

На правах рукописи

Власова Мария Андреевна

**ПРОТЕОМНАЯ ДИАГНОСТИКА РАКА ЯИЧНИКА
С ПРИМЕНЕНИЕМ SELDI-МАСС-СПЕКТРОМЕТРИИ**

03.00.04 – биохимия

АВТОРЕФЕРАТ

**ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ
КАНДИДАТА БИОЛОГИЧЕСКИХ НАУК**

Москва 2007

Работа выполнена в Государственном Учреждении Научно-исследовательском институте биомедицинской химии имени В.Н. Ореховича Российской академии медицинских наук.

Научный руководитель:

кандидат биологических наук
Мошковский Сергей Александрович

Официальные оппоненты:

доктор биологических наук
Прозоровский Владимир Николаевич
доктор биологических наук
Ковалев Леонид Иванович

Ведущая организация: Государственное Учреждение Российский онкологический научный центр им. Н.Н. Блохина Российской академии медицинских наук

Защита диссертации состоится 11 октября в 11 часов на заседании диссертационного совета Д.001.010.01 при ГУ НИИ биомедицинской химии им. В. Н. Ореховича РАМН по адресу: 119121, Москва, ул. Погодинская, 10.

С диссертацией можно ознакомиться в библиотеке ГУ НИИ биомедицинской химии им. В. Н. Ореховича РАМН по адресу: 119121, Москва, ул. Погодинская, 10.

Автореферат разослан « » сентября 2007 г.

Ученый секретарь Диссертационного совета
кандидат химических наук

Е.А. Карпова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

В современном обществе одной из основных причин смертности являются злокачественные опухоли. Эффективность лечения рака во многом определяется возможностью его ранней диагностики, поэтому важнейшей задачей является разработка простых и эффективных способов диагностики рака на ранних стадиях. Проблема ранней диагностики в особенности актуальна в отношении рака яичника. Ранние стадии рака яичника протекают практически бессимптомно, в результате почти в 90% случаев рак яичника диагностируют только на поздних стадиях, когда заболевание уже плохо поддается лечению [Jacobs, Menon (2004) Mol. Cell. Proteomics. 3, 355-366].

Одним из распространенных и перспективных способов малоинвазивной диагностики рака является диагностика по концентрации в крови опухолевых биомаркеров. К настоящему времени для клинической диагностики рака яичника используют только один биомаркер – СА125. Однако этот биомаркер обладает невысокими чувствительностью и специфичностью (при специфичности 97% чувствительность составляет всего 65%) [Zhang et al (2004), Cancer Res., 64, 5882-5890].

Интенсивное развитие протеомных технологий открыло новые перспективы для поиска биомаркеров заболеваний. Одной из новых и наиболее высокопроизводительных и перспективных платформ для поиска биомаркеров является масс-спектрометрическая технология белковых чипов SELDI-TOF (Surface-enhanced laser desorption/ionization time-of-flight, усиленная поверхностью времяпролетная лазерная десорбция/ионизация). Применение масс-спектрометрии SELDI-TOF в этих целях началось около пяти лет назад и сразу показало многообещающие результаты. В ряде работ была продемонстрирована точность диагностики, близкая к 100 % [Petricoin et al (2002) Lancet, 359, 572-577; Kozak et al (2003) Proc. Natl. Acad. Sci. USA, 100, 12343-12348]. Однако неудовлетворительная воспроизводимость результатов между разными

лабораториями и во многих случаях неизвестная природа дискриминаторных масс-спектрометрических пиков порождало сомнения в возможности внедрения результатов в практическую медицину.

Анализ идентифицированных с использованием технологии SELDI-TOF биомаркеров рака яичника показывает, что все они являются белками, присутствующими в сыворотке в достаточно высоких (микромольных) концентрациях, а изменение их уровня обусловлено системным ответом организма на сопутствующее опухоли воспаление. Особое место среди найденных биомаркеров занимает идентифицированный в лаборатории диагностической протеомики научно-исследовательского института биомедицинской химии РАМН в 2004 году биомаркер рака яичника сывороточный амилоид А острой фазы (A-SAA - acute-phase serum amyloid A). Несмотря на то, что A-SAA давно известен как маркер воспаления, он обладает рядом уникальных и важных для потенциального биомаркера рака свойств.

Во-первых, его концентрация в сыворотке крови в норме в 10-100 раз ниже, чем у других кандидатных маркеров. Во-вторых, при воспалении концентрация A-SAA увеличивается в 100 и более раз, тогда как концентрация других маркеров, идентифицированных с использованием технологии SELDI-TOF, меняется сравнительно слабо. В-третьих, A-SAA синтезируется непосредственно в тканях некоторых злокачественных опухолей [Gutfeld *et al* (2006) *J. Histochem. Cytochem.*, **54**, 63-73; Kovacevic *et al* (2006) *FEBS Lett.*, **580**, 161-167]. И наконец, на основании последних данных о молекулярных эффектах и взаимодействиях A-SAA, повышение его концентрации при злокачественных опухолях может играть активную роль в развитии последних. Так, показана способность A-SAA индуцировать экспрессию транскрипционного фактора NFκB [He *et al* (2003) *Blood*, **101**, 1572-1581], играющего ведущую роль в блокировании апоптоза, и матриксных металлопротеиназ MMP1, MMP3 и MMP9 [O'Hara *et al* (2004) *Arthritis Rheum.* **50**, 1788-1799; Lee *et al* (2005) *BBRC*, **330**, 989-998], вызывающих деградацию межклеточного матрикса и тем самым способствующих ангиогенезу и

метастазированию. Способность А-SAA стимулировать пролиферацию клеток, ингибировать апоптоз, а также стимулировать ангиогенез была напрямую продемонстрирована на синовиоцитах больных ревматоидным артритом [*Lee et al*, (2006) *J. Immunol.*, Oct 15;177(8), p. 5585-5594].

