

(19)



**Евразийское
патентное
ведомство**

(11) **041076**(13) **B1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2022.09.08

(21) Номер заявки
201591742

(22) Дата подачи заявки
2014.03.13

(51) Int. Cl. **C12Q 1/00** (2006.01)
G01N 33/50 (2006.01)
G01N 31/00 (2006.01)

(54) СПОСОБ УЛУЧШЕНИЯ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ С ИСПОЛЬЗОВАНИЕМ ИЗМЕРЯЕМЫХ АНАЛИТОВ

(31) **61/851,867**

(32) **2013.03.14**

(33) **US**

(43) **2016.09.30**

(86) **PCT/US2014/000041**

(87) **WO 2014/158287 2014.10.02**

(71)(73) Заявитель и патентовладелец:
ОТРЕЙСИЗ ИНК. (US)

(72) Изобретатель:
**Красик Галина, Марефат Мохсен,
Лингенфелтер Кит (US)**

(74) Представитель:
Медведев В.Н. (RU)

(56) SONG Q. et al. Identification Of Serum Biomarkers For Lung Cancer Using Magnetic Bead-Based SELDI-TOF-MS. Acta Pharmacologica Sinica. 24 October 2011. Vol. 32; pages 1537-1542; Abstract: Methods, Results, Conclusion; page 1538, right column, fourth paragraph; page 1540, left column, first paragraph; page 1540, left column, second paragraph; Table 1; Tables 2, 3 DOI. 10.1038/aps.2011.137.

MALIK I. et al. Soluble Adhesion Molecules And Prediction Of Coronary Heart Disease A Prospective Study And Meta-Analysis The Lancet 22 September 2001 Vol 358; pages 971-975.

(57) Изобретение относится к способам повышения точности диагностики заболеваний и связанным диагностическим тестам, включающим в себя корреляцию измеряемых аналитов с исходами с двумя состояниями. Согласно одному аспекту изобретения предусмотрен способ диагностики рака, содержащий этапы: а) определения концентраций predetermined аналитов в слепой пробе субъекта, при этом predetermined аналиты выбирают из группы, содержащей IL-6, IL-8, VEGF, TNF α и PCA; б) выбора метапеременной, ассоциированной с субъектом, которая меняется в популяции, связанной с субъектом, для членов упомянутой популяции, которые, как известно, либо болеют, либо не болеют раком, при этом метапеременной является возраст; в) преобразования концентраций аналитов в зависимости от характеристик распределения в популяции и метапеременной для вычисления псевдоконцентрации, которая представляет каждый аналит, при этом упомянутое преобразование включает в себя коррекцию пространственного смещения; d) сравнения псевдоконцентраций с моделью обучающей последовательности псевдоконцентраций посредством использования линейного показателя ("CSI") и q-показателя ("CSq"), определяемых для членов популяции, которые, как известно, либо болеют, либо не болеют раком; e) определения, болен ли субъект раком или нет на основе того, попадают ли показатели CSI и CSq псевдоконцентраций субъекта в группу с наличием рака или в группу без рака на графике CSI/CSq модели обучающей последовательности. Согласно другому аспекту изобретения предусмотрен способ подготовки упомянутой модели обучающей последовательности, способной диагностировать рак.

B1**041076****041076****B1**

Область техники

Изобретение относится к способам повышения точности диагностики заболеваний и связанным диагностическим тестам, включающим в себя корреляцию измеряемых аналитов с исходами с двумя состояниями.

Уровень техники

Корреляционные методы, где три или более независимых переменных используются для корреляции с некоторым исходом с двумя состояниями (например, наличие или отсутствие некоторого определенного заболевания), обычно используются с методом поиска кластера или окрестности, регрессионным методом и вейвлет-методами. В случаях прогнозирования заболеваний, общие составляющие крови или сыворотки измеряются, и корреляция пытается использовать их концентрации в качестве независимых переменных для прогнозов различных состояний болезни. В случае некоторого определенного состояния заболевания, когда исходом является либо "наличие заболевания", либо "отсутствие заболевания", обычно используется метод логистической регрессии. Другие методы включают, например, генетические алгоритмы. Прогностическая сила этих методов сильно зависит от составляющих аналитов, выбранных для метода. Специалисты в данной области поймут, что многие аналиты и параметры, которые, казалось бы, могут улучшить прогностическую силу, на самом деле не улучшают диагностическую и аналитическую силу.

Регрессионный метод использует тенденции в независимых переменных для корреляции с исходами. Линейный метод основан на линейных тенденциях, а логистическая регрессия основана на логарифмических тенденциях. При прогнозировании биологического заболевания обычно используется логистическая регрессия.

Метод групповой кластеризации обследует переменную корреляционную топологию на предмет группировки аналогичных исходов.

Метод кластеризации имеет преимущество в том, что он может найти корреляции, где тенденции являются не смежными, а имеют локальные топологические обращения. Этот метод, хоть и является весьма нелинейным и восприимчивым к локальным крайне изменчивым исходам с малыми ошибками измерений, может обладать большей прогностической силой при использовании в биологии. Кроме того, оба метода обычно могут быть объединены с кластерным методом, применяемом на небольшом масштабе в общем регрессионном методе.

Тем не менее, некоторые независимые переменные, которые, казалось бы, должны обладать некоторой корреляцией, на практике не демонстрируют прогностической тенденции. Таким образом, необходим подход, который улучшает диагностическую точность за счет использования специфических переменных пациента и специфических переменных популяции, которые до сих пор не внесли полезную информацию в диагностику состояний заболевания.

Было проведено значительное исследование, чтобы найти биомаркеры, которые по отдельности или в сочетании могут прогнозировать состояния заболевания с воспроизводимостью и прогностической силой, достаточными для клинического применения. Это исследование имело ограниченный успех или не имело его вовсе. Высокораспространенные белки (НАР) были в значительной степени исследованы, чтобы найти некоторый единственный белок, который мог бы быть использован для такого прогноза. Многочисленные примеры были найдены, но никакой из них не имел достаточно низких уровней ложноположительных результатов, чтобы использоваться для скрининга пациентов с помощью такого маркера. В результате, такие отдельные биомаркеры используются только для контроля за терапией, за исключением ПСА при раке предстательной железы. Этот тест требует, чтобы концентрация, которая указывает, что биопсия была бы подходящей, сильно искажалась, чтобы снизить ложноположительные результаты, что приводит к очень высоким уровням ложноположительных результатов. Более 80% мужчин, которые указываются как нуждающиеся биопсии, на самом деле рака простаты не имеют.

Также, было установлено, что ДНК-маркеры очень хороши в некоторых случаях для некоторого подтипа рака, но опять же не подходят для скрининга по тем же причинам, что и НАР, указанные выше.

С использованием многочисленных белков также были изучены подходы протеомики. Эта работа была сосредоточена на, опять же, НАР или эффекторных белках высокого уровня. Эта работа в основном была проведена мультиплексными методами измерения белка, такими как иммунологические анализы, чипы и масс-спектрофотометрия. Некоторые ранние исследования завершились с некоторыми успехами для рака яичников. Тем не менее, проблема всех этих методов заключается в том, что многие из выбранных белков не имеют сильной корреляции с прогрессированием болезни от здоровых к больным (и многие не имеют известной биологической связи со стадией болезни, что, например, является обычным для масс-спектрометрии). Кроме того, масс-спектрометрия страдает от серьезной проблемы передискретизации в связи с тем, что вся проба сыворотки детально исследуется спектрофотометром на предмет уровней белков и, таким образом, обучение алгоритма корреляции затрудняется. В случае масс-спектрометрии в пробе сыворотки может быть найдено более 200 белков и 10000 пиков масс-спектра.

В области диагностики также необходимы способы, которые используют менее распространенные белки, которые являются более пригодными для диагностических целей, чем НАР, а также аналитические способы, которые подходят для анализа малораспространенных биомаркеров.

Сущность изобретения

Изобретение и различные варианты его осуществления заявлены в формуле изобретения, которая является частью этой патентной заявки. Без ограничения вышеизложенным, в предпочтительном аспекте, в предпочтительном варианте осуществления, изобретение относится к улучшению прогностической силы и диагностической точности способов для прогнозирования состояний заболевания с помощью способов корреляции с многочисленными переменными (мультивариантных способов). Эти способы включают в себя способы протеомики, метаболомики и другие способы, которые включают определение уровней различных биомаркеров, обнаруживаемых в телесных жидкостях и пробах тканей.

Различные варианты, раскрытые и обсуждаемые изобретателями в этой заявке, включают использование метапеременных, в частности, использование способов, которые регулируют влияние измеряемых аналитов биомаркера на показатель корреляции. Такие метапеременные могут быть идентифицированы, опираясь на специальные знания о реакции иммунной системы и понимание возможных ошибок измерения. Эти способы могут применяться к построению модели обучающей последовательности или к слепым пробам, по которым проводится диагностика.

В одном варианте осуществления данное изобретение относится к способу диагностики заболевания, включающему следующие этапы: а) определение концентраций по меньшей мере трех predetermined аналитов в слепой пробе субъекта; b) выбор одной или более метапеременных, связанных с субъектом, которая(ые) меняются в популяции, связанной с этим субъектом, для членов этой популяции, которые, как известно, либо имеют, либо не имеют заболевания; c) преобразование концентраций аналитов в зависимости от одной или более характеристик распределения в популяции и одной или более метапеременных для вычисления псевдоконцентрации, которая представляет каждый аналит; d) сравнение псевдоконцентраций с моделью обучающей последовательности для псевдоконцентраций, определенных для членов упомянутой популяции, которые, как известно, либо имеют, либо не имеют этого заболевания; и e) определение, указывает ли упомянутое сравнение, что субъект имеет данное заболевание. Предполагается, что этап (а) определения концентраций (или уровней) predetermined аналитов может быть выполнен в отдельное время и в отдельном месте от остальных этапов способа. Аналогичным образом, другой(ие) этап(ы) способа может быть осуществлен на практике полностью или частично в разное время и в разных местах. Соответственно, авторы настоящего изобретения также предусматривают, что изобретенный ими способ может содержать меньше этапов, в частности только этапы (b)-(e).

В одном аспекте настоящего изобретения вышеуказанные способы используют по меньшей мере три, по меньшей мере четыре, по меньшей мере пять или по меньшей мере шесть или более аналитов, которые измеряют или уровни которых определяют в биологической пробе, полученной от субъекта или пациента. В другом аспекте вышеуказанные способы включают оценку или прогноз наличия или отсутствия некоторого определенного заболевания, например солидных раков тканей, в том числе, но без ограничения упомянутым, рака молочной железы, рака предстательной железы и рака легких.

В некоторых вариантах осуществления метапеременной является возраст. В некоторых вариантах осуществления метапеременную выбирают из группы, состоящей из статуса пре-, пери- и постменопаузы, полового созревания, массы тела, географического расположения источника пробы, процента жира в организме, расы или расовой смеси, или этнической принадлежности, рода или эры (или продолжительности) периода времени.

В другом варианте осуществления "сравнительный" этап, как описано в данном документе, включает использование способа корреляции, выбираемого из способов, включающих, но не ограничивающихся, кластеризацию, поиск окрестности, регрессионный или вейвлет-анализ. Способ также может включать в себя использование инконгруэнтной модели обучающей последовательности. Такие инконгруэнтные модели обучающей последовательности могут быть использованы, по мере необходимости, с любыми из изобретенных способов, например в связи с этапами преобразования, сравнения и определения, которые можно повторить со второй моделью обучающей последовательности, способной выявить не связанные с заболеванием состояния в популяции субъекта, которые частично имитируют изменения в аналитах сыворотки при заболевании, но вызваны не заболеванием, в отличие от состояний или патологий самого заболевания. Таким образом, связанный вариант осуществления включает в себя вторую модель обучающей последовательности и оценку и прогноз заболевания для трех состояний: отсутствие заболевания, состояние отсутствия заболевания, которое частично имитирует заболевание, и наличие заболевания.

В другом аспекте настоящего изобретения способы изобретения реализуются посредством компьютера с использованием микропроцессора и необязательно содержат дополнительный этап вывода показателя в форме, пригодной для медицинского работника, например, врача, который осуществляет диагностику заболевания.

Некоторые варианты осуществления настоящего изобретения используют математические методы для нормализации и сглаживания нерегулярностей или несмежных распределений концентраций, которые включают использование логарифма соотношений измеряемых концентраций и скорректированных по возрасту средних значений концентраций белков для состояний отсутствия заболевания и наличия заболевания, для которых индивидуальная проба является прогностической и соотношение концентра-

ций белков для состояний отсутствия заболевания и наличия заболеваний является таким, что распределение получаемой в результате новой независимой переменной, подлежащей использованию в корреляции, сжимается для содействия корреляционному вычислению.

В другом аспекте настоящего изобретения взаимосвязь между независимыми переменными и мета-переменной охватывает характеристики распределения в популяции независимых переменных, связанных со степенью нелинейности взаимосвязи между состояниями наличия заболевания и отсутствия заболевания, одной или более группами (гауссовыми или негауссовыми), средними групповыми значениями, усредненными групповыми значениями, медианными групповыми значениями и групповыми значениями динамического диапазона.

Некоторые варианты осуществления настоящего изобретения включают в себя корректировку в модели обучающей последовательности веса влияния отдельных биомаркеров на основе традиционного (или специализированного) знания о характеристиках усиления или ослабления экспрессии отдельных биомаркеров, например, о подгруппах или степени нелинейности, в ходе прогрессирования заболевания у типичного субъекта в релевантной популяции.

Некоторые другие варианты включают корректировку в модели обучающей последовательности веса влияния отдельных биомаркеров на основе традиционного (или специализированного) знания о нестабильностях топологии плоскости бимаркера на достаточном количестве плоскостей бимаркеров, чтобы значительно изменить показатель риска или прогноз состояния болезни, при этом такие нестабильности обусловлены сильными уклонами или глубокими пиками или впадинами в топологии каждой плоскости бимаркера.

В других вариантах осуществления модель обучающей последовательности корректируется по весу влияния отдельных биомаркеров на основе традиционного (или специализированного) знания о неопределенностях биомаркерного анализа, например, неопределенностях, которые могут возникать при очень низких или очень высоких уровнях кривой результатов анализа.

В другом аспекте изобретения инконгруэнтная модель обучающей последовательности используется для корректировки или регулировки отдельных слепых (контрольных) проб, которые демонстрируют нестабильности в прогнозе исхода вследствие топологических нестабильностей на достаточном количестве плоскостей бимаркера, чтобы существенно изменить показатель риска для некоторой определенной слепой пробы, при этом нестабильности обусловлены сильными уклонами или глубокими пиками или впадинами в топологии некоторой плоскости бимаркера.

Другой вариант осуществления относится к более персонализированным медицинским подходам к диагностике и терапии, в которых базовые значения концентраций индивидуальных белков (или другого анализата, такого как метаболит) определяются для субъекта в течение периода времени, включающего период времени, когда у субъекта отсутствует заболевание, вместо значения популяции для заболевания, для которого желателен прогноз наличия заболевания или его диагностика.

Еще один аспект настоящего изобретения относится к измерению малораспространенных биомаркеров, в том числе сигнальных белков, которые включают в себя по меньшей мере один биомаркер в каждой по меньшей мере из трех категорий, которые включают маркеры воспаления иммунной системы, маркеры ангиогенеза опухоли, маркеры клеточного апоптоза, связанные с белками васкуляризации маркеры и маркеры ткани. В одном из вариантов изобретения, малораспространенные биомаркеры являются белками с крайне низким распространением с уровнями концентраций ниже около 1 пг/мл в пробах, полученных от по меньшей мере приблизительно 20% релевантной популяции для некоторого определенного субъекта.

Другой вариант осуществления изобретения включает определение концентраций в биологической пробе по меньшей мере трех биомаркеров, выбранных из категорий, которые включают маркеры воспаления иммунной системы, маркеры ангиогенеза опухоли, маркеры клеточного апоптоза, маркеры ткани и белков васкуляризации, при этом любые один или более из упомянутых по меньшей мере трех биомаркеров, отличные от маркеров ткани, являются малораспространенными белками с определяемыми концентрациями ниже около 1 пг/мл для по меньшей мере 20% релевантной популяции для некоторого определенного субъекта в субпопуляции, которая имеет заболевание, для которого желателен диагноз или прогноз вероятности заболевания.

В предпочтительном варианте осуществления болезнью является рак и более предпочтительно солидная опухоль.

