. 2022, doi: 10.3390/v14102182.
157. X. Wang, M. A. Sanborn, Y. Dai, and J. Rehman, “Temporal transcriptomic analysis using TrendCatcher identifies early and persistent neutrophil activation in severe COVID-19.,” *JCI insight*, vol. 7, no. 7, pp. 1–18, Apr. 2022, doi: 10.1172/jci.insight.157255.
158. Z. Dong, Q. Yan, W. Cao, Z. Liu, *et al.*, “Identification of key molecules in COVID-19 patients significantly correlated with clinical outcomes by analyzing transcriptomic data.,” *Frontiers in immunology*, vol. 13, no. August, p. 930866, 2022, doi: 10.3389/fimmu.2022.930866.
159. D. Das and S. Podder, “Deregulation of ceRNA Networks in Frontal Cortex and Choroid Plexus of Brain during SARS-CoV-2 Infection Aggravates Neurological Manifestations: An Insight from Bulk and Single-Cell Transcriptomic Analyses.,” *Advanced biology*, vol. 6, no. 8, p. e2101310, Aug. 2022, doi: 10.1002/adbi.202101310.
160. M. Fasan, A. Collese, V. Angerilli, M. Sbaraglia, *et al.*, “Multi-Design Differential Expression Profiling of COVID-19 Lung Autopsy Specimens Reveals Significantly Deregulated Inflammatory Pathways and SFTPC Impaired Transcription.,” *Cells*, vol. 11, no. 6, Mar. 2022, doi: 10.3390/cells11061011.
161. S. Maleknia, M. J. Tavassolifar, F. Mottaghitlab, M. R. Zali, *et al.*, “Identifying novel host-based diagnostic biomarker panels for COVID-19: a whole-blood/nasopharyngeal transcriptome meta-analysis.,” *Molecular medicine (Cambridge, Mass.)*, vol. 28, no. 1, p. 86, Aug. 2022, doi: 10.1186/s10020-022-00513-5.
162. F. C. Walker, P. R. Sridhar, and M. T. Baldrige, “Differential roles of interferons in innate responses to mucosal viral infections.,” *Trends in immunology*, vol. 42, no. 11, pp. 1009–1023, Nov. 2021, doi: 10.1016/j.it.2021.09.003.

163. P. Mertowska, K. Smolak, S. Mertowski, and E. Grywalska, “Immunomodulatory Role of Interferons in Viral and Bacterial Infections.,” *International journal of molecular sciences*, vol. 24, no. 12, p. 10115, Jun. 2023, doi: 10.3390/ijms241210115.
164. W. M. Schneider, M. D. Chevillotte, and C. M. Rice, “Interferon-stimulated genes: a complex web of host defenses.,” *Annual review of immunology*, vol. 32, pp. 513–45, 2014, doi: 10.1146/annurev-immunol-032713-120231.
165. A. Vanderheiden, P. Ralfs, T. Chirkova, A. A. Upadhyay, *et al.*, “Type I and Type III Interferons Restrict SARS-CoV-2 Infection of Human Airway Epithelial Cultures.,” *Journal of virology*, vol. 94, no. 19, Sep. 2020, doi: 10.1128/JVI.00985-20.
166. N. D. Maulding, S. Seiler, A. Pearson, N. Kreusser, *et al.*, “Dual RNA-Seq analysis of SARS-CoV-2 correlates specific human transcriptional response pathways directly to viral expression.,” *Scientific reports*, vol. 12, no. 1, p. 1329, Jan. 2022, doi: 10.1038/s41598-022-05342-4.
167. R. Alberts, S. C. Chan, Q.-F. Meng, S. He, *et al.*, “Integration and Reanalysis of Four RNA-Seq Datasets Including BALF, Nasopharyngeal Swabs, Lung Biopsy, and Mouse Models Reveals Common Immune Features of COVID-19.,” *Immune network*, vol. 22, no. 3, p. e22, Jun. 2022, doi: 10.4110/in.2022.22.e22.
168. M. Z. Hasan, S. Islam, K. Matsumoto, and T. Kawai, “SARS-CoV-2 infection initiates interleukin-17-enriched transcriptional response in different cells from multiple organs.,” *Scientific reports*, vol. 11, no. 1, p. 16814, Aug. 2021, doi: 10.1038/s41598-021-96110-3.
169. Y. Tanaka, K. Higashihara, M. A. Nakazawa, F. Yamashita, *et al.*, “Dynamic changes in gene-to-gene regulatory networks in response to SARS-CoV-2 infection.,” *Scientific reports*, vol. 11, no. 1, p. 11241, May 2021, doi: 10.1038/s41598-021-90556-1.
170. S. Pence, B. Caykara, H. H. Pence, S. Tekin, *et al.*, “Transcriptomic analysis of asymptomatic and symptomatic severe Turkish patients in SARS-CoV-2 infection.,” *Northern clinics of Istanbul*, vol. 9, no. 2, pp. 122–130, 2022, doi:

