

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ».

Физический факультет

Кафедра общей физики

Мартюшев Алексей Антонович

КУРСОВАЯ РАБОТА

Автоматизация обработки и анализа метаболомных данных

Электромагнитный практикум, 2 курс, группа № 18345

Научный руководитель:

Оценка научного руководителя

«_____» _____ 20__ г.

Преподаватель практикума

Оценка преподавателя практикума

«_____» _____ 20__ г.

Куратор практикума:

к.т.н. Астрелин В.Т.

Итоговая оценка

«_____» _____ 20__ г.

Новосибирск, 2019 г.

Оглавление

1. Введение.....	3
2. Теоретическая часть.....	5
2.1. Масс-спектрометрия	5
2.2. Обработка метаболомных данных.....	7
3. Экспериментальная часть.....	10
3.1. Описание экспериментальной установки	10
3.2. Результаты.....	10
4. Заключение	14
Список литературы	15

1. Введение

Метаболомом называется полный набор метаболитов – низкомолекулярных химических соединений (с молекулярной массой менее 1 кДа), содержащихся в биологическом образце (клетке, органе, биологической жидкости и т. д.). Содержание определенных метаболитов в конкретных биологических образцах находится в непосредственной зависимости от внешних и внутренних факторов: от пола, возраста, образа жизни, окружающей его среды, а также от наличия различных патологий. Таким образом, анализ метаболома и изменений в нем интересен, прежде всего, как метод исследования развития заболеваний и патологий, а также как способ создания более полного представления о живых организмах в целом.

При анализе метаболома существует множество технических сложностей. Одна из них заключается в том, что при регистрации метаболитов, приходится обрабатывать весьма большие объемы данных. Например, на данный момент в плазме крови человека зарегистрировано около 4200 метаболитов, в моче — 3000 [1]. Такое количество метаболитов невозможно зарегистрировать в одном эксперименте (обычно в эксперименте количество метаболитов не превышает нескольких сотен). При этом метаболомный состав разных биологических образцов сильно различается, а сами метаболиты могут быть веществами с абсолютно различными свойствами, их концентрации также могут значительно варьироваться. Все эти факторы приводят к необходимости создания наиболее точных методов регистрации метаболитов и алгоритмов для обработки большого объема полученных данных.

Одним из основных методов регистрации метаболитов является высокоэффективная жидкостная хроматография с масс-спектрометрическим детектированием (ВЭЖХ-МС). Процесс получения и регистрации метаболитов состоит из следующих этапов: экстракция метаболитов из биологических образцов, разделение сложной смеси метаболитов на более простые фракции с помощью жидкостной хроматографии, ионизация полученных соединений,

детектирование ионов с помощью масс-спектрометрии, обработка и анализ данных.

Объектом исследования в данной работе являлись эмбриональные стволовые клетки (ESC), индуцированные плюрипотентные стволовые клетки (iPSC) и эмбриональные фибробласты (EF). Эмбриональные стволовые клетки выделяют из внутренней клеточной массы бластоцисты. Индуцированные плюрипотентные стволовые клетки получают из дифференцированных клеток путем их репрограммирования, что является более простым и удобным методом. Регенеративный потенциал стволовых клеток предоставляет возможность лечения заболеваний и восстановления, после перенесенных болезней. Потому, в данном случае крайне важна простота и скорость получения стволовых клеток в большом объеме, при этом важно правильно понимать и оценивать сходства и различия индуцированных плюрипотентных и эмбриональных стволовых клеток.

Целью данной работы является обработка и анализ метаболомных данных, полученных с помощью метода ВЭЖХ-МС. Для этого были поставлены следующие задачи: изучение метода ВЭЖХ-МС, изучение методов обработки полученных данных с помощью пакета прикладных программ на языках программирования R и python и применение данных методов непосредственно для анализа и обработки данных.