В других вариантах осуществления оцениваются по меньшей мере два или более, по меньшей мере три или более, по меньшей мере четыре или более, по меньшей мере пять или более, по меньшей мере шесть или более, по меньшей мере семь или более, по меньшей мере восемь или более, по меньшей мере девять или более, или по меньшей мере десять или более биомаркеров (или определяются их уровни).

Еще один аспект изобретения включает оценку или анализ, при которых значения концентрации по меньшей мере одного из определенных (или измеренных) анализатов ниже, чем LOD, когда значение концентрации для такого анализата(анализов) определяется прямой или другим соответствующим стандартным методом подбора кривой между LOD и самым низким зарегистрированным значением для этого анализата. Предпочтительно никакому анализу не присваивается нулевое или отрицательное значение, и

никакому аналиту не присваивается значение, которое меньше приблизительно минимально допустимого значения для этого аналита в аналогичных пробах.

В других своих вариантах осуществления настоящее изобретение включает в себя диагностический набор с реагентами для обнаружения одного или более аналитов, двух или более аналитов, трех или более аналитов, четырех или более аналитов, пяти или более аналитов, шести или более аналитов, семи или более аналитов, восьми или более аналитов, девяти или более аналитов или десяти или более аналитов, которые ниже их LOD.

Другой вариант осуществления изобретения включает в себя компьютерные системы и микропроцессорное оборудование и системы, пригодные для реализации любого из способов, диагностических прогнозов и анализов, описанных в данном документе, включающих один или более этапов, рассмотренных в данном описании.

Описание чертежей

Следующие фигуры, которые описаны ниже и которые включены и составляют часть данного описания, иллюстрируют примерные варианты осуществления в соответствии с настоящим раскрытием, и их не следует рассматривать как ограничивающие объем изобретения, поскольку изобретение может допускать другие в равной степени эффективные варианты осуществления. Фигуры не обязательно выполнены в масштабе и некоторые особенности и некоторые виды на этих фигурах могут быть показаны в увеличенном масштабе или схематично в интересах ясности и краткости.

Фиг. 1 представляет собой блок-схему, которая представляет собой процесс построения модели обучающей последовательности (или диагностической модели), а затем обеспечения диагностических показателей для слепых проб, которые оценивают риск наличия заболевания.

Фиг. 2 представляет собой типичное распределение в популяции, в данном случае для цитокина Интерлейкин-6.

Фиг. 3 представляет собой плоскость бимаркера для одной из десяти таких плоскостей, показывающую псевдоконцентрации двух биомаркеров, используемых в диагностическом способе.

Фиг. 4 показывает плоскость бимаркера с точками данных обучающей последовательности.

Фиг. 5 показывает плоскость бимаркера без точек данных обучающей последовательности.

Фиг. 6 показывает плоскость бимаркера с затененной областью, где влияние снижено для реакции иммунной системы.

Фиг. 7 показывает плоскость бимаркера с затененной областью, где влияние снижено для проблем стабильности топологии.

Фиг. 8 показывает плоскость бимаркера с затененной областью, где влияние снижено для известной неопределенности измерения анализа.

Фиг. 9 показывает результаты слепых тестов с двумя пробами, которые не прошли тест на нестабильность топологии и были скорректированы с помощью инконгруэнтного алгоритма.

Фиг. 10 показывает результаты клинического исследования для рака молочной железы; в этом случае показатели рака в обучающей последовательности показаны для модели I обучающей последовательности с использованием 10 плоскостей бимаркера.

Фиг. 11 показывает результаты клинического исследования для рака молочной железы; в этом случае показатели рака в обучающей последовательности показаны для модели II обучающей последовательности с использованием 105 плоскостей бимаркера.

Фиг. 12 показывает результаты фактической диагностики для слепых проб, используемых при клиническом исследовании.

Фиг. 13 показывает калибровочную кривую для белка TNF α с точками данных слепой пробы.

Фиг. 14 показывает величины ошибок анализа белка TNF α для 5% ошибки показателя рака.

Табл. 1 показывает вычисление стабильности и показатель качества для алгоритма I из клинических исследований.

Табл. 2 показывает результаты доклинического исследования, проведенного на 868 женщинах на предмет рака молочной железы.

Табл. 3 показывает относительную прогностическую силу для различных способов корреляции для упомянутого исследования рака молочной железы для 868 женщин.

Табл. 4 показывает результаты доклинического исследования для 107 женщин на предмет рака яичников.

Табл. 5 показывает результаты доклинического исследования для 259 мужчин на предмет рака простаты.

Подробное описание

Изобретение описано ниже со ссылкой на подробные иллюстративные варианты осуществления. Станет понятно, что изобретение может быть воплощено в широком разнообразии форм, некоторые из которых могут быть совершенно отличными от раскрытых в данной заявке вариантов осуществления. Следовательно, конкретные структурные и функциональные особенности, раскрытые ниже, являются лишь примерными и не ограничивают объем настоящего изобретения.

В предпочтительном варианте осуществления настоящее изобретение относится к способу диагностики заболевания, который описан более подробно ниже. В качестве введения в то, что описано ниже, способ в общем использует измеряемые концентрации предопределенных аналитов в слепой пробе пациента, для которого желателен прогнозный диагноз в отношении того, имеет ли он заболевание или нет. Согласно способам настоящего изобретения, каждая концентрация аналита преобразуется в псевдоконцентрацию на основе использования по меньшей мере одной выбранной метапеременной, которая связана с пациентом и также изменяется в выбранной популяции, связанной с этим пациентом. В этой выбранной популяции, концентрации измеряются для тех же самых предопределенных аналитов для членов популяции, которые, как известно, либо имеют, либо не имеют заболевания. Для диагностики заболеваний, псевдоконцентрации обрабатываются с помощью способов и алгоритмов, описанных ниже. Обработанные значения псевдоконцентраций сравнивают с диагностической моделью (или моделью обучающей последовательности) для псевдоконцентраций, определенных и аналогичным образом обработанных для участников популяции, которые, как известно, либо имеют, либо не имеют этого заболевания. В конечном счете, производится определение, действительно ли оценка пробы, взятой у пациента, показывает состояние пациента, указывающее на нахождение этого пациента в некоторой группе популяции, имеющей заболевание или не имеющей заболевание. Это определение можно просмотреть, например, как результат, выводимый из компьютеризированной системы для использования медицинскими работниками.

Процесс поэтапно показан на блок-схеме на фиг. 1. В первую очередь производится построение модели обучающей последовательности, конечный продукт которой позволяет производить результативную диагностику неизвестных проб пациентов, называемых слепыми пробами, когда правильный диагноз не известен во время анализа этих слепых проб. В общем, настоящее изобретение обеспечивает показатель риска для медицинского работника, который затем учитывает этот показатель наряду с другими факторами пациента, чтобы сделать медицинское заключение о наличии или отсутствии некоторого определенного патологического состояния.

Определения

"Аналитическая чувствительность" определяется как три среднеквадратических отклонения выше нулевого калибратора. Диагностические показатели не считаются точными для концентраций ниже этого уровня. Таким образом, клинически-значимые концентрации ниже этого уровня не считаются точными и не используются для диагностических целей в клинической лаборатории.

"Бимаркер" представляет собой набор из двух псевдоконцентраций, которые нормируются и функционально связаны с изменением метапеременной по отношению к биологическому переходу от состояния отсутствия заболевания к состоянию наличия заболевания, когда нанесены на двухосевой график (или координатную сетку) как показано, например, на фиг. 3, и далее упоминаются как "плоскости бимаркера".

"Биологическая проба" означает ткань или телесную жидкость, такую как кровь или плазма, которая берется у субъекта и на основании которой концентрации или уровни диагностически-информативных аналитов (также называемых маркерами или биомаркерами) могут быть определены.

"Биомаркер" или "Маркер" означает биологическую составляющую биологической пробы субъекта, которая обычно представляет собой белок или метаболический аналит, измеряемый в телесной жидкости, например белок сыворотки крови. Примеры включают в себя цитокины, опухолевые маркеры и т.п.

"Слепая проба" представляет собой биологическую пробу, взятую у субъекта без известного диагноза некоторого определенного заболевания, и для которого желателен прогноз наличия или отсутствия этого заболевания.

"Предел обнаружения" (LOD) определяется как 2 среднеквадратических отклонения значения концентрации выше значения "нулевого" калибратора концентрации. Обычно нулевой калибратор работает с 20 или более экземплярами, чтобы получить точное представление о среднеквадратическом отклонении измерения. Определения концентрации ниже этого уровня считаются равными нулю или отсутствующими, например, для обнаружения вирусов или бактерий. Для целей настоящего изобретения, 1,5 среднеквадратических отклонения могут быть использованы, когда пробы дублируются, хотя использование 20 экземпляров является предпочтительным.

Диагностические представления, требующие некоторого значения концентрации ниже этого уровня, не воспроизводятся.

"Малораспространенные белки" являются белками в сыворотке на очень низком уровне. Определение этого уровня четко не описано в литературе, но при использовании в настоящем описании, уровень будет меньше, чем примерно 1 пг/мл в сыворотке крови или плазме и в других телесных жидкостях организма, у которого эти пробы были взяты.

"Метапеременная" означает информацию, которая является характеристикой некоторого определенного субъекта, отличной от концентраций или уровней аналитов и биомаркеров, но которая не обязательно является индивидуализированной или уникальной для данного субъекта. Примеры таких метапеременных включают в себя, но без ограничения, возраст субъекта, (пре-, пери-, пост-) менопаузальное состояние и другие состояния и характеристики, такие как половое созревание, масса тела, географическое местоположение или регион проживания пациента, географический источник биологической пробы,

процент жира в организме, возраст, раса или расовая смесь или период времени.

"Распределение в популяции" означает диапазон концентраций некоторого конкретного аналита в биологических пробах в некоторой определенной популяции субъектов. "Популяция" в данном описании означает, но без ограничения: лиц, выбираемых из некоторой географической области, некоторой конкретной расы, или некоторого конкретного пола. Характеристика распределения в популяции, выбираемая для использования как описано в данной заявке, дополнительно предполагает использование двух различных субпопуляций в пределах большей определенной популяции, которые являются членами популяции, которые были диагностированы как имеющие некоторое заболевание (субпопуляция с заболеванием) и не имеющие этого заболевания (субпопуляция без заболевания). Популяцией может быть любая группа, в которой желателен прогноз заболевания. Кроме того, предполагается, что соответствующие популяции включают тех субъектов, которые имеют заболевание, которое развилось до некоторой конкретной клинической стадии по отношению к другим стадиям развития для некоторого определенного заболевания.

"Характеристики распределения в популяции" являются определяемыми в распределении в популяции некоторого биомаркера, например, среднее значение концентрации некоторого конкретного аналита или его медианное значение концентрации, или динамический диапазон концентрации, или как распределение в популяции разделяется на группы, которые могут быть распознаны по характерным пикам, поскольку на степень усиления или ослабления экспрессии различных представляющих интерес биомаркеров и метапеременных влияет возникновение и прогрессирование заболевания по мере того как пациент претерпевает биологический переход или прогрессирование от состояния отсутствия заболевания в состояние наличия заболевания.

"Прогностическая сила" означает среднее значение чувствительности и специфичности для некоторого диагностического анализа или теста.

"Псевдоконцентрация" означает заместитель или замещающее значение для концентрации некоторого измеряемого биомаркера и, в сущности, является новой независимой переменной, которая может быть использована в диагностическом корреляционном анализе.

Псевдоконцентрация связана и вычисляется из концентрации измеряемых аналитов биомаркеров, такие аналиты обладают присущей прогностической силой для состояния некоторого определенного заболевания. Псевдоконцентрация вычисляется с использованием представляющей интерес характеристики распределения в популяции, скорректированной на основе метапеременной, для преобразования фактической измеренной концентрации прогностического биомаркера для некоторого определенного пациента, для которого желателен диагноз.

"Нестабильность топологии" представляет собой область на координатных сетках большинства или всех плоскостей биомаркера, в которой все или большинство точек находятся в участках сильных уклонов топологии. Топология представляет собой форму многомерного корреляционного вычисления, которое учитывает все измеренные независимые переменные (то есть, определяемые концентрации биомаркеров) и метапеременную. Это топология, для одного значения метапеременной, имеет по меньшей мере пять измерений для измерения пяти биомаркеров (но может быть и больше). Топология также смещается по форме по мере изменения значения метапеременной. Такая многомерная топология может быть наглядно визуализирована посредством взятия десяти биплоскостных срезов, проходящих через топологию. Это приводит к тому, что вычисленные показатели заболевания "в группе риска" могут быть неверными из-за шума измерения. Показатель может быть получен путем взвешивания отдельных графиков биомаркеров для прогностической силы в отношении состояния наличия заболевания и состояния отсутствия заболевания, и с учетом других факторов, таких как нестабильность измерения топологии и простая ошибка измерения. Диапазон показателя является произвольным и известным специалистам в данной области, а значение представляет собой процентную вероятность того, что пациент имеет заболевание или не имеет заболевания.

"Обучающая последовательность" является группой пациентов (200 или больше, как правило, для достижения статистической значимости) с известными концентрациями биомаркеров, известными значениями метапеременных и известными диагнозами. Обучающая последовательность используется для определения значений осей "псевдоконцентраций" в плоскостях "биомаркера", а также точек координатной сетки показателя из кластерного анализа, который будет использоваться для определения показателей для отдельных слепых проб.

"Модель обучающей последовательности" является алгоритмом или группой алгоритмов, которые строятся из обучающей последовательности и позволяют оценить слепые пробы в отношении прогнозного исхода относительно вероятности того, что субъект (или пациент) имеет заболевание или не имеет заболевания. Затем "модель обучающей последовательности" используется для вычисления показателей для слепых проб для клинических и диагностических целей. Для этой цели, показатель обеспечивается в произвольном диапазоне, который указывает процентную вероятность наличия заболевания или отсутствия заболевания, или другой показатель, который может быть предпочтительным для медицинского работника, который ставит пациенту диагноз.

"Инконгруэнтная модель обучающей последовательности" (или "вторичный алгоритм") является

вторичной моделью обучающей последовательности, которая использует другой феноменологический метод отбора и обработки данных, так что отдельные точки на координатных сетках плоскостей бимаркера вряд ли будут нестабильными как в первичной корреляционной модели обучающей последовательности, так и в этом вторичном алгоритме.

Обсуждение

Некоторые аспекты изобретения, в том числе определенные термины, будут рассмотрены ниже более подробно как руководство специалистам в данной области техники для осуществления настоящего изобретения на практике.

Метапеременные.

Заявленное изобретение относится, в частности, к улучшенным способам диагностики для корреляционных диагностических анализов, использующих метапеременные. Такие метапеременные могут улучшать прогнозную способность для диагностических целей, когда они демонстрируют значительный диапазон изменения "нормальных" значений среди членов популяции, которые, как известно, либо имеют, либо не имеют заболевания. Как указано в данном описании, метапеременные используются для преобразования или конвертации измеренных уровней аналитов в "псевдоконцентрации". Предполагается, что эти уровни или концентрации различных аналитов могут быть измерены или определены способами, известными специалистам в данной диагностической области техники.

Метапеременная может быть относительно более информативной, когда она имеет физиологическую или физико-химическую связь с биологическим статусом субъекта, которая отражает изменение с течением времени развития заболевания, хотя, эта метапеременная сама по себе и не является особо прогностической. Например, индекс массы тела (BMI, ИМТ) является допустимой метапеременной, а сама масса тела влияет на уровни различных сигнальных белков при сердечных заболеваниях. В способах по настоящему изобретению ИМТ может быть значительно более полезным при прогнозном анализе, когда он используется в качестве метапеременной, а не как другая независимая переменная, такая как независимые переменные измеренных уровней различных циркулирующих в крови белков. Настоящее изобретение основано отчасти на обнаружении того факта, что изменение массы тела во всей популяции, связанной с некоторым субъектом-человеком, дополнительно связано с определяемыми моделями распределения в популяции измеряемых уровней белков в сыворотке крови. Эти уровни белка (или биомаркера) являются независимыми переменными, измеряемыми для диагностических целей по мере того, как некоторый определенный субъект претерпевает биологический переход (или прогрессирование) состояния от здорового к больному.

Аналогичным образом авторы настоящего изобретения показали, что возраст субъекта в диагностике заболевания, например, рака, при использовании его в качестве независимой переменной вместе с измеряемыми уровнями аналита в традиционных корреляционных способах не является клинически-прогнозным. Однако, когда возраст используется в качестве метапеременной в соответствии со способами настоящего изобретения, его использование ведет к улучшению точности диагностики.

