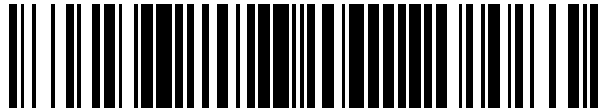


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 660 127**

51 Int. Cl.:

**G01N 33/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **08.03.2013 PCT/EP2013/054764**

87 Fecha y número de publicación internacional: **12.09.2013 WO13132075**

96 Fecha de presentación y número de la solicitud europea: **08.03.2013 E 13708798 (7)**

97 Fecha y número de publicación de la concesión europea: **13.12.2017 EP 2823315**

54 Título: **Análisis de enriquecimiento de sustrato de quinasa**

30 Prioridad:

**09.03.2012 GB 201204278**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**20.03.2018**

73 Titular/es:

**QUEEN MARY UNIVERSITY OF LONDON (100.0%)  
Mile End Road  
London Greater London E1 4NS, GB**

72 Inventor/es:

**RODRIGUEZ CUTILLAS, PEDRO**

74 Agente/Representante:

**ISERN JARA, Jorge**

ES 2 660 127 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

**DESCRIPCIÓN**

Análisis de enriquecimiento de sustrato de quinasa

5 CAMPO DE LA INVENCION

La presente invención se refiere a un método para cuantificar la actividad de una enzima modificadora de proteínas y encuentra un uso particular en la cuantificación de la actividad de una proteína quinasa.

10 ANTECEDENTES DE LA INVENCION

15 Los lípidos y las proteínas quinasas median procesos de señalización celular que son importantes para la biología normal y de enfermedad. La fosfoproteómica a gran escala, ahora rutinaria en muchos laboratorios de espectrometría de masas (MS), debería permitir la cuantificación de la señalización sin una preconcepción de las rutas dentro de la red que puedan estar activas. Ahora se pueden medir varios miles de sitios de fosforilación con alta precisión mediante el uso de técnicas cuantitativas basadas en MS (Thingholm y col., Proteomics 9, 1451 (marzo de 2009)).

20 Dado que, por definición, cada sitio de fosforilación es el resultado de una actividad de quinasa (con la oposición de una actividad de fosfatasa), en teoría debería ser posible utilizar datos de fosfoproteómica para obtener una estimación de actividad para cada quinasa expresada en el sistema bajo investigación (Cutillas & Jorgensen, Biochem J 434 (marzo de 2011)). Esto implicaría medir sustratos de quinasa conocidos (es decir, sitios de fosforilación específicos) que podrían tomarse entonces como marcadores de actividades de tales quinasas. Sin embargo, el uso de datos de fosfoproteómica para inferir las actividades de las quinasas no es sencillo. Las bases de datos de las relaciones sustrato-quinasa están disponibles públicamente y, aunque no son exhaustivas, un subconjunto de los sitios cuantificables por la fosfoproteómica a gran escala está representado en estas bases de datos. El desafío al usar esta información es que varias quinasas diferentes pueden fosforilar los mismos sustratos y las proteínas fosforiladas en un tipo de célula pueden no expresarse o ser sustratos pobres en otras. Además, la naturaleza dinámica de la fosforilación proteica significa que esta modificación puede cambiar rápidamente durante el curso de un experimento y las variables difíciles de controlar, como el reloj circadiano, la confluencia celular y el esfuerzo cortante introducido como resultado del manejo de cultivos celulares, pueden afectar la actividad de proteína quinasa, contribuyendo así a datos fosfoproteómicos con ruido de fondo. Por lo tanto, debido a los efectos estocásticos, un experimento de fosfoproteómica puede mostrar niveles inconsistentes de fosforilación de los marcadores de sustrato conocidos de una actividad de quinasa dada.

35 Por lo tanto, existe una necesidad en la técnica de un método para deducir de manera fiable la actividad de proteína quinasa con base en fosfoproteómica basada en MS.

40 RESUMEN DE LA INVENCION

Los presentes inventores han identificado por primera vez una forma de analizar datos obtenidos a partir de experimentos de fosfoproteómica basados en MS con el fin de inferir la actividad de enzimas modificadoras de proteínas, por ejemplo, la actividad de proteínas quinasas.

45 Los aspectos de la invención para los que se busca protección son como se definen en las reivindicaciones.

De acuerdo con esto, en un primer aspecto, la presente invención proporciona un método *in vitro* para determinar la activación diferencial de una enzima en una primera muestra biológica con respecto a una segunda muestra biológica, en donde la enzima cataliza una reacción que implica la adición de un grupo funcional a una proteína o péptido, comprendiendo el método:

50 (1) agrupar los péptidos modificados endógenos de la primera muestra biológica y los péptidos modificados endógenos de la segunda muestra biológica en un único grupo de acuerdo con uno de los siguientes parámetros:

- 55 o péptidos modificados endógenos que tienen un sitio de modificación que es modificado por la misma enzima; o
- o péptidos modificados endógenos que tienen un sitio de modificación que es parte de la misma secuencia específica de aminoácidos que es modificada en la misma posición por la misma enzima;

60 (2) calcular el enriquecimiento de los péptidos modificados endógenos de la primera muestra biológica en comparación con los péptidos modificados endógenos de la segunda muestra biológica en el grupo, en el que se compara el aumento en abundancia o frecuencia de los péptidos modificados endógenos de la primera muestra biológica con el aumento o disminución en la abundancia o frecuencia de los péptidos modificados endógenos de la segunda muestra biológica; y

65

(3) calcular el significado estadístico de dicho enriquecimiento;

en el que un enriquecimiento estadísticamente significativo es indicativo de que la enzima se activa en la primera muestra biológica en comparación con la segunda muestra biológica, y en el que el método comprende además identificar péptidos modificados endógenos en la primera muestra biológica y la segunda muestra biológica usando espectrometría de masas (MS) antes de la etapa (1), en la que la identificación de péptidos modificados endógenos en cada una de la primera muestra biológica y la segunda muestra biológica se lleva a cabo usando un método que comprende las siguientes etapas:

(a) obtener péptidos endógenos a partir de la muestra biológica;

(b) añadir péptidos modificados de referencia en una cantidad conocida a los péptidos obtenidos en la etapa (a) para producir una mezcla de péptidos endógenos y péptidos modificados de referencia;

(c) llevar a cabo espectrometría de masas (MS) en dicha mezcla de péptidos endógenos y péptidos modificados de referencia y normalizar la señal de los péptidos modificados endógenos a la señal de los péptidos modificados de referencia para obtener datos relativos a los péptidos endógenos en la muestra biológica; y

(d) comparar los datos relacionados con los péptidos endógenos en la muestra biológica con datos en una base de datos de péptidos modificados utilizando un programa informático;

en donde la base de datos de péptidos modificados se compila por un método que comprende:

i obtener péptidos a partir de una muestra biológica;

ii enriquecer péptidos modificados a partir de los péptidos obtenidos en el paso i;

iii realizar cromatografía líquida-espectrometría de masas en tándem (LC-MS/MS) sobre los péptidos modificados enriquecidos obtenidos en la etapa ii;

iv comparar los péptidos modificados detectados en la etapa iii con una base de datos de referencia conocida para identificar los péptidos modificados; y

v compilar datos relacionados con los péptidos modificados identificados en el paso iv en una base de datos.

El enriquecimiento puede calcularse por medio de:

- contar el número de péptidos modificados endógenos en el grupo que son más abundantes en la primera muestra biológica que en la segunda muestra biológica y el número de péptidos modificados endógenos en el grupo que son menos abundantes en la primera muestra biológica que en la segunda muestra biológica; y restar el número de péptidos modificados endógenos en el grupo que son menos abundantes en la primera muestra biológica que en la segunda muestra biológica a partir del número de péptidos modificados endógenos en el grupo que son más abundantes en la primera muestra biológica que en la segunda muestra biológica muestra, o

- calcular la abundancia media de todos los péptidos modificados endógenos en el grupo de la primera muestra biológica y la abundancia media de todos los péptidos modificados endógenos en el grupo de la segunda muestra biológica; y dividir la abundancia media de todos los péptidos modificados endógenos en el grupo de la primera muestra biológica por la abundancia media de todos los péptidos modificados endógenos en el grupo de la segunda muestra biológica, o

- calcular la media de todos los cambios de repetición de los péptidos modificados endógenos en el grupo y la media de todos los cambios de repetición de los péptidos modificados endógenos en todo el experimento; y luego dividir la media de todos los cambios de repetición de los péptidos modificados endógenos en el grupo por cambio de repetición medio de todos los péptidos modificados endógenos en todo el experimento.