2. Теоретическая часть

2.1. Масс-спектрометрия

Масс-спектрометрия – метод исследования вещества, основанный на определении отношения массы иона к его заряду. Данные масс-спектрометрии представляются в виде масс-спектров – зависимости интенсивности ионного сигнала от отношения массы к заряду иона.

Почти все масс-спектрометры — вакуумные приборы, так как ионы крайне нестабильны в присутствии других молекул.

Масс-спектрометр состоит из трех компонент: ионного источника (ионизатора), масс-анализатора и детектора. Ионизатор преобразует частицы исследуемого вещества в ионы. При использовании метода ВЭЖХ-МС наиболее распространенным источником ионизации является электроспрей или ионизация распылением в электрическом поле при атмосферном давлении (ESI, electrospray ionization).

Рассмотрим стадии формирования ионов с помощью электроспрея:

изначально посредством
приложенного напряжения
происходит

перераспределение заряда в
растворе, в результате
которого, положительные
ионы скапливаются у выхода
капилляра. При достаточно
сильном приложенном поле

образуется струя из вершины конуса, которая далее разлетается на мелкие капли. Происходит сокращение размеров заряженных капель за счет испарения растворителя и последующего распада капель на более мелкие, за счет выравнивания сил кулоновского отталкивания ионов и поверхностного натяжения капли. Так происходит вплоть до получения ионов.

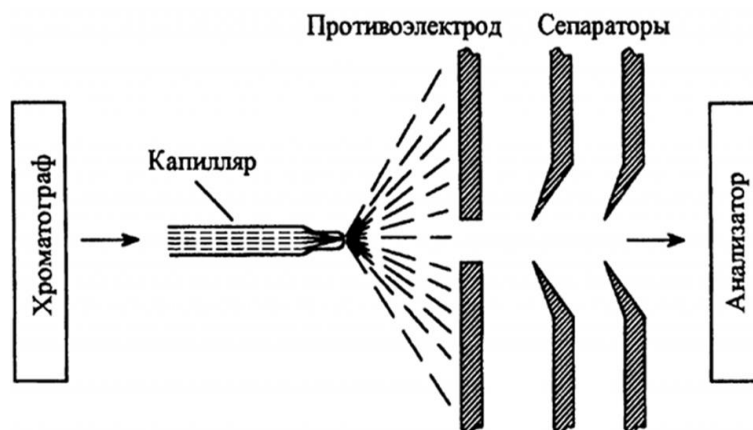


Рис. 1. Ионизация электрораспылением (электроспрей).

Масс-анализатор производит сортировку ионов по отношению их массы к заряду. Масс-анализаторы подразделяются на непрерывные и импульсные, разница между ними заключается в том, что на первые ионы подаются непрерывным потоком, а на вторые – порциями. Также можно подразделить масс-анализаторы на секторные, времяпролетные и ортогональные времяпролетные. Последние позволяют сочетать непрерывный источник ионов (электроспрей) с времяпролетным масс-анализатором, что позволяет получать масс-спектры высокого разрешения.

Простейший принцип работы времяпролетного масс-анализатора основывается на том, что среди ионов, ускоренных электрическим полем до одинаковой кинетической энергии, ионы с наименьшей массой, будут иметь наибольшую скорость и достигнут детектора раньше ионов с большей массой. Таким образом, отношение массы к заряду ионов определяется исходя из времени полета иона до детектора в дрейфовой вакуумной камере (без электрического поля). Отсчет времени начинается с момента подачи импульса и заканчивается в момент попадания иона на детектор. Время полета для каждого отдельного иона уникально и легко определяется из закона сохранения энергии:

$$\frac{mv^2}{2} = ZeU,$$

учитывая:

$$v = \frac{L}{t},$$

очевидно:

$$\frac{m}{Z} = t^2 \frac{2eU}{L^2},$$

где U и L – значения разности потенциалов и расстояния, которые необходимо преодолеть иону.