В общем, метапеременные, которые определены и описаны в данной заявке, будут являться прогнозными для диагностических целей, если характеристики распределения в популяции у измеряемого аналита демонстрируют значительное функциональное разделение (или дисперсию) при сравнении субъектов с заболеванием и без заболевания. Это функциональное разделение означает, что отношение метапеременной (возраста) и интересующей характеристики в популяции (среднее значение в популяции) для болезненного состояния и здорового состояния отличается существенно.

График на фиг. 2 показывает распределение IL-6 в популяции, в частности, показывает четыре различные подгруппы популяции, помеченные от I до IV, которые отражают иммунный ответ на прогрессирование рака. Группы с более высокими концентрациями, как полагают, могут быть результатом более сильной иммунной реакции на иммунные стимулы, такие как инфекции, раны, аллергии, и конечно рак, а высшая группа (IV) показывает решительную нелинейную реакцию IL-6 на эти стимулы. Эти распределения характеристик в популяции и такое разделение на группы измеряемых аналитов могут быть использованы для улучшения прогнозной способности диагностической модели обучающей последовательности.

Диагностические способы с метапеременными согласно данному изобретению основаны на исследовании, которое включало оценку проб примерно 868 пациентов. В этих пробах, концентрации пяти испытываемых сигнальных белков низкого уровня (ПСА, IL-6, IL-8, TNF α и VEGF) были измерены у субъектов, которые клинически диагностировались на имеющих рак молочной железы или на не имеющих рак молочной железы. Уровни белка (или аналита) являются классическими независимыми переменными для диагностического тестирования на заболевание.

Изобретатели также получили информацию о возрасте каждого из этих субъектов. Классический логистический регрессионный анализ пяти биомаркеров достиг примерно 82% прогнозной способности, а использование возраста в качестве шестой независимой переменной в этом аналитическом способе, как оказалось, давало незначительное улучшение прогнозной способности. Способ кластеризации данных, использующий только определенные биомаркеры, также достиг немного более высокой прогнозной способности в примерно 88%. Опять же, использование возраста в качестве независимой переменной суще-

ственно не улучшало точность прогнозирования.

Аналогично, использование классического кластерного анализа со значениями концентрации, преобразованными в логарифмы, достигло прогнозной способности в примерно 92%, а использование возраста в качестве независимой переменной добавило менее 0,5% к такой прогнозной способности. В таком виде анализа, как известно, логарифм концентраций пробы использовался, потому что концентрации этих аналитов могут распространяться на четыре или более участка динамического диапазона. Известно также, что уровни пяти аналитов в крови пациентов с раком имеют тенденцию к прогрессированию к сильно увеличенным концентрациям, но не всегда. Таким образом, этот подход приводит точки в модели обучающей последовательности на многомерных кластерных графиках в непосредственную близость, устраняя тенденцию к передискретизации точек кластера при низких концентрациях. Этот логарифмический способ сжатия используется обычно, поскольку он сокращает пространственное смещение.

Изобретатели установили, что возраст отдельного субъекта, в качестве примера, может быть использован для создания того, что в данном описании именуют метапеременной. Эта метапеременная используется в свою очередь для создания новых независимых переменных, которые именуются в данном описании как псевдоконцентрации, которые вычисляются из характеристик распределения в популяции измеряемых независимых переменных (в данном случае, концентраций белка). Поступая таким образом, который более подробно описан в данном описании, в двух аналитических моделях была получена прогнозная способность от примерно 97% до более чем 100%. Для целей настоящей заявки, прогнозная способность, которая превышает 100%, означает, что состояния наличия заболевания и отсутствия заболевания разделяются значительным разрывом в корреляционных показателях.

Характеристики распределения в популяции: примеры характеристик распределения в популяции, которые были найдены авторами данного изобретения в качестве доказательных, когда возраст выбирается в качестве метапеременной, являются среднее значение концентрации каждого измеряемого аналита для пациентов, либо имеющих, либо не имеющих заболевание; и срединное значение концентрации для подгрупп значений концентрации для отдельных возрастных групп (или для популяции в целом) для групп здоровых или больных пациентов, как показано на фиг. 2. Подгруппы, показанные на графике распределения в популяции, корректируются математически по-разному, другими словами, высокоуровневые концентрации с повышенным уровнем экспрессии выше 5 пг/мл сильно сжимаются. Эти отношения обеспечили новые основанные на возрасте независимые переменные, которые затем были использованы непосредственно в кластерном анализе, вместо фактических уровней концентрации измеряемых биомаркерных аналитов.

Средние значения концентраций определяются в зависимости от возраста для пациентов, диагностированных как находящиеся в состоянии наличия заболевания или отсутствия заболевания. Это именуется скорректированным по возрасту средним значением для каждого состояния. Значения псевдоконцентраций вычисляются из этих средних значений и фактической концентрации для пациента, используя уравнение 1 (которое описано в подразделе, относящемся к псевдоконцентрациям).

Как это часто бывает с корреляционным анализом с множеством независимых переменных, это множество независимых переменных идут в паре на многомерных графиках на основании классического анализа близости кластеров. Кроме того, эти переменные часто сжимаются чтобы вместить относительно большие разбросы динамического диапазона измеряемых переменных (то есть, аналитов). В этом случае, в данной заявке мы опишем способ, в котором независимые переменные являются значениями "псевдоконцентраций", основанными на возрасте или других метапеременных и фактических концентрациях. Предполагается, что специалисты в данной области легко смогут идентифицировать и выбрать характеристики распределения в популяции, которые улучшают прогножную способность.

Псевдоконцентрации.

Чтобы извлечь прогножную способность из возраста пациента, который сам по себе также может быть рассмотрен в качестве независимой переменной в других диагностических подходах, авторы настоящей заявки использовали возраст как метапеременную, чтобы вычислить "псевдоконцентрации", которые затем использовались в качестве заменителей фактических измеренных концентраций определяемых аналитов. Важно, чтобы метапеременная обладала разделением в интересующей популяции между состоянием наличия заболевания и состоянием отсутствия заболевания, чтобы извлечь ее прогножную способность для всех значений метапеременной, например, возраста, в случае рака молочной железы. Под "разделением" подразумевается, что наблюдаемая взаимосвязь между характеристикой популяции для субъектов в субпопуляциях отсутствия заболевания и наличия заболевания отличалась для некоторой конкретной метапеременной. Это может быть определено эмпирически путем измерений и вычислений. Интересующая популяция определяется по характеру желаемой группы прогнозного исхода и может быть, но без ограничения упомянутым: географическим регионом, например, континентальной частью Соединенных Штатов; определенной расой или этнической группой; или некоторым конкретным полом, например, женщины.

При осуществлении на практике способа согласно настоящему изобретению для прогнозирования рака молочной железы, концентрации пяти или шести сигнальных белков крайне низкого уровня предпочтительно являются независимыми переменными для диагностической корреляции. Каждый из этих

пяти биомаркеров был измерен, и его изменения с возрастом были определены в контрольной популяции, в которой находились женщины в возрасте от 35 до 80, половина членов упомянутой популяции были здоровы, а у другой половины был диагностирован рак молочной железы. В то время как субъект в популяции как правило рассматривается как имеющий или не имеющий заболевания, уровни биомаркеров, определяемых для отдельного субъекта, изменяются по мере прогрессирования заболевания этого субъекта и не все субъекты имеют одинаковый уровень биомаркеров до и во время начала и прогрессирования их заболевания на его различных стадиях. Таким образом, характеристики распределения в популяции отражают нормальную изменчивость биомаркера в субпопуляции в состоянии отсутствия заболевания, а также нормальную изменчивость биомаркера в субпопуляции в состоянии наличия заболевания.

Характеристики распределения в популяции в этом случае были: 1) средние значения в зависимости от возраста для состояния отсутствия заболевания и наличия заболевания (в данном случае, рака); и 2) степень и нелинейность повышения экспрессии биомаркера при переходе от отсутствия заболевания к наличию заболевания. Например, повышенная экспрессия интерлейкина-6 проиллюстрирована на фиг. 2, поскольку уровень этого цитокина отражает реакции субъектов на иммунный вызов в результате рака. IL-6 известен как провоспалительный респондер, экспрессия которого повышается как сигнал иммунной системе на включение своего общего ответа. Четыре отдельные группы продемонстрировали разные уровни повышенной экспрессии. Эти уровни были приняты во внимание при вычислении псевдоконцентраций. Например, степень сжатия данных при переходе от фактического сжатия к псевдоконцентрации варьируется в зависимости от местоположения группы и является очень серьезной для верхней группы 4.

Для того, чтобы реализовать вышеописанное, необходимо измерить популяцию с равными количествами субъектов без заболевания и с заболеванием. Размер такой обучающей последовательности оптимально определяется количеством используемых маркеров. Предпочтительный размер определяется, когда прогнозная способность модели обучающей последовательности находится в пределах 95% точности для аналогичной или большей последовательности неизвестной популяции. Тогда могут быть определены скорректированные по возрасту средние значения этих двух состояний, а влияние заболевания можно увидеть по повышенной или пониженной экспрессии биомаркера.

Фиг. 2 показывает характеристику распределения в популяции для одного биомаркера, IL-6, поскольку иммунная система повышает экспрессию этого белка, когда иммунная система борется с наличием рака или других провоспалительных состояний. Преобразование измеренной концентрации в псевдоконцентрацию включает нормализацию этой концентрации с помощью скорректированных по возрасту средних значений для, опять же, популяций с заболеванием и без заболевания и сжатие динамического диапазона измеряемых значений концентраций. Высоко рассеянные отдаленные концентрации в группе выше 5 пг/мл, которые простираются вплоть до 100 пг/мл, например, являются сильно сжатыми. Это улучшает прогнозную способность. Результатом является некоторая новая независимая переменная, именуемая в данной заявке как псевдоконцентрация, которая является безразмерной, нормализованной и отражает, в одном предпочтительном варианте осуществления, возрастные изменения в распределениях биомаркеров в популяции.

Отношение, которое включает в себя скорректированное по возрасту среднее значение для отсутствия заболевания и наличия заболевания и фактическую концентрацию в пробе пациента, используется в следующей форме.

Уравнение 1: псевдо-концентрация $\alpha = \text{натуральный логарифм } ((C_1/C_{(c \text{ или } h)}) - (C_h/C_c))^2$, где

C_1 = измеренная концентрация аналита фактического пациента;

$C_{(c \text{ или } h)}$ = скорректированная по возрасту пациента концентрация аналита данного пациента; значение корректируется в зависимости от того, находится ли пациент в состоянии отсутствия заболевания или наличия заболевания;

C_h = скорректированная по возрасту пациента средняя концентрация аналита среди пациентов без заболевания;

C_c = скорректированная по возрасту пациента средняя концентрация аналита среди пациентов с заболеванием.

Это уравнение 1 предназначено для корректировки сжатия и расширения в зависимости от распределения по группам с повышенной экспрессией, см. пики, например, на фиг. 2. Вышеуказанная формула удовлетворяет этому требованию; однако, многие другие формы этого уравнения могут быть реализованы как будет очевидно специалистам в уровне техники. Например, C_1 , C_h и C_c могут быть непосредственно концентрациями или расстояниями концентраций от среднего, медианного значения, или расстоянием от медианных значений подгрупп или краев динамического диапазона, как описано выше.

Псевдоконцентрации (безразмерные и не являющиеся, в силу этого, концентрациями или уровнями) затем используются на многомерном графике корреляционных кластеров для анализа. Кроме того, все графики приведены к общим характеристикам распределения в популяции; средним значениям возраста отсутствия заболевания или наличия заболевания (скорректированным по возрасту, либо нет), медианному значению или динамическому диапазону подгрупп. Эти способы могут обеспечить улучшения прогнозной способности в 5 или более процентных пунктах.

Индивидуализированная медицина набирает популярность и эффективность. Также предполагается, что способы прогнозирования заболевания, описанные выше, могут быть персонализированы путем замены характеристик распределения в популяции, выработанных и описанных в данной заявке для состояния отсутствия заболевания, измерениями базового уровня отсутствия заболевания для индивидуума. Другими словами, значение C_h в уравнении, указанном выше, было бы фактическим значением базового уровня для отдельного пациента, а не средним значением в популяции для состояния отсутствия заболевания. Тогда, оценка заболевания соответственно основывалась бы на переходе индивидуума от таких измеренных характеристики к характеристикам заболевания, которые указывают на популяцию в целом.

Плоскости бимаркера.

Для анализа, использующего пять биомаркеров (и одну метапеременную), таких плоскостей бимаркера будет десять. График на фиг. 3 иллюстрирует красные точки сетки с заболеваниями и желтые точки сетки без заболевания. Пробы обучающей последовательности, используя независимые измеренные переменные (концентрации) и метапеременную (возраст), определяют возрастные характеристики популяции, вычисляют расстояния псевдоконцентраций на осях, и эти псевдоконцентрации применяются к каждой из десяти плоскостей бимаркера. График делится на 2000 элементов сетки по каждой оси для 40000 точек сетки в общем.

Определение относительно того, относится ли некоторая точка сетки к наличию заболевания или отсутствию заболевания, вычисляется путем определения расстояния от каждой отдельной точки сетки до ближайших точек измеренных данных для проб обучающей последовательности. На фиг. 3 приводится пример, в котором двумя биомаркерами являются IL-6 и VEGF, а в качестве метапеременной используется возраст. Как на оси ординат, так и на оси абсцисс приведены псевдоконцентрации, определенные как описано выше. Метапеременная и измеренные независимые переменные встроены таким образом в псевдоконцентрацию на этих графиках. Каждой из точек сетки присваивается значение, отражающее отсутствие заболевания и наличие заболевания, и назначается некоторый соответствующий числовой показатель (например, +1 и -1, хотя фактические числа являются произвольными). Этот показатель определяется вычисленным расстоянием до двух точек данных обучающей последовательности, относящихся к наличию заболевания или отсутствию заболевания. Наиболее короткое расстояние определяет этот показатель. Количество проб обучающей последовательности может меняться для выполнения такого определения расстояния (например, приблизительно от 4 до 6) см. фиг. 3. Относительно низкое число сравниваемых проб, скорее всего, приведет к относительно меньшей прогнозной способности. Точно так же увеличенное число сравниваемых проб может уменьшить прогнозную способность, поскольку "охват" точки сетки простирается далеко в нелокальные области топологии. Лучшее число определяется путем экспериментального расчета.

Предпочтительное число сравниваемых выборочных точек определяется там, где модель обучающей последовательности наибольшим образом согласуется с фактическим диагнозом. Фиг. 3 показывает процесс этого вычисления для двух биомаркеров, IL-6 и VEGF. Неизвестной точке сетки (небольшой квадрат, находящийся в районе 12,00 по оси y и в районе 4,00 по оси x на плоскости) назначается состояние отсутствия заболевания или наличия заболевания путем определения ее расстояния до, в данном случае, трех ближайших точек данных обучающей последовательности для каждого здорового и больного. Эти расстояния суммируются, а затем точке сетки назначается соответствующее состояние, отсутствие заболевания или наличие заболевания (вычислительный показатель +1 или -1, соответственно). Любой слепой пробе, диагностируемой в некоторый момент времени в будущем, показатель состояния будет назначаться, основываясь на том, куда она падает на этой сетке. Каждой слепой пробе также будет присваиваться показатель для всех плоскостей бимаркера. Общее количество проб обучающей последовательности может быть по меньшей мере 200 и более.

Определение общего показателя рака для слепых проб выполняется по всем плоскостям бимаркера с использованием определения отдельной точки сетки для отдельной пробы, умноженной на общую прогнозную способность для отдельной плоскости бимаркера. Отдельное значение точки сетки слепой пробы (+1 или -1, например) умножается на прогнозную способность (или чувствительность) отдельной плоскости бимаркера. Все десять плоскостей затем суммируются. Стандартный линейный способ и/или способ нахождения квадратного корня из суммы квадратов используются для получения конечного общего показателя для всех плоскостей бимаркера. Показатели нормируются и сдвигаются для получения показателей в диапазоне от 0 до 200, что является выходными данными, используемыми поставщиком медицинских услуг. Этот диапазон является произвольным.

Большие наборы плоскостей бимаркера могут быть построены из той же самой группы биомаркеров посредством проведения над ней математической манипуляции. Эти большие наборы бимаркера вполне могут обладать большей прогнозной способностью или они могут составлять инконгруэнтную модель обучающей последовательности (или вторичный алгоритм) для дальнейших улучшений прогнозной способности. Например, используя соотношения концентраций 5 биомаркеров, а не сами концентрации, для построения каждой псевдоконцентрации, будет создано 10 значений псевдоконцентраций и 45 плоскостей бимаркера. Построение псевдоконцентраций и плоскостей бимаркера вполне может обладать

более высокой прогнозной способностью, но, скорее всего, потребует больших обучающих последовательностей для выполнения точной корреляции с общей популяцией. Может быть также использовано, например, следующее соотношение: каждая концентрация, деленная на 1 за минусом другого значения концентрации. Специалист в данной области сможет легко определить, обладают ли эти альтернативные способы подготовки данных для многомерного кластерного анализа лучшей прогнозной способностью, посредством проведения тестирования способа(ов) на предмет прогнозной способности с наборами слепых проб.

