

Автономная некоммерческая образовательная организация высшего  
образования  
«Сколковский институт науки и технологий»

*на правах рукописи*

Ткачев Анна Игоревна

**МЕТОДОЛОГИЯ АНАЛИЗА ДАННЫХ ПАНОРАМНЫХ  
ИЗМЕРЕНИЙ ЛИПИДНОГО СОСТАВА С ИСПОЛЬЗОВАНИЕМ  
ВЭЖХ-МС И ЕЕ ПРИМЕНЕНИЕ В ИССЛЕДОВАНИИ  
ПСИХИЧЕСКИХ РАССТРОЙСТВ**

**РЕЗЮМЕ**

диссертации на соискание учёной степени  
кандидата компьютерных наук

Москва - 2022

Диссертационная работа была выполнена в Сколковском институте науки и технологий.

Научный руководитель: Филипп Ефимович Хайтович,  
PhD, Сколковский институт науки и технологий

# Введение

## Значимость работы

Прогресс в биологических науках неразрывно связан с технологическими достижениями. До развития методов масс-спектрометрии, провести обширный количественный анализ молекулярного состава являлось затруднительным. Среди прочего, совершенствование этих методов позволило вынести анализ липидов — основных строительных элементов всех живых клеток и ключевых участников энергетического метаболизма — на один план с другими молекулярными измерениями. В свете недостаточности генетических данных для всестороннего понимания некоторых заболеваний, их изучение с новой точки зрения, такой, как липидный фенотип, представляется перспективным.

Психические заболевания — это тяжелые состояния, поражающие существенную долю населения. Эти расстройства являются примерами нарушений, для которых еще не обнаружена удовлетворительная генетическая модель, несмотря на их известную наследуемость. Действительно, хотя общепризнано, что психические расстройства неразрывно связаны с их физическими проявлениями в мозге и теле, природа этих заболеваний до сих пор неясна. Неизвестны причины развития психических заболеваний, как и неизвестны конкретные механизмы, посредством которых лекарства содействуют улучшению симптоматики. Принято разделять отдельные психические расстройства, такие как шизофрения, депрессия и биполярное расстройство, но четкие границы между заболеваниями, также, как и однородность внутри расстройств, не являются однозначными. Поскольку психические заболевания в настоящее время характеризуются на основе определенных поведенческих симптомов, для прогресса в понимании психических расстройств требуются более объективные маркеры заболевания.

Липидные профили как в мозге, так и в крови, актуальны в контексте этих расстройств. Растет понимание значимости функций, выполняемых липидами в головном мозге, для которого доля липидов среди прочих соединений особенно велика. В то же время психические заболевания — это не просто расстройства головного мозга, они неразрывно связаны с физическим здоровьем, о чем свидетельствует высокая распространенность сопутствующих сердечно-сосудистых заболеваний среди психиатрических больных. Плазма крови содержит важные индикаторы метаболического здоровья, среди которых липиды крови, вероятно, играют важную роль. Из-за относительной новизны области и связанных технологических сложностей, исследования, посвященные липидам в мозге и крови пациентов с психическими расстройствами, по-прежнему малочисленны. В этой

диссертации представлены результаты анализа данных для липидных профилей крови в контексте психических заболеваний и отклика на лечение, а также результаты анализа данных для липидного состава мозга макак вследствие лечения в контексте терапевтических и побочных эффектов антидепрессантов на головной мозг человека.

Липидомика – развивающаяся область, стремительный рост которой только начинается. Среди различных аналитических подходов, именно панорамный, или нецелевой, анализ липидов связан с более высокой сложностью данных, поскольку он направлен на количественное определение липидных соединений в биологических системах максимально обширно, насколько это практически достижимо. Нецелевой характер анализа является одновременно преимуществом и источником потенциальных трудностей. Хотя такой анализ потенциально чрезвычайно информативен, вычислительные методы и стандарты обработки данных отстают от технологических достижений. Поскольку соответствующие методы анализа данных еще недостаточно хорошо отработаны, их совершенствование представляют собой отдельную актуальную тему для исследования. Значительная часть этой диссертации связана с обработкой данных, специфичной для нецелевой липидомики. Развитие методов липидомики, включая инструменты обработки данных, способствует расширению сферы ее применения за пределами своей узкой области, что может привести к лучшему пониманию роли липидов как в психических заболеваниях, так и в здоровых и патологических процессах в целом.

## Цели и задачи диссертации

Диссертация посвящена разработке и применению методов анализа данных нетаргетированной липидомики к ряду конкретных исследований, связанных с изучением липидного профиля при психических расстройствах. Конкретные задачи исследования, затронутыми в этой работе, состояли в следующем:

- Методологические аспекты разработки и оптимизации протоколов анализа нецелевых липидных данных, полученных методами высокоэффективной жидкостной хроматографией с масс-спектрометрией (ВЭЖХ-МС).
- Исследование липидного профиля плазмы крови у индивидуумов с диагнозами шизофрения, депрессия и биполярное расстройство.
- Связь между индивидуальными изменениями профилей липидов в крови и тяжестью симптомов у лиц с шизофренией до и после стационарного лечения.
- Вызванные лечением долгосрочные изменения в молекулярном составе мозга молодых макак, включая изменения содержания липидов, в контексте терапевтических и побочных эффектов антидепрессантов у детей и молодых людей.

## Резюме основных результатов

В результате исследований, включенных в эту диссертацию, выдвинуты следующие предложения:

- Описаны основные этапы обработки в протоколах анализа данных нецелевой липидомики и идентифицированы существующие пробелы. Вычислительные подходы, разработанные для этой диссертации, были неотъемлемой частью результатов липидомных исследований, включенных в эту диссертацию.
- Дублирование сигнала в экспериментах по нецелевой липидомике остается одним из упускаемых из виду шагов в большинстве предлагаемых методологий анализа. Основанный на данных подход, предложенный в этой диссертации для сокращения таких дубликаций, удалил в два раза больше избыточных признаков по сравнению со стандартным подходом фильтрации аддуктов.
- У пациентов с шизофренией наблюдаются робастные изменения уровней отдельных липидных соединений по сравнению со здоровыми индивидуумами.
- Эти изменения являются общими для других психических расстройств, такими как биполярное и депрессивное расстройства.
- У пациентов с шизофренией липидный профиль высоко специфичен относительно индивидуумов без психических расстройств, о чем свидетельствует высокая предсказательная сила предложенной классификационной модели.
- Плохой ответ на лечение у больных шизофренией ассоциирован с изменением уровня определенных липидных соединений в крови.
- В результате лечения флуоксетином, липиды в мозге молодых макаков изменяются. Характерный липидный профиль связан со снижением полиненасыщенных жирных кислот (ПНЖК).

## Научная новизна

- Предложен новый метод для учета дублирования сигналов и образования неожиданных аддуктов в нетаргетированной липидомике.
- Проведенный анализ липидных профилей в крови пациентов с психическими расстройствами был самым полным на момент исследования с точки зрения размеров выборок, липидного покрытия, а также наличия сравнительного анализа для нескольких расстройств.
- Маркеры ответа на лечение при шизофрении немногочисленны, а описанный липидный профиль триглицеридов ранее не сообщался в контексте ответа на лечение и шизофрении.

- Изучение долгосрочного влияния флуоксетина на липидный профиль мозга молодых макак было первым исследованием долгосрочного влияния флуоксетина на молекулярный состав мозга у приматов.

## Практическая ценность

Практическая ценность полученных результатов определяется необходимостью более объективных показателей болезни для развития исследований в области психического здоровья.

Для предложенной диагностической модели шизофрении на основе липидов, валидированной на нескольких тестовых выборках, была продемонстрирована потенциальная применимость на практике. Несмотря на то, что модель для диагностики шизофрении, сама по себе, имеет ограниченное практическое применение, вероятно, она может быть расширена до пограничных состояний, таких как менее тяжелые психические заболевания или предсказание развития болезни в будущем.

Изучение реакции на психофармакологию является весьма актуальной проблемой в связи с непредсказуемостью исходов лечения. Путем исследования связи между изменениями индивидуальных профилей липидов и улучшением симптомов после лечения был описан конкретный профиль липидов, связанный с плохим ответом на лечение. Дальнейшее изучение маркеров ответа на лечение потенциально может улучшить не только понимание эффектов лекарств, но и позволить в будущем более персонализированный выбор терапии.

Анализ влияния антидепрессанта флуоксетина на мозг молодых макак был связан со снижением ПНЖК. Считается, что эти липидные компоненты важны как для психического здоровья, так и для правильного развития мозга у детей. Хотя результаты, основанные на модельных организмах, не могут быть непосредственно распространены на людей, они указывают на важность изучения липидов и ПНЖК в контексте депрессии.

Наконец, анализ данных нетаргетированной липидомики сам по себе является сложной задачей из-за ограниченности соответствующей литературы. Это можно объяснить новизной области и ее междисциплинарным характером на стыке аналитической химии и анализа данных. Глава по обработке данных нетаргетированной липидомики, представленная в диссертации, может послужить полезным практическим руководством, а также указателем для дальнейших исследований и разработок.

## Личный вклад

Весь статистический анализ, основанный на липидных данных, был выполнен соискателем. Вся обработка липидных данных, за исключением нескольких конкретных шагов программного обеспечения, была выполнена соискателем.