La importancia estadística de dicho enriquecimiento puede calcularse mediante la prueba hipergeométrica, la prueba t pareada, la prueba Z, la prueba de Kolmogorov-Smirnov, la prueba de Chi cuadrado, la prueba exacta de Fisher o la prueba de rango con signo de Wilcoxon y opcionalmente comprende adicionalmente llevar a cabo una corrección de prueba múltiple, en la que opcionalmente la corrección de prueba múltiple es el método de Tasa de Descubrimiento Falso (FDR) de Benjamini Hochberg.

La enzima puede ser una proteína quinasa.

La etapa (b) puede comprender adicionalmente péptidos modificados enriquecidos de dicha mezcla de péptidos endógenos y péptidos modificados de referencia para producir una mezcla de péptidos modificados enriquecidos y la etapa (c) comprende llevar a cabo la espectrometría de masas (MS) en dicha mezcla de péptidos modificados enriquecidos para obtener datos relacionados con los péptidos modificados endógenos en la muestra. La etapa de

enriquecimiento de péptidos modificados se puede llevar a cabo usando cromatografía, opcionalmente en la que la cromatografía se selecciona del grupo que consiste en cromatografía de afinidad de iones metálicos inmovilizados (IMAC), cromatografía en dióxido de titanio (TiO<sub>2</sub>) y cromatografía en dióxido de zirconio (ZrO<sub>2</sub>); o la etapa de enriquecimiento de péptidos modificados se puede llevar a cabo usando métodos basados en anticuerpos. Los datos relativos a los péptidos en la muestra biológica pueden comprender la relación de masa a carga (m/z), carga (z) y tiempo de retención relativo de los péptidos; y/o dicha espectrometría de masas (MS) en la etapa (c) puede ser cromatografía líquida-espectrometría de masas (LC-MS). La etapa ii puede llevarse a cabo usando cromatografía multidimensional, opcionalmente en la que la cromatografía multidimensional se lleva a cabo usando cromatografía líquida de alta resolución de intercambio catiónico fuerte (SCX-HPLC), cromatografía de afinidad de iones metálicos inmovilizados (IMAC) y cromatografía en dióxido de titanio (TiO<sub>2</sub>) o cromatografía líquida de alta resolución de intercambio de aniones (SAX-HPLC), cromatografía de afinidad de iones metálicos inmovilizados (IMAC) y cromatografía en dióxido de titanio (TiO<sub>2</sub>); o la etapa ii puede llevarse a cabo usando métodos basados en anticuerpos. El paso iv puede llevarse a cabo utilizando el motor de búsqueda MASCOT. Los datos relacionados con los péptidos modificados identificados en la etapa iv pueden seleccionarse del grupo que consiste en la identidad del péptido modificado, la relación masa a carga (m/z), la carga (z) y el tiempo de retención relativa del péptido modificado. La técnica MS puede usar etiquetas de isótopos para la cuantificación. La técnica de EM puede usar marcación metabólica o derivación química. La marcación metabólica puede comprender el uso de aminoácidos marcados con isótopos estables en cultivo (SILAC) o en el que la derivación química es iTRAQ, ICAT o TMT.

## 20 DESCRIPCIÓN DETALLADA DE LA INVENCION

La presente descripción se refiere a un método para cuantificar la actividad de enzimas modificadoras de proteínas tales como proteínas quinasas en una muestra. El método se basa en el análisis de péptidos modificados, por ejemplo, péptidos fosforilados, que se identifican usando técnicas basadas en MS.

Como se describe en el presente documento, el método es un método para cuantificar la actividad de una enzima modificadora de proteínas en una muestra. La mayoría de las proteínas se modifican de alguna manera mediante la adición de grupos funcionales y tales modificaciones se efectúan mediante enzimas modificadoras de proteínas. Las modificaciones proteicas que pueden detectarse mediante espectrometría de masas incluyen fosforilación, glicosilación, acetilación, metilación y lipidación. Estas modificaciones proteicas tienen varios roles biológicos en la célula. Por "enzima modificadora de proteínas" se entiende una enzima que cataliza una reacción que implica la adición de un grupo funcional a una proteína o péptido.

El método descrito en el presente documento puede aplicarse a la cuantificación de la actividad de cualquier enzima modificadora de proteínas cuya actividad puede detectarse usando métodos basados en MS. Tales enzimas incluyen proteínas quinasas, proteínas glicosiltransferasas, proteínas acetiltransferasas, proteínas metiltransferasas y proteínas palmitoiltransferasas. La actividad de estas enzimas da como resultado la fosforilación, la acetilación, la glicosilación, la metilación y la lipidación de sustratos proteicos o peptídicos, respectivamente. Todas estas modificaciones proteicas pueden detectarse mediante espectrometría de masas.

El método puede ser un método para cuantificar la actividad de una proteína quinasa. El método puede basarse en el análisis de péptidos fosforilados. Los péptidos fosforilados contienen uno o más aminoácidos que están fosforilados (es decir, un fosfato).

El grupo (PO<sub>4</sub>) se ha agregado a ese aminoácido. Dichos aminoácidos fosforilados se denominan en el presente documento "sitios de fosforilación". En relación con esto, el término "fosfoproteína" se usa en el presente documento para referirse a una proteína fosforilada y el término "fosfopéptido" se usa en el presente documento para referirse a un péptido fosforilado.

Las proteínas quinasas humanas se pueden dividir en varios grupos que incluyen AGC quinasas, por ejemplo, proteína quinasa A (PKA), proteína quinasa B (PKB) (también conocida como Akt), proteína quinasa C (PKC) y proteína quinasa G (PKG); tirosina quinasas; tirosina-quinasa como quinasas; proteínas quinasas dependientes de calcio/calmodulina; el grupo caseína quinasa 1; Grupo CMGC, por ejemplo, quinasas CDK, MAPK, GSK3 y CLK; y STE, los homólogos de levadura Estéril 7, Estéril 11 y Estéril 20 quinasas.

El método descrito en el presente documento es un método para cuantificar la actividad de una enzima modificadora de proteínas en una muestra e implica hacer una comparación entre péptidos modificados de una primera muestra y péptidos modificados de una segunda muestra. La primera y la segunda muestras usadas en los métodos descritos en el presente documento pueden ser cualquier muestra que contenga péptidos. La muestra es una muestra biológica y, por lo tanto, puede ser cualquier tipo de muestra obtenida a partir de una fuente biológica, por ejemplo, una muestra obtenida de un ser humano, animal, planta o bacteria. La invención abarca por lo tanto el uso de muestras obtenidas de fuentes humanas y no humanas.

Las muestras usadas en los métodos de la presente invención pueden ser de cualquier especie de interés. Típicamente, las muestras son de un humano o animal. El animal es típicamente un mamífero, por ejemplo, un roedor como un ratón, rata o conejillo de indias, o un ungulado tal como una vaca, oveja o cabra. El animal es

alternativamente un ave, como un pollo, un pez, como un pez cebra, un nematodo, como el gusano *Caenorhabditis elegans*, o un insecto, como la mosca de la fruta *Drosophila melanogaster*. Las muestras usadas en los métodos de la invención también pueden ser de otras formas de vida tales como bacterias y levaduras. Las muestras usadas en los métodos de la invención son típicamente muestras de una especie experimentalmente importante de bacteria tal como *Escherichia coli*, *Salmonella enterica*, *Streptococcus pneumoniae* o *Staphylococcus aureus*, o de levadura tal como la levadura de panadería *Saccharomyces cerevisiae* o la levadura de fisión *Schizosaccharomyces pombe*. Las muestras usadas en los métodos de la invención pueden ser alternativamente de una planta u hongo o un virus.

Típicamente, la muestra biológica se deriva de un ser humano, y puede ser, por ejemplo, una muestra de un fluido corporal tal como orina o sangre u otro tejido. Típicamente, la muestra biológica es una línea celular o un tejido, típicamente un tejido primario. Por ejemplo, la muestra puede ser un tejido de un humano o animal. El humano o animal puede estar sano o enfermo. Alternativamente, la muestra puede ser una línea celular derivada de células humanas o animales sanas o enfermas.

El método de la invención es un método *in vitro* y, por lo tanto, no comprende la etapa de obtener una muestra de un organismo tal como un animal.

En una primera etapa (1), el método descrito en la presente comprende agrupar péptidos modificados de una primera muestra y péptidos modificados de una segunda muestra en un único grupo de acuerdo con uno de los siguientes parámetros:

(a) péptidos modificados que tienen un sitio de modificación que es modificado por la misma enzima; o

(b) péptidos modificados que tienen un sitio de modificación que es parte del mismo motivo de modificación.

"Agrupación" significa que los péptidos modificados tales como los péptidos fosforilados se colocan en un grupo o conjunto. Los péptidos modificados de una primera muestra y los péptidos modificados de una segunda muestra se colocan en un solo grupo basándose en uno de los parámetros expuestos anteriormente. En los ejemplos del presente documento, el grupo también se denomina "grupo de sustrato". De forma similar, un péptido modificado también se denomina aquí "sustrato".