При определении значения t также необходимо учитывать разницу во времени между отправкой сигнала о стартовом импульсе и действительном моментом подачи импульса, и, соответственно, между моментом попадания иона на детектор и регистрацией сигнала. Для этого обычно используют калибровочную кривую, которая строится по соединениям с известным отношением массы к заряду.

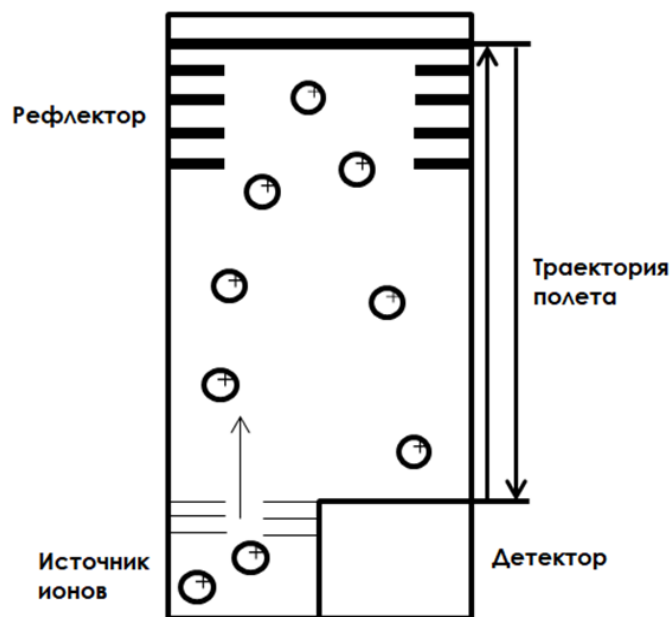


Рис. 2. Схема работы ортогонального времяпролетного масс-спектрометра.

Современные масс-спектрометры обладают точностью определения отношения массы к заряду на уровне 1 ppm. Такая высокая точность позволяет, во-первых определять химическую формулу изучаемого соединения, а во-вторых одновременно регистрировать большое число соединений, даже при незначительной разнице в массе. Таким образом, масс-спектрометрия является довольно мощным способом для определения химических формул исследуемых соединений.

2.2. Обработка метаболомных данных

Как упоминалось, одной из основных проблем, возникающих при обработке и анализе метаболомных данных, связана с большим количеством детектируемых в образцах соединений. Также в хроматограммах могут наблюдаться крайне разнообразные формы пиков, не похожие на форму гауссовой функции, что приводит к трудностям поиска пиков. Стоит учитывать, что при ионизации соединения образуются различные ионы, например, $M+H$, $M+Na$, а также возможно наличие изотопов. Из-за этого увеличивается количество различных сигналов, регистрируемых у одного соединения.

В результате ВЭЖХ-МС эксперимента получается трехмерный массив данных, одной координатой в котором является время удерживания на хроматографической колонке, другой - отношение массы к заряду, третьей - интенсивность сигнала

(рис. 3). По этим данным можно построить хроматограмму по интересующему иону, проинтегрировать пик, получить тем самым его площадь. Площадь пика будет пропорциональна концентрации данного соединения в исследуемом образце. Однако осуществление всех этих действий для каждого соединения в каждом образце без использования специализированных средств обработки данных является крайне трудоёмкой задачей.

Обработка данных состоит из следующих этапов: нахождение пиков и их интегрирование, коррекция времен удерживания на хроматографической колонке, поиск изотопов и ионов, принадлежащих одному соединению, и сопоставление пиков между образцами для создания таблицы данных, которая содержит все найденные пики во всех исследуемых образцах, далее данные анализируются с помощью методов математической статистики.