В целях дальнейшего улучшения прогнозной способности, эти скорректированные по возрасту или по группам концентрации приводят в надлежащее состояние для их нормализации и уменьшения или устранения пространственного смещения в кластеризации на многомерных сгруппированных графиках маркеров для анализа близости кластеров. См. фиг. 3, на которой представлена плоскость бимаркера для IL-6 и VEGF. Есть десять таких плоскостей для испытательной панели рака молочной железы с пятью биомаркерами. В этом случае вычисленные значения псевдоконцентраций нормализуют и сдвигают для получения произвольных значений в диапазоне от нуля до двадцати с сильным сжатием выпадающих концентраций с высокой экспрессией.

Каждая из проекций бимаркера из многомерных плоскостей маркера на то же самое нормализованное расстояние между концентрациями из анализа возраста/разбиения на группы сжимаются и нормализуются относительно скорректированных по возрасту средних значений, а также скорректированных по возрасту (или целым популяциям) подгрупп.

Улучшения прогнозной способности модели обучающей последовательности с использованием корректируемых уровней влияния на плоскость бимаркера:

Обычно плоскость бимаркера будет оценена по двоичным числам на предмет отсутствия заболевания и наличия заболевания (например, +1 и -1). Способ с псевдоконцентрациями описанный в данном документе поддается дальнейшим совершенствованиям прогнозной способности за счет избирательной регулировки уровней влияния этих двух двоичных чисел. Описанные ниже способы разрабатываются в модели обучающей последовательности и после их установки фиксируются в этой модели.

Фиг. 4 и 5 ниже иллюстрируют проекции одной плоскости бимаркера для случая из пяти биомаркеров, используемых для прогнозирования наличия болезненного состояния, в данном случае рака молочной железы, с помощью пяти маркеров: IL-6, IL-8, TNF α , VEGF и ПСА. Фиг. 4 показывает модель обучающей последовательности с данными, используемыми, чтобы оценить точки сетки на графике посредством анализа с поиском кластера. Фиг. 5 показывает модель обучающей последовательности без данных. Это составляет модель обучающей последовательности. Данные обучающей последовательности, использованные для создания модели не являются необходимыми как только каждой из 40000 точек сетки присвоен показатель, и показатель слепой пробе присваивается в зависимости от того, куда эта проба попадает на сетке. Топология показывает красным цветом рак-положительные области и синим цветом рак-отрицательные области. При вычислении общего показателя в этом случае точки сетки с отсутствием заболевания устанавливаются в +1, а точки сетки с заболеванием (раком) устанавливаются в -1. Каждый бимаркер в данном примере с пятью биомаркерами анализируют в пяти ортогональных пространствах, из которых на фиг. 5 проиллюстрирована одна проекция двух измерений. На этом графике показана топология различных подгрупп ответа иммунной системы. В этом случае все точки сетки (2000×2000 или 40000 в данном случае) оценивают в обычном порядке и значение -1 назначают заболевание(рак молочной железы)-положительной точке и значение +1 назначают заболевание-отрицательной точке. Эта плоскость бимаркера нормализуется с помощью расстояний псевдоконцентраций и для метапеременной возраста, как отмечено выше.

Фиг. 6 показывает ту же самую модель бимаркера и дополнительно группы иммунного ответа (см. фиг. 2) внутри серых областей. Влияние серых областей корректируют, чтобы отразить тот факт, что каждая серая заблокированная область имеет несколько иное влияние на вероятность того, что пациент здоров или болен. Эта корректировка может осуществляться либо по оценке человека с проверкой обучающей последовательности (обеспечила ли корректировка уточненные результаты обучающей последовательности), либо с помощью точного компьютерного инкрементального анализа с множеством переменных. Две отдельные плоскости бимаркера создаются для двух исходов, которые являются состояниями наличия заболевания и отсутствия заболевания. В этом случае более вероятно, что точки слепых данных в группе IV иммунного ответа указывают на заболевание, и, таким образом, влияние (абсолютное значение) будет незначительно увеличено (например, путем изменения показателя с -1 до -1,1). Фактическая величина этого приращения предпочтительно будет определяться с помощью компьютерного анализа или, возможно, точных ручных способов. Этот способ является работоспособным для способа поиска кластера в рамках корреляционного анализа, но и другие средства могут быть использованы для достижения того же самого эффекта. Эти способы взвешивания влияния относительно связи с заболеванием могут обеспечивать улучшение прогнозной способности в примерно 1%. При прогнозных способностях свыше 95% такое улучшение является существенным.

Фиг. 7 снова показывает ту же самую плоскость бимаркера с серой заключенной в окружность зо-

ной в сложной области нелинейного, быстрого изменения топологии наличия заболевания в сравнении с отсутствием заболевания. Такие области могут быть идентифицированы путем введения тестовых значений слепых проб с введенным шумом (скажем $\pm 10\%$) в модель, а затем введения измеренной величины шума.

Большинство этих слепых точек существенно не изменится в оценке заболевания (рак). Тем не менее, может быть обнаружено, что некоторые точки сетки выполнили резкий переход из показателя отсутствия заболевания в показатель наличия заболевания после такого вида регулировки шума. Эти точки относятся к областям, в которых большинство или все плоскости бимаркера обладают быстро изменяющейся топологией, которая перекрывает общие многомерные плоскости биомаркера. Путем аккуратного уменьшения влияния в таких областях, может быть увеличен вес нескольких релевантных плоскостей бимаркера, в которых зашумленные данные находятся на широкой плоскости, не находясь при этом вблизи границ изменения исхода. Этот способ продемонстрировал возможность корректировки ошибочных прогнозов. В данном случае влияние (абсолютное значение) красных, рак-позитивных областей будет уменьшено, например, с $-1,0$ до $-0,9$. Или влияние синих не связанных с заболеванием областей будет уменьшено с $+1,0$ до $-0,9$. Уровень оптимального сдвига может быть определен точным компьютерным анализом.

Шум в анализе может оказать негативное влияние на точность корреляционного анализа. Этот шум может быть особенно проблематичным на уровне или ниже предела обнаружения анализа. Такой шум также может быть уменьшен за счет сокращения влияния измеренных точек для отдельных биомаркеров, которые находятся в этих неустойчивых зонах. Фиг. 8 снова показывает плоскость бимаркера для ПСА и IL-6 для панели для рака молочной железы. Все области в пределах серой прямоугольной области в нижнем левом углу фигуры находятся ниже традиционного предела обнаружения (LOD) в анализе. Традиционно LOD определяется как два среднеквадратических отклонения 20 нулевых калибраторов плюс среднее значение значений этих двадцати нулевых калибраторов. Статистическая достоверность для значений на этом уровне составляет 95% в пределах двух среднеквадратических отклонений, и, конечно, достоверность измерений снижается по мере того, как измеряемая проба уходит ниже LOD. Такие данные все еще могут содержать полезную информацию, но применять их для анализа следует с меньшим влиянием. В этом случае, влияние на точки данных слепых проб в пределах серой области снижается, например, с $+1,0$ до $-0,9$ для точек сетки модели обучающей последовательности в пределах серой области. Это увеличивает влияние для точек данных этой исследуемой пробы, которые находятся выше предела обнаружения на их, других плоскостях бимаркера.

Вышеизложенные способы дополняют друг друга и могут реализовываться в tandem путем объединения сдвигов влияния.

Способы улучшения прогнозной способности.

Путем тестирования слепых проб на предмет нестабильности: после того как модель обучающей последовательности настроена и завершена она используется для вычисления показателей рака для слепых проб пациентов. Авторы используют два предпочтительных способа для получения показателей рака. Первый способ, именуемый линейным (CSI), берет показатель расположения в топологии ($+1$ или -1), умноженный на прогнозируемую способность для этой плоскости бимаркера. Затем они суммируются, масштабируются и сдвигаются, чтобы получить оценку от 0 до 200. Второй показатель, именуемый q-показателем (CSq), вычисляют с помощью квадратного корня суммы квадратов тех же самых значений. Этот второй способ подчеркивает разницу в отдельном показателе бимаркера и является пригодным для постановки общего конечного диагноза врачом.

Нестабильность топологии все еще остается в плоскостях бимаркера из-за высоко нелинейного характера способа кластеризации при корреляции и не может быть устранена полностью. Расположения этих нестабильностей могут быть найдены обширной и точной оценкой для вычисленного показателя рака каждой точки сетки посредством пошагового увеличения значений псевдоконцентраций по каждой плоскости бимаркера. Эта процедура будет включать обширные компьютерные вычисления 40000 точек сетки по 10 плоскостям бимаркера по числу биомаркеров (2000000 для 5 биомаркеров). Неустойчивые области будут выявлены по значительным колебаниям в показателях рака в соседних точках сетки. Это также может быть сделано менее строгим способом посредством визуального наложения всех десяти плоскостей бимаркера с поиском областей близкого перехода от здорового состояния к больному (для например, рак, по приблизительно 5 или более плоскостям). Эти визуально найденные области затем могут быть проверены меньшим числом компьютерных верификационных вычислений.

Согласно другому аспекту настоящего изобретения, тест на стабильность и способы, включающие введение шума, могут быть применены к набору слепых данных. А инконгруэнтная модель обучающей последовательности может быть использована для арбитража или корректировки показателей рака. Для этого аспекта настоящего изобретения, фиксированный уровень шума вводят для каждого набора слепых данных пациентов (например, плюс или минус 10%). Если набор слепых проб состоит из примерно 100 пациентов, то исполнение компьютером фактической модели обучающей последовательности будет проводиться для набора из 300 проб, каждая из которых присутствует в трех экземплярах (в исходных данных, плюс шум и минус шум). Результирующий набор данных в трех экземплярах затем тестируется на

предмет стабильности (а составляет -10%, b составляет +10%, а точка с является исходными данными). Табл. 1 показывает результат теста на стабильность для данных из клинического исследования. Следует обратить внимание, что три пробы демонстрируют очень высокую нестабильность в показателях рака. Пробы 138, 207, 34 и 29 все демонстрируют очень высокий критерий достоверности. Критерий достоверности (чем ниже, тем достовернее) должен охватывать как степень смещения показателя и, в особенности, сдвигается ли показатель для прогноза перехода из здорового состояния к больному или наоборот. Эти наборы данных из слепых проб имеют высокий риск некорректности в прогнозном диагнозе.

Инконгруэнтная модель обучающей последовательности может быть использована для арбитража наборов данных проб пациентов "в группе риска", которые не прошли тест на достоверность с введением шума. Эти точки находятся в группе риска из-за неизбежного шума измерения, либо случайного, либо систематического, в сочетании с крайней нестабильностью топологии, вызванной тем фактом, что точка данных слепой пробы находится на очень крутом уклоне большинства, если не всех, плоскостей бимаркера, так что малые возмущения дают в результате большие колебания в показателе. В табл. 1 приведены пробы с введенным шумом. Каждая проба имеет три значения: 1) плюс шум, 2) минус шум и 3) исходные данные без шума. Эти пробы демонстрируют показатели рака, которые осуществляют резкий переход с заболевания на отсутствие заболевания и обратно с введением $\pm 10\%$ шума. Данные этих проб в этом случае оцениваются как нестабильные. Уровень нестабильности не является точно определенным и корректировки могут быть сделаны для различных уровней вводимого шума. В этом случае они корректируются с помощью шума $\pm 10\%$ и показателя стабильности, превышающего 200 (следует отметить, что показатель стабильности и показатель рака являются двумя заметно различными числами разного содержания).

Шум измерения может подвергаться арбитражу с помощью этого второго инконгруэнтного алгоритма. Инконгруэнтный алгоритм, используемый для арбитража, может использоваться для корректировки упомянутых наборов проб пациентов "в группе риска" даже если эти он обладает немного меньшей прогнозной способностью нежели основной алгоритм, поскольку он улучшает шансы того, что точка является корректной. В данном случае (см. фиг. 9) две пробы были скорректированы: проба 138 имела показатель 85 отсутствия заболевания здоровый и была скорректирована до 195 с помощью инконгруэнтного алгоритма (эта точка была стабильной с алгоритмом I), проба 34 имела показатель 102 (линейный способ) и была скорректирована до 198 снова с помощью алгоритма II. Пробы 29 и 207 инконгруэнтным алгоритмом изменены не были.

Инконгруэнтная модель обучающей последовательности (алгоритм II) использовала 105 плоскостей бимаркера и была инконгруэнтной по отношению к первичной модели обучающей последовательности (алгоритм I) в том отношении, что те же самые пробы демонстрировали стабильность в тесте на стабильность алгоритма II. Тестирование инконгруэнтной модели обучающей последовательности проводится точно так же, как и для первичной модели обучающей последовательности. Следует отметить, что способ логистической регрессии может быть также использован для вычисления этих показателей проб. Алгоритм II обладает высокой прогнозной способностью, поэтому он был использован. Выполняющая арбитраж модель обучающей последовательности может быть использована даже если ее прогнозная способность ниже (предпочтительно все же не ниже 50%-ой прогнозной способности) таковой у основного алгоритма до тех пор, пока она обладает вероятно правильным результатом, не демонстрирующим нестабильность. Следует обратить внимание, что коррекция является разительной для интересующих слепых проб, которые не прошли тест с шумом. Все эти пробы на самом деле указывали на рак с высокими показателями. Восемь из десяти плоскостей бимаркера для этих проб находились в топологии с крайне нестабильными точками сетки. Таким образом, эти показатели находились "в группе риска" и в действительности не были правильными (один был неправильным, а другой был ненадежным с показателями 100/120). В этом случае одна проба была скорректирована для улучшения прогнозной способности с 97% до 98%, что является очень значительным снижением ошибки (50%). Одна проба, хоть и была ненадежной, была изменена на указание рака и также скорректирована.

Способ улучшения прогнозной способности корреляции двоичного исхода с состоянием заболевания за счет исключения независимого состояния, которое частично имитирует одно из конечных состояний при первичном анализе на заболевание:

Кластерный анализ широко использует три или более независимых переменных, зачастую это концентрации белка в сыворотке крови пациента. Корреляционный алгоритм может действовать только на двоичный результат либо заболевание отсутствует, либо заболевание присутствует, но он производит непрерывную оценку, которая более точно относится к вероятности фактического исхода, являющегося упомянутыми двумя двоичными условиями. В некоторых случаях, имеются другие условия, номинально классифицируемые как отсутствие заболевания, которые частично имитируют состояние заболевания в пределах распределений в популяции используемых биомаркеров. В некоторых из этих случаев это "имитирующее" отсутствие заболевания состояние может приводить к ложноположительному результату корреляционного анализа. Решение для исправления этого вида ложноположительного результата заключается в создании дополнительного нового корреляционного анализа, полностью отдельного от анализа на предмет отсутствия заболевания или наличия заболевания. Этот новый корреляционный ана-

лиз предпочтительно использует те же самые измеряемые данные биомаркера, что и в корреляции на предмет отсутствия заболевания или наличия заболевания, или он может использовать некоторые или все другие биомаркеры. Этот новый корреляционный анализ обеспечивает результат, указывающий на "имитацию отсутствия заболевания" или "наличие заболевания" или по меньшей мере обеспечивает показатель, позволяющий выполнить суждение о реальном состоянии пациента. Неопределенный показатель или показатель вблизи перехода для анализа на предмет отсутствия заболевания или наличия заболевания в сочетании с очень низким или высоким показателем в корреляции имитации отсутствия заболевания или наличия заболевания может помочь практикующему врачу улучшить суждение о состоянии заболевания и сократить число ложноположительных показателей.