Los péptidos modificados pueden colocarse en un grupo sobre la base de que tienen un sitio de modificación que es modificado por la misma enzima. Por lo tanto, cada péptido modificado en el grupo tiene al menos un sitio de modificación que es modificado por la misma enzima. Por ejemplo, los péptidos fosforilados se colocan en un grupo sobre la base de que tienen un sitio de fosforilación fosforilado por la misma quinasa, donde se sabe que un sitio de fosforilación dentro de cada uno de los péptidos fosforilados en el grupo está fosforilado por la misma quinasa específica. De forma similar, si la enzima es una acetilasa, los péptidos modificados se colocan en un grupo sobre la base de que tienen un sitio de acetilación acetilado por la misma acetilasa, en el que se sabe que un sitio de acetilación dentro de cada uno de los péptidos acetilados del grupo ser acetilado por la misma acetilasa específica.

La información sobre las relaciones quinasa-sustrato y por lo tanto sobre los sitios de fosforilación que son fosforilados por una quinasa particular puede obtenerse de bases de datos disponibles públicamente, por ejemplo, PhosphoSite (Hornbeck et al., *Proteomics* 4, 1551 (Junio, 2004)) y PhosphoElm (Dinkel et al., *Nucleic Acids Res* 39, D261 (enero de 2011)). De forma similar, la información sobre otros sitios de modificación puede obtenerse de bases de datos disponibles públicamente y de trabajos de investigación individuales obtenidos de la literatura.

Los péptidos modificados pueden colocarse en un grupo sobre la base de que tienen un sitio de modificación que es parte del mismo motivo de modificación. Por ejemplo, los péptidos fosforilados se colocan en un grupo sobre la base de que tienen un sitio de fosforilación que es parte del mismo motivo de fosforilación. Por "motivo de modificación" se entiende una secuencia específica de aminoácidos que se modifica en la misma posición por la misma enzima. Por ejemplo, un "motivo de fosforilación" es una secuencia específica de aminoácidos que se fosforila en la misma posición por la misma proteína quinasa o por quinasas relacionadas funcionalmente, en donde un sitio de fosforilación dentro de cada uno de los péptidos fosforilados en el grupo es parte de un motivo de fosforilación predefinido.

La información sobre motivos de modificación tales como motivos de fosforilación se pueden obtener, por ejemplo, a partir de la bibliografía o a partir de un análisis de un conjunto de datos usando un programa informático tal como Motif-X (Schwartz and Gygi, *Nat Biotechnol* 23, 1391 (Nov, 2005)).

Cuando la enzima modificadora de proteínas es una proteína quinasa, la etapa (1) del método descrito en la presente comprende agrupar péptidos fosforilados de una primera muestra y péptidos fosforilados de una segunda muestra en un único grupo de acuerdo con uno de los siguientes parámetros:

- péptidos fosforilados que tienen un sitio de fosforilación que es fosforilado por la misma proteína quinasa; o

- péptidos fosforilados que tienen un sitio de fosforilación que es parte del mismo motivo de fosforilación;

Antes de agrupar los péptidos modificados, los péptidos modificados se pueden seleccionar para usar en el método descrito en este documento de acuerdo con el significado estadístico de la aparición de la modificación.

5 El método de la divulgación implica agrupar péptidos modificados de una primera muestra y de una segunda muestra en un solo grupo. Por lo tanto, el método encuentra uso en la comparación de la actividad de una enzima modificadora de proteína tal como una proteína quinasa entre al menos dos muestras, por ejemplo, la comparación de dos muestras que son de diferentes fuentes o que han sido tratadas con diferentes sustancias de prueba.

10 Alternativamente, el método se puede usar para comparar una muestra de prueba y una muestra de control. Una de las dos muestras puede ser una muestra de control. La primera y la segunda muestra pueden ser de la misma fuente, pero la primera o la segunda muestra se tratan con una sustancia de prueba, mientras que la otra muestra no se trata de esta manera.

15 El método puede usarse para comparar la actividad de una enzima modificadora de proteínas tal como una proteína quinasa entre más de dos muestras, por ejemplo 3, 4, 5, 6, 7, 8, 9, 10 o incluso más muestras. La etapa (1) del método puede implicar la agrupación de péptidos modificados de todas las muestras en un solo grupo de acuerdo con uno de los parámetros como se describe en este documento. Los pasos (2) y (3) del método se llevan a cabo luego para comparar el enriquecimiento de los péptidos modificados de una de las muestras en comparación con los péptidos modificados de otra muestra en el grupo. Por ejemplo, cuando hay 3 muestras, se puede comparar el enriquecimiento de los péptidos modificados en la primera muestra versus la segunda, en la segunda, versus la  
20 tercera muestra y en la primera versus la tercera muestra.

25 La propia muestra o el organismo del que se obtiene la muestra se pueden tratar con una sustancia de prueba antes de llevar a cabo el método de la invención. Por lo tanto, una línea celular o un organismo del que se obtiene un tejido se puede tratar con una sustancia de prueba antes de llevar a cabo el método de la invención. La sustancia de prueba es típicamente un producto químico o fármaco exógeno, como inhibidores de moléculas pequeñas, ARNi, péptidos terapéuticos y anticuerpos. Esto permite la investigación de los efectos de la sustancia de ensayo sobre la actividad de una enzima modificadora de proteínas y la comparación de tales efectos en diferentes muestras.

30 Por ejemplo, una línea celular puede tratarse con agonistas de vías y/o inhibidores de quinasas antes de llevar a cabo el método de la invención. Los inhibidores de quinasas típicos incluyen inhibidores de src y fosfoinosítido 3-quinasa (PI3K), tales como PP2 y PI-103. Otros inhibidores de PI3K incluyen wortmanina. Al menos 80 inhibidores de quinasas están en diferentes etapas de desarrollo clínico (Zhang, J. ; et al. Nat Rev Cancer 2009, 9, (1), 28-39). La técnica también es útil para investigar otros tipos de inhibidores que se sospecha que tienen un efecto en las vías  
35 de la quinasa, como los inhibidores de HSP90, los inhibidores de fosfatasa y los fármacos con anticuerpos.

40 Un "péptido" como se define en este documento es una secuencia de aminoácidos corta e incluye oligopéptidos y polipéptidos. Típicamente, tales péptidos tienen entre aproximadamente 5 y 30 aminoácidos de longitud, por ejemplo de 6 o 7 a 25, 26 o 27 aminoácidos, de 8, 9 o 10 a 20 aminoácidos, de 11 o 12 a 18 aminoácidos o de 14 a 16 aminoácidos, por ejemplo 15 aminoácidos. Sin embargo, también se pueden usar péptidos más cortos y más largos, tales como entre aproximadamente 2 y aproximadamente 50, por ejemplo de aproximadamente 3 a aproximadamente 35 o 40 o de aproximadamente 4 a aproximadamente 45 aminoácidos. Típicamente, el péptido es adecuado para el análisis de espectrometría de masas, es decir, la longitud del péptido es tal que el péptido es adecuado para el análisis de espectrometría de masas. La longitud del péptido que puede analizarse está limitada  
45 por la capacidad del espectrómetro de masas para secuenciar tales péptidos largos. En ciertos casos, se pueden analizar polipéptidos de hasta 300 aminoácidos, por ejemplo de 50 a 250 aminoácidos, de 100 a 200 aminoácidos o de 150 a 175 aminoácidos.

50 Como se describe en este documento, el método se basa en el análisis de péptidos modificados identificados usando técnicas basadas en MS. Por consiguiente, los péptidos modificados de una primera muestra y de una segunda muestra que se agrupan en la primera etapa del método se identifican y/o cuantifican típicamente usando técnicas basadas en MS. El método de la divulgación por lo tanto incluye una etapa de identificación de péptidos modificados en una primera muestra y/o una segunda muestra usando espectrometría de masas (MS), antes del paso (1) de agrupar los péptidos modificados de una primera muestra y de un segundo muestra. También se describe aquí un método para cuantificar la actividad de una enzima modificadora de proteínas en una muestra, que comprende identificar péptidos modificados en una primera muestra y una segunda muestra usando espectrometría de masas (MS) y:

60 (1) agrupar los péptidos modificados de una primera muestra y los péptidos modificados de una segunda muestra en un solo grupo de acuerdo con uno de los siguientes parámetros:

- o péptidos modificados que tienen un sitio de modificación que es modificado por la misma enzima modificadora de proteína; o
- 65 o (b) péptidos modificados que tienen un sitio de modificación que es parte del mismo motivo de modificación;

(2) calcular el enriquecimiento de los péptidos modificados de la primera muestra en comparación con los péptidos modificados de la segunda muestra en el grupo; y

(3) calcular el significado estadístico de dicho enriquecimiento;

5 en el que un enriquecimiento estadísticamente significativo es indicativo de que se está activando una enzima modificadora de proteína en la primera muestra en comparación con la segunda muestra.