Этап поиска и интегрирования пиков является алгоритмически сложным из-за большого количества пиков различной формы и интенсивности. Также существуют проблемы из-за различия времени удерживания компонент на хроматографической колонке от эксперимента к эксперименту. Таким образом, спектры различных образцов могут быть заметно сдвинуты относительно друг друга. Для лучшего соотнесения спектров используется коррекция времен

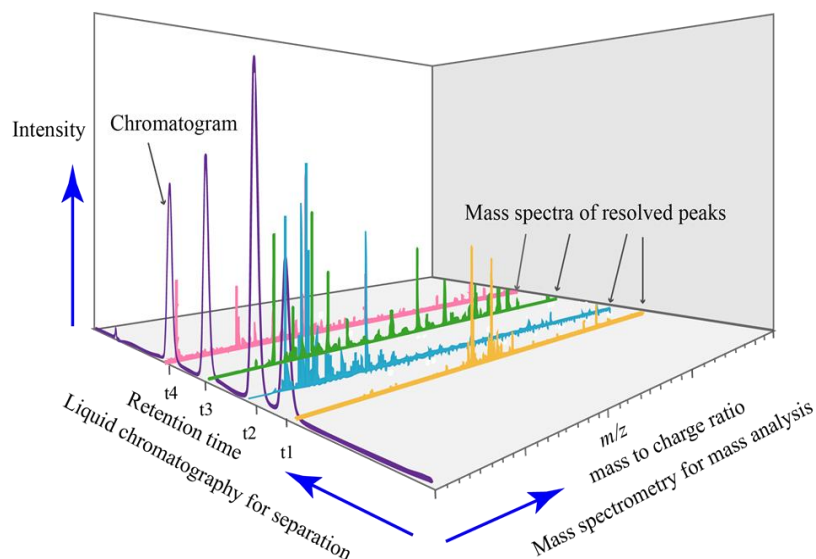


Рис. 3. Данные, получаемые в результате ВЭЖХ-МС.

удержания. Этот процесс обеспечивает выравнивание пиков по времени, что упрощает сопоставление пиков среди образцов.

Для визуализации данных и понимания зависимостей между образцами, используется анализ с помощью метода главных компонент (PCA, Principal Component Analysis). Он использует ортогональное преобразование для превращения системы возможно коррелирующих переменных в набор линейно не коррелирующих. При этом базисные векторы пространства признаков переходят в новые базисные векторы, называемые главными компонентами. То есть этот метод является линейным подходом к понижению размерности, в котором вводятся новые признаки, представляющие собой линейную комбинацию исходных признаков. Таким образом, вектор $\vec{x}$ переходит в вектор $\vec{y}$ следующим образом:

$$y_i = \sum_{j=1}^n w_{ij} x_j,$$

где w_{ij} – весовой коэффициент, с которым старый признак i входит в новый признак j . Это линейное преобразование определяется таким образом, чтобы проекция векторов $\vec{y}_i$ на первую главную компоненту имела наибольшую возможную дисперсию, следующей по величине дисперсию имела проекция векторов $\vec{y}_i$ на вторую главную компоненту, ортогональную первой и т.д.

Далее для полученных данных применяются статистические критерии однородности, проверяющие гипотезу о том, что концентрация данного метаболита отличается между группами. К этим критериям относятся U-критерий Манна – Уитни и t-критерий Стьюдента. Последний предполагает, что данные имеют нормальное распределение, потому более корректным для биологических данных считается использование U-критерия Манна – Уитни.

Для первичной обработки данных ВЭЖХ-МС использовалась библиотека XCMS, написанная на языке программирования «R». Библиотека включает в себя алгоритмы по нахождению и интегрированию пиков, коррекции времени

удержания соединений на хроматографической установке и сопоставлению пиков принадлежащих одному образцу.

3. Экспериментальная часть

3.1. Описание экспериментальной установки

Для анализа образцов в данной работе использовался масс-спектрометр Bruker Daltonics Maxis 4G с ионизатором ESI и времяпролетным анализатором в сочетании с высокоэффективным жидкостным хроматографом UltiMate 3000RS.