## Публикации по теме диссертации

1. Tkachev A., Stekolshchikova E., Anikanov N., et al. Shorter Chain Triglycerides Are Negatively Associated with Symptom Improvement in Schizophrenia. *Biomolecules*. 2021;11(5):720. Published 2021 May 11. doi:10.3390/biom11050720
2. Tkachev A., Stekolshchikova E., Bobrovskiy D.M., et al. Long-Term Fluoxetine Administration Causes Substantial Lipidome Alteration of the Juvenile Macaque Brain. *Int J Mol Sci*. 2021;22(15):8089. Published 2021 Jul 28. doi:10.3390/ijms22158089

## Другие публикации автора

Публикации автора, не включенные в данную диссертацию:

3. Tkachev A, Stepanova V, Zhang L, et al. Differences in lipidome and metabolome organization of prefrontal cortex among human populations. *Sci Rep*. 2019;9(1):18348. Published 2019 Dec 4. doi:10.1038/s41598-019-53762-6
4. Kurochkin I, Khrameeva E, Tkachev A, et al. Metabolome signature of autism in the human prefrontal cortex. *Commun Biol*. 2019;2:234. Published 2019 Jun 21. doi:10.1038/s42003-019-0485-4
5. Khrameeva E, Kurochkin I, Han D, et al. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res*. 2020;30(5):776-789. doi:10.1101/gr.256958.119
6. Stepanova V, Moczulska KE, Vacano GN, et al. Reduced purine biosynthesis in humans after their divergence from Neandertals. *Elife*. 2021;10:e58741. Published 2021 May 4. doi:10.7554/eLife.58741

## Структура диссертации

Первая **глава** диссертации знакомит с темой исследования и покрывает цели и задачи диссертационной работы.

**Глава 2** содержит вводную информацию и разделена на три основных раздела. Раздел 2.1 содержит общее введение о психических расстройствах. В нем представлен беглый обзор для трех психических расстройств: шизофрении, биполярного расстройства и большого депрессивного расстройства. В нем обсуждаются современные подходы к диагностике, доступные маркеры заболевания, сходства и различия между расстройствами, а также сопутствующие соматические заболевания. Также представлен краткий обзор психофармакологии. Раздел 2.2 знакомит с темой липидов и липидомики. Дается краткое, но достаточно полное введение в соответствующую терминологию. Обсуждаются результаты исследований липидных профилей в различных тканях млекопитающих, а также дается обзор метаболизма липидов, позволяющий читателям, не знакомым с этой темой, лучше понять значимость результатов представленных липидомных исследований.

Представлен обзор исследований в области липидомики связанных с психическими расстройствами, для плазмы крови и головного мозга, в частности. В разделе 2.3 представлена общая информация об экспериментальных методах липидомики, в частности, о высокоэффективной жидкостной хроматографии и масс-спектрометрии (ВЭЖХ-МС). Он включает обзор доступной литературы, касающейся обработки данных нецелевой липидомики на основе ВЭЖХ-МС.

**Главы 3, 4 и 5** содержат основные результаты исследований профилей липидов при психических расстройствах. Каждая глава содержит также обсуждение представленных результатов и раздел методов, описывающий методологию эксперимента и статистический анализ.

**Глава 6** посвящена обработке данных нецелевой липидомики на основе ВЭЖХ-МС. В разделе 6.1 описаны основные этапы протоколов обработки данных нецелевой липидомики и идентифицированы существующие пробелы. В частности, обсуждаются удаление признаков со слабым биологическим сигналом, высокой технической вариабельностью, коррекция влияния экспериментальных факторов, а также сопоставление разных наборов данных для сравнения результатов нескольких отдельных экспериментов. В разделе 6.2 предлагается основанный на данных подход для решения задачи, которая часто упускается в существующих методологиях анализа данных нецелевой липидомики – задача фильтрации дублированных сигналов и неожиданных аддуктов в основанной на ВЭЖХ-МС нецелевой липидомике.

**Глава 7** содержит выводы.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Межкогортный анализ липидных профилей у лиц с тяжелыми психическими расстройствами.

### Аннотация

Есть отрывочные свидетельства о том, что распространенные психические расстройства – шизофрения (SCZ), биполярное расстройство (BPD) и большое депрессивное расстройство (MDD) – существенно изменяют как липидный состав головного мозга, так и липидный состав плазмы крови. В данном исследовании представлен обширный анализ изменений липидов плазмы крови в этих трех заболеваниях, включающий измерения 1361 липидных признаков в 1354 образцах плазмы крови у индивидуумов из трех когорт, собранных в разных точках мира: в Китае, Западной Европе (Германия и Австрия) и России. Обнаружен и описан липидный профиль плазмы крови из 77 липидов, количественно отличающий больных шизофренией от контрольных лиц и являющийся общим для трех когорт. Этот липидный профиль также присутствовал у пациентов с двумя другими психическими

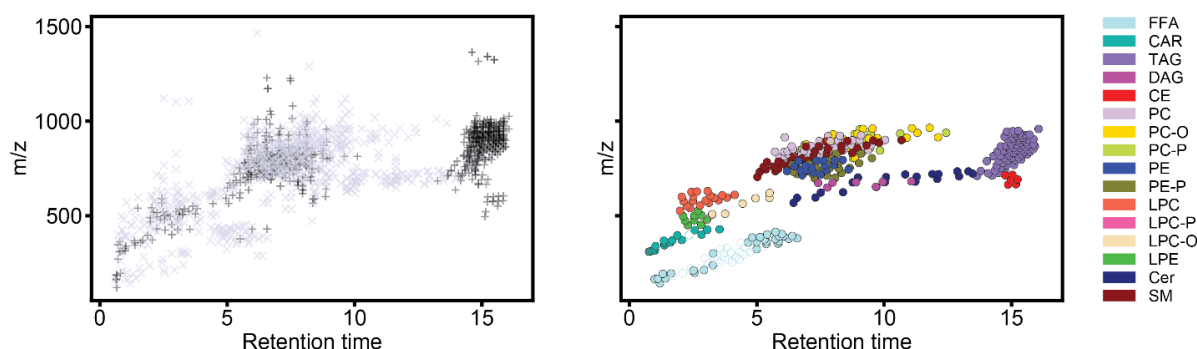


расстройствами, что согласуется с данными о сходстве этих трех расстройств по симптоматике и генетическим данным. Кроме того, на основе содержаний липидов была построена классификационная модель, отделяющая пациентов с SCZ от контрольной группы (CTL) с высокой диагностической способностью (площадь под кривой рабочей характеристики приёмника ROC AUC = 0.86-0.95), которая была дополнительно валидирована на двух различных наборах данных.

### Постановка исследования

Чтобы оценить, изменяется ли надежным и воспроизводимым образом состав липидома крови при распространенных психических расстройствах, шизофрении (SCZ), большом депрессивном расстройстве (MDD) и биполярном расстройстве (BPD), было проведено исследование с участием 876 пациентов с психическими расстройствами и 478 здоровых индивидуумов (CTL), соответствующие трем когортам, собранных в разных точках мира: в Чунцин, в Китае (CN), в нескольких городах в Германии и Австрии (DE-AT) и в Москве, в России (RU) (CN:  $n = 170, 222, 153$  для SCZ, MDD и CTL соответственно; DE-AT:  $n = 184, 148, 187$  для SCZ, BPD и CTL соответственно; RU:  $n = 82, 36, 34, 138$  для SCZ, BPD, MDD и CTL соответственно). Кроме того, была собрана дополнительная группа образцов, состоящая из 104 пациентов с первым эпизодом психоза (FEP) (набор данных *ferRU*). Всего было оценено содержания липидов в 1552 образцах плазмы.

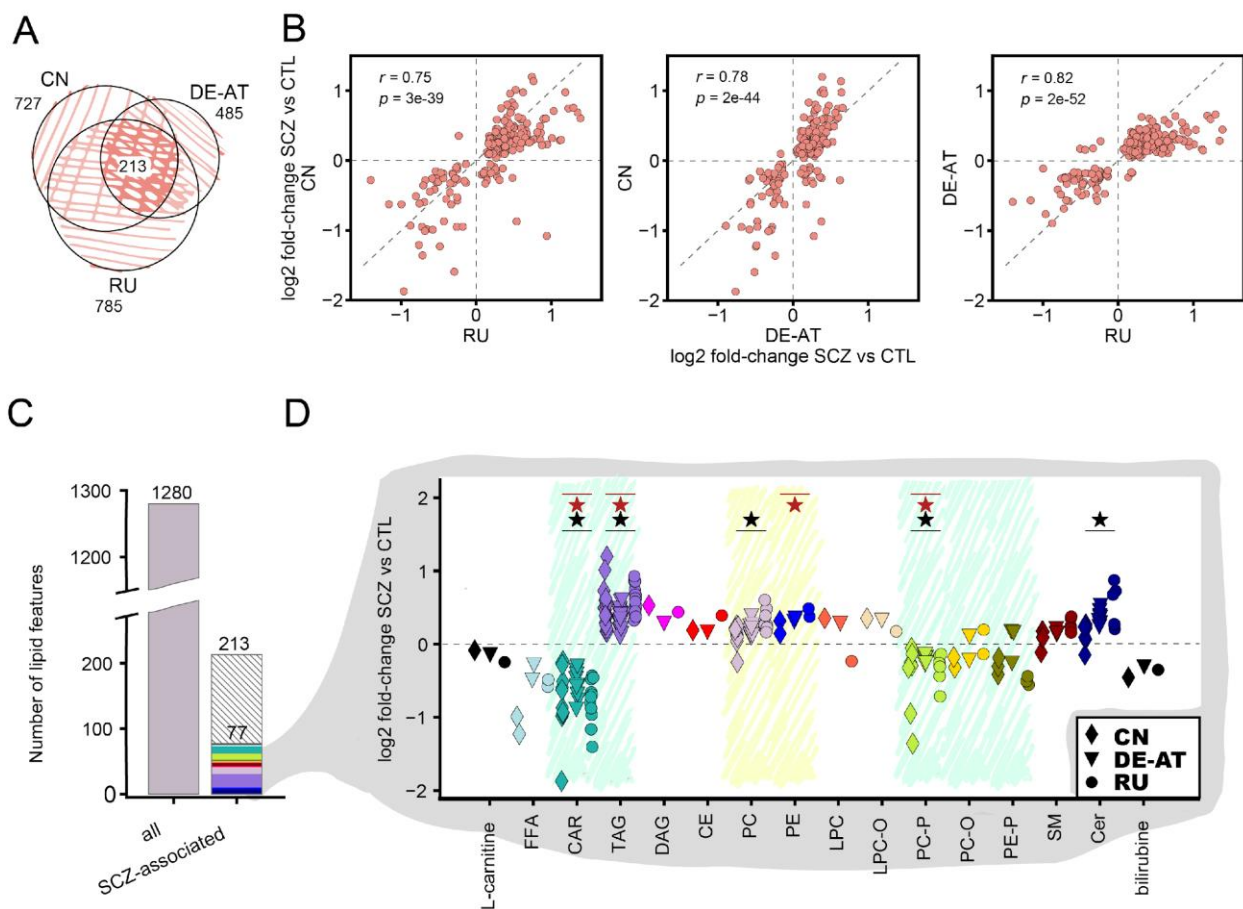
Чтобы получить систематический охват липидома плазмы крови, образцы были анализированы нецелевым методом используя высокоэффективную жидкостную хроматографию с масс-спектрометрией (ВЭЖХ-МС) в отрицательном и положительном режимах ионизации. Этот подход позволил оценить содержания для 1361 воспроизводимых липидных признаков. Среди них было автоматически аннотировано 395 липидных соединений на основе их времени удерживания и значений массы-на-заряд, включая, для подмножества этих соединений, подтверждение аннотации на уровне их спектров фрагментации. Аннотированные липиды охватывают 16 классов липидов, что хорошо согласуется с ранее описанными компонентами липидома плазмы крови (рис. 1).