10 La identificación y cuantificación de péptidos modificados se puede llevar a cabo usando cualquier método adecuado. Típicamente, la cuantificación se puede llevar a cabo por cualquier método que involucre espectrometría de masas (MS), tal como cromatografía líquida-espectrometría de masas (LC-MS). El LC-MS o el LC-MS/MS son típicamente MS sin etiquetas, pero las técnicas que usan el etiquetado isotópico como base para la cuantificación también pueden usarse como base para el análisis.

15 En los métodos de la presente descripción, la cuantificación de una modificación de proteína tal como fosforilación se lleva a cabo típicamente usando la técnica TIQUAS (cuantificación dirigida y en profundidad de señalización), como se describe en el documento WO 2010/119261 (Solicitud de Patente Internacional No. PCT/GB2010/000770). Esta técnica permite la cuantificación sensible, rápida y completa de péptidos modificados. El método puede, en un simple ensayo, medir simultáneamente las cantidades de miles de sitios de fosforilación en proteínas. Como se establece en el documento WO 2010/119261, la técnica TIQUAS también puede usarse para cuantificar péptidos modificados distintos de los péptidos fosforilados. De hecho, la técnica TIQUAS se puede usar para cuantificar péptidos que contienen cualquier modificación que pueda detectarse mediante espectrometría de masas.

20 En el método de la divulgación, la etapa de identificar péptidos modificados utilizando espectrometría de masas (MS) antes de la etapa (1) puede llevarse a cabo usando un método que comprende las siguientes etapas:

(a) obtención de péptidos a partir de una muestra;

30 (b) añadir péptidos modificados de referencia a los péptidos obtenidos en la etapa (a) para producir una mezcla de péptidos y péptidos modificados de referencia;

(c) realizar una espectrometría de masas (MS) sobre dicha mezcla de péptidos y péptidos modificados de referencia para obtener datos relativos a los péptidos en la muestra; y

35 (d) comparar los datos relacionados con los péptidos en la muestra con los datos en una base de datos de péptidos modificados utilizando un programa informático;

en el que la base de datos de péptidos modificados se compila mediante un método que comprende:

40 i obtener péptidos a partir de una muestra;

ii enriquecer péptidos modificados a partir de los péptidos obtenidos en el paso i;

45 iii realizar cromatografía líquida-espectrometría de masas en tándem (LC-MS/MS) sobre los péptidos modificados enriquecidos obtenidos en la etapa ii;

iv comparar los péptidos modificados detectados en la etapa iii con una base de datos de referencia conocida para identificar los péptidos modificados; y

50 v compilar datos relacionados con los péptidos modificados identificados en el paso iv en una base de datos.

Cuando la enzima es una proteína quinasa y la modificación es fosforilación, la etapa de identificación de péptidos modificados usando espectrometría de masas (MS) antes de la etapa (1) puede llevarse a cabo utilizando un método que comprende las siguientes etapas:

55 (a) obtención de péptidos fosforilados a partir de una muestra;

(b) añadir péptidos fosforilados de referencia a los péptidos obtenidos en la etapa (a) para producir una mezcla de péptidos y péptidos fosforilados de referencia;

60 (c) llevar a cabo una espectrometría de masas (MS) sobre dicha mezcla de péptidos y péptidos fosforilados de referencia para obtener datos relativos a los péptidos en la muestra; y

65 (d) comparar los datos relacionados con los péptidos en la muestra con datos en una base de datos de péptidos fosforilados usando un programa informático;

en el que la base de datos de péptidos fosforilados se compila mediante un método que comprende:

i obtener péptidos a partir de una muestra;

5 ii enriquecer péptidos fosforilados a partir de los péptidos obtenidos en el paso i;

iii realizar cromatografía líquida-espectrometría de masas en tándem (LC-MS/MS) sobre los péptidos fosforilados enriquecidos obtenidos en la etapa ii;

10 iv comparar los péptidos fosforilados detectados en la etapa iii con una base de datos de referencia conocida para identificar los péptidos fosforilados; y

v compilar datos relacionados con los péptidos fosforilados identificados en el paso iv en una base de datos.

15 En relación con esto, la palabra "péptido" se usa de forma intercambiable con la palabra "polipéptido".

La etapa (a) implica la obtención de péptidos a partir de una muestra. Los péptidos se pueden obtener de la muestra usando cualquier método adecuado conocido en la técnica. El paso (a) del método puede comprender:

20 (a1) someter a lisis en la muestra;

(a2) extraer las proteínas de las células lisadas obtenidas en la etapa (a1); y

(a3) escindir dichas proteínas en péptidos.

25 En la etapa (a1), las células en la muestra se lisan o se abren. Las células se pueden lisar usando cualquier medio adecuado conocido en la técnica, por ejemplo usando métodos físicos tales como lisis mecánica (por ejemplo usando un mezclador Waring), homogeneización de líquidos, sonicación o lisis manual (por ejemplo, usando una mano y un mortero) o detergente basados en métodos como CHAPS o Triton-X. Típicamente, las células se lisan usando un regulador de desnaturalización tal como un regulador basado en urea.

30 En la etapa (a2), las proteínas se extraen de las células lisadas obtenidas en la etapa (a1). En otras palabras, las proteínas están separadas de los otros componentes de las células lisadas.

35 En el paso (a3), las proteínas de las células lisadas se escinden en péptidos. En otras palabras, las proteínas se descomponen en péptidos más cortos. La descomposición de proteínas también se conoce comúnmente como digestión. La escisión de proteínas se puede llevar a cabo en la presente descripción usando cualquier agente adecuado conocido en la técnica.

40 La escisión o digestión de proteínas se lleva a cabo típicamente usando una proteasa. Se puede usar cualquier proteasa adecuada. La proteasa es típicamente tripsina, quimotripsina, Arg-C, pepsina, V8, Lys-C, Asp-C y/o AspN. Alternativamente, las proteínas se pueden escindir químicamente, por ejemplo usando hidroxilamina, ácido fórmico, bromuro de cianógeno, BNPS-skatol, ácido 2-nitro-5-tiocianobenzoico (NTCB) o cualquier otro agente adecuado.

45 En la etapa (b), se añaden péptidos modificados de referencia (típicamente péptidos fosforilados de referencia) a los péptidos obtenidos en la etapa (a) para producir una mezcla de péptidos y péptidos modificados de referencia (típicamente péptidos fosforilados de referencia). La etapa (b) da como resultado una mezcla de péptidos (incluidos los modificados, típicamente fosforilados) por muestra. Los péptidos modificados de referencia (típicamente péptidos fosforilados de referencia) también se denominan aquí "estándares internos" (IS). Típicamente, se añaden de 5 a 10, por ejemplo de 6 a 9 o de 7 a 8, péptidos modificados de referencia (típicamente péptidos fosforilados de referencia).

50 Los péptidos modificados de referencia son típicamente péptidos fosforilados de referencia y típicamente se derivan de una proteína de referencia de naturaleza y concentración definidas, a menudo denominada proteína de patrón interno (IS). Los IS pueden ser proteínas disponibles comercialmente, por ejemplo caseína. Alternativamente, los IS se sintetizan específicamente para usar en el método. Los péptidos fosforilados de referencia típicamente se sintetizan con la misma secuencia que algunos de los péptidos fosforilados que se desea cuantificar pero que están enriquecidos en isótopos pesados estables de carbono y nitrógeno. Los péptidos típicamente se sintetizan usando química en fase sólida en la que se agrega un aminoácido a la vez para formar una cadena de aminoácidos o polipéptido. Típicamente, dichos péptidos están enriquecidos en <sup>13</sup>C y <sup>15</sup>N que sustituyen a los <sup>12</sup>C y <sup>14</sup>N comunes.

60 Este enriquecimiento da como resultado que los péptidos fosforilados de referencia sean aproximadamente de 6 a 10 daltons más pesados que los péptidos fosforilados endógenos con la misma secuencia, de modo que puedan distinguirse usando un espectrómetro de masas.



Cuando la enzima modificadora de proteínas es una proteína acetiltransferasa y los péptidos acetilados están siendo cuantificados, los péptidos modificados de referencia son péptidos acetilados de referencia. Dichos péptidos acetilados de referencia son típicamente péptidos sintéticos que contienen aminoácidos acetilados.

5 Los péptidos modificados de referencia (típicamente péptidos fosforilados de referencia) se añaden típicamente a una cantidad conocida en cada una de las muestras por comparar. Las señales de los péptidos modificados endógenos (típicamente péptidos fosforilados) se normalizan a la señal de los péptidos modificados de referencia (típicamente péptidos fosforilados de referencia) en el análisis corriente abajo.