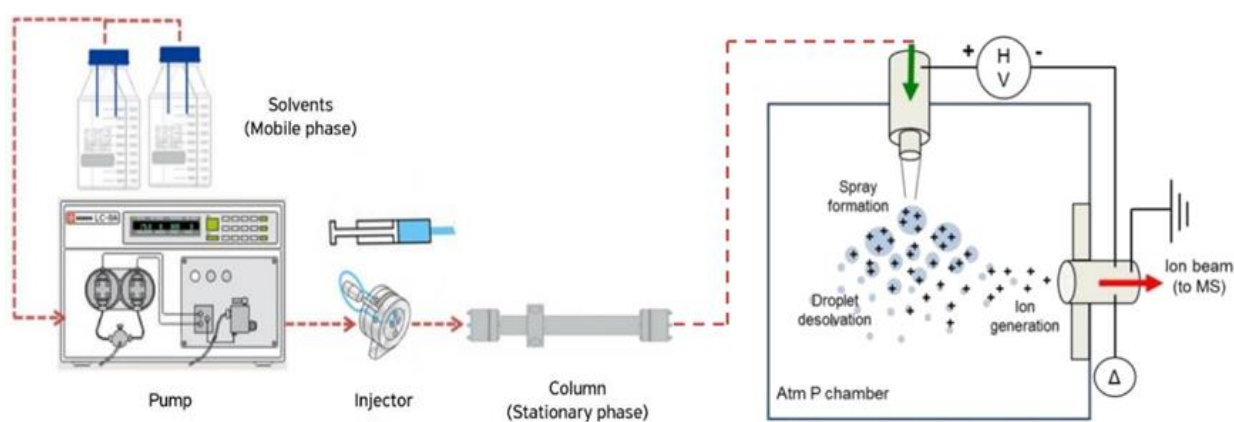


Рис. 4. Схема высокоэффективной жидкостной хроматографии с масс-спектрометрическим детектированием.

3.2. Результаты

В эксперименте исследовались 5 образцов эмбриональных стволовых клеток (ESC), 5 образцов индуцированных плюрипотентных стволовых клеток (iPSC) 4 образца эмбриональных фибробласт (EF). На рис. 5 представлена хроматограмма (зависимость интенсивности ионного тока от времени удержания) полученных необработанных данных масс-спектрометрии.