**Рисунок 1:** Детектированные признаки и аннотированные липидные соединения. Слева: время удерживания и значения массы-на-заряд для всех 1361 детектированных липидных признаков в положительном (черные, «плюс») и отрицательном (серые, «х» маркеры) режимах ионизации. Справа: то же самое для автоматически аннотированных липидов. Пустые кружки обозначают липиды, удаленные из последующего анализа в связи с их ассоциацией с периодом голодания перед сдачей крови. Цвета соответствуют липидным классам, обозначенными справа.

### Изменения липидома крови, ассоциированных с шизофренией

Сравнение содержаний липидов между 436 образцами SCZ и 478 образцами CTL после исключения эффектов, потенциально связанных с искажающими факторами, такими как возраст, пол, период голодания перед сдачей крови, индекс массы тела (ИМТ), выявило значимые различия для 38-61% признаков в каждой из трех когорт (критерий Вилкоксона, поправка Бенджамини-Хохберг на множественное тестирование, ожидаемая доля ложно-положительных FDR = 10%; перестановочное P-значение  $p < 0.001$ ). Среди них, 213 признаков показали значимые различия независимо во всех трех когортах — CN, DE-AT и RU (SCZ-ассоциированные признаки; перестановочное P-значение  $p < 0.00001$ ; рис. 2A). Кроме того, направление изменений содержаний липидов между SCZ и CTL показали значимые корреляции между всеми парами когорт (корреляция Пирсона,  $r \geq 0.75$ ,  $p \leq 0.00001$ ; рис. 2B).



**Рисунок 2:** SCZ-ассоциированные липиды. (A) Количество липидных признаков, показывающих статистические различия в каждой из когорт - CN, DE-AT и RU - и их пересечение (SCZ-ассоциированные признаки). (B) Парное сравнение различий в содержаниях липидов между шизофренией (SCZ) и контрольными (CTL) образцами для когорт. Различия в интенсивностях отображены в виде кратного изменения в логарифмической шкале по основанию 2 (определенная как разница между средними значениями интенсивностей в логарифмической шкале по основанию 2). Точки представляют 213 липидных признаков, ассоциированных с SCZ. Коэффициенты корреляции Пирсона и соответствующие P-значения отмечены в верхнем левом углу. Горизонтальная и вертикальная пунктирные линии указывают на прямую  $y = 0$  и  $x = 0$ , диагональная пунктирная линия указывает на прямую  $y = x$ . (C) Количество всех проанализированных липидных признаков, после исключения признаков, связанных с периодом голодания перед сдачей крови (слева), количество SCZ-ассоциированных признаков (справа, заштриховано), и количество аннотированных липидных соединений, ассоциированных с SCZ (справа, цветом). Цвета соответствуют классам соединений, отображенными на панели D. (D) Различия в интенсивностях липидов между SCZ и CTL-индивидуумами для 77 соединений, ассоциированных с SCZ, сгруппированные в соответствии с их липидным классом и представленные для каждой когорты (отмечены формой метки). Маркеры в виде звезд сверху обозначают значимость различия на уровне липидных классов, соответствующий анализ описан в подразделе 3.1.2 диссертации.

213 липидных признаков, ассоциированных с SCZ, включали 77 уникальных липидов, подтвержденных спектрами фрагментации, и охватывали 14

больших классов липидов (рис. 2С; таблица 1). Анализ этих 77 липидов выявил системный эффект шизофрении, приводящий к скоординированному сдвигу липидов внутри класса в сторону меньшей или большей интенсивности во всех трех когортах (рис. 2С).

Lipid class	Lipid species	Molecular subspecies, when available
TAG	TAG(47:1) TAG(47:2) TAG(48:1) TAG(48:3) TAG(49:0) TAG(49:1) TAG(49:2) TAG(49:3) TAG(50:1) TAG(50:2) TAG(50:3) TAG(51:1) TAG(51:2) TAG(51:3) TAG(52:1) TAG(52:2) TAG(54:2) TAG(54:3) TAG(55:6) TAG(56:5) TAG(56:6)	
CAR	CAR(10:0) CAR(10:1) CAR(11:1) CAR(12:0) CAR(12:2) CAR(13:0) CAR(13:1) CAR(14:0) CAR(14:2) CAR(16:2) CAR(18:2)	
FFA	FFA(12:2) FFA(13:1)	
CE	CE(22:5)	
DAG	DAG(36:2)	DAG(18:1_18:1) DAG(18:0_18:2)
PC	PC(33:1) PC(34:1) PC(36:1) PC(37:3) PC(38:3) PC(38:4) PC(38:5) PC(40:4) PC(40:5) PC(42:5)	PC(18:1_15:0) PC(16:1_17:0) PC(16:0_17:1) PC(16:0_18:1) PC(18:0_18:1) PC(17:0_20:3), PC(19:1_18:2) PC(18:0_20:3) PC(16:0_22:4) PC(18:0_20:4), PC(18:1_20:3) PC(16:0_22:5) PC(18:0_22:4) PC(20:0_20:4) PC(18:0_22:5) PC(18:0_24:5)
PE	PE(36:4) PE(40:5)	PE(16:0_20:4) PE(18:0_22:5)
PC-O, PC-P	PC(O-34:0) PC(O-38:6) PC(P-34:2) PC(P-34:3) PC(P-35:3) PC(P-38:4) PC(P-38:5) PC(P-38:6) PC(P-40:7)	PC(O-18:0/16:0) PC(O-16:0/22:6) PC(P-16:0/18:2) PC(P-16:1/18:2) PC(P-35:3)[unresolved fatty acid composition] PC(P-18:0/20:4) PC(P-16:0/22:5) PC(P-16:0/22:6) PC(P-18:1/22:6)
PE-P	PE(P-36:4) PE(P-38:5) PE(P-38:6)	PE(P-16:0/20:4) PE(P-18:1/20:4) PE(P-16:0/22:6)
LPC	LPC(22:4)	
LPC-O	LPC(O-26:1)	
Cer	Cer(d34:1) Cer(d34:2) Cer(d36:1) Cer(d36:2) Cer(d40:2) Cer(d40:3) Cer(d42:2) Cer(d42:3)	Cer(d18:1/16:0) Cer(d18:2/16:0) Cer(d18:1/18:0) Cer(d18:2/18:0) Cer(d18:2/22:0) Cer(d40:3)[unresolved fatty acid composition] Cer(d18:1/24:1) Cer(d18:2/24:1)
SM	SM(d35:2) SM(d36:1) SM(d36:2) SM(d37:2) SM(d38:2)	
other	bilirubine, L-carnitine	

**Таблица 1:** 77 липидов, ассоциированных с SCZ, включая их жирнокислотный состав.

Диагностическая модель и оценка воспроизводимости между когортами  
Хотя большинство различий в содержаниях липидов между SCZ и CTL имели умеренную амплитуду, была построена диагностическая модель, способная отделить пациентов с шизофренией от контроля. В частности, модель логистической регрессии, обученная на трех объединенных когортах и 395 автоматически аннотированных липидах, отделила SCZ от CTL с хорошей диагностической способностью (площадь под кривой рабочей характеристики приемника (ROC AUC) =  $0.95 \pm 0.02$  для когорты RU,  $0.88 \pm 0.03$  для когорты DE-AT и  $0.86 \pm 0.03$  для когорты CN). Кроме того, модель так же хорошо показала себя при классификации образцов FEP и CTL из набора данных ferRU (ROC AUC =  $0.89 \pm 0.025$ ). Подраздел 3.1.3 диссертации также включает анализ воспроизводимости липидных изменений в SCZ между когортами на уровне индивидуальных липидов, а также на уровне диагностической модели.

Изменения липидов при SCZ коррелируют с другими психическими расстройствами

Далее были изучены изменения содержания липидов для расстройств BPD и MDD, каждое из которых было представлено двумя независимыми когортами (CN:  $n = 222$  для MDD; DE-AT:  $n = 148$  для BPD; RU:  $n = 36, 34$  для BPD и MDD, соответственно). Было определено 97 липидных признаков, демонстрирующих значимые различия в содержаниях в обеих когортах для MDD (липиды, ассоциированные с MDD; критерий Вилкоксона, поправка Бенджамини-Хохберг на множественное тестирование, ожидаемая доля ложноположительных FDR = 10%), и аналогично 47 липидов для BPD (липиды, ассоциированные с BPD; критерий Вилкоксона, поправка Бенджамини-Хохберг на множественное тестирование, ожидаемая доля ложноположительных FDR = 10% в каждой из двух когорт).

Сравнение ассоциированных с различными расстройствами липидов в разделе 3.1.4 показало, что, хотя разные липиды оказались статистически значимыми для расстройств, изменения липидных интенсивностей по сравнению с контрольной группой были сильно скоррелированы для трех расстройств (корреляция Пирсона,  $r = 0.94$  и  $r = 0.88$ ,  $p < 0.00001$  для сравнения MDD-SCZ в когортах CN и RU соответственно; корреляция Пирсона,  $r = 0.66$  и  $r = 0.93$ ,  $p < 0.00001$  для сравнения BPD-SCZ в когортах DE-AT и RU соответственно), что указывало на общий профиль липидов для расстройств SCZ, BPD и MDD. Анализ воспроизводимости эффекта SCZ между когортами (описанный в подразделе 3.1.3 диссертации) показал, что обнаружение липидных изменений, специфичных только для конкретного заболевания, является затруднительным. Сравнение с доступной литературой также подтвердило общие изменения для трех расстройств. Подраздел 3.1.5 диссертации включал анализ взаимосвязи между генами, ассоциированными с заболеванием, и

генами, связанными с метаболизмом липидов, на основе опубликованных данных. Результаты показали, что гены, связанные с вариацией липидома плазмы крови, сильно перекрываются с генетическими маркерами, общими для трех расстройств.

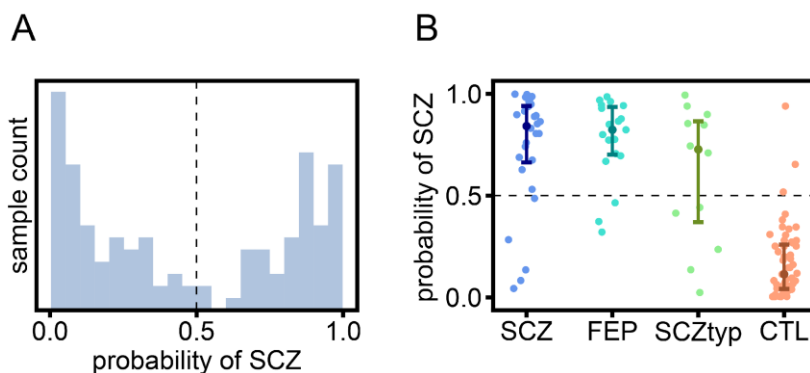
Взаимосвязь между изменениями липидома в SCZ и приемом лекарств.

Большинство пациентов с SCZ, принимавших участие в нашем исследовании, получали длительную терапию антипсихотическими препаратами, которые, как было показано, могут оказывать влияние на отдельные липиды плазмы крови. Чтобы оценить степень влияния лекарств на представленные результаты, в разделе 3.1.6 был включен дополнительный анализ, посвященный пациентам с первым эпизодом психоза, в также индивидуумам с SCZ, которые не получали антипсихотические препараты в течение как минимум шести месяцев до забора крови.

Валидация предсказательной модели

В приведенных выше результатах была определена диагностическая модель на основе содержаний липидов, способная отделить SCZ от CTL с высокой диагностической способностью (площадь под кривой рабочей характеристики приёмника ROC AUC = 0.86-0.95). Хотя результаты модели воспроизводились для различных наборов данных, панель биомаркеров на основе липидов еще далека от применений. В качестве первого шага к решению этого вопроса было проведено следующее исследование. В общей сложности, 119 образцов плазмы крови индивидуумов с психическими расстройствами и без были проанализированы используя те же методы липидного анализа, что и в исходном исследовании. Примечательно, что метки образцов были скрыты на протяжении всего анализа данных, чтобы исключить оптимизацию гиперпараметров модели и имитировать реальную обобщающую способность модели.

В исходном исследовании, диагностическая модель имела чувствительность 0.77 и специфичность 0.86. Соответственно, используя модель, обученную на содержании липидов в 914 образцах SCZ и CTL из исходного исследования, был предсказан статус заболевания в 119 тестовых образцах. Для определения меток предсказания SCZ и CTL, было установлено пороговое значение 0.5 для предсказанных моделью вероятностей заболевания (рис. 3А).



**Рисунок 3:** Вероятность предсказания SCZ для тестовых образцов. (А) Распределение вероятностей предсказания SCZ для 119 образцов. Черная пунктирная линия соответствует установленному порогу для предсказания меток. (В) Распределение вероятностей предсказания заболевания отдельно для образцов SCZ, FEP, SCZ<sub>тип</sub> и CTL. Четыре более темные точки соответствуют медианным значениям, а усы распространяются на межквартильные диапазоны.

Из 119 проверочных образцов, 28 соответствовали лицам с SCZ, 19 — лицам с FEP, 55 — здоровым CTL и 12 — лицам с шизотипическим расстройством личности (SCZ<sub>тип</sub>), являющимся расстройством схожим с SCZ, но без ярко выраженных психических симптомов. Для пяти образцов метки болезней восстановить не удалось. Доля правильно идентифицированных пациентов с SCZ и FEP (чувствительность) составила 0.82 и 0.84 соответственно, в то время как специфичность оказалась равной 0.95, что выше показателей, ожидаемых от исходного исследования (рис. 3В). Интересно, что индивидуумы с шизотипическим расстройством личности показали высокую вариабельность в вероятностных предсказаниях, и их распределение отличалось от распределения как для SCZ/FEP, так и для CTL (доля индивидуумов с SCZ<sub>тип</sub>, предсказанных как SCZ: 0.58; рис. 3В). Эти результаты соответствуют классификации данного расстройства как промежуточного фенотипа шизофренического спектра.

Триглицериды с более короткой цепью отрицательно связаны с улучшением симптомов шизофрении.

Данная глава следует результатам, опубликованным в [1]. Некоторые отрывки были дословно процитированы из [1], а рисунки воспроизведены или изменены с разрешением из [1].

#### Аннотация

Шизофрения — серьезное психическое расстройство, требующее пожизненного лечения. Хотя доступны лекарства, которые эффективны в лечении некоторых пациентов, индивидуальная реакция на лечение может варьироваться, при этом у некоторых пациентов может наблюдаться устойчивость к одному или нескольким препаратам. В настоящее время мало что известно о причинах различий в ответе на лечение, наблюдаемых у лиц с шизофренией, а удовлетворительные маркеры плохого ответа недоступны для

клинической практики. В этой работе были изучены изменения уровней 322 липидов плазмы крови между двумя временными точками, измеренными у 92 человек с диагнозом шизофрения во время их стационарного лечения, и их связь со степенью улучшения симптоматики. Было обнаружено, что 20 триглицеридов были повышены у лиц с наименьшим улучшением показателей по шкале позитивных и негативных симптомов (Positive and Negative Syndrome Scale, PANSS), но не у лиц с наибольшим снижением симптоматики. Данные триглицериды отличались от остальных присутствующих в плазме крови. Они содержали малое количество атомов углерода в остатках жирных кислот и были низкоинтенсивными по сравнению с основными триглицеридами плазмы крови.