- 10.14744/nci.2022.28000.
171. L. T. de F. Cavalcante, G. C. da Fonseca, L. A. Amado Leon, A. L. Salvio, *et al.*, “Buffy Coat Transcriptomic Analysis Reveals Alterations in Host Cell Protein Synthesis and Cell Cycle in Severe COVID-19 Patients.,” *International journal of molecular sciences*, vol. 23, no. 21, p. 13588, Nov. 2022, doi: 10.3390/ijms232113588.
 172. S. Cohen, P. E. Bigazzi, and T. Yoshida, “Commentary. Similarities of T cell function in cell-mediated immunity and antibody production.,” *Cellular immunology*, vol. 12, no. 1, pp. 150–9, Apr. 1974, doi: 10.1016/0008-8749(74)90066-5.
 173. T. Rahman, A. Das, M. H. Abir, I. H. Nafiz, *et al.*, “Cytokines and their role as immunotherapeutics and vaccine Adjuvants: The emerging concepts,” *Cytokine*, vol. 169, p. 156268, Sep. 2023, doi: 10.1016/J.CYTO.2023.156268.
 174. R. Channappanavar and S. Perlman, “Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology.,” *Seminars in immunopathology*, vol. 39, no. 5, pp. 529–539, Jul. 2017, doi: 10.1007/s00281-017-0629-x.
 175. B. J. Webb, I. D. Peltan, P. Jensen, D. Hoda, *et al.*, “Clinical criteria for COVID-19-associated hyperinflammatory syndrome: a cohort study.,” *The Lancet. Rheumatology*, vol. 2, no. 12, pp. e754–e763, Dec. 2020, doi: 10.1016/S2665-9913(20)30343-X.
 176. M. Tian, W. Liu, X. Li, P. Zhao, *et al.*, “HIF-1 α promotes SARS-CoV-2 infection and aggravates inflammatory responses to COVID-19.,” *Signal transduction and targeted therapy*, vol. 6, no. 1, p. 308, Aug. 2021, doi: 10.1038/s41392-021-00726-w.
 177. M. G. Ferrarini, A. Lal, R. Rebollo, A. J. Gruber, *et al.*, “Genome-wide bioinformatic analyses predict key host and viral factors in SARS-CoV-2 pathogenesis.,” *Communications biology*, vol. 4, no. 1, p. 590, May 2021, doi: 10.1038/s42003-021-02095-0.

178. H. K. Lee, L. Knabl, L. Pipperger, A. Volland, *et al.*, “Immune transcriptomes of highly exposed SARS-CoV-2 asymptomatic seropositive versus seronegative individuals from the Ischgl community.” *Scientific reports*, vol. 11, no. 1, p. 4243, Feb. 2021, doi: 10.1038/s41598-021-83110-6.
179. Y. Zhang, X. Guo, C. Li, Z. Kou, *et al.*, “Transcriptome Analysis of Peripheral Blood Mononuclear Cells in SARS-CoV-2 Naïve and Recovered Individuals Vaccinated With Inactivated Vaccine.” *Frontiers in cellular and infection microbiology*, vol. 11, no. March 2020, p. 821828, 2021, doi: 10.3389/fcimb.2021.821828.
180. K. A. Overmyer, E. Shishkova, I. J. Miller, J. Balnis, *et al.*, “Large-Scale Multi-omic Analysis of COVID-19 Severity.” *Cell systems*, vol. 12, no. 1, pp. 23-40.e7, Jan. 2021, doi: 10.1016/j.cels.2020.10.003.
181. A. C. Aschenbrenner, M. Mouktaroudi, B. Krämer, M. Oestreich, *et al.*, “Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients.” *Genome medicine*, vol. 13, no. 1, p. 7, Jan. 2021, doi: 10.1186/s13073-020-00823-5.
182. Q. S. Wang, R. Edahiro, H. Namkoong, T. Hasegawa, *et al.*, “The whole blood transcriptional regulation landscape in 465 COVID-19 infected samples from Japan COVID-19 Task Force.” *Nature communications*, vol. 13, no. 1, p. 4830, Aug. 2022, doi: 10.1038/s41467-022-32276-2.
183. P. P. Freire, A. H. Marques, G. C. Baiocchi, L. F. Schimke, *et al.*, “The relationship between cytokine and neutrophil gene network distinguishes SARS-CoV-2-infected patients by sex and age.” *JCI insight*, vol. 6, no. 10, pp. 1–17, May 2021, doi: 10.1172/jci.insight.147535.
184. J. Fu, J. Kong, W. Wang, M. Wu, *et al.*, “The clinical implication of dynamic neutrophil to lymphocyte ratio and D-dimer in COVID-19: A retrospective study in Suzhou China.” *Thrombosis research*, vol. 192, pp. 3–8, Aug. 2020, doi: 10.1016/j.thromres.2020.05.006.
185. B. J. Barnes, J. M. Adrover, A. Baxter-Stoltzfus, A. Borczuk, *et al.*, “Targeting

- potential drivers of COVID-19: Neutrophil extracellular traps,” *The Journal of experimental medicine*, vol. 217, no. 6, Jun. 2020, doi: 10.1084/jem.20200652.
186. R. M. Golonka, P. Saha, B. S. Yeoh, S. Chattopadhyay, *et al.*, “Harnessing innate immunity to eliminate SARS-CoV-2 and ameliorate COVID-19 disease,” *Physiological genomics*, vol. 52, no. 5, pp. 217–221, May 2020, doi: 10.1152/physiolgenomics.00033.2020.
187. M. Laforge, C. Elbim, C. Frère, M. Hémadi, *et al.*, “Tissue damage from neutrophil-induced oxidative stress in COVID-19,” *Nature reviews. Immunology*, vol. 20, no. 9, pp. 515–516, Sep. 2020, doi: 10.1038/s41577-020-0407-1.
188. P. L. Meroni, S. Croci, P. A. Lonati, F. Pregnolato, *et al.*, “Complement activation predicts negative outcomes in COVID-19: The experience from Northern Italian patients,” *Autoimmunity reviews*, vol. 22, no. 1, p. 103232, Jan. 2023, doi: 10.1016/j.autrev.2022.103232.
189. Y. Xiong, Y. Liu, L. Cao, D. Wang, *et al.*, “Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients,” *Emerging Microbes and Infections*, vol. 9, no. 1, pp. 761–770, 2020, doi: 10.1080/22221751.2020.1747363.
190. T. Välikangas, S. Junntila, K. T. Rytönen, A. Kukkonen-Macchi, *et al.*, “COVID-19-specific transcriptomic signature detectable in blood across multiple cohorts,” *Frontiers in genetics*, vol. 13, no. August, p. 929887, Aug. 2022, doi: 10.3389/fgene.2022.929887.
191. U. Fahnøe, A. Ronit, R. M. G. Berg, S. E. Jørgensen, *et al.*, “A Distinct Dexamethasone-Dependent Gene Expression Profile in the Lungs of COVID-19 Patients,” *The Journal of infectious diseases*, vol. 226, no. 12, pp. 2137–2141, Dec. 2022, doi: 10.1093/infdis/jiac218.
192. J. Mellors, T. Tipton, S. Longet, and M. Carroll, “Viral Evasion of the Complement System and Its Importance for Vaccines and Therapeutics,” *Frontiers in immunology*, vol. 11, p. 1450, 2020, doi: 10.3389/fimmu.2020.01450.
193. B. Afzali, M. Noris, B. N. Lambrecht, and C. Kemper, “The state of complement in