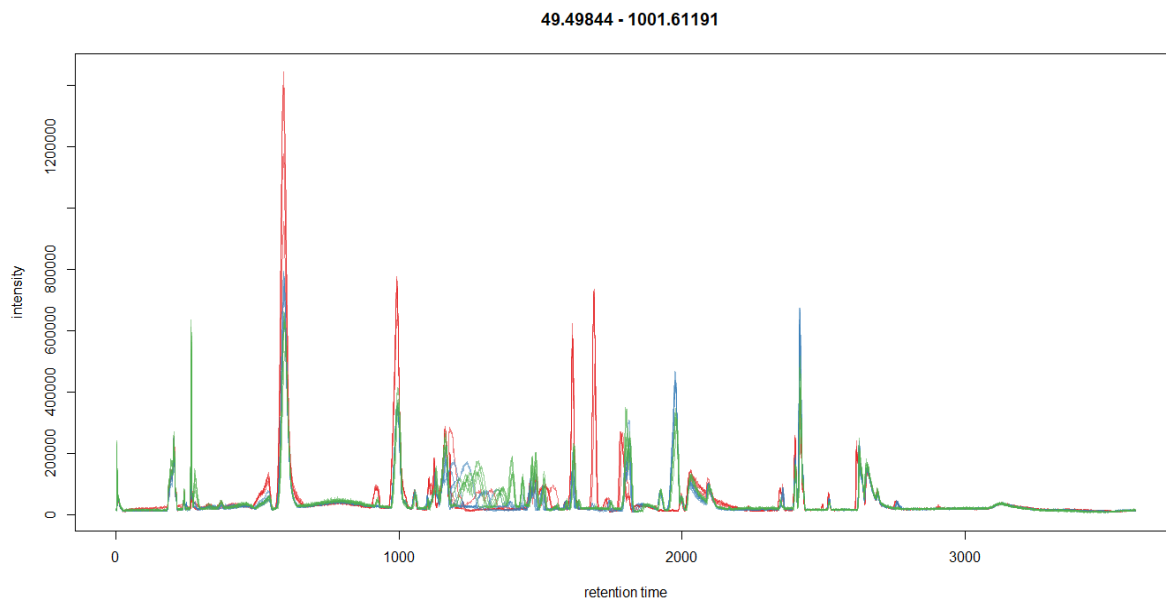


Рис. 5. Общая хроматограмма, полученная в результате эксперимента. Красным цветом обозначены EF, синим - ESC, зеленым - iPSC

Далее, с помощью XCMS был произведен поиск пиков, их интегрирование, коррекция времени удержания, группировка пиков, принадлежащих одному соединению. В результате была получена таблица с данными отношения массы к заряду для каждого иона, времени удержания соответствующего иону соединения, и интенсивностью пика для каждого из 14 исследуемых образцов.

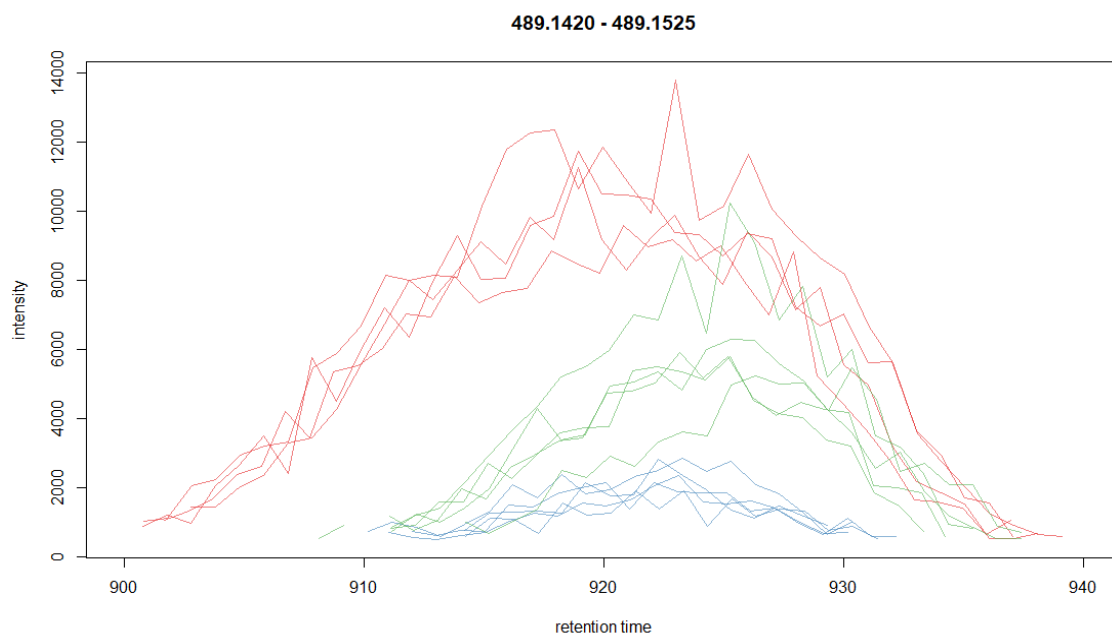


Рис. 6. Хроматограммы по выделенному иону для $m/z = 489.146 \pm 0.005$ по всем образцам. Красным цветом изображены хроматограммы образцов EF, синим - ESC, зеленым - iPSC.

Также была проведена визуализация данных методом главных компонент (PCA). PCA показал довольно четкое разделение образцов на 3 группы (рис. 7). Однако между группами ESC и iPSC различия оказались менее значительными, чем между ESC и EF, и iPSC и EF.

Затем был произведен поиск различий между образцами. Сравнивались интенсивности пиков, относящихся к одному и тому же иону, между группами (EF и ESC, EF и iPSC, ESC и iPSC). Сравнение производилось для каждого найденного иона. Для этого к каждой паре групп был применен U-критерий Манна-Уитни, с помощью которого рассчитывалось

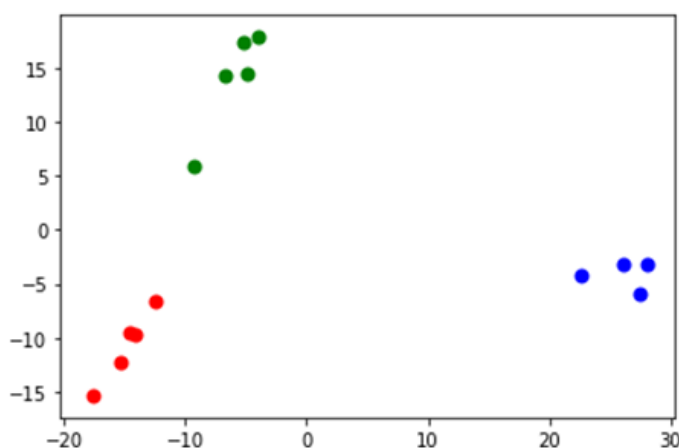


Рис. 7. Метод главных компонент, примененный к полученным данным. Красные точки – образцы iPSC, зеленые - ESC, синие - EF.

значение достоверности критерия p-value. Значение p-value $\ll 0.05$ означало, что две сравниваемые группы различны, то есть признавалась статистическая значимость различий между значениями интенсивности в сравниваемых группах. В дополнение к этому вычисляли само отношение средних значений интенсивностей между группами. Наибольшему вниманию были подвергнуты ионы, которые, во-первых, статистически различаются, и, во-вторых, различаются минимум в 2 раза между двумя группами по среднему значению интенсивности. Также, производился отбор по среднему значению интенсивности, которое хотя бы в одной из групп должно было превышать установленный минимальный порог, в нашем случае – значения средней интенсивности 50000. В конечном итоге были определены метаболиты, концентрации которых значительно различаются между группами, названия и формулы которых определялись с помощью базы данных метаболома человека (Human Metabolome Database).

Прежде всего, исследовались различия между группами эмбриональных стволовых клеток и индуцированных плюрипотентных стволовых клеток.

Таблица 1. Список метаболитов, концентрации которых заметно различаются между группами ESC и iPSC.

Предположительное название метаболита	m/z иона, относящегося к соединению (Да)	Химическая формула	Отношение средних интенсивностей (ESC к iPSC)	p- value
Adenosine	268.1041	C ₁₀ H ₁₃ N ₅ O ₄	0.175	0.0122
Glutathione disulfide	307.0833	C ₁₀ H ₁₀ O ₇ S	2.848	0.0102
Guanosine	284.0989	C ₁₀ H ₁₃ N ₅ O ₅	0.245	0.0122
L-Phenylalanine	166.0863	C ₉ H ₁₁ NO ₂	1879.998	0.0097

Затем рассматривались сходства между ESC и iPSC и их различия с группой EF.

Таблица 2. Список метаболитов, концентрации которых схожи между группами ESC и iPSC и заметно различаются с группой EF.

Название метаболита	m/z, (Да)	Хим. формула	Отношение средних интенсивностей (EF к ESC)	Отношение средних интенсивностей (EF к iPSC)	p- value
Creatine	170.0326	C ₄ H ₉ N ₃ O ₂	20.667	21.223	0.0199
Proline	116.0706	C ₅ H ₉ NO ₂	2.361	3.190	0.0200
Taurine	163.9779	C ₂ H ₇ NO ₃ S	109.364	131.289	0.0199

Для метаболитов из Таблицы 1 и Таблицы 2, а также, для некоторых других отличающихся метаболитов, название которых в точности определить не удалось, были построены диаграммы интенсивностей в каждой группе образцов. Пример диаграммы для m/z = 307.083 ± 0.005 Da (Glutathione disulfide) приведён на Рис. 8.