Примером такой ситуации, когда состояние отсутствия заболевания имитирует болезненное состояние, является доброкачественная гипертрофия предстательной железы (ДГПЖ). Это состояние обычно демонстрирует высокие уровни по меньшей мере одного биомаркера, используемого для диагностики рака предстательной железы. Например, биомаркер, простатический специфический антиген, будет повышен у мужчин с ДГПЖ, а также с раком предстательной железы. Табл. 5 показывает, что этот дополнительный способ корреляционного анализа может различать мужчин с ДГПЖ и раком предстательной железы и, аналогичным образом с использованием тех же самых биомаркеров, но с помощью другой модели обучающей последовательности, может различать мужчин, которые предположительно находятся в состоянии отсутствия заболевания, и мужчин с подтвержденным раком предстательной железы в состоянии заболевания. У небольшой части мужчин, ложноположительный результат приведет к модели обучающей последовательности "отсутствие заболевания" против "рака", но данная ситуация будет дискриминирована моделью обучающей последовательности "ДГПЖ" против "рака". В таких случаях, два показателя, один для, предположительно, "отсутствия заболевания" против "рака", и один для "ДГПЖ" против "рака", поможет врачу или другому работнику здравоохранения решить, каким будет следующий диагностический этап. Например, для общей оценки (для любого из CSI или CSq) от 0 до 200 для обеих моделей показатель равный 110 для ОТСУТСТВИЯ РАКА ПРЕДСТАТЕЛЬНОЙ ЖЕЛЕЗЫ ИЛИ НАЛИЧИЯ РАКА ПРЕДСТАТЕЛЬНОЙ ЖЕЛЕЗЫ является слабым показателем, чтобы рассматриваться в качестве рак-положительного, но дополнительный учет второго показателя, равного 30 для ДГПЖ или рака, укажет практикующему врачу высокую вероятность ДГПЖ, а не рака. Практикующий врач будет использовать эту дополнительную информацию наряду с другой медицинской информацией и историей болезни пациента, чтобы решить, какими будут следующие диагностические этапы.

Некоторые способы для улучшения прогнозной способности традиционных основанных на протеомике корреляционных способов для диагностики заболевания были описаны в данном описании. К ним относятся: 1) использование метапеременной и значений псевдоконцентраций для корреляции, и 2) использование специального знания о стабильности топологии и аналитических характеристиках измерения для корректировки влияния плоскости бимаркера в модели обучающей последовательности. Кроме того, описаны способы нахождения и корректировки проблем со стабильностью слепых проб, которые являются уникальными для конкретной модели обучающей последовательности, использующие инконгруэнтную модель обучающей последовательности. Дополнительно, описаны способы нахождения и исправления состояний отсутствия заболевания, которые частично имитируют модель обучающей последовательности для состояния некоторого определенного заболевания. Все эти способы дополняют друг друга и могут быть использованы вместе. Например, регулировка модели обучающей последовательности для областей с высокой вероятностью нестабильности не может полностью устранить эту проблему из прогнозных вычислений по слепым пробам, таким образом оба способа могут быть использованы для улучшений прогнозной способности. Авторы настоящего изобретения обнаружили, что сочетание этих способов может обеспечить прогнозные способности, превышающие 95%, а исследование на предмет рака молочной железы, обсуждаемое в примере 1, обеспечило прогнозную способность сверх 98% (100% чувствительность, 97,5% специфичность).

Пример 1. Клиническое исследование - оценка исследования крови на предмет рака молочной железы.

Производительность тестового набора OTraces BC Sera Dx и системы иммунохимических инструментов Otraces CDx (www.otraces.com) оценивалась в эксперименте по оценке риска наличия рака молочной железы. Тестовый набор измеряет концентрации пяти цитокинов очень низкого уровня и тканевых маркеров, и использует модель обучающей последовательности, которая была разработана в соответствии с вышеописанным, для вычисления показателей CSI и CSq, чтобы оценить риск рака молочной железы. IL-6, IL-8, VEGF, TNF α и ПСА были измеряемыми белками. Эксперимент состоял в измерении порядка 300 проб пациентов, грубо 50% которых были случаи рака молочной железы, диагностированного по биопсии, а другие 50% были получены от пациентов, которые предположительно рассматривались как не имеющие заболевания (или, в этом случае, не имеющие рака молочной железы). У 50% результатов биопсии для 200 проб заболевание отсутствовало, а у других 50% имелся рак молочной железы, каждая группа была дополнительно подразделена на группы по конкретным возрастам.

Результаты анализа проб были использованы для создания модели обучающей последовательности, которая прогнозировала состояние заболевания. Остальные пробы (около 110) затем обрабатывались

этой моделью обучающей последовательности как слепые пробы, чтобы получить результирующие численные показатели риска рака и эти показатели были раскрыты принимающему клиническому центру. Показатели этих слепых проб впоследствии анализировались упомянутым клиническим центром для оценки клинической точности результатов.

Для этого эксперимента были созданы две диагностические модели, которые именуются в данном описании алгоритм I (или модель I обучающей последовательности) и алгоритм II (или модель II обучающей последовательности), как описано выше. Способ анализа окрестности кластера был использован для обоих алгоритмов. Возраст испытуемых был использован не как независимая переменная, а скорее как метAPERЕМЕННАЯ для преобразования измеряемых концентраций в новые независимые переменные, именуемые в этом описании как псевдоконцентрации, которые были использованы непосредственно в корреляционном анализе. Различие между алгоритмом I и алгоритмом II состояло в числе новых независимых переменных, использованных в корреляции. Алгоритм I использовал пять переменных псевдоконцентраций в десятимерном кластерном пространстве. Это пространство может наблюдаться человеческим глазом с помощью проекции или срезов через это многомерное пространство, чтобы взглянуть на двумерную плоскость бимаркера. В алгоритме I таких плоскостей было десять.

Алгоритм II использовал в десять раз больше созданных независимых переменных, так что в нем имелось около 100 плоскостей бимаркера. Предполагалось, что 200 проб достаточно для модели обучающей последовательности так, чтобы она достаточно близко моделировала основную популяцию. Вторичная или инконгруэнтная модель обучающей последовательности была создана из набора обучающих данных тех же самых 200 проб. Модель обучающей последовательности являлась первичным способом оценки, использованным для описания результатов в данном описании. Инконгруэнтная модель обучающей последовательности использовалась для арбитража показателей рака, вычисленных первичной моделью обучающей последовательности, которые рассматривались в качестве нестабильных, т.е. показателей, которые находились в области нестабильности топологии. Несмотря на то, что инконгруэнтная модель обучающей последовательности является несколько менее точной на слепых пробах, она все еще может выполнять арбитраж первичной модели обучающей последовательности и, тем самым, улучшать прогнозную способность.

Вышеописанный способ кластерного анализа имеет существенное преимущество, по отношению к логистической регрессии, в том, что он может работать с высоко нелинейными тенденциями в независимых переменных используемых для создания результата вычисления. Результатом (исходом) является либо наличие заболевания, либо отсутствие заболевания (в данном случае наличие рака или отсутствие рака), и он основан на псевдоконцентрациях для вычислений модели обучающей последовательности. Недостаток данного способа заключается в том, что высоко нелинейные области могут быть связаны с очень резкими уклонами топологии. Таким образом, неизвестная (или слепая) проба может располагаться на крутом пике или в глубокой и резкой впадине, что имеет эффект усиления небольших ошибок в вычисленных псевдоконцентрациях. Авторы настоящей заявки оценивали стабильность вычисленного показателя с помощью собственного теста на стабильность, а затем использовали алгоритм II для арбитража алгоритма I, если алгоритм II для проб, которые демонстрировали стабильность.

Фиг. 10, 11 и 12 показывают результаты обучающей последовательности алгоритма I. Сама модель состоит из 10 плоскостей бимаркера с 40000 точек топологии, каждой из которых способом кластеризации присваивается показатель для отсутствия заболевания и наличия заболевания (в данном случае, рака молочной железы). Способность модели разделять два набора отсутствия заболевания и наличия заболевания продемонстрирована на этих фигурах. Модель должна быть построена из равных частей для двух определяемых состояний, предпочтительно ровно 50 на 50% или очень близко к такому соотношению. Кроме того, способ использует возраст как преобразующую метAPERЕМЕННУЮ. Пробы обучающей последовательности включали пробы, распределенные по всем представляющим интерес возрастным группам. Модель (с фиг. 10) для алгоритма I была построена из 100 здоровых женщин и 98 женщин с раком молочной железы.

Сводная таблица на фиг. 10 иллюстрирует численные результаты, где N=198 - количество проб, CI - корректно названные пробы, FI - ложно названные пробы, а 4 пробы были признаны неопределенными.

Вторичная модель обучающей последовательности была разработана, чтобы различить упомянутые четыре неопределенные пробы, которые появились в результате использования первичной модели обучающей последовательности. Эта модель является инконгруэнтной моделью обучающей последовательности. Эта вторичная модель использует те же самые данные обучающей последовательности что и первичная модель.

Фиг. 11 демонстрирует результаты для вычислений инконгруэнтной модели обучающей последовательности. Алгоритм II демонстрирует 100% разделение с более чем 60 точками разделения.

Результаты тестирования слепых проб в исследовании рака молочной железы.

На фиг. 12 показаны результаты для слепых проб, оцененных в данном клиническом испытании. Результаты показывают 100% чувствительность и 97,5% специфичность. Онкологи в центре клинического исследования установили значение диагностического перехода так, что все пробы положительные на рак молочной железы были идентифицированы корректно. Таким образом, две пробы (предполагаемого)

отсутствия заболевания были названы рак-положительными. Это было важным сигналом, поскольку пробы, в отношении которых было вынесено рак-положительное суждение, все получали следующий диагностический этап - маммографию. Многие женщины не проходят маммографию, поскольку они не живут достаточно близко к учреждениям с соответствующим медицинским оборудованием. Тем не менее, их кровь можно забрать удаленно от клинической лаборатории и отправить на льду в лабораторию в крупный город.

Пример 2. Использование метапеременной "Возраст" для улучшения диагностической точности.

Табл. 2 демонстрирует сведенные в таблицу результаты клинического исследования проб 868 субъектов на предмет рака молочной железы. Табл. 3 показывает сравнение различных способов для вычисления корреляции. Стандартный способ, логистическая регрессия, продемонстрировал лишь 82% прогнозную способность. Стандартный анализ окрестности кластера улучшил ее, показав примерно 88% прогнозную способность. Способы, описанные в настоящем описании и использующие метапеременную и подходы с взвешиванием, обработку стабильности топологии, группировку по реакциям иммунной системы и обработку взвешиванием для обеспечения эффективности анализа - в сочетании с тестированием на нестабильность слепых проб и корректировкой инконгруэнтным алгоритмом, обеспечили более чем 97% прогнозную способность.

Пример 3. Использование метапеременной "Возраст" для улучшения диагностической точности в исследовании на рак яичников.

В табл. 4 приведены результаты исследования 107 женщин, имеющих рак яичников или не имеющих рак яичников, с использованием способа с метапеременной, описанного в настоящей заявке. В этом исследовании не использовались все возможные улучшения прогнозной способности, описанные в данной заявке, но этот факт все же не помешал достигнуть относительно высокой прогнозной способности примерно в 95%.

Пример 4. Использование метапеременной "Возраст" для улучшения точности диагностики на рак предстательной железы.

В табл. 5 приведены результаты исследования 259 мужчин, имеющих либо рак предстательной железы, либо доброкачественную гиперплазию предстательной железы (ДГПЖ), с использованием способа с метапеременной, описанного в настоящей заявке. В этом исследовании также не использовались все возможные улучшения прогнозной способности, описанные в данной заявке, но этот факт все же не помешал достигнуть относительно высокой прогнозной способности примерно в 94%. Следует обратить внимание, что ДГПЖ является на сегодняшний день наиболее распространенным состоянием, которое вызывает ложноположительные результаты в текущем тесте ПСА на предмет рака предстательной железы. Примерно 4 мужчины из пяти, у которых рак предстательной железы был диагностирован традиционным образом, имеют ДГПЖ, что подтверждается биопсиями на рак предстательной железы, большинство из которых оказываются негативными. Способ с метапеременной в состоянии исправить эти неправильные диагнозы, как описано выше.

В приведенных результатах в примерах 3 и 4 (для рака яичников и рака предстательной железы, соответственно) не применялись использующие метапеременную или корректировку влияния способы (LOD, группировка по субпопуляциям и нестабильность), а также не применялся учитывающий стабильность слепых проб способ, поскольку они еще не были обнаружены авторами настоящего изобретения на момент измерения упомянутых данных.

II. Диагностические способы, использующие предпочтительные категории аналитов и аналиты, измеряемые ниже традиционных пределов обнаружения.

Настоящее изобретение также основано на открытии того, что некоторые белки иммунной системы позволяют диагностировать риск некоторого определенного заболевания при измеряемых концентрациях существенно ниже тех, которые в настоящее время используются для коммерческого диагностического тестирования. К ним относятся цитокины, функциональность которых, в первую очередь, но не полностью, определяет их в качестве сигнальных белков и находится в некоторых группах; маркеры воспаления иммунной системы, маркеры ангиогенеза опухоли, клеточного апоптоза и васкуляризации опухоли, а также известные маркеры опухолевой ткани.

Авторы продемонстрировали, что выбор нескольких белков очень низкого содержания (LAP) и использование нетрадиционных способов корреляционного анализа и определения концентрации из способа иммуноанализа значительно улучшает прогнозную способность. Эти белки низкого содержания, именуемые сигнальными белками (т.е. белками, которые выполняют функцию одного или нескольких типов в сигнальной сети), действуют либо исходя из непосредственного ответа иммунной системы на наличие опухоли, либо вследствие направления организма опухолью на обеспечение необходимых физиологических реакций, которые необходимы опухоли для развития. Дополнительный выбор нескольких LAP (предпочтительно около 6 или менее) решает трудно решаемую проблему передискретизации, а также проблемы размера обучающей последовательности. Тот факт, что данные белки находились на уровне или ниже традиционных определенных пределов обнаружения, препятствовал ранее исследованию на предмет их полезности.

Настоящее изобретение также основано на неожиданном открытии того, что значения concentra-

ции ниже традиционно определяемого уровня обнаружения, обеспечивают значимую информацию при оценке соответствующего риска заболевания. Такая информация традиционно не использовалась при клинической диагностике из-за ее неточности. Тем не менее, авторы настоящего изобретения обнаружили, что она является эффективной для обеспечения аппроксимации прямой линией от точки калибровки LOD до наименьшего значения сигнала в тестовом прогоне и для использования этих значений. Поступив таким образом было обеспечено плавное Гауссово распределение для анализа распределения в популяции, а также, неожиданно, точные прогнозы показателя рака. В этом диагностическом способе, не следует сообщать никакое показание ниже LOD, которое находится ниже того, что обычно просматривается в крупномасштабных оценках популяции данного рынка. В качестве примера, если уровень сигнала (или измеренная концентрация) находится ниже LOD, использование описанного в данном документе способа целесообразно вплоть до низшего уровня, обнаруженного в нормальной сыворотке. Таким образом, когда LOD для IL-6 составляет около 250 фг/мл, но сообщаемые значения обнаруживались в сыворотке всего лишь со значением 10 фг/мл, этот уровень должен быть наиболее низким используемым в настоящем изобретении уровнем. Кроме того, никакое значение не может равняться нулю или быть отрицательным. Данный подход работает с многообразием традиционных стратегий создания стандартной кривой.

Авторы настоящего изобретения неожиданно обнаружили, что использование белков иммунной системы, цитокинов, чья функциональность находится в некоторых группах, маркеров воспаления иммунной системы, ангиогенеза опухоли, клеточного апоптоза и васкуляризации опухоли, а также известных маркеров опухолевой ткани достигает такой прогнозной способности, которая обладает характеристикой ложноотрицательных результатов корреляционной модели, превышающей 95%, и характеристикой ложноположительных результатов, также превышающей 95%. Таким белкам необходимы способы для извлечения полезной информации о концентрации на уровнях значительно ниже 1 пг/мл для некоторых маркеров. Для некоторой конкретной основанной на протеомике испытательной панели для рака молочной железы, которая, например, использует ПСА для тканевого маркера, IL-6 для воспалительной реакции, IL-8 для воспаления и васкуляризации, VEGF для васкуляризации и TNF α для противоопухолевого генеза, прогнозные способности были обеспечены выше 98%. Некоторые из этих маркеров имеют значительное распределение в популяции на уровне ниже 1 пг/мл (до менее чем 50 фг/мл). Это препятствовало их исследованию исследователями на предмет возможности использования этих белков для способов клинической диагностики.