- COVID-19.,” *Nature reviews. Immunology*, vol. 22, no. 2, pp. 77–84, Feb. 2022, doi: 10.1038/s41577-021-00665-1.
194. C. E. Samuel, “Interferon at the crossroads of SARS-CoV-2 infection and COVID-19 disease.,” *The Journal of biological chemistry*, p. 104960, Jun. 2023, doi: 10.1016/j.jbc.2023.104960.
 195. S. Eshetie, P. Jullian, B. Benyamin, and S. H. Lee, “Host genetic determinants of COVID-19 susceptibility and severity: A systematic review and meta-analysis.,” *Reviews in medical virology*, p. e2466, Jun. 2023, doi: 10.1002/rmv.2466.
 196. R. N. Vuille-dit-Bille, S. M. Camargo, L. Emmenegger, T. Sasse, *et al.*, “Human intestine luminal ACE2 and amino acid transporter expression increased by ACE-inhibitors.,” *Amino acids*, vol. 47, no. 4, pp. 693–705, Apr. 2015, doi: 10.1007/s00726-014-1889-6.
 197. Severe Covid-19 GWAS Group, D. Ellinghaus, F. Degenhardt, L. Bujanda, *et al.*, “Genomewide Association Study of Severe Covid-19 with Respiratory Failure.,” *The New England journal of medicine*, vol. 383, no. 16, pp. 1522–1534, Oct. 2020, doi: 10.1056/NEJMoa2020283.
 198. COVID-19 Host Genetics Initiative, “Mapping the human genetic architecture of COVID-19.,” *Nature*, vol. 600, no. 7889, pp. 472–477, Dec. 2021, doi: 10.1038/s41586-021-03767-x.
 199. J. E. Horowitz, J. A. Kosmicki, A. Damask, D. Sharma, *et al.*, “Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease.,” *Nature genetics*, vol. 54, no. 4, pp. 382–392, Apr. 2022, doi: 10.1038/s41588-021-01006-7.
 200. A. Kousathanas, E. Pairo-Castineira, K. Rawlik, A. Stuckey, *et al.*, “Whole-genome sequencing reveals host factors underlying critical COVID-19.,” *Nature*, vol. 607, no. 7917, pp. 97–103, Jul. 2022, doi: 10.1038/s41586-022-04576-6.
 201. E. Pairo-Castineira, S. Clohisey, L. Klaric, A. D. Bretherick, *et al.*, “Genetic mechanisms of critical illness in COVID-19,” *Nature*, vol. 591, no. 7848, pp. 92–98, Mar. 2021, doi: 10.1038/s41586-020-03065-y.

202. H. Zaidi, M. Bader-El-Den, and J. McNicholas, “Using the National Early Warning Score (NEWS/NEWS 2) in different Intensive Care Units (ICUs) to predict the discharge location of patients.,” *BMC public health*, vol. 19, no. 1, p. 1231, Sep. 2019, doi: 10.1186/s12889-019-7541-3.
203. J. G. Jang, J. Hur, K. S. Hong, W. Lee, *et al.*, “Prognostic Accuracy of the SIRS, qSOFA, and NEWS for Early Detection of Clinical Deterioration in SARS-CoV-2 Infected Patients.,” *Journal of Korean medical science*, vol. 35, no. 25, p. e234, Jun. 2020, doi: 10.3346/jkms.2020.35.e234.
204. M. E. Kanof, P. D. Smith, and H. Zola, “Isolation of whole mononuclear cells from peripheral blood and cord blood.,” *Current protocols in immunology*, vol. Chapter 7, no. 1, pp. 7.1.1-7.1.7, May 2001, doi: 10.1002/0471142735.im0701s19.
205. Y. Liao, G. K. Smyth, and W. Shi, “featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.,” *Bioinformatics (Oxford, England)*, vol. 30, no. 7, pp. 923–30, Apr. 2014, doi: 10.1093/bioinformatics/btt656.
206. M. Schubert, S. Lindgreen, and L. Orlando, “AdapterRemoval v2: Rapid adapter trimming, identification, and read merging,” *BMC Research Notes*, 2016, doi: 10.1186/s13104-016-1900-2.
207. G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, *et al.*, “ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks,” *Bioinformatics*, 2009, doi: 10.1093/bioinformatics/btp101.
208. G. Bindea, J. Galon, and B. Mlecnik, “CluePedia Cytoscape plugin: Pathway insights using integrated experimental and in silico data,” *Bioinformatics*, 2013, doi: 10.1093/bioinformatics/btt019.
209. M. Martin, “Cutadapt removes adapter sequences from high-throughput sequencing reads,” *EMBnet.journal*, vol. 17, no. 1, p. 10, May 2011, doi: 10.14806/ej.17.1.200.
210. T. Coxe, D. J. Burks, U. Singh, R. Mittler, *et al.*, “Benchmarking RNA-Seq Aligners at Base-Level and Junction Base-Level Resolution Using the Arabidopsis thaliana Genome,” *Plants*, vol. 13, no. 5, p. 582, Feb. 2024, doi: 10.3390/plants13050582.