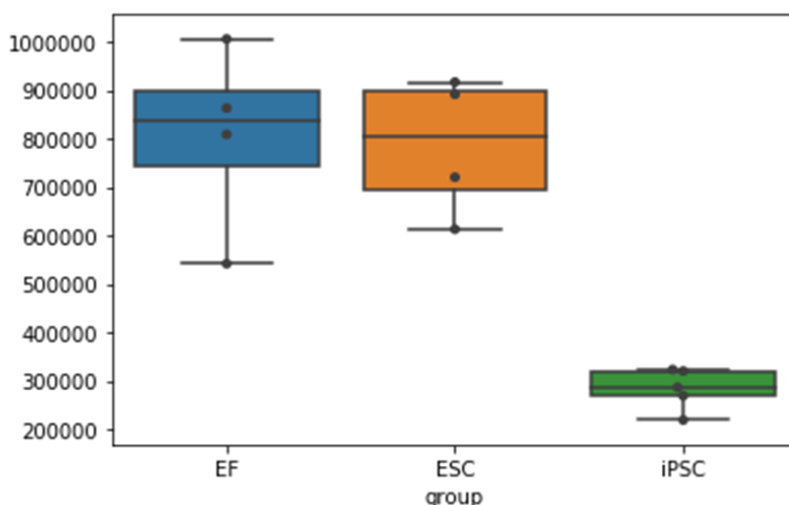


Рис. 8. Диаграмма интенсивностей в каждом образце для иона с m/z = 307.083 ± 0.005 Da (Glutathione disulfide).

Таким образом, была получена возможность наглядно оценить различие между образцами.

В результате сравнения, оказалось, что образцы ESC и iPSC, несмотря на некоторые различия в составе основных низкомолекулярных составляющих, все равно оказались более схожи между собой, чем каждая из этих групп с EF. Это может служить поводом для проведения дальнейших исследований, в результате которых станет известно, насколько незначительные изменения в метаболомном составе индуцированных плюрипотентных стволовых клеток скажутся на составе и функционировании уже дифференцированной клетки.

4. Заключение

Был проведен метаболомный анализ 5 образцов эмбриональных стволовых клеток, 5 образцов индуцированных плюрипотентных стволовых клеток и 4 образцов эмбриональных фибробласт, с помощью метода ВЭЖХ-МС. Удалось обнаружить значительное изменение концентраций некоторых метаболитов между образцами ESC и iPSC, однако таких метаболитов оказалось гораздо меньше, по сравнению с общим числом обнаруженных, что может свидетельствовать о схожести метаболомного состава ESC и iPSC.

Список литературы

1. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, Young N, Xia J, Knox C, Dong E, Huang P, Hollander Z, Pedersen TL, Smith SR, Bamforth F, Greiner R, McManus B, Newman JW, Goodfriend T, Wishart DS (Feb 2011). "The human serum metabolome". *PLoS ONE*. **6**(2): e16957.
2. Talrose V. L., Ljubimova A. K. Secondary Processes in the Ion Source of a Mass Spectrometer (Reprint from 1952). *J. Mass Spectrom.* 1998, 33, 502—504.
3. Malcolm Dole, L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson. Molecular Beams of Macroions // *The Journal of Chemical Physics*. — 1968-09-01. — Т. 49, вып. 5. — С. 2240–2249.
4. Sas KM, Karnovsky A, Michailidis G, Pennathur S. Metabolomics and diabetes: analytical and computational approaches. *Diabetes*. 2015 Mar 1;64(3):718-32.
5. Downard, Kevin (2004). *Mass Spectrometry - A Foundation Course*. Royal Society of Chemistry.
6. Nordström, A; O'Maille G; Qin C; Siuzdak G. (2006). "Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum". *Analytical Chemistry*. **78** (10): 3289–95.