

**(12) FASCÍCULO DE PATENTE DE INVENÇÃO**

(22) Data de pedido: <b>2011.04.05</b>	(73) Titular(es): <b>PROGNOSYS BIOSCIENCES, INC.</b> <b>505 COAST BLVD. SOUTH LA JOLLA, CA 92037</b> <b>US</b>
(30) Prioridade(s): <b>2010.04.05 US 321124 P</b> <b>2011.04.05 US 80616</b>	
(43) Data de publicação do pedido: <b>2013.02.13</b>	(72) Inventor(es): <b>MARK S. CHEE</b> <b>US</b>
(45) Data e BPI da concessão: <b>2015.09.02</b> <b>248/2015</b>	(74) Mandatário: <b>JOÃO LUÍS PEREIRA GARCIA</b> <b>RUA CASTILHO, 167 2º 1070-050 LISBOA</b> <b>PT</b>

(54) Epígrafe: **ENSAIOS BIOLÓGICOS CODIFICADOS ESPACIALMENTE**

(57) Resumo:

A PRESENTE INVENÇÃO PROPORCIONA ENSAIOS E SISTEMAS DE ENSAIO PARA UTILIZAÇÃO EM ENSAIOS BIOLÓGICOS CODIFICADOS ESPACIALMENTE. A INVENÇÃO PROPORCIONA UM SISTEMA DE ENSAIO QUE COMPREENDE UM ENSAIO CAPAZ DE ELEVADOS NÍVEIS DE MULTIPLEXAÇÃO ONDE REAGENTES SÃO PROPORCIONADOS A UMA AMOSTRA BIOLÓGICA EM PADRÕES ESPACIAIS DEFINIDOS; INSTRUMENTOS CAPAZES DE ADMINISTRAÇÃO CONTROLADA DE REAGENTES DE ACORDO COM OS PADRÕES ESPACIAIS E UM ESQUEMA DE DESCODIFICAÇÃO QUE PROPORCIONA UMA LEITURA QUE É DE NATUREZA DIGITAL.

## **RESUMO**

### **ENSAIOS BIOLÓGICOS CODIFICADOS ESPACIALMENTE**

A presente invenção proporciona ensaios e sistemas de ensaio para utilização em ensaios biológicos codificados espacialmente. A invenção proporciona um sistema de ensaio que compreende um ensaio capaz de elevados níveis de multiplexação onde reagentes são proporcionados a uma amostra biológica em padrões espaciais definidos; instrumentos capazes de administração controlada de reagentes de acordo com os padrões espaciais e um esquema de descodificação que proporciona uma leitura que é de natureza digital.

## DESCRIÇÃO

### ENSAIOS BIOLÓGICOS CODIFICADOS ESPACIALMENTE

#### CAMPO DA INVENÇÃO

Esta invenção refere-se a ensaios de moléculas biológicas, e mais particularmente a ensaios para determinar as distribuições espaciais de um grande número de moléculas biológicas numa amostra sólida simultaneamente.

#### ANTECEDENTES DA INVENÇÃO

Na seguinte discussão determinados artigos e métodos serão descritos para fins de antecedentes e de apresentação. Nada contido no presente documento será considerado como uma "admissão" de técnicas anteriores. O requerente reserva expressamente o direito a demonstrar, se adequado, que os artigos e métodos referenciados no presente documento não constituem técnicas anteriores sob as disposições legais aplicáveis.

A análise da expressão génica compreensiva e a análise proteica têm sido ferramentas úteis na compreensão dos mecanismos da biologia. A utilização destas ferramentas permitiu a identificação de genes e proteínas envolvidas no desenvolvimento e em várias doenças tais como o cancro e doença autoimune. Métodos convencionais tais como hibridação *in situ* e outra deteção multiplexada de diferentes transcritos revelaram padrões espaciais de expressão génica e auxiliaram a esclarecer a base molecular do desenvolvimento e da doença. Outras tecnologias que permitiram a análise quantitativa de várias sequências de ARN por amostra incluem microarranjo (veja-se Shi, *et al.*, Nature Biotechnology, 24(9): 1151-61 (2006); e Slonim e Yanai, Plos Computational Biology, 5(10):e1000543 (2009)); análise seriada da expressão génica (SAGE) (veja-se Velculescu, *et al.*, Science, 270(5235):484-87 (1995)), implementação de alto rendimento de qPCR (veja-se Spurgeon, *et al.*, Plos ONE, 3(2):e1662 (2008), e PCT *in situ*

(veja-se Nuovo, *Genome Res.*, 4:151-67 (1995)). Embora estes métodos sejam úteis, no entanto, não permitem medição simultânea da expressão de vários genes ou a presença e/ou atividade de várias proteínas em várias localizações espaciais numa amostra. A microdissecção por captura de laser permitiu a análise de vários genes num pequeno número de localizações, mas é muito extensiva, trabalhosa, e não é bem representada à escala. Determinados ensaios de PCR num formato 2D conservam informação espacial (veja-se Armani, *et al.*, *Lab on a Chip*, 9(24): 3526-34 (2009)), mas estes métodos têm baixa resolução espacial já que se apoiam na transferência física de tecido a poços, o que também previne o acesso aleatório às amostras de tecido e níveis elevados de multiplexação.

Atualmente, não existe qualquer método prático para analisar a resolução elevada aos padrões de expressão espacial de um grande número de genes, proteínas, ou outras moléculas biologicamente ativas simultaneamente. Existe consequentemente uma necessidade de mapas espaciais reproduzíveis, de alta resolução de moléculas biológicas em tecidos. A presente invenção aborda esta necessidade.

O documento WO2010/019826 descreve a deteção e a quantificação de moléculas alvo individuais em amostras biomoleculares utilizando sondas nanorepórter estáveis com sinal diferenciadamente detetável.

MIR KALIM *et al.*, *NUCLEIC ACIDS RESEARCH*, OXFORD UNIVERSITY PRESS, GB, vol. 37, nº 1, 1 de janeiro de 2009 (201-01-2009), páginas e5-1, descreve Sequenciação por Ligação e Clivagem Cíclicas (CycLiC) diretamente num molde capturado em microarranjo.

## **SUMÁRIO DA INVENÇÃO**

Este Sumário é proporcionado para apresentar uma seleção de conceitos numa forma simplificada que são adicionalmente descritos abaixo na Descrição Detalhada. Este Sumário não pretende identificar características chave ou essenciais do assunto reivindicado, nem pretende ser utilizado para limitar

o âmbito do assunto reivindicado. Outras características, detalhes; utilidades, e vantagens do assunto reivindicado serão aparentes a partir da seguinte Descrição Detalhada escrita incluindo aqueles aspetos ilustrados nos desenhos acompanhantes e definidos na reivindicação anexa.

Esta divulgação refere-se a sistemas de ensaio que proporcionam maias espaciais de alta resolução de atividade biológica em tecidos. O sistema de ensaio compreende um ensaio capaz de níveis elevados de multiplexação onde são proporcionadas sondas codificadas a uma amostra biológica em padrões espaciais definidos; instrumentos capazes de administração controlada de reagente de acordo com os padrões espaciais; e um esquema de decodificação proporcionando uma leitura de natureza digital. Em resumo, a presente invenção proporciona a capacidade de visualizar vários alvos biológicos em várias localizações, proporcionando a resolução da hibridação *in situ* com a análise de dados altamente paralela da sequenciação.

Consequentemente, a invenção proporciona um método de acordo com a reivindicação 1 e um método de acordo com a reivindicação 2.

Em aspetos particulares da invenção os alvos biológicos compreendem ácidos nucleicos e as sondas codificadas são oligonucleótidos, e nalguns aspetos, existem duas sondas codificadas para cada um dos múltiplos alvos de ácidos nucleicos. Em alguns aspetos, os múltiplos alvos biológicos compreendem proteínas, as regiões de sonda das sondas codificadas são proteínas e as etiquetas de codificação compreendem oligonucleótidos. Nalguns aspetos os múltiplos alvos biológicos compreendem enzimas. Nalguns aspetos as regiões de sonda das sondas codificadas compreendem anticorpos, aptâmeros ou moléculas pequenas.

Alguns aspetos do sistema de ensaio compreendem adicionalmente uma etapa de amplificação entre a etapa de separação e a etapa de determinação. Em alguns aspetos, a etapa

de determinação é realizada através de sequenciação de ácidos nucleicos, e em aspetos preferidos, a sequenciação é sequenciação de ácidos nucleicos digital de alto rendimento.

Nalguns aspetos da invenção; o produto dos múltiplos alvos biológicos sendo submetido a ensaio e os múltiplos sítios na amostra é maior do que 20, o produto dos múltiplos alvos biológicos sendo submetidos a ensaio e dos múltiplos sítios na amostra é maior do que 20, do que 50, nalguns aspetos o produto dos múltiplos alvos biológicos sendo submetidos a ensaio e os múltiplos sítios na amostra é maior do que 75, 100, 150, 500, 750, 1.000, 5.000, 10.000, 25.000, 50.000, 100.000, 500.000, ou 1.000.000 ou mais. Noutros aspetos, a sequência de pelo menos cinquenta mil sondas de codificação é determinada em paralelo, noutros aspetos as sequências de pelo menos cem mil sondas de codificação é determinada em paralelo, nalguns aspetos a sequência de pelo menos quinhentas mil sondas de codificação é determinada em paralelo, e nalguns aspetos a sequência de pelo menos um milhão, dez milhões, cem milhões, mil milhões, dez mil milhões, cem mil milhões ou mais sondas de codificação é determinada em paralelo.

Nalguns aspetos, o padrão espacial conhecido é determinado através de características histológicas da amostra. Também nalguns aspetos, *hardware* programado com *software* realiza pelo menos duas etapas da etapa de administração, a etapa de separação, a etapa de determinação e a etapa de associação.

Nalguns aspetos, as regiões de sonda das sondas codificadas são proteínas e a etapa de separação é alcançada através de sondas codificadas que interagem com os alvos biológicos sendo capturados por parte de um agente de captura por afinidade. Nalguns aspetos as regiões de sonda das sondas codificadas são ácidos nucleicos e a etapa de separação é alcançada através de uma lavagem da amostra.

A divulgação também se refere a um sistema de ensaio para determinar padrões espaciais de abundância ou atividade ou

ambos de múltiplos alvos proteicos em múltiplos sítios numa amostra, onde o sistema de ensaio realiza as seguintes etapas: proporcionar uma amostra fixada a um suporte, administrar sondas codificadas para os múltiplos alvos proteicos aos múltiplos sítios na amostra numa padrão espacial conhecido, onde cada sonda codificada compreende uma região de sonda proteica que pode interagir com os alvos proteicos e uma etiqueta de codificação que identifica uma localização do sítio ao qual a sonda codificada foi administrada e a região de sonda proteica da sonda codificante da qual é parte a etiqueta de codificação; permitir que as sondas codificadas interajam com os alvos proteicos; separar as sondas codificadas que interagem com os alvos proteicos das sondas codificadas que não interagem com os alvos proteicos; determinar a totalidade ou uma porção de uma sequência das sondas codificadas através de sequenciação de alto rendimento, e associar a abundância ou atividade ou ambas dos múltiplos alvos proteicos às localizações dos múltiplos sítios na amostra.

A divulgação também se refere a um sistema de ensaio para determinar padrões espaciais de abundância ou atividade ou ambos de múltiplos alvos biológicos em múltiplos sítios numa amostra, onde o sistema de ensaio realiza as seguintes etapas: proporcionar uma amostra fixada a um suporte, administrar sondas codificadas para os múltiplos alvos biológicos aos múltiplos sítios na amostra numa padrão espacial conhecido, onde cada sonda codificada compreende uma região de sonda proteica que pode interagir com os alvos biológicos e uma etiqueta de codificação que identifica uma localização do sítio ao qual a sonda codificada foi administrada e a identifica o alvo biológico; permitir que as sondas codificadas interajam com os alvos biológicos; determinar a totalidade ou uma porção de uma sequência das sondas codificadas, e associar a abundância ou atividade ou ambas dos múltiplos alvos biológicos às localizações dos sítios na amostra.

A invenção pode utilizar vários mecanismos de detecção, com

base nas moléculas a ser detetadas e os reagentes necessários para tal sistema de deteção. Métodos exemplares que podem ser utilizados com a invenção são descritos mais detalhadamente abaixo.

### **DESCRIÇÃO DAS FIGURAS**

A Figura 1 proporciona uma vista geral simplificada do sistema de ensaio utilizando o mesmo.

A Figura 2 proporciona uma vista geral simplificada de uma forma de realização dos sistemas de ensaio utilizando a presente invenção para detetar ácidos nucleicos.

A Figura 3 é uma ilustração representacional de uma forma de realização do ensaio visto geralmente na Figura 2. A Figura 4 ilustra um mecanismo geral para uma forma de realização de um esquema codificante combinatório dos sistemas de ensaio utilizando a invenção.

A Figura 5 proporciona um exemplo específico simplificado da forma de realização de um esquema codificado combinatório mostrado na Figura 4.

### **DEFINIÇÕES**

Os termos utilizados no presente documento pretendem ter o significado simples e vulgar conforme entendido por parte dos peritos na especialidade. As seguintes definições pretendem auxiliar o leitor na compreensão da presente invenção, mas não pretendem varias ou de outra forma limitar o significado de tais termos a menos que especificamente indicado.

O termo "anticorpo" conforme utilizado no presente documento pretende referir uma imunoglobulina ou anticorpo inteiro ou qualquer fragmento funcional de uma molécula de imunoglobulina que é capaz de ligar especificamente a um antigénio (anticorpos e antigénios são "parceiros de ligação" conforme definido no presente documento). "Anticorpo" conforme utilizado no presente documento pretende incluir o anticorpos inteiro bem como quaisquer fragmentos de anticorpo capazes de ligar ao antigénio ou fragmento antigénico de interesse. Exemplos de tais péptidos incluem moléculas de anticorpo



completas, fragmentos de anticorpos, tal como Fab, F(ab')<sub>2</sub>, CDRS, VL, VH, e qualquer outra porção de um anticorpo que é capaz de ligar especificamente a um antigénio. Os anticorpos para ensaios da invenção são imunorreativos ou imunoespecíficos para, e como tal ligam especificamente e seletivamente a, proteínas ou detetadas (isto é, alvos biológicos) ou utilizadas para deteção (isto é, sondas) nos ensaios da invenção.

O termo "agente de ligação" conforme utilizado no presente documento refere-se a qualquer agente que liga especificamente a uma molécula biológica de interesse.

"Complementar" ou "substancialmente complementar" refere-se à hibridação ou emparelhamento de bases ou à formação de um duplex entre nucleótidos ou ácidos nucleicos, tais como, por exemplo, entre as duas cadeias de uma molécula de ADN de cadeia dupla ou entre um iniciador oligonucleotídico e um sítio de ligação de iniciador num ácido nucleico de cadeia simples. Nucleótidos complementares são, geralmente, A e T (ou A e U), ou C e G. É dito que duas moléculas de ARN ou ADN de cadeia simples são substancialmente complementares quando os nucleótidos de uma cadeia, alinhados e comparados de forma ótima e com inserções ou eliminações nucleotídicas adequadas, se emparelham com pelo menos cerca de 80% da outra cadeia, habitualmente pelo menos cerca de 90% até cerca de 95%, e inclusive cerca de 98% até cerca de 100%.

"Hibridação" refere-se ao processo no qual dois polinucleótidos de cadeia simples ligam não covalentemente para formar um polinucleótido de cadeia dupla estável. O polinucleótido (habitualmente) de cadeia dupla resultante é um "híbrido" ou "duplex". "Condições de hibridação" incluirão tipicamente concentrações de sal de aproximadamente menos de 1M, frequentemente menos de cerca de 500 mM e podem ser menos de cerca de 200 mM. Um "tampão de hibridação" é uma solução de sal de tampão tal como SSPE a 5%, ou outros tais tampões conhecidos na técnica. As temperaturas de hibridação podem ser tão baixas como 5 °C, mas são tipicamente maiores do que 22 °C,

e mais tipicamente maiores do que cerca de 30 °C, e tipicamente superiores a 37 °C. As hibridações são frequentemente realizadas sob condições restritivas, isto é, condições sob as quais um iniciador hibridará com a sua sequência alvo mas não hibridará com as outras sequências não complementares. As condições restritivas são dependentes de sequência e são diferentes em diferentes circunstâncias. Por exemplo, fragmentos mais longos podem requerer temperaturas de hibridação mais elevadas para hibridação específica do que os fragmentos curtos. Já que outros fatores podem afetar a restringência da hibridação, incluindo composição base e comprimento das cadeias complementares, presença de solventes orgânicos, e a medida da não correspondência de bases, a combinação de parâmetros é mais importante que a medida absoluta de qualquer parâmetro por si só. São geralmente selecionadas condições restritivas para ser cerca de 5 °C menores do que a  $T_m$  para a sequência específica a uma força iônica e pH definidos. Condições restritivas exemplares incluem uma concentração de sal de pelo menos 0,01 M até não mais do que 1M de concentração de íão sódio (ou outro sal) a um pH de cerca de 7,0 até cerca de 8,3 e uma temperatura de pelo menos 25 °C. Por exemplo, as condições de 5xSSPE (NaCl a 750 mM, fosfato de sódio a 50 mM, EDTA a 5 mM a pH 7,4) e uma temperatura de aproximadamente 30 °C são adequadas para hibridações específicas de alelo, embora uma temperatura adequada dependa do comprimento e/ou teor GC da região hibridada.

"Ligação" significa formar uma ligação covalente ou ligação entre os terminais de dois ou mais ácidos nucleicos, por exemplo, oligonucleótidos e/ou polinucleótidos, numa reação conduzida por molde. A natureza da ligação pode variar amplamente o a ligação pode ser levada a cabo enzimaticamente ou quimicamente. Conforme utilizado no presente documento, as ligações são habitualmente levadas a cabo enzimaticamente para formar uma ligação fosfodiéster entre um nucleótido de carbono

terminal 5' de um oligonucleótido com um carbono 3' de outro nucleótido.

"Ácido nucleico", "oligonucleótido", "oligo" ou equivalentes gramaticais utilizados no presente documento referem-se geralmente a pelo menos dois nucleótidos ligados covalentemente. Um ácido nucleico conterá geralmente ligações fosfodiéster, embora nalguns casos possam ser incluídos análogos de ácidos nucleicos que tenham estruturas principais alternativas tais como ligações fosforamidite, fosforoditioato, ou metilfosforoamidite; ou estruturas principais e ligações de ácidos nucleicos peptídicos. Outros ácidos nucleicos análogos incluem aqueles com estruturas bicíclicas incluindo ácidos nucleicos bloqueados, estruturas principais positivas, estruturas principais não iônicas e estruturas principais não de ribose. Podem ser realizadas modificações da estrutura principal ribose-fosfato para aumentar a estabilidade das moléculas; por exemplo, híbridos ANP:ADN podem exibir estabilidade mais elevada nalguns ambientes.

"Iniciador" significa um oligonucleótido, quer natural quer sintético, que é capaz, após formar um duplex com um molde polinucleotídico, de atuar como um ponto de iniciação da síntese de ácidos nucleicos e ser estendido a partir do seu terminal 3' ao longo do molde de forma a ser formado um duplex estendido. A sequência de nucleótidos adicionada durante o processo de extensão é determinado através da sequência do polinucleótido molde. Os iniciadores são habitualmente estendidos por parte de uma ADN polimerase.

O termo "SNP" ou "polimorfismo de nucleótido único" refere-se a uma variação genética entre indivíduos; por exemplo, uma posição de base azotada única no ADN de organismos que é variável. Os SNPs são encontrados ao longo do genoma; muita da variação genética entre indivíduos é devida à variação em *loci* de SNP, e frequentemente esta variação genética resulta em variação fenotípica entre indivíduos. SNPs para utilização

na presente invenção e os seus respetivos alelos podem ser derivados a partir de qualquer número de fontes, tais como bases de dados públicas (U.C. Santa Cruz Human Genome Browser Gateway (<http://genome.ucsc.edu/cgi-bin/hgGateway>) ou a página web NCBI dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>), ou podem ser determinados experimentalmente conforme descrito na Patente U.S. N.º 6.969.589; e Publicação U.S. N.º 2006/0188875 intitulada "Human Genomic Polymorphisms." Embora a utilização de SNPs seja descrita nalgumas das formas de realização apresentadas no presente documento, será entendido que podem ser utilizados outros marcadores genéticos bialélicos ou multialélicos. Um marcador genético bialélico é aquele que tem duas formas polimórficas, ou alelos. Conforme mencionado acima, para um marcador genético bialélico que é associado com uma característica, o alelo que é mais abundante na composição genética de um grupo de casos conforme comparado com um grupo de controlo é denominado o "alelo associado", e o outro alelo pode ser referido como o "alelo não associado". Assim, para cada polimorfismo bialélico que é associado com uma característica determinada (por exemplo, uma doença ou resposta a fármaco), existe um alelo associado correspondente. Outros polimorfismos bialélicos que podem ser utilizados com os métodos apresentados no presente documento incluem, mas não estão limitados a alterações, inserções, deleções, e translocações multinucleotídicas. Será adicionalmente apreciado que as referências a ADN no presente documento podem incluir ADN genómico, ADN mitocondrial, ADN epissómico, e/ou derivados de ADN tais com amplicões, transcritos de ARN, ADNc, análogos de ADN, etc. Os loci polimórficos que são rastreados num estudo de associação podem encontrar-se num estado haploide ou diploide e, idealmente, seriam a partir de sítios ao longo do genoma.

O termo "liga seletivamente", "ligação seletiva" e semelhantes conforme utilizados no presente documento, ao referir um parceiro de ligação (por exemplo, proteína, ácido

nucleico, anticorpo ou outro agente de captura por afinidade, etc.), refere-se a uma reação de ligação de dois ou mais parceiros de ligação com afinidade elevada e/ou complementaridade para garantir hibridação seletiva sob condições de ensaio designadas. Tipicamente, a ligação específica será pelo menos três vezes o desvio padrão do sinal de fundo. Assim, sob condições designadas o parceiro de ligação liga à sua molécula "alvo" particular e não liga em quantidade significativa a outras moléculas presente na amostra.

"Sequenciação", "determinação de sequência" e semelhantes significa determinação de informação relacionada com a sequência de bases nucleotídicas de um ácido nucleico. Tal informação pode incluir a identificação ou determinação de informação de sequências parciais bem como inteiras do ácido nucleico. A informação de sequências pode ser determinada com graus variantes de fiabilidade ou confiança estatística. Num aspeto, o termo inclui a determinação da identidade e ordenação de uma série de nucleótidos contíguos num ácido nucleico. "Sequenciação digital de alto rendimento" ou "sequenciação de nova geração" significa determinação de sequências utilizando métodos que determinam muitos (tipicamente milhares até milhares de milhões) de sequências de ácidos nucleicos de uma forma intrinsecamente paralela, isto é, onde são preparados moldes de ADN para sequenciação não um de cada vez, mas num processo em bruto, e onde muitas sequências são lidas preferentemente em paralelo, ou alternativamente utilizando um processo em série de rendimento ultra alto que em si mesmo pode ser paralelizado. Tais métodos incluem mas não estão limitados a pirosequenciação (por exemplo, conforme comercializado por 454 Life Sciences, Inc., Branford, CT); sequenciação através de ligação (por exemplo, conforme comercializado na tecnologia SOLiD™, Life Technology, Inc., Carlsbad, CA); sequenciação através de síntese utilizando nucleótidos modificados (tal como comercializado na tecnologia TruSeq™ e HiSeq™ por Illumina, Inc., San Diego, CA, HeliScope™ por Helicos

Biosciences Corporation, Cambridge, MA, e PacBio RS por Pacific Biosciences of California, Inc., Menlo Park, CA), sequenciação através de detecção de iões (Ion Torrent, Inc., South San Francisco, CA); sequenciação de nanobolas de ADN (Complete Genomics, Inc., Mountain View, CA); tecnologias de sequenciação à base de nanoporos (por exemplo, conforme desenvolvido por Oxford Nanopore Technologies, LTD, Oxford, RU), e métodos de sequenciação altamente paralelizada semelhantes.

O termo " $T_m$ " é utilizado em referência à "temperatura de fusão". A temperatura de fusão é a temperatura à qual uma população de moléculas de ácido nucleico de cadeia dupla se dissocia pela metade em cadeias simples. São bem conhecidas na técnica várias equações para calcular a  $T_m$  de ácidos nucleicos. Conforme indicado através de referências padrão, uma estimativa simples do valor da  $T_m$  pode ser calculada através da equação,  $T_m = 81,5 + 0,41 (\% G+C)$ , quando um ácido nucleico se encontra em solução aquosa a NaCl a 1M (veja-se por exemplo, Anderson e Young, Quantitative Filter Hybridization, em Nucleic Acid Hybridization (1985)). Outras referências, (por exemplo, Allawi e SantaLucia, Jr., Biochemistry, 36:10581-94 (1997)) incluem métodos alternativos de imputação que têm em conta características estruturais e ambientais, bem como de sequência para o cálculo da  $T_m$ .

#### **DESCRIÇÃO DETALHADA DA INVENÇÃO**

A prática das técnicas descritas no presente documento pode empregar, a menos que indicado de outro modo, técnicas e descrições convencionais de química orgânica, tecnologia de polímeros, biologia molecular (incluindo técnicas recombinantes), biologia celular, bioquímica, e tecnologia de sequenciação, que se encontram dentro da perícia dos peritos na especialidade. Tais técnicas convencionais incluem síntese de matrizes de polímeros hibridação e ligação de polinucleótidos, e detecção de hibridação utilizando uma etiqueta. Ilustrações específicas de técnicas adequadas podem

ser tidas como referência aos exemplos no presente documento. Contudo, outros procedimentos convencionais equivalentes podem, naturalmente, ser também utilizados. Tais técnicas e descrições convencionais podem ser encontradas em manuais de laboratório padrão tais como Green, *et al.*, Eds., *Genome Analysis: A Laboratory manual Series* (Vols. I-IV) (1999); Weiner, Gabriel, Stephens, Eds., *Generic Variation: A Laboratory Manual* (2007); Dieffenbach, Dveksler, Eds., *PGR Primer: A Laboratory Manual* (2003); Bowtell e Sambrook, *DNA Microarrays: A Molecular Cloning Manual* (2003); Mount, *Bioinformatics: Sequence and Genome Analysis* (2004); Sambrook e Russell, *Condensed Protocols from Molecular Cloning: A Laboratory Manual* (2006); e Sambrook e Russell, *Molecular Cloning: Laboratory Manual* (2002) (todos da Cold Spring Harbour Laboratory Press); Stryer, *Biochemistry* (4th Ed.) (1995) W.H. Freeman; Nova Iorque N.I.; Gait, *"Oligonucleotide Synthesis: A Practical Approach"* (2002) IRL Press, Londres; Nelson e Cox, Lehninger. *Principles of Biochemistry* (2000) 3rd Ed., W. H. Freeman Pub., Nova Iorque, N.I.; e Berg, *et al.*, *Biochemistry* (2002) 5th Ed., W.H. Freeman Pub., Nova Iorque, N.Y.

Note-se que conforme utilizado no presente documento e nas reivindicações anexas, as formas singulares "uma/um" e "o/a" incluem os respectivos plurais a não ser que o contexto claramente dite o contrário. Consequentemente, por exemplo, referência a um "ácido nucleico" refere-se a um ou mais ácidos nucleicos, e referência a "o ensaio" inclui referência a etapas e métodos equivalentes conhecidos por parte dos peritos na especialidade, e assim por diante.

A menos que definido de outra forma, todos os termos técnicos e científicos utilizados no presente documento possuem o mesmo significado como habitualmente entendido por um perito na especialidade à qual esta invenção pertence.

Onde é proporcionado um intervalo de valores, é entendido que cada valor interveniente, entre o limite superior e inferior desse intervalo e qualquer outro valor estabelecido

ou interveniente nesse intervalo estabelecido é abrangido dentro da invenção. Os limites superior e inferior destes intervalos mais pequenos podem, independentemente, ser incluídos nos limites mais pequenos, e estão também englobados na presente invenção, sujeitos a qualquer limite especificamente excluído no intervalo indicado. Quando o intervalo indicado inclui um ou ambos os limites, intervalos excluindo qualquer dos limites incluídos estão também incluídos na invenção.

Na seguinte descrição, são estabelecidos vários detalhes específicos para proporcionar um entendimento mais minucioso da presente invenção. Contudo, será aparente para um perito na especialidade que a presente invenção pode ser posta em prática sem um ou mais destes detalhes específicos. Noutros casos, não foram descritas características e procedimentos bem conhecidos por parte dos peritos na especialidade de forma a evitar tornar a invenção menos clara.

**A Invenção Em Geral** (A invenção encontra-se definida nas reivindicações).

A invenção refere-se a ensaios multiplexados espacialmente codificados compreendendo 1) um ensaio capaz de níveis elevados de multiplexação com um esquema de codificação espacial eficaz; 2) instrumentos capazes de administrar reagentes de acordo com um padrão espacial; e 3) descodificação determinada através de uma leitura de natureza digital. Os sistemas de ensaio detetam a presença ou ausência e quantidade relativa de um alvo biológico ou atividade biológica indicativos de um alvo biológico, bem como a localização do alvo biológico ou atividade biológica numa amostra, por exemplo, uma secção de tecido ou outra estrutura biológica disposta sobre um suporte tal como uma lâmina de microscópio ou placa de cultura.

O sistema de ensaio proporciona adicionalmente instrumentos com uma capacidade de administrar reagentes num padrão definido espacialmente. Estes instrumentos, em conjunto



com *software*, reagentes e protocolos, proporciona um componente chave do sistema de ensaio altamente inovador, permitindo a medição de numerosos alvos ou atividades biológicas num ambiente espacial significativo, incluindo expressão génica e localização de péptidos. Um esquema de codificação utilizado nestes sistemas de ensaio permite que seja determinada a localização de alvos ou atividade biológicos (ou ausência dos mesmos nas amostras biológicas após os produtos do ensaio multiplexado serem removidos a partir da amostra biológica e agrupados para análise. A descodificação do esquema de codificação pode ser realizada através de, por exemplo, de alto rendimento, que proporciona facilmente milhões a milhões de milhões de pontos de dados a baixo custo. Os resultados dos ensaios tais como a quantidade ou atividade de alvos biológicos pode então ser mapeada para a localização específica na amostra biológica. Os sistemas de ensaio abrem uma nova janela analítica para os padrões espaciais complexos da função e regulação celular nas amostras biológicas.

É proporcionada uma vista geral simplificada do sistema de ensaio 100 utilizando a presente invenção na Figura 1. Na etapa 110, é proporcionada uma amostra biológica fixada a um suporte. A amostra biológica contém alvos biológicos de interesse. Os alvos biológicos podem incluir qualquer molécula de interesse, tais como ácidos nucleicos (incluindo, por exemplo, transcritos de ARN, sequências genómicas de ADN, ADNcs, amplicões, ou outras sequências de ácidos nucleicos) e proteínas, enzimas e semelhantes. Na etapa 120, sondas codificadas são administradas à amostra biológica de acordo com um padrão espacial conhecido. As sondas codificadas compreendem sondas, que podem interagir com alvos biológicos de interesse, e etiquetas de codificação, que identificam as posições na amostra dos alvos biológicos sendo ensaiados, e consequentemente podem ser utilizadas para ligar resultados de ensaios a localizações na amostra. As etiquetas de codificação na maioria das formas de realização são oligonucleótidos.

Contudo, as etiquetas de codificação podem também ser etiquetas de massa, marcadores fluorescentes, ou outras frações.

Nalgumas formas de realização, as porções de sonda e etiqueta de codificação da sonda codificada são pré-emparelhadas anteriormente a serem administradas à amostra biológica. Por exemplo, no caso onde as sondas codificadas são oligonucleótidos, tanto a sequência de sonda como a de etiqueta de codificação podem ser sintetizadas como um único oligonucleótido. Alternativamente, as porções de sonda e de etiqueta de codificação das sondas codificantes podem ser sintetizadas ou obtidas separadamente e combinadas anteriormente à administração à amostra biológica (por exemplo, podem ser sintetizados dois oligonucleótidos separados e emparelhados através de, por exemplo, ligação, ou um anticorpo e um oligonucleótido podem ser preparados separadamente e conjugados anteriormente à administração à amostra biológica). Também, conforme é descrito nas Figuras 2-5, as sondas e as etiquetas de codificação (em oligonucleótidos codificantes) são sintetizadas separadamente, e são administradas à amostra biológica em etapas diferentes (por exemplo, sondas primeiro e etiquetas de codificação posteriormente, ou vice-versa) no ensaio.

Na etapa 130, as sondas codificadas são permitidas reagir ou interagir com os alvos biológicos, isto é, são proporcionadas condições para permitir por exemplo, que os oligonucleótidos hibridem com alvos de ácidos nucleicos, que as enzimas catalisem reações com alvos proteicos, que anticorpos liguem a epítomos, etc. No caso onde os alvos biológicos são ácidos nucleicos, as sondas codificadas são tipicamente oligonucleótidos e hibridam com os ácidos nucleicos alvo. No caso de que os alvos biológicos são proteínas, as sondas codificadas são tipicamente aptâmeros, moléculas pequenas, ou proteínas conjugadas com oligonucleótidos que interagem com proteínas alvo através de ligação com as mesmas ou através de reação com as mesmas (isto

é, uma das proteínas é um substrato para a outra). Oligonucleótidos codificantes podem ser emparelhados com as sondas (proteínas) através de conjugação, foto-reticulação ou reticulação química por meio de grupo adequados e semelhantes.

Após as sondas codificadas interagirem com os alvos biológicos, as sondas codificadas que interagiram com os alvos biológicos devem ser separadas das sondas codificadas que não interagiram com os alvos biológicos da etapa 140. No caso onde os alvos biológicos são ácidos nucleicos e as sondas codificadas são oligonucleótidos, a separação pode ser alcançada através de, por exemplo, lavagem das sondas codificadas não hibridadas a partir da amostra. De forma semelhante, para outros ensaios que são baseados em ligação por afinidade, incluindo aqueles utilizando aptâmero, molécula pequena, e sondas proteicas, podem ser utilizadas etapas de lavagem para remover ligantes de baixa afinidade. No caso onde a sonda é transformada através de interação com o alvo, por exemplo, no caso de um péptido, por exemplo, através de clivagem por parte de uma protease ou fosforilação por parte de uma quinase, é conveniente recolher todas as sondas codificadas -- tanto sondas codificadas que interagiram com os alvos biológicos e foram transformadas como sondas codificadas que não foram transformadas. Após recolha ou agrupação, pode ser utilizado um anticorpo ou outro agente de captura por afinidade para capturar sondas que foram transformadas através da adição de um fração (por exemplo, um grupo fosfato). Em casos onde as sondas foram transformadas através de clivagem, as sondas transformadas podem ser separadas, por exemplo, através da captura das sondas não transformadas através de uma etiqueta que é removida a partir das sondas transformadas durante a transformação (por exemplo, através de clivagem), ou através da adição de um nova etiqueta no sítio de clivagem.

Após as sondas codificadas reagidas (transformadas) ou interagidas serem separadas a partir das sondas codificadas não reagidas ou não interagidas, a sequência das sondas codificadas

reagidas e/ou interagidas é determinada através de, preferentemente, sequenciação. A sequência das sondas codificadas permite o mapeamento dos resultados do ensaio para as localizações na amostra biológica.

A Figura 2 proporciona uma vista geral simplificada de um sistema de ensaio utilizando a presente invenção incorporando uma implementação eficaz de um esquema de codificação combinatório para a codificação de informação espacial. Para propósitos desta vista geral, as sondas são oligonucleótidos, mas conforme explicado noutro ponto, podem também ser utilizados outros tipos de sondas. Na etapa 210, é proporcionada uma amostra biológica fixada a um suporte, por exemplo, uma amostra de tecido ou outra estrutura biológica. Na etapa 220 uma ou mais sondas oligonucleotídicas são administradas à amostra biológica, onde as sondas oligonucleotídicas são capazes de hibridar com alvos biológicos na amostra biológica. Na etapa 230, as sondas oligonucleotídicas são permitidas interagir com (hibridar com) os alvos de ácidos nucleicos; ou seja, são proporcionadas condições adequadas onde as sondas oligonucleotídicas podem hibridar com os alvos de ácidos nucleicos.

Na etapa 240, as sondas oligonucleotídicas que não hibridaram com alvos de ácidos nucleicos são removidas, e desta forma separadas a partir de sondas oligonucleotídicas que hibridaram com alvos de ácidos nucleicos. Nesta forma de realização, a separação pode ser alcançada através de, por exemplo, lavagem da amostra para remover sondas oligonucleotídicas não hibridadas. A seguir, Na etapa 250, são administrados oligonucleótidos de codificação (os agentes de codificação) à amostra biológica de acordo com um padrão espacial conhecido, onde os oligonucleótidos de codificação compreendem etiquetas de codificação que são utilizadas para codificar a localização de alvos biológicos na amostra biológica. Note-se que em contraste com o sistema de ensaio da Figura 1, aqui as sondas e agentes de codificação

(oligonucleótidos de codificação) são administrados em etapas separadas. Na etapa 260, os oligonucleótidos de codificação são acoplados às sondas oligonucleotídicas para criar sondas codificadas. Neste caso onde as sondas são oligonucleótidos, os oligonucleótidos de codificação podem ser acoplados às sondas oligonucleotídicas através de, por exemplo, ligação. Alternativamente, a informação nos oligonucleótidos de codificação pode ser transferida através da utilização de uma ADN polimerase para estender um oligonucleótido de sonda que atua como um iniciador, e desta forma copiar e incorporar a sequência dos oligonucleótidos de codificação.

Na etapa 270, é determinada a sequência das etiquetas de codificação nas sondas codificadas bem como a sequência ou uma porção da sequência da própria sonda, e na etapa 280, os ácidos nucleicos alvo são mapeados de volta para a mostra biológica. Em algumas formas de realização, a abundância de sequência revela a quantidade relativa de alvos biológicos na localização. Embora esta forma de realização mostre as etapas individuais numa ordem particular, de forma a explicar melhor a invenção, a ordem precisa das etapas pode ser variada. Por exemplo, as etapas 220 e 250 podem ser combinadas, de forma a ser administrada uma mistura das sondas e oligonucleótidos de codificação de acordo com um padrão espacial eleito. A etapa de acoplamento 260 pode então ser levada a cabo imediatamente após as etapas combinadas 220 e 250, ou concomitantemente com as mesmas. Neste caso, a etapa 240 ocorreria então após a etapa 260. Pode desta forma ser apreciado que os dois resultados chave desta série de etapas, isto é, a codificação específica de localização de moléculas de sonda e a separação de moléculas de sonda com base na sua capacidade de interagir com moléculas alvo correspondentes, podem ser alcançados com alguma flexibilidade na implementação das etapas particulares. De forma semelhante, existe flexibilidade considerável no desenho do esquema de codificação. Conforme descrito abaixo, os ensaios da invenção são particularmente passíveis a métodos

combinatórios.

Assim, a presente invenção proporciona uma capacidade de visualizar vários alvos biológicos diferentes em muitas localizações, proporcionando a resolução da hibridação *in situ* com a análise de dados altamente paralela da sequenciação. Em algumas formas de realização, a soma dos múltiplos alvos biológicos sendo submetido a ensaio e os múltiplos sítios na amostra biológica é maior do que 20, noutras formas de realização, a soma dos múltiplos alvos biológicos sendo submetido a ensaio e os múltiplos sítios na amostra biológica é maior do que 50, noutras formas de realização, a soma dos múltiplos alvos biológicos sendo submetido a ensaio e os múltiplos sítios na amostra biológica é maior do que 100, maior do que 500 nm, 1.000, 10.000, 25.000, 100.000, 500.000, 1.000.000. Será apreciado que, devido à dimensão de codificação espacial da invenção, números ainda maiores podem ser contemplados. Por exemplo, realizar o ensaio de 10.000 alvos por localização x 10.000 localizações geraria  $10^8$  ensaios diferentes, e números ainda maiores que estes podem ser facilmente contemplados, particularmente se são utilizadas localizações espaciais com resolução na ordem daquela de células individuais. Além disso, nas formas de realização onde a sequenciação digital de alto rendimento é empregue, as sequências de pelo menos 1.000 sondas de codificação são, tipicamente, determinadas em paralelo. Mais tipicamente, utilizando uma leitura digital, é desejável obter múltiplas leituras de sequências múltiplas para cada ensaio (definidas por uma sonda e um código de localização espacial). É desejável obter uma média de pelo menos 3 cópias por ensaio, e mais tipicamente pelo menos 10 ou pelo menos 30 cópias por ensaio, dependendo do desenho da experiência e requisitos do ensaio. Para uma leitura quantitativa com um intervalo dinâmico adequado, pode ser desejável obter pelo menos 1.000 leituras por ensaio. Como tal, se 1.000.000 ensaios são levados a cabo, o número de leituras de sequências pode ser de 1 bilião ou mais.

Com a sequenciação digital de alto rendimento, e permitindo redundância, a sequência de pelo menos 10.000 sondas de codificação são determinadas em paralelo ou a sequência de pelo menos 100.000, 500.000, 1.000.000, 10.000.000, 100.000.000, 1.000.000.000 ou mais sondas de codificação são determinadas em paralelo.

#### Ensaaios

A porção de ensaio do sistema de ensaio compreende as seguintes etapas gerais: administração de sondas e agentes de codificação onde os agentes de codificação, (em algumas formas de realização pré-acoplados às sondas) são administrados à amostra de acordo com um padrão espacial conhecido, permitindo às sondas interagir ou reagir com alvos biológicos na amostra, e, se as sondas e os agentes de codificação não foram pré-acoplados, acoplamento dos agentes de codificação às sondas.

As amostras incluem virtualmente qualquer amostra ou amostras biológicas que podem ser fixas a um suporte ou proporcionadas essencialmente de um modo bidimensional, onde a capacidade de vincular um alvo ou atividade biológica submetida a ensaio de volta à localização dentro da amostra biológica é importante. Amostras biológicas exemplares incluem secções de tecido (por exemplo, incluindo o seccionamento de animais inteiros e biópsias de tecido), populações celulares em lâminas ou placas de cultura, e semelhantes. Os sistemas de ensaio são particularmente vantajosos no sentido em que são compatíveis com numerosos tipos de amostras biológicas, incluindo amostras frescas, tais como secções de tecidos primárias, e amostras preservadas (incluindo, mas não limitadas a, amostras congeladas e amostras fixas em paraformalina, e incorporadas em parafina (FFPE)). Um aspeto importante dos sistemas de ensaio consiste em que as amostras biológicas são imobilizadas numa superfície de substrato tendo áreas discretas e independentemente mensuráveis.

Os alvos biológicos a serem detetados podem ser quaisquer

moléculas biológicas incluindo, mas não limitadas a, proteínas, ácidos nucleicos, lípidos, hidratos de carbono, iões, ou complexos multicompetentes contendo qualquer dos anteriores. Exemplos de alvos subcelulares incluem organelos, por exemplo, mitocôndrias, aparelho de Golgi, retículo endoplasmático, cloroplastos, vesículas endocíticas, vesículas exocíticas, vacúolos, lisossomas, etc.

Em algumas formas de realização particulares, o sistema de ensaio é utilizado para analisar ácidos nucleicos, por exemplo, através de genotipagem, quantificação do número de cópias de ADN ou transcritos de ADN, localização de transcritos particulares dentro das amostras, e semelhantes. A Figura 3 ilustra um esquema geral para um ensaio exemplar para, por exemplo, detetar polimorfismos de nucleótido único (SNPs) que podem ser utilizados com o sistema de ensaio. Na Figura 3, são proporcionadas duas sondas oligonucleotídicas. Cada sonda oligonucleotídica compreende uma região específica de alvo (localizada em cada lado do SNP a ser analisado) observada em 305 e 307, e regiões de ligação, observadas em 301 e 303. Permite-se às sondas oligonucleotídicas hibridar com um ácido nucleico alvo (não mostrado) na amostra biológica. Na etapa 302, uma das sondas oligonucleotídicas é estendida para incorporar a sequência de SNP e ligada à outra sonda para formar uma sonda estendida que compreende a região de ácido nucleico alvo 309 e regiões de ligação 301 e 303.

Dois agentes de codificação, ambos compreendendo uma etiqueta de codificação (observada em 315 e 317), uma região de ligação (observada em 311 e 313), e uma primeira região (observada em 319 e 321) são combinadas com e ligadas à sonda estendida na etapa 304 para formar um oligonucleótido específico de alvo codificado. Novamente, em contraste com a Figura 1, as sondas e agentes de codificação são administradas em etapas separadas. Proceder desta forma permite a utilização das formas de realização combinatórias descritas abaixo. Em formas de realização preferidas, Os oligonucleótidos



codificante dentro de um par de oligonucleótidos de codificação ligam-se especificamente a um lado da sequência alvo ou ao outro (ou seja, 5' ou 3' da sequência alvo). Também, tipicamente, as regiões de ligação e de iniciador dos oligonucleótidos de codificação e sondas são universais; ou seja, o conjunto de ligação e regiões de iniciador utilizadas na construção das sondas e oligonucleótidos de codificação são constantes, e apenas diferem as regiões específicas de alvo das sondas e as etiquetas de codificação dos oligonucleótidos de codificação. Contudo, novamente em formas de realização alternativas, as regiões de ligação e iniciador não são universais e diferem entre sondas e agentes de codificação.

Após a ligação, as sondas codificadas são eluídas, agrupadas, e, opcionalmente, adaptadores de sequenciação são adicionados às sondas codificadas através de PCR. Em formas de realização alternativas, os iniciadores de sequenciação podem ser ligados aos oligonucleótidos de codificação, ou sequências de iniciador de sequenciação podem ser incluídas como parte do oligonucleótido codificante. Como observado na Figura 3, cada adaptador de sequenciação compreende uma região de iniciador 319 ou 321, compatível com as regiões de iniciador 319 e 321 nas sondas codificadas. A construção final que compreende o primeiro adaptador 327, primeira região de iniciador 319, primeira etiqueta de codificação 315, regiões de ligação 311 e 301, região alvo 309, regiões de ligação 313 e 303, segunda etiqueta de codificação 317, segunda região de iniciador 325 e segundo adaptador 329 está agora pronta para entrar num processo de sequenciação digital de alto rendimento.

Uma combinação de reações de extensão e ligação são exemplificadas na Figura 3, mas deve ser apreciado que uma variedade de reações podem ser utilizadas para acoplar os oligonucleótidos de codificação aos oligonucleótidos específicos de alvo, incluindo apenas ligação (por exemplo, para oligonucleótidos que hibridam com porções contínuas da sequência de ácidos nucleicos alvo). Alternativamente, um

ensaio utilizando um oligonucleótido adicional, tal como no ensaio GOLD-ENGATE® (veja-se Fan, *et al.*, Cold Spring Symp. Quant. Biol., 68:69-78 (2003); (Illumina, Inc., San Diego, CA)), pode ser empregue.

Noutras formas de realização, o sistema de ensaio também pode ser utilizado para analisar péptidos ou proteínas, a presença de anticorpos, atividades enzimáticas e de outras proteínas, modificações pós-tradução, formas de péptidos ativas e não ativas, bem como isoformas de péptidos numa amostra biológica. Consequentemente, as sondas podem compreender uma região ativa de uma enzima, um domínio de ligação de uma imunoglobulina, domínios definidos de proteínas, proteínas completas, péptidos sintéticos, péptidos com mutações introduzidas, aptâmeros e semelhantes.

Em determinados aspetos, as sondas são substratos para enzimas e proenzimas, por exemplo, cinases, fosfatases, zimogénios, proteases, ou fragmentos dos mesmos. Em certos aspetos, as sondas são substratos de fosforilação utilizados para detetar proteínas envolvidas em uma ou mais vias de transdução de sinal, por exemplo, uma cinase ou uma fosfatase. Noutra abordagem específica as sondas são substratos de protease específicos que se associam apenas com proteases ou classes de proteases individuais. Noutras abordagens, as sondas são diferentes formas processadas, isoformas e/ou domínios de uma enzima. Sondas baseadas em enzimas são tipicamente conjugadas ou, de outra forma, ligadas a agentes de codificação de oligonucleótidos. Os agentes de codificação de oligonucleótidos neste caso poderiam também incluir um componente de sequência nucleotídica que permite a identificação da sonda proteica.

Em determinados aspetos, a presente invenção proporciona ensaios para avaliar diferenças na quantidade e/ou atividade de alvos biológicos entre diferentes localizações numa amostra e/ou entre amostras. O método inclui determinar uma pluralidade de resultados codificados a partir da amostra biológica e

avaliar as diferenças na quantidade dos alvos biológicos em cada localização na amostra biológica.

### **Formas de Realização Combinatórias**

Para maximizar a eficiência da codificação, pode ser utilizada uma abordagem combinatória utilizando pares de etiquetas de codificação nos oligonucleótidos de codificação. Através do desacoplamento da informação específica de alvo e das etiquetas de codificação, o número de oligonucleótidos necessários é dramaticamente reduzido, com uma concomitante diminuição no custo.

A Figura 4 ilustra um mecanismo geral para uma forma de realização de um esquema de codificação combinatório da invenção, onde ácidos nucleicos numa secção de tecido representativa (mostrado em 416) são submetidos a ensaio. A Figura 4 em A mostra duas construções oligonucleotídicas específicas de alvo/de codificação 420 e 422 (por exemplo, formadas entre as etapas 302 e 304 da Figura 3) especialmente ligadas a um ácido nucleico alvo 402 de interesse. A primeira sonda de codificação 420 compreende a etiqueta de codificação 408, associada com, por exemplo, um sítio de iniciação universal para a amplificação dos produtos de ensaio ou um adaptador para permitir a identificação dos identificadores de codificação utilizando tecnologias de sequenciação 404. A segunda sonda codificada 422 compreende a etiqueta de codificação 406, associada com, por exemplo, um sítio de iniciação universal para a amplificação dos produtos de ensaio ou um adaptador para permitir a identificação dos identificadores de codificação utilizando tecnologias de sequenciação 410.

A Figura 4 em B mostra o padrão espacial que pode ser utilizado para vinte etiquetas de codificação diferentes, al até a10 (etiqueta de codificação 406 na sonda codificada 420) e b1 até b10 (etiqueta de codificação 408 na sonda codificada 422). A etiqueta de codificação a1, por exemplo, é depositada na amostra biológica em dez áreas ou pontos diferentes

(mostrados como a primeira linha horizontal de pontos em 412). A etiqueta de codificação a2 é depositada na amostra biológica em dez pontos na segunda linha horizontal em 412. A etiqueta de codificação a3 é depositada na amostra biológica em dez pontos na terceira linha horizontal em 412, e assim por diante. Enquanto as etiquetas "a" são depositadas em dez filas horizontais, as etiquetas "b" são depositadas em dez linhas verticais como mostrado em 414. Por exemplo, a etiqueta de codificação b1 é depositada na amostra biológica em dez pontos na primeira linha vertical em 414, a etiqueta de codificação b2 é depositada na amostra biológica em dez pontos na segunda linha vertical em 414, e assim por diante. A utilização de tal configuração permite que vinte etiquetas de codificação definam unicamente 100 localizações diferentes na amostra biológica.

A Figura 4 em C mostra uma secção de tecido representativa 416 coincidente com a grelha de etiqueta de codificação 418. As setas mostram como as etiquetas de codificação "a" e as etiquetas de codificação "b" são depositadas na grelha 418 que é coincidente com a secção de tecido 416. Se, uma vez sequenciadas, as etiquetas de codificação a1 e b4, por exemplo, são associadas com uma sequência de ácidos nucleicos alvo, então a sequência de ácidos nucleicos alvo (ou seja, o alvo biológico) estava presente na secção de tecido na localização a1, b4.

A Figura 5 proporciona um exemplo específico e simplificado do esquema de codificação dos sistemas de ensaio utilizando a invenção. A Figura 5 mostra os oligonucleótidos de codificação 510, compreendendo a1, a2, a3, a4 e b1, b3, b3 e b4. Os oligonucleótidos específicos de alvo (TSOs) (sondas) 1 e 2 são mostrados em 520. Um esquema de depósito ou deposição é mostrado em 530. Tal como a grelha exemplificada na Figura 4, os oligonucleótidos de codificação a1 até a4 são depositados em pontos num padrão (aqui, num padrão vertical), os oligonucleótidos de codificação b1 até b4 são depositados em

pontos num padrão (aqui, um padrão horizontal). No entanto, a grelha mostrada como um quadrado com pontos é, realmente, um padrão de deposição sobre uma amostra biológica (não mostrado) tal como uma secção de tecido 416 mostrada na Figura 4.

Os oligonucleótidos específicos de alvo são administrados à amostra biológica, onde os oligonucleótidos específicos de alvo hibridam com os ácidos nucleicos alvo na amostra biológica se os ácidos nucleicos alvo estão presentes. Os oligonucleótidos específicos de alvo não hibridados são depois removidos, por exemplo, por lavagem. Os oligonucleótidos de codificação são depois administrados à amostra biológica de acordo com o padrão espacial mostrado em 530. Os oligonucleótidos de codificação são ligados (ou, por exemplo, estendidos e ligados) a quaisquer oligonucleótidos específicos de alvo que hibridam com o ácido nucleico alvo na amostra biológica, as construções ligadas são depois eluidas a partir da amostra biológica, agrupadas, e adaptadores de sequenciação são adicionados através de, por exemplo, PCR ou ligação, se as sequências não estavam previamente incluídas nos oligonucleótidos de codificação. As construções ligadas são sequenciadas através de, por exemplo, sequenciação de alto desempenho ou de "próxima geração".

O agrupamento das sequências resultantes é mostrado em 540. A leitura de sequência foi obtida para o oligonucleótido específico de alvo 1 apenas em a4b1, a4b2, alb3, a2b3, a3b3, a4b3 e a4b4 (posições mostradas em linhas horizontais). Uma leitura de sequência foi obtida para o oligonucleótido específico de alvo 2 apenas em alb1 (posição mostrada com linhas verticais). Uma leitura de sequência foi obtida para ambos os oligonucleótidos específicos de alvo 1 e 2 nas posições a2b1, a3b1, alb2, a2b2, e a3b2 (posições mostradas com sombreado transversal). Não foi obtida qualquer leitura de sequência para quaisquer oligonucleótidos específicos de alvo em alb4, a2b4 ou a3b4 (posições mostradas sem sombreado). Assim, na amostra biológica na qual o ensaio foi realizado, o primeiro ácido

nucleico alvo foi detetado numa grande porção do lado esquerdo e na parte de baixo da amostra biológica, o segundo ácido nucleico alvo foi detetado apenas na porção superior esquerda da amostra biológica, e nenhum ácido nucleico alvo foi detetado na porção superior direita da amostra biológica. A expressão diferencial dos dois ácidos nucleicos alvo pode agora ser mapeada de volta à amostra biológica e às estruturas biológicas ou tipos celulares nestas localizações na amostra biológica.

Adicionalmente à informação de localização, pode ser obtida informação que se refere à abundância relativa das etiquetas codificadas. Por exemplo, se se verifica que existem dez vezes mais sequências a4T1b1 ocorrendo no conjunto de dados em comparação com as sequências a4T1b2, isto indicaria que a sequência de ácidos nucleicos alvo 1 é dez vezes mais abundante na localização a4T1b1 do que na localização a4T 1b2.

No caso da análise de nucleótidos como mostrada na Figura 3, através da ligação das etiquetas de codificação a oligonucleótidos específicos de alvo, apenas são necessários  $2n$  oligonucleótidos específicos de alvo para  $h$  alvos. Por exemplo, utilizando a abordagem combinatória delineada na Figura 2, submeter a ensaio 100 alvos diferentes em 10.000 localizações espaciais exigiria  $2 \times 100$  oligonucleótidos específicos de alvo e  $2 \times 100$  oligonucleótidos de codificação: A contagem total dos oligonucleótidos do ensaio seria apenas de 400 (200 específicos de alvo e 200 de codificação), não contando os iniciadores universais. Em contraste, se os oligonucleótidos de codificação foram desacoplados por calor dos oligonucleótidos específicos de alvo,  $(n \times X \text{ códigos de posição}) + (n \times Y \text{ códigos de posição})$  seriam necessários, ou no exemplo anterior, 20.000 oligonucleótidos, não contando as sequências de iniciadores universais. Além disso, embora as formas de realização mostradas nas Figuras 2-5 ilustrem um esquema combinatório utilizando dois agentes de codificação (etiquetas de codificação), três, quatro ou mais agentes de codificação e etiquetas de codificação podem ser utilizados,

e ligados à sonda ou um ao outro através de vários meios e em várias combinações de etapas.

Devido ao aspeto de codificação espacial do sistema de ensaio, uma grande quantidade de informação pode ser gerada mesmo com um número modesto de ensaios. Por exemplo, cinco ou mais alvos biológicos submetidos a ensaio em cinco ou mais posições na amostra geram 25 ou mais combinações. Ao utilizar a sequenciação digital como uma leitura, o número ótimo de leituras de sequências por combinação depende da sensibilidade e intervalo dinâmico necessários, e pode ser ajustado. Por exemplo, se por cada combinação são amostradas, em média, 100 leituras, o total de 25 combinações são 25.000 leituras. Se 1.000 alvos são submetidos a ensaio em 1.000 localizações com uma profundidade de amostragem média de 1.000, então são necessárias  $10^9$  leituras. Estes números, embora grandes, encontram-se dentro da capacidade dos métodos de sequenciação digital intrinsecamente paralelos, que podem gerar conjuntos de dados de biliões ou mesmo triliões de leituras num período de tempo razoável a um custo muito baixo por leitura. Como tal, através da variação dos números de posições interrogadas ou alvos biológicos submetidos a ensaio, ou ambos, e utilizando sequenciação digital, grandes quantidades de informação podem ser obtidas. Em aspetos específicos; múltiplas localizações são interrogadas para duas ou mais moléculas biológicas.

### **Sistemas de Administração de Reagentes**

O sistema de administração de reagentes inclui instrumentos que permitem a administração de reagentes a porções discretas da amostra biológica, mantendo a integridade dos padrões espaciais do esquema de codificação. Os sistemas de administração de reagentes dos sistemas de ensaio compreendem meios de produção de imagem opcionais; *hardware* e *software* de controlo de administração de reagentes. A administração de reagentes pode ser alcançada num número de diferentes formas. Deve notar-se que a administração de reagentes pode ser a várias amostras biológicas diferentes ao

mesmo tempo. Uma única secção de tecido foi exemplificada no presente documento; no entanto, múltiplas amostras biológicas podem ser fixadas e analisadas simultaneamente. Por exemplo, secções em série de uma amostra de tecido podem ser analisadas em paralelo e os dados combinados para construir um mapa em 3D.

Os instrumentos que permitem a padronização espacial de reagentes sobre a amostra biológicas são integrais com o sistema de ensaio. Tecnologias para formular e administrar ambos moléculas biológicas (por exemplo, oligonucleótidos ou anticorpos) e reagentes químicos (por exemplo, pequenas moléculas ou dNTPs) são conhecidas na técnica, e utilizações destes sistemas de instrumento são conhecidos para um perito na especialidade e são facilmente adaptáveis aos sistemas de ensaio. Um exemplos de um sistema de administração de reagentes é o manipulador de líquidos acústico Labcyte™ Echo, que pode ser utilizado para administrar gotículas na escala dos nanolitros contendo moléculas biológicas com elevada precisão e reprodutibilidade. Um perito na especialidade poderia incorporar este dispositivo de administração de reagentes no sistema global, utilizando *software* para especificar as localizações às quais os reagentes devem ser administrados.

Outros instrumentos que podem ser utilizados para a deposição de agentes e/ou identificadores de codificação sobre as amostras biológicas incluem, mas não estão limitados a, marcação por jato de tinta; marcação mecânica através de pino, caneta ou capilar; impressão de micro contacto; métodos fotoquímicos ou fotolitográficos; e semelhantes. Para várias aplicações, pode ser preferido segmentar ou sequestrar determinadas áreas das amostras biológicas em uma ou mais áreas de ensaio para diferentes distribuições de reagente e/ou determinação de alvos biológicos. As áreas de ensaio podem ser fisicamente ser fisicamente separada utilizando barreiras ou canais.

Num aspeto exemplar, o sistema de administração de reagentes pode ser um sistema baseado em fluxo. Os sistemas



baseados em fluxo para a administração de reagentes na presente invenção podem incluir instrumentos tais como uma ou mais bombas, válvulas, reservatórios de fluido, canais, e/ou células de armazenamento de reagentes. Os sistemas de administração de reagentes estão configurados para mover fluido para contactar com uma secção discreta da amostra biológica. O movimento dos reagentes pode ser conduzido por uma bomba disposta, por exemplo, a jusante dos reagentes fluidos. A bomba pode conduzir cada reagente fluido para (e através) do compartimento de reação. Alternativamente, os reagentes podem ser conduzidos através do fluido por gravidade. As Publicações U.S. N.ºs 20070166725 e 20050239192 divulgam certas ferramentas fluídicas de uso geral que podem ser utilizadas com os sistemas de ensaio, permitindo a manipulação precisa de gases, líquidos e sólidos para alcançar manipulações analíticas muito complexas com *hardware* relativamente simples.

Num exemplo mais específico, uma ou mais células de fluxo podem ser ligadas à amostra biológica fixa ao substrato desde cima. A célula de fluxo pode incluir tubos de entrada e saída conectados à mesma e opcionalmente uma bomba externa é utilizada para administrar reagentes à célula de fluxo e através da amostra biológica. As câmaras de fluxo estão configuradas para administrar reagentes apenas a determinadas porções da amostra biológica, restringindo a quantidade e tipo de reagente administrado a qualquer secção específica da amostra biológica.

Noutro aspeto, um sistema microfluídico pode ser integrado no substrato sobre o qual a amostra biológica é colocada ou externamente ligado em cima do substrato. As passagens microfluídicas para manter e carregar fluido podem ser formadas sobre e/ou acima do substrato planar através de uma camada fluídica contígua com o substrato. Os reagentes fluidos podem ser seleccionados e administrados de acordo com a abertura e encerramento seletivos de válvulas dispostas entre os reservatórios de reagentes.

As bombas geralmente incluem qualquer mecanismo para mover fluido e/ou reagentes dispostos em fluido. Em alguns exemplos, a bomba pode ser configurada para mover fluido e/ou reagentes através de passagens com pequenos volumes (ou seja, estruturas microfluídicas). A bomba pode operar mecanicamente exercendo uma pressão positiva ou negativa sobre o fluido e/ou sobre uma estrutura carregando fluido, eletricamente através da aplicação apropriada de um ou mais campos elétricos, ou ambos, entre outros meios. Bombas mecânicas exemplares podem incluir bombas de seringa, bombas peristálticas, bombas rotatórias, gás pressurizado, pipetas, etc. As bombas mecânicas podem ser microusinadas, moldadas, etc. Bombas elétricas exemplares podem incluir eletrodos e podem funcionar por eletroforese, eletroosmose, eletrocapilaridade, dieletroforese (incluindo formas de onda migratória das mesmas), e/ou semelhantes.

As válvulas incluem, geralmente, qualquer mecanismo para regular a passagem de fluido através de um canal. As válvulas podem incluir, por exemplo, elementos deformáveis que podem ser seletivamente deformados para parcial ou completamente encerrar um canal, uma projeção móvel que podem ser seletivamente estendida para dentro de um canal para bloquear parcial ou completamente um canal, uma estrutura eletrocapilar, e/ou semelhantes.

Uma junta aberta pode ser ligada ao topo da amostra biológica e a amostra e reagentes podem ser injetados na junta. Materiais de junta adequados incluem, mas não estão limitados a, neopreno, nitrilo e borracha de silicone. Alternativamente, uma câmara de reação estanque à água pode ser formada por uma junta interposta entre a amostra biológica sobre o substrato e um material quimicamente inerte e resistente à água tal como, mas não limitado a, alumínio negro anodizado, termoplásticos (por exemplo, poliestireno, policarbonato, etc.), vidro, etc.

Numa forma de realização opcional, o sistema de ensaio compreende meios de produção de imagem para determinar as

características e organização da amostra biológica de interesse. As imagens obtidas, por exemplo, podem ser utilizadas para desenhar o padrão de deposição dos reagentes. Os meios de produção de imagem são opcionais, já que um indivíduo pode, por outro lado, visualizar a amostra biológica utilizando, por exemplo, um microscópio, analisar a organização da amostra biológica, e especificar um padrão espacial para administração de reagentes de ensaio. Se incluído, o sistema de administração pode compreender uma disposição de microcircuito incluindo um gerador de imagens, tal como baseado em CCD ou IGFET (por exemplo, baseado em CMOS) e um pulverizador ultrassónico para a administração de reagentes conforme descrito na Publicação U.S. N.º 20090197326. Também, deve notar-se que embora as Figuras 4 e 5 ilustram utilizando uma configuração de grelha x,y, outras configurações podem ser utilizadas, tais como, por exemplo, seguindo a topologia de uma amostra de tecido; direcionando determinados grupos celulares, camadas celulares e/ou tipos celulares num tecido, e semelhantes.

Ainda noutra alternativa, o sistema de administração de reagentes controla a administração de reagentes para padrões específicos na superfície de uma amostra biológica utilizando técnicas de semicondutor tais como mascaramento e pulverização. Áreas específicas de uma amostra biológica podem ser protegidas da exposição a reagentes através da utilização de uma máscara para proteger áreas específicas da exposição. Os reagentes podem ser introduzidos na amostra biológica utilizando técnicas convencionais tais como pulverização ou fluxo de fluido. A utilização da administração mascarada resulta num esquema de administração padronizado sobre a superfície do substrato.

Numa abordagem preferida, os instrumentos de administração de reagente são baseados em tecnologia de impressão por jato de tinta. Existe uma variedade de diferentes mecanismos de jato de tinta (por exemplo, térmicos e

piezoelétricos) e a compatibilidade foi mostrada com formulações de tinta aquosas e orgânicas. Conjuntos de bocais independentemente atuados podem ser utilizados para administrar múltiplos reagentes ao mesmo tempo, e resoluções muito elevadas podem ser alcançadas.

De modo a atingir sítios alvo específicos de interesse, uma imagem informativa da amostra biológica a ser submetida a ensaio pode ser utilizada para auxiliar nos métodos de administração de reagentes e esquema de codificação associado. Regiões de amostra da amostra biológica podem ser identificadas utilizando processamento de imagens (por exemplo, imagens de tipos celulares diferenciadas por imunohistoquímica ou outras químicas de coloração) integradas com outras características do sistema de ensaio. Em alguns aspetos, o *software* é utilizado para traduzir automaticamente a informação das imagens num padrão de administração de reagentes. Um mecanismo para registar e alinhar de forma muito precisa a amostra para a administração de reagentes é, portanto, um componente importante dos sistemas de ensaio utilizando a invenção. Mecanismos tais como a utilização de marcadores fiduciais sobre lâminas e/ou outros sistemas de posicionamento físico preciso podem ser adaptados para este propósito.

A invenção é preferentemente utilizada com um conjunto de aplicações de *software* adaptado ao sistema de ensaio. Opcionalmente, o *software* de desenho de oligonucleótidos é utilizado para desenhar os nucleótidos de codificação (e em formas de realização onde os ácidos nucleicos são submetidos a ensaio, os oligonucleótidos específicos de alvo) para o ensaio específico a ser realizado, e pode ser integrado como uma parte do sistema. Também opcionalmente, algoritmos e *software* para administração de reagentes e análise de dados (ou seja, análise de sequência) podem ser integrados para determinar os resultados do ensaio. A análise de dados integrada é particularmente útil, uma vez que o tipo de conjunto de dados que é gerado pode ser massivo como uma consequência

da escala. Algoritmos e ferramentas de *software* que são especialmente desenhados para análise dos dados espacialmente associados gerados pelos sistemas de ensaio, incluindo ferramentas de visualização e *software* de análise de padrão, potenciam o valor dos dados gerados pelos sistemas de ensaio.

Em determinados aspetos, o sistema de ensaio compreende processos para fazer e levar a cabo o controlo de qualidade dos reagentes, por exemplo, a integridade e fidelidade da sequência de agrupamentos de oligonucleótidos: Em particular, os reagentes são formulados de acordo com fatores tais como a volatilidade, estabilidade em temperaturas chave, e compatibilidade química para compatibilidade com os instrumentos de administração de reagentes e podem ser analisados através de instrumentos integrados dentro do sistema de ensaio.

### **Sequenciação**

Numerosos métodos podem ser utilizados para identificar as etiquetas de codificação e sequências de sondas nas sondas codificadas dos sistemas de ensaio. As etiquetas de codificação podem ser detetadas utilizando técnicas tais como espectroscopia de massa (por exemplo, Maldi T de, LC-MS/MS), produção de imagem por ressonância magnética nuclear, ou, preferentemente, sequenciação de ácidos nucleicos. Exemplos de técnicas para descodificar as etiquetas de codificação podem ser encontradas, por exemplo, na Publicação U.S. N.º 20080220434. Por exemplo, as etiquetas de codificação podem ser etiquetas de massa de oligonucleótidos (OMTs ou *massTags*). Tais etiquetas são descritas, por exemplo, na Publicação U.S. N.º 20090305237. Ainda noutra alternativa, as sondas codificadas podem ser amplificadas e hibridadas com um microarranjo. Isto necessitaria que reações de amplificação separadas fossem levadas a cabo, nas quais cada amplificação é específica para um código ou subconjunto de códigos espaciais particulares, alcançada através da utilização de iniciadores específicos de código. Cada amplificação também poderia incorporar um

marcador resolúvel diferente (por exemplo, fluoróforo). Após a hibridação, as quantidades relativas ou um alvo particular mapeando em diferentes localizações espaciais na amostra podem ser determinados pelas abundâncias relativas dos marcadores resolúveis.

Num aspeto preferido particular, as etiquetas de codificação resultantes de acordo com o sistema de ensaio são substratos para a sequenciação de alto desempenho de nova geração, e são utilizados métodos de sequenciação de nova geração altamente paralelos para conformar a sequência das etiquetas de codificação, por exemplo, com tecnologia SOLiD™ (Life Technologies, Inc.) ou Genome Analyzer (Illumina, Inc.). Tais métodos de sequenciação de nova geração podem ser levados a cabo, por exemplo, utilizando um método de sequenciação de uma passagem ou utilizando sequenciação de terminais emparelhados. Métodos de sequenciação de nova geração incluem, mas não estão limitados a, métodos à base de hibridação, tais como divulgado em, por exemplo, Drmanac, Patentes U.S. N.ºs 6.864.052; 6.309.824; e 6.401.267; e Drmanac *et al*, Publicação de Patente U.S. 2005/0191656; métodos de sequenciação por síntese, por exemplo, Patente U.S. N.ºs 6.210.891; 6.828.100; 6.969.488; 6.897.023; 6.833.246; 6.911.345; 6.787.308; 7.297.518; 7.462.449 e 7.501.245; Pedidos de Publicação U.S. N.ºs 20110059436; 20040106110; 20030064398; e 20030022207; Ronaghi, *et al.*, Science, 281: 363-365 (1998); e Li, *et al.*, Proc. Natl. Acad. Sci. 100: 414-419 (2003); métodos à base de ligação, por exemplo, Patentes U.S. N.ºs 5.912.148 e 6.130.073; e as Patente U.S. N.ºs 20100105052, 20070207482 e 20090018024; sequenciação por nanoporos, por exemplo, Pedidos de Patente U.S. N.ºs 20070036511; 20080032301; 20080128627; 20090082212; e Soni e Meller, Clin Chem 53: 1996-2001 (2007)), bem como outros métodos, por exemplo, Pedidos de Patente U.S. N.ºs 20110033854; 20090264299; 20090155781; e 20090005252; além disso, veja-se, McKernan, *et al.*, Genome Res., 19:1527-41 (2009) e Bentley, *et al.*, Nature 456:53-59 (2008).

### **Aplicações do Sistema de Ensaio**

Será aparente para um perito na especialidade após a leitura da presente divulgação que existem várias áreas importantes de pesquisa biológica, diagnósticos, e desenvolvimento de fármacos que beneficiarão de um sistema de ensaio multiplexado de alto rendimento que pode medir simultaneamente a quantidade e localização espacial de um alvo biológico numa amostra biológica. Por exemplo, a combinação da capacidade de estimar a abundância relativa de diferentes transcritos de ARN com a capacidade de reconstruir uma imagem de padrões espaciais de abundância ao longo de várias localizações, que podem ser tão pequenas como ou inclusive mais pequenas que células individuais, num tecido permite várias áreas diferentes de pesquisa básica. As seguintes são utilizações exemplares e não se destinam de nenhuma forma a ser limitantes do âmbito.

Num exemplo, são determinados padrões tridimensionais da expressão génica através da análise de uma série de secções de tecido, de forma análoga à reconstrução de imagens por tomografia computadorizada. Um tal método pode ser utilizado para medir alterações da expressão génica em patologia de doenças, por exemplo, em tecido canceroso e/ou um tecido após lesão, inflamação ou infeção. Com os sistemas de ensaio utilizando a invenção, é obtida informação mais detalhada sobre expressão génica e localização de proteínas em tecidos complexos, levando a novas visões na função e regulação tanto em estados normais como de doença, e proporciona novas hipóteses que podem ser testadas. Por exemplo, um sistema de ensaio utilizando a invenção pode permitir algumas das visões adquiridas a partir de vários estudos individuais e que programas de maior tamanho como ENCODE (Birney, *et al.*, *Nature*, 447:799-816 (2007)) e modENCODE sejam integrados ao nível dos tecidos. Os sistemas de ensaio também auxiliam nos esforços computacionais de modelar redes de interação da expressão génica no campo da biologia de sistemas.

Os sistemas de ensaio proporcionam também uma abordagem inovadora da análise da variação somática, por exemplo, mutações somáticas no cancro de organismos com variabilidade de resposta a infeções. Por exemplo, os tumores são tipicamente heterogéneos, contendo células cancerígenas bem como células geneticamente normais num ambiente local anormal. As células cancerígenas são submetidas a mutação e seleção, e neste processo não é incomum o desenvolvimento de clones locais. A identificação de mutações somáticas relativamente raras no contexto de tumores pode permitir o estudo do papel das mutações chave na seleção de variantes clonais. Podem ser analisados padrões de transcrição associados com a angiogénese, inflamação, ou outros processos relacionados com o cancro tanto em células cancerígenas como geneticamente normais para visões na biologia do cancro e auxiliar no desenvolvimento de novos agentes terapêuticos para o tratamento de cancros. Noutro exemplo, os indivíduos têm suscetibilidade variada a organismos infecciosos, e os sistemas de ensaio utilizando a invenção podem ser utilizados para estudar a interação entre micróbios e tecidos ou com os vários tipos celulares dentro do tecido.

De forma importante, adicionalmente a proporcionar informação associada com o espaço, a invenção permite um grande aumento da sensibilidade de deteção de mutações raras, já que o sinal para ruído pode ser dramaticamente aumentado já que somente é ensaiada uma localização pequena em qualquer reação determinada. Num ensaio típico para mutações raras numa amostra mista, a amostra é tratada em bruto, isto é, os ácidos nucleicos são extraídos a partir de várias células para um único agrupamento. Assim, se uma mutação está presente numa célula em 10.000, deve ser detetada contra um fundo de ADN normal a partir de ~10.000 células. Em contraste, com os sistemas de ensaio utilizando a invenção podem ser analisadas várias células, mas células individuais ou pequenos grupos de células seriam identificados por parte do sistema de codificação



espacial. Como tal, nos sistemas de ensaio utilizando a presente invenção, o fundo é reduzido por ordens de magnitude, aumentando grandemente a sensibilidade. Além disso, pode ser observada a organização espacial de células mutantes, que pode ser particularmente importante na detecção de mutações chave em secções de tecido no cancro. Análises histológicas moleculares já proporcionam visões na biologia do cancro e podem ter potencial para utilização em diagnósticos. A tecnologia da invenção promete aumentar grandemente o poder de tais abordagens.

#### **EXEMPLOS**

Os exemplos seguintes são apresentados de modo a proporcionar aos peritos na especialidade uma divulgação e descrição completas de como preparar e utilizar a presente invenção, e não pretendem limitar o âmbito do que o inventor contempla como a sua invenção, nem pretendem representar ou implicar que as experiências abaixo são a totalidade ou as únicas experiências realizadas. As presentes formas de realização devem, como tal, ser consideradas em todos os aspetos como ilustrativas e não restritivas.

Foram realizados esforços para assegurar precisão com respeito aos números utilizados (por exemplo, quantidades, temperatura, etc.) mas alguns erros e desvios experimentais deveriam ser tomados em conta. A menos que seja indicado de outro modo, partes são partes em peso, massa molecular é massa molecular de massa média, temperatura é em graus centígrados, e a pressão é a atmosférica ou próxima.

#### **EXEMPLO 1: Prova Inicial do conceito de Esquema de Codificação**

Como uma prova inicial do conceito, é desenvolvido um sistema modelo utilizando um microarranjo para demonstrar um ensaio uniplexo funcional. O desenho básico valida o conceito do ensaio, e estabelece um ensaio funcional anteriormente à abordagem de questões relacionadas com a análise de uma amostra biológica mais complicada. É utilizada sequenciação convencional como uma leitura para esta prova de conceito.

É utilizado um microarranjo como um suposto para uma secção de tecido. As sequências alvo do microarranjo são especificadas totalmente, para que a composição dos alvos seja conhecida e possa ser variada sistematicamente. Moldes oligonucleotídicos sintéticos são ligados a uma lâmina de vidro através de uma modificação amino 5'. Cada lâmina tem uma única sequência molde oligonucleotídica, e os ensaios que são levados a cabo podem empregar qualquer ligação, ou extensão seguida por ligação já que isto pode ser útil na determinação de determinados polimorfismos.

Após a parte *in situ* do ensaio estar completa, os produtos de reação são eluídos e analisados através de qPCR para determinar a presença ou ausência de um produto e rendimento estimado, e através de sequenciação convencional para determinar a estrutura dos produtos do ensaio. Os ensaios uniplexo que são testados incluem controlos positivos e negativos adequados, e uma variante nucleotídica única (SNV) para verificar a capacidade de discriminação de variantes de base única.

#### **EXEMPLO 2: Escalabilidade**

A complexidade do sistema de ensaio é aumentada para estabelecer a escalabilidade do ensaio para utilização em estudos de alto rendimento. A escalabilidade dos sistemas tanto de codificação espacial como de ensaio é demonstrada através da realização de um ensaio 24-plexo x 24-sítios utilizando um sistema modelo de micro arranjo.

A quantidade de alvo biológico, aqui uma sequência alvo de ADN, em cada ensaio é variada sistematicamente no substrato do microarranjo. Por exemplo, num microarranjo com tamanho de ponto de 50 micra (de centro a centro), uma área de 1 mm<sup>2</sup> contém ~400 pontos. A região à volta de cada sítio é opcionalmente ocupada por uma região que é isenta destes pontos para permitir a resolubilidade das sequências alvo. Alternativamente, os pontos podem ser agrupados com dois ou mais pontos adjacentes rodeados por ou adjacentes a uma região que é isenta de

sequências alvo.

De forma a demonstrar que a codificação espacial é precisa, os sítios compreendem diferentes composições alvo para mostrar que a leitura do ensaio corresponde à composição esperada de cada sítio. Com 24 sequências alvo, é realizado um padrão digital simples com cada sítio tendo um conjunto diferente de 12 alvos presentes e 12 alvos ausentes, para fabricar um código binário (0 = ausente, 1 = presente). A leitura do ensaio é então determinada para mostrar que as regiões detetadas correspondem ao sinal esperado após a descodificação espacial. Neste exemplo particular, o espaço do código é suficientemente grande ( $2^{24}$ ) para que inclusive alguns erros não resultariam na mistura de códigos diferentes. Além disso, este desenho permite a identificação de erros e permite uma estimativa não só da precisão da codificação espacial mas também da precisão da sinalização da presença ou ausência de sequências alvo.

A capacidade de detetar diferenças quantitativas é avaliada através da geração de curvas de dose-resposta para cada um dos 24 ensaios que são levados a cabo em cada sítio num ensaio de 24 sítios. Isto permite estimar o limite de deteção, intervalo dinâmico e potência para detetar uma dada alteração de vezes através do intervalo.

Noutro aspeto, um desenho de quadrado latino é utilizado para representar alvos individuais em razões diferentes através da variação do número de características para cada alvo. Por outras palavras, com múltiplos pontos num sítio, o número de sítios atribuídos a cada uma das 24 sequências alvo pode ser variado e cada um dos 24 sítios pode ter uma composição diferente. Um microarranjo de 1 x 3 polegadas é suficientemente grande para permitir múltiplas réplicas. Este conjunto maior de 24 sequências necessitará desconvolução, e isto é alcançado utilizando técnicas de alto desempenho tais como tecnologias de sequenciação de próxima geração (por exemplo, tecnologia SOLiD™ (Life Technologies, Inc., Carlsbad, CA) ou Genome

Analyzer (Illumina, Inc., San Diego, CA). A utilização do ensaio 24-plexo demonstra tanto a precisão da codificação e descodificação espacial, bem como a resposta quantitativa do sistema de ensaio.

**EXEMPLO 3: Adaptação do ensaio a amostras preservadas.**

ADN genómico é submetido a ensaio como uma prova de conceito para testar ARN, uma vez que proporciona uma forma para estabelecer um sinal de referência de cópia única. Uma vez que um ensaio de trabalho é desenvolvido para amostras FFPE, é adaptado a um ensaio de ARN. Para esta finalidade, as concentrações de oligonucleótidos de ensaio são submetidas a ensaio para assegurar a compatibilidade com multiplexação elevada. Assumindo um diâmetro celular de 10 micra, e administração de uma gotícula de reagente de 10 micra de diâmetro a uma célula individual, o volume da gotícula será ~500  $\mu$ l e pode conter  $\sim 3 \times 10^{11}$  moléculas a uma concentração de 1  $\mu$ M. Para submeter a ensaio 1.000 sequências alvo em 10.000 células, ~2.000 oligonucleótidos de direcionamento seriam necessários numa gotícula. Como tal, cada gotícula poderia conter ~160 milhões de cópias de cada oligo de ensaio, um vasto excesso em relação a alguns milhares de sequências alvo numa célula.

A manipulação de pequenos números absolutos de moléculas de produto geradas a partir de amostras muito pequenas ou comprometidas é potenciada para contrariar o problema de uma eficiência de baixa recuperação; ou seja, a eficiência de eluição e perdas resultantes da absorção de moléculas para superfícies são prevenidas. Uma abordagem para tratar o último problema consiste em incluir um material portador, tal como glicogénio ou ácidos nucleicos portadores.

**Exemplo 4: Adaptação do Ensaio a uma Amostra Biológica.**

Um molde de ARN de controlo é imobilizado a um suporte sólido de forma a criar um sistema artificial. O ensaio é realizado utilizando ADN ligase T4, que pode reparar entalhes em híbridos de ADN/ARN. Os ensaios são levados a cabo em lâminas

correspondentes, ou secções diferentes da mesma lâmina, onde num caso ADNg é submetido a ensaio e no outro ARN é submetido a ensaio. Quando se submete a ensaio ADNg a lâmina pode ser pré-tratada com ARNase, e quando se submete a ensaio ARN a lâmina é pré-tratada com ADNase. Os resultados do ensaio são confirmados pela extração de ADNd ou ARN e quantificação das quantidades relativas por qPCR ou RT-qPCR respetivamente.

De modo a tornar os ensaios de ARN da secção de tecido o mais informativos possível, as informações preexistentes sobre os níveis de expressão em tecidos específicos para atingir transcritos através de um intervalo de abundâncias são utilizadas no desenho do ensaio. Tanto os transcritos de elevada abundância, como alguns transcritos de média e baixa abundância, são direcionados para permitir uma avaliação inicial das características de desempenho quantitativas do ensaio.

O precedente meramente ilustra os princípios da invenção. Além disso, todos os exemplos e linguagem condicional recitada no presente documento destinam-se principalmente a auxiliar o leitor na compreensão dos princípios da invenção e os conceitos contribuídos pelos inventores para aprofundar a técnica, e devem ser considerados como não tendo limitação a tais exemplos e condições especificamente indicados. Além disso, todas as afirmações no presente documento que recitam princípios, aspetos e formas de realização da invenção bem como exemplos específicos da mesma, destinam-se a englobar equivalentes estruturais e funcionais das mesmas. Adicionalmente, pretende-se que tais equivalentes incluam ambos equivalentes atualmente conhecidos e equivalentes desenvolvidos no futuro, isto é, quaisquer elementos desenvolvidos que desempenhem a mesma função, independentemente da estrutura. O âmbito da presente invenção, como tal, não se destina a ser limitado às formas de realização exemplares apresentadas e descritas no presente documento.

**DOCUMENTOS REFERIDOS NA DESCRIÇÃO**

Esta lista de documentos referidos pelo autor do presente pedido de patente foi elaborada apenas para informação do leitor. Não é parte integrante do documento de patente europeia. Não obstante o cuidado na sua elaboração, o IEP não assume qualquer responsabilidade por eventuais erros ou omissões.

**Documentos de patente referidos na descrição**

- WO 2010019826 A [0005]
- US 6969589 B [0027]
- US 20060188875 A [0027]
- US 20070166725 A [0070]
- US 20050239192 A [0070]
- US 20090197326 A [0076]
- US 20080220434 A [0082]
- US 20090305237 A [0082]
- US 6864052 B, Drmanac [0083]
- US 6309824 B [0083]
- US 6401267 B [0083]
- US 20050191656 A, Drmanac [0083]
- US 6210891 B [0083]
- US 6828100 B [0083]
- US 6969488 B [0083]
- US 6897023 B [0083]
- US 6833246 B [0083]
- US 6911345 B [0083]
- US 6787308 B [0083]
- US 7297518 B [0083]
- US 7462449 B [0083]
- US 7501245 B [0083]
- US 20110059436 A [0083]
- US 20040106110 A [0083]
- US 20030064398 A [0083]

- US 20030022207 A [0083]
- US 5912148 A [0083]
- US 6130073 A [0083]
- US 20100105052 A [0083]
- US 20070207482 A [0083]
- US 20090018024 A [0083]
- US 20070036511 A [0083]
- US 20080032301 A [0083]
- US 20080128627 A [0083]
- US 20090082212 A [0083]
- US 20110033854 A [0083]
- US 20090264299 A [0083]
- US 20090155781 A [0083]
- US 20090005252 A [0083]

#### **Documentos de não patente citados na descrição**

- **SHI et al.** *Nature Biotechnology*, 2006, vol. 24 (9), 1151-61 [0003]
- **SLONIM; YANAI.** *Plos Computational Biology*, 2009, vol. 5 (10), 1000543 [0003]
- **VELCULESCU et al.** *Science*, 1995, vol. 270 (5235), 484-87 [0003]
- **SPURGEON et al.** *Plos ONE*, 2008, vol. 3 (2), 1662 [0003]
- **NUOVO.** *Genome Res*, 1995, vol. 4, 151-67 [0003]
- **ARMANI et al.** *Lab on a Chip*, 2009, vol. 9 (24), 3526-34 [0003]
- **MIR KALIM et al.** NUCLEIC ACIDS RESEARCH. OXFORD UNIVERSITY PRESS, GB, 01 January 2009, vol. 37, e5-1 [0006]
- **ANDERSON ; YOUNG.** Quantitative Filter Hybridization. *Nucleic Acid Hybridization*, 1985 [0030]
- **ALLAWI ; SANTALUCIA, JR.** *Biochemistry*, 1997, vol. 36, 10581-94 [0030]
- **Genome Analysis: A Laboratory manual Series.** 1999, vol. I-IV [0031]
- **Generic Variation: A Laboratory Manual.** 2007 [0031]

- PGR Primer: A Laboratory Manual. 2003 [0031]
- **BOWTELL ; SAMBROOK**. DNA Microarrays: A Molecular Cloning Manual. 2003 [0031]
- **MOUNT**. *Bioinformatics: Sequence and Genome Analysis*, 2004 [0031]
- **SAMBROOK ; RUSSELL**. Condensed Protocols from Molecular Cloning: A Laboratory Manual. 2006 [0031]
- **SAMBROOK ; RUSSELL**. Molecular Cloning: Laboratory Manual. Cold Spring Harbour Laboratory Press, 2002 [0031]
- **STRYER**. Biochemistry. W.H. Freeman, 1995 [0031]
- **GAIT**. Oligonucleotide Synthesis: A Practical Approach. IRL Press, 2002 [0031]
- **NELSON ; COX**. Lehninger. Principles of Biochemistry. W. H. Freeman Pub, 2000 [0031]
- **BERG et al**. Biochemistry. W.H. Freeman Pub, 2002 [0031]
- **FAN et al**. *Cold Spring Symp. Quant. Biol.*, 2003, vol. 68, 69-78 [0053]
- **RONAGHI et al**. *Science*, 1998, vol. 281, 363-365 [0083]
- **LI et al**. *Proc. Natl. Acad. Sci.*, 2003, vol. 100, 414-419 [0083]
- **SONI; MELLER**. *Clin Chem*, 2007, vol. 53, 1996-2001 [0083]
- **MCKERNAN et al**. *Genome Res.*, 2009, vol. 19, 1527-41 [0083]
- **BENTLEY et al**. *Nature*, 2008, vol. 456, 53-59 [0083]
- **BIRNEY et al**. *Nature*, 2007, vol. 447, 799-816 [0085]

Lisboa, 1 de Dezembro de 2015



## REIVINDICAÇÕES

1. Um método para determinar padrões espaciais de abundância ou atividade ou ambos de múltiplos alvos de ácidos nucleicos em múltiplos sítios numa amostra, compreendendo as seguintes etapas:

- proporcionar uma amostra fixa a um suporte;
- administrar sondas oligonucleotídicas a múltiplos alvos de ácidos nucleicos aos múltiplos sítios na amostra;
- permitir que as sondas oligonucleotídicas hibridem com os alvos de ácidos nucleicos;
- lavar as sondas oligonucleotídicas não hibridadas da amostra;
- administrar agentes de codificação a localizações dos múltiplos sítios na amostra de acordo com um padrão espacial conhecido, em que pelo menos dois agentes de codificação são administrados a cada um dos múltiplos sítios e a combinação dos agentes de codificação administrada a cada sítio é diferente;
- acoplar os agentes de codificação às sondas oligonucleotídicas hibridadas com os alvos de ácidos nucleicos para formar sondas codificadas;
- determinar a totalidade ou uma porção de uma sequência das sondas codificadas; e
- associar a abundância ou atividade ou ambas dos múltiplos alvos biológicos às localizações de múltiplos sítios na amostra.

2. Um método para determinar padrões espaciais de abundância ou atividade ou ambos de múltiplos alvos biológicos em múltiplos sítios numa amostra, compreendendo as seguintes etapas:

- proporcionar uma amostra fixa a um suporte;
- administrar sondas codificadas a múltiplos alvos biológicos num padrão espacial conhecido aos múltiplos sítios na

amostra, em que cada sonda codificada compreende uma região de sonda capaz de interação com um alvo biológico e uma etiqueta de codificação compreendendo um oligonucleótido que identifica a localização do sítio ao qual a sonda codificada é administrada;

permitir que as sondas codificadas interajam com os alvos biológicos;

determinar uma sequência das sondas codificadas, em que a sequência compreende a sequência oligonucleotídica da etiqueta de codificação, desta forma identificando a localização do sítio ao qual a sonda codificada é administrada; e

associar a abundância ou atividade ou ambas dos múltiplos alvos biológicos às localizações dos sítios na amostra.

3. O método de acordo com a reivindicação 2, compreendendo adicionalmente após a etapa de permissão e antes da etapa de determinação uma etapa de separação das sondas codificadas que interagem com os alvos biológicos das sondas codificadas que não interagem com os alvos biológicos.

4. O método de acordo com a reivindicação 3, em que os alvos biológicos são ácidos nucleicos e as sondas codificadas são sondas oligonucleotídicas.

5. O método de acordo com a reivindicação 1 ou 4, em que existem duas sondas oligonucleotídicas para cada um dos alvos biológicos de ácidos nucleicos.

6. O método de acordo com a reivindicação 3, em que os múltiplos alvos biológicos são proteínas, as regiões de sonda das sondas de codificação são proteínas e as etiquetas de codificação compreendem oligonucleótidos.

7. O método de acordo com a reivindicação 3, em que os

múltiplos alvos biológicos compreendem enzimas.

8. O método de acordo com a reivindicação 3, em que as regiões de sonda das sondas codificadas compreendem anticorpos, aptâmeros, moléculas pequenas, substratos enzimáticos, substratos enzimáticos putativos, ou agentes de captura por afinidade.

9. O método de acordo com qualquer uma das reivindicações 1, 3 e 6, compreendendo adicionalmente uma etapa de amplificação entre a etapa de separação ou a etapa de acoplamento e a etapa de determinação.

10. O método de acordo com qualquer uma das reivindicações 1-3, em que a etapa de determinação é realizada através de sequenciação de ácidos nucleicos ou sequenciação de alto rendimento.

11. O método de acordo com a reivindicação 1, em que a etapa de acoplamento é realizada através de ligação ou através de extensão seguida por ligação.

12. O método de acordo com qualquer uma das reivindicações 1-3, em que o produto dos múltiplos alvos biológicos a ser submetido a ensaio e os múltiplos sítios na amostra é maior do que 1.000.000.

13. O método de acordo com qualquer uma das reivindicações 1-3 e 6, em que as sequências de pelo menos um milhão de sondas de codificação são determinadas em paralelo.

14. O método de acordo com qualquer uma das reivindicações 1-3, em que o padrão espacial conhecido é determinado através de características histológicas da amostra.

15. O método de acordo com qualquer uma das reivindicações 1-3, em que o *hardware* programado com *software* realiza pelo menos duas etapas da etapa de administração, a etapa de separação, a etapa de determinação e a etapa de associação.

Lisboa, 1 de Dezembro de 2015

100

FIG. 1



FIG. 2

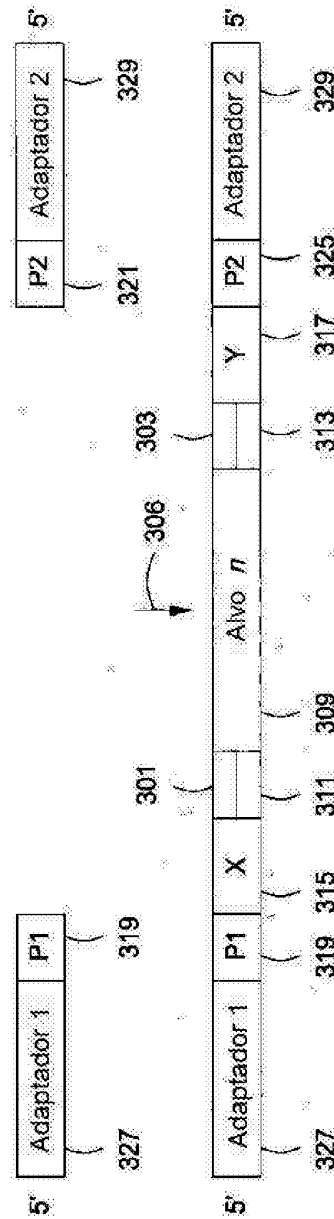
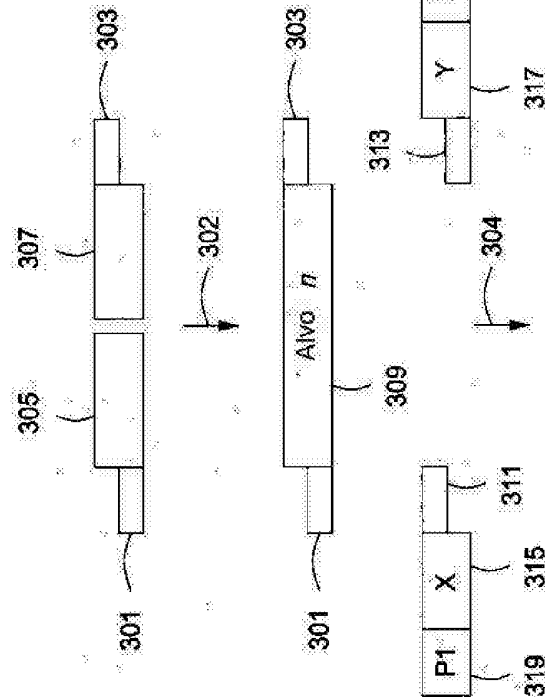


FIG. 3

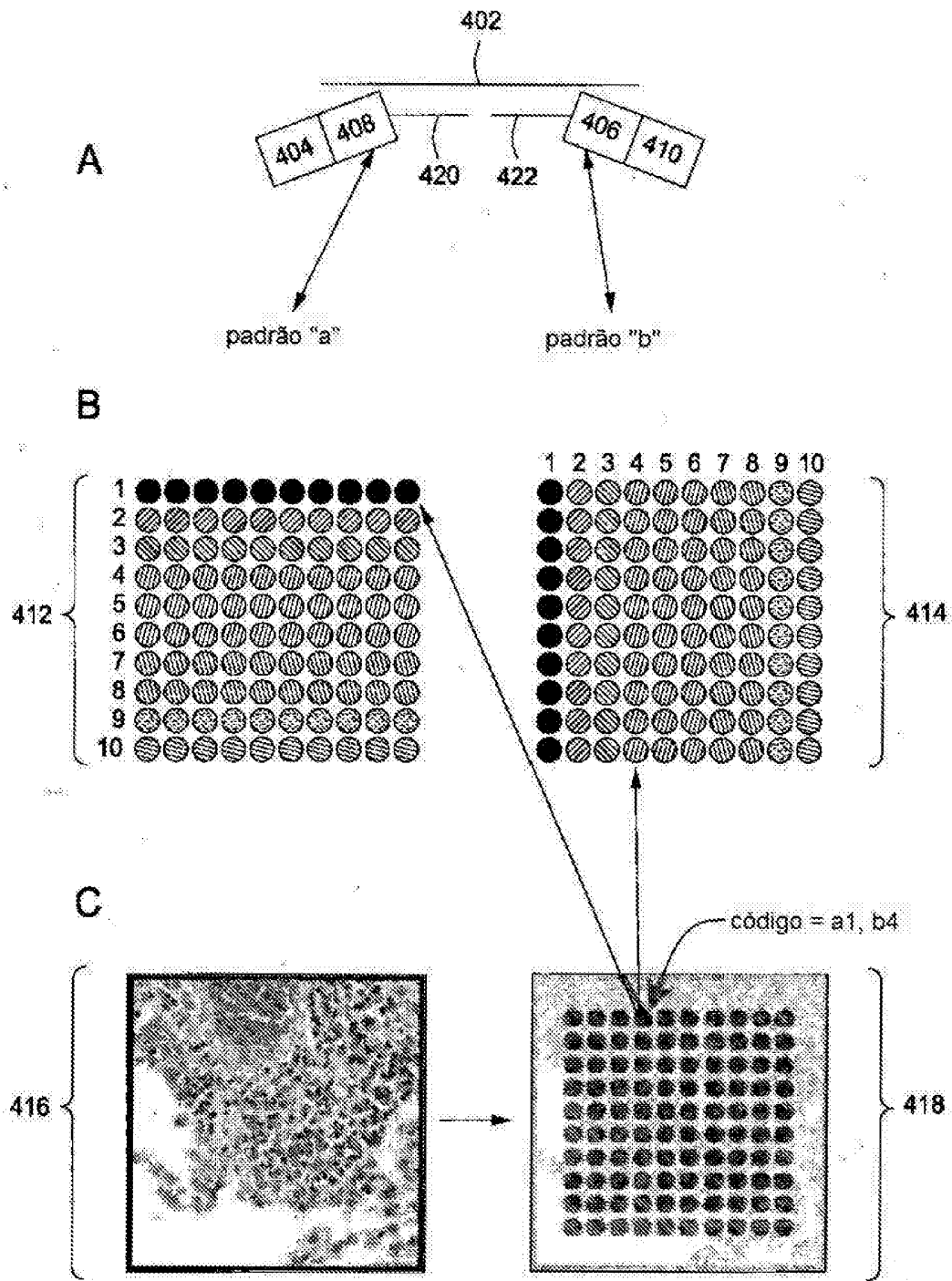


FIG. 4



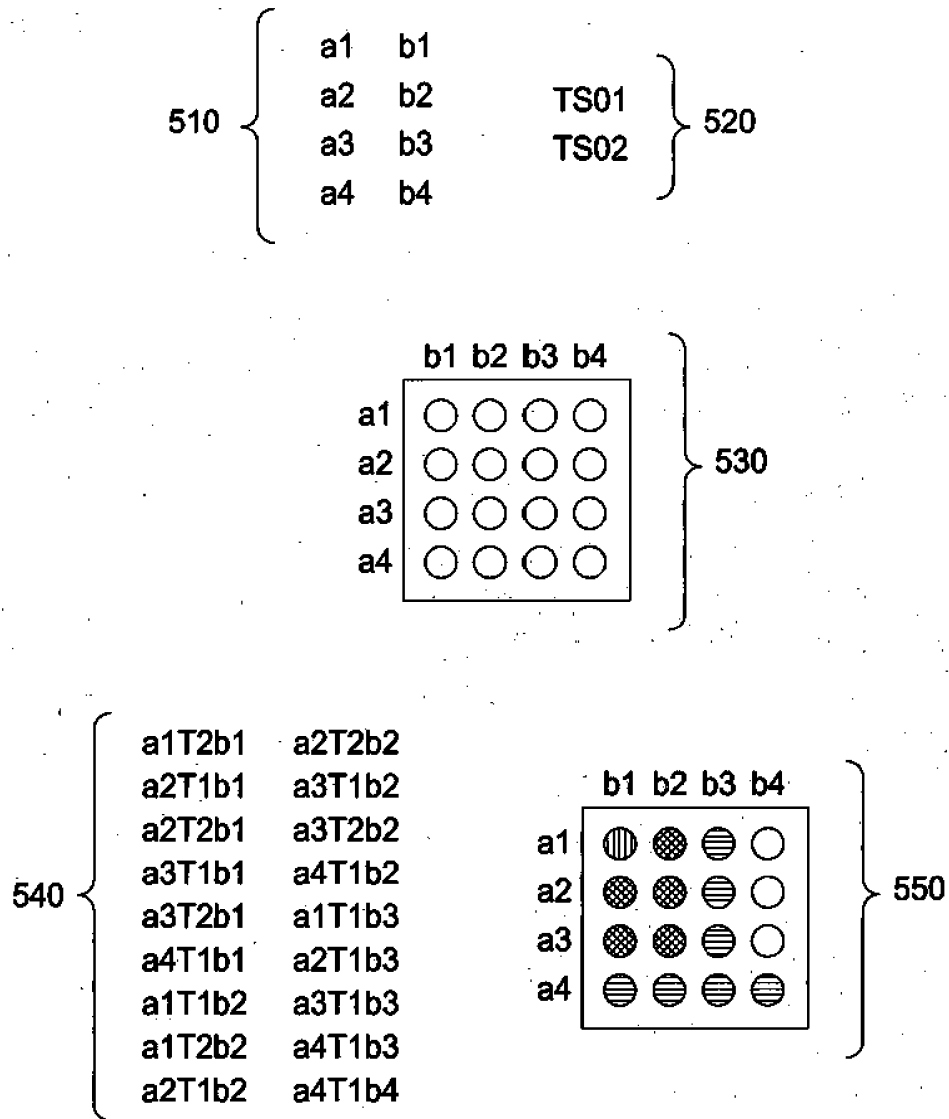


FIG. 5