211. Y. Chen, A. T. L. Lun, and G. K. Smyth, “From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline.,” *F1000Research*, vol. 5, p. 1438, Aug. 2016, doi: 10.12688/f1000research.8987.2.
212. J. L. Goldstein and M. S. Brown, “The LDL receptor.,” *Arteriosclerosis, thrombosis, and vascular biology*, vol. 29, no. 4, pp. 431–8, Apr. 2009, doi: 10.1161/ATVBAHA.108.179564.
213. S. Andrei, M. Olivier, C. De Tymowski, A. Snauwaert, *et al.*, “Relationship between high-density lipoprotein cholesterol (HDL-C) concentration and ventilator-associated pneumonia in ICU COVID-19 patients.,” *Anaesthesia, critical care & pain medicine*, p. 101535, May 2025, doi: 10.1016/j.accpm.2025.101535.
214. F. Begue, B. Veeren, P. Rondeau, A.-M. Florence, *et al.*, “HDL proteome and apolipoproteins concentrations in severe ICU COVID-19 patients.,” *Lipids in health and disease*, vol. 24, no. 1, p. 32, Jan. 2025, doi: 10.1186/s12944-024-02381-w.
215. M. L. Apalama, F. Begue, S. Tanaka, M. Cournot, *et al.*, “High-density lipoproteins and COVID-19: preparing the next pandemic.,” *Journal of lipid research*, vol. 66, no. 5, p. 100779, May 2025, doi: 10.1016/j.jlr.2025.100779.
216. X. Wei, W. Zeng, J. Su, H. Wan, *et al.*, “Hypolipidemia is associated with the severity of COVID-19.,” *Journal of clinical lipidology*, vol. 14, no. 3, pp. 297–304, May 2020, doi: 10.1016/j.jacl.2020.04.008.
217. A. V Sorokin, S. K. Karathanasis, Z.-H. Yang, L. Freeman, *et al.*, “COVID-19-Associated dyslipidemia: Implications for mechanism of impaired resolution and novel therapeutic approaches.,” *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, vol. 34, no. 8, pp. 9843–9853, Aug. 2020, doi: 10.1096/fj.202001451.
218. J. Fan, H. Wang, G. Ye, X. Cao, *et al.*, “Letter to the Editor: Low-density lipoprotein is a potential predictor of poor prognosis in patients with coronavirus disease 2019.,” *Metabolism: clinical and experimental*, vol. 107, p. 154243, Jun. 2020, doi: 10.1016/j.metabol.2020.154243.

219. X. Hu, D. Chen, L. Wu, G. He, *et al.*, “Declined serum high density lipoprotein cholesterol is associated with the severity of COVID-19 infection,” *Clinica Chimica Acta*, vol. 510, pp. 105–110, Nov. 2020, doi: 10.1016/j.cca.2020.07.015.
220. D. W. Sanders, C. C. Jumper, P. J. Ackerman, D. Bracha, *et al.*, “SARS-CoV-2 requires cholesterol for viral entry and pathological syncytia formation.,” *eLife*, vol. 10, Apr. 2021, doi: 10.7554/eLife.65962.
221. S. Orłowski, J.-J. Mourad, A. Gallo, and E. Bruckert, “Coronaviruses, cholesterol and statins: Involvement and application for Covid-19.,” *Biochimie*, vol. 189, pp. 51–64, Oct. 2021, doi: 10.1016/j.biochi.2021.06.005.
222. F. W. Charlton, S. Hover, J. Fuller, R. Hewson, *et al.*, “Cellular cholesterol abundance regulates potassium accumulation within endosomes and is an important determinant in bunyavirus entry.,” *The Journal of biological chemistry*, vol. 294, no. 18, pp. 7335–7347, May 2019, doi: 10.1074/jbc.RA119.007618.
223. J. Petersen, M. J. Drake, E. A. Bruce, A. M. Riblett, *et al.*, “The major cellular sterol regulatory pathway is required for Andes virus infection.,” *PLoS pathogens*, vol. 10, no. 2, p. e1003911, Feb. 2014, doi: 10.1371/journal.ppat.1003911.
224. R. Wang, C. R. Simoneau, J. Kulsuptrakul, M. Bouhaddou, *et al.*, “Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses.,” *Cell*, vol. 184, no. 1, pp. 106–119.e14, Jan. 2021, doi: 10.1016/j.cell.2020.12.004.
225. W. Lee, J. H. Ahn, H. H. Park, H. N. Kim, *et al.*, “COVID-19-activated SREBP2 disturbs cholesterol biosynthesis and leads to cytokine storm.,” *Signal transduction and targeted therapy*, vol. 5, no. 1, p. 186, Sep. 2020, doi: 10.1038/s41392-020-00292-7.
226. H. I. Lim and K. A. Hajjar, “Annexin A2 in Fibrinolysis, Inflammation and Fibrosis.,” *International journal of molecular sciences*, vol. 22, no. 13, Jun. 2021, doi: 10.3390/ijms22136836.
227. R. H. Fairwozy, J. Cooper, J. White, C. Giambartolomei, *et al.*, “Identifying low density lipoprotein cholesterol associated variants in the Annexin A2 (ANXA2) gene.,” *Atherosclerosis*, vol. 261, pp. 60–68, Jun. 2017, doi:

- 10.1016/j.atherosclerosis.2017.04.010.
228. K. Ly, Y. G. L. Saavedra, M. Canuel, S. Routhier, *et al.*, “Annexin A2 reduces PCSK9 protein levels via a translational mechanism and interacts with the M1 and M2 domains of PCSK9.” *The Journal of biological chemistry*, vol. 289, no. 25, pp. 17732–46, Jun. 2014, doi: 10.1074/jbc.M113.541094.
229. J. R. Taylor, J. G. Skeate, and W. M. Kast, “Annexin A2 in Virus Infection.” *Frontiers in microbiology*, vol. 9, no. DEC, p. 2954, Dec. 2018, doi: 10.3389/fmicb.2018.02954.
230. J. Kzhyshkowska and L. Krusell, “Cross-talk between endocytic clearance and secretion in macrophages.” *Immunobiology*, vol. 214, no. 7, pp. 576–93, Jul. 2009, doi: 10.1016/j.imbio.2009.03.007.
231. J. Kzhyshkowska, “Multifunctional receptor stabilin-1 in homeostasis and disease.” *TheScientificWorldJournal*, vol. 10, pp. 2039–53, Oct. 2010, doi: 10.1100/tsw.2010.189.
232. A. Erol, “Role of oxidized LDL-induced ‘trained macrophages’ in the pathogenesis of COVID-19 and benefits of pioglitazone: A hypothesis.” *Diabetes & metabolic syndrome*, vol. 14, no. 4, pp. 713–714, Jul. 2020, doi: 10.1016/j.dsx.2020.05.007.
233. M. Febbraio, D. P. Hajjar, and R. L. Silverstein, “CD36: a class B scavenger receptor involved in angiogenesis, atherosclerosis, inflammation, and lipid metabolism.” *The Journal of clinical investigation*, vol. 108, no. 6, pp. 785–91, Sep. 2001, doi: 10.1172/JCI14006.
234. L. Park, J. Zhou, P. Zhou, R. Pistick, *et al.*, “Innate immunity receptor CD36 promotes cerebral amyloid angiopathy,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 8, pp. 3089–3094, Feb. 2013, doi: 10.1073/pnas.1300021110.
235. A. Cabrera, D. Neculai, and K. C. Kain, “CD36 and malaria: friends or foes? A decade of data provides some answers.” *Trends in parasitology*, vol. 30, no. 9, pp. 436–44, Sep. 2014, doi: 10.1016/j.pt.2014.07.006.
236. L. K. Erdman, G. Cosio, A. J. Helmers, D. C. Gowda, *et al.*, “CD36 and TLR

- interactions in inflammation and phagocytosis: implications for malaria.,” *Journal of immunology (Baltimore, Md. : 1950)*, vol. 183, no. 10, pp. 6452–9, Nov. 2009, doi: 10.4049/jimmunol.0901374.
237. T. K. Means, E. Mylonakis, E. Tampakakis, R. A. Colvin, *et al.*, “Evolutionarily conserved recognition and innate immunity to fungal pathogens by the scavenger receptors SCARF1 and CD36.,” *The Journal of experimental medicine*, vol. 206, no. 3, pp. 637–53, Mar. 2009, doi: 10.1084/jem.20082109.
238. J. F. C. Glatz and J. J. F. P. Luiken, “From fat to FAT (CD36/SR-B2): Understanding the regulation of cellular fatty acid uptake.,” *Biochimie*, vol. 136, pp. 21–26, May 2017, doi: 10.1016/j.biochi.2016.12.007.
239. J. F. C. Glatz and J. J. F. P. Luiken, “Dynamic role of the transmembrane glycoprotein CD36 (SR-B2) in cellular fatty acid uptake and utilization.,” *Journal of lipid research*, vol. 59, no. 7, pp. 1084–1093, Jul. 2018, doi: 10.1194/jlr.R082933.
240. Y. Li, P. Yang, L. Zhao, Y. Chen, *et al.*, “CD36 plays a negative role in the regulation of lipophagy in hepatocytes through an AMPK-dependent pathway.,” *Journal of lipid research*, vol. 60, no. 4, pp. 844–855, Apr. 2019, doi: 10.1194/jlr.M090969.
241. L. G. Gardinassi, C. O. S. Souza, H. Sales-Campos, and S. G. Fonseca, “Immune and Metabolic Signatures of COVID-19 Revealed by Transcriptomics Data Reuse.,” *Frontiers in immunology*, vol. 11, p. 1636, Jun. 2020, doi: 10.3389/fimmu.2020.01636.
242. S. Chevrier, Y. Zurbuchen, C. Cervia, S. Adamo, *et al.*, “A distinct innate immune signature marks progression from mild to severe COVID-19.,” *Cell reports. Medicine*, vol. 2, no. 1, p. 100166, Jan. 2021, doi: 10.1016/j.xcrm.2020.100166.
243. P. Tontonoz, L. Nagy, J. G. Alvarez, V. A. Thomazy, *et al.*, “PPARgamma promotes monocyte/macrophage differentiation and uptake of oxidized LDL.,” *Cell*, vol. 93, no. 2, pp. 241–52, Apr. 1998, doi: 10.1016/s0092-8674(00)81575-5.
244. H. Garg and P. Khanna, “Covid and cholesterol (C&C): Something to worry about or much ado about nothing?,” *Trends in anaesthesia & critical care*, vol. 36, pp. 39–40, Feb. 2021, doi: 10.1016/j.tacc.2020.09.003.

245. M. Hernandez-Quiles, M. F. Broekema, and E. Kalkhoven, “PPAR γ in Metabolism, Immunity, and Cancer: Unified and Diverse Mechanisms of Action.,” *Frontiers in endocrinology*, vol. 12, p. 624112, Feb. 2021, doi: 10.3389/fendo.2021.624112.
246. J. Huckriede, S. B. Anderberg, A. Morales, F. de Vries, *et al.*, “Evolution of NETosis markers and DAMPs have prognostic value in critically ill COVID-19 patients.,” *Scientific reports*, vol. 11, no. 1, p. 15701, Aug. 2021, doi: 10.1038/s41598-021-95209-x.
247. H. Patel, N. J. Ashton, R. J. B. Dobson, L.-M. Andersson, *et al.*, “Proteomic blood profiling in mild, severe and critical COVID-19 patients.,” *Scientific reports*, vol. 11, no. 1, p. 6357, Mar. 2021, doi: 10.1038/s41598-021-85877-0.
248. S.-J. Cho, M.-C. Kook, J. H. Lee, J.-Y. Shin, *et al.*, “Peroxisome proliferator-activated receptor γ upregulates galectin-9 and predicts prognosis in intestinal-type gastric cancer.,” *International journal of cancer*, vol. 136, no. 4, pp. 810–20, Feb. 2015, doi: 10.1002/ijc.29056.
249. S.-U.-S. Mattoo, S.-J. Kim, D.-G. Ahn, and J. Myoung, “Escape and Over-Activation of Innate Immune Responses by SARS-CoV-2: Two Faces of a Coin.,” *Viruses*, vol. 14, no. 3, Mar. 2022, doi: 10.3390/v14030530.
250. T. Sapir, Z. Averch, B. Lerman, A. Bodzin, *et al.*, “COVID-19 and the Immune Response: A Multi-Phasic Approach to the Treatment of COVID-19.,” *International journal of molecular sciences*, vol. 23, no. 15, Aug. 2022, doi: 10.3390/ijms23158606.
251. B. Witholt, H. V Heerikhuizen, and L. De Leij, “How does lysozyme penetrate through the bacterial outer membrane?,” *Biochimica et biophysica acta*, vol. 443, no. 3, pp. 534–44, Sep. 1976, doi: 10.1016/0005-2736(76)90471-5.
252. T. Ganz, M. E. Selsted, D. Szklarek, S. S. Harwig, *et al.*, “Defensins. Natural peptide antibiotics of human neutrophils.,” *The Journal of clinical investigation*, vol. 76, no. 4, pp. 1427–35, Oct. 1985, doi: 10.1172/JCI112120.
253. M. Tollin, P. Bergman, T. Svenberg, H. Jörnvall, *et al.*, “Antimicrobial peptides in

- the first line defence of human colon mucosa.,” *Peptides*, vol. 24, no. 4, pp. 523–30, Apr. 2003, doi: 10.1016/s0196-9781(03)00114-1.
254. K. R. Loeb and A. L. Haas, “The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins.,” *The Journal of biological chemistry*, vol. 267, no. 11, pp. 7806–13, Apr. 1992, [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/1373138>.
255. Y. J. Jeon, J. S. Choi, J. Y. Lee, K. R. Yu, *et al.*, “ISG15 modification of filamin B negatively regulates the type I interferon-induced JNK signalling pathway.,” *EMBO reports*, vol. 10, no. 4, pp. 374–80, Apr. 2009, doi: 10.1038/embor.2009.23.
256. Y.-C. Perng and D. J. Lenschow, “ISG15 in antiviral immunity and beyond.,” *Nature reviews. Microbiology*, vol. 16, no. 7, pp. 423–439, Jul. 2018, doi: 10.1038/s41579-018-0020-5.
257. A. Okumura, G. Lu, I. Pitha-Rowe, and P. M. Pitha, “Innate antiviral response targets HIV-1 release by the induction of ubiquitin-like protein ISG15.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 5, pp. 1440–5, Jan. 2006, doi: 10.1073/pnas.0510518103.
258. C. D. Swaim, A. F. Scott, L. A. Canadeo, and J. M. Huibregtse, “Extracellular ISG15 Signals Cytokine Secretion through the LFA-1 Integrin Receptor.,” *Molecular cell*, vol. 68, no. 3, pp. 581-590.e5, Nov. 2017, doi: 10.1016/j.molcel.2017.10.003.
259. Y. Tang, J. Sun, H. Pan, F. Yao, *et al.*, “Aberrant cytokine expression in COVID-19 patients: Associations between cytokines and disease severity.,” *Cytokine*, vol. 143, p. 155523, Jul. 2021, doi: 10.1016/j.cyto.2021.155523.
260. N. A. P. Lieberman, V. Peddu, H. Xie, L. Shrestha, *et al.*, “In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age.,” *PLoS biology*, vol. 18, no. 9, p. e3000849, Sep. 2020, doi: 10.1371/journal.pbio.3000849.
261. B. Vastrad, C. Vastrad, and A. Tengli, “Bioinformatics analyses of significant genes, related pathways, and candidate diagnostic biomarkers and molecular targets in SARS-CoV-2/COVID-19.,” *Gene reports*, vol. 21, p. 100956, Dec. 2020, doi: 10.1016/j.genrep.2020.100956.

262. L. Fang, T. Tang, and M. Hu, “Identification of Differentially Expressed Genes in COVID-19 and Integrated Bioinformatics Analysis of Signaling Pathways.,” *Genetics research*, vol. 2021, p. 2728757, 2021, doi: 10.1155/2021/2728757.
263. T. Islam, M. R. Rahman, B. Aydin, H. Beklen, *et al.*, “Integrative transcriptomics analysis of lung epithelial cells and identification of repurposable drug candidates for COVID-19.,” *European journal of pharmacology*, vol. 887, p. 173594, Nov. 2020, doi: 10.1016/j.ejphar.2020.173594.
264. T. Klemm, G. Ebert, D. J. Calleja, C. C. Allison, *et al.*, “Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2,” *Embo j*, vol. 39, no. 18, p. e106275, 2020, doi: 10.15252/embj.2020106275.
265. D. Shin, R. Mukherjee, D. Grewe, D. Bojkova, *et al.*, “Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity.,” *Nature*, vol. 587, no. 7835, pp. 657–662, Nov. 2020, doi: 10.1038/s41586-020-2601-5.
266. M. Zhang, J. Li, H. Yan, J. Huang, *et al.*, “ISGylation in Innate Antiviral Immunity and Pathogen Defense Responses: A Review.,” *Frontiers in cell and developmental biology*, vol. 9, p. 788410, 2021, doi: 10.3389/fcell.2021.788410.
267. G. Liu, J.-H. Lee, Z. M. Parker, D. Acharya, *et al.*, “ISG15-dependent activation of the sensor MDA5 is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity.,” *Nature microbiology*, vol. 6, no. 4, pp. 467–478, Apr. 2021, doi: 10.1038/s41564-021-00884-1.
268. A. C. Tecalco Cruz, “Free ISG15 and Protein ISGylation Emerging in SARS-CoV-2 Infection.,” *Current drug targets*, vol. 23, no. 7, pp. 686–691, 2022, doi: 10.2174/1389450123666220316094720.
269. A. Java, A. J. Apicelli, M. K. Liszewski, A. Coler-Reilly, *et al.*, “The complement system in COVID-19: friend and foe?,” *JCI insight*, vol. 5, no. 15, Aug. 2020, doi: 10.1172/jci.insight.140711.
270. A. G. Savitt, S. Manimala, T. White, M. Fandaros, *et al.*, “SARS-CoV-2 Exacerbates COVID-19 Pathology Through Activation of the Complement and Kinin Systems.,” *Frontiers in immunology*, vol. 12, p. 767347, 2021, doi:

- 10.3389/fimmu.2021.767347.
271. Y. M. Ali, M. Ferrari, N. J. Lynch, S. Yaseen, *et al.*, “Lectin Pathway Mediates Complement Activation by SARS-CoV-2 Proteins,” *Frontiers in immunology*, vol. 12, p. 714511, Jul. 2021, doi: 10.3389/fimmu.2021.714511.
272. P. Macor, P. Durigutto, A. Mangogna, R. Bussani, *et al.*, “Multi-organ complement deposition in COVID-19 patients.,” *medRxiv: the preprint server for health sciences*, Jan. 2021, doi: 10.1101/2021.01.07.21249116.
273. F. Defendi, C. Leroy, O. Epaulard, G. Clavarino, *et al.*, “Complement Alternative and Mannose-Binding Lectin Pathway Activation Is Associated With COVID-19 Mortality.,” *Frontiers in immunology*, vol. 12, p. 742446, 2021, doi: 10.3389/fimmu.2021.742446.
274. Y. Wu, X. Huang, J. Sun, T. Xie, *et al.*, “Clinical Characteristics and Immune Injury Mechanisms in 71 Patients with COVID-19.,” *mSphere*, vol. 5, no. 4, Jul. 2020, doi: 10.1128/mSphere.00362-20.
275. J. C. Holter, S. E. Pischke, E. de Boer, A. Lind, *et al.*, “Systemic complement activation is associated with respiratory failure in COVID-19 hospitalized patients.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 40, pp. 25018–25025, Oct. 2020, doi: 10.1073/pnas.2010540117.
276. G. Zilow, J. A. Sturm, U. Rother, and M. Kirschfink, “Complement activation and the prognostic value of C3a in patients at risk of adult respiratory distress syndrome.,” *Clinical and experimental immunology*, vol. 79, no. 2, pp. 151–7, Feb. 1990, doi: 10.1111/j.1365-2249.1990.tb05171.x.
277. D. Annane, N. Heming, L. Grimaldi-Bensouda, V. Frémeaux-Bacchi, *et al.*, “Eculizumab as an emergency treatment for adult patients with severe COVID-19 in the intensive care unit: A proof-of-concept study.,” *EClinicalMedicine*, vol. 28, p. 100590, Nov. 2020, doi: 10.1016/j.eclinm.2020.100590.
278. W. M. Zelek, J. Cole, M. J. Ponsford, R. A. Harrison, *et al.*, “Complement Inhibition with the C5 Blocker LFG316 in Severe COVID-19.,” *American journal of*