Следует понимать, что настоящее изобретение включает в себя диагностические тесты, в которых, например, рак молочной железы прогнозируют с использованием маркеров воспаления иммунной системы (IL-6, IL-8), васкуляризации (IL-8, VEGF), белков противоопухолевого генеза (TNF α), а также тканевого маркера (ПСА). Могут быть использованы другие маркеры, которые попадают в эти категории, такие как СА 19.9 для ПСА, тканевых маркеров; или IL-1 в дополнение или как замена IL-6. Для рака предстательной железы предполагается, что будет использован прогнозный анализ, который включает маркеры воспаления иммунной системы (IL-6, IL-18), маркеры васкуляризации (IL-8, VEGF), белки противоопухолевого генеза (TNF α) и тканевый маркер (ПСА). Могут быть использованы другие маркеры, которые подпадают в эти категории (например, IL-1 в дополнение или как замена IL-6). Прогнозы предполагаемого статуса рака яичников выполняют с использованием маркеров воспаления иммунной системы (IL-6, IL-18), васкуляризации (IL-8, VEGF), белков противоопухолевого генеза (IL-12) и тканевого маркера (СА 125). Могут быть использованы другие маркеры, которые попадают в эти категории (например, TNF α для IL-12).

Авторы также обнаружили, что корреляционный анализ требует, чтобы измерения 100% популяции были однозначно точные, иначе прогнозная способность компрометируется. "Однозначно точные" не означает, что точность этих измерений должна быть эквивалентна ожидаемым измерениям при постановках клинических диагнозов, применяемых сегодня в клинических лабораториях. В клинической лаборатории, когда значение концентрации необходимо для диагноза точки в массиве, калибровочная кривая должна быть выше аналитической чувствительности, что означает 99,7% достоверность того, что результат находится в пределах 3 среднеквадратических отклонений от действительного значения.

Любые пробы, имеющие неопределенные или 0 значения для маркера, не принуждают корреляционное вычисление к рассмотрению этой пробы в качестве полностью некорректной. Сочетание нескольких сигнальных белков LAP со способом извлечения измерения концентрации крайне низкого уровня обеспечивает существенное улучшение результатов. При крайне низких уровнях ниже приемлемого традиционным анализом предела обнаружения, точность получается за счет простого использования прямой линии от LOD до наименьшего сигнала пробы и использования наименьшего физиологического уровня, обнаруженного в сыворотке в качестве концентрации для данной точки при тестовом прогоне многих проб. Точки проб между LOD и наименьшим показанием оцениваются по этой прямой линии. Также могут быть использованы другие стандартные способы аппроксимации кривой. Эти улучшения являются достаточно значительными, чтобы измерительная панель могла использоваться для скрининга на рак (обеспечивая прогнозную способность в 98% или более).

Авторы настоящего изобретения обнаружили, что использование белков иммунной системы, цитокинов, чья функциональность находится в некоторых группах; маркеров воспаления иммунной системы, антигенеза опухоли, клеточного апоптоза и васкуляризации опухоли, а также известных маркеров опухолевой ткани обеспечивает достижение такой прогнозной способности, которая обладает характеристикой ложноотрицательных результатов корреляционной модели, превышающей 95%, и характеристикой ложноположительных результатов, также превышающей 95%. Таким белкам необходимы способы для извлечения полезной информации о концентрации на уровнях значительно ниже 1 пг/мл для некоторых маркеров.

Для некоторой конкретной основанной на протеомике испытательной панели для, например, рака молочной железы, которая использует ПСА для тканевого маркера, IL-6 для воспалительной реакции, IL-8 для воспаления и васкуляризации, VEGF для васкуляризации и TNF α для противоопухолевого генеза, прогнозные способности были обеспечены выше 95%. Некоторые из этих маркеров имеют значительные распределения в популяции на уровне ниже 1 пг/мл (до менее чем 100 фг/мл). Корреляционный анализ требует, чтобы измерения крайне высокого процента (100%) популяции были однозначно точными измерениями, иначе корреляция найдена не будет. Любые пробы, имеющие неопределенные или 0 значения для маркера, не принуждают корреляционное вычисление к рассмотрению этой пробы в качестве полностью некорректной. Сочетание нескольких сигнальных белков LAP со способом извлечения измерения концентрации крайне низкого уровня обеспечивает существенное улучшение результатов. Этого вполне достаточно, чтобы измерительная панель могла быть пригодной для скрининга на рак.

Фиг. 13 иллюстрирует типичную калибровочную кривую ELISA для TNF α . В этом случае предел обнаружения (LOD) составляет около 1,0 пг/мл, что практически является наилучшим результатом, который может быть достигнут с помощью нынешних измерительных технологий быстрого скрининга. Следует обратить внимание, что значительный процент популяции находится ниже LOD. Неожиданно, но эти точки данных являются очень полезными и в действительности необходимыми для достижения корреляции с высокой прогнозной способностью.

Согласно обычной практике клинической лаборатории предел обнаружения определяется как два среднеквадратических отклонения выше нулевого калибратора, а среднеквадратическое отклонение как правило вычисляется с использованием двадцати нулевых стандартов. Измеряемые результаты, которые находятся ниже этого уровня, как правило не сообщаются или, если сообщаются, помечаются как находящиеся ниже LOD. Когда анализ с единственным аналитом используют для целей клинической диагностики, он должен иметь точное отдельно стоящее число над LOD для того, чтобы предложить практикующему врачу правильную диагностическую информацию, и требуется традиционный способ.

В случае диагностических анализов с использованием сигнальных белков низкого уровня, как описано в данной заявке, измерения с целью присвоения надежных показателей вероятности состояний отсутствия заболевания против наличия заболевания используют эти находящиеся ниже LOD измерения. На фиг. 14 приведены данные с фиг. 13, но со шкалами ошибок, демонстрирующими величину ошибки допустимой для поддержания ошибки вычисленного показателя рака менее чем в 5%. Крайне низкие уровни обнаружения на уровне или ниже LOD допускают некоторую относительную величину ошибки. То же самое допускается при крайне высоких концентрациях.

Критическая область для калибровочной кривой в диагностическом тесте находится там, где прогнозный показатель диагноза находится в диапазоне от показателей, которые явно указывают на состояние отсутствия заболевания, до показателей, которые явно указывают на состояние наличия заболевания. Кроме того, следует обратить внимание, что если данные, находящиеся ниже LOD, либо устранены, либо сообщены как 0, присвоение показателя для этого пациента может привести к серьезным ошибкам, здоровый пациент с низким показателем может быть отнесен к больному раком пациенту с высоким показателем. Так происходит поскольку алгоритм должен быть способен к "привязке" всех пяти маркеров к некоторому общему уровню в крайних положениях и должен быть точным на переходах. Если один маркер не привязан вычисленные результаты основаны лишь на 4-х других маркерах. Тем не менее вполне могут иметься состояния, при которых повышенные результаты для этих 4-х маркеров не указывают на рак, но все пять маркеров являются необходимыми и должны быть повышены для указания на рак и для выработки точного показателя. Каждый отдельный маркер оказывает ограниченный эффект на отдельный показатель рака и этот эффект взвешивается с помощью реального влияния маркеров и его вероятного уровня шума на основе его расположения относительно калибровочной кривой.

В качестве примера, раскрытая в данной заявке тестовая панель для рака молочной железы, которая включает ПСА, демонстрирует лишь 3-4% улучшение прогнозной способности общей модели обучающей последовательности для общего анализа большой обучающей последовательности (200 проб). Тем не менее, его исключение из одной слепой пробы может исказить показатель от модели обучающей последовательности настолько, что эта проба будет смещена относительно здорового состояния или рака (например, показатель от 50 до 180 на шкале от 0 до 200). Эти слепые пробы могут быть редкими, но лишь одна такая проба из 100 приводит к падению в прогнозной способности на 1%. С помощью раскрытых способов авторы добились 98% прогнозной способности, таким образом 1% потеря является очень

значительной. Точка данных слепой пробы для ПСА, в качестве примера, если оценена на очень низких уровнях около 10 фг/мл, может иметь 20-кратное смещение, но не вызывать при этом существенной ошибки показателя рака. С другой стороны, если ПСА не используется или назначен как 0, показатель рака становится "непривязанным" низким уровнем ПСА и, если не включается, может сместить показатель рака так, что он будет рассмотрен в качестве полностью некорректного.

Понимание неопределенностей, связанных с этими низкими уровнями измерений, очень важно. Для измеряемых проб, прогоняемых в двух экземплярах, уровень достоверности при среднеквадратическом отклонении, составляющем 1,5 от номинальной точки измерения, составляет 95%. При среднеквадратическом отклонении, составляющем 0,75 от номинальной точки измерения, достоверность составляет 67%. Этого уровня точности более чем достаточно при очень низких уровнях в этих способах и, полное отсутствие значения может сделать способ непригодным.

Способ, в соответствии с настоящим изобретением, для оценки значения измеряемых точек для корреляционных оценок регулируется следующим образом.

1) Традиционная калибровочная кривая должна распространяться от LOD для анализа, как он оценивался во время разработки, до максимально высокого значения, которое может иметь место при применении типичных способов иммуноанализа, например, ELISA, через динамическую область популяции.

2) Ниже LOD все точки, как предполагается, имеют существенный шум и, таким образом, фактические точки измерения могут быть ниже наименьшего калибратора. В этом случае сообщаемые результаты должны быть:

а) выше нуля, отрицательные концентрации не возможны, а нулевые значения в корреляционном алгоритме так же вредны для точности результатов, как и полное отсутствие значения;

б) определенными, а не просто иметь значение, назначенное всем неопределенным точечным концентрациям проб пациентов, что искажало бы распределения в популяции, необходимые для построения модели обучающей последовательности по подгруппам;

с) сообщены не ниже, чем обычно обнаруживается в крупномасштабных оценках популяции данного рынка.

Изобретатели неожиданно обнаружили, что простая аппроксимация прямой линией от калибровочной точки LOD до наименьшего значения сигнала в тестовом прогоне является адекватной. Это обеспечивает плавное Гауссово распределение для анализа распределения в популяции и точные вычисления показателя рака. Используя этот подход, любое число различных стратегий создания стандартных кривых будет работать, если следовать вышеуказанным правилам.

Предполагается, что способы, раскрытые для использования концентраций аналита ниже LOD, могут применяться для любого анализа, который использует измеряемые концентрации маркеров, например, в пробах пациентов. Предпочтительные варианты осуществления включают диагностические анализы на предмет различных заболеваний, таких как солидные опухоли, в том числе рак предстательной железы, легких, молочной железы и яичника. Эти способы необязательно, но предпочтительно, могут быть объединены с другими способами диагностики и анализа данных, описанными в других местах в настоящем описании.

Кроме того, специалисту в данной области техники будет понятно, что ошибки анализа, вызванные использованным корреляционным способом, например, нестабильности топологии плоскости бимаркера, или неадекватным размером обучающей последовательности, которые описаны в данной заявке, вызваны совершенно разными явлениями и требуют совершенно различных способов их исправления. Например, некоторые раскрытые ошибки и способы их исправления подходят для экстремальных нелинейных (или очень резких) уклонов, на которых располагается тестируемая точка пробы в корреляционной топологии бимаркера. Такие ошибки могут быть найдены путем введения искусственного шума и арбитража с помощью инконгруэнтной модели обучающей последовательности. Шум, к которому обращаются в данной заявке, присущ аналитическому измерению, которое является результатом экспериментальной ошибки, но крайне важно, чтобы его эффект на получаемый в результате показатель рака был понятен.

III. Реализация с помощью компьютерных систем.

Реализации различных способов, описанных в данной заявке, могут быть выполнены в цифровой электронной схеме или компьютерном аппаратном обеспечении, прошивке, программном обеспечении или в их комбинациях. Реализации могут осуществляться в виде компьютерного программного продукта, например, компьютерной программы, материально воплощенной на носителе информации, например, в машиночитаемом запоминающем устройстве или в распространяемом сигнале, для исполнения посредством или для управления работой устройства обработки данных, например, программируемого процессора, компьютера или нескольких компьютеров. Компьютерная программа, такая как вышеописанная компьютерная программа(ы), может быть написана на любом языке программирования, в том числе компилируемые или интерпретируемые языки, а также может быть развернута в любой форме, в том числе как автономная программа или как модуль, компонент, подпрограмма или другой блок, пригодный для использования в вычислительной среде. Компьютерная программа может быть развернута для исполнения на одном компьютере или на нескольких компьютерах в одном месте или распределенных по нескольким местам и взаимосвязанным коммуникационной сетью.

Этапы способа могут быть выполнены с помощью одного или нескольких программируемых процессоров, выполняющих компьютерную программу для выполнения функций посредством обработки входных данных и формирования выходных данных. Этапы способа также могут быть выполнены посредством, а устройство может быть реализовано как специализированная логическая схема, например, FPGA (программируемая пользователем вентиляционная матрица) или ASIC (интегральная схема специального назначения).

Процессоры, подходящие для выполнения компьютерной программы включают в себя, в качестве примера, общие и специальные микропроцессоры, а также любой один или более процессоров любого типа цифрового компьютера. Как правило процессор будет принимать инструкции и данные от постоянного запоминающего устройства или оперативного запоминающего устройства или обоих. Элементы компьютера может включать в себя по меньшей мере один процессор для выполнения инструкций, а также одно или более запоминающих устройств для хранения инструкций и данных. Как правило компьютер также может включать в себя или быть выполненным с возможностью принимать данные от и передавать данные на, или выполнять и то и другое, один или более съемных накопителей для сохранения данных, например, магнитные, магнитооптические диски или оптические диски. Информационные носители, пригодные для воплощения инструкций компьютерной программы и данные включают в себя все формы энергонезависимой памяти, в том числе, в качестве примера, устройства полупроводниковой памяти, например, EPROM, EEPROM, а также устройства флэш-памяти; магнитные диски, например, внутренние жесткие диски или извлекаемые диски; магнитно-оптические диски; и CD-ROM и DVD-ROM диски. Процессор и память могут быть дополнены или включены в логическую схему специального назначения.

Для обеспечения взаимодействия с пользователем, реализации могут быть выполнены на компьютере, имеющем устройство отображения, например, электронно-лучевую трубку (ЭЛТ) или жидкокристаллический (LCD) монитор для отображения информации пользователю, и клавиатуру и указательное устройство, например, мышь или трекбол, с помощью которых пользователь может осуществить компьютерный ввод. Другие виды устройств также могут быть использованы для обеспечения взаимодействия с пользователем; например, обратная связь, предоставляемая пользователю может принимать любую форму, относящуюся к ощущениям, например, зрительную обратную связь, акустическую обратную связь или тактильную обратную связь; а ввод от пользователя может быть принят в любой форме, в том числе акустический, речевой или тактильный ввод.

Реализации могут быть выполнены в вычислительной системе, которая включает в себя внутренний компонент, например, в качестве сервера данных, или которая включает в себя компонент промежуточных программных средств, например, сервер приложений, или которая включает в себя клиентский компонент, например, клиентский компьютер, имеющий графический пользовательский интерфейс или веб-браузер, через который пользователь может взаимодействовать с реализацией, или любую комбинацию таких внутренних, промежуточных или клиентских компонентов. Компоненты могут быть взаимосвязаны с помощью любой формы или носителя передачи цифровых данных, например, сети связи. Примеры сетей связи включают в себя локальную сеть (LAN) и глобальную сеть (WAN), например, Интернет.

Изобретение, описанное в данной заявке в целом относится к способам улучшения диагностической точности или прогнозной способности основанных на протеомике и метаболомике корреляционных способов для прогнозирования патологических состояний. Несмотря на то, что некоторые примерные варианты осуществления были подробно описаны выше и проиллюстрированы на прилагаемых чертежах, следует понимать, что такие варианты осуществления лишь иллюстрируют, но не ограничивают широкое изобретение. В частности, следует признать, что принципы изобретения применимы к широкому кругу заболеваний. Кроме того, хотя предпочтительные варианты осуществления настоящего изобретения относятся к диагностике на предмет человеческих заболеваний в субъекте (или пациенте), для которого требуется получить диагноз, здесь явно предполагается, что способы и системы, раскрытые в данном описании, также пригодны для целей диагностики нечеловеческих видов, в частности, приматов и других млекопитающих, и, таким образом, являются частью данного изобретения.

Несмотря на то, что некоторые особенности описанных реализаций были проиллюстрированы, как описано в данной заявке, многие модификации, замены, изменения и эквиваленты будут понятны специалистам в данной области техники. Таким образом, следует понимать, что изобретение не ограничивается конкретными вариантами осуществления или раскрытыми компоновками, а скорее предназначено для охвата любых изменений, адаптаций или модификаций, которые находятся в пределах объема и сущности изобретения, определенного в приложенной формуле изобретения.

Таким образом, в первом аспекте настоящего изобретения предусмотрен способ диагностики рака, содержащий этапы: а) определения концентраций по меньшей мере трех предопределенных аналитов в слепой пробе субъекта, при этом предопределенные аналиты выбирают из группы, содержащей IL-6, IL-8, VEGF, TNF α и ПСА; б) выбора одной или более метапеременной, ассоциированной с субъектом, которая меняется в популяции, связанной с субъектом, для членов упомянутой популяции, которые, как известно, либо болеют, либо не болеют раком, при этом метапеременной является возраст; в) преобразования концентраций аналитов в зависимости от одной или более характеристик распределения в популя-

ции и упомянутой одной или более метапеременных для вычисления псевдоконцентрации, которая представляет каждый аналит, при этом упомянутое преобразование включает в себя коррекцию пространственного смещения; d) сравнения псевдоконцентраций с моделью обучающей последовательности псевдоконцентраций посредством использования линейного показателя ("CSI") и q-показателя ("CSq"), определяемых для членов популяции, которые, как известно, либо болеют, либо не болеют раком; e) определения, болен ли субъект раком или нет на основе того, попадают ли показатели CSI и CSq псевдоконцентраций субъекта в группу с наличием рака или в группу без рака на графике CSI/CSq модели обучающей последовательности.

Во втором аспекте настоящего изобретения предусмотрен способ подготовки модели обучающей последовательности, способной диагностировать рак, содержащий этапы: а) определения концентраций по меньшей мере трех предопределенных аналитов в обучающей последовательности проб группы субъектов, при этом предопределенные аналиты выбирают из группы, содержащей IL-6, IL-8, VEGF, TNF α и ПСА; б) выбора метапеременной, связанной с субъектами, которая меняется в популяции, связанной с субъектами, для членов упомянутой популяции, которые, как известно, либо болеют, либо не болеют раком, при этом метапеременной является возраст; с) преобразования концентраций аналитов в зависимости от одной или более характеристик распределения в популяции и метапеременной для вычисления псевдоконцентрации, которая представляет каждый аналит, при этом упомянутое преобразование включает в себя коррекцию пространственного смещения; и d) подготовки модели обучающей последовательности из псевдоконцентраций посредством использования линейного показателя ("CSI") и q-показателя ("CSq"), определенных для членов популяций, которые, как известно, либо болеют, либо не болеют раком.

В способе согласно любому из вышеуказанных аспектов упомянутые по меньшей мере три предопределенных аналита выбирают из группы, состоящей из по меньшей мере четырех, по меньшей мере пяти и по меньшей мере шести аналитов.

В способе согласно любому из вышеуказанных аспектов рак выбирают из группы, состоящей из солидных тканевых опухолей, в том числе, но без ограничения упомянутыми, рак молочной железы, рак предстательной железы и рак легких.

Способ согласно любому из вышеуказанных аспектов содержит этап сравнения, при этом этап сравнения включает в себя способ корреляции, выбранный из группы, состоящей из способов кластеризации, поиска окрестности, регрессионного или вейвлет-анализа.

Способ согласно любому из вышеуказанных аспектов содержит этап сравнения, при этом этап сравнения дополнительно содержит использование инконгруэнтной модели обучающей последовательности.

Способ согласно любому из вышеуказанных аспектов реализуется компьютером и, необязательно, дополнительно содержит этап вывода показателя.

В способе согласно любому из вышеуказанных аспектов этапы преобразования, сравнения и определения повторяются со второй моделью обучающей последовательности, способной выявлять состояния отсутствия заболевания в популяции субъекта, которые частично имитируют изменения в сывороточном анализе, но не являются состоянием наличия заболевания.

В способе согласно любому из вышеуказанных аспектов, включающих в себя вторую обучающую последовательность, при этом этап определения обеспечивает оценку риска для трех состояний: состояние отсутствия заболевания, состояние отсутствия заболевания, которое частично имитирует состояние наличия заболевания, и состояние наличия заболевания.

В способе согласно любому из вышеуказанных аспектов математический способ для нормализации и сглаживания нерегулярностей или несмежных распределений концентраций является логарифмом соотношений измеренных концентраций и скорректированных по возрасту средних значений концентраций белков для состояний отсутствия заболевания и наличия заболевания, для которых индивидуальная проба является прогнозной и соотношение концентраций белков для состояний отсутствия заболевания и наличия заболевания является таким, что распределение результирующей новой независимой переменной, подлежащей использованию в корреляции, сжимается для содействия корреляционному вычислению.

В способе согласно любому из вышеуказанных аспектов метапеременная выбирается из групп, состоящих из: пре-, пери- и пост-менопаузального состояния, полового созревания, массы тела, географического местоположения источника пробы, процента жира в организме, возраста, расы или расовой смеси или этнической принадлежности, вида или временной эпохи.

В способе согласно любому из вышеуказанных аспектов взаимосвязь между независимыми переменными и метапеременной охватывает характеристики распределения в популяции независимых переменных, выбираемых из группы, состоящей из степени нелинейности взаимосвязи между состояниями наличия заболевания и отсутствия заболевания, одной или более групп (гауссовыми или негауссовыми), средних групповых значений, усредненных групповых значений, медианных групповых значений и групповых значений динамического диапазона.

Способ согласно любому из вышеуказанных аспектов дополнительно содержит этап коррекции для

модели обучающей последовательности, выполняемый посредством подстройки весового влияния индивидуальных биомаркеров на основании знания о характеристиках усиления или ослабления экспрессии индивидуальных биомаркеров, например, подгруппах или степени нелинейности, в процессе прогрессирования заболевания у типичного субъекта в релевантной популяции.

Способ согласно любому из вышеуказанных аспектов дополнительно содержит этап коррекции для модели обучающей последовательности, выполняемый посредством подстройки весового влияния индивидуальных биомаркеров на основании знания о нестабильностях топологии плоскости би-маркера на достаточном количестве плоскостей би-маркера, для существенного изменения показателя риска, причем такие нестабильности вызваны резкими уклонами или глубокими пиками или впадинами в топологии каждой плоскости би-маркера.

Способ согласно любому из вышеуказанных аспектов дополнительно содержит этап коррекции для модели обучающей последовательности, выполняемый посредством подстройки весового влияния индивидуальных биомаркеров на основании знания о неопределенностях анализа биомаркера, например, неопределенностях, которые могут возникать при очень низких или очень высоких уровнях на кривой результатов анализа.

Способ согласно любому из вышеуказанных аспектов дополнительно содержит этап коррекции для модели обучающей последовательности, выполняемый посредством использования инконгруэнтной модели обучающей последовательности для коррекции индивидуальных слепых проб, которые демонстрируют нестабильности при прогнозировании исхода вследствие нестабильностей топологии на достаточном количестве плоскостей би-маркера для существенного изменения показателя риска, причем такие нестабильности вызваны резкими уклонами или глубокими пиками или впадинами в топологии каждой плоскости би-маркера.

В способе согласно любому из вышеуказанных аспектов базовые значения индивидуальных концентраций белка определяются для субъекта с течением времени, в том числе периода времени, когда субъект находится в состоянии отсутствия заболевания, вместо значения популяции.

В способе согласно любому из вышеуказанных аспектов измеряемые аналиты представляют собой сигнальные белки низкого содержания, которые включают по меньшей мере один аналит в каждой из по меньшей мере трех категорий, выбираемых из группы, состоящей из: белков воспаления иммунной системы, ангиогенеза опухоли, клеточного апоптоза, васкуляризации и тканевых маркеров.

В способе согласно любому из вышеуказанных аспектов один или более аналитов представляют собой белки крайне низкого содержания с уровнями концентрации ниже, приблизительно, 1 пг/мл в пробах, извлеченных из по меньшей мере, приблизительно, 20 процентов релевантной популяции для некоторого определенного субъекта.

Источники

Все журнальные статьи и все другие публикации, патенты и тексты, упомянутые в данном описании, включены в данное описание по ссылке во всей своей полноте, в том числе следующие источники:

- (1) Drukier, et al., "High-Sensitivity Blood-Based Detection of Breast Cancer by Multi Photon Detection Diagnostic Proteomics," Journal of Proteome Research 2006, 5:1908,1915.
- (2) Lokshin et al., "Multimarker assay for early diagnosis of ovarian cancer," American Association for Cancer Research, Amer Assoc Cancer Res 2006, 47:653. CME: Disclosure.
- (3) Drukier, et al., Ultra-Sensitive Immunoassays Using Multi Photon Detection in Diagnostic Proteomics of Blood," Journal of Proteome Research 2005, 4:2375-2378.
- (4) Drukier, "Supersensitive Immunoassays," U.S. Patent No. 7,604,956 (2009).

Таблица 1

Проба	Возраст	CSL	CSQ	Качественный показатель стабильности
138a	68	104,375	129,943	438
138b		43,244	26,996	
138c		85,23	69,261	
207a	70	82,546	61,476	24039
207b		82,546	61,476	
207c		142,396	166,928	
29a	69	79,4	53,971	351
29b		161,089	178,912	
29c		161,089	178,912	
34a	67	102,019	120,426	853
34b		59,671	36,265	
34c		102,019	120,426	
65a		200	200	22
65b		180,581	190,11	
65c		180,581	190,11	
72a		181,757	191,335	11
72b		181,757	191,335	
72c		181,757	191,335	
74a		141,818	166,105	95
74b		141,649	165,886	
74c		121,016	147,092	
77a		181,073	190,646	36
77b		161,759	179,731	
77c		181,073	190,646	
80a		141,398	165,473	64
80b		161,395	179,26	
80c		141,398	165,473	
286_1a		19,165	9,629	36
286_1b		38,382	20,474	
286_1c		19,165	9,629	
287_1a		0	0	0
287_1b		0	0	
287_1c		0	0	
288_1a		0	0	0
288_1b		0	0	
288_1c		0	0	
289_1a		0	0	0
289_1b		0	0	
289_1c		0	0	
290_1a		0	0	0
290_1b		0	0	
290_1c		0	0	
291_1a		0	0	0
291_1b		0	0	
291_1c		0	0	
245_1a		0	0	124
245_1b		81,565	58,963	
245_1c		40,353	22,961	
246_1a		42,152	25,392	25
246_1b		42,152	25,392	
246_1c		42,152	25,392	

Таблица 2

Состояние	Группа	Правильно идентифицировано	Неопределенно	Ошибочно идентифицировано
Рак молочной железы	495	98,0%	1,0%	1,0%
Здоровые женщины	373	98,0%	0,5%	1,5%

Таблица 3

Способ		Прогностическая сила
Логистическая регрессия		82%
Кластерный анализ стандартной окрестности		
	Линейный	88%
	Логарифмический	90%
Способ с метапеременной		> 95%

Таблица 4

Состояние	Группа	Правильно идентифицировано	Неопределенно	Ошибочно идентифицировано
Рак яичников	51	94,1%	3,9%	0,0%
Здоровые женщины	56	96,4%	3,6%	0,0%

Таблица 5

Состояние	Группа	Правильно идентифицировано	Неопределенно	Ошибочно идентифицировано
Рак простаты	111	93,70%	0,90%	5,40%
Доброкачественная гиперплазия/гипертрофия предстательной железы	148	95,90%	0,00%	4,10%

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ диагностики рака, содержащий этапы:

а) определения концентраций по меньшей мере трех предопределенных аналитов в слепой пробе субъекта, при этом предопределенные аналиты выбирают из группы, содержащей IL-6, IL-8, VEGF, TNF α и ПСА;

б) выбора одной или более метапеременной, ассоциированной с субъектом, которая меняется в популяции, связанной с субъектом, для членов упомянутой популяции, которые, как известно, либо болеют, либо не болеют раком, при этом метапеременной является возраст;

с) преобразования концентраций аналитов в зависимости от одной или более характеристик распределения в популяции и упомянутой одной или более метапеременных для вычисления псевдоконцентрации, которая представляет каждый аналит, при этом упомянутое преобразование включает в себя коррекцию пространственного смещения;

д) сравнения псевдоконцентраций с моделью обучающей последовательности псевдоконцентраций посредством использования линейного показателя ("CSI") и q-показателя ("CSq"), определяемых для членов популяции, которые, как известно, либо болеют, либо не болеют раком;

е) определения, болен ли субъект раком или нет на основе того, попадают ли показатели CSI и CSq псевдоконцентраций субъекта в группу с наличием рака или в группу без рака на графике CSI/CSq модели обучающей последовательности.

2. Способ подготовки модели обучающей последовательности, способной диагностировать рак, содержащий этапы:

а) определения концентраций по меньшей мере трех предопределенных аналитов в обучающей последовательности проб группы субъектов, при этом предопределенные аналиты выбирают из группы, содержащей IL-6, IL-8, VEGF, TNF α и ПСА;

б) выбора метапеременной, связанной с субъектами, которая меняется в популяции, связанной с субъектами, для членов упомянутой популяции, которые, как известно, либо болеют, либо не болеют раком, при этом метапеременной является возраст;

с) преобразования концентраций аналитов в зависимости от одной или более характеристик распределения в популяции и метапеременной для вычисления псевдоконцентрации, которая представляет каждый аналит, при этом упомянутое преобразование включает в себя коррекцию пространственного смещения; и

д) подготовки модели обучающей последовательности из псевдоконцентраций посредством исполь-

зования линейного показателя ("CSI") и q-показателя ("CSq"), определенных для членов популяций, которые, как известно, либо болеют, либо не болеют раком.

3. Способ по предыдущему пункту, в котором рак выбирают из группы, состоящей из рака молочной железы, рака предстательной железы и рака легких.

4. Способ по любому из предшествующих пунктов, имеющий этап сравнения, при этом этап сравнения включает в себя способ корреляции, выбранный из группы, состоящей из способов кластеризации, поиска окрестности, регрессионного или вейвлет-анализа.

5. Способ по любому из предшествующих пунктов, имеющий этап сравнения, при этом этап сравнения дополнительно содержит использование инконгруэнтной модели обучающей последовательности.

6. Способ по любому из предшествующих пунктов, при этом способ реализуется компьютером и, необязательно, дополнительно содержит этап вывода показателя.

7. Способ по любому из предшествующих пунктов, в котором этапы преобразования, сравнения и определения повторяются со второй моделью обучающей последовательности, способной выявлять состояние отсутствия заболевания в популяции субъекта, которые частично имитируют изменения в сыов-точном анализе, но не являются состоянием наличия заболевания.

8. Способ по любому из предшествующих пунктов, которые включают в себя вторую обучающую последовательность, при этом этап определения обеспечивает оценку риска для трех состояний: состояние отсутствия заболевания, состояние отсутствия заболевания, которое частично имитирует состояние наличия заболевания, и состояние наличия заболевания.

9. Способ по любому из предшествующих пунктов, в котором математический способ для нормализации и сглаживания нерегулярностей или несмежных распределений концентраций является

натуральным логарифмом $((C_1/C_{(c \text{ или } h)}) - (C_h/C_c))^2$, где

C_1 - измеренная концентрация аналита фактического пациента;

$C_{(c \text{ или } h)}$ - скорректированная по возрасту пациента концентрация аналита данного пациента; значение корректируется в зависимости от того, находится ли пациент в состоянии отсутствия заболевания или наличия заболевания;

C_h - скорректированная по возрасту пациента средняя концентрация аналита среди пациентов без заболевания;

C_c - скорректированная по возрасту пациента средняя концентрация аналита среди пациентов с заболеванием.

10. Способ по п.9, при этом взаимосвязь между концентрациями и метапеременной охватывает характеристики распределения в популяции независимых переменных, выбираемых из группы, состоящей из степени нелинейности взаимосвязи между состояниями наличия заболевания и отсутствия заболевания, одной или более групп (гауссовыми или негауссовыми), средних групповых значений, усредненных групповых значений, медианных групповых значений и групповых значений динамического диапазона.

11. Способ по любому из предшествующих пунктов, дополнительно содержащий этап коррекции для модели обучающей последовательности, выполняемый посредством подстройки весового влияния аналитов на основании знания о характеристиках усиления или ослабления экспрессии аналитов, например, подгруппах или степени нелинейности, в процессе прогрессирования заболевания у типичного субъекта в релевантной популяции.

12. Способ по любому из предшествующих пунктов, дополнительно содержащий этап коррекции для модели обучающей последовательности, выполняемый посредством подстройки весового влияния аналитов на основании знания о нестабильностях топологии плоскости би-маркера на достаточном количестве плоскостей би-маркера, для существенного изменения показателя риска, причем такие нестабильности вызваны резкими уклонами или глубокими пиками или впадинами в топологии каждой плоскости би-маркера.

13. Способ по любому из предшествующих пунктов, дополнительно содержащий этап коррекции для модели обучающей последовательности, выполняемый посредством подстройки весового влияния аналитов на основании знания о неопределенностях анализа аналитов, например, неопределенностях, которые могут возникать при очень низких или очень высоких уровнях на кривой результатов анализа.

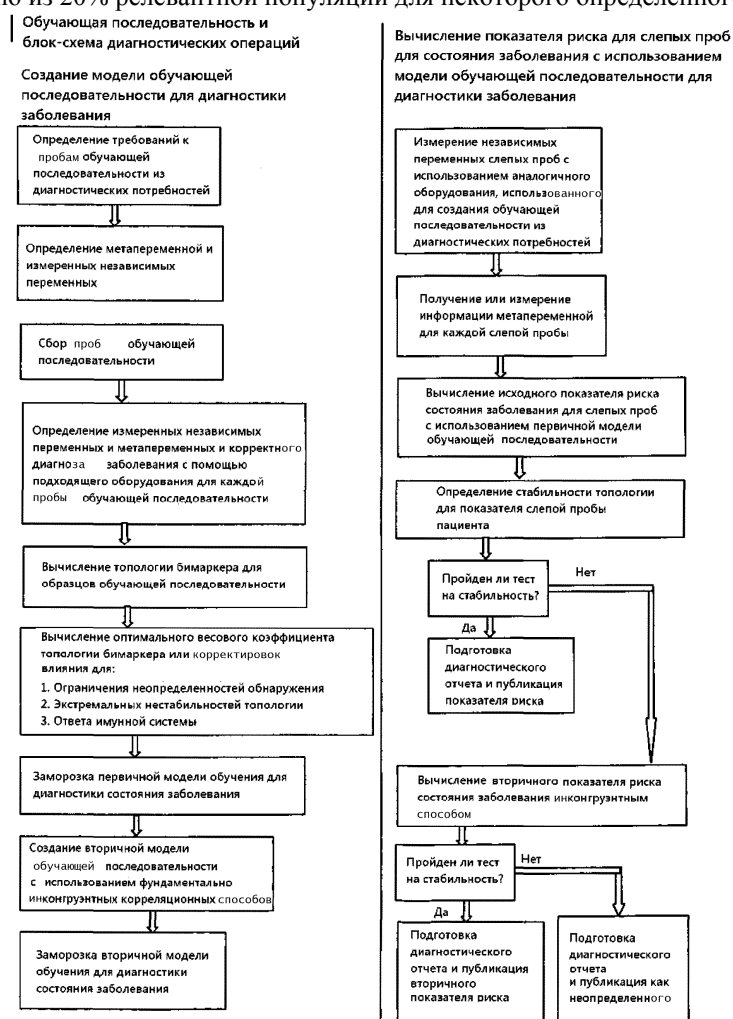
14. Способ по любому из предшествующих пунктов, дополнительно содержащий этап коррекции для модели обучающей последовательности, выполняемый посредством использования инконгруэнтной модели обучающей последовательности для коррекции индивидуальных слепых проб, которые демонстрируют нестабильности при прогнозировании исхода вследствие нестабильностей топологии на достаточном количестве плоскостей би-маркера для существенного изменения показателя риска, причем такие нестабильности вызваны резкими уклонами или глубокими пиками или впадинами в топологии каждой плоскости би-маркера.

15. Способ по любому из предшествующих пунктов, в котором базовые значения индивидуальных концентраций аналитов определяются для субъекта с течением времени, в том числе периода времени, когда субъект находится в состоянии отсутствия заболевания, вместо значения популяции.

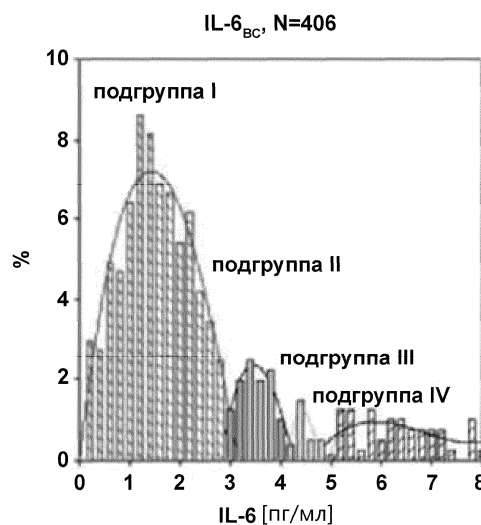
16. Способ по любому из предшествующих пунктов, в котором измеряемые аналиты представляют собой сигнальные белки низкого содержания, которые включают по меньшей мере один аналит в каждой

по меньшей мере из трех категорий, выбираемых из группы, состоящей из белков воспаления иммунной системы, ангиогенеза опухоли, клеточного апоптоза, васкуляризации и тканевых маркеров.

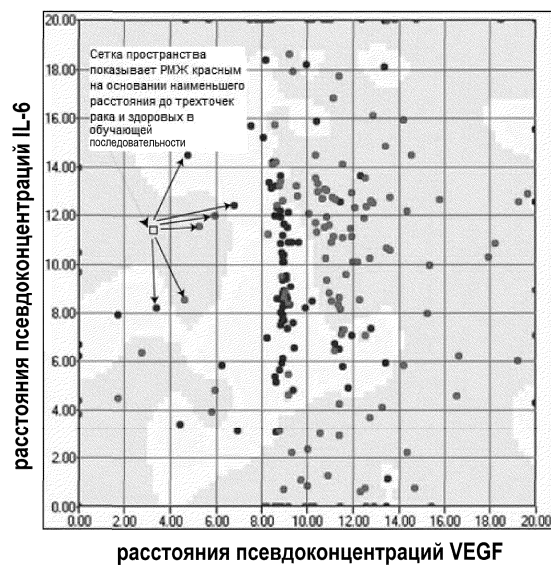
17. Способ по п.16, при этом один или более аналитов представляют собой белки крайне низкого содержания с уровнями концентрации ниже приблизительно 1 пг/мл в пробах, извлеченных, по меньшей мере, приблизительно из 20% релевантной популяции для некоторого определенного субъекта.



Фиг. 1

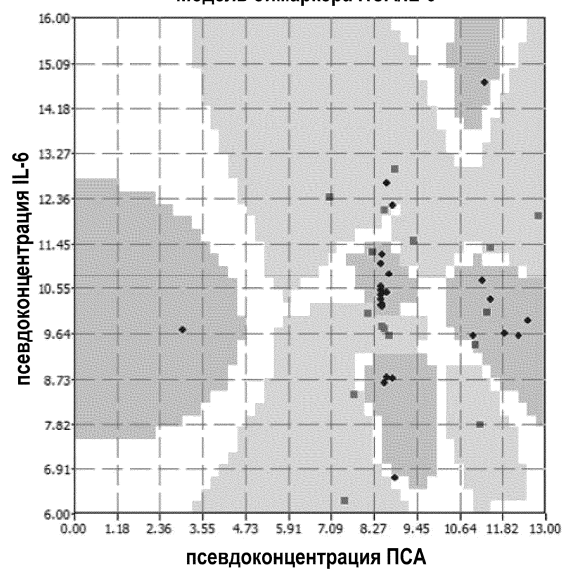


Фиг. 2



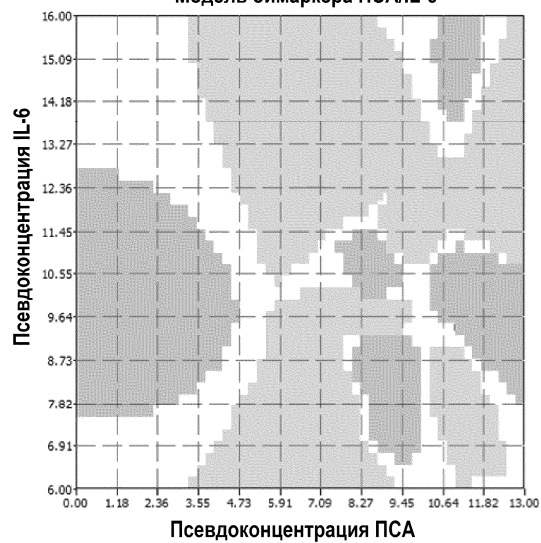
Фиг. 3

Модель бимаркера ПСА/IL-6

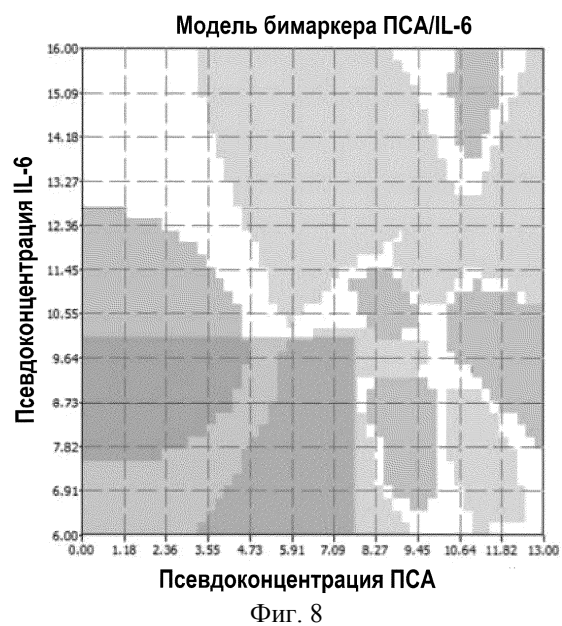
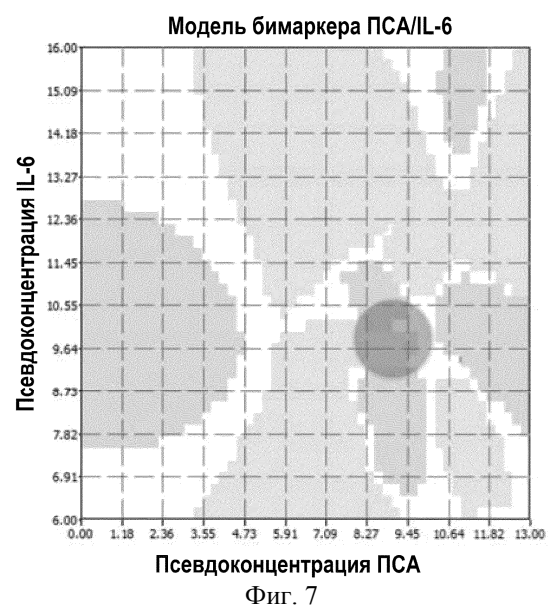
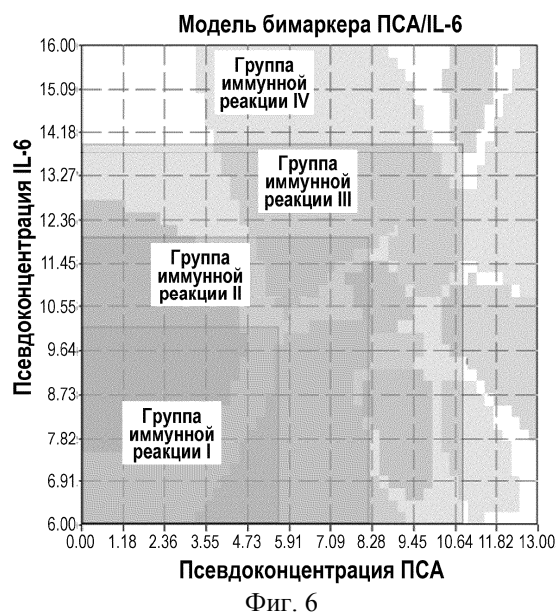


Фиг. 4

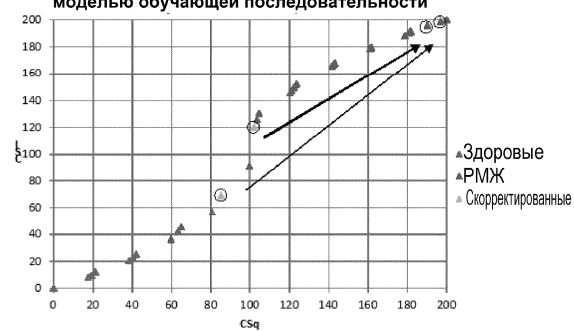
Модель бимаркера ПСА/IL-6



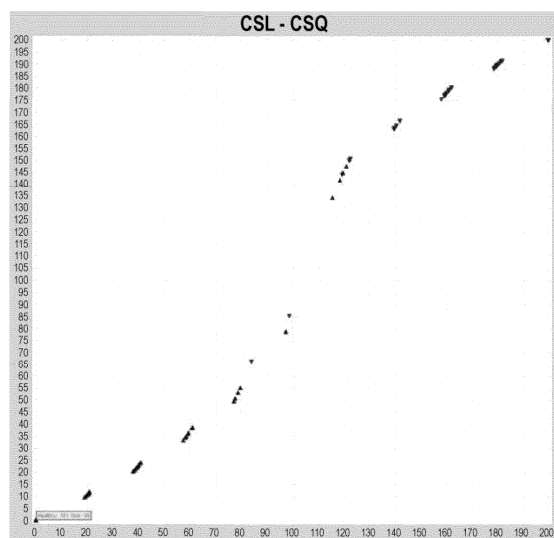
Фиг. 5



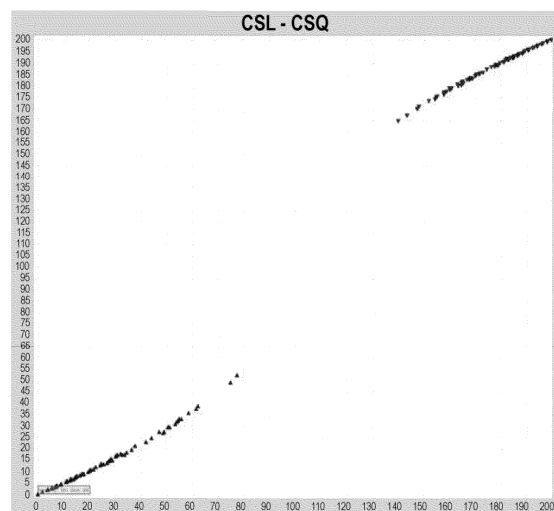
Данная фигура показывает две нестабильные точки данных, скорректированные инконгруэнтной моделью обучающей последовательности



Фиг. 9

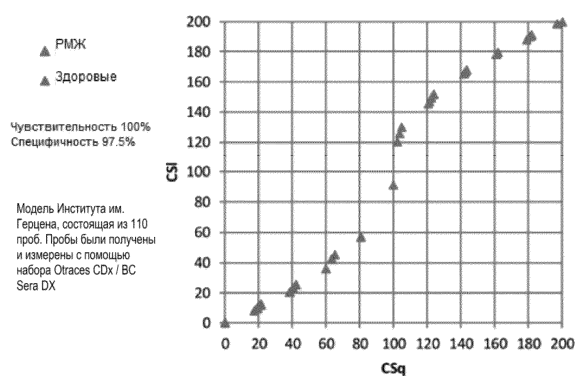


Фиг. 10

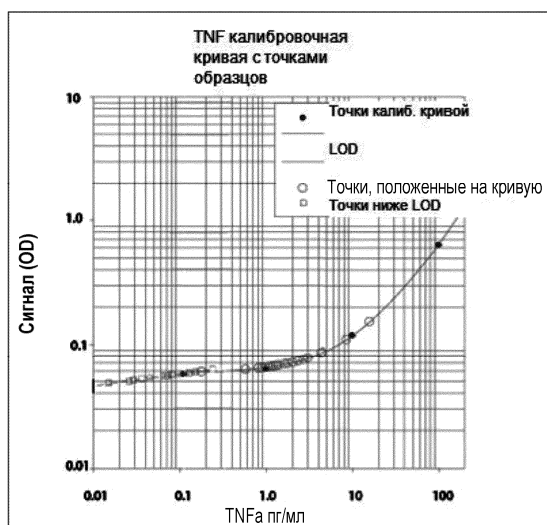


Фиг. 11

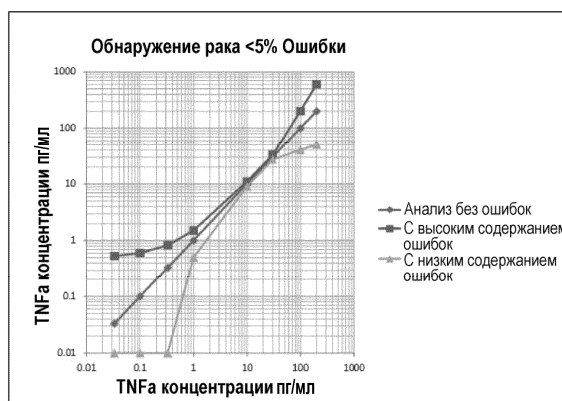
Определение показателя РМЖ для слепых проб из Института им. Герцена



Фиг. 12



Фиг. 13



Фиг. 14



Евразийская патентная организация, ЕАПВ

Россия, 109012, Москва, Малый Черкасский пер., 2