### Постановка исследования

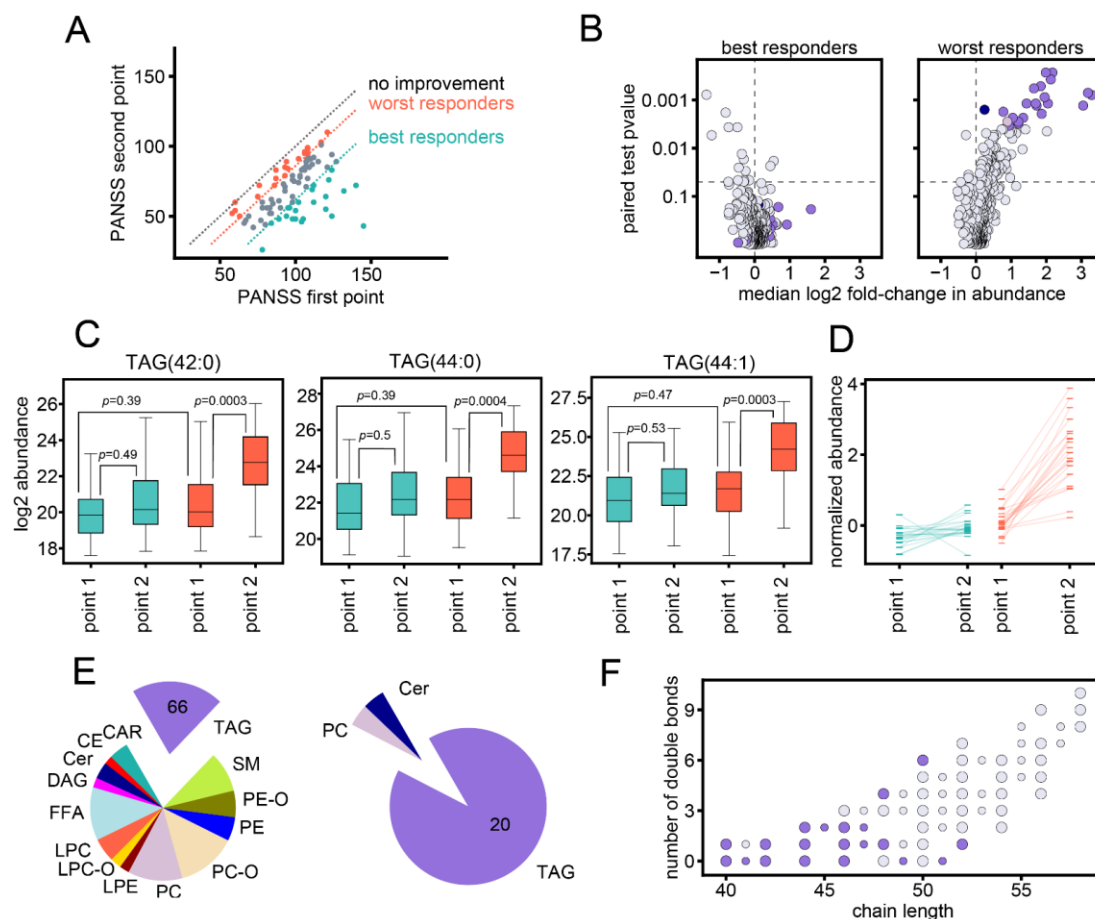
Мы оценили содержания 322 липидов в плазме крови у 92 человек с диагнозом шизофрения, собранной в двух временных точках: в начале и в конце их госпитализации в психиатрической клинике (среднее значение  $\pm$  стандартное нормальное отклонение:  $37 \pm 19$  дней). Выборка была представлена индивидуумами женского и мужского пола (58 % женщин) в возрасте от 17 до 43 лет, у которых также была собрана информация о принимаемых лекарственных препаратах. Тяжесть симптомов оценивали по шкале позитивных и негативных симптомов (Positive and Negative Syndrome Scale, PANSS) в двух временных точках. Липидные измерения были произведены с использованием масс-спектрометрии в сочетании с жидкостной хроматографией в отрицательном и положительном режимах ионизаций. Из полученных липидных признаков был аннотирован набор из 322 уникальных липидных соединений на основе их значений массы к заряду, времени удерживания, а также спектров фрагментации. Аннотированные соединения охватывали 14 классов липидов и хорошо соответствовали ожидаемому составу липидома плазмы крови.

Связь между изменениями содержания липидов и улучшением симптоматики

У всех пациентов, кроме одного, наблюдалось улучшение симптоматики во второй временной точке по сравнению с первой, о чем свидетельствовало снижение баллов по шкале PANSS (рис. 4A). Тем не менее, изменение значений PANSS сильно варьировалось. 25% пациентов с наибольшим улучшением симптоматики показали изменение оценок PANSS от -102 до -39 баллов ( $n = 23$ ), а 25% респондентов с наименьшим улучшением показали изменение оценок PANSS от -14 до 0 баллов ( $n = 24$ ) (рис. 4A). Степень изменений интенсивностей липидов между временными точками различались у пациентов в зависимости от улучшения их симптоматики. В то время как для 25% пациентов с наименьшим улучшением симптоматики 22 липида показали значимые



изменения между первой и второй временными точками (липиды, ассоциированные с наихудшим ответом; критерий Вилкоксона, поправка Бенджамини-Хохберг на множественное тестирование, ожидаемая доля ложно-положительных FDR = 5%; рис. 4B), эффект у 25 % пациентов с наибольшим улучшением симптоматики был ниже (без значимых изменений, критерий Вилкоксона, поправка Бенджамини-Хохберг на множественное тестирование, ожидаемая доля ложно-положительных FDR = 5%; рис. 4B). Соответственно, хотя уровни липидов, ассоциированных с наихудшим ответом, в первой временной точке были схожими у лиц с наибольшим и наименьшим улучшением симптоматики, у лиц с наилучшим ответом не наблюдалось статистически значимого увеличения от первой ко второй точки (соответственно: U-критерий Манна-Уитни, 1 из 22  $p < 0.05$ ; ранговый критерий Вилкоксона, все 22  $p > 0.05$ ; рис. 4C-D). Среди 22 липидов, ассоциированных с наихудшим ответом, 20 были триглицеридами, 30% от общего количества триглицеридов (рис. 4E). Наблюдаемый статистический эффект касался триглицеридов с меньшим числом атомов углерода (40-48 атомов углерода в остатках жирных кислот) (рис. 4F).



**Рис. 4.** Значимые изменения содержаний липидов для пациентов с наименьшим улучшением симптоматики. (A) Значения PANSS в первой и второй временной точках для 92 индивидуумов.

25% пациентов с наименьшим улучшением по шкале PANSS ( $n = 24$ ) и с наибольшим улучшением по шкале PANSS ( $n = 23$ ) отмечены оранжевым и зеленым соответственно. Оранжевая и зеленая пунктирные линии разграничивают верхний и нижний квартиль разницы в значениях PANSS, серая пунктирная линия соответствует идентичным значениям PANSS в двух временных точках. (B) P-значения для критерия Вилкоксона, построенные в зависимости от медианного кратного изменения интенсивностей липидов между двумя временными точками (в логарифмической шкале по основанию 2,  $\log_2FC$ ) для индивидуумов с наибольшим улучшением симптоматики (слева) и наименьшим улучшением симптоматики (справа). 22 липида, ассоциированных с наихудшим ответом, отмечены цветом соответствующего липидного класса: TAG (фиолетовый), PC (пыльно-розовый), Cer (темно-синий). Пунктирные линии разграничивают  $\log_2FC = 0$  и номинальное значение  $p = 0.05$ . (C) Интенсивности определенных липидов (в логарифмической шкале по основанию 2) для пациентов с наилучшим и наихудшим ответом для двух временных точек. Три липида, ассоциированных с наихудшим ответом и показывающие самый сильный статистический эффект ( $p < 0.0005$  для критерия Вилкоксона для оценки изменений у пациентов с наихудшим ответом) изображены на графике: TAG(42:0), TAG(44:0), TAG(44:1). Отмеченные P-значения соответствуют критерию Вилкоксона и U-критерию Манна-Уитни для сравнения между группами. Усы диаграммы размаха соответствуют стандартному определению. (D) Медианные значения нормализованных интенсивностей липидов (в логарифмической шкале по основанию 2) для пациентов с наилучшим и наихудшим ответом в двух временных точках для всех 22 липидов, ассоциированных с наихудшим ответом. Для каждого липида интенсивности были нормализованы согласно среднему значению липида у всех пациентов в первый момент времени. (E) Слева: 322 аннотированных липида, сгруппированных по липидным классам. Справа: 22 липида, ассоциированных с наихудшим ответом, сгруппированных по липидным классам. Количество триглицеридов указано. (F) Количество атомов углерода в остатках жирных кислот (длина цепи) и количество двойных связей для аннотированных триглицеридов. Липиды, ассоциированные с наихудшим ответом, окрашены в фиолетовый цвет. Большие и меньшие размеры кружков соответствуют триглицеридам с четной и нечетной длиной цепи, соответственно.

## Влияние лекарств и пола

Поскольку связь между изменениями концентрации липидов и тяжестью симптомов может быть искажена схемами лечения, подраздел 4.1.3 диссертации содержит анализ возможного влияния эффекта лекарственного лечения на полученные результаты. Далее, подраздел 4.1.4 содержит анализ возможного влияния дисбаланса по половому признаку на полученные результаты.

Подраздел 4.1.5 диссертации также рассматривает липидные профили в первой временной точке в качестве прогностических маркеров улучшения симптомов после лечения, однако статистически значимых результатов получено не было.

Длительный прием флуоксетина вызывает изменение липидов в мозге молодых макаков.

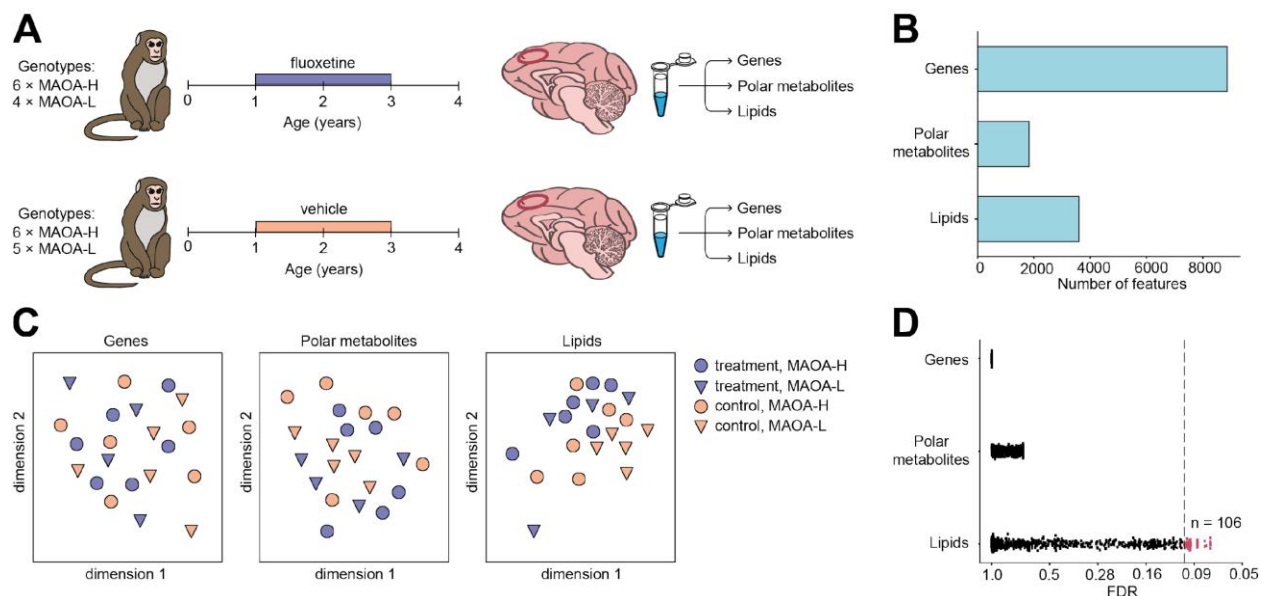
Данная глава следует результатам, опубликованным в [2]. Некоторые отрывки были дословно процитированы из [2], а рисунки воспроизведены или изменены с разрешением из [2].