Перспективным подходом к разработке нового метода диагностики рака яичника может стать объединение данных масс-спектрометрических методов с данными иммуноферментного анализа о концентрациях классического биомаркера рака яичника СА125 и белка А-SAA.

Цель и задачи исследования

Целью данной работы является разработка экспериментальной системы диагностики рака яичника путем комбинирования данных иммуноферментного анализа о концентрациях СА125 и А-SAA и протеомных профилей сыворотки крови с применением различных методов статистической обработки результатов.

В соответствии с указанной целью были поставлены следующие задачи:

1. Измерить методом иммуноферментного анализа концентрации СА125 и А-SAA в выборке сывороток больных раком яичника на разных стадиях, больных доброкачественными гинекологическими опухолями и здоровых лиц.

2. Провести масс-спектрометрическое профилирование (SELDI-TOF) всех сывороток в условиях, оптимизированных для измерения уровня А-SAA, и определить чувствительность масс-спектрометрической детекции А-SAA в сыворотке.

3. Провести статистический анализ полученных данных, разработать диагностический алгоритм и определить его точность, чувствительность и специфичность.

Научная новизна работы

Впервые определена чувствительность масс-спектрометрии SELDI-TOF для детекции белка А-SAA в составе сыворотки крови. Ранее чувствительность метода при работе с такой сложной смесью, как сыворотка крови, была неизвестна.

С помощью кластерного анализа определена природа большей части дискриминаторных пиков на масс-спектрах.

Создана экспериментальная диагностическая система на основе комбинации масс-спектрометрических данных с данными о концентрации белка А-SAA и биомаркера рака яичника СА125.

К масс-спектрометрическим данным впервые применен статистический метод пар с наибольшим счетом.

Практическая значимость работы

Предложена новая экспериментальная диагностическая система для рака яичника, разработанная с применением современных статистических методов на основе комбинации данных о концентрации классического биомаркера рака яичника СА125, концентрации сывороточного амилоида А острой фазы и данных масс-спектрометрии с точностью диагностики 95,2%. Путем анализа выборки сывороток больных и здоровых людей российского населения показана применимость масс-спектрометрического метода для диагностики рака яичника.

Положения диссертации, выносимые на защиту

1. Определение чувствительности масс-спектрометрии SELDI-TOF для детекции белка сывороточного амилоида А острой фазы в составе сыворотки крови.

2. Разработка алгоритмов распознавания сывороток больных раком яичника и сывороток женщин, не страдающих раком яичника с помощью статистических методов логистической регрессии, опорных векторов и пар с наибольшим счетом при различных комбинациях исходных данных.

Апробация работы

Основные положения работы были представлены на следующих конференциях:

1. 4-ый Ежегодный Всемирный Конгресс международной организации «Протеом человека» (HUPO 4-th Annual World Congress), Мюнхен, Германия, 28 августа-1 сентября 2005 года;

2. Международная школа-конференция молодых ученых «Системная биология и биоинженерия», (Звенигород, 28 ноября - 2 декабря 2005 г.);

3. 3-я Международная Конференция «Геномика, Протеомика, Биоинформатика и Нанотехнологии для Медицины» (3rd International Conference “Genomics, Proteomics, Bioinformatics and Nanotechnologies for Medicine”), Новосибирск, 12-16 июля 2006 года;

4. 5-ый Ежегодный Всемирный Конгресс международной организации «Протеом человека» (HUPO 5th Annual World Congress), Лонг Бич, Калифорния, 28 октября-1 ноября 2006 года.

Публикации

Материалы диссертации опубликованы в 4 статьях и 4 публикациях в сборниках докладов научных конференций

Структура и объем работы

Диссертация выполнена на 102 страницах, включает в себя введение, литературный обзор, результаты и их обсуждение, заключение, выводы и список литературы. Диссертация содержит 9 таблиц и 11 рисунков. Список литературы состоит из 134 наименований.

МАТЕРИАЛЫ И МЕТОДЫ

Объекты исследования

В работе анализировали сыворотки 34 женщин, больных раком яичника, в том числе 7 сывороток пациенток с раком в ранней (1 и 2-ой) стадии; 13 сывороток женщин больных доброкачественными опухолями яичника и 16 – миомой матки; одну сыворотку больной миомой матки и фибромой яичника и 26 сывороток здоровых женщин. Образцы сывороток предоставлены кафедрой акушерства и гинекологии Российского Государственного Медицинского Университета.

Иммуноферментный анализ

Для определения концентрации А-SAA использовали набор для иммуноферментного анализа Human SAA (Biosource, США). Для определения концентрации СА125 использовали набор для иммуноферментного анализа СА125 EIA (CanAg, Канада). Анализ проводили в соответствии с протоколами производителей.