- respiratory and critical care medicine*, vol. 202, no. 9, pp. 1304–1308, Nov. 2020, doi: 10.1164/rccm.202007-2778LE.
279. B. Afzali, M. Noris, B. N. Lambrecht, and C. Kemper, “The state of complement in COVID-19.,” *Nature reviews. Immunology*, vol. 22, no. 2, pp. 77–84, Feb. 2022, doi: 10.1038/s41577-021-00665-1.
280. J. Q. He, C. Wiesmann, and M. van Lookeren Campagne, “A role of macrophage complement receptor CR1g in immune clearance and inflammation.,” *Molecular immunology*, vol. 45, no. 16, pp. 4041–7, Oct. 2008, doi: 10.1016/j.molimm.2008.07.011.
281. M. van Lookeren Campagne and A. Verschoor, “Pathogen clearance and immune adherence ‘revisited’: Immuno-regulatory roles for CR1g.,” *Seminars in immunology*, vol. 37, pp. 4–11, Jun. 2018, doi: 10.1016/j.smim.2018.02.007.
282. A. G. Small, M. Al-Baghdadi, A. Quach, C. Hii, *et al.*, “Complement receptor immunoglobulin: a control point in infection and immunity, inflammation and cancer.,” *Swiss medical weekly*, vol. 146, p. w14301, 2016, doi: 10.4414/smw.2016.14301.
283. Y. Liu, H. Liu, C. Li, C. Ma, *et al.*, “Proteome Profiling of Lung Tissues in Chronic Obstructive Pulmonary Disease (COPD): Platelet and Macrophage Dysfunction Contribute to the Pathogenesis of COPD.,” *International journal of chronic obstructive pulmonary disease*, vol. 15, pp. 973–980, 2020, doi: 10.2147/COPD.S246845.
284. C. Zhao, J. Mo, X. Zheng, Z. Wu, *et al.*, “Identification of an Alveolar Macrophage-Related Core Gene Set in Acute Respiratory Distress Syndrome.,” *Journal of inflammation research*, vol. 14, pp. 2353–2361, 2021, doi: 10.2147/JIR.S306136.
285. Y. Liao, S. Guo, Y. Chen, D. Cao, *et al.*, “VSIG4 expression on macrophages facilitates lung cancer development.,” *Laboratory investigation; a journal of technical methods and pathology*, vol. 94, no. 7, pp. 706–15, Jul. 2014, doi: 10.1038/labinvest.2014.73.
286. T. Guo, H. Ma, and Y. Zhou, “Bioinformatics analysis of microarray data to identify

- the candidate biomarkers of lung adenocarcinoma.,” *PeerJ*, vol. 7, p. e7313, 2019, doi: 10.7717/peerj.7313.
287. M. A. Hausburg, J. S. Williams, K. L. Banton, C. W. Mains, *et al.*, “C1 esterase inhibitor-mediated immunosuppression in COVID-19: Friend or foe?,” *Clinical immunology communications*, vol. 2, pp. 83–90, Dec. 2022, doi: 10.1016/j.clicom.2022.05.001.
288. A. E. Davis, P. Mejia, and F. Lu, “Biological activities of C1 inhibitor.,” *Molecular immunology*, vol. 45, no. 16, pp. 4057–63, Oct. 2008, doi: 10.1016/j.molimm.2008.06.028.
289. H. Jiang, E. Wagner, H. Zhang, and M. M. Frank, “Complement 1 inhibitor is a regulator of the alternative complement pathway.,” *The Journal of experimental medicine*, vol. 194, no. 11, pp. 1609–16, Dec. 2001, doi: 10.1084/jem.194.11.1609.
290. M. A. Hausburg, K. L. Banton, M. Roshon, and D. Bar-Or, “Clinically distinct COVID-19 cases share notably similar immune response progression: A follow-up analysis.,” *Heliyon*, vol. 7, no. 1, p. e05877, Jan. 2021, doi: 10.1016/j.heliyon.2020.e05877.
291. E. Roy-Vallejo, S. Fernández De Córdoba-Oñate, P. Delgado-Wicke, A. Triguero-Martínez, *et al.*, “Occurrence of SARS-CoV-2 viremia is associated with genetic variants of genes related to COVID-19 pathogenesis.,” *Frontiers in medicine*, vol. 10, p. 1215246, 2023, doi: 10.3389/fmed.2023.1215246.
292. I. Aksamentov, C. Roemer, E. Hodcroft, and R. Neher, “Nextclade: clade assignment, mutation calling and quality control for viral genomes,” *Journal of Open Source Software*, vol. 6, no. 67, p. 3773, Nov. 2021, doi: 10.21105/joss.03773.
293. C. J. Hemmer, M. Löbermann, and E. C. Reisinger, “[COVID-19: epidemiology and mutations : An update].,” *Der Radiologe*, vol. 61, no. 10, pp. 880–887, Oct. 2021, doi: 10.1007/s00117-021-00909-0.
294. K. Tao, P. L. Tzou, J. Nouhin, R. K. Gupta, *et al.*, “The biological and clinical significance of emerging SARS-CoV-2 variants.,” *Nature reviews. Genetics*, vol. 22, no. 12, pp. 757–773, Dec. 2021, doi: 10.1038/s41576-021-00408-x.

295. R. Arya, S. Kumari, B. Pandey, H. Mistry, *et al.*, “Structural insights into SARS-CoV-2 proteins.,” *Journal of molecular biology*, vol. 433, no. 2, p. 166725, Jan. 2021, doi: 10.1016/j.jmb.2020.11.024.
296. A. Kumar, P. Prasoon, C. Kumari, V. Pareek, *et al.*, “SARS-CoV-2-specific virulence factors in COVID-19.,” *Journal of medical virology*, vol. 93, no. 3, pp. 1343–1350, Mar. 2021, doi: 10.1002/jmv.26615.
297. X. Wei, W. Zeng, J. Su, H. Wan, *et al.*, “Hypolipidemia is associated with the severity of COVID-19,” *Journal of Clinical Lipidology*, vol. 14, no. 3, pp. 297–304, May 2020, doi: 10.1016/j.jacl.2020.04.008.
298. A. Zinellu, P. Paliogiannis, A. G. Fois, P. Solidoro, *et al.*, “Cholesterol and Triglyceride Concentrations, COVID-19 Severity, and Mortality: A Systematic Review and Meta-Analysis With Meta-Regression.,” *Frontiers in public health*, vol. 9, p. 705916, Aug. 2021, doi: 10.3389/fpubh.2021.705916.
299. A. Rubio-Casillas, E. M. Redwan, and V. N. Uversky, “SARS-CoV-2: A Master of Immune Evasion.,” *Biomedicines*, vol. 10, no. 6, Jun. 2022, doi: 10.3390/biomedicines10061339.