10 La etapa (b) de esta descripción puede comprender adicionalmente péptidos modificados enriquecidos (típicamente péptidos fosforilados) a partir de la mezcla de péptidos y péptidos modificados de referencia (típicamente péptidos fosforilados de referencia) obtenidos en la etapa (b) para producir una mezcla de péptidos modificados enriquecidos (típicamente péptidos fosforilados). Esta etapa adicional da como resultado una única mezcla de péptidos modificados enriquecidos (típicamente péptidos fosforilados) por muestra. La etapa (c) comprende así llevar a cabo la espectrometría de masas (MS) en la mezcla de péptidos modificados enriquecidos (típicamente péptidos fosforilados) para obtener datos relacionados con los péptidos en la muestra. La etapa (b) típicamente da como resultado una mezcla de péptidos modificados enriquecidos (típicamente péptidos fosforilados).

20 La etapa de enriquecimiento de péptidos modificados (típicamente péptidos fosforilados) se lleva a cabo típicamente usando cromatografía. En una realización, la cromatografía es cromatografía de afinidad de iones metálicos inmovilizados (IMAC), cromatografía en dióxido de titanio (TiO<sub>2</sub>), y/o cromatografía en dióxido de zirconio (ZrO<sub>2</sub>). Típicamente, la cromatografía es IMAC y cromatografía en TiO<sub>2</sub>.

25 Alternativamente, la etapa de enriquecimiento de péptidos modificados (típicamente péptidos fosforilados) se lleva a cabo usando métodos basados en anticuerpos.

Cuando la enzima modificadora de proteínas es una proteína quinasa y los péptidos que se cuantifican son péptidos fosforilados, los anticuerpos con afinidad por aminoácidos fosforilados tales como tirosina, treonina, serina o histidina están unidos (inmovilizados) a una matriz sólida. Los péptidos fosforilados están enriquecidos por la capacidad de estos anticuerpos para unir específicamente péptidos fosforilados. Los péptidos no fosforilados se eliminan mediante lavado mientras que los péptidos fosforilados se retienen en las matrices revestidas de anticuerpo. La elución de péptidos fosforilados a partir del anticuerpo inmovilizado se lleva a cabo típicamente usando disolventes de pH bajo o mediante cualquier otro método adecuado que desnaturalice la interacción entre el anticuerpo y los péptidos fosforilados.

35 Cuando la enzima modificadora de proteínas es una proteína acetiltransferasa y los péptidos que se cuantifican son péptidos acetilados, los péptidos acetilados se enriquecen mediante el uso de anticuerpos específicos contra residuos de aminoácidos acetilados. Tales anticuerpos se unen a una matriz sólida y luego se enriquecen por la capacidad de los anticuerpos para unir específicamente residuos de aminoácidos acetilados. Los péptidos no acetilados se eliminan después por lavado mientras que los péptidos acetilados se retienen en el anticuerpo inmovilizado.

45 En la etapa (c) de esta descripción, la espectrometría de masas (MS) puede llevarse a cabo en la mezcla de péptidos y péptidos modificados de referencia (típicamente péptidos fosforilados de referencia) obtenidos en la etapa (b) para obtener datos relacionados con los péptidos en la muestra. Normalmente, estos datos tienen la forma de un archivo de datos MS para la muestra. Cuando la etapa (b) de esta realización comprende además enriquecer péptidos modificados (típicamente péptidos fosforilados) a partir de la mezcla de péptidos y péptidos modificados de referencia (típicamente péptidos fosforilados de referencia) obtenidos en la etapa (b) para producir una mezcla de péptidos modificados enriquecidos (típicamente péptidos fosforilados), la etapa (c) puede comprender realizar espectrometría de masas (MS) en dicha mezcla de péptidos modificados enriquecidos (típicamente péptidos fosforilados) para obtener datos relacionados con los péptidos en la muestra, típicamente un archivo de datos MS para la muestra. Típicamente, la espectrometría de masas es cromatografía líquida-espectrometría de masas (LC-MS). El paso (c) por lo tanto, normalmente resulta en un archivo de datos LC-MS (uno de cada muestra).

55 Los datos relativos a los péptidos en la muestra comprenden típicamente la relación de masa a carga (m/z), carga (z) y/o tiempo de retención relativa de los péptidos.

60 En la etapa (d) de esta descripción, los datos relacionados con los péptidos en la muestra (típicamente en la forma de un archivo de datos MS y más típicamente un archivo de datos LC-MS) se pueden comparar con datos en una base de datos de péptidos modificados (típicamente péptidos fosforilados) usando un programa de ordenador. Por ejemplo, la relación masa a carga (m/z), la carga (z) y el tiempo de retención relativo de los péptidos en la muestra se comparan con la relación masa a carga (m/z), carga (z) y tiempo de retención relativo de los péptidos modificados (típicamente péptidos fosforilados) en la base de datos. Esto permite la identificación y cuantificación de cada péptido modificado (típicamente péptido fosforilado) en la muestra usando la base de datos de péptidos modificados (típicamente péptidos fosforilados).

65

Típicamente, el programa informático es el programa denominado PESCAL (Cutillas, P.R., Vanhaesebroeck, B. Mol Cell Proteomics 6(9), 1560-73, 2007). PESCAL construye cromatogramas de iones extraídos (XIC, es decir, un perfil de elución) para cada uno de los péptidos modificados (típicamente péptidos fosforilados) presentes en la base de datos en todas las muestras que se van a comparar. Esto se hace centrando el XIC en la  $m/z$  y el tiempo de retención del péptido previamente identificado para ser modificado (típicamente fosforilado) (es decir, presente en la base de datos construida en el primer paso del procedimiento). PESCAL también considera la carga del péptido para ayudar en la correcta asignación de identidad. El programa también calcula la altura del pico y el área bajo la curva de cada XIC. Los datos se normalizan dividiendo la lectura de intensidad (áreas de pico o alturas) de cada péptido modificado (típicamente péptido fosforilado) que se analiza por los péptidos modificados de referencia (típicamente péptidos fosforilados de referencia).

La base de datos de péptidos modificados se compila por un método que comprende los siguientes pasos:

i obtener péptidos a partir de una muestra;

ii enriquecer péptidos modificados a partir de los péptidos obtenidos en el paso i;

iii realizar cromatografía líquida-espectrometría de masas en tándem (LC-MS/MS) sobre los péptidos modificados enriquecidos obtenidos en la etapa ii;

iv comparar los péptidos modificados detectados en la etapa iii con una base de datos de referencia conocida para identificar los péptidos modificados; y

v compilar datos relacionados con los péptidos modificados identificados en el paso iv en una base de datos.

La etapa i implica la obtención de péptidos a partir de una muestra. Los péptidos se pueden obtener de la muestra usando cualquier método adecuado conocido en la técnica y como se describe en este documento.

La muestra es una muestra biológica y, por lo tanto, puede ser cualquier tipo de muestra obtenida a partir de una fuente biológica, como se describió anteriormente. Típicamente, la muestra es una línea celular o un tejido.

Cuando la muestra utilizada en la etapa i es una línea celular, la muestra puede tratarse con un inhibidor antes de llevar a cabo la etapa i. El inhibidor puede ser cualquier tipo adecuado de inhibidor. Típicamente, cuando se cuantifican los péptidos fosforilados, el inhibidor es un inhibidor de fosfatasa. El tratamiento con inhibidores de la fosfatasa aumenta la estequiometría de la fosforilación y da como resultado un mayor número de péptidos fosforilados que se pueden incluir en la base de datos. Además, los inhibidores de la metiltransferasa o de la acetilhidrolasa pueden usarse cuando el objetivo es cuantificar los péptidos metilados y acetilados, respectivamente.

El paso i de este método comprende:

(i1) someter a lisis células en una muestra;

(i2) extraer las proteínas de las células lisadas obtenidas en el paso (i1); y

(i3) escindir dichas proteínas en péptidos.

Estos son como se describió anteriormente. Sin embargo, el paso (i3) se lleva a cabo típicamente usando el mismo método que en el paso (a) descrito anteriormente.

En la etapa ii, los péptidos modificados (típicamente péptidos fosforilados) se enriquecen a partir de los péptidos obtenidos en la etapa i. El paso ii da lugar así a varias fracciones enriquecidas en péptidos modificados (típicamente péptidos fosforilados).

El enriquecimiento de péptidos modificados (típicamente péptidos fosforilados) en la etapa ii se lleva a cabo típicamente usando cromatografía multidimensional. En una realización, la cromatografía multidimensional se lleva a cabo usando cromatografía líquida de alta resolución de intercambio catiónico fuerte (SCX-HPLC), cromatografía de afinidad de iones metálicos inmovilizados (IMAC) y cromatografía en dióxido de titanio ( $\text{TiO}_2$ ). En otra realización, la cromatografía multidimensional se lleva a cabo usando cromatografía líquida de alta resolución de intercambio aniónico (SAX-HPLC), cromatografía de afinidad por ion metálico inmovilizado (IMAC) y cromatografía en dióxido de titanio ( $\text{TiO}_2$ ). En estas realizaciones de la invención, las técnicas cromatográficas se llevan a cabo secuencialmente.