Профилирование сывороток с применением SELDI-TOF

Для масс-спектрометрического профилирования использовали нормальнофазовые чипы NP20 (Ciphergen, США). На пятна белковых чипов наносили по 1 мкл сыворотки, разведенной в 10 раз деионизованной водой, высушивали на воздухе и дважды промывали деионизованной водой. Затем пятна снова высушивали и наносили раствор матрицы дважды по 0,5 мкл. В качестве матрицы использовали насыщенный раствор α -циано-4-гидроксикоричной кислоты (Ciphergen, США) в 50% (об./об.) ацетонитриле, содержащем 0,5% (об./об.) трифторуксусной кислоты (ТФУ), разведенный в два раза тем же растворителем. Профилирование проводили на масс-спектрометре SELDI-TOF Protein Biology System II (PBS II) (Ciphergen, США). Спектры снимали в автоматическом режиме в диапазоне масс 7000-70000 Да при интенсивности лазера 230, чувствительности детектора 9 и 90 импульсов лазера на пятно. Для калибровки спектров использовали следующие внешние стандарты: инсулин (5435 Да), убиквитин (8565 Да), цитохром С лошади (12361 Да), миоглобин спермы кита (16952 Да) и бычий сывороточный альбумин (66431 Да). Спектры для каждой сыворотки были сняты, по крайней мере, в двух повторностях для получения воспроизводимых результатов.

Анализ масс-спектров

Были проанализированы пики масс-спектров в пределах от 5500 до 17500 Да. Для анализа была использована программа Biomarker Wizard™ (Ciphergen Biosystems) со следующими настройками: соотношение сигнал/шум (первая ступень) 10, соотношение сигнал/шум (вторая ступень) 5, пороговая

интенсивность пика 0%, допустимая ошибка в определении массы 0,2%. Массы пиков и их интенсивности экспортировали в таблицы MS Excel, и значения интенсивностей в повторных измерениях усредняли.

Статистическая обработка результатов

В качестве исходных данных были использованы значения интенсивностей 48 пиков, идентифицированных с помощью программы Biomarker Wizard™, и данные иммуноферментного анализа о концентрациях А-SAA и СА125. Эти данные были обработаны как по отдельности, так и совместно.

Метод опорных векторов и рекурсивный отбор признаков

Метод опорных векторов (support vector machine, SVM) представляет собой метод распознавания образов с учителем, нашедший широкое применение в статистической обработке биомедицинских данных.

Целью метода опорных векторов является поиск гиперплоскости с максимальной шириной границы, разделяющей обучающую выборку на два класса. В исследовании стояла задача наилучшего разделения множества исследованных сывороток на «рак» и «отсутствие рака». Перед применением SVM был проведен отбор признаков для улучшения параметров модели. Для отбора значимых признаков применили алгоритм рекурсивного исключения признаков (Recursive Feature Elimination algorithm, RFE, [Guyon, I. et al (2002) Machine Learning, **46**, 389-422]). Сначала алгоритм применяли на всех данных, затем переменную, получившую наименьший вес, исключали, и эту операцию повторяли, пока все переменные не были ранжированы в соответствии с порядком их исключения. Затем отбирали тот набор признаков, который позволял добиться наилучшей точности классификации на всей исследуемой выборке (точность определяли с помощью 10-кратной перекрестной проверки достоверности, см. ниже).

Логистическая регрессия и информационный критерий Акаике

В качестве альтернативы SVM для разработки диагностического алгоритма применяли метод логистической регрессии (logistic regression, LR), который

позволяет оценивать вероятность того, что исследуемый образец относится к раку. Как и в методе опорных векторов, для улучшения модели предварительно проводили отбор признаков. Для отбора признаков использовали ступенчатую модель отбора, основанную на информационном критерии Акаике. Применение информационного критерия Акаике позволяет достигнуть баланса между сложностью модели (количеством входящих в уравнение регрессии переменных) и ее эффективностью для классификации.

Определение эффективности диагностического алгоритма

Для проверки чувствительности и специфичности разработанных диагностических моделей, использовали 10-кратную перекрестную проверку достоверности. Все имеющиеся образцы разделили на 10 групп, 9 групп использовали в качестве обучающей выборки для разработки диагностического алгоритма, а оставшуюся одну группу использовали в качестве тестовой выборки для его проверки. 10-кратную перекрестную проверку достоверности проводили по 100 раз для диагностических алгоритмов, полученных как методом SVM, так и методом LR. Для каждого диагностического алгоритма были вычислены точность, чувствительность и специфичность с доверительными интервалами.

Метод пар с наибольшим счетом (top scoring pairs, TSP)

Сравнительно новым, простым и эффективным подходом к классификации является метод пар с наибольшим счетом TSP [Tan et al (2005) *Bioinformatics*, 21, 3896-3904]. Этот метод основан на относительных, а не абсолютных интенсивностях пиков и позволяет получать алгоритмы классификации, которые легко интерпретировать.

Реальные значения интенсивностей пиков внутри каждого профиля заменяют рангами. Затем выявляют такие пары пиков, у которых ранги в наибольшей степени различаются между исследуемыми классами. Для обработки данных с помощью TSP и проверки эффективности модели путем перекрестной проверки достоверности с исключением по одной пробе использовали программу,

созданную разработчиками метода TSP и доступную на их сайте (<http://www.bme.jhu.edu/~actan/KTSP/>), с неизменными настройками.

Кластерный анализ

Для кластеризации масс-спектрометрических данных использовали метод средней связи.

Программное обеспечение

Всю статистическую обработку, включая применение метода логистической регрессии, метода опорных векторов, кластерного анализа, а также отбор признаков проводили с помощью находящегося в открытом доступе R-языка (www.r-project.org).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Определение концентрации белка A-SAA в сыворотках методом иммуноферментного анализа

Концентрация A-SAA была измерена методом иммуноферментного анализа во всех сыворотках. Во многих сыворотках больных раком яичника уровень A-SAA был повышен вплоть до 3-4 г/л (3×10^{-4} М), тогда как в норме концентрация A-SAA составляет 0,010-0,030 г/л ($8,5 \times 10^{-7}$ - $2,5 \times 10^{-6}$ М). Концентрации A-SAA для каждой исследованной сыворотки приведены на рисунке 1.

Надо отметить, что при раке наблюдается довольно резкое увеличение концентрации A-SAA, во многих случаях на два - три порядка. Повышение концентрации A-SAA в ряде случаев не связано со стадией рака, так как из семи сывороток пациенток с раком яичника в ранней стадии у четырех уровень A-SAA чрезвычайно высокий (рис. 1). В двух из 14 сывороток пациенток с доброкачественными опухолями яичника уровень A-SAA также достигает 3 - 4 г/л (около 3×10^{-4} М). Оба случая соответствуют серозным цистаденомам. Не исключено, что настолько повышенная концентрация A-SAA может предшествовать малигнизации опухоли, но эта гипотеза нуждается в подтверждении.

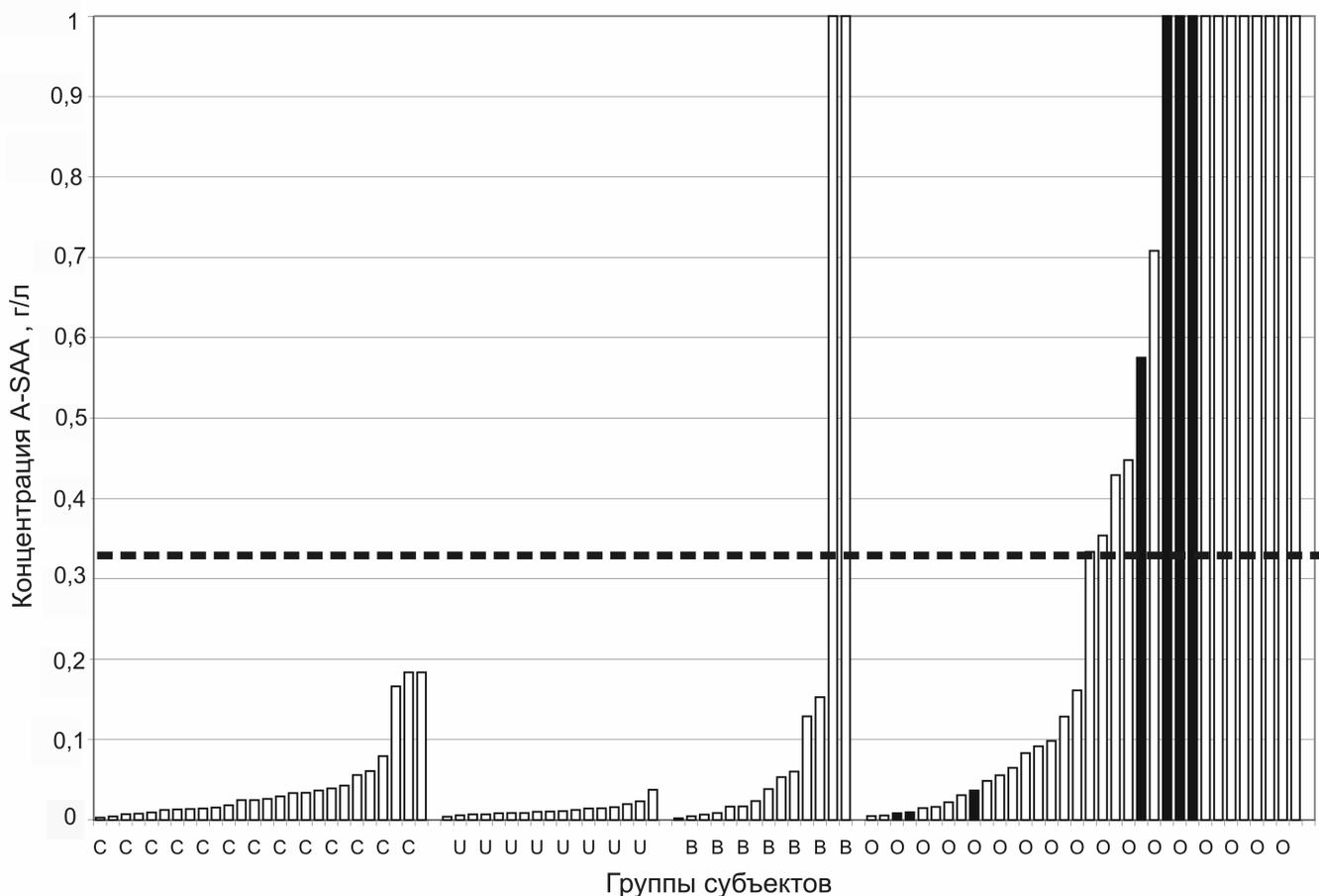


Рисунок 1. Концентрации А-SAA в исследованных сыворотках. Буквами по горизонтальной оси обозначены группы субъектов: О - рак яичника; В - доброкачественная опухоль яичника; U - миома матки; С - здоровые женщины. Конкретные значения концентрации выше 1 г/л не показаны. Пунктирной линией показано примерное критическое значение А-SAA для диагностики рака яичника, которое совпадает с минимальной концентрацией А-SAA, детектируемой с помощью SELDI-TOF. Столбики, соответствующие ранней стадии рака, выделены черным цветом.

В исследование были включены также пациентки с миомой матки. Известно, что у пациенток с миомой матки концентрация СА125 часто превышает критический уровень, равный 35-40 единиц, что заставляет заподозрить наличие молчащей злокачественной опухоли. В то же время, концентрация А-SAA во всех исследованных случаях миомы очень низкая.

А-SAA как биомаркер рака яичника при критическом значении концентрации 0,3 г/л ($2,6 \times 10^{-5} \text{M}$) обладает чувствительностью 50% (17 определенных случаев рака из 34) и специфичностью 96,4% (54/56 сывороток

женщин, не больных раком). Отсюда ясно, что А-SAA не может быть использован в качестве индивидуального биомаркера рака яичника. Однако его использование в комбинации с СА125 и/или масс-спектрометрическими данными может привести к улучшению качества диагностики.

Определение чувствительности детекции А-SAA с помощью SELDI-TOF

Для наилучшей детекции пика, соответствующего А-SAA, был проведен подбор условий профилирования с использованием сывороток крови с известными концентрациями А-SAA. Были испробованы различные типы чипов (анионообменный, катионообменный, обращеннофазовый). Лучшие результаты были получены при применении нормальнофазовых чипов NP20 и наиболее простой и быстрой процедуры профилирования. В итоге, было проведено профилирование всех сывороток на NP20 чипах с помощью протокола снятия спектров, сфокусированного на массах 10-12 кДа.

Все пики в спектрах автоматически детектировали с помощью программы Ciphergen Protein с базовыми настройками. При этом пик массой 11,68 кДа был детектирован во всех сыворотках с концентрацией А-SAA 0,333 г/л и более (по данным иммуноферментного анализа). Другими словами, все 19 сывороток с концентрацией А-SAA выше 0,3 г/л имеют хорошо выраженный пик в интересующем нас интервале (рисунок 2).

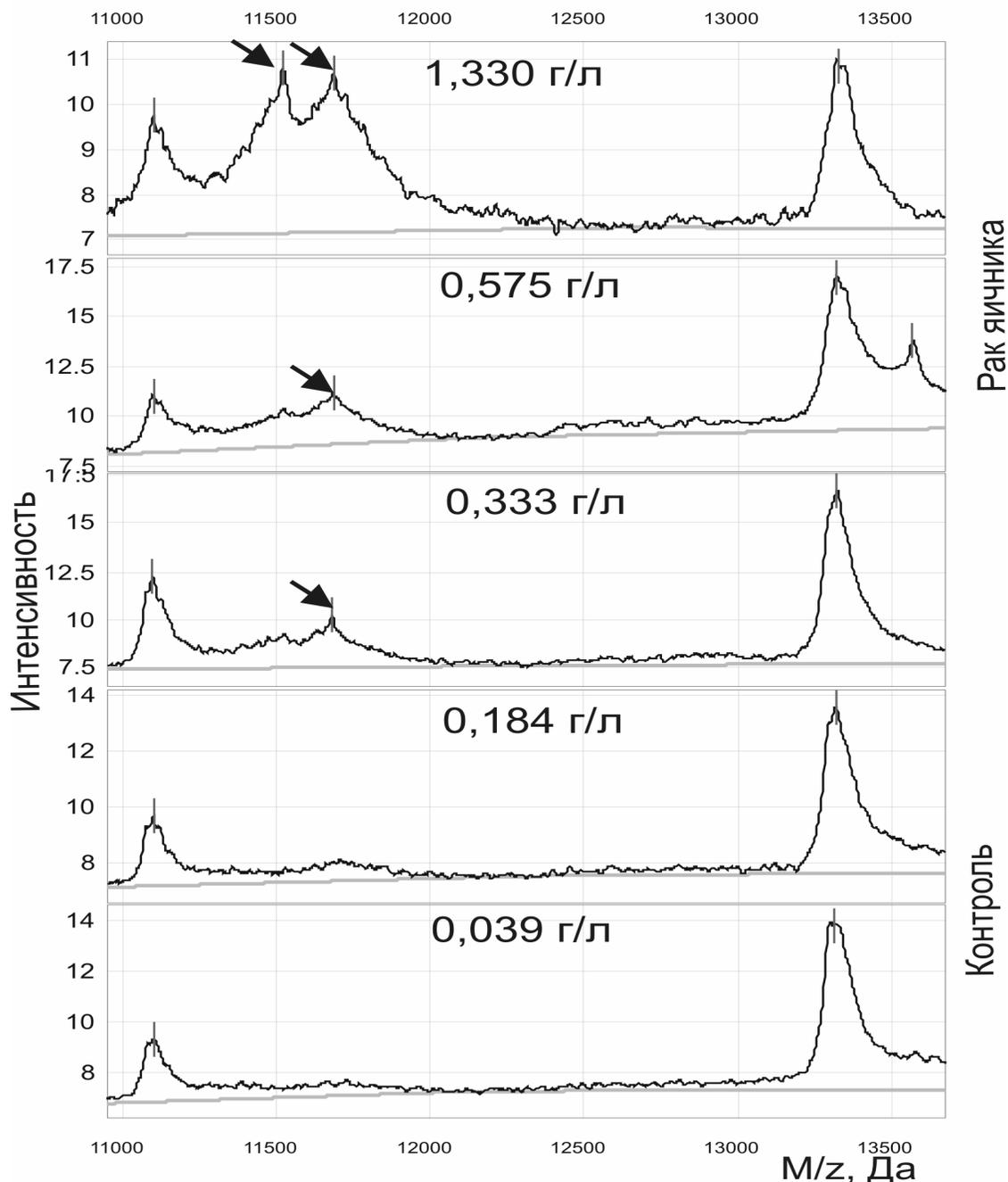


Рисунок 2. Чувствительность определения А-SAA в сыворотке с помощью масс-спектрометрии SELDI-TOF. Для каждого спектра показаны значения концентрации А-SAA, измеренные методом иммуноферментного анализа. Стрелками показаны пики А-SAA на спектрах.

Среди исследованных образцов не было таких, концентрация А-SAA в которых находилась бы в пределах от 0,184 до 0,333 г/л. На основании имеющихся данных можно утверждать, что чувствительность масс-спектрометрического определения А-SAA в сыворотке составляет 0,2-0,3 г/л ($1,7-2,6 \times 10^{-5}$ М). Истинная

молярная чувствительность, по крайней мере, вдвое ниже, так как А-SAA присутствует в сыворотке как минимум в двух основных формах: полноразмерной, массой 11683 Да, и с отщепленным аргинином на N-конце, массой 11525 Да. Масс-спектрометрическую детекцию А-SAA нельзя считать количественной, так как интенсивность пиков, соответствующих близким концентрациям А-SAA, в разных спектрах была различной; кроме того, при концентрациях выше 1 г/л наступало насыщение (разница в интенсивности сигнала между концентрациями 1 и 3 г/л очень слабая). Однако масс-спектрометрия SELDI-TOF позволяет детектировать А-SAA в сыворотке только при концентрациях 0,3 г/л и выше. Важно подчеркнуть, что критический уровень А-SAA, детектируемый с помощью масс-спектрометра SELDI-TOF, очень близок к критическому уровню этого белка для ряда воспалительных заболеваний [*Casl et al*, Eur. J. Clin. Chem. Clin. Biochem., 1996 Jan; 34(1), p. 31-35], и поэтому прямое масс-спектрометрическое профилирование может быть использовано для быстрого определения повышенного уровня А-SAA в сыворотке.

Обработка масс-спектров SELDI-TOF, полученных на чипах с нормальнофазовой поверхностью

На масс-спектрах было детектировано 48 пиков в интервале m/z 5500-17500 Да, каждый из которых имеет высокую интенсивность (соотношение сигнал/шум > 5) по крайней мере в одной исследованной пробе. Список интенсивностей полученных пиков для всех сывороток, разделенных на два класса (рак и отсутствие рака), был использован в качестве исходных данных для разработки диагностических алгоритмов с использованием различных статистических методов.

Классификация спектров SELDI-TOF с использованием метода пар с наибольшим счетом (TSP)

Классификатор TSP был предложен для обработки данных ДНК-микрочипов и идентификации генов-маркеров [*Xu et al* (2005) *Bioinformatics*, **21**, 3905-3911]. Этот алгоритм классифицирует образцы на основе разностей интенсивностей

пиков, а не абсолютных значений интенсивностей, отбирая одну или несколько (k-TSP) пар биомаркеров с наибольшей разницей интенсивностей. Так как масс-спектрометрические данные в целом аналогичны данным ДНК-микрочипов, мы применили TSP-классификацию к спектрам исследуемой выборки. Применение алгоритма позволило выявить только одну пару с наибольшим счетом, состоящую из интенсивностей пиков 11681 Да и 13769 Да. В случае, если разность интенсивностей указанных пиков положительна, исследуемый образец классифицируют как рак, если отрицательна, то как отсутствие рака. Несмотря на то, что процент правильной классификации с применением TSP (точность метода классификации) составляет всего 79,9% даже при применении перекрестной проверки достоверности с исключением по одному образцу, стоит отметить, что биологический смысл полученных результатов легко интерпретировать. Пик с m/z 11681 Да соответствует основной форме A-SAA [*Moshkovskii et al* (2005) *Proteomics*, **5**, 3790-3797], концентрация которого при раке яичника повышается, а пик с m/z около 13769 Да соответствует транстиретину (TTR), концентрация которого понижается при раке яичника [*Gericke et al* (2005) *BMC Cancer*, **5**, 133].

Таким образом, TSP-классификатор заслуживает внимания как один из способов обработки данных протеомных исследований с целью выявления биомаркеров. Пара биомаркеров, отобранная в данном исследовании (A-SAA-TTR), представляет собой новую бинарную переменную (разность интенсивностей, принимающую положительные или отрицательные значения), которую можно комбинировать с данными иммуноферментного анализа для улучшения классификации.

Метод опорных векторов и метод логистической регрессии в классификации на основе комбинированных протеомных данных

Различные типы данных, полученных в данном исследовании, были использованы для разработки диагностического алгоритма с применением методов SVM и LR, чтобы определить, какой набор данных оптимален для распознавания рака. Диагностический алгоритм разрабатывали на основе (1)

данных ИФА о концентрации стандартного биомаркера рака яичника CA125; (2) данных ИФА о концентрациях CA125 и A-SAA; (3) данных масс-спектрометрии, то есть значений интенсивностей 48 SELDI-MS m/z пиков; (4) комбинированных данных ИФА о концентрациях CA125 и A-SAA и масс-спектрометрических данных и (5) данных ИФА о концентрациях CA125 и A-SAA в комбинации с бинарной переменной, полученной после применения TSP анализа и представляющую собой формализованное представление данных масс-спектрометрии. В каждом случае эффективность классификации проверяли с применением перекрестной проверки достоверности. Точность, чувствительность и специфичность полученных классификаций приведены в таблице 1.

Надо отметить, что добавление данных иммуноферментного анализа о концентрации A-SAA к данным о концентрации CA125 не приводит к улучшению диагностики, несмотря на то, что в нескольких исследованных случаях при раке уровень A-SAA был существенно повышен на фоне нормального уровня CA125. При работе с данными масс-спектрометрии, метод SVM дал лучшие результаты, чем LR, и точность диагностического алгоритма, разработанного методом SVM для масс-спектрометрических данных была выше, чем при использовании только концентрации CA125 (89,5% вместо 86,1%). LR-классификатор, напротив, дал существенно лучшие результаты, чем SVM при обработке бинарных данных TSP в комбинации с данными ИФА, причем данные TSP существенно повысили точность классификации (90,7% по сравнению с 85-86% для одного CA125). Характерные свойства переменной TSP нуждаются в дальнейшем исследовании, так как она может быть более воспроизводима между различными масс-спектрометрами или исследуемыми выборками, чем абсолютные значения интенсивностей масс-спектрометрических пиков.

Наконец, наилучшей точности классификации (95,2%) рака и отсутствия рака удалось достичь при использовании SVM для обработки комбинированных данных ИФА и масс-спектрометрии.

Таблица 1. Результаты применения метода опорных векторов и логистической регрессии для разработки диагностического алгоритма. Чувствительность и специфичность определяли с использованием 10-кратной перекрестной проверки достоверности

Метод	Точность, %	Специфичность, %	Чувствительность, %
SVM (CA125)^а	86,2±0,7^с	98,8±0,3	64,7±1,7
SVM (ИФА)^б	86,4±0,7	96,5±0,5	70,3±1,8
SVM RFE (MS)^в	89,5±0,7	93,3±0,7	83,6±1,6
SVM (TSP(MS)+ИФА)^г	86,7±0,7	92,8±0,8	77,5±1,6
SVM RFE (ИФА+MS)^д	95,2±0,4	98,1±0,4	90,8±1,1
LR (CA125)	85,1±0,7	95,6±0,5	67,5±1,7
LR (ИФА)	86,6±0,7	94,3±0,6	74,2±1,7
LR (MS)	86,0±0,7	87,5±1,3	83,7±1,5
LR AIC (TSP(MS)+ИФА)	90,7±0,6	96,9±0,5	81,2±1,4
LR AIC (ИФА+MS)	91,9±0,6	92,7±0,7	90,7±1,1

^а В качестве единственной переменной применяли уровень CA125.

^б В качестве переменных использовали концентрации CA125 и A-SAA, измеренные методом иммуноферментного анализа.

^в В качестве переменных использовали 48 значений интенсивностей пиков масс-спектров SELDI-TOF.

^г Два значения концентраций, измеренных с помощью ИФА, совместили с бинарной переменной, полученной путем применения TSP к масс-спектрометрическим данным.

^д 48 переменных спектров SELDI-TOF комбинировали с двумя переменными ИФА.

^с Доверительный интервал вычисляли для доверительной вероятности $\gamma = 0,95$.

Аннотация компонентов масс-спектров

Несмотря на то, что к настоящему моменту идентифицировано большое количество биомаркеров, обнаруженных с помощью SELDI-TOF, систематических исследований по аннотации масс-спектрометрических профилей не проводилось. Мы предприняли попытку сопоставления наблюдаемых масс-спектрометрических пиков с известными компонентами сыворотки крови. Для этого была проанализирована литература, в которой описаны идентифицированные с помощью SELDI-TOF биомаркеры, а также проведен кластерный анализ данных с целью выявления групп взаимосвязанных белков. Среди пиков, детектированных нами на масс-спектрах, предположительно присутствуют несколько форм А-SAA и транстретина, аполипопротеин С1, фрагменты комплемента С3а и С4а, альфа- и бета-субъединицы гемоглобина.

Дискриминаторные пики, отобранные различными классификаторами

В таблице 2 приведен список дискриминаторных пиков, отобранных разными статистическими методами, часть из которых аннотирована. Из таблицы видно, что основные дискриминаторные пики относятся к А-SAA, транстретину и их модификациям. Возможно, именно данные о соотношении различных модификаций биомаркеров позволили существенно повысить эффективность диагностики.

Таблица 2. Дискриминаторные пики, отобранные методом опорных векторов, логистической регрессии и пар с наибольшим счетом при применении этих методов на масс-спектрометрических данных и комбинации масс-спектрометрических данных с данными ИФА

#	m/z, Да	Название белка	SVM RFE (MS)	SVM RFE (MS+ИФА)	LR (MS)	LR (MS+ИФА)
1	5675		+	+	+	
2	6441	Аполипопротеин С1 с отщепленными N- концевыми треонином и пролином	+			

#	m/z, Да	Название белка	SVM RFE (MS)	SVM RFE (MS+ИФА)	LR (MS)	LR (MS+ИФА)
3	6454		+	+	+	
4	7566		+			+
5	7651				+	
6	7769			+		
7	8208			+		
8	8766				+	
9	8829				+	
10	10265		+	+	+	+
11	11304		+			
12	11370				+	+
13	11649	Сывороточный амилоид А 2 бета	+			
14	11681	Сывороточный амилоид А1 альфа	+		+	
15	11728	Сывороточный амилоид А1 бета	+	+		
16	12168		+	+	+	+
17	13769	Транстиретин		+	+	
18	13870	Цистеинилированный транстиретин	+	+	+	+
19	14685		+	+		
20	15310		+	+		
21	17017		+	+		

ЗАКЛЮЧЕНИЕ

В данной работе предложена новая диагностическая система, разработанная с применением современных статистических методов на основе комбинации данных о концентрации классического биомаркера рака яичника СА125, концентрации сывороточного амилоида А и данных масс-спектрометрии SELDI-TOF. Если вклад концентрации сывороточного амилоида А, измеренной методом иммуноферментного анализа, в эффективность диагностики оказался несущественным, то совмещение масс-спектрометрических данных с данными о концентрации СА125 позволило значительно повысить качество диагностики,

достигнув точности диагностики 95,2%. Полученные результаты могут служить предпосылками для внедрения в перспективе в медицинскую практику нового способа диагностики рака яичников.

ВЫВОДЫ

1. Предел чувствительности масс-спектрометрической детекции белка А-SAA в сыворотке крови составляет 0,2-0,3 г/л ($1,7-2,6 \times 10^{-5}$ М).

2. На выборке из российского населения показана применимость метода масс-спектрометрии SELDI-TOF для классификации сывороток больных раком яичника, доброкачественными гинекологическими опухолями и здоровых женщин. Наиболее высокая точность диагностики была получена при применении к масс-спектрометрическим данным метода опорных векторов и составила 89,5 %.

3. Добавление к масс-спектрометрическим данным данных иммуноферментного анализа о концентрации СА125 и А-SAA привело к повышению точности диагностики до 95%.

4. На основе результатов кластерного анализа масс-спектрометрических данных и анализа литературных данных часть пиков на масс-спектре аннотирована. Показано, что несколько дискриминаторных масс-спектрометрических пиков соответствуют сывороточному амилоиду А острой фазы, транстиретину и их модификациям.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. S. Moshkovskii, M. Vlasova, M. Safarova, O. Makarov, and A. Archakov “Serum amyloid A as ovarian cancer biomarker” // Abstr. HUPO 4th Annual World Congress, (August 28-September 1, 2005, Munich). – Molecular & Cellular Proteomics. – August 2005.– Vol. 4, Number 8 (Suppl. 1);
2. М.А. Власова, С.А. Мошковский, М.Р. Сафарова, О.В. Макаров, А.И.Арчаков “Молекулярная диагностика рака яичника с использованием протеомных технологий” // Биомедицинская химия. – 2005. – т. 51 № 4. – с. 367-383;
3. М. А. Власова, С. А. Мошковский, А. И. Арчаков «Комбинирование данных масс-спектров SELDI и концентрации СА125 для диагностики рака яичника» // Материалы международной школы-конференции молодых ученых «Системная биология и биоинженерия» (28 ноября - 2 декабря 2005 г., Москва). – МАКС Пресс. – 2005. – с. 18;
4. Е.И. Гоуфман, С.А. Мошковский, О.В.Тихонова, П.Г.Лохов, В.Г. Згода, М.В. Серебрякова, И.Ю. Торопыгин, М.А. Власова, М.Р.Сафарова, О.В.Макаров, А.И. Арчаков Протеомное исследование термостабильной фракции сыворотки пациентов с различными опухолями с применением двумерного электрофореза. // Биохимия. – 2006. – т. 71 №4. – с. 354-360;
5. М.А. Власова, С.А. Мошковский. Молекулярные взаимодействия сывороточного амилоида А острой фазы: возможная связь со злокачественными опухолями. // Биохимия. – 2006. – т. 71 №10 – с. 1051-1059;
6. S.A. Moshkovskii, M. A. Vlasova, M.A. Pyatnitskiy, A.I. Archakov Acute phase serum amyloid A in ovarian cancer as an important component of proteome diagnostic profile. // Abstr. 3rd International Conference “Genomics, Proteomics, Bioinformatics and Nanotechnologies for Medicine” (July 12-16, 2006, Novosibirsk, Russia) – p. 64;
7. Moshkovskii S.A., Vlasova M.A., Pyatnitskiy M.A., Archakov A.I. Serum amyloid A and tranthyretin forms constitute discriminatory SELDI profile for ovarian cancer. //

Abstr. HUPO 5th Annual World Congress (October 28-November 1, 2006, Long Beach, California). – Molecular & Cellular Proteomics. – 2006. – Vol. 5, Number 10 (Suppl.);
8. Moshkovskii S.A., Vlasova M.A., Pyatnitskiy M.A., Tikhonova O.V., Safarova M.R., Makarov O.V., Archakov A.I. Acute phase serum amyloid A in ovarian cancer as an important component of proteome diagnostic profiling. // Proteomics. Clinical Applications. – 2007. – Vol. 1 (1). – p. 107-117.