## Аннотация

Флуоксетин — антидепрессант, назначаемый не только взрослым, но и детям для лечения депрессии, обсессивно-компульсивного расстройства и нарушений развития нервной системы. Побочные эффекты длительного лечения, отмеченные у некоторых пациентов, особенно у молодых людей, требуют детального изучения молекулярных изменений, вызванных лечением данным препаратом. В представленном исследовании было оценено остаточное влияние двухлетнего приема флуоксетина на экспрессию генов, содержания липидов и полярных метаболитов в прелимбической коре для 10 макак, получавших препарат, и 11 контрольных макак. Анализ 8871 транскриптов мРНК, 3608 липидов и 1829 полярных метаболитов выявил существенные изменения содержания липидов в головном мозге, включая значимые изменения для 106 липидных признаков, а также минорные изменения в экспрессии генов. Изменения липидов у животных, получавших антидепрессант, были наиболее очевидными для полиненасыщенных жирных кислот (ПНЖК). Снижение уровня ПНЖК наблюдалось для всех липидных классов за исключением сфинголипидов, которые обычно не содержат ПНЖК, что свидетельствует о системных изменениях метаболизма жирных кислот.

## Постановка исследования

В этом исследовании были оценены изменения экспрессии генов, полярных метаболитов и липидов в прелимбической коре (ПЛК; часть медиальной префронтальной коры) макак, получавших флуоксетин, используя методы секвенирования РНК, масс-спектрометрии ион-циклотронного резонанса с преобразованием Фурье и высокоэффективной жидкостной хроматографией с масс-спектрометрией (ВЭЖХ-МС), соответственно (рис. 5А). Прием флуоксетина и носителя (плацебо) начинался в возрасте одного года, что эквивалентно возрасту 4–6 лет у людей, и продолжался непрерывно в течение двух лет, после чего следовал период постдозирования в течение одного года, который заканчивался в возрасте четырех лет (приблизительное начало полового созревания). Эксперимент включал 10 обезьян, получавших флуоксетин, и 11 контрольных обезьян, которые получали носитель.



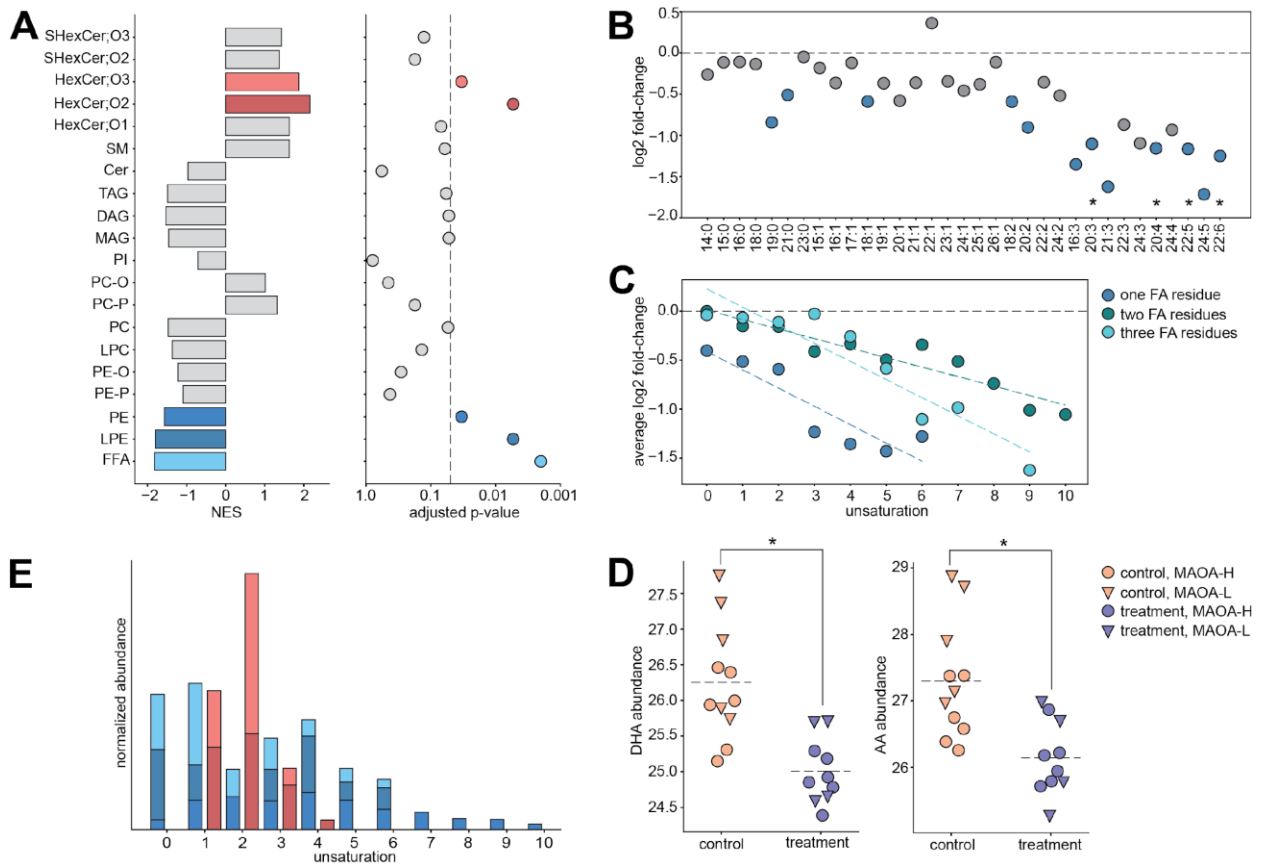
**Рисунок 5:** Дизайн эксперимента и количественные молекулярные фенотипы. (A) Схема экспериментального дизайна. (B) Количество измеренных генов, полярных метаболитов и липидов. (C) Графики многомерного масштабирования, отображающие различия между образцами, рассчитанные на основе уровней содержания 8871 гена, 1829 полярных метаболитов и 3608 липидов. Цвет соответствует статусу лечения (флуоксетин/носитель), форма - MAOA генотипу. (D) Распределение значений P-значений для t-критерия Стьюдента после поправки на множественное тестирование, рассчитанных для 8871 гена, 1829 полярных метаболитов и 3608 липидов при сравнении между животными получавших флуоксетин и носитель. Пунктирная линия соответствует отсечке FDR 10%. Количество липидов, прошедших пороговое значение FDR (красный), отмечено справа.

РНК-секвенирование предоставило количественные измерения экспрессии для 8871 белок кодирующих генов, аннотированных в геноме макаки. Масс-спектрометрические анализы выявили содержания для 1829 полярных метаболитов и 3608 липидных признаков, при этом из их числа 514 и 373 признака соответственно были автоматически аннотированы (рис. 5B). Визуализация вариации среди образцов на основе этих трех типов данных указывала на разделение животных, получавших флуоксетин или носитель, на уровне липидов, но не на уровне генов и полярных метаболитов (рис. 5C).

### Влияние флуоксетина на липиды

Статистический анализ выявил 106 липидных признака (липиды, ассоциированные с лечением), демонстрирующие значимые различия между обезьянами, получавшими флуоксетин, и контрольной группой (поправка Бенджамини-Хохберга, ожидаемая доли ложно-положительных FDR = 10%; пермутационное P-значение = 0.008). Напротив, эффект лечения был значительно слабее для экспрессии генов и содержания полярных метаболитов. Статистическая значимость была слишком низкой, чтобы определить метаболиты или гены, ассоциированные с лечением (рис. 7D; минимальное наблюдаемое P-значение после поправки на множественное

тестирование  $p = 0.9986$  и  $0.68740$  для экспрессии генов и содержаний полярных метаболитов, соответственно). Далее мы провели групповой анализ для оценки значимости эффекта лечения на уровне липидных классов. Этот подход выявил значимые различия для пяти классов липидов: свободных жирных кислот (FFA), фосфатидилэтаноламинов (PE), лизофосфатидилэтаноламинов (LPE) и гексозилцерамидов (HexCer;O2 и HexCer;O3) (рис. 6А; анализ обогащения наборов генов, Gene Set Enrichment Analysis, GSEA, скорректированное Р-значение  $p < 0.05$ ). Соответственно, липиды из этих классов имели скоординированный ответ на лечение, демонстрируемый увеличением количества липидов HexCer;O2 и HexCer;O3 в мозге обезьян, получавших лечение, и снижением липидов FFA, LPE и PE.



**Рисунок 6:** Влияние флуоксетина на липиды головного мозга. (A) Результаты анализа обогащения липидных классов, основанные на сравнении получавших лечение макак и контрольных животных. Слева: нормализованные оценки обогащения (Normalized Enrichment Score, NES). Положительный NES соответствует увеличению содержания липидов у получавших лечение животных, а отрицательный NES соответствует уменьшению. Цвета обозначают классы липидов, демонстрирующие значимое обогащение различий между лечением и контролем. Справа: скорректированные Р-значения. Пунктирная линия соответствует пороговому Р-значению после коррекции на множественное тестирование. (B) Кратные изменения в

содержании липидов (в логарифмической шкале по основанию два,  $\log_2 FC$ ), рассчитанные между макаками, получавшими лечение, и контрольными животными для липидов из класса свободных жирных кислот (FFA). Ось  $x$  указывают длину цепи жирной кислоты и количество двойных связей. Свободные жирные кислоты упорядочены по возрастанию ненасыщенности. Синими символами отмечены соединения с номинальным  $P$ -значением для  $t$ -теста  $p < 0.05$ . Звездочками отмечены статистически значимые соединения ( $t$ -критерий; ожидаемая доля ложно-положительных FDR = 10%). Пунктирная линия указывает  $\log_2 FC = 0$ . (C) Значения  $\log_2 FC$ , рассчитанные между макаками, получавшими лечение, и контрольными животными, усредненные для всех липидов с одинаковым уровнем ненасыщенности (количеством двойных связей). Сначала липиды были разделены на группы в зависимости от количества остатков жирных кислот: один (темно-синий), два (зеленый) или три (светло-синий), а сфинголипиды были исключены. Черная пунктирная линия указывает  $\log_2 FC = 0$ . Цветные пунктирные линии обозначают прямые линейной регрессии, соответствующие каждой группе ( $p = 0.0054$ ,  $0.54 \times 10^{-5}$  и  $0.00022$  для классов липидов, содержащих один, два и три остатка жирных кислот, соответственно). (D) Уровни содержания докозагексаеновой кислоты (DHA) и арахидоновой кислоты (AA) у макаков, получавших лечение, и контрольных животных. Символы представляют отдельные образцы. Цвета соответствуют статусу лечения, формы - MAOA генотипу. Звездочками отмечены статистически значимые различия между лечением и контролем ( $t$ -критерий; FDR = 10%). (E) Суммарные интенсивности липидов, содержащихся в липидных классах FFA, LPE, PE, HexCer;O2 и HexCer;O3 (цвета как на панели A) для каждого уровня насыщенности (количество двойных связей на соединение). Кумулятивное содержание было нормализовано между классами липидов.

Эффект флуоксетина был неоднородным в пределах одного липидного класса, так как зависел от состава остатков жирных кислот. В частности, липиды, содержащие жирные кислоты с множественными двойными связями (или ПНЖК) подвергались большим изменениям в результате лечения. Этот эффект был наиболее очевиден для свободных жирных кислот (рис. 6B), но также наблюдался для других классов глицерофосфо- и глицеролипидов (F-критерий для линейной регрессии,  $p = 0.0054$ ,  $0.00004$  и  $0.00022$  для липидных классов, содержащих один, два и три остатка жирных кислот соответственно; рис. 6C). Примечательно, что две ПНЖК, составляющие до 25% всего содержания жирных кислот в головном мозге, докозагексаеновая кислота (DHA, FFA 22:6) и арахидоновая кислота (AA, FFA 20:4), значимо различались между макаками получавшими лечение и контрольными животными (FDR с поправкой Benjamini-Hochberg = 10%; рис. 8D). В отличие от глицерофосфолипидов и глицеролипидов, которые содержат значительное количество остатков ПНЖК, липидные классы, обедненные ПНЖК, а именно, HexCer;O2 и HexCer;O3, значимо увеличились в результате лечения (рис. 6A,E; GSEA, скорректированный  $p < 0.05$ ). Подраздел 5.1.3 диссертации также включает анализ эффектов взаимодействия между MAOA генотипом и лечением флуоксетином на содержания липидов в головном мозге исследуемых животных.

### Групповой анализ данных экспрессии генов и полярных метаболитов

Хотя мы не обнаружили значимых различий в экспрессии отдельных генов или содержания полярных метаболитов в мозге макаков, получавших флуоксетин, подпороговые эффекты все еще могли быть информативными. Используя

групповой анализ, предназначенный для выявления таких эффектов, были обнаружены значимые эффекты лечения на уровне экспрессии генов, но не на уровне полярных метаболитов. В частности, 87 групп генов, определенных через онтологию генов (Gene Ontology, GO) и более подробно описанных в подразделе 5.1.4 диссертации, различались у макаков, получавших флуоксетин (анализ обогащения набора генов, GSEA, P-значения после поправки на множественное тестирование  $p < 0.05$ ).

## Анализ данных липидомики

Раздел 6.1 диссертации описывает предлагаемую методологию анализа данных нецелевой липидомики.

Ниже представлена сокращенная версия раздела 6.2 диссертации «Обширная дубликация сигнала в ВЭЖХ-МС липидомике».

## Аннотация

В нецелевой липидомике, основанной на ВЭЖХ-МС, одно и то же липидное соединение может быть представлено в данных несколько раз. Источником этого дублицированного сигнала являются различные аддукты и другие типы модификаций соединений, которые могут образовываться при электроспрей ионизации. Ожидаются хорошо описанные аддукты, такие как  $[M + NH_4]^+$ ,  $[M + Na]^+$ ,  $[M + K]^+$  в режиме положительной ионизации или  $[M-H+HCOO]^-$ ,  $[M-H+CH_3COOH]^-$ ,  $[M-Cl]^-$  в режиме отрицательной ионизации, но эти аддукты составляют лишь небольшую часть наблюдаемых дублицированных сигналов. Трудность определения возможных аддуктов и прочих модификаций заключается в их зависимости от условий эксперимента. В данном разделе представлен основанный на данных подход к их оценке. В частности, был использован набор из 29 различных внутренних стандартов для определения сигналов в данных, связанных с их добавлением к образцу. Была предложена процедура фильтрации, которая привела к существенному сокращению числа признаков при сохранении большинства автоматически аннотированных признаков. Предложенный метод имел преимущества по сравнению со стандартным подходом удаления только хорошо описанных аддуктов.

## Результаты

Набор различных внутренних стандартов ( $n = 29$ ; подробный список можно найти в подразделе 6.2.1 диссертации) использовался для оценки образования аддуктов (и других дублицированных сигнала) в нецелевом методе ВЭЖХ-МС. Пять повторов образцов плазмы были обогащены этими стандартами и сравнивались с пятью повторами без добавок.

Сначала, для сравнения содержания липидов в образцах с добавленными стандартами (образцы с добавкой) и без (образцы без добавки) использовался t-критерий Стьюдента. Признаки, которые имели более низкую среднюю

интенсивность в образцах с добавкой по сравнению с образцами без добавки, а также Р-значения, проходящие порог в 1% после поправки на множественное тестирование, были определены как признаки, связанные с добавкой ( $n = 4790$  и  $1561$  для режимов положительной и отрицательной ионизации, соответственно; t-критерий Стьюдента, поправка Бенджамини-Хохберга).

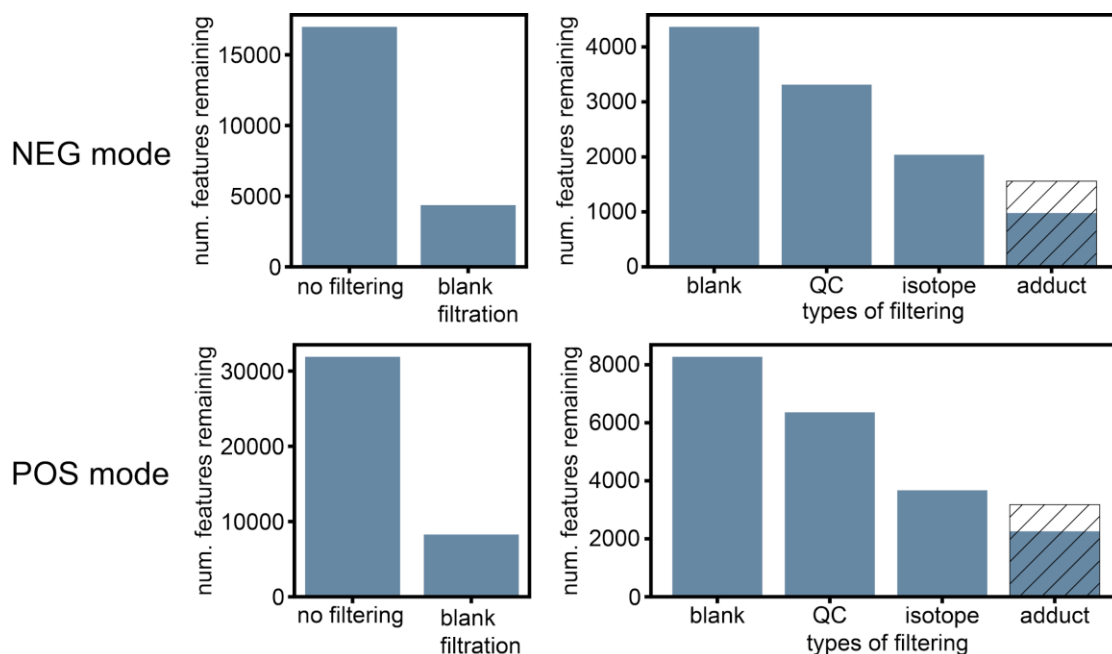
Далее, связанные с добавкой признаки были использованы для определения попарных различий в  $m/z$  для последующей фильтрации. Для каждого стандарта были рассчитаны разницы в значениях  $m/z$  между его основным аддуктом и связанных с добавкой признаками, имеющими то же время удержания в хроматографе (в интервале 0.02 минуты). Объединение этих разниц в значениях  $m/z$  для всех стандартов использовалась для определения разниц в  $m/z$ , общих для разных стандартов. Для это цели была применены гистограмма со скользящим окном вдоль значений  $m/z$  совместно с функцией обнаружения пиков, подробно описанные в подразделе 6.2.2 диссертации. Парные разницы в  $m/z$ , присутствующие по меньшей мере для 3 различных стандартов, сохранялись для последующей фильтрации ( $n = 69$  и  $23$  для положительного и отрицательного режимов ионизации, соответственно; таблица 2). В то время как некоторые значения  $m/z$  были ожидаемые, другие были получены на основе данных и не могли быть охарактеризованы из литературы.

Отрицательный режим ионизации		Положительный режим ионизации	
$m/z$ разница	Кол-во стандартов	$m/z$ разница	Кол-во стандартов
82.00199997	13	21.9819999	19
164.006	11	58.05199997	14
67.986999999	10	83.9519999999	11
246.009	9	106.9999	9
149.99	8	157.957	8
-24.04400002	8	11.949	8
-14.01500002	8	89.9699997	7
157.989	8	219.926	7
84.9799997	8	79.9399997	7
- 774.93700002	7	179.936	7
231.994	7	43.96299998	7
239.993	7	233.943	7
		225.944	7



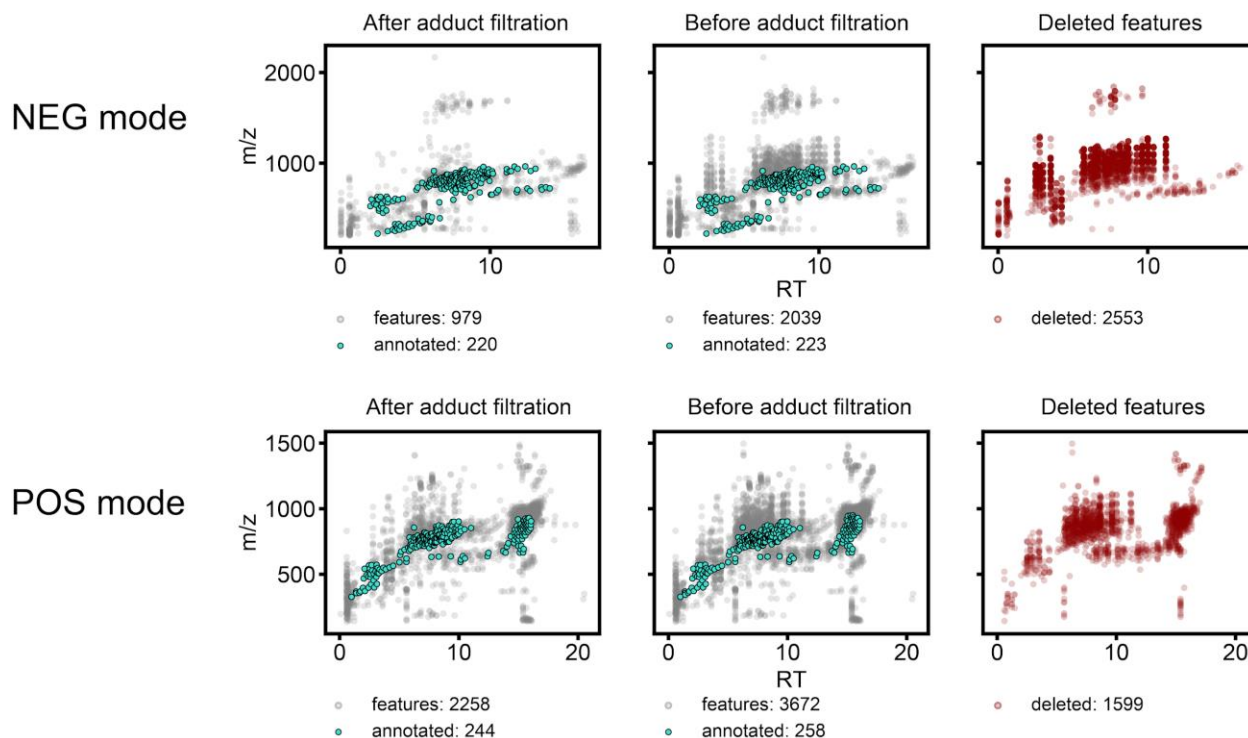
**Таблица 2:** Самые частые попарные разницы в  $m/z$ , определенные через сравнение образцов с добавкой и без.

Удаление дублицированных сигналов выполнялось путем поиска совместно элюирующих соединений (признаков с одинаковым временем удержания в хроматографе, в интервале 0.02 минуты) с этими полученными разницами в  $m/z$ . Поскольку различия в значениях  $m/z$  были рассчитаны на основе одного основного аддукта и одного неаннотированного признака, выбор того, какой из двух признаков следует фильтровать, был прямолинейным. После выполнения стандартных процедур фильтрации данных, включая фильтрацию по холостым образцам, фильтрацию по техническим репликам и удаление  $C^{13}$  изотопов, описанный метод фильтрации дублицированных сигналов удалил 1414 и 1060 признаков в положительном и отрицательном режимах ионизации, соответственно (рис. 7). Важно отметить, что стандартный подход, когда рассматриваются только хорошо описанные и распространенные аддукты, работал хуже (рис. 7). Только 35 и 45 % признаков, удаленных вышеописанным методом, были бы найдены при стандартной фильтрации аддуктов в двух режимах, соответственно.



**Рисунок 7:** Количество признаков, оставшихся в образцах плазмы без добавок после различных процедур фильтрации. Слева: без фильтрации и после фильтрации по холостым образцам (удаление контаминантов/ложно-обнаруженных признаков). Справа: после фильтрации по холостым пробами, техническим повторам (удаление признаков с технической вариабельностью), фильтрации изотопов и фильтрации аддуктов/ дублицированных сигналов. Заштрихованный столбец показывает количество признаков, оставшихся после фильтрации только по обычным аддуктам, более подробно описанной в подразделе 6.2.1 диссертации.

Для оценки эффективности описанной процедуры фильтрации был рассмотрен набор автоматически аннотированных соединений. Из 258 и 223 аннотированных липидов в положительном и отрицательном режимах ионизации, 15 и 3 были удалены как дублицированные сигналы (рис. 8).



**Рисунок 8:** Значения  $m/z$  и времени удерживания (retention time, RT) измеренных признаков в положительном и отрицательном режимах ионизации. Слева направо: признаки, оставшиеся после очистки данных с удалением нежелательных дублицированных признаков, неаннотированные (серые точки) и аннотированные (синие точки) признаки; признаки, оставшиеся после стандартной очистки данных до удаления избыточных аддуктов, неаннотированные (серые точки), и аннотированные (синие точки) признаки; удаленные фильтрацией аддуктов признаки, красные точки. Количество признаков указано под графиками.

В качестве альтернативы предложенному здесь методу фильтрации дублицированных сигналов, попарные различия в  $m/z$  признаков могут быть определены из всего набора измеренных признаков [3]. В подразделе 6.2.1 далее обсуждаются преимущества предложенного здесь метода по сравнению с существующими альтернативами.

## Выводы

Основные результаты диссертации представлены ниже:

- Был предложен метод обработки данных нецелевого анализа липидома с изложением и критической оценкой основных этапов обработки, обеспечивающий практическое руководство для подобного анализа.
- Был предложен основанный на данных подход для уменьшения количества дублицированных сигналов в нецелевых экспериментах по

липидомике, который удалял в два раза и более избыточных признаков, чем стандартный подход фильтрации аддуктов, сохраняя при этом большинство аннотированных признаков.

- Был обнаружен профиль изменений липидов в плазме крови больных шизофренией, надежно воспроизведенный в нескольких независимых когортах.
- Было показано, что описанный липидный профиль также характерен для других психических расстройств, а именно, большому депрессивному расстройству и биполярному расстройству.
- Была предложена диагностическая модель на основе липидных измерений, отделяющая людей с шизофренией от контрольной группы с высокой диагностической способностью (площадь под кривой рабочей характеристики приемника, ROC AUC = 0.86-0.95). Модель была валидирована на двух дополнительных независимых наборах данных.
- Был описан особый профиль липидов, связанный с плохим ответом на лечение при шизофрении, а именно, было показано, что увеличение триглицеридов с более короткими цепочками в остатках жирных кислот было связано с плохим симптоматическим ответом после лечения.
- Было показано, что липиды мозга молодых макак изменяются в результате приема флуоксетина, в то время как другие модальности данных, экспрессия генов и концентрации полярных метаболитов, не показали значимых изменений. Было показано, что среди липидов, именно полиненасыщенные жирные кислоты (ПНЖК) были снижены в мозге макак, подвергшихся лечению флуоксетином. При этом для свободных жирных кислот эффект был наиболее сильным, однако снижение уровней других липидов, содержащих ПНЖК, было заметным.

## Список литературы

1. Tkachev A, Stekolshchikova E, Anikanov N, Zozulya S, Barkhatova A, Klyushnik T, et al. Shorter chain triglycerides are negatively associated with symptom improvement in schizophrenia. *Biomolecules*. 2021. 2021. <https://doi.org/10.3390/biom11050720>.
2. Tkachev A, Stekolshchikova E, Bobrovskiy DM, Anikanov N, Ogurtsova P, Park DI, et al. Long-term fluoxetine administration causes substantial lipidome alteration of the juvenile macaque brain. *Int J Mol Sci*. 2021. 2021. <https://doi.org/10.3390/ijms22158089>.
3. Mahieu NG, Patti GJ. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Anal Chem*. 2017. 2017. <https://doi.org/10.1021/acs.analchem.7b02380>.