

(19)



Евразийское  
патентное  
ведомство

(21) 201991019 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки  
2020.01.22

(22) Дата подачи заявки  
2017.10.30

(51) Int. Cl. C12N 15/10 (2006.01)  
C12N 15/62 (2006.01)  
G01N 33/68 (2006.01)  
C07K 16/12 (2006.01)

(54) СПОСОБ СКРИНИНГА И ДЕТЕКЦИИ БЕЛКА

(31) 16196571.0

(32) 2016.10.31

(33) EP

(86) PCT/EP2017/077816

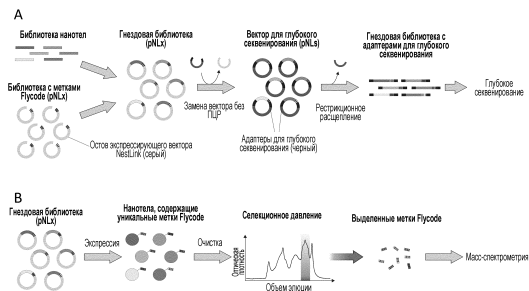
(87) WO 2018/078167 2018.05.03

(71) Заявитель:  
УНИВЕРСИТЕТ ЦЮРИХ (CH)

(72) Изобретатель:  
Зегер Маркус, Эглофф Паскаль,  
Циммерманн Иван (CH)

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

(57) Изобретение относится к способу идентификации и количественной оценки полипептида из библиотеки полипептидов. Способ включает стадии: 1 - получения библиотеки полипептидов и библиотеки меток для детекции, 2 - получения гнездовой библиотеки, содержащей полипептиды и метки для детекции, 3 - секвенирования гнездовой библиотеки, 4 - отбора члена гнездовой библиотеки на одной или более стадиях отбора, которые являются независимыми от физической связи генотип-фенотип, 5 - выделения метки для детекции из отобранного полипептида, 6 - идентификации и количественной оценки метки для детекции посредством масс-спектрометрии, 7 - получения последовательности отобранного полипептида. Изобретение также относится к коллекции полипептидов, к коллекции меток для детекции и к коллекции плазмидных векторов.



A1

201991019

201991019

A1

## ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-556443EA/026

### СПОСОБ СКРИНИНГА И ДЕТЕКЦИИ БЕЛКА

Настоящее изобретение относится к способу присоединения меток для детекции к белковой библиотеке и к последующему использованию меток для идентификации и количественного определения белков, удовлетворяющих определенным биофизическим или фармакологическим критериям.

#### Описание

##### Уровень техники

Способы скрининга белков и белкового дисплея представляют собой способы существующего уровня техники для идентификации или обогащения белков, имеющих конкретные характеристики (например, высокую аффинность связывания с молекулой-мишенью).

При скрининге, белки анализируют по одному. Это является очень трудоемким и ограниченным сравнительно небольшим количеством тестов. При скрининге по связыванию белков, например, индивидуальные кандидаты на связывание идентифицируют посредством ELISA, и положительные наилучшие кандидаты в ELISA характеризуют далее, например, их характеризуют биофизически посредством эксклюзионной хроматографии, экспериментов разворачивания, и их терапевтический потенциал тестируют *in vivo* в моделях на животных.

В способах дисплея, полные пулы белков (происходящие из библиотек) обогащают после нескольких циклов отбора. Обработка пулов позволяет огромную производительность без большой трудоемкости. Способы дисплея, такие как фаговый, рибосомный или дрожжевой дисплей, однако, требуют физической связи между фенотипом (белком) и генотипом (кодирующими его нуклеиновыми кислотами). Это серьезное ограничение для большинства анализов, поскольку физические объекты, необходимые для проведения дисплея (т.е. фаг, рибосома и кодирующая ДНК или РНК), как правило, более чем в 100 раз больше, чем фактическая связывающая молекула (например, фрагмент антитела). Это неминуемо вызывает сдвиг отбора и ограничивает возможное селекционное давление до небольшой подгруппы из всех воображимых видов селекционного

давления – только виды селекционного давления, на которые не оказывает критического влияния огромный размер частицы для дисплея, можно использовать в настоящее время (например, связывание).

На основании вышеуказанного существующего уровня техники, целью настоящего изобретения является предоставление средств и способов для идентификации индивидуальных белков, удовлетворяющих определенным биофизическим или фармакологическим критериям, из полных библиотек белков в отсутствие физической связи генотип-фенотип. Это цель достигнута посредством пунктов формулы изобретения по настоящему описанию.

#### Термины и определения

Специалисту в данной области понятно, что в настоящем описании, число, показывающее размер библиотеки, относится к разнообразию членов библиотеки. Библиотека I, которая больше чем библиотека II, соответствует библиотеке I, которая содержит более высокое количество уникальных членов библиотеки, чем библиотека II. Библиотека нуклеиновых кислот с 100000 членами может содержать несколько миллионов молекул нуклеиновой кислоты, но только каждый из 100000 отдельных членов библиотеки характеризуется последовательностью нуклеиновой кислоты, уникальной в указанной библиотеке. Подобным образом, библиотека полипептидов с 1000 членов может содержать миллионы молекул полипептидов, но только 1000 уникальных членов библиотеки полипептидов. Выражение «один член библиотеки» относится к одному конкретному члену библиотеки, который может присутствовать во множестве идентичных копий.

В контексте настоящего описания, выражение «две последовательности нуклеиновой кислоты находятся в рамке считывания» означает, что количество пар оснований между последним кодоном первой последовательности нуклеиновой кислоты и первым кодоном второй последовательности нуклеиновой кислоты кратно трем.

В контексте настоящего описания, выражения «полипептид является ассоциированным с меткой для детекции», соответственно «полипептид/метка для детекции являются ассоциированными с

аффинной меткой», означает, что оба вышеупомянутых члена содержатся в одной первичной аминокислотной последовательности, т.е. одной непрерывной полипептидной цепи. В частности, указанная метка для детекции и указанный полипептид могут быть разделены одной или более аминокислотами). Указанная метка для детекции и указанная аффинная метка могут также быть разделены одной или более аминокислотами.

В контексте настоящего описания, термин «разделяемый элемент» относится к пептидной последовательности, поддающейся разделению посредством химических средств или посредством ферментных средств, например, посредством протеаз. Протеазы могут являться специфическими для последовательности (например, тромбин) или иметь ограниченную специфичность для последовательности (например, трипсин). Разделяемые элементы I и II могут также содержаться внутри аминокислотной последовательности метки для детекции или полипептида, в частности, в случаях, когда последняя аминокислота метки для детекции или полипептида представляет собой K или R.

В контексте настоящего описания, термин «аффинная метка» относится к группе, присоединенной к полипептиду для обеспечения очистки указанного полипептида из биохимической смеси. Очистка (аффинная очистка) основана на высоко специфическом взаимодействии (с константой диссоциации  $\leq 10^{-5}$ ) между аффинной меткой и партнером по связыванию аффинной метки. Аффинные метки могут состоять из аминокислотной последовательности, или могут содержать аминокислотную последовательность, к которой химическая группа присоединена посредством посттрансляционной модификации. В качестве неограничивающего примера, аффинная метка выбрана из группы, содержащей His-метку, CBP-метку (CBP: связывающий кальмодулин белок), CYD-метку (CYD: ковалентный, но способный к диссоциации пептид NorpD), Strep-метку, StrepII-метку, FLAG-метку, HPC-метку (HPC: тяжелая цепь белка C), GST-метку (GST: глутатион-S-трансфераза), Avi-метку, биотинилированную метку, Muc-метку, 3xFLAG-метку и MBP-метку (MBP: связывающий мальтозу белок).

Дополнительные примеры аффинных меток можно обнаружить в Kimple et al., Curr Protoc Protein Sci. 2013 Sep 24;73:Unit 9.9.

В контексте настоящего описания, термин «глубокое секвенирование» относится к параллельному секвенированию нескольких тысяч различных молекул нуклеиновой кислоты с перекрытием  $\geq 5x$ , в частности  $\geq 40x$ . Термин «перекрытие» относится к количеству раз, когда данный нуклеотид считывается в ходе процесса глубокого секвенирования в среднем.

В контексте настоящего описания, термин антитело используют в его значении, известном в области клеточной биологии и иммунологии. Полноразмерное антитело представляет собой гликопротеин, содержащий по меньшей мере две тяжелых (H) цепи и две легких (L) цепи, соединенные между собой дисульфидными связями. Каждая тяжелая цепь состоит из переменной области тяжелой цепи (VH) и константной области тяжелой цепи (CH). Каждая легкая цепь состоит из переменной области легкой цепи (кратко обозначенной в настоящем описании как VL) и константной области легкой цепи (CL). Переменные области тяжелых и легких цепей содержат связывающий домен, который взаимодействует с антигеном. Константные антител могут опосредовать связывание иммуноглобулина с тканями или факторами хозяина, включая различные клетки иммунной системы (например, эффекторные клетки) и первый компонент классической системы комплемента.

В контексте настоящего описания, термин «нанотело» относится к «однодоменному антителу», т.е. фрагменту антитела, состоящему из одиночного переменного домена антитела. Нанотело является способным избирательно связываться со специфическим антигеном. Оно имеет молекулярную массу только 12-15 кДа (Harmsen et al., Appl. Microbiol. Biotechnol. 77 (1): 13-22). Обычно, нанотела получают посредством иммунизации дромедаров, верблюдов, лам, альпак или акул желательным антигеном и последующего выделения мРНК, кодирующей антитела из тяжелой цепи. Нанотела можно также получать из обычных мышинных или человеческих IgG с четырьмя цепями.

В контексте настоящего описания, термин «синтетело» относится к синтетическому нанотелу. Синтетела не получают посредством иммунизации с использованием антигена, но отбирают *in vitro* из синтетической библиотеки.

В контексте настоящего описания, термин «обогащение» относится к способу увеличения относительного количества определенного соединения в смеси соединений.

В контексте настоящего описания, термин «библиотека с меткой Glycode» относится к библиотеке аминокислотных последовательностей по изобретению, содержащей множество вариантов последовательностей.

В контексте настоящего описания, термин «NestLink» относится к способу, в котором метки для детекции присоединяют к библиотеке белков. Затем, метки используют для идентификации и количественной оценки индивидуальных белков, удовлетворяющих определенным биофизическим или фармакологическим критериям, в библиотеке. NestLink объединяет ключевые преимущества способов скрининга и дисплея.

В контексте настоящего описания, термин «уровень гидрофобности» относится к прогнозируемому значению, характеризующему пептид. Уровень гидрофобности рассчитывают способом, описанным в Krokhin et al., Mol Cell Proteomics. 2004 Sep;3(9):908-19, в соответствии с формулами:

$$H = K_L * (\sum R_c + 0,42R_{cNT}^1 + 0,22R_{cNT}^2 + 0,05R_{cNT}^3), \text{ если } H < 38$$

и

$$H = K_L * (\sum R_c + 0,42R_{cNT}^1 + 0,22R_{cNT}^2 + 0,05R_{cNT}^3) - 0,3(K_L * (\sum R_c + 0,42R_{cNT}^1 + 0,22R_{cNT}^2 + 0,05R_{cNT}^3) - 38), \text{ если } H \geq 38;$$

если  $H < 38$ ,  $H_{\text{конечный}} = H$ ;

если  $H \geq 38$ ,  $H_{\text{конечный}} = H - 0,3 * (H - 38)$ ;

где  $H_{\text{конечный}}$  представляет собой уровень гидрофобности, и  $R_c$  представляют собой характеристики коэффициентов удержания для типов аминокислот в соответствии со следующей таблицей:

	$R_c$
Trp	11
Phe	10,5
Leu	9,6

Ile	8,4
Met	5,8
Val	5
Tyr	4
Ala	0,8
Thr	0,4
Pro	0,2
Glu	0
Asp	-0,5
Cys	-0,8
Ser	-0,8
Gln	-0,9
Gly	-0,9
Asn	-1,2
Arg	-1,3
His	-1,3
Lys	-1,9

$R_{CNT}$  аминокислоты X определяют как:

$$R_{CNT}^X = (\sum R_c / 20) - R_c^X$$

N соответствует номеру остатка метки для детекции, начиная с 1 от N-конца.  $K_L$  определяют как:

$$\text{если } N < 10, K_L = 1 - 0,027 * (10 - N)$$

$$\text{если } N > 20, K_L = 1 - 0,014 * (N - 20),$$

в ином случае  $K_L = 1$ .

Аминокислотные последовательности приведены от амино- до карбокси-конца. Заглавные буквы в положениях последовательности относятся к L-аминокислотам в однобуквенном коде (Stryer, Biochemistry, 3rd ed. p. 21).

#### Подробное описание изобретения

##### *Способ отбора полипептида из библиотеки полипептидов*

В соответствии с первым аспектом, изобретение относится к способу отбора полипептида из библиотеки полипептидов. Способ включает следующие стадии:

а. Получают первую библиотеку нуклеиновых кислот. Каждый член первой библиотеки нуклеиновых кислот содержит кодирующую полипептид последовательность, кодирующую член первой библиотеки полипептидов. Каждый член первой библиотеки нуклеиновых кислот отличается от любого другого члена первой библиотеки нуклеиновых кислот.

b. Получают вторую библиотеку нуклеиновых кислот. Вторая библиотека содержит множество членов. Каждый член содержит кодирующую метку последовательность, кодирующую метку для детекции. Каждая метка для детекции имеет следующие характеристики:

i. Метка характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой второй библиотекой нуклеиновых кислот.

ii. Метка характеризуется молекулярной массой между 200 и 5000 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 500 и 2500 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 900 и 2200 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 903 и 2180 Да.

iii. Метка содержит первый разделяемый элемент.

Спецификация массы, приведенная в ii, относится к массе метки после ее выделения, т.е. после разделения первого разделяемого элемента.

c. Кодирующую полипептид последовательность, содержащуюся в члене первой библиотеки нуклеиновых кислот, вставляют в член указанной второй библиотеки нуклеиновых кислот. Таким образом, получают библиотеку меченых нуклеиновых кислот, кодирующую библиотеку меченых полипептидов. Каждый член библиотеки меченых полипептидов содержит полипептид и метку для детекции. Метка для детекции отделена от полипептида первым разделяемым элементом.

Библиотека меченых полипептидов представляет собой «гнездовую библиотеку», поскольку кодирующие полипептиды последовательности первой библиотеки нуклеиновых кислот «гнездятся» внутри членов второй библиотеки нуклеиновых кислот. Вторая библиотека нуклеиновых кислот является в несколько раз большей, чем библиотека меченых нуклеиновых кислот. Библиотека меченых нуклеиновых кислот является в несколько раз большей, чем первая библиотека нуклеиновых кислот.



В библиотеке меченых нуклеиновых кислот, каждая кодирующая полипептид последовательность первой библиотеки нуклеиновых кислот является ассоциированной с кодирующей метку последовательностью второй библиотеки нуклеиновых кислот. Ассоциация присутствует в рамке считывания. Кодирующая полипептид последовательность вставлена в положение, где она может подвергаться транскрипции и последующей трансляции в подходящем хозяине после того как член библиотеки меченых нуклеиновых кислот вводят в подходящего хозяина. Введение в бактериальные клетки можно осуществлять посредством трансформации. Введение в не относящиеся к бактериальным клетки можно осуществлять посредством трансфекции. Специалисту в данной области известно, что хозяин не обязательно является необходимым для трансляции: можно использовать также способы трансляции *in vitro*. Обзоры бесклеточных систем экспрессии см. в Rosenblum, FEBS Lett. 2014 Jan 21; 588(2):261-8 и Zemella, ChemBioChem. 2015 Nov; 16(17):2420-31. Кодирующая полипептид последовательность и кодирующая метку последовательность экспрессируются внутри одной и той же экспрессированной последовательности.

Библиотека меченых нуклеиновых кислот содержит все кодирующие полипептиды последовательности первой библиотеки нуклеиновых кислот, но только подгруппу кодирующих метку последовательностей второй библиотеки нуклеиновых кислот. Каждый член библиотеки меченых нуклеиновых кислот содержит только одну кодирующую полипептид последовательность и одну кодирующую метку последовательность. Каждая кодирующая метку последовательность содержится только в одном члене библиотеки меченых нуклеиновых кислот. Иными словами, каждая кодирующая метку последовательность является уникальной в библиотеке меченых нуклеиновых кислот. Каждая кодирующая полипептид последовательность может, однако, содержаться в нескольких членах библиотеки меченых нуклеиновых кислот (избыточное мечение). В конкретных вариантах осуществления, каждая кодирующая полипептид последовательность первой библиотеки нуклеиновых кислот является ассоциированной по меньшей мере с одной кодирующей метку последовательностью второй библиотеки

нуклеиновых кислот. В конкретных вариантах осуществления, каждая кодирующая полипептид последовательность первой библиотеки нуклеиновых кислот является ассоциированной по меньшей мере с двумя кодирующую метку последовательностями второй библиотеки нуклеиновых кислот. В конкретных вариантах осуществления, каждая кодирующая полипептид последовательность первой библиотеки нуклеиновых кислот является ассоциированной по меньшей мере с пятью различными кодирующими метку последовательностями второй библиотеки нуклеиновых кислот. В конкретных вариантах осуществления, каждая кодирующая полипептид последовательность первой библиотеки нуклеиновых кислот является ассоциированной по меньшей мере с десятью различными кодирующими метку последовательностями второй библиотеки нуклеиновых кислот. В конкретных вариантах осуществления, каждая кодирующая полипептид последовательность первой библиотеки нуклеиновых кислот является ассоциированной в среднем с 10-30 различными кодирующими метку последовательностями второй библиотеки нуклеиновых кислот. В конкретных вариантах осуществления, каждая кодирующая полипептид последовательность первой библиотеки нуклеиновых кислот является ассоциированной в среднем с приблизительно двадцатью различными кодирующими метку последовательностями второй библиотеки нуклеиновых кислот.

d. Множество последовательностей нуклеиновых кислот получают из библиотеки меченых нуклеиновых кислот. В частности, последовательность нуклеиновой кислоты получают для каждого члена библиотеки меченых нуклеиновых кислот. Каждая из указанного множества последовательностей нуклеиновых кислот содержит кодирующую полипептид последовательность и кодирующую метку последовательность.

На основании информации о секвенировании, полученной на стадии d, создают базу данных. База данных содержит последовательности всех полипептидов и всех меток для детекции, содержащихся в библиотеке меченых полипептидов. Специалисту в данной области понятно, что база данных может не содержать каждый отдельный член библиотеки меченых нуклеиновых кислот, по техническим причинам. Последовательности могут находиться в

форме последовательностей нуклеиновой кислоты и/или аминокислотных последовательностей. База данных содержит информацию о том, какая подгруппа кодирующих метку последовательностей второй библиотеки нуклеиновых кислот содержится в библиотеке меченых нуклеиновых кислот. База данных содержит также информацию о том, какая кодирующая метку последовательность, или соответственно, какие кодирующие метку последовательности, ассоциированы с данной кодирующей полипептид последовательностью.

e. Паттерн фрагментации по масс-спектрометрии прогнозируют для каждой метки для детекции, кодируемой кодирующей метку последовательностью, полученной на стадии d. Специалисту в данной области понятно, что паттерн фрагментации прогнозируют для *выделенной* метки для детекции, то есть, для метки для детекции, освобожденной от ассоциированного с ней полипептида посредством разделения первого разделяемого элемента. Специалисту в данной области понятно, что прогнозирование паттерна фрагментации включает также прогнозирование общей массы выделенной метки для детекции.

f. Библиотеку меченых полипептидов экспрессируют из библиотеки меченых нуклеиновых кислот. Вследствие использования способа избыточного мечения, описанного на стадии c, библиотека меченых полипептидов может содержать данный член указанной первой библиотеки полипептидов, меченый несколькими различными метками для детекции (но только одной меткой на молекулу). Избыточное мечение является предпочтительным, поскольку оно облегчает однозначную детекцию члена первой библиотеки полипептидов посредством множества меток для детекции и минимизирует потенциальные влияния меток для детекции на биофизические свойства членов библиотеки меченых полипептидов. Избыточность представляет собой добавление, необходимое по техническим причинам: некоторые метки для детекции могут не быть детектированы, поскольку они уменьшают уровни экспрессии, они теряются в ходе подготовки образца или они не элюируются внутри окна гидрофобности обращеннофазовой колонки, анализируемого посредством масс-спектрометрии.

g. Член библиотеки меченых полипептидов отбирают на стадии отбора, получая отобранный полипептид. Эта стадия отбора включает выделение тех членов библиотеки меченых полипептидов, которые удовлетворяют определенным биохимическим критериям. Иными словами, селекционное давление прилагают к библиотеке меченых полипептидов. Это селекционное давление должно приводить к физическому разделению белков, так что физически разделенные подпулы получают и собирают. Ключевым преимуществом способа по изобретению является то, что диапазон возможных критериев отбора намного выше, чем в способах белкового дисплея. В качестве неограничивающего примера, критерии можно выбирать из группы критериев, включающей способность связываться с молекулой-мишенью с определенной аффинностью, стабильность полипептида в определенных условиях, определенное поведение при агрегации (например, преобладающая встречаемость в форме мономера) в определенных условиях, устойчивость по отношению к протеазам, способность к проникновению в ткани, быстрое или медленное выведение из кровотока, способность к проникновению через гематоэнцефалический барьер и способность к накоплению в опухолях.

h. Первый разделяемый элемент разрезают. Таким образом, метку для детекции отделяют от отобранного полипептида и получают выделенную метку для детекции.

i. Выделенную метку для детекции идентифицируют и количественно оценивают следующим образом:

i. Паттерн фрагментации выделенной метки для детекции регистрируют посредством масс-спектрометрии. Паттерн фрагментации предоставляет информацию относительно массы и гидрофобности фрагментов выделенной метки для детекции. Паттерн фрагментации предоставляет информацию относительно аминокислотной последовательности выделенной метки для детекции.

ii. Массу и паттерн фрагментации, полученные на стадии i, приводят в соответствие с массами и паттернами фрагментации, прогнозированными на стадии e. Таким образом, идентифицируют выделенную метку для детекции. Комбинация информации, полученной посредством масс-спектрометрии, с информацией, полученной

посредством секвенирования библиотеки меченых нуклеиновых кислот, позволяет однозначную идентификацию данной метки для детекции.

Точность совпадения прогнозируемого и зарегистрированного паттернов фрагментации может быть оценена в баллах и позволяет ранжирование членов библиотеки полипептидов. Сравнение ранжирования полипептидов для различных условий отбора можно использовать в качестве относительного показателя различных характеристик полипептидов (например, скорости диссоциации, распределения в тканях, специфического для конформации связывания и т.д.). Сравнение является наиболее точным для членов библиотеки избыточно меченых полипептидов, где различия в эффективности регистрации паттерна фрагментации индивидуальных меток усредняют.

Показатель точности совпадения прогнозируемых и зарегистрированных паттернов фрагментации можно использовать в качестве показателя относительных количеств членов библиотеки полипептидов после отбора. Относительные количества является наиболее точными для членов библиотеки избыточно меченых полипептидов для членов библиотеки избыточно меченых полипептидов, где различия в эффективности регистрации паттерна фрагментации индивидуальных меток усредняют.

j. Последовательность нуклеиновой кислоты, содержащую кодирующую метку последовательность, кодирующую метку для детекции, идентифицированную на стадии i, выбирают из множества последовательностей нуклеиновых кислот, полученных на стадии d. Таким образом, идентифицируют член библиотеки меченых полипептидов, ассоциированный с меткой для детекции, идентифицированной на стадии i.

Специалисту в данной области понятно, что стадии g-j проводят для ряда различных членов указанной библиотеки меченых полипептидов параллельно. Пул нескольких полипептидов, удовлетворяющих определенным критериям, отбирают на стадии g, и все из этих полипептидов идентифицируют посредством анализа масс-спектрометрии их меток для детекции. Специалисту в данной

области понятно, что по техническим причинам, не каждый отдельный полипептид можно идентифицировать на этой стадии.

Анализ масс-спектрометрии, проведенный на стадии i, является количественным, таким образом, способ по изобретению позволяет не только идентификацию полипептида, но также количественную оценку количества этого полипептида в образце.

Для обеспечения избыточного и уникального мечения, является важным, чтобы

i) первая библиотека имела ограниченный, определенный объем. В конкретных вариантах осуществления, первая библиотека нуклеиновых кислот имеет размер 5-100000. В конкретных вариантах осуществления, первая библиотека нуклеиновых кислот имеет размер 100-50000. В конкретных вариантах осуществления, первая библиотека нуклеиновых кислот имеет размер 500-5000.

ii) вторая библиотека нуклеиновых кислот имеет размер  $10^3$ - $10^{11}$ , в частности,  $10^5$ - $10^{10}$ , более конкретно,  $10^6$ - $10^9$ , даже более конкретно, приблизительно  $10^8$ , до стадии вставки первой библиотеки

iii) после стадии вставки, выбранная подгруппа из плазмид с множеством комбинаций полипептид/метка, составляет по меньшей мере 3x, в частности, по меньшей мере 5x, более конкретно, по меньшей мере 15x, даже более конкретно, по меньшей мере 25x количество членов указанной первой библиотеки нуклеиновых кислот.

iv) выбранная подгруппа из плазмид с множеством комбинаций полипептид/метка, составляет менее 50%, в частности, менее 5%, более конкретно, менее 0,5%, даже более конкретно, менее 0,05% от количество членов указанной второй библиотеки нуклеиновых кислот.

Размер библиотеки можно контролировать посредством стадии ограничения разнообразия перед стадией a, в которых первую библиотеку отбирают как подгруппу из предварительной библиотеки большего размера.

Способ по изобретению позволяет анализ библиотек белков в отсутствие физической связи генотип-фенотип, необходимой для способов белкового дисплея. Это исключает недостаток наличия

больших физических объектов (например, фага или рибосомы и кодирующей ДНК или РНК), присоединенных к членам библиотеки белков. Можно проводить скрининг полноразмерных библиотек белков в форме пулов по выбранным критериям, вместо тестирования индивидуальных белков, как обычно в случае скрининга белков. Однако, даже несмотря на переработку целых пулов белков, считывание является сходным со скринингами, как если бы каждый отдельный белок характеризовали индивидуально. Это имеет особенную важность в области разработки связывающих белков (лекарственных средств, диагностических средств, исследовательских инструментов и т.д.). Ряд характеристик белков можно анализировать для тысяч кандидатов за один раз. Иллюстративным вопросом может являться следующий: Какие кандидаты на связывающий член являются стабильными, растворимыми и мономерными?

Способ по изобретению позволяет задавать соответствующий вопрос непосредственно в начале цепи разработки белкового лекарственного средства: «Какой связывающий член имеет наибольший терапевтический потенциал *in vivo*?» Вопросами относительно терапевтического потенциала являются следующие: Какой связывающий член переживает жесткие условия в кишечнике при пероральном введении? Какой связывающий член пересекает гематоэнцефалический барьер? Какой связывающий член имеет оптимальные свойства выведения из крови посредством почек? Какой связывающий член, из тысяч, имеет хорошее проникновение в ткани для соответствующей ткани?

В конкретных вариантах осуществления, метка для детекции характеризуется уровнем гидрофобности между -27 и 128. В конкретных вариантах осуществления, метка для детекции характеризуется уровнем гидрофобности между -1 и 70. Уровень гидрофобности относится к массе метки для детекции после ее выделения, т.е. после разделения первого разделяемого элемента. Уровень гидрофобности не включает ассоциированную аффинную метку.

В конкретных вариантах осуществления, член библиотеки меченых полипептидов является ассоциированным с аффинной меткой.

Такая аффинная метка может облегчать очистку отобранного члена библиотеки меченых полипептидов и/или собственно метки для детекции перед масс-спектрометрией. Аффинная метка и член библиотеки меченых полипептидов содержатся внутри одной первичной аминокислотной последовательности. Каждый член библиотеки меченых полипептидов содержит полипептид и метку для детекции. Аффинная метка может являться ассоциированной либо с полипептидом, либо с меткой для детекции.

В конкретных вариантах осуществления, аффинная метка выбрана из группы, содержащей His-метку, CBP-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, HPC-метку, GST-метку, Avi-метку, биотинилированную метку, Мус-метку, 3xFLAG-метку и MBP-метку.

В конкретных вариантах осуществления, метка для детекции является ассоциированной с аффинной меткой. В этих случаях, аффинная метка локализована на С-конце метки для детекции. Это расположение имеет то дополнительное преимущество, что метка для детекции защищена от деградации пептидазами и обеспечивает то, что только недеградированные полипептиды, ассоциированные с полными метками для детекции, выделяют в ходе очистки белка. Специалисту в данной области понятно, что выражение «аффинная метка локализована на С-конце метки для детекции» не обязательно подразумевает, что аффинная метка локализована непосредственно на С-конце метки для детекции, но может присутствовать линкер из нескольких аминокислот, разделяющий аффинную метку и метку для детекции.

В конкретных вариантах осуществления, аффинная метка отделена от указанной метки для детекции вторым разделяемым элементом, и указанный второй разделяемый элемент разрезают перед стадией *i*. Таким образом, только метку для детекции без ассоциированной аффинной метки анализируют посредством масс-спектрометрии.

Спецификации массы и паттерна фрагментации метки для детекции относятся к массе и паттерну фрагментации метки после ее отделения от ассоциированного полипептида и аффинной метки, т.е. после разделения первого и второго разделяемых элементов.



Специалисту в данной области понятно, что в случаях, когда метку для детекции не освобождают от ассоциированной аффинной метки перед масс-спектрометрией, это может влиять на результаты анализа масс-спектрометрии. Если все метки для детекции ассоциированы с одной и той же аффинной меткой, изменения массы и паттерна фрагментации можно учитывать, таким образом, все еще является возможной идентификация метки для детекции, хотя и не настолько эффективно и четко, как в случаях, когда метку для детекции отделили от аффинной метки посредством разделения второго разделяемого элемента.

В конкретных вариантах осуществления, аффинная метка представляет собой His-метку.

В конкретных вариантах осуществления, стадия h включает анализ выделенной метки для детекции посредством жидкостной хроматографии в сочетании с масс-спектрометрией с ионизацией электрораспылением (LC-MS). В конкретных вариантах осуществления, эта стадия включает жидкостную обращеннофазовую хроматографию. выделенные метки для детекции разделяют в соответствии с их гидрофобностью посредством обращеннофазовой хроматографии для уменьшения сложности образца. Затем их массу и паттерн фрагментации пептида регистрируют посредством масс-спектрометрии.

В конкретных вариантах осуществления, стадия d включает секвенирование полной меченой экспрессирующей библиотеки с перекрытием  $\geq 5x$ . В конкретных вариантах осуществления, стадия d включает глубокое секвенирование меченой экспрессирующей библиотеки.

В конкретных вариантах осуществления, стадия d включает вставку кодирующей полипептид последовательности и кодирующей метку последовательности, содержащихся в библиотеке меченых нуклеиновых кислот, вместе в вектор для секвенирования. Глубокое секвенирование обычно включает стадию амплификации ПЦР. Авторы настоящего изобретения отметили, что амплификация ПЦР приводит к значительному количеству событий рекомбинации между фрагментами генов членов меченой библиотеки. Таким образом, они

сконструировали набор плазмид для глубокого секвенирования, который позволяет присоединение элементов последовательности, необходимым для глубокого секвенирования, посредством рестрикционного расщепления и лигирования, таким образом, исключая необходимость амплификации ПЦР гнездовой библиотеки перед глубоким секвенированием.

В конкретных вариантах осуществления, выделенная метка для детекции состоит из 5-30 непрерывных аминокислот, и содержит одну и только одну аминокислоту, имеющую положительно заряженную боковую цепь. В конкретных вариантах осуществления, выделенная метка для детекции состоит из 7-21 непрерывных аминокислот и содержит одну и только одну аминокислоту, имеющую положительно заряженную боковую цепь. В конкретных вариантах осуществления, выделенная метка для детекции состоит из 11-15 непрерывных аминокислот и содержит одну и только одну аминокислоту, имеющую положительно заряженную боковую цепь.

В конкретных вариантах осуществления, аминокислота, имеющая положительно заряженную боковую цепь, локализована на С-конце выделенной метки для детекции. В конкретных вариантах осуществления, аминокислота, имеющая положительно заряженную боковую цепь, выбрана из аргинина (R) и лизина (K). В конкретных вариантах осуществления, аминокислота, имеющая положительно заряженную боковую цепь, представляет собой аргинин (R), локализованный на С-конце выделенной метки для детекции.

Специалисту в данной области понятно, что, в дополнение к аминокислоте, имеющей положительно заряженную боковую цепь, выделенная метка для детекции несет другой положительный заряд при нейтральном pH, представляющий собой первичный амин на N-конце выделенной метки для детекции.

В конкретных вариантах осуществления, выделенная метка для детекции содержит элемент последовательности I, выбранный из коллекции элементов последовательности I, где указанный элемент последовательности I состоит из 5-10, в частности, 7 аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R.

В конкретных вариантах осуществления, одна и только одна аминокислота, имеющая положительно заряженную боковую цепь, локализована на С-конце выделенной метки для детекции, и остальные аминокислоты независимо выбраны из А, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R. В конкретных вариантах осуществления, одна и только одна аминокислота, имеющая положительно заряженную боковую цепь, представляет собой R, локализованный на С-конце выделенной метки для детекции.

Выделенные метки для детекции оптимально поддаются детекции посредством масс-спектрометрии, в частности посредством LC-MS (жидкостной обращеннофазовой хроматографии в сочетании с ESI-MS). Аминокислоты C и M исключены из дизайна метки для детекции, поскольку они подвержены окислению. Аминокислоты K, R и H исключены из элемента последовательности I, поскольку они могут добавлять дополнительную аминокислоту с положительно заряженной боковой цепью к метке, что является нежелательным, поскольку метка будет нести дополнительный заряд во время детекции ESI-MS и выпадает из оптимального диапазона детекции. K и R могут добавлять дополнительные участки расщепления трипсином в последовательность метки, что является нежелательным.

Добавление K в аминокислотную последовательность метки для детекции добавляет другой первичный амин, что может затруднять мечение метки для детекции изобарическими метками для относительной и абсолютной количественной оценки посредством масс-спектрометрии с использованием химических реакций NHS.

В конкретных вариантах осуществления, выделенная метка для детекции содержит

а. элемент последовательности I, где элемент последовательности I состоит из 5-10, в частности 7, непрерывных аминокислот, независимо друг от друга выбранных из А, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R; и

б. элемент последовательности II, выбранный из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

В конкретных вариантах осуществления, выделенная метка для детекции состоит из

а. элемента последовательности III: GS;

б. элемента последовательности I, где элемент последовательности I состоит из 5-10, в частности 7, непрерывных аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R; и

с. элемента последовательности II, выбранного из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

Порядок элементов последовательности от N-конца к C-концу представляет собой: элемент последовательности III, элемент последовательности I, элемент последовательности II. Эти метки для детекции попадают в диапазон масс между 903 и 2180 Да, который является оптимальным для чувствительной детекции посредством ESI-MS. Выделенные метки несут два положительных заряда при физиологическом pH и ниже, а именно, R на C-конце и N-концевой первичный амин. Положительный заряд на C-конце выделенной метки для детекции облегчает ионизацию метки для масс-спектрометрической детекции и действует как уникальный участок расщепления трипсином. Пептиды с C-концевыми остатками аргинина или лизина особенно хорошо поддаются детекции посредством масс-спектрометрии (обеспечивающие преимущества свойства ионизации). В каждой выделенной метке для детекции N-концевой амин представляет собой только первичный амин, который используют для присоединения по аминогруппе посредством химических реакций NHS. Это позволяет присоединение меток для проведения количественной масс-спектрометрии, например, iTRAQ (изобарических меток для относительной и абсолютной количественной оценки). Метки для детекции конструируют для демонстрации диапазона вариантов гидрофобности, идеально подходящего для разделения пептидов посредством стандартных колонок для обращеннофазовой хроматографии.

В конкретных вариантах осуществления, все элементы последовательности I, содержащиеся в первой библиотеке нуклеиновых кислот, совместно составляют коллекцию элементов последовательности I. Внутри коллекции элементов

последовательности I, каждая аминокислота встречается с частотой, указанной в таблице 1.

Таблица 1

A	S	T	N	Q	D	E	V	L	F	Y	W	G	P	Всего
18	6	12	1	1	11	11	12	2	1	4	1	8	12	100

В конкретных вариантах осуществления, один из указанного первого и/или указанного второго разделяемого элемента представляет собой или содержит последовательность узнавания протеазой. В конкретных вариантах осуществления, указанный первый и указанный второй разделяемые элементы оба представляют собой или содержат последовательность узнавания протеазой.

В конкретных вариантах осуществления, первый разделяемый элемент представляет собой или содержит последовательность узнавания тромбином, и/или второй разделяемый элемент представляет собой или содержит последовательность узнавания трипсином.

#### *Коллекция полипептидов*

В соответствии с вторым аспектом, настоящее изобретение относится к коллекции полипептидов. Каждый член коллекции полипептидов является ассоциированным с меткой для детекции. В конкретных вариантах осуществления, каждый член коллекции полипептидов является ассоциированным по меньшей мере с одной меткой для детекции. Выражение «ассоциированный по меньшей мере с одной меткой для детекции» относится к тому факту, что каждый член коллекции полипептидов может являться ассоциированным с более, чем одной меткой для детекции, но только с одной меткой на молекулу полипептида. Иными словами, коллекция полипептидов может содержать полипептид А, ассоциированный с меткой для детекции X, и полипептид А, ассоциированный с меткой для детекции Y, но не полипептид А, ассоциированный с обеими метками для детекции X и Y. В конкретных вариантах осуществления, каждый член коллекции полипептидов является ассоциированным по меньшей мере с двумя метками для детекции. В конкретных вариантах осуществления, каждый член коллекции полипептидов является ассоциированным по меньшей мере с пятью метками для детекции. В

конкретных вариантах осуществления, каждый член коллекции полипептидов является ассоциированным по меньшей мере с десятью метками для детекции. В конкретных вариантах осуществления, каждый член коллекции полипептидов является ассоциированным с приблизительно двадцатью метками для детекции. Каждая метка для детекции имеет следующие характеристики:

а. Метка характеризуется аминокислотной последовательностью метки, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой множеством экспрессирующих векторов.

б. Метка характеризуется молекулярной массой между 200 и 5000 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 500 и 2500 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между приблизительно 900 и приблизительно 2200 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 903 и 2180 Да.

с. Метка отделена от указанного члена указанной коллекции полипептидов первым разделяемым элементом.

В конкретных вариантах осуществления второго аспекта изобретения, метка для детекции характеризуется уровнем гидрофобности между -27 и 128. В конкретных вариантах осуществления, метка для детекции характеризуется уровнем гидрофобности между -1 и 70.

В конкретных вариантах осуществления второго аспекта изобретения, член коллекции полипептидов является ассоциированным с аффинной меткой.

В конкретных вариантах осуществления второго аспекта изобретения, метка для детекции является ассоциированной с аффинной меткой. Аффинная метка и метка для детекции содержатся внутри одной и той же первичной аминокислотной последовательности. Аффинная метка отделена от метки для детекции вторым разделяемым элементом. Метку для детекции можно освобождать от аффинной метки посредством разделения второго разделяемого элемента. В конкретных вариантах осуществления, аффинная метка выбрана из группы, содержащей His-метку, СВР-

метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, НРС-метку, GST-метку, Avi-метку, биотинилированную метку, Мус-метку, 3xFLAG метку и MBP-метку. В конкретных вариантах осуществления, аффинная метка представляет собой His-метку.

В конкретных вариантах осуществления второго аспекта изобретения, выделенная метка для детекции состоит из 5-30 непрерывных аминокислот и содержит одну и только одну аминокислоту, имеющую положительно заряженную боковую цепь. В конкретных вариантах осуществления, выделенная метка для детекции состоит из 7-21 непрерывных аминокислот и содержит одну и только одну аминокислоту, имеющую положительно заряженную боковую цепь. В конкретных вариантах осуществления, выделенная метка для детекции состоит из 11-15 непрерывных аминокислот и содержит одну и только одну аминокислоту, имеющую положительно заряженную боковую цепь.

В конкретных вариантах осуществления, аминокислота, имеющая положительно заряженную боковую цепь, локализована на С-конце выделенной метки для детекции. В конкретных вариантах осуществления, аминокислота, имеющая положительно заряженную боковую цепь, выбрана из аргинина (R) и лизина (K). В конкретных вариантах осуществления, аминокислота, имеющая положительно заряженную боковую цепь, представляет собой аргинин (R), локализованный на С-конце выделенной метки для детекции.

В конкретных вариантах осуществления второго аспекта изобретения, метка для детекции содержит

а. элемент последовательности I, где элемент последовательности I состоит из 5-10, в частности 7, непрерывных аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R; и

б. элемент последовательности II, выбранный из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

#### *Метка для детекции*

В соответствии с третьим аспектом, настоящее изобретение относится к пептидной метке для детекции, разработанной для

оптимальной детекции посредством масс-спектрометрии. Метка для детекции состоит из 4-20 аминокислот и имеет следующие признаки:

а. Метка для детекции содержит только одну аминокислоту, имеющую положительно заряженную боковую цепь.

б. Метка для детекции характеризуется молекулярной массой между 200 и 5000 Да. В конкретных вариантах осуществления, метка для детекции характеризуется молекулярной массой между 500 и 2500 Да. В конкретных вариантах осуществления, метка для детекции характеризуется молекулярной массой между 900 и 2200 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 903 и 2180 Да.

В конкретных вариантах осуществления третьего аспекта изобретения, метка для детекции состоит из 7-18 аминокислот. В конкретных вариантах осуществления третьего аспекта изобретения, метка для детекции состоит из 11-15 аминокислот.

В конкретных вариантах осуществления третьего аспекта изобретения, метка для детекции в основном состоит из

а. элемента последовательности I, где указанный элемент последовательности I состоит из 5-10, в частности 7, непрерывных аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R; и

б. элемента последовательности II, выбранного из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

#### *Коллекция меток для детекции*

В соответствии с другим аспектом, настоящее изобретение относится к коллекции пептидных меток. Коллекция пептидных меток содержит пептидные метки в соответствии с третьим аспектом изобретения. Каждая метка для детекции, содержащаяся в коллекции пептидных меток, состоит из 4-20 аминокислот и характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, содержащейся в указанной коллекции меток для детекции. В конкретных вариантах осуществления, каждая метка для детекции состоит из 7-18 аминокислот. В конкретных вариантах осуществления, каждая метка для детекции состоит из 11-15 аминокислот. В конкретных



вариантах осуществления, коллекция пептидных меток содержит по меньшей мере 96 пептидных меток. В конкретных вариантах осуществления, коллекция пептидных меток содержит по меньшей мере 500000 пептидных меток. В конкретных вариантах осуществления, коллекция пептидных меток содержит по меньшей мере  $10^7$  пептидных меток. В конкретных вариантах осуществления, коллекция пептидных меток содержит приблизительно  $10^8$  пептидных меток.

В конкретных вариантах осуществления этого аспекта изобретения, метка для детекции содержит только одну аминокислоту, имеющую положительно заряженную боковую цепь, и остальные аминокислоты выбраны из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R.

В конкретных вариантах осуществления этого аспекта изобретения, метка характеризуется уровнем гидрофобности между -27 и 128. В конкретных вариантах осуществления, метка для детекции характеризуется уровнем гидрофобности между -1 и 70.

В конкретных вариантах осуществления этого аспекта изобретения, метка для детекции является ассоциированной с аффинной меткой. В конкретных вариантах осуществления, аффинная метка выбрана из группы, содержащей His-метку, CBP-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, HPC-метку, GST-метку, Avi-метку, биотинилированную метку, Мус-метку, 3xFLAG-метку и MBP-метку. В конкретных вариантах осуществления аффинная метка представляет собой His-метку. Аффинная метка и метка для детекции содержатся внутри одной и той же первичной аминокислотной последовательности. Аффинная метка отделена от метки для детекции разделяемым элементом.

В конкретных вариантах осуществления этого аспекта изобретения, метка для детекции в основном состоит из

а. элемента последовательности I, где указанный элемент последовательности I состоит из 5-10, в частности 7, непрерывных аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R; и

b. элемента последовательности II, выбранного из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

*Коллекция плазмидных векторов*

В соответствии с другим аспектом, настоящее изобретение относится к коллекции плазмидных векторов. Каждый член указанной коллекции плазмидных векторов содержит последовательность нуклеиновой кислоты, кодирующую метку для детекции. Каждая метка для детекции состоит из 4-20 аминокислот и характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой указанной коллекцией плазмидных векторов. В конкретных вариантах осуществления, каждая метка для детекции состоит из 7-18 аминокислот. В конкретных вариантах осуществления, каждая метка для детекции состоит из 11-15 аминокислот. В конкретных вариантах осуществления, коллекция плазмидных векторов содержит по меньшей мере 96 плазмидных векторов. В конкретных вариантах осуществления, коллекция плазмидных векторов содержит по меньшей мере 500000 плазмидных векторов. В конкретных вариантах осуществления, коллекция плазмидных векторов содержит по меньшей мере  $10^7$  плазмидных векторов. В конкретных вариантах осуществления, коллекция плазмидных векторов содержит приблизительно  $10^8$  плазмидных векторов.

В конкретных вариантах осуществления этого аспекта изобретения, метка для детекции содержит только одну аминокислоту, имеющую положительно заряженную боковую цепь.

В конкретных вариантах осуществления этого аспекта изобретения, метка для детекции характеризуется молекулярной массой между 200 и 5000 Да. В конкретных вариантах осуществления, метка для детекции характеризуется молекулярной массой между 500 и 2500 Да. В конкретных вариантах осуществления, метка для детекции характеризуется молекулярной массой между 900 и 2200 Да. В конкретных вариантах осуществления, метка для детекции характеризуется молекулярной массой между 903 и 2180 Да.

В конкретных вариантах осуществления этого аспекта изобретения, метка характеризуется уровнем гидрофобности между -27 и 128. В конкретных вариантах осуществления, метка для детекции характеризуется уровнем гидрофобности между -1 и 70.

В конкретных вариантах осуществления этого аспекта изобретения, метка для детекции является ассоциированной с аффинной меткой. В конкретных вариантах осуществления, аффинная метка выбрана из группы, содержащей His-метку, СВР-метку, СУD-метку, Strep-метку, StrepII-метку, FLAG-метку, НРС-метку, GST-метку, Avi-метку, биотинилированную метку, Мус-метку, 3xFLAG-метку и МВР-метку. В конкретных вариантах осуществления аффинная метка представляет собой His-метку. Аффинная метка и метка для детекции содержатся внутри одной и той же первичной аминокислотной последовательности. Аффинная метка отделена от метки для детекции вторым разделяемым элементом.

В конкретных вариантах осуществления этого аспекта изобретения, метка для детекции в основном состоит из

а. элемента последовательности I, где указанный элемент последовательности I состоит из 5-10, в частности 7, непрерывных аминокислот, независимо друг от друга выбранных из А, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R; и

б. элемента последовательности II, выбранный из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

В конкретных вариантах осуществления этого аспекта изобретения, каждый член коллекции плазмидных векторов содержит

а. кассету для отрицательного отбора, фланкированную на 5' первым участком рестрикции эндонуклеазой и на 3' вторым участком рестрикции эндонуклеазой;

б. промотор, локализованный на 5' от первого участка рестрикции эндонуклеазой;

с. последовательность нуклеиновой кислоты метки, кодирующую метку для детекции, локализованную на 3' от второго участка рестрикции эндонуклеазой. В конкретных вариантах осуществления, последовательность нуклеиновой кислоты, кодирующая метку для детекции, и второй участок рестрикции эндонуклеазой разделены

менее, чем 100 парами оснований. В конкретных вариантах осуществления, последовательность нуклеиновой кислоты, кодирующая метку для детекции, и второй участок рестрикции эндонуклеазой разделены менее, чем 50 парами оснований. В конкретных вариантах осуществления, последовательность нуклеиновой кислоты, кодирующая метку для детекции, и второй участок рестрикции эндонуклеазой разделены приблизительно 20 парами оснований. В конкретных вариантах осуществления, пары оснований, локализованные между последовательностью нуклеиновой кислоты, кодирующей метку для детекции, и вторым участком рестрикции эндонуклеазой, кодируют первый разделяемый элемент.

В конкретных вариантах осуществления этого аспекта изобретения, каждый член коллекции плазмидных векторов содержит

a. последовательность нуклеиновой кислоты метки, кодирующая метку для детекции, ассоциированную внутри одной и той же рамки считывания с последовательностью нуклеиновой кислоты, кодирующей полипептид;

b. обеспечивающий разнообразие элемент, содержащий неидентичные основания, для предотвращения перегрузки сигнала в ходе секвенирования;

c. участок связывания праймера для связывания праймеров для секвенирования;

d. индексный элемент, содержащий одну из нескольких определенных последовательностей нуклеиновой кислоты, для обеспечения мультиплексности;

e. адаптерный элемент для иммобилизации молекулы ДНК в ходе секвенирования и

f. два участка рестрикции эндонуклеазами, фланкирующие элементы a-e, для высвобождения фрагмента ДНК из плазмидного вектора перед секвенированием.

Плазмидные векторы, описанные в предшествующем варианте осуществления, служат в качестве плазмид для глубокого секвенирования. Предпочтительно, эти векторы не содержат аффинную метку, чтобы уменьшить длину фрагмента, подлежащего секвенированию.

*Способ детекции белка*

В соответствии с другим аспектом, настоящее изобретение относится к способу детекции белка. Способ включает следующие стадии:

а. Получают библиотеку нуклеиновых кислот, кодирующую библиотеку полипептидов. Каждый полипептид, содержащийся в библиотеке полипептидов, является ассоциированным с меткой для детекции. Полипептид и метка для детекции содержатся внутри одной и той же первичной аминокислотной последовательности. Каждая метка для детекции имеет следующие характеристики:

i. Метка характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой библиотекой нуклеиновых кислот.

ii. Метка характеризуется молекулярной массой между 200 и 5000 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 500 и 2500 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между приблизительно 900 и приблизительно 2200 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 903 и 2180 Да.

iii. Метка отделена от ассоциированного полипептида первым разделяемым элементом.

Каждая метка для детекции, кодируемая библиотекой нуклеиновых кислот, является уникальной по отношению к любой другой метке для детекции, кодируемой библиотекой нуклеиновых кислот. Каждый полипептид, содержащийся в библиотеке полипептидов, является ассоциированным по меньшей мере с одной меткой для детекции. В конкретных вариантах осуществления, каждый полипептид, содержащийся в библиотеке полипептидов, является ассоциированным по меньшей мере с двумя метками для детекции. В конкретных вариантах осуществления, каждый полипептид, содержащийся в библиотеке полипептидов, является ассоциированной по меньшей мере с пятью метками для детекции. В конкретных вариантах осуществления, каждый полипептид, содержащийся в библиотеке полипептидов, является ассоциированным по меньшей мере с десятью метками для детекции. В конкретных

вариантах осуществления, каждый полипептид, содержащийся в библиотеке полипептидов, является ассоциированным с приблизительно двадцатью метками для детекции. Каждая молекула полипептида содержит только одну метку для детекции.

b. Получают базу данных. База данных содержит следующую информацию:

i. Множество последовательностей нуклеиновых кислот и/или аминокислот. Множество последовательностей содержит последовательности всех членов библиотеки нуклеиновых кислот. Каждая из последовательностей содержит последовательность, определяющую полипептид, и последовательность, определяющую метку для детекции.

ii. Паттерн фрагментации при масс-спектрометрии для каждой метки для детекции, кодируемой библиотекой нуклеиновых кислот.

c. Библиотеку полипептидов экспрессируют из библиотеки нуклеиновых кислот.

d. Отбирают член библиотеки полипептидов на стадии отбора, получая отобранный полипептид.

e. Первый разделяемый элемент разрезают. Таким образом, метку для детекции отделяют от отобранного полипептида и получают выделенную метку для детекции.

f. Выделенную метку для детекции идентифицируют следующим образом:

i. Паттерн фрагментации выделенной метки для детекции регистрируют посредством масс-спектрометрии.

ii. Паттерн фрагментации, полученный на стадии i, приводят в соответствие с прогнозированным паттерном фрагментации в полученной базе данных. Таким образом, идентифицируют выделенную метку для детекции. Комбинация информации, полученной посредством масс-спектрометрии, с информацией, полученной посредством секвенирования библиотеки меченых нуклеиновых кислот, позволяет однозначную идентификацию данной метки для детекции.

g. Последовательность, определяющую метку для детекции, идентифицированную на стадии f, выбирают из множества последовательностей, содержащихся в базе данных. Таким образом,

идентифицируют член библиотеки полипептидов, ассоциированный с меткой для детекции, идентифицированной на стадии f.

В конкретных вариантах осуществления, каждый член указанной библиотеки полипептидов является ассоциированным с аффинной меткой.

В конкретных вариантах осуществления, каждая метка для детекции является ассоциированной с аффинной меткой.

В конкретных вариантах осуществления, аффинная метка выбрана из группы, содержащей His-метку, CBP-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, NPC-метку, GST-метку, Avi-метку, биотинилированную метку, Muc-метку, 3xFLAG-метку и MBP-метку.

В конкретных вариантах осуществления, аффинная метка отделена от указанной метки для детекции вторым разделяемым элементом, и указанный второй разделяемый элемент разрезают до стадии f. Таким образом, только метку для детекции без ассоциированной аффинной метки анализируют посредством масс-спектрометрии.

Спецификации массы и паттерна фрагментации метки для детекции относятся к массе и паттерну фрагментации метки после ее отделения от ассоциированного полипептида и аффинной метки, т.е. после разделения первого и второго разделяемых элементов. Специалисту в данной области понятно, что в случаях, когда метку для детекции не освобождают от ассоциированной аффинной метки перед масс-спектрометрией, это может влиять на результаты анализа масс-спектрометрии. Поскольку все метки для детекции ассоциированы с одной и той же аффинной меткой, изменения массы и паттерна фрагментации можно учитывать, таким образом, все еще является возможной идентификация метки для детекции, хотя и не настолько эффективно и четко, как в случаях, когда метку для детекции отделили от аффинной метки посредством разделения второго разделяемого элемента.

В конкретных вариантах осуществления, аффинная метка представляет собой His-метку.

Специалисту в данной области понятно, что стадии d-g проводят для ряда различных членов библиотеки полипептидов

параллельно. Пул из нескольких полипептидов отбирают на стадии g, и все из этих полипептидов идентифицируют посредством анализа масс-спектрометрии их меток для детекции. Специалисту в данной области понятно, что по техническим причинам, не каждый отдельный полипептид можно идентифицировать на этой стадии.

Анализ масс-спектрометрии, проведенный на стадии f, является количественным, таким образом, способ по изобретению позволяет не только идентификацию полипептида, но также количественную оценку количества этого полипептида в образце.

*Способ ассоциации полипептида с уникальной меткой для детекции*

В соответствии с другим аспектом, настоящее изобретение относится к способу ассоциации полипептида с уникальной меткой для детекции. Способ включает следующие стадии:

a. Получают первую библиотеку нуклеиновых кислот. Каждый член первой библиотеки нуклеиновых кислот содержит кодирующую полипептид последовательность, кодирующую член первой библиотеки полипептидов;

b. Получают вторую библиотеку нуклеиновых кислот. Каждый член второй библиотеки нуклеиновых кислот содержит кодирующую метку последовательность, кодирующую метку для детекции. Каждая метка для детекции имеет следующие характеристики:

i. Метка характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой второй библиотекой нуклеиновых кислот;

ii. Метка характеризуется молекулярной массой между 200 и 5000 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 500 и 2500 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между приблизительно 900 и приблизительно 2200 Да. В конкретных вариантах осуществления, метка характеризуется молекулярной массой между 903 и 2180 Да.

c. Кодирующую полипептид последовательность, содержащуюся в члене первой библиотеки нуклеиновых кислот, вставляют в член



второй библиотеки нуклеиновых кислот. Таким образом, получают плазмиды с множеством комбинацией полипептид-метка.

Первая библиотека нуклеиновых кислот имеет размер 5-100000. В конкретных вариантах осуществления, первая библиотека нуклеиновых кислот имеет размер 100-50000. В конкретных вариантах осуществления, первая библиотека нуклеиновых кислот имеет размер 500-5000.

Вторая библиотека нуклеиновых кислот имеет размер  $10^3$ - $10^{11}$ . В конкретных вариантах осуществления, вторая библиотека нуклеиновых кислот имеет размер  $10^5$ - $10^{10}$ . В конкретных вариантах осуществления, вторая библиотека нуклеиновых кислот имеет размер  $10^6$ - $10^9$ . В конкретных вариантах осуществления, вторая библиотека нуклеиновых кислот имеет размер приблизительно  $10^8$ .

Внутри плазмид с множеством комбинаций полипептид/метка, каждая кодирующая полипептид последовательность первой библиотеки нуклеиновых кислот является ассоциированной с кодирующей метку последовательностью второй библиотеки нуклеиновых кислот. Ассоциация возникает внутри одной и той же рамки считывания.

d. Отбирают подгруппу из плазмид с множеством комбинаций полипептид-метка. Эта стадия отбора включает отбор определенного количества клонов, где каждый клон содержит один член из множества плазмид с комбинациями полипептид-метка. Таким образом, получают библиотеку меченых нуклеиновых кислот, кодирующую библиотеку меченых полипептидов. Каждый член библиотеки меченых полипептидов содержит полипептид и метку для детекции. Каждая метка содержится только в одном члене библиотеки меченых полипептидов. Иными словами, каждая метка для детекции является уникальной в библиотеке меченых полипептидов. Каждый полипептид может, однако, содержаться в нескольких членах библиотеки меченых полипептидов (избыточное мечение).

В конкретных вариантах осуществления, каждый полипептид является ассоциированным по меньшей мере с одной меткой для детекции. В конкретных вариантах осуществления, каждый полипептид является ассоциированным по меньшей мере с двумя метками для детекции. В конкретных вариантах осуществления,

каждый полипептид является ассоциированным по меньшей мере с пятью метками для детекции. В конкретных вариантах осуществления, каждый полипептид является ассоциированным по меньшей мере с десятью метками для детекции. В конкретных вариантах осуществления, каждый полипептид является ассоциированным с приблизительно двадцатью метками для детекции.

В конкретных вариантах осуществления этого аспекта изобретения, отобранная подгруппа из множества плазмид с комбинациями полипептид-метка составляет по меньшей мере 10х количество членов первой библиотеки нуклеиновых кислот. В конкретных вариантах осуществления, отобранная подгруппа из множества плазмид с комбинациями полипептид-метка составляет по меньшей мере 20х количество членов первой библиотеки нуклеиновых кислот.

В конкретных вариантах осуществления этого аспекта изобретения, отобранная подгруппа из множества плазмид с комбинациями полипептид-метка составляет менее 50% от количества членов второй библиотеки нуклеиновых кислот. В конкретных вариантах осуществления этого аспекта изобретения, отобранная подгруппа из множества плазмид с комбинациями полипептид-метка составляет менее 5% от количества членов второй библиотеки нуклеиновых кислот. В конкретных вариантах осуществления этого аспекта изобретения, отобранная подгруппа из множества плазмид с комбинациями полипептид-метка составляет менее 0,05% от количества членов второй библиотеки нуклеиновых кислот.

Посредством выбора оптимального размера отобранной подгруппы из множества плазмид с комбинациями полипептид-метка, обеспечивают, чтобы в библиотеке меченых полипептидов, каждая метка для детекции являлась уникальной (присутствующей только один раз), но каждый полипептид присутствовал несколько раз, каждый раз ассоциированный с другой меткой для детекции.

Во всех случаях, когда альтернативы для одиночных отделяемых признаков представлены в настоящем описании как «варианты осуществления», следует понимать, что такие альтернативы можно свободно комбинировать для формирования

отдельных вариантов осуществления изобретения, описанных в настоящем документе.

Изобретение далее проиллюстрировано посредством следующих примеров и фигур, из которых можно вывести дополнительные варианты осуществления и преимущества. Эти примеры предназначены для иллюстрации изобретения, а не для ограничения его объема.

#### **Краткое описание чертежей**

На фигуре 1 показан обзор технологии NestLink. А) Библиотеку нанотел заключают гнездовым способом в библиотеку с метками Flycode, кодируемую экспрессирующим вектором pNLx. Затем, последовательности нанотел, снабженные меткой flycode, вырезают посредством рестрикционного расщепления и вставляют в pNLs, что приводит к присоединению необходимых адаптерных последовательностей для глубокого секвенирования. Адаптеры, связанные с снабженными меткой flycode нанотелами, затем вырезают посредством рестрикционного расщепления и подвергают глубокому секвенированию в линейной форме. В) Гнездовую библиотеку, кодируемую в pNLx, экспрессируют и очищают. Прилагают селекционное давление (в этом конкретном случае, белки с кажущейся молекулярной массой мономера нанотела отбирают посредством эксклюзионной хроматографии), и метки Flycode из отобранных нанотел выделяют посредством отщепления протеазой. С) Данные глубокого секвенирования позволяют получение базы данных, приписывающей все метки Flycode соответствующим им нанотелам. Метки Flycode для каждого нанотела соединяют последовательно. Ранее выделенные метки Flycode (см. В) подвергают LC-MS, и получают списки пиков из зарегистрированных данных MS/MS. Для данных MS/MS проводят поиск в базе данных, содержащей соединенные последовательно метки Flycode, что позволяет идентификацию и относительную количественную оценку отобранных нанотел.

На фигуре 2 показаны соответствующие варианты дизайна плазмид для технологии NestLink до (верхние цепи) и после вставки библиотеки (нижние цепи). А) Фагмида, использованная для отборов в фаговом дисплее нанотел против молекул-мишеней. Фагмида несет два участка рестрикции SapI, позволяющие вставку

библиотек нанотел и их эффективный перенос в экспрессирующий вектор NestLink pNLx после обогащения посредством фагового дисплея. В) экспрессирующий вектор NestLink pNLx, несущий разнообразие Flycode из приблизительно  $10^8$  вариантов. Участки SapI разработаны для исчезновения после вставки библиотеки нанотел. Снабженные меткой flycode нанотела можно специфически вырезать из экспрессирующего вектора посредством рестрикции SfiI. Размещение участков SfiI обеспечивает глубокое секвенирование полноразмерных нанотел, присоединенных к соответствующим им меткам Flycode, но минимизирует длину прочтения глубокого секвенирования посредством исключения избыточных последовательностей, таких как PelB и His-метка. С) Получен набор векторов для глубокого секвенирования (pLNs) с различными индексами, где каждый несет все необходимые последовательности для секвенирования Illumina MiSeq. Снабженные меткой flycode нанотела вставляют в этот вектор посредством рестрикции SfiI и лигирования. Затем их высвобождают в форме линейного фрагмента, содержащего все адаптерные области MiSeq, посредством рестрикции BseRI. По этому способу, для получения фрагментов ДНК для анализа MiSeq не требуется ПЦР, которая может приводить к событиям рекомбинации в последовательностях нанотело-метка Flycode и таким образом, нарушать связь между меткой Flycode и последовательностью нанотела. D) Адаптеры для глубокого секвенирования можно также присоединять посредством синтетических двухцепочечных адаптерных олигонуклеотидов с помощью соответствующих одноцепочечных выступающих концов, комплементарных участку рестрикции SfiI, кодируемому в pNLx.

На фигуре 3 показана абсолютная количественная оценка PLOI-членов посредством меток Flycode с использованием LC-MS. Семь известных количеств (x-ось) снабженных меткой flycode синтетел добавляли в два различных образца, содержащие лизат либо из *E. coli*, либо из *M. smegmatis*, соответственно (фон). Снабженные меткой flycode синтетела добавляли при 0,2, 0,4, 1,3, 4,1, 8,5, 18,0 и 27,5 единиц оптической плотности (280 нм), и они содержали 28, 56, 112, 56, 112, 84 и 112 меток Flycode, как определено посредством глубокого секвенирования. Выделенные

метки Flycode анализировали посредством LC-MS. Интенсивности MS1 для всех меток Flycode каждого синтетела суммировали с использованием программного обеспечения Progenesis.

На фигуре 4 показана идентификация синтетел, для которых показаны наилучшие скорости диссоциации из 1080 кандидатов на связывание посредством NestLink. А: Мономерные синтетела, элюированные совместно с биотинилированным белком-мишенью в растворе (SEC), иммобилизовали на двух эквивалентных колонках с стрептавидин-сефарозой. Одну колонку промывали буфером, другую колонку промывали избытком небитинилированной мишени в течение 3 мин. Затем метки Flycode из оставшихся связанными синтетел выделяли и оценивали количественно по интенсивности в LC-MS1. В: Интенсивности в LC-MS1 (сумму для всех меток flycode) определяли для каждого члена пула, и соотношение между двумя колонками наносили на график по y-оси для каждого индивидуального синтетела (x-ось). Синтетела, которые не экспрессировались, не являлись мономерными или не связывались с мишенью в растворе, не поддавались детекции ни на одной из колонок, поскольку были удалены в результате предварительного селекционного давления, описанного в эксперименте для проверки и подтверждения концепции (синтетела 320-1080). Слабо связывающиеся синтетела поддавались детекции только после промывки буфером, но не при конкуренции с избытком мишени (синтетела 187-320). Синтетела 1-186 детектировали на обеих колонках и ранжировали в соответствии с их скоростью диссоциации. Наиболее многообещающими синтетелами для нижестоящих применений являлись синтетела с наиболее медленной скоростью диссоциации, приводящей к соотношению, близкому к 1. С: Корреляция считывания NestLink и экспериментов SPR для индивидуально отобранных синтетел. Последовательности ДНК для 11 синтетел, анализированных в В), синтезировали (генный синтез) и соответствующие связывающие молекулы экспрессировали, очищали и анализировали посредством поверхностного плазмонного резонанса по одной. Данные SPR наносили на график в форме остаточного сигнала связывания после 3 минут промывки (в качестве показателя скорости диссоциации) по x-оси против

соотношения синтетел, определенного посредством NestLink, как показано в В) по у-оси.

Фигура 5: Анализ 3469 нанотел из иммунизированной альпаки и идентификация тех, для которых показано наиболее сильное связывание антигена в растворе. После исключения членов пула с плохими уровнями экспрессии (стадия 1) и растворимостью (стадия 2, отбор мономерных нанотел), мономерную фракцию пула инкубировали с мембранным белковым антигеном в трех различных стехиометрических соотношениях и анализировали посредством SEC. Образцы для LC-MS собирали после стадии 1 (регистрация уровня экспрессии каждого индивидуального члена пула), на стадии 2 (регистрация растворимости каждого индивидуального члена пула) и для всех пиков комплекса мишень/нанотело на стадии 3. Круговые диаграммы представляют относительное количество каждого нанотела в пуле (несвязывающие молекулы или слабо связывающие молекулы совместно окрашены светло серым, общее количество членов пула соответствует 100%) на различных стадиях способа отбора, как определено по сумме всех интенсивностей MS1 для каждого нанотела (100%=сумма всех интенсивностей MS1 для всех меток flycode из всех нанотел). Как ожидали для стадии 3, увеличение соотношения пула к антигену приводит к увеличению внутренней конкуренции множества связывающих членов пула за ограниченное количество антигена. Фракция членов пула с наиболее сильной аффинностью, таким образом, увеличивается при более высокой конкуренции за ограниченное количество эпитопов.

Фигура 6: А: Анализ 1456 синтетел из пула, полученного посредством отборов *in vitro* (стадия 1) против очищенного белка-мишени наружной мембраны, для связывания с поверхностью клеток представляющих интерес грамотрицательных бактерий (стадия 2). На стадии 2 (NestLink), члены пула с плохими уровнями экспрессии и растворимостью сначала исключали из совокупности, с последующими 4 различными экспериментами осаждения с использованием 4 различных представляющих интерес штаммов бактерий. После удаления посредством промывки членов, не связавшихся с клетками с высокой аффинностью, все метки Flycode из пула выделяли и анализировали посредством LC-MS. Затем сумму всех интенсивностей

MS1 для всех меток Flycode на синтетело можно было использовать в качестве показателя относительной концентрации каждого индивидуального синтетела в пуле на каждой из клеток-мишеней. Это позволяло однозначное специфическое для клеток считывание (B), регистрирующее для каждого синтетела (x-ось) его относительную концентрацию (по отношению к целому пулу) на каждом из 4 типов клеток. Из соображений ясности, только 25% из всех анализированных синтетел показано в B.

### **Примеры**

#### *Библиотека последовательностей с метками flycode*

Рандомизированная библиотека кодируемых короткими ДНК пептидов разработана для оптимальной детекции посредством масс-спектрометрии (MS), в частности, посредством LC-MS (жидкостной обращеннофазовой хроматографии в сочетании с ESI-MS). Пептиды попадают в диапазон масс между 903 и 2180 Да, который является оптимальным для чувствительной детекции посредством ESI-MS. Метки Flycode несут два положительных заряда при физиологическом pH и ниже, а именно, R на C-конце и N-концевой первичный амин. Положительный заряд на C-конце метки flycode облегчает ионизацию пептида для масс-спектрометрической детекции и действует как уникальный участок расщепления трипсином. В каждой метке flycode N-концевой амин представляет собой только первичный амин, который используют для присоединения по аминогруппе посредством химических реакций NHS. Это позволяет присоединение меток для проведения количественной масс-спектрометрии, например, iTRAQ (изобарических меток для относительной и абсолютной количественной оценки). Метки метки flycode сконструированы для демонстрации диапазона вариантов гидрофобности, идеально подходящего для разделения пептидов посредством стандартных колонок для обращеннофазовой хроматографии.

Библиотека с метками flycode состоит из двух частей плюс фланкирующие аминокислоты, которые являются постоянными, а именно, GS на N-конце и R на C-конце. N-концевая последовательность «GS» является частью участка расщепления протеазой тромбином, которая остается на метке flycode после расщепления.

Часть 1: Область штрих-кода охватывает 7 последовательных рандомизированных положений аминокислот. Средняя частота аминокислот приведена в таблице 1 выше (в %).

Не все двадцать природных аминокислот присутствуют в штрих-коде (С, М, К, R, Н и I отсутствуют). С и М исключены, поскольку они подвержены окислению. К, R и Н исключены, поскольку они могут добавлять дополнительный положительный заряд к последовательности метки flycode, что является нежелательным, поскольку пептид будет в таком случае нести дополнительный заряд во время детекции ESI-MS и выпадать из оптимального диапазона детекции. К и R могут добавлять дополнительные участки расщепления трипсином в последовательность метки flycode, что является нежелательным. К добавляет другой первичный амин, что может затруднять мечение пептида с использованием химических реакций NHS. Изолейцин исключен, поскольку его невозможно отличить по массе от лейцина.

Часть 2: С-конец сконструирован в 5 различных вариантах, которые являются одинаково частыми в библиотеке с метками flycode и которые все оканчиваются на R. Они также лишены С, М, К, Н и I. Метки flycode состоят, таким образом, минимум из 11 аминокислот и максимум из 15 аминокислот (GS+7 рандомизированных остатков+2-6 С-концевых остатков). 5 различных окончаний С-конца перечислены ниже:

SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR),  
SEQ ID NO 04 (WLTVR), SEQ ID NO 05 (WQEGGR).

*Экспрессирующий вектор NestLink pLNx, содержащий библиотеку с метками flycode.*

Экспрессирующий вектор NestLink pLNx несет библиотеку с метками flycode с разнообразием  $10^8$  вариантов последовательностей (фигура 2) и позволяет введение библиотеки представляющих интерес белков (PLOI) в рамке считывания с метками flycode. Результатом этой стадии является «гнездовая библиотека», поскольку две библиотеки (PLOI и библиотека с метками flycode) вставлены друг в друга гнездовым способом. Экспрессирующий вектор позволяет также опосредованное рестрикционным ферментом вырезание гнездовой библиотеки (PLOI, слитой с метками flycode),



таким образом, что либо ее можно вставлять в плазмиду для глубокого секвенирования, либо можно проводить прямое лигирование адаптеров Illumina MiSeq с использованием двухцепочечных олигонуклеотидов (адаптеров). Следует отметить, что PLOI может представлять собой любую генетически кодируемую библиотеку.

PLOI вводят в экспрессирующий вектор посредством рестрикционного расщепления источника ДНК, кодирующего библиотеку, с последующим лигированием в экспрессирующий вектор. Авторы настоящего изобретения используют для этой цели рестрикционный фермент типа IIS (SapI). Источник ДНК как правило, происходит из фагмиды, полученной после отборов с использованием фагового дисплея, содержащей участки SapI, ориентированные таким образом, что PLOI можно субклонировать в экспрессирующий вектор NestLink без амплификации ПЦР (описание этого вектора см. ниже). Когда PLOI вставляют, она заменяет кассету для отрицательного отбора (ccdB), что сильно улучшает эффективность стадии вставки.

Метку flycode отщепляют от PLOI посредством тромбина, и His-метку удаляют из метки flycode посредством трипсина. Эти отщепления обеспечивают то, что пептиды с оптимальной массой, оптимальной гидрофобностью и оптимальным зарядом выделяют для масс-спектрометрии (см. описание метки flycode выше). Понятно также, что любые другие комбинации протеаз можно использовать для той же самой цели.

Следует отметить, что С-концевой аргинин (R) метки flycode играет важную роль: во-первых, он является единственной положительно заряженной аминокислотой метки flycode, поскольку остатки лизина или другие остатки аргинина исключены из библиотеки с метками flycode. По этой причине, трипсин - протеаза, которая осуществляет расщепление после положительно заряженных остатков, и которую, таким образом, считают довольно неспецифической, - можно использовать для специфического расщепления пептидной связи между аргинином и His-меткой (иначе метка flycode с His-меткой становится слишком тяжелой для анализа масс-спектрометрии, и His-метка может уменьшать

разделение при обращеннофазовой хроматографии до масс-спектрометрии). Во-вторых, известно, что пептиды с С-концевыми остатками аргинина особенно хорошо поддаются детекции посредством масс-спектрометрии (обеспечивающие преимущества свойства ионизации). И в-третьих, из-за этой единственной положительно заряженной аминокислоты, присутствующей в метке flycode, общий заряд постоянно составляет 2+ (N-конец+аргинин, все другие остатки являются нейтральными при низком pH при детекции), что облегчает анализ данных.

Важным аспектом этой технологии является тот факт, что является возможным (и необходимым) присоединять несколько уникальных меток flycode к одному и тому же члену библиотеки представляющих интерес белков. Например, для анализа пула из 100 различных белков, 2000 меток flycode присоединяют к этим 100 белкам, так что в среднем, каждый белок из пула связан 20 раз с различными метками flycode (соотношение между членами пула и метками flycode можно фактически менять, как желательно). Избыточное мечение облегчает однозначную детекцию членов пула посредством множества последовательностей меток flycode и усредняет потенциальное влияние последовательности метки flycode на биофизические свойства анализируемых представляющих интерес белков. Избыточное мечение позволяет также определение относительных количеств различных членов библиотеки белков в отобранном образце или одного и того же члена библиотеки белков в различным образом отобранных образцах. Избыточность представляет собой добавление, необходимое по техническим причинам: несмотря на то, что метки flycode разработаны для оптимальной детекции посредством масс-спектрометрии, некоторые метки flycode могут не быть детектированы, поскольку они либо теряются в ходе подготовки образца, либо не элюируются внутри окна гидрофобности обращеннофазовой колонки, анализируемого посредством масс-спектрометрии.

Кроме того, экспрессирующий вектор NestLink содержит два участка рестрикции SfiI, позволяющие вырезание гнездовой библиотеки (PLOI, слитой с метками flycode), таким образом, что либо ее можно вставлять в плазмиду для глубокого секвенирования,

либо можно проводить прямое лигирование адаптеров Illumina MiSeq с использованием двухцепочечных олигонуклеотидов (адаптеров). Обоснование этой критической стадии представлено ниже.

Следует отметить, что участки рестрикции SfiI и\или другие участки рестрикции либо внутри PLOI, либо между PLOI и меткой flycode, можно использовать для добавления дополнительных последовательностей к гнездовой библиотеке. Эти дополнительные последовательности можно, таким образом, экспрессировать как слитые с гнездовой библиотекой (между меткой flycode и PLOI или смежные с гнездовой библиотекой). Важно, что такие последовательности не увеличивают длину прочтения глубокого секвенирования (которая ограничена по техническим причинам), поскольку перенос в плазмиду для глубокого секвенирования (или прямое лигирование адаптеров для глубокого секвенирования посредством олигонуклеотидов) проводят до введения этих дополнительных последовательностей. Кроме того, добавление дополнительных последовательностей таким образом сохраняет физическую связь между меткой flycode и PLOI, что является абсолютно критическим для корректного приписывания меток flycode PLOI-членам.

#### *Плазмиды для глубокого секвенирования*

Плазмиды для глубокого секвенирования представляют собой набор векторов, которые несут все необходимые последовательности для глубокого секвенирования посредством Illumina MiSeq и которые позволяют вставку пулов членов гнездовой библиотеки из экспрессирующего вектора NestLink.

Перенос гнездовой библиотеки в плазмиду для глубокого секвенирования (фигуры 1 и 2) проводят посредством рестриционного расщепления и лигирования. Авторы настоящего изобретения использовали для этой цели рестриционный фермент SfiI, поскольку он имеет достаточно высокую специфичность, что является критическим при расщеплении полноразмерных библиотек, которые могут случайно кодировать участки рестрикции. Кроме того, выбранный участок узнавания SfiI транслируется в целесообразно гибкие и гидрофильные аминокислоты, которые можно

использовать в качестве линкерных аминокислот в экспрессирующей конструкции.

Авторы настоящего изобретения смогли экспериментально показать, что для NestLink является критическим, чтобы стадия переноса из экспрессирующего вектора NestLink в плазмиду для глубокого секвенирования не включала стадию амплификации ПЦР гнездовой библиотеки. Амплификация ПЦР последовательностей белок-метка flycode неминуемо приводит к рекомбинации негомологичных областей (например, CDR) между членами библиотеки и метками flycode (непреднамеренному присоединению метки flycode одного представляющего интерес белка к другому, куда она не присоединена в экспрессирующем векторе NestLink). Таким образом, связь между меткой flycode и белком нарушается.

Как описано выше, гнездовую библиотеку вырезают из экспрессирующего вектора посредством SfiI. Затем ее лигируют в плазмиду для глубокого секвенирования. Она заменяет кассету для отрицательного отбора (ccdB), что является критическим для эффективности стадии вставки. После вставки, ее фланкируют последовательностями, которые являются необходимыми (и часто используемыми) для глубокого секвенирования посредством Illumina MiSeq. Секвенирование происходит с обоих концов к центру. Соответствующие области, таким образом, присутствуют по обеим сторонам от вставки в противоположном направлении (обратно комплементарные последовательности, за исключением индекса).

Ниже приведено описание последовательности от внутренней части (вставки) до внешних областей:

**Участки SfiI:** Используют для замены ccdB гнездовой библиотекой.

**Разнообразие:** По технологии Illumina MiSeq получают первый сигнал секвенирования на основании последовательности рядом с участком связывания праймера. Первые несколько оснований должны быть разнообразными (не идентичными) для предотвращения перегрузки сигнала в каналах детекции и прекращения выполнения секвенирования.

**Участок связывания праймера:** В нем связывается праймер для секвенирования.

**Индекс** (отмеченный номерами 501 и 701): Технология Illumina MiSeq обеспечивает мультиплексность, т.е. несколько образцов можно анализировать при одном выполнении секвенирования. Для определения того, какие прочтения принадлежат какому образцу, считывают также индекс (вариабельный фрагмент из 8 п.о.). Для обеспечения возможности секвенирования для нескольких экспериментов NestLink при одном выполнении глубокого секвенирования, авторы настоящего изобретения получили набор из 11 плазмид для глубокого секвенирования, где каждая несет различную пару индексов (следует отметить, что в данном случае присутствует индексная последовательность по обеим сторонам от ставки).

**Адаптер:** Его используют для иммобилизации ДНК-матрицы для глубокого секвенирования в проточной ячейке Illumina MiSeq.

**Участок рестрикции BseRI:** Его используют для получения линейного фрагмента ДНК, необходимого для глубокого секвенирования Illumina MiSeq. Тот факт, что BseRI представляет собой рестрикционный фермент типа IIS (расщепляет вне своей последовательности узнавания), является особенно полезным для минимизации выступающего конца адаптера.

В традиционном способе, все эти элементы последовательности Illumina MiSeq присоединяют к ДНК, подлежащей секвенированию, либо посредством ПЦР, либо посредством лигирования адаптеров Illumina, с последующей амплификацией ПЦР, либо посредством набора для подготовки образцов ДНК без ПЦР TRuSeq (Illumina). В способе авторов настоящего изобретения, ДНК, подлежащую секвенированию (в данном случае, последовательности белок-метка flycode), субклонируют из донорного вектора (в данном случае, экспрессирующего вектора NestLink) в вектор для глубокого секвенирования посредством рестрикции и лигирования, таким образом, избегая ПЦР. На конечной стадии, вектор для глубокого секвенирования расщепляют с использованием BseRI. Это высвобождает полную матрицу для секвенирования Illumina MiSeq, которую отделяют от остова вектора в агарозном геле для ДНК и очищают посредством экстракции из геля.

### *Двухцепочечные адаптерные олигонуклеотиды для глубокого секвенирования*

Второй способ, позволяющий независимое от ПЦР присоединение необходимых адаптерных последовательностей для глубокого секвенирования Illumina MiSeq к PLOI, основан на двухцепочечных олигонуклеотидах, несущих такой же набор адаптерных последовательностей, как описано в случае плазмид для глубокого секвенирования, которые можно получать посредством генного синтеза для комплементарных одноцепочечных олигонуклеотидов и последующей реакции гибридизации. Комплементарные одиночные цепи синтезируют с разницей в длине, приводящей к липким выступающим концам адаптера после гибридизации. Эти выступающие концы соответствуют комплементарным последовательностям разрезанных участков рестрикции SfiI, которые образуются, когда снабженную меткой flycode PLOI вырезают из экспрессирующего вектора NestLink. Олигонуклеотиды после гибридизации можно, таким образом, лигировать с высокой эффективностью с снабженной меткой flycode PLOI для присоединения адаптерных последовательностей, необходимых для глубокого секвенирования Illumina MiSeq. Продукт лигирования очищают в агарозном геле до глубокого секвенирования.

Ниже приведено описание последовательности конечной матрицы для глубокого секвенирования от внутренней части (вставки) до внешних областей:

**Снабженная меткой flycode PLOI:** Снабженную меткой flycode PLOI вырезают из экспрессирующего вектора NestLink посредством рестрикционного расщепления SfiI.

**Остаток участков рестрикции SfiI:** Этот фермент позволяет вырезание из экспрессирующего вектора NestLink, и полученные липкие концы используют для сайт-специфического присоединения адаптеров для глубокого секвенирования.

**Разнообразие:** По технологии Illumina MiSeq получают первый сигнал секвенирования на основании последовательности рядом с участком связывания праймера. Первые несколько оснований должны быть разнообразными (не идентичными) для предотвращения

перегрузки сигнала в каналах детекции и прекращения выполнения секвенирования.

**Участок связывания праймера:** В нем связывается праймер для секвенирования.

**Индекс** (отмеченный номерами 501 и 701): Технология Illumina MiSeq обеспечивает мультиплексность, т.е. несколько образцов можно анализировать при одном выполнении секвенирования. Для определения того, какие прочтения принадлежат какому образцу, считывают также индекс (вариабельный фрагмент из 8 п.о.). Для обеспечения возможности секвенирования для нескольких экспериментов NestLink при одном выполнении глубокого секвенирования, авторы настоящего изобретения получили 7 адаптеров для глубокого секвенирования (3 для одного конца и 4 для другого конца), позволяющие получение 12 различных пар индексов.

**Адаптер:** Его используют для иммобилизации ДНК-матрицы для глубокого секвенирования в проточной ячейке Illumina MiSeq.

*Количественная оценка PLOI-членов посредством меток Flycode*

Множество применений NestLink требуют абсолютной количественной оценки снабженных меткой flycode PLOI-членов. В то время как LC-MS является неточной при количественной оценке индивидуальных пептидов в протеомике, NestLink имеет преимущество множества меток Flycode, присоединенных к каждому PLOI-члену, и гомогенной библиотеки Flycode, разработанной для оптимальной детекции посредством масс-спектрометрии. На основании этих соображений, авторы настоящего изобретения выдвинули гипотезу, что суммарная интенсивность MS1 всех меток Flycode для любого данного PLOI-члена должна быть пропорциональна количеству этого PLOI-члена в образце. Авторы настоящего изобретения тестировали эту гипотезу посредством добавления известных количеств восьми синтетел, связанных с меняющимися количествами меток Flycode, в два образца, содержащие лизаты из *E. coli* и *M. smegmatis*, соответственно (фигура 3). Наблюдаемая линейная взаимосвязь между суммарной интенсивностью MS1 всех меток Flycode для каждого снабженного меткой flycode синтетела и его добавленным количеством

подтвердила правильность гипотезы и показала, что способ NestLink, описанный в настоящем документе, можно использовать для количественной оценки индивидуальных PLOI-членов в пуле. Абсолютное количество индивидуальных PLOI-членов можно определять, если известные количества одного или более снабженных меткой flycode белков (стандартов) добавляются в образец перед выделением flycode для LC-MS.

*Фагмида для отборов с использованием фагового дисплея (до NestLink)*

В большинстве текущих применений авторами настоящего изобретения, PLOI представляет собой пул обогащенных синтетических нанотел, называемых синтетелами. Как правило, большую библиотеку синтетел обогащают с использованием фагового дисплея для связывания с белком-мишенью. Чтобы избежать рекомбинации негомологичных областей (т.е., CDR), PLOI не должна подвергаться амплификации посредством ПЦР после отборов с использованием фагового дисплея. С этой целью, фагмидный вектор (фигура 2A) конструировали таким образом, что PLOI можно субклонировать посредством участков рестрикции SapI в экспрессирующий вектор NestLink. Следует отметить, что участки SapI являются частью продукта трансляции, экспонированного на поверхности фага. Авторы настоящего изобретения смогли экспериментально показать, что эти дополнительные аминокислоты, происходящие из участков SapI, не мешают эффективности фагового дисплея.

Помимо участков SapI, вектор для фагового дисплея содержит все элементы, как правило, присутствующие в фагмидах, использованных для дисплея белков на фаге M13, и является производным вектора pMESy4 (genbank KF415192).

Дополнительное общее замечание, относящееся ко всем векторам, описанным в настоящем описании: Чтобы обеспечивать эффективный перенос вставки от одного вектора в другой, является критическим, чтобы векторы несли различные гены устойчивости к антибиотикам. Таким образом, экспрессирующий вектор NestLink несет маркер устойчивости к хлорамфениколу, и векторы для глубокого секвенирования - маркер устойчивости к канамицину.



Кроме того, фагмида для отборов с использованием фагового дисплея содержит маркер устойчивости к ампициллину.

*Эксперимент для проверки и подтверждения концепции*

В этом эксперименте, авторы настоящего изобретения показали, что NestLink можно использовать для характеристики индивидуальных белков внутри большого пула белков-кандидатов не имеющим аналогов способом, и что можно идентифицировать члены пула с наиболее многообещающими характеристиками для выбранных нижестоящих применений.

Более конкретно, эксперимент для проверки и подтверждения концепции, описанный ниже, показывает, что i) разработан эффективный способ получения библиотеки гнездовым способом при хорошо контролируемом разнообразии библиотеки, и ii) что гнездовые библиотеки могут служить основой для селекционного давления на пулы связывающих членом не имеющим аналогов способом.

В этом примере, авторы настоящего изобретения работали с PLOI, состоящей из пула синтетел, которую предварительно обогащали посредством рибосомного и фагового дисплея (не описано) против связывающего мальтозу белка (MBP)

Авторы настоящего изобретения использовали способ NestLink, описанный в этом патенте, для приложения следующих видов селекционного давления к разнообразному пулу синтетел за один раз: i) отбор синтетел с наивысшей экспрессией, ii) отбор синтетел с наивысшей растворимостью и iii) отбор синтетел, которые связываются с мишенью в анализе связывания в растворе.

С использованием способа, описанного в разделе материалы и способы, авторы настоящего изобретения намеревались связать приблизительно 1200 отдельных членом пула синтетел с приблизительно 17000 уникальными метками Flycode, получая в результате так называемую «гнездовую библиотеку». Это проводили посредством первого культивирования соответствующего количества клонов клеток, содержащих кодирующий синтетела фагмиды, в одном сосуде, с последующим выделением их плазмидной ДНК. Вместо индивидуального отбора клонов синтетел, количество колониеобразующих единиц (КОЕ) на объем выделенных бактерий

определяли после трансформации посредством рассева на чашки с агаром. Таким образом, соответствующий объем выделенных бактерий (приблизительно 1200 КОЕ) использовали для инокуляции культуры, которую затем собирали для выделения плазмидной ДНК. Затем вставки ДНК из этих ограниченных по разнообразию фагмид лигировали в экспрессирующий вектор pNLx, содержащий библиотеку с меткой Flycode из приблизительно  $10^8$  отдельных вариантов. С использованием оценок КОЕ, как описано выше, количество клонов ограничивали до приблизительно 17000. Поскольку использовали только приблизительно 17000 содержащих метку Flycode векторов (как определено по оценке КОЕ) из  $10^8$  вариантов, авторы настоящего изобретения рассчитали, что 99,974% меток Flycode являются уникальными, и таким образом, подавляющее большинство меток Flycode метят одно уникальное синтетело. Кроме того, поскольку они вставили гнездовым способом приблизительно 1200 генов синтетел в приблизительно 17000 содержащих метки Flycode векторов, они ожидали, что среднее синтетело мечено 14 различными метками Flycode.

Гнездовую библиотеку в векторе pNLx экспрессировали в бактериях в одном флаконе и очищали в форме пула снабженных меткой flycode связывающих членов для проведения экспериментов отбора (см. ниже). Для секвенирования гнездовой библиотеки, снабженные меткой flycode синтетела переносили в вектор для глубокого секвенирования pNLs, несущий все соответствующие последовательности для глубокого секвенирования Illumina с использованием устройства MiSeq. Глубокое секвенирование гнездовой библиотеки обеспечило однозначное приписывание каждой метки Flycode соответствующему ей синтетелу. Данные глубокого секвенирования находились в соответствии с ожидаемыми количествами синтетела и Flycode в гнездовой библиотеке, поскольку 1080 отдельных последовательностей синтетел, связанных с 13620 уникальными метками Flycode, получены после фильтрации данных. В среднем каждое синтетело было, таким образом, связано 12,61 раз с различными метками Flycode. Авторы настоящего изобретения не наблюдали неоднозначного связывания метки Flycode с синтетелами после фильтрации данных секвенирования (т.е. одной

и той же метки Flycode, присоединенной к двум или более различным синтетелам). Эта успешная попытка вставки библиотек друг в друга гнездовым способом с использованием хорошо контролируемого разнообразия не имеет аналогов, насколько известно авторам настоящего изобретения.

С использованием данных глубокого секвенирования, базу данных, содержащую информацию о полной последовательности гнездовой библиотеки, конструировали посредством последовательного соединения всех меток Flycode каждого синтетела в теоретическую непрерывную белковую последовательность с использованием соответствующего синтетела в качестве идентификатора. Эту базу данных затем загрузили на сервер Mascot, для более позднего использования в поисках ионов MS/MS.

В качестве примера нового применения этой технологии, авторы настоящего изобретения использовали гнездовую библиотеку и специфически отбирали и идентифицировали синтетела с определенным кажущимся гидродинамическим радиусом и синтетела, для которых показано высокоаффинное взаимодействие с МВР в растворе. Обе этих характеристики определяли посредством эксклюзионной хроматографии (SEC), и они являлись недоступными с использованием систем дисплея из современного уровня техники, требующих связи генотип-фенотип, поскольку генотип увеличивает размер экспонируемого белка обычно более, чем в 100 раз, делая частицу для дисплея нечувствительной к небольшим изменениям размера на уровне белка.

С этой целью, гнездовую библиотеку экспрессировали, и снабженные меткой flycode связывающие члены очищали посредством смолы Ni-NTA и подвергали SEC. Элюированные фракции синтетел, которые соответствовали мономерным белкам (кандидаты на связывающие члены с наивысшей растворимостью), пулировали и разделяли на равные аликвоты. Одну аликвоту инкубировали с МВР, а другую только с буфером. Два образца анализировали по отдельности в SEC (анализ без МВР использовали в качестве контроля), и собирали фракции после элюции, соответствующие размеру комплекса синтетело-МВР. Затем метки Flycode из

собранных фракций для анализа с МВР и контрольного анализа выделяли и либо подвергали двум отдельным анализам LC-MS, либо объединяли в одном анализе LC-MS/MS после мечения выделенных меток Flycode изобарической меткой. Полученную ранее базу данных глубокого секвенирования (приписывания меток Flycode синтетелам) можно затем использовать для идентификации меток Flycode в поиске Mascot, таким образом, однозначно идентифицируя синтетела, элюированные с размером комплекса синтетело-МВР. Этот эксперимент позволил авторам настоящего изобретения идентифицировать более 300 уникальных синтетел, которые все хорошо экспрессировались, являлись мономерными и связывали белок-мишень в растворе.

#### *Применение NestLink для определения скорости диссоциации*

Для оценки в баллах специфических для МВР синтетел, которые были идентифицированы в вышеописанном эксперименте для проверки и подтверждения концепции, в соответствии с их скоростями диссоциации при связывании, авторы настоящего изобретения иммобилизовали равные количества выделенных комплексов МВР-синтетело посредством биотинилированного МВР на двух колонках с стрептавидин-сефарозой (фигура 4). Отбор по скорости диссоциации с использованием избытка небитинилированного МВР (промывка в течение 3 мин) затем проводили на одной колонке, в то время как другую колонку промывали только буфером. После промывок, оставшиеся синтетела с обеих колонок элюировали, и их метки Flycode подвергали двум анализам LC-MS/MS. Подобно вышеописанному эксперименту связывания в растворе (анализы SEC), базу данных глубокого секвенирования использовали в поисках Mascot для идентификации синтетел посредством меток Flycode. Кроме того, интенсивность MS1 для всех идентифицированных меток Flycode суммировали для каждого синтетела с использованием программного обеспечения Progenesis. Из-за количественного характера интенсивности пиков MS1, как определено выше, авторы настоящего изобретения ожидали, что соотношение между суммами интенсивности для меток flycode для каждого синтетела между двумя колонками может соответствовать их относительным концентрациям до и после отбора по скорости диссоциации с

использованием избытка небiotинилированной мишени. Принимая, что каждая реакция диссоциации следует одно-экспоненциальному спаду и используя известное время промывки избытком мишени (3 мин), авторы настоящего изобретения, таким образом, смогли определить приблизительные скорости диссоциации для более, чем 300 связывающих членов за один раз. Этот анализ подтвержден измерением скоростей диссоциации 11 индивидуальных связывающих членов с использованием поверхностного плазмонного резонанса. Определение скоростей диссоциации внутри пула кандидатов на связывание в одном эксперименте не имеет аналогов, насколько известно авторам настоящего изобретения. Способ, который ранее требовал несколько недель из-за необходимости обработки индивидуальных белков, можно в настоящее время осуществлять за один раз с использованием технологии, описанной в настоящем документе.

*Применение NestLink для идентификации связывающих молекул из иммунизированных верблюдовых*

NestLink использовали для пула природных нанотел, полученного посредством выделения кДНК из В-клеток иммунизированной альпаки (верблюдовые). Антиген, используемый для иммунизации, представлял собой TM287/288, транспортер ABC (интегральный мембранный белок) из *Thermotoga maritima*. В отличие от традиционного способа получения нанотел верблюдовых, этот пул нанотел не обогащали против мишени с использованием фагового дисплея.

Нанотела подвергали амплификации ПЦР, ограничивали по разнообразию и вставляли в библиотеку с метками Flycode с получением 3469 уникальных последовательностей нанотел, присоединенных к 59974 уникальных меток Flycode, как определено посредством глубокого секвенирования Illumina MiSeq (см. раздел материалы и способы). Гнездовую библиотеку экспрессировали и очищали посредством Ni-NTA, с последующим выделением мономерных членов пула посредством SEC. Аналогично эксперименту для проверки и подтверждения концепции (описанному выше), не обеспечивающие преимуществ кандидаты на связывание, которые не экспрессировались или которые были нерастворимыми, исключали на

этих стадиях предварительного отбора. Образец для LC-MS собирали после элюции с колонки Ni-NTA и из мономерной фракции после разделения SEC. Затем увеличивающиеся количества пула инкубировали с TM287/288 в соотношениях приблизительно 0,1:1, 2:1 и 100:1, и смеси антиген/пул снова подвергали трем разделением SEC (фигура 5). Фракции, соответствующие размеру комплекса мишень/нанотело, собирали. Метки Flycode из всех собранных образцов выделяли отдельно и анализировали посредством LC-MS/MS, что позволяло сравнение уровней экспрессии, растворимости (мономерные при SEC) и силы связывания с антигеном в растворе для всех связывающих членов за один раз.

В этом анализе 3469 уникальных нанотел иммунизированного верблюдового, авторы настоящего изобретения идентифицировали 27 семейств высокоаффинных связывающих членов с обеспечивающими преимуществами стабильностью, уровнями экспрессии и растворимостью. Примечательно, что NestLink являлся более эффективным, чем отборы с использованием фагового дисплея и избыточные общепринятые скрининги, с использованием ELISA и секвенирования по Сэнгеру, в которых идентифицировали только 5 из этих семейств в том же самом пуле за значительно более длительный период времени обработки. В общем, таким образом, можно утверждать, что NestLink можно использовать для идентификации наиболее многообещающих биомолекул-кандидатов из иммунизированных верблюдовых, с производительностью и точностью, нереализованными в способах современного уровня техники.

*Применение NestLink для идентификации связывающих молекул, нацеленных на белок на поверхности клеток*

Эксперименты, описанные выше, проводили с целью идентификации связывающих белков против очищенных мишеней/антигенов в растворе, с получением обеспечивающих преимуществ исследовательских инструментов для применений *in vitro*, таких как кристаллография. В настоящей работе, авторы настоящего изобретения намеревались исключить основной сдерживающий фактор разработки лекарственных средств, представляющий собой идентификацию связывающих мембранный белок молекул, узнающих белок-мишень с высокой специфичностью и

аффинностью на поверхности клеток. Разработка биомолекулярного лекарственного средства против мембранного белка-мишени, как правило, требует двух последовательных стадий, которые фундаментально отличаются. Во-первых, разнообразный пул кандидатов на связывание получают способами дисплея или иммунизации. Во-вторых, проводят скрининг разнообразного пула по связыванию и функции в клеточных анализах. Последний является заведомо неэффективным и медленным, поскольку он требует анализа кандидатов на связывание по одному (как правило, в миниатюризованном формате). В этом эксперименте, авторы настоящего изобретения заменили вторую стадию (скрининга) на NestLink, для идентификации связывающихся с поверхностью клетки молекул, специфических против интегрального мембранного белка-мишени без трудоемкого анализа индивидуальных кандидатов на связывание по одному.

Авторы настоящего изобретения сначала провели отбор в дисплее *in vitro* библиотеки синтетел против чистого, солюбилизованного в детергенте белкового антигена наружной мембраны грамотрицательной бактерии (стадия 1, получение разнообразного пула кандидатов на связывание). Вместо тестирования каждого индивидуального кандидата на связывание из этого разнообразного пула индивидуально по связыванию с клеточной поверхностью (обычно, стадия 2), авторы настоящего изобретения провели NestLink и тестировали большой пул кандидатов за один раз (фигура 6А). 1456 синтетел вставляли в библиотеку с метками flycode, получая связывание 31500 меток Flycode (в среднем 22 метки Flycode/синтетело). Как описано выше, приписывание метки flycode связывающему члену получали посредством глубокого секвенирования, и гнездовую библиотеку экспрессировали, очищали, и мономерные члены пула выделяли (отрицательный отбор/исключение нежелательных кандидатов на связывание). Таким образом, сначала исключали члены пула с плохими уровнями экспрессии и плохой растворимостью, и мониторировали уровни экспрессии и характеристики растворимости каждого члена пула. Способ NestLink, таким образом, отбирал исключительно многообещающие кандидаты на связывание в отборе на

поверхности клеток, который проводили следующим образом: мономерные члены пула разделяли на 4 эквивалентные фракции, и каждую фракцию инкубировали с другим бактериальным штаммом. Несвязывающиеся синтетела-кандидаты удаляли посредством осаждения и ресуспендирования/промывки с использованием буфера. Затем все метки Flycode из синтетел, связавшихся с одним из бактериальных штаммов, выделяли и подвергали анализу LC-MS. Сумму всех интенсивностей MS1 для всех меток Flycode на синтетело использовали в качестве показателя относительной концентрации каждого индивидуального синтетела в пуле на каждой из клеток-мишеней. Это обеспечивало точное специфическое для клеток считывание (фигура 6B).

Из 1456 кандидатов на связывание в пуле, идентифицировали 6 хорошо экспрессирующихся и растворимых синтетел, которые специфически узнавали белковую мишень в ее нативной форме, погруженную в наружную мембрану представляющей интерес грамотрицательной бактерии (штамм 4). Авторы настоящего изобретения подтвердили это обнаружение посредством индивидуального анализа этих 6 идентифицированных синтетел посредством проточной цитометрии против 4 штаммов (после их флуоресцентного мечения). Для всех тестируемых кандидатов показан одинаковый профиль специфичности в этом анализе отдельных клонов, как его наблюдали посредством NestLink. Следует отметить, что каждый из идентифицированных связывающих членов присутствовал только при  $< 0,05\%$  в гнездовом пуле, как определено посредством глубокого секвенирования Illumina MiSeq. С учетом того, что скрининги современного уровня техники принимают во внимание только одну характеристику кандидата на связывание (например, связывание с мишенью), но не могут обеспечивать регистрацию уровней экспрессии или подверженности растворимости/олигомеризации, является маловероятным, чтобы любой из этих шести связывающих членов мог быть идентифицирован классическим способом скрининга отдельных клонов. Таким образом, этот эксперимент показывает, что NestLink позволяет скрининг библиотек связывающих членов с не имеющей аналогов глубиной,



благодаря отсутствию связи генотип-фенотип и сочетанию двух библиотек.

*Применение NestLink для мониторинга биораспределения и фармакокинетических параметров в модельных организмах*

В предшествующих примерах, авторы настоящего изобретения показали, что отбор NestLink позволяет не имеющее аналогов селекционное давление, благодаря отсутствию связи генотип-фенотип (например, отбор мономерных членов пула/библиотеки в SEC). В настоящем описании, приведено другое селекционное давление, которого невозможно достигать в случае физической связи генотип-фенотип: отбор белков с конкретными биораспределением и фармакокинетическими свойствами в живых организмах. Гнездовой (меченный flycode) пул биомолекулярных лекарственных средств-кандидатов можно инъецировать в модели на животных, и относительную концентрацию каждого члена пула можно измерять по истечению определенного времени в каждой локализации в организме (например, в различных органах, тканях или опухолях и т.д.) посредством LC-MS. Анализ этого типа может обеспечивать исчерпывающий/глобальный анализ биораспределения для каждого индивидуального члена пула в организме в одной конкретной временной точке. Если несколько аналогичных индивидуумов одного вида подвергают этому анализу в меняющихся различных временных точках, анализ биораспределения NestLink можно расширять на измерение времени, таким образом, обеспечивая получение фармакокинетических данных с низким или средним разрешением по времени для каждого кандидата.

Авторы настоящего изобретения разработали основу для такого типа анализов посредством тестирования и оптимизации способов экстракции меток Flycode из гомогенизированных тканей мышей, в которые предварительно добавлены различные количества снабженных меткой flycode синтетел. Подробно, сначала несколько синтетел связывали с небольшим количеством меток Flycode (20-30), и приписывание синтетела метке flycode определяли посредством глубокого секвенирования Illumina MiSeq. Затем меченные Flycode синтетела экспрессировали и очищали индивидуально, и их концентрацию определяли посредством измерений оптической

плотности. Затем индивидуальные синтетела комбинируют в различных концентрациях (перекрывающих порядок величины).

Параллельно, замороженные органы (печень, легкие, почку) и кровь мышей размораживали и гомогенизировали с использованием условий денатурирующего буфера и гомогенизатора Поттера. Ранее подготовленную смесь для титрования добавляли в гомогенаты и инкубировали в течение 30 мин при комнатной температуре, чтобы позволить подействовать потенциальным протеазам или модифицирующим flycode ферментам. Затем синтетела вместе с их оставшимися метками Flycode экстрагировали, метки Flycode выделяли посредством отщепления протеазой и анализировали посредством LC-MS. На основании детекции индивидуальных синтетел из смеси для титрования, авторы настоящего изобретения обнаружили, что детекция синтетел посредством LC-MS из таких гомогенизированных органов и тканей, является, как правило, надежной при снижении до количества 30-100 нг (синтетела). Принимая во внимание, что вплоть до 1 мг лекарственных средств можно, как правило, инъецировать в модели на мышах, понятно, что в наиболее важных локализациях в организме, дюжины микрограмм могут присутствовать после инъекции гнездового пула. Таким образом, присутствует достаточное количество недеградированных и немодифицированных меток Flycode для мониторинга глобального биораспределения и для проведения фармакокинетических анализов пула связывающих членов.

### **Материалы и способы**

Ниже представлен общий протокол способа NestLink. Он включает все стадии, необходимые для проведения экспериментов, как описано выше, и предоставляет детали применительно к введению библиотеки гнездовым способом, глубокому секвенированию, экспрессии и очистке снабженных меткой flycode пулов связывающих членов, выделению метки Flycode, LC-MS и анализу данных.

*Клонирование снабженных меткой flycode нанотел посредством введения библиотеки гнездовым способом*

#### 1. Ограничение разнообразия пула синтетел/нанотел

Эксперименты NestLink проводили с использованием пулов синтетел или природных нанотел, полученных в отборах *in vitro* связывающих членов посредством фагового дисплея или иммунизаций, соответственно. В случае использования фагового дисплея для отбора связывающих членов, 200 нг отобранного *in vitro* пула потенциальных связывающих членов, кодируемого в фагмиде, трансформировали в 50 мкл химически компетентных клеток *E. coli* MC1061 (компетентность обеспечивали по протоколу из Promega Corporation, Subcloning Notebook 2004). Серии разведений рассеивали на чашки с агаром, содержащие 120 мкг/мл ампициллина, и инкубировали в течение ночи при 30°C. Колонии с чашки, содержащей желательное количество колониеобразующих единиц (в вышеуказанных примерах, количество лежало в диапазоне между 1000 и 1500 КОЕ), ресуспендировали в 2 мл среды LB, содержащей 100 мкг/мл ампициллина, и суспензию переносили в 200 мл культуру в среде LB, содержащей 100 мкг/мл ампициллина. Эту культуру выращивали в течение ночи при 37°C и использовали для получения ДНК (Набор: #740412.10, MACHEY-NAGEL). 15 мкг полученной фагмиды расщепляли посредством 100 единиц BspQI (New England Biolabs, # R0712L) в буфере NEB 3.1 (New England Biolabs, # B7203S) в реакционном объеме 140 мкл при 50°C в течение 1 час, с последующей тепловой инактивацией фермента при 80°C в течение 20 мин. Проводили электрофорез в 2% (масс./об.) агарозном геле, и полосу, соответствующую пулу связывающих членов, вырезали и экстрагировали (Набор: #740609,250, MACHEY-NAGEL). В случае иммунизированных альпак, последовательности нанотел амплифицировали с кДНК из В-клеток, как описано (Pardon et al., Nat Protoc., 2014 Mar;9(3):674-93) и амплифицировали с праймерами, содержащими участки рестрикции BspQI. 5 мкг очищенного продукта ПЦР расщепляли посредством 100 единиц BspQI (New England Biolabs, # R0712L) в буфере NEB 3.1 (New England Biolabs, # B7203S) в реакционном объеме 140 мкл при 50°C в течение 1 час, с последующей тепловой инактивацией фермента при 80°C в течение 20 мин. Проводили электрофорез в 2% (масс./об.) агарозном геле, и полосу, соответствующую пулу связывающих

членов, вырезали и экстрагировали (Набор: #740609,250, MACHERY-NAGEL). Расщепленный фрагмент после ПЦР клонировали в вектор для начального клонирования FX с маркером устойчивости к канамицину (Geertsma et al., *Biochemistry*, 2011 Apr 19;50(15):3272-8), и 3500 КОЕ ресуспендировали в 2 мл среды LB, содержащей 50 мкг/мл канамицина, и суспензию переносили в 200 мл культуру в среде LB, содержащей 50 мкг/мл канамицина. Эту культуру выращивали в течение ночи при 37°C и использовали для получения ДНК (Набор: #740412.10, MACHERY-NAGEL). 15 мкг полученной фагмиды расщепляли посредством 100 единиц BspQI (New England Biolabs, # R0712L) в буфере NEB 3.1 (New England Biolabs, # B7203S) в реакционном объеме 140 мкл при 50°C в течение 1 час, с последующей тепловой инактивацией фермента при 80°C в течение 20 мин. Проводили электрофорез в 2% (масс./об.) агарозном геле, и полосу, соответствующую пулу связывающих членов, вырезали и экстрагировали (Набор: #740609.250, MACHERY-NAGEL).

2. Присоединение меток Glycode к пулам синтетел/нанотел и ограничение разнообразия меток Glycode

Вектор pNLx, содержащий библиотеку с метками Glycode, расщепляли посредством BspQ1, как описано выше для фагмиды, и проводили электрофорез в 1% (масс./об.) агарозном геле. Полосу, соответствующую открытому вектору, вырезали и экстрагировали (Набор: #740609.250, MACHERY-NAGEL). 200 нг пула связывающих членов лигировали с 400 нг расщепленного pNLx с использованием 2,5 единиц T4-лигазы (Fermentas #EL0011) в буфере для T4-лигазы (Fermentas #B69) в реакционном объеме 28 мкл при 37°C в течение 1 час, с последующей тепловой инактивацией при 65°C в течение 10 мин. 25 мкл реакционной смеси после лигирования использовали для трансформации в 150 мкл электрокомпетентных клеток *E. coli* MC1061 (полученных в соответствии с Howard and Kaser 2007, *Making and using antibodies*, page 170). Клетки восстанавливали в течение 30 мин при 37°C в среде SOC, и 200 мл культуры, содержащей 25 мкг/мл хлорамфеникола, инокулировали объемом восстановленных бактерий, соответствующим желательному количеству колониеобразующих единиц, как определено по рассеvu

разведенного образца на чашку с агаром, содержащим 25 мкг/мл хлорамфеникола (в вышеуказанных примерах количество КОЕ лежало в диапазоне между 13000 и 30000). Культуру выращивали в течение ночи при 37°C, с последующим получением ДНК (Набор: #740412.10, MACHERY-NAGEL) и получением препарата для хранения в глицерине, содержащего 1 мл культуры в стационарной фазе, смешанной с 1 мл 50% (об./об.) глицерина.

#### *Глубокое секвенирование*

1. Присоединение адаптерных последовательностей Illumina

15 мкг рNLx, содержащего снабженные меткой flycode связывающие члены, расщепляли посредством 120 единиц SfiI (Fermentas # ER1821) в Буфере G (Fermentas # BG5) в реакционном объеме 140 мкл при 50°C в течение 3 час, с последующим добавлением 12 мкл 0,5 М ЭДТА для инактивации фермента. Проводили электрофорез в 2% агарозном геле, и полосу, соответствующую пулу связывающих членов, связанных с метками Flycode, вырезали и экстрагировали (Набор: #740609.250, MACHERY-NAGEL). Для первого примера с синтетелами против MBP, вектор рNLs, содержащий адаптеры, подходящие для глубокого секвенирования ДНК посредством Illumina MiSeq с соответствующим индексом (в этом случае, 502 и 703 использовали для двойного индексирования), расщепляли посредством SfiI, как описано для рNLx выше, и проводили электрофорез в 1% агарозном геле. Полосу, соответствующую остову вектора, вырезали и экстрагировали (Набор: #740609.250, MACHERY-NAGEL). 400 нг пула снабженных меткой flycode связывающих членов лигировали в 300 нг расщепленного рNLx с использованием 2,5 единиц T4-лигазы (Fermentas #EL0011) в буфере для T4-лигазы (Fermentas #B69) в реакционном объеме 28 мкл при 37°C в течение 1 час, с последующей тепловой инактивацией при 65°C в течение 10 мин. 25 мкл реакционной смеси после лигирования использовали для трансформации в 250 мкл электрокомпетентных клеток *E. coli* MC1061 (полученных в соответствии с Howard and Kaser 2007, Making and using antibodies, page 170). Клетки восстанавливали в течение 45 мин при 37°C в среде SOC, и 200 мл культуры,

содержащей 30 мкг/мл канамицина, инокулировали всеми восстановленными клетками. Тестируемый образец рассевали на чашки с селективным агаром с канамицином, чтобы подтвердить, что эффективность лигирования и трансформации являлась достаточной для переноса всей гнездовой библиотеки (всего > 200000 КОЕ). Культуру выращивали в течение ночи при 37°C, с последующим получением ДНК (Набор: #27106, QUIAGEN). Рестрикционное расщепление 1 мкг полученного pNLs, содержащего пул снабженных меткой flycode связывающих членов, проводили с использованием 5 единиц BseRI (New England Biolabs, #R0581S) в буфере CutSmart (New England Biolabs, #B7204S) в общем реакционном объеме 20 мкл при 37°C в течение 2 час, с последующей тепловой инактивацией при 80°C в течение 20 мин. Следует отметить, что в этой точке несколько снабженных меткой flycode пулов против различных мишеней можно пулировать (перед расщеплением BseRI), где каждый помещен в снабженный различным индексом pNLs. Вставку, содержащую пул снабженных меткой flycode связывающих членов, присоединенных к адаптерам MiSeq, затем экстрагировали из 1% агарозного геля.

Для других примеров, представленных выше, 300-400 нг гибризованных олигонуклеотидов, содержащих липкие выступающие концы SfiI, смешивали с 600 нг пула снабженных меткой flycode связывающих членов, вырезанного из pNLx посредством SfiI, с использованием 5 единиц T4-лигазы (Fermentas #EL0011) в буфере для T4-лигазы (Fermentas #B69) в реакционном объеме 20 мкл при 37°C в течение 1 час, с последующей тепловой инактивацией при 65°C в течение 10 мин. Пул снабженных меткой flycode связывающих членов, присоединенных к адаптерам MiSeq, затем вырезали из 2% агарозного геля (Набор: #740609.250, MACHERY-NAGEL). Следует отметить, что в этой точке несколько снабженных меткой flycode пулов против различных мишеней можно пулировать, где каждый содержит различную пару лигированных адаптеров.

## 2. Определение связей нанотело-метка Flycode

Глубокое секвенирование проводили в устройстве MiSeq от Illumina с использованием способа спаренных концов (Набор

реагентов v2 MiSeq (300-циклов)). На первой стадии анализа, прочтения для спаренных концов сшивали вместе с использованием стандартного программного обеспечения (Illumina). Для любой данной пары индексов, получали всего 800000-8 миллионов прочтений, что соответствует средней избыточности прочтения 25-70 (это количество равно общему количеству прочтений, деленному на общее ожидаемое количество меток flycode для данной гнездовой библиотеки). С использованием разработанного на заказ сценария, полученные необработанные прочтения фильтровали посредством применения следующих положительных критериев: i) корректный фланкирующий паттерн неизменных частей метки Flycode, ii) корректный фланкирующий паттерн неизменных частей нанотел, iii) последовательность не содержит N, iv) последовательность лежит в ожидаемом диапазоне размеров возможных слитых белков нанотело-метка Flycode, v) последовательность слитого белка нанотело-метка Flycode находится в рамке считывания (т.е. является кратной 3) vi) последовательность лишена стоп-кодонов. После фильтрации, получали список уникальных меток Flycode. Метки Flycode, прочтенные по меньшей мере пять раз, считали корректными. Для каждой корректной метки Flycode, получали консенсусную последовательность всех связанных последовательностей нанотела. Способ консенсусной последовательности являлся необходимым для коррекции ошибок секвенирования в последовательности нанотела. Консенсусный показатель вводили для мониторинга изменчивости среди последовательностей нанотела, присоединенных к одной и той же метке Flycode. Этот показатель вводит большие штрафы в случае, когда одно или более нанотел, присоединенных к идентичной метке Flycode, являются явно отличными друг от друга, таким образом, удаляя метки Flycode, связанные с двумя или более различными нанотелами, из дальнейшего анализа. Только пары нанотело-метка Flycode с высоким консенсусным показателем учитывали далее. На конечной стадии, идентифицировали идентичные (консенсусные) последовательности нанотела и все связанные с ним метки Flycode (в среднем 12-40 меток Flycode на нанотело в вышеуказанных примерах). Все метки Flycode, связанные с одним и тем же

нанотелом, соединяли последовательно в гипотетическую белковую последовательность с использованием последовательности нанотела в качестве идентификатора, и эту базу данных сохраняли в формате файла fasta.

*Экспрессия и очистка мономерных снабженных меткой flycode синтетел/нанотел*

Препараты *E. coli* MC1061 для хранения в глицерине, содержащие рNLx, несущие пул снабженных меткой flycode связывающих членов, использовали для инокуляции 50 мл предварительной культуры LB, содержащей 1% глюкозу, которую культивировали в течение ночи при 37°C. 600 мл культуры ТВ инокулировали посредством предварительной культуры до OD 0,05 и культивировали в течение 1,5 час при 37°C с последующим культивированием в течение ночи при 20°C. Индукцию проводили при OD<sub>600</sub> 0,8 посредством 0,05% (масс./об.) арабинозы. Клетки собирали посредством центрифугирования при 5000 g в течение 20 мин. Супернатант декантировали, и клетки ресуспендировали в 25 мл 50 mM Tris-HCl pH 7,5 (20°C), 150 mM NaCl, 15 mM имидазола pH 8,0 (20°C), дополненного небольшим количеством ДНКазы I (SIGMA #DN25). Клетки лизировали с использованием микрофлюидизатора (Microfluidics #110P) при 30000 фунт/кв. дюйм (207 МПа) за два цикла, при охлаждении на льду. Клеточный дебрис осаждали при 5000 g в течение 30 мин, и супернатант наносили на колонку 1,5 мл Ni-NTA superflow (QUIAGEN # 1018142) под действием силы тяжести. Колонку промывали посредством 30 мл буфера для промывки, содержащего 20 mM Tris-HCl pH 7,5 (20°C), 150 mM NaCl и 30 mM имидазол pH 8 (20°C). Элюцию проводили посредством 6 мл 20 mM Tris-HCl pH 7,5 (20°C), 150 mM NaCl и 300 mM имидазола pH 8 (20°C). 5 мл элюата наносили на HiLoad 16/600 Superdex 200 pg (GE Healthcare Life Sciences #28989335), и область, соответствующую мономерной фракции, собирали и концентрировали до объема 1,2 мл при оптической плотности (280 нм) 2,1 в Nanodrop 2000c (Thermo Scientific) против буфера для экспериментов с дальнейшим отбором, как описано в вышеуказанных примерах.



*Выделение меток Flycode*

Образцы, содержащие снабженную меткой flycode PLOI, разводили в 10-20 раз буфером Eх (20 мМ Tris-HCl pH 8,5, 150 мМ NaCl, 0,5% (об./об.) Triton X-100, 0,125% (масс./об.) дезоксихолат натрия, 10 мМ имидазол pH 8,0, 4,5 мМ GdmCl), фильтровали (шприцевой фильтр с отсечением 0,2 мкм) и инкубировали с 100 мкл взвеси Ni-NTA superflow (QUIAGEN # 1018142) в течение 2 час при комнатной температуре. Затем смолу осаждали при 500 g в течение 10 мин и переносили в миниколонку для хроматографии bio-spin, с последующими 3×500 мкл промывками с использованием буфера Eх, 3×500 мкл с использованием буфера TH (20 мМ TEAB pH 8,0, 150 мМ NaCl, 2,5 мМ CaCl<sub>2</sub>), содержащего 30 мМ имидазол pH 8,0, и 3×500 мкл буфера TH. После закрытия нижнего конца колонки смолу ресуспендировали в 100 мкл буфера TH, содержащего 2,4 ед. тромбина (MILLIPORE #69671-3), с последующей инкубацией в течение ночи при комнатной температуре. Затем из колонки спускали жидкость, и промывали ее 3×500 мкл буфера TH, содержащего 30 мМ имидазол pH 8,0, с последующими 3×500 мкл буфера TRY (20 мМ TEAB pH 8,0, 50 мМ NaCl, 2,5 мМ CaCl<sub>2</sub>) и элюцией посредством буфера TRY, содержащего 300 мМ имидазол, pH 8,0. Элват центрифугировали (15000 g) через предварительно уравновешенный (H<sub>2</sub>O) концентратор Microcon с порогом отсечения 10 кДа (AMICON: YM-10), и 1 мкг трипсина (PROMEGA #V5113) добавляли к фильтрату с последующей инкубацией в течение ночи при 37°C.

Затем элжированные образцы меток Flycode подвергали очистке способом ZipTip (MILLIPORE #ZTC18S960). Наконечники ZipTip предварительно промывали 200 мкл метанола, 200 мкл 60% (об./об.) ацетонитрила (ACN) и 200 мкл 3% (об./об.) ацетонитрила, содержащего 0,1% (об./об.) трифторуксусной кислоты. Наносили 100 мкл образца Flycode, с последующей промывкой с использованием 200 мкл 3% (об./об.) ацетонитрила, содержащего 0,1% (об./об.) трифторуксусной кислоты, и элюцией посредством 2×40 мкл 60% (об./об.) ацетонитрила, содержащего 0,1% (об./об.) трифторуксусной кислоты. Затем растворитель выпаривали (speedvac), и метки Flycode ресуспендировали в 15 мкл 3%

(об./об.) ацетонитрила, содержащего 0,1% (об./об.) муравьиной кислоты.

#### *LC-MS*

С использованием системы Easy-nLC 1000 HPLC, 2 мкл раствора ресуспендированной метки Flycode инъецировали в изготовленную в собственной лаборатории капиллярную колонку, набитую обращеннофазовым материалом (ReproSil-Pur 120 C18-AQ, 1,9 мкм; размеры колонки 150 мм x 0,075 мм). Колонку уравнивали растворителем А (0,1% муравьиная кислота (ФА) в воде). Пептиды элюировали при скорости потока 0,3 мкл/мин с использованием следующего градиента: 0-60 мин; 5-20% В (0,1% ФА в АСН), 60-70 мин; 20-97% В. После 10 мин промывки посредством 97% В, колонку повторно уравнивали растворителем А в течение 5 мин. Высокоточные масс-спектры получали с использованием масс-спектрометра Orbitrap Fusion (Thermo Scientific) с использованием следующих параметров: диапазон сканирования 300-1500 m/z, AGC-мишень 5e5, разрешение 120000 (при m/z 190) и максимальное время инъекции 100 мс. Зависимые от данных спектры MS/MS регистрировали в режиме наивысшей скорости в линейной ионной ловушке с использованием выделения квадруполь (окно 1,6 m/z), AGC-мишень 1e4, максимальное время инъекции 35 мс, HCD-фрагментация с 30% энергией столкновения, максимальное время цикла 3 с, и разрешены все возможные параллелизуемые периоды времени. Сигналы моноизотопных предшественников отбирали для MS/MS с состояниями заряда между 2 и 6, и минимальной интенсивностью сигнала 5e4. Динамическое исключение устанавливали на 25 с, и окно исключения - на 10 м.д. После сбора данных, списки пиков получали с использованием Proteome Discoverer 1.4 (Thermo Scientific).

#### *Анализ данных и количественная оценка*

Анализы LC-MS (один анализ на выделение метки flycode/образец) предварительно проверяли посредством программного обеспечения Xcalibur, и необработанные файлы Xcalibur импортировали и конвертировали посредством Progenesis в файлы mznlд. Progenesis затем использовали для выравнивания представляющих интерес результатов анализов LC-MS (показатель

выравнивания > 80%) и удаляли пептидные ионы с зарядами +1 и +5 - +20 из анализа. Объединенный файл mgf для всех выровненных результатов анализов LC-MS затем экспортировали из Progenesis (порог ранга < 5, количество фрагментов ионов > 1000, удаление изотопов и деконволюция заряда) и загружали на сервер mascot вместе с ранее определенным приписыванием метки flycode PLOI-члену (база данных глубокого секвенирования в формате файла fasta, см. выше). Идентификации Mascot напрямую импортировали в программное обеспечение Scaffold, с последующим переводом данных и экспортом регистрации спектра, которые затем импортировали в Progenesis, позволяющее приписывание признаков соответствующим им меткам flycode. С использованием Progenesis, признаки интенсивности, как правило, нормализовали по добавленным стандартам, и все уникальные метки flycode для каждого PLOI-члена использовали для количественной оценки. Необработанные и нормализованные интенсивности затем экспортировали (формат CSV) и далее анализировали посредством Excel.

**ФОРМУЛА ИЗОБРЕТЕНИЯ**

1. Способ отбора полипептида из библиотеки полипептидов, включающий стадии

а. получения первой библиотеки нуклеиновых кислот, где каждый член указанной первой библиотеки нуклеиновых кислот содержит кодирующую полипептид последовательность, кодирующую член первой библиотеки полипептидов;

б. получения второй библиотеки нуклеиновых кислот, где указанная вторая библиотека нуклеиновых кислот содержит множество членов, где каждый член содержит кодирующую метку последовательность, кодирующую метку для детекции, где указанная метка для детекции:

i. характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой указанной второй библиотекой нуклеиновых кислот;

ii. характеризуется молекулярной массой между 200 и 5000 Да, в частности, между 500 и 2500 Да, более конкретно, между приблизительно 900 и 2200 Да; и

iii. содержит первый разделяемый элемент;

с. вставки указанной кодирующей полипептид последовательности, содержащейся в указанном члене указанной первой библиотеки нуклеиновых кислот, в член указанной второй библиотеки нуклеиновых кислот, таким образом, получения библиотеки меченых нуклеиновых кислот, кодирующей библиотеку меченых полипептидов, где каждый член указанной библиотеки меченых полипептидов содержит полипептид и метку для детекции, отделенную от указанного полипептида указанным первым разделяемым элементом;

д. получения множества последовательностей нуклеиновых кислот из указанной библиотеки меченых нуклеиновых кислот, где каждая из указанного множества последовательностей нуклеиновых кислот содержит кодирующую полипептид последовательность и кодирующую метку последовательность;

e. прогнозирования паттерна фрагментации при масс-спектрометрии для каждой метки для детекции, кодируемой кодирующей метку последовательностью, полученной на стадии d;

f. экспрессии указанной библиотеки меченых полипептидов из указанной библиотеки меченых нуклеиновых кислот;

g. отбора члена указанной библиотеки меченых полипептидов на стадии отбора с получением отобранного полипептида;

h. разделения указанного первого разделяемого элемента, таким образом, отделения указанной метки для детекции от указанного отобранного полипептида с получением выделенной метки для детекции;

i. идентификации указанной выделенной метки для детекции посредством

i. регистрации паттерна фрагментации указанной выделенной метки для детекции посредством масс-спектрометрии;

ii. приведения указанного паттерна фрагментации, полученного на стадии i, в соответствие с указанными паттернами фрагментации, прогнозированными на стадии e, таким образом, идентификации указанной выделенной метки для детекции;

j. отбора из указанного множества последовательностей нуклеиновых кислот, полученных на стадии d, последовательности нуклеиновой кислоты, содержащей кодирующую метку последовательность, кодирующую указанную метку для детекции, идентифицированную на стадии i, таким образом, идентификации члена указанной библиотеки меченых полипептидов, ассоциированного с указанной меткой для детекции, идентифицированной на стадии i.

2. Способ по п.1, где указанная выделенная метка для детекции характеризуется уровнем гидрофобности между -27 и 128, в частности, между -1 и 70.

3. Способ по любому из вышеуказанных пп., где указанный член указанной библиотеки меченых полипептидов является ассоциированным с аффинной меткой, в частности, с аффинной меткой, выбранной из группы, содержащей His-метку, СВР-метку, СYD-метку, Strep-метку, StrepII-метку, FLAG-метку, НРС-метку,

GST-метку, Avi-метку, биотинилированную метку, Мус-метку, 3xFLAG-метку и MBP-метку.

4. Способ по любому из вышеуказанных пп., где указанная метка для детекции является ассоциированной с аффинной меткой, в частности, с аффинной меткой, выбранной из группы, содержащей His-метку, CBP-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, NPC-метку, GST-метку, Avi-метку, биотинилированную метку, Мус-метку, 3xFLAG-метку и MBP-метку.

5. Способ по п.4, где указанная аффинная метка отделена от указанной метки для детекции вторым разделяемым элементом, и указанный второй разделяемый элемент разрезают до стадии *i*.

6. Способ по любому из вышеуказанных пп., где стадия *i* включает анализ указанной выделенной метки для детекции посредством жидкостной хроматографии в сочетании с масс-спектрометрией с ионизацией электрораспылением (LC-MS).

7. Способ по любому из вышеуказанных пп., где стадия *d* включает секвенирование указанной полной меченой экспрессирующей библиотеки с перекрытием  $\geq 5x$ .

8. Способ по любому из вышеуказанных пп., где указанная выделенная метка для детекции состоит из 5-30, в частности, 7-21, более конкретно, 11-15 аминокислот и содержит только одну аминокислоту, имеющую положительно заряженную боковую цепь.

9. Способ по любому из вышеуказанных пп., где указанная выделенная метка для детекции содержит элемент последовательности *I*, выбранный из коллекции элементов последовательности *I*, где указанный элемент последовательности *I* состоит из 5-10, в частности, 7 аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R.

10. Способ по любому из вышеуказанных пп., где указанная одна аминокислота, имеющая положительно заряженную боковую цепь, локализована на C-конце указанной выделенной метки для детекции, в частности, указанная одна аминокислота, имеющая положительно заряженную боковую цепь, представляет собой C-концевой аргинин, и остальные аминокислоты, содержащиеся в выделенной метке для

детекции, независимо выбраны из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и P.

11. Способ по любому из вышеуказанных пп., где указанная выделенная метка для детекции содержит

а. указанный элемент последовательности I, где указанный элемент последовательности I состоит из 5-10, в частности, 7 аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и P; и

б. элемент последовательности II, выбранный из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

12. Способ по любому из вышеуказанных пп., где указанная выделенная метка для детекции состоит из

а. элемента последовательности III, где указанный элемент последовательности III представляет собой GS;

б. указанного элемента последовательности I, где указанный элемент последовательности I состоит из 5-10, в частности, 7 аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и P; и

с. указанного элемента последовательности II, выбранного из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR);

где, в частности, порядок указанных элементов последовательности от N-конца к С-концу представляет собой элемент последовательности III, элемент последовательности I, элемент последовательности II.

13. Способ по любому из вышеуказанных пп. 9-12, где все элементы последовательности I, содержащиеся в указанной первой библиотеке нуклеиновых кислот, совместно составляют коллекцию элементов последовательности I, и где внутри указанной коллекции элементов последовательности I, каждая аминокислота встречается с частотой, указанной в таблице 1.

14. Способ по любому из вышеуказанных пп., где указанный первый и/или указанный второй разделяемый элемент представляет собой или содержит последовательность узнавания протеазой.

15. Способ по любому из вышеуказанных пп., где

а. указанный первый разделяемый элемент представляет собой или содержит последовательность узнавания тромбином и/или

б. указанный второй разделяемый элемент представляет собой или содержит последовательность узнавания трипсином.

16. Коллекция полипептидов, где каждый член указанной коллекции полипептидов является ассоциированным с меткой для детекции, в частности, по меньшей мере с одной, более конкретно, по меньшей мере с двумя, даже более конкретно, по меньшей мере с пятью, даже более конкретно, по меньшей мере с 10, даже более конкретно, с приблизительно двадцатью метками для детекции, и где указанная метка для детекции

а. характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой указанным множеством экспрессирующих векторов;

б. характеризуется молекулярной массой между 200 и 5000 Да, в частности, между 500 и 2500 Да, более конкретно, между приблизительно 900 и 2200 Да;

с. отделена от указанного члена указанной коллекции полипептидов первым разделяемым элементом.

17. Коллекция полипептидов по п.16, где указанная выделенная метка для детекции характеризуется уровнем гидрофобности между -27 и 128, в частности, между -1 и 70.

18. Коллекция полипептидов по любому из пп. 16-17, где каждый член указанной коллекции полипептидов является ассоциированным с аффинной меткой, в частности, с аффинной меткой, выбранной из группы, содержащей His-метку, CBP-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, HPC-метку, GST-метку, Avi-метку, биотинилированную метку, Muc-метку, 3xFLAG-метку и MBP-метку.

19. Коллекция полипептидов по любому из пп. 16-18, где указанная метка для детекции является ассоциированной с аффинной меткой, в частности, с аффинной меткой, выбранной из группы, содержащей His-метку, CBP-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, HPC-метку, GST-метку, Avi-метку, биотинилированную метку, Muc-метку, 3xFLAG-метку и MBP-метку,



где указанная аффинная метка отделена от указанной метки для детекции вторым разделяемым элементом.

20. Коллекция полипептидов по любому из пп. 16-19, где указанная метка для детекции состоит из 4-20, в частности, 7-18, более конкретно, 11-15 аминокислот и содержит только одну аминокислоту, имеющую положительно заряженную боковую цепь.

21. Коллекция полипептидов по любому из пп. 16-20, где указанная метка для детекции содержит

а. элемент последовательности I, где указанный элемент последовательности I состоит из 5-10, в частности, 7 аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R; и

б. элемент последовательности II, выбранный из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

22. Метка для детекции, состоящая из 4-20, в частности, 7-18, более конкретно, 11-15 аминокислот, где указанная метка для детекции

а. содержит только одну аминокислоту, имеющую положительно заряженную боковую цепь;

б. характеризуется молекулярной массой между 200 и 5000 Да, в частности, между 500 и 2500 Да, более конкретно, между 900 и 2200 Да.

23. Метка для детекции по п.22, где указанная метка для детекции в основном состоит из

а. элемента последовательности I, где указанный элемент последовательности I состоит из 5-10, в частности, 7 аминокислот, независимо друг от друга выбранных из A, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R; и

б. элемента последовательности II, выбранного из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

24. Коллекция меток для детекции, содержащая по меньшей мере 96, более конкретно, по меньшей мере 500000, даже более конкретно, по меньшей мере  $10^7$  меток для детекции, даже более конкретно, приблизительно  $10^8$  меток для детекции по любому из пп.

19 или 20, где каждая метка для детекции состоит из 4-20, в частности, 7-18, более конкретно, 11-15 аминокислот и характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, содержащейся в указанной коллекции меток для детекции.

25. Коллекция меток для детекции по п.24, где каждая метка для детекции содержит только одну аминокислоту, имеющую положительно заряженную боковую цепь, и остальные аминокислоты выбраны из А, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R.

26. Коллекция меток для детекции по любому из пп. 24 или 25, где каждая метка для детекции характеризуется уровнем гидрофобности между -27 и 128, в частности, между -1 и 70.

27. Коллекция меток для детекции по любому из пп. 24-26, где каждая метка для детекции является ассоциированной с аффинной меткой, в частности, с аффинной меткой, выбранной из группы, содержащей His-метку, CBP-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, HPC-метку, GST-метку, Avi-метку, биотинилированную метку, Muc-метку, 3xFLAG-метку и MBP-метку, более конкретно, His-метку, где указанная аффинная метка отделена от указанной метки для детекции разделяемым элементом.

28. Коллекция плазмидных векторов, в частности, по меньшей мере из 96, более конкретно, по меньшей мере из 500000, даже более конкретно, по меньшей мере из  $10^7$  плазмидных векторов, даже более конкретно, приблизительно из  $10^8$  плазмидных векторов, где каждый член указанной коллекции плазмидных векторов содержит кодирующую метку последовательность нуклеиновой кислоты, кодирующую метку для детекции, где каждая метка для детекции состоит из 4-20, в частности, 7-18, более конкретно, 11-15 аминокислот и характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой указанной коллекцией плазмидных векторов.

29. Коллекция плазмидных векторов по п.28, где указанная метка для детекции содержит только одну аминокислоту, имеющую положительно заряженную боковую цепь.

30. Коллекция плазмидных векторов по любому из пп. 28-29, где указанная метка для детекции характеризуется массой между 200 и 5000 Да, в частности, между 500 и 2500 Да, более конкретно, между приблизительно 900 Да и приблизительно 2200 Да.

31. Коллекция плазмидных векторов по любому из пп. 28-30, где указанная кодируемая метка для детекции характеризуется уровнем гидрофобности между -27 и 128, в частности, между -1 и 70.

32. Коллекция плазмидных векторов по любому из пп. 28-31, где указанная метка для детекции является ассоциированной с аффинной меткой, в частности, с аффинной меткой, выбранной из группы, содержащей His-метку, СВР-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, НРС-метку, GST-метку, Avi-метку, биотинилированную метку, Мус-метку, 3xFLAG-метку и МВР-метку, более конкретно, His-метку, где указанная аффинная метка отделена от указанной метки для детекции разделяемым элементом.

33. Коллекция плазмидных векторов по любому из пп. 28-32, где указанная метка для детекции в основном состоит из

а. элемента последовательности I, где указанный элемент последовательности I состоит из 5-10, в частности, 7 аминокислот, независимо друг от друга выбранных из А, S, T, N, Q, D, E, V, L, I, F, Y, W, G и R; и

б. элемента последовательности II, выбранного из SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) и SEQ ID NO 05 (WQEGGR).

34. Коллекция плазмидных векторов по любому из пп. 28-33, где каждый член указанной коллекции плазмидных векторов содержит

а. кассету для отрицательного отбора, фланкированную на 5' первым участком рестрикции эндонуклеазой и на 3' вторым участком рестрикции эндонуклеазой;

б. промотор, локализованный на 5' от указанного первого участка рестрикции эндонуклеазой;

с. указанную последовательность нуклеиновой кислоты метки кодирующую указанную метку для детекции, локализованную на 3' от указанного второго участка рестрикции эндонуклеазой.

35. Коллекция плазмидных векторов по любому из пп. 28-33, где каждый член указанной коллекции плазмидных векторов содержит

a. указанную последовательность нуклеиновой кислоты метки, кодирующую указанную метку для детекции [ассоциированную внутри одной и той же рамки считывания с последовательностью нуклеиновой кислоты, кодирующей полипептид]

b. обеспечивающий разнообразие элемент, в частности, обеспечивающий разнообразие элемент, содержащий неидентичные основания, для предотвращения перегрузки сигнала в ходе секвенирования

c. участок связывания праймера, в частности, для связывания праймеров для секвенирования

d. индексный элемент, в частности, участок связывания праймера, содержащий одну из нескольких определенных последовательностей нуклеиновой кислоты, для обеспечения мультиплексности

e. адаптерный элемент, в частности, адаптерный элемент для иммобилизации молекулы ДНК в ходе секвенирования, и

f. два участка рестрикции эндонуклеазами, в частности, два участка рестрикции эндонуклеазами, фланкирующие элементы a-e, для высвобождения фрагмента ДНК из плазмиды.

36. Способ детекции белка, включающий

a. получение библиотеки нуклеиновых кислот, кодирующей библиотеку полипептидов,

где указанная библиотека полипептидов содержит множество членов, и каждый член является ассоциированным с меткой для детекции, и где указанная метка для детекции

i. характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой указанной библиотекой нуклеиновых кислот;

ii. характеризуется молекулярной массой между 200 и 5000 Да, в частности, между 500 и 2500 Да, более конкретно, между приблизительно 900 и 2200 Да; и

iii. отделена от указанного члена указанной коллекции полипептидов первым разделяемым элементом;

b. получения базы данных, содержащей

i. множество последовательностей нуклеиновых кислот и/или аминокислот, где указанное множество последовательностей содержит последовательности всех членов указанной библиотеки нуклеиновых кислот, и где каждая из указанных последовательностей содержит последовательность, определяющую полипептид, и последовательность, определяющую метку для детекции;

ii. прогнозированный паттерн фрагментации при масс-спектрометрии для каждой метки для детекции, кодируемой указанной библиотекой нуклеиновых кислот;

c. экспрессии указанной библиотеки полипептидов из указанной библиотеки нуклеиновых кислот;

d. отбора члена указанной библиотеки полипептидов на стадии отбора, с получением отобранного полипептида;

e. разделения указанного первого разделяемого элемента, таким образом, отделения указанной метки для детекции от указанного отобранного полипептида, с получением выделенной метки для детекции;

f. идентификации указанной выделенной метки для детекции посредством

i. регистрации паттерна фрагментации указанной выделенной метки для детекции посредством масс-спектрометрии;

ii. приведения указанного паттерна фрагментации, полученного на стадии i, в соответствие с указанными паттернами фрагментации, прогнозированными в указанной базе данных, таким образом, идентификации указанной выделенной метки для детекции;

g. отбора из указанного множества последовательностей, содержащихся в указанной базе данных, последовательности, определяющей указанную метку для детекции, идентифицированную на стадии f, таким образом, идентификации члена указанной библиотеки полипептидов, ассоциированного с указанной меткой для детекции, идентифицированной на стадии f.

37. Способ по п.36, где каждый член указанной библиотеки полипептидов является ассоциированным с аффинной меткой, в частности, с аффинной меткой, выбранной из группы, содержащей

His-метку, СВР-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, НРС-метку, GST-метку, Avi-метку, биотинилированную метку, Мус-метку, 3xFLAG-метку и МВР-метку.

38. Способ по п.36, где каждая метка для детекции является ассоциированной с аффинной меткой, в частности, с аффинной меткой, выбранной из группы, содержащей His-метку, СВР-метку, CYD-метку, Strep-метку, StrepII-метку, FLAG-метку, НРС-метку, GST-метку, Avi-метку, биотинилированную метку, Мус-метку, 3xFLAG-метку и МВР-метку.

39. Способ по п.38, где указанная аффинная метка отделена от указанной метки для детекции вторым разделяемым элементом, и указанный второй разделяемый элемент разрезают до стадии f.

40. Способ ассоциации полипептида с уникальной меткой для детекции, включающий стадии

а. получения первой библиотеки нуклеиновых кислот, где каждый член указанной первой библиотеки нуклеиновых кислот содержит кодирующую полипептид последовательность, кодирующую член первой библиотеки полипептидов;

б. получения второй библиотеки нуклеиновых кислот, где каждый член указанной второй библиотеки нуклеиновых кислот содержит кодирующую метку последовательность, кодирующую метку для детекции, где указанная метка для детекции:

i. характеризуется аминокислотной последовательностью, отличной от аминокислотной последовательности любой другой метки для детекции, кодируемой указанной второй библиотекой нуклеиновых кислот;

ii. характеризуется молекулярной массой между 200 и 5000 Да, в частности, между 500 и 2500 Да, более конкретно, между приблизительно 900 и 2200 Да;

с. вставки указанной кодирующей полипептид последовательности, содержащейся в указанном члене указанной первой библиотеки нуклеиновых кислот, в член указанной второй библиотеки нуклеиновых кислот, где

i. указанная первая библиотека нуклеиновых кислот имеет размер 5-100000, в частности, 100-50000, более конкретно, 500-5000, и

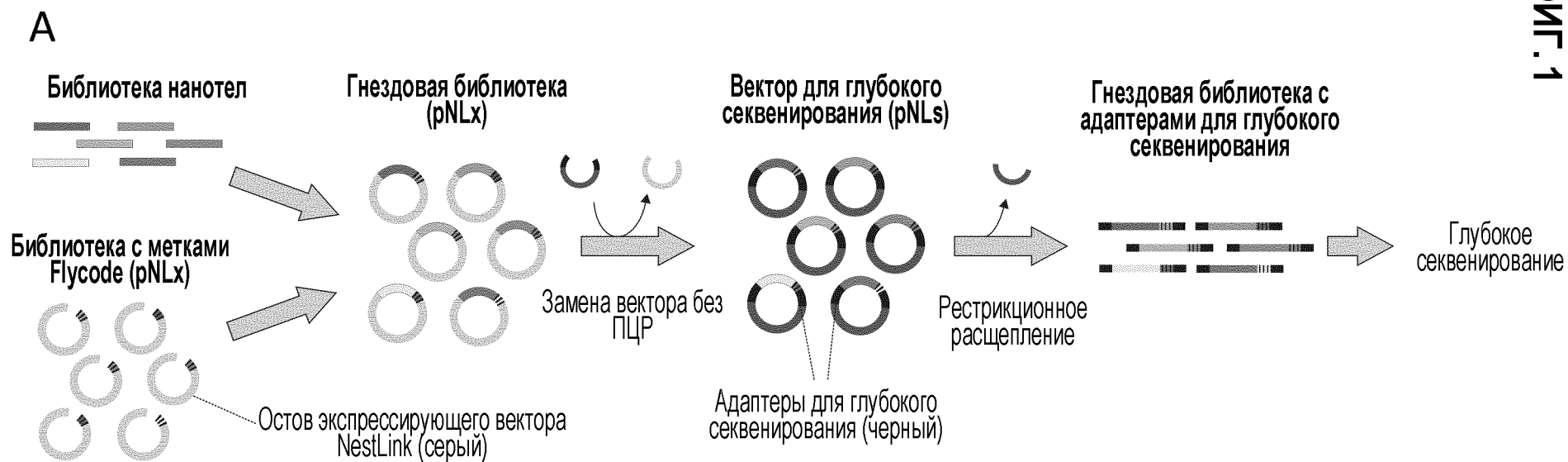
ii. указанная вторая библиотека нуклеиновых кислот имеет размер  $10^3$ - $10^{11}$ , в частности,  $10^5$ - $10^{10}$ , более конкретно,  $10^6$ - $10^9$ , даже более конкретно, приблизительно  $10^8$ ,

таким образом, получения плазмид с множеством комбинаций полипептид/метка;

d. отбора подгруппы из указанного множества плазмид с комбинациями полипептид/метка, таким образом, получения библиотеки меченых нуклеиновых кислот, кодирующей библиотеку меченых полипептидов.

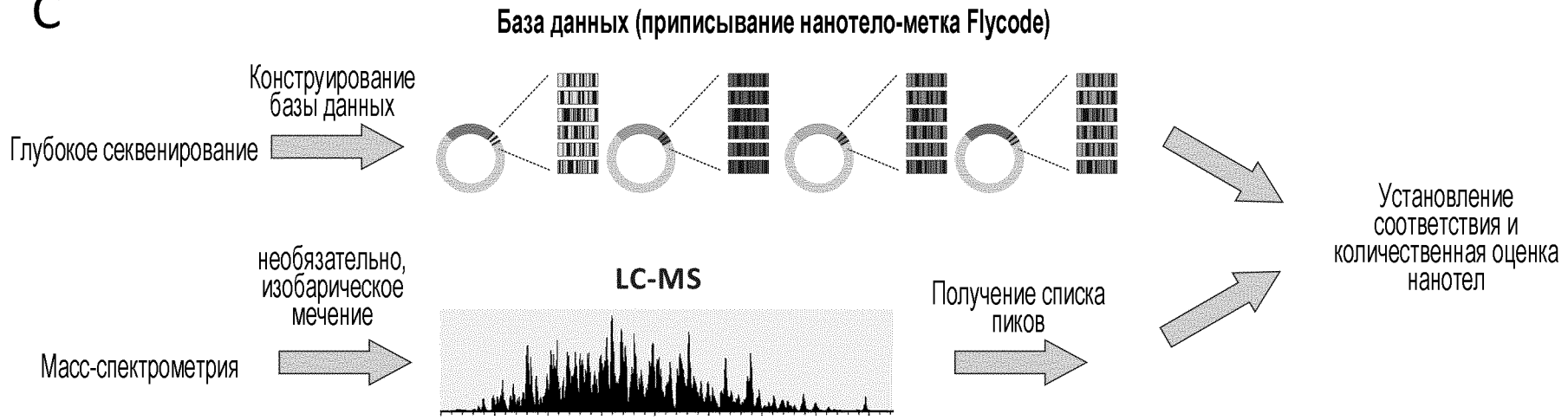
41. Способ по п.40, где указанная подгруппа из указанного множества плазмид с комбинациями полипептид/метка составляет по меньшей мере 3x, в частности, по меньшей мере 5x, более конкретно, по меньшей мере 15x, даже более конкретно, по меньшей мере 25x количество членов указанной первой библиотеки нуклеиновых кислот.

42. Способ по любому из пп. 40 или 41, где указанная подгруппа из указанного множества плазмид с комбинациями полипептид/метка составляет менее 50%, в частности, менее 5%, более конкретно, менее 0,5%, даже более конкретно, менее 0,05% от количества членов указанной второй библиотеки нуклеиновых кислот.





С

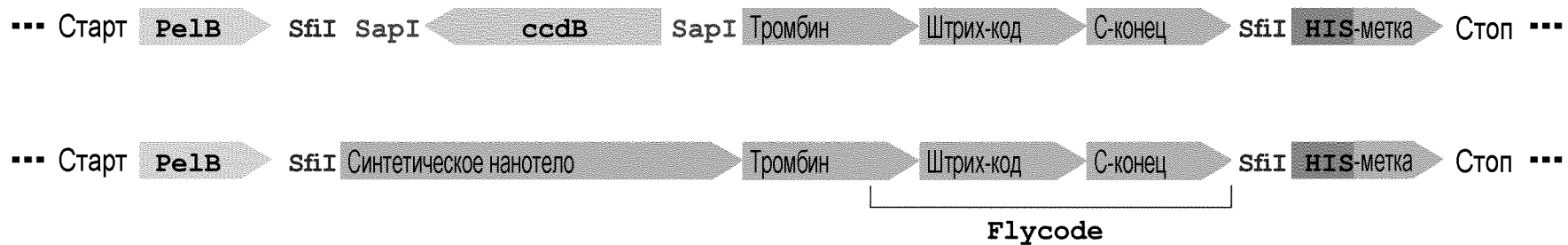


ФИГ. 1 (продолжение)

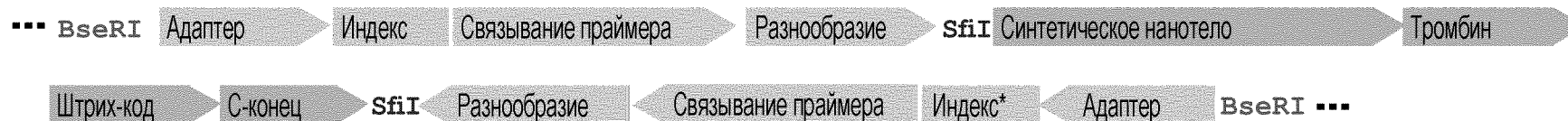
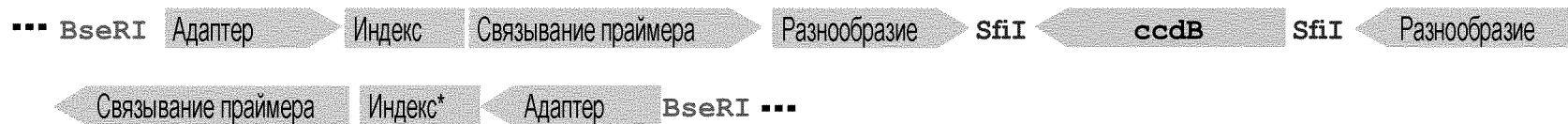
**A** (фагмида)



**B** (Библиотека с метками Flycode и экспрессирующий вектор, pNLx)



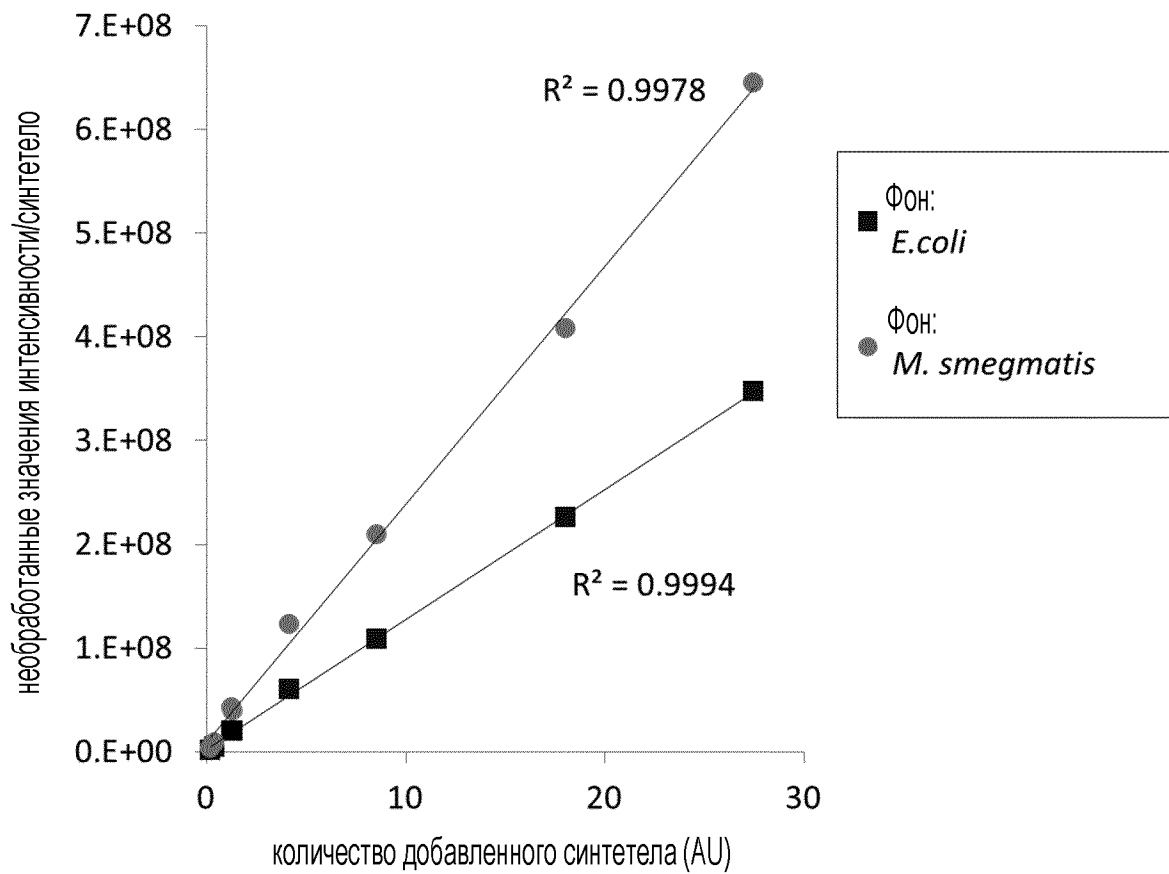
C (Вектор для глубокого секвенирования, pNLs)



D (двухцепочечные адаптерно-олигонуклеотиды для глубокого секвенирования)



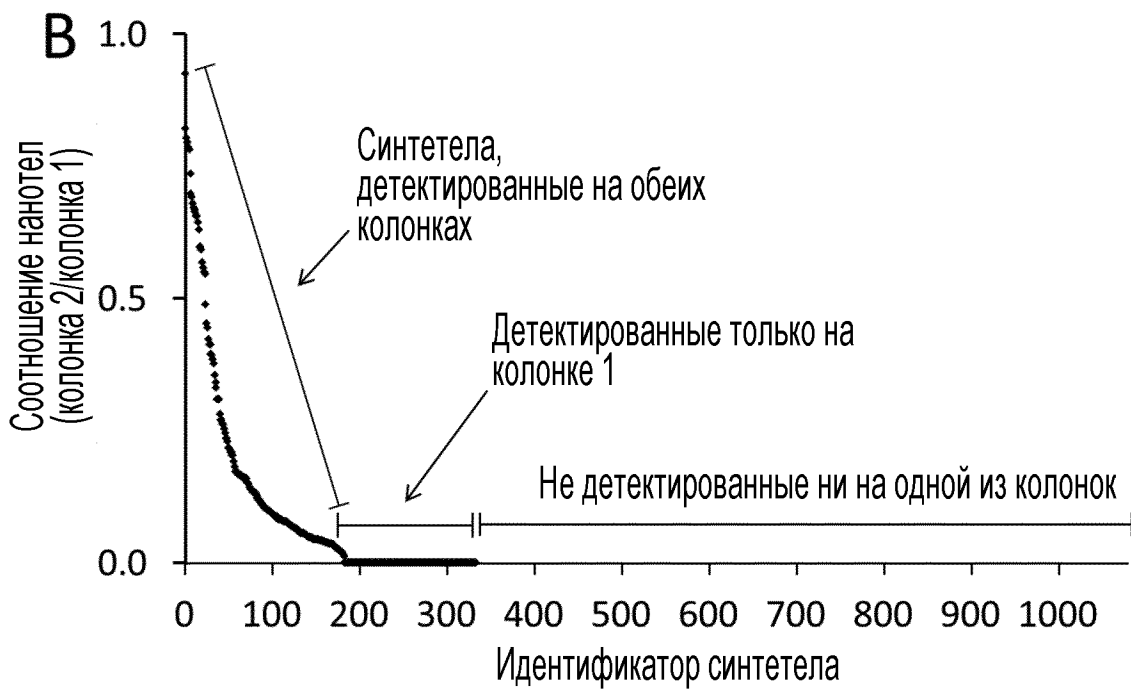
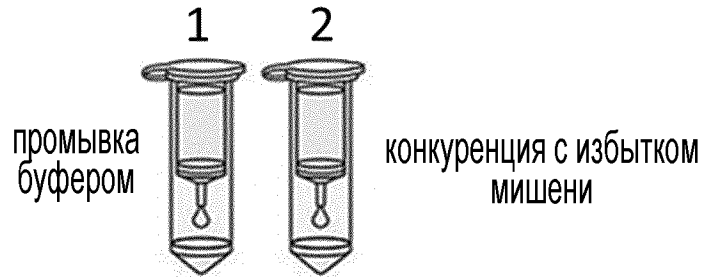
ФИГ.3



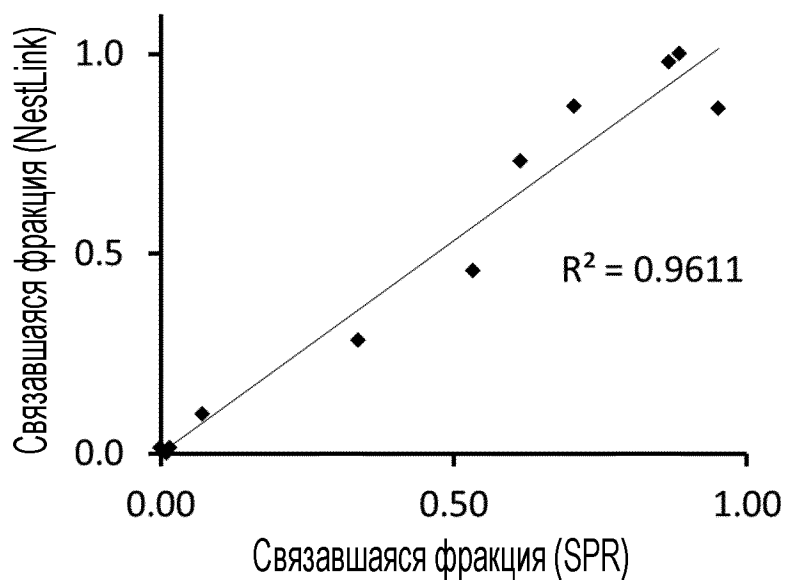
ФИГ. 4

А

Иммобилизация членов пула в комплексе с биотинилированной мишенью на двух колонках, с последующим отбором по скорости диссоциации



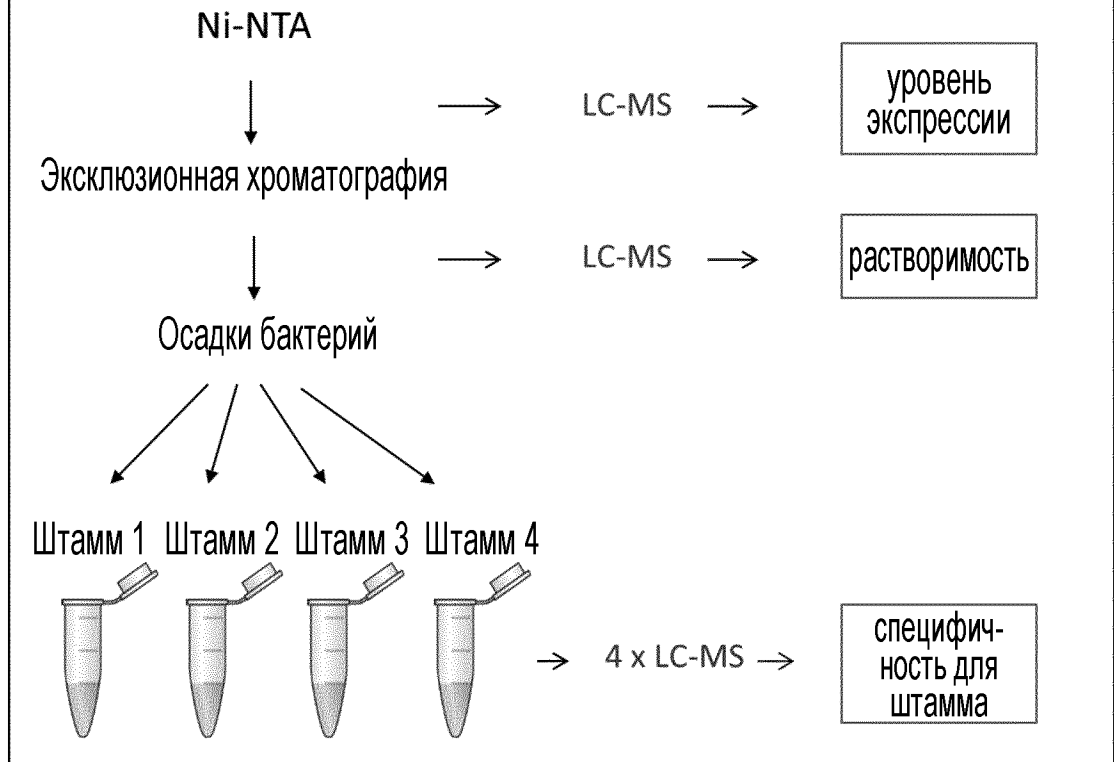
С



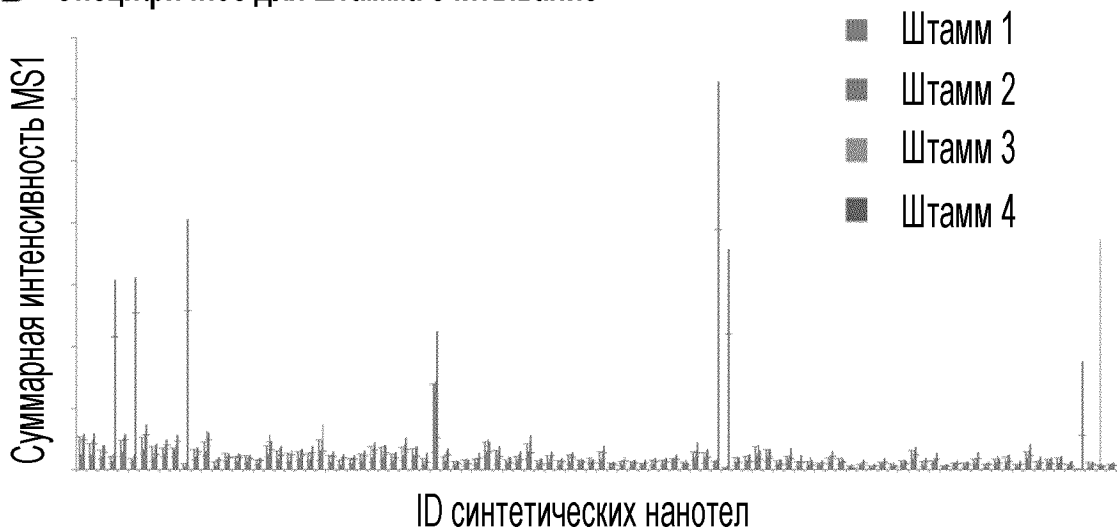
ФИГ. 5

**А** 1. Дисплей *in vitro* (Получение пула кандидатов на связывание)

**2. Nestlink** (Идентификация членов, специфически связывающихся с поверхностью клеток)



**В** Специфичное для штамма считывание



## ФИГ. 6