Alternativamente, el enriquecimiento de péptidos modificados (típicamente péptidos fosforilados) en la etapa ii se lleva a cabo usando métodos basados en anticuerpos, como se describió anteriormente.

En la etapa iii, se lleva a cabo una espectrometría de masas en tándem por cromatografía líquida (LC-MS/MS) en los péptidos modificados enriquecidos (típicamente péptidos fosforilados) obtenidos en la etapa ii.

En la etapa iv, los péptidos modificados (típicamente péptidos fosforilados) detectados en la etapa iii se comparan con una base de datos de referencia conocida para identificar los péptidos modificados (típicamente péptidos fosforilados). Normalmente, este paso se lleva a cabo utilizando un motor de búsqueda disponible en el mercado, como, entre otros, los motores de búsqueda MASCOT, ProteinProspector o Sequest.

En el paso v, los datos relacionados con los péptidos modificados (típicamente péptidos fosforilados) identificados en el paso iv se compilan en una base de datos. Esta base de datos enumera todos los parámetros necesarios para la cuantificación de péptidos fosforilados en experimentos biológicos posteriores. Típicamente, los datos relacionados con los péptidos modificados (típicamente péptidos fosforilados) incluyen la identidad de los péptidos modificados (típicamente péptido fosforilado), la relación masa a carga (m/z), la carga y/o el tiempo de retención relativo. Esto permite que los datos relacionados con los péptidos en la muestra, típicamente la relación masa a carga (m/z), carga (z) y tiempo de retención relativa de los péptidos en la muestra, se comparen con los valores de los péptidos modificados (típicamente péptidos fosforilados) en la base de datos y así permite la identificación y cuantificación de los péptidos modificados (típicamente péptidos fosforilados) en la muestra.

La compilación de la base de datos no necesita llevarse a cabo simultáneamente con el método de la invención. La compilación de la base de datos se puede llevar a cabo por separado, antes de que se use la técnica TIQUAS en el método de la invención para identificar el péptido en la muestra.

La base de la técnica TIQUAS es la construcción de una base de datos de péptidos modificados (típicamente péptidos fosforilados) que pueden detectarse y cuantificarse por LC-MS. Esta base de datos enumera todos los parámetros necesarios para la cuantificación de péptidos modificados (típicamente péptidos fosforilados) en experimentos biológicos posteriores que incluyen la identidad del péptido modificado (típicamente péptido fosforilado), relación masa a carga (m/z), carga y retención relativa hora. La base de datos se puede construir enriqueciendo péptidos modificados (típicamente péptidos fosforilados) usando cromatografía multidimensional (como intercambio catiónico fuerte, IMAC y TiO<sub>2</sub>). Las fracciones de péptidos modificados enriquecidos (típicamente péptidos fosforilados) se pueden analizar a continuación mediante LC-MS/MS para la identificación de péptidos modificados (típicamente péptidos fosforilados).

El programa informático denominado PESCAL (Cutillas and Vanhaesebroeck, Molecular & Cellular Proteomics 6, 1560-1573 (2007)) automatiza la cuantificación de cada uno de los péptidos modificados (típicamente péptidos fosforilados) enumerados en la base de datos en series de LC-MS de péptidos modificados (típicamente péptidos fosforilados) tomados de experimentos biológicos. Para estos experimentos biológicos, las proteínas en los lisados celulares se digieren usando tripsina u otras proteasas adecuadas. Los patrones internos de péptidos (tales como fosfopéptidos), que son péptidos modificados de referencia (típicamente péptidos fosforilados de referencia), se enriquecen en cantidades conocidas en todas las muestras que se van a comparar.

Los péptidos modificados (típicamente péptidos fosforilados) en la mezcla de péptidos resultante se enriquecen usando una etapa de extracción de IMAC o TiO<sub>2</sub> simple de realizar. Los péptidos modificados enriquecidos (típicamente péptidos fosforilados) se analizan en una sola ejecución de LC-MS típicamente pero no restringida a aproximadamente 120 minutos (ciclo total). Luego, PESCAL construye cromatogramas de iones extraídos (XIC, es decir, un perfil de elución) para cada uno de los péptidos modificados (típicamente péptidos fosforilados) presentes en la base de datos en todas las muestras que se van a comparar. El programa también calcula la altura del pico y el área bajo la curva de cada XIC. Los datos se normalizan dividiendo la lectura de intensidad (áreas de pico o alturas) de cada analito peptídico modificado (típicamente fosfopéptido) por los de los péptidos modificados (típicamente fosfopéptido) IS.

Como alternativa al uso de la técnica TIQUAS, en los métodos de la invención, la cuantificación de modificaciones tales como la fosforilación también se puede llevar a cabo usando técnicas de EM que usan marcadores de isótopos para la cuantificación, como el etiquetado metabólico (por ejemplo, aminoácidos marcados con isótopos estables en cultivo, (SILAC); Olsen, JV et al., Cell 127, 635-648 (2006)), y derivación química (por ejemplo, iTRAQ (Ross, PL, et al., Mol Cell Proteomics 2004, 3, (12) ), 1154-69), ICAT (Gygi, SP et al., Nat Biotechnol 17, 994-999 (1999)), TMT (Dayon L et al., Anal Chem. 2008 Apr 15; 80 (8): 2921-31) Las modificaciones de proteínas pueden cuantificarse con técnicas de LC-MS que miden las intensidades de iones no fragmentados o con técnicas de LC-MS/MS que miden las intensidades de los iones de fragmentos (como el control de reacción seleccionado (SRM), también llamado monitoreo de reacción múltiple (MRM)).

Una vez que los péptidos modificados se han agrupado de acuerdo con uno de los parámetros descritos en el presente documento, el siguiente paso en el método de la divulgación es (2) calcular el enriquecimiento de los péptidos modificados de la primera muestra en comparación con los péptidos modificados a partir del segundo muestra en el grupo.

Por "enriquecimiento" de los péptidos modificados se entiende un aumento en la abundancia o frecuencia. Por consiguiente, en esta etapa del método de la divulgación, el aumento en la abundancia o frecuencia de los péptidos modificados de una muestra (la primera muestra) se compara con el aumento (o disminución) en abundancia o frecuencia de los péptidos modificados de otra muestra (la segunda muestra). Como se establece anteriormente,

cuando se usan más de dos muestras en el método de la divulgación, esta etapa se lleva a cabo para comparar el enriquecimiento de péptidos modificados entre dos muestras cualquiera a la vez. El cálculo del enriquecimiento se puede hacer usando cualquier método apropiado.

5 El enriquecimiento se puede calcular contando el número de péptidos modificados en el grupo que son más abundantes en la primera muestra que en la segunda muestra y el número de péptidos modificados en el grupo que son menos abundantes en la primera muestra que en la segunda muestra; y restando el número de péptidos modificados en el grupo que son menos abundantes en la primera muestra que en la segunda muestra a partir del número de péptidos modificados en el grupo que son más abundantes en la primera muestra que en la segunda muestra. Esto puede expresarse alternativamente como restar el número de péptidos modificados en el grupo que son más abundantes en la segunda muestra que en la primera muestra a partir del número de péptidos modificados en el grupo que son más abundantes en la primera muestra que en la segunda muestra.

15 Por lo tanto, este método implica contar el número de péptidos modificados en el grupo de la primera muestra cuyas intensidades se incrementan o disminuyen con respecto a la segunda muestra. Opcionalmente, solo se tienen en cuenta los péptidos modificados en el grupo de la primera muestra cuyas intensidades se incrementan o disminuyen estadísticamente significativamente con respecto a la segunda muestra. El enriquecimiento de los péptidos modificados se calcula a continuación mediante un parámetro al que se hace referencia como "recuentos delta", que se define como el número de péptidos modificados en un grupo de sustrato que aumentan su intensidad con respecto a la segunda muestra menos aquellos que disminuyen su intensidad. La ventaja de este enfoque es que no implica divisiones y, por lo tanto, es aplicable a situaciones en las que el número de péptidos modificados en un grupo particular es cero en algunas de las muestras que se comparan. Un ejemplo de este enfoque se muestra en la Figura 6B.

25 El enriquecimiento puede calcularse comparando la abundancia media de todos los péptidos modificados en el grupo de la primera muestra con la abundancia media de todos los péptidos modificados en el grupo de la segunda muestra. Esto se hace típicamente calculando la abundancia media (promedio aritmético) de todos los péptidos modificados en el grupo de la primera muestra y la abundancia media de todos los péptidos modificados en el grupo de la segunda muestra; y dividir la abundancia media de todos los péptidos modificados en el grupo de la primera muestra por la abundancia media de todos los péptidos modificados en el grupo de la segunda muestra. La cifra resultante puede transformarse opcionalmente en  $\log_2$ . Por lo tanto, este método implica comparar las medias (promedio aritmético) de las intensidades de todos los péptidos modificados en el grupo a partir de una primera muestra con respecto a las intensidades de todos los péptidos modificados en el grupo de la segunda muestra. Un ejemplo de este enfoque se muestra en la Figura 6A. Esto se denomina aquí método de "repetición".

35 El enriquecimiento se puede calcular calculando la media de todos los cambios de repetición de los péptidos modificados en el grupo y la media de todos los cambios de repetición de los péptidos modificados en todo el experimento; y luego dividiendo la media de todos los cambios de repetición de los péptidos modificados en el grupo por cambio de repetición medio de todos los péptidos modificados en todo el experimento. La cifra resultante puede transformarse opcionalmente en  $\log_2$ . Un ejemplo de este enfoque se muestra en la Figura 6C. Esto se denomina aquí método de "enriquecimiento".

45 La etapa final (3) del método de la invención comprende calcular el significado estadístico del enriquecimiento. El método de la divulgación implica calcular el enriquecimiento de los péptidos modificados de la primera muestra en comparación con los péptidos modificados de la segunda muestra en el grupo. Por consiguiente, un enriquecimiento estadísticamente significativo en este método es indicativo de que se está activando una enzima modificadora de proteína en la primera muestra en comparación con la segunda muestra. Si el enriquecimiento no es estadísticamente significativo, esto indica que la enzima modificadora de proteínas no se activa en la primera muestra. Se entenderá que también se aplicará lo contrario, es decir, que si hay enriquecimiento de los péptidos modificados de la segunda muestra en comparación con los péptidos modificados de la segunda muestra del grupo, entonces la enzima modificadora de proteínas se activa en la segunda muestra en comparación a la primera muestra.

55 El significado estadístico se puede calcular usando cualquier método estadístico adecuado, que estará dentro de la capacidad de una persona experta en la técnica. Los métodos estadísticos adecuados incluyen la prueba hipergeométrica, la prueba t pareada, la prueba Z, la prueba de Kolmogorov-Smirnov, la prueba de Chi cuadrado, la prueba exacta de Fisher o la prueba de rango con signo de Wilcoxon. Cualquiera de estas pruebas puede ser seguido opcionalmente por un método de corrección de pruebas múltiples, como el método de Tasa de Descubrimiento Falso de Benjamini Hochberg (FDR).

60 Por ejemplo, la significancia del enriquecimiento puede calcularse mediante la prueba hipergeométrica. Esto puede ser seguido opcionalmente por un método de corrección de prueba múltiple como el método Benjamini Hochberg FDR. Este método de cálculo de la significancia se usa típicamente en combinación con el método de "recuentos delta" descrito en este documento. Alternativamente, el montante del enriquecimiento puede calcularse usando la prueba t pareada. Nuevamente, esto puede ser seguido opcionalmente por una corrección de prueba múltiple tal como el método Benjamini Hochberg FDR. Este método para calcular la significancia se usa típicamente en

combinación con el método de "repetición" descrito en este documento. En el método de "enriquecimiento" descrito aquí, el significado estadístico del enriquecimiento puede calcularse en una realización usando un puntaje  $Z = (mS - mP) * m^{1/2} / \delta$  (donde mS: log2 significan intensidades de grupo de sustratos; mP: intensidades medias de Log2; conjunto de datos completo; m: tamaño del grupo de sustrato;  $\delta$ : desviación estándar, intensidades medias, conjunto de datos completo). El puntaje z se puede convertir a un valor p. Se puede determinar un enriquecimiento estadísticamente significativo basándose en un valor de p adecuado, por ejemplo,  $p < 0.05$ ,  $p < 0.01$  o  $p < 0.001$ . El significado estadístico se puede calcular usando un programa de ordenador apropiado.

Un enriquecimiento estadísticamente significativo es indicativo de que se está activando una enzima modificadora de proteína en la primera muestra en comparación con la segunda muestra. Esto también se puede describir como un enriquecimiento diferencial de péptidos modificados entre muestras. Por lo tanto, el método de la divulgación es como un método para determinar la activación diferencial de una enzima modificadora de proteínas tal como una quinasa en una primera muestra en comparación con una segunda muestra. Sin embargo, la cuantificación de los resultados se puede obtener, por ejemplo, analizando el enriquecimiento usando diferentes concentraciones de una sustancia de prueba y observando los efectos dependientes de la concentración.

Los presentes inventores han ideado una técnica para inferir sistemáticamente la activación de la ruta de proteína quinasa a partir de datos de fosfoproteómica basados en MS. La técnica se denomina análisis de enriquecimiento de sustrato de quinasa (KSEA).

El método puede ser un método para cuantificar la actividad de proteína quinasa en una muestra, que comprende:

(i) agrupar los péptidos fosforilados de una primera muestra y los péptidos fosforilados de una segunda muestra en un solo grupo de acuerdo con uno de los siguientes parámetros:

(a) péptidos fosforilados que tienen un sitio de fosforilación que está fosforilado por la misma proteína quinasa; o

(b) péptidos fosforilados que tienen un sitio de fosforilación que es parte del mismo motivo de fosforilación;

(ii) calcular el enriquecimiento de los péptidos fosforilados de la primera muestra en comparación con los péptidos fosforilados de la segunda muestra en el grupo; y

(iii) calcular el significado estadístico de dicho enriquecimiento;

en el que un enriquecimiento estadísticamente significativo es indicativo de que se está activando una proteína quinasa en la primera muestra en comparación con la segunda muestra.

El enriquecimiento se puede calcular contando el número de péptidos fosforilados en el grupo que son más abundantes en la primera muestra que en la segunda muestra y el número de péptidos fosforilados en el grupo que son menos abundantes en la primera muestra que en la segunda muestra; y restando el número de péptidos fosforilados en el grupo que son menos abundantes en la primera muestra que en la segunda muestra a partir del número de péptidos fosforilados en el grupo que son más abundantes en la primera muestra que en la segunda muestra.

El enriquecimiento puede calcularse calculando la abundancia media (promedio aritmético) de todos los péptidos fosforilados en el grupo de la primera muestra y la abundancia media de todos los péptidos fosforilados en el grupo de la segunda muestra; y dividiendo la abundancia media de todos los péptidos fosforilados en el grupo de la primera muestra por la abundancia media de todos los péptidos fosforilados en el grupo de la segunda muestra.

El enriquecimiento puede calcularse calculando la media de todos los cambios de repetición de los péptidos fosforilados en el grupo y la media de todos los cambios de repetición de los péptidos fosforilados en todo el experimento; y luego dividiendo la media de todos los cambios de repetición de los péptidos fosforilados en el grupo por cambio de repetición medio de todos los péptidos fosforilados en todo el experimento.

La presente invención se describirá ahora adicionalmente a modo de referencia a los siguientes ejemplos que están presentes solo con fines de ilustración. En los ejemplos, se hace referencia a una serie de figuras en las que:

La Figura 1 muestra que los fosfopéptidos están regulados diferencialmente en las células P31/Fuj y Kasumi-1. Las células P31/Fuj y Kasumi-1 se sembraron 24 h antes de la recolección y se procesaron para el análisis por MS libre de etiqueta como se describe en la sección de Materiales y Métodos de los Ejemplos en este documento. (A) Gráfico de volcán del pliegue y valores p para cada sitio de fosforilación que muestra los niveles de fosfopéptidos en células P31/Fuj relativas a las células Kasumi-1. La media se refiere al promedio aritmético de los datos de repetición Log2. (B) Resumen de fosfopéptidos regulados diferencialmente entre las células P31/Fuj y Kasumi-1. (C) Distribución de frecuencia del cambio de repeticiones en la fosforilación entre líneas celulares. (D) Análisis de componentes principales de células P31/Fuj y Kasumi-1 basadas en datos de fosforilación global. (E) Ejemplos ilustrativos de expresión de grupo de sustrato a través de las dos líneas celulares. Cada punto de datos es la diferencia de

repetición Log<sub>2</sub> de un fosfopéptido que contiene un sitio de fosforilación que se sabe que es el sustrato de la quinasa nombrada. (F) Datos de enriquecimiento del grupo de sustrato de quinasa representadas con más de tres entradas en el grupo de sustrato.

5 La Figura 2 muestra la determinación de quinasa activadas diferencialmente entre las células P31/Fuj y Kasumi-1 usando el análisis de enriquecimiento del sustrato de quinasa (KSEA) tomando fosfoSite como la base de datos de las relaciones quinasa-sustrato. Los fosfopéptidos regulados diferencialmente entre las células P31/Fuj y Kasumi-1 se usaron para inferir la actividad de quinasa distintiva entre ambas líneas celulares. Las actividades quinasa se infirieron usando algoritmos basados en el conteo de fosfopéptidos (A) o mediante la comparación de los promedios de las intensidades de fosfopéptidos con los grupos de sustratos (B).

15 La Figura 3 muestra la determinación de quinasa activadas diferencialmente entre las células P31/Fuj y Kasumi-1 usando el análisis de enriquecimiento del sustrato de quinasa (KSEA) tomando phosphoElm como la base de datos de las relaciones quinasa-sustrato. Los fosfopéptidos regulados diferencialmente entre las células P31/Fuj y Kasumi-1 se usaron para inferir la actividad de quinasa distintiva entre ambas líneas celulares. Las actividades quinasa se infirieron usando algoritmos basados en el conteo de fosfopéptidos (A) o mediante la comparación de los promedios de las intensidades de fosfopéptidos con los grupos de sustratos (B).

20 La Figura 4 es una inmunoprecipitación Western que confirma los resultados obtenidos por KSEA y muestra la fosforilación de proteínas en sitios que se sabe que se correlacionan con la actividad de quinasa mostradas con una flecha en la Figura 1F.

25 La Figura 5 muestra KSEA comparativa de líneas celulares P31/Fuj y Kasumi-1 basadas en motivos de fosforilación. (A) Descripción general del enriquecimiento del grupo de sustrato basado en motivos de fosforilación comunes. (B) Ejemplos ilustrativos de grupos de sustrato enriquecidos en cualquiera de las líneas celulares.

30 La Figura 6 muestra una comparación de estrategias de enriquecimiento. Se probaron tres métodos matemáticos diferentes para inferir el enriquecimiento de grupos de sustrato en los datos de fosfoproteómica. Los diferentes valores se calcularon de la siguiente manera. (A) Repetición = log<sub>2</sub> (intensidades medias del grupo de sustrato en P31-Fuji/intensidades del grupo de sustrato medio en Kasumi-1). El montante del enriquecimiento se calculó mediante la prueba t comparada en Excel. (B) Recuentos delta = el número de sustratos en el grupo de sustrato aumentado significativamente en P31-Fuj menos los aumentados en Kasumi-1. El montante del enriquecimiento se calculó mediante la prueba hipergeométrica en Excel. (C) Enriquecimiento = Log<sub>2</sub> Media del grupo de sustrato dividido por Log<sub>2</sub> Media de todos los datos. El montante del enriquecimiento se calculó usando un puntaje Z = (mS-mP)\*m<sup>1/2</sup>/δ (donde mS: Log<sub>2</sub> significa grupo de sustrato de intensidades; mP: Log<sub>2</sub> significa intensidades conjunto de datos; m: tamaño del grupo de sustrato; δ: estándar desviación media intensidades conjunto de datos completo). El puntaje z se convirtió a un valor p en Excel. La importancia de los enriquecimientos se denota como \* p <0.05; \*\* p <0.01; \*\*\* p <0.001.

40 La Figura 7 muestra la heterogeneidad del enriquecimiento del grupo de sustrato en los blastos de AML primarios. (a) La agrupación jerárquica (correlación de Spearman, enlace completo) de los valores de enriquecimiento del sustrato identificó 12 grupos principales. (b, c) Ejemplos de grupos de sustrato corregulados (m, número de fosfopéptidos cuantificados en el grupo de sustrato nombrado). (d) Correlación negativa entre actividades en el grupo 5a y aquellas en los grupos 9, 6 y 7 que comprenden sustratos para CK2a, PKB/PKC, PAK/MAPKAPK2 y Lyn/Syk/Tyr, respectivamente. (e) Grupos de sustratos de quinasa enriquecidos con mayor frecuencia en la LMA primaria.

### Ejemplo 1 - Análisis de enriquecimiento de sustrato de quinasa (KSEA) en células P31/Fuj y Kasumi-1

#### 50 Materiales y métodos

##### Cultivo de células

55 Se hicieron crecer células P31/Fuj y Kasumi-1 en RPMI-1640 suplementado con 10% de SFB, 100 unidades/mL de penicilina/estreptomina y b-Mercaptoetanol 50 μM a 37°C en una atmósfera humidificada con 5% de CO<sub>2</sub>. Las células se mantuvieron a una confluencia de 0.5-2x10<sup>6</sup> células/mL.

##### Lisis celular y digestión de proteínas

60 Las células se dividieron en una confluencia de 0.5x10<sup>6</sup> células/mL y para cada condición se realizaron 3 réplicas biológicas independientes por experimento y todos los experimentos se realizaron dos veces para un total de al menos 6 réplicas biológicas por condición. Se recogieron un total de 10x10<sup>6</sup> células por condición por centrifugación, se lavaron dos veces con PBS frío suplementado con Na<sub>3</sub>VO<sub>4</sub> 1 mM y NaF 1 mM y se sometieron a lisis en 1 mL de regulador de urea (Urea 8 M en Hepes 20 mM pH 8.0 suplementado con Na<sub>3</sub>VO<sub>4</sub> 1 mM, NaF 1 mM, β-glicerofosfato 1 mM y ácido okadaico 1mM). Los lisados celulares se homogeneizaron adicionalmente por sonicación (3 pulsos de 15 s) y el material insoluble se eliminó por centrifugación. La proteína se cuantificó usando el ensayo de Bradford.

Se redujeron y alquilaron 0.5 mg de proteína por muestra mediante incubación secuencial con DTT 4.1 mM y yodoacetamida 8.3 mM durante 15 minutos a temperatura ambiente en la oscuridad. Para la digestión de proteínas, la concentración de urea se redujo a 2M mediante la adición de Hepes 20 mM, pH 8.0. Se añadió TLCK-tripsina inmovilizada (20 unidades de TAME/mg) y las muestras se incubaron durante la noche a 37°C. La digestión se detuvo añadiendo 1% de concentración final de TFA y las perlas de tripsina se eliminaron por centrifugación. Las soluciones peptídicas resultantes se desalaron usando cartuchos C18-Oasis según lo indicado por el fabricante con ligeras modificaciones. En resumen, los cartuchos de Oasis se acondicionaron con 1 ml de ACN y se equilibraron con 1 mL de solución de lavado (0.1% de TFA/2% de ACN). Los péptidos se cargaron en los cartuchos y se lavaron con 1.5 mL de solución de lavado. Finalmente, los péptidos se eluyeron con 0.5 mL de regulador de ácido glicólico (ácido glicólico 1 M/5% de TFA/80% de ACN).

#### Enriquecimiento de fosfopéptidos

El fosfoenriquecimiento peptídico se realizó usando TiO<sub>2</sub> como se describió previamente (Montoya et al, Methods 54, 370, 2011). En resumen, los eluyentes peptídicos se normalizaron en 1 mL con regulador de ácido glicólico y se incubaron con 25 µL de regulador de TiO<sub>2</sub> (suspensión al 50% en X% de TFA) durante 5 min a temperatura ambiente. Las esferas de TiO<sub>2</sub> se empaquetaron por centrifugación en columnas de C<sub>18</sub> Spin previamente equilibradas con regulador de ácido glicólico. Las columnas se lavaron secuencialmente con 300 µl de ácido glicólico, 50% de ACN y regulador de bicarbonato de amonio (NH<sub>4</sub>HCO<sub>3</sub> 20 mM, pH 6.8 en 50% de ACN). Para la elución de fosfopéptidos, las perlas se incubaron durante 1 minuto a temperatura ambiente con 50 µL de NH<sub>4</sub>OH al 5% en 50% de ACN y se centrifugaron. Este paso fue repetido 3 veces. Los eluyentes de la misma muestra se combinaron y se acidificaron con ácido fórmico hasta una concentración final del 10%. Finalmente, las muestras se secaron usando Speed-Vac y los sedimentos se almacenaron a -80°C.

#### Análisis LC-MS/MS

Los péptidos secos se reconstituyeron en TFA al 1% que contenía 20 fmol/µl de una digestión de enolasa de levadura. LC-MS/MS se realizó como se describe previamente (Casado and Cutillas, Mol Cell Proteomics 10, M110 003079 (enero de 2011)). En resumen, los sedimentos de fosfopéptidos se resuspendieron en 20 µL de TFA al 0,1% y se cargaron 4 mL en un sistema LC-MS/MS, que consiste en una cromatografía líquida de alta presión con nanoflujo (UPLC, nanoAccuity, Waters) acoplada en línea a un espectrómetro de masas Orbitrap-XL (Thermo Fisher Scientific, Hemel Hempstead, Reino Unido). Los 5 mejores iones de carga múltiple más intensos se seleccionaron para la fragmentación de CID en el modo de activación de múltiples etapas. La resolución de MS1 se estableció en 60.000.

#### Procesamiento de datos y análisis estadístico

La identificación de péptidos se realizó mediante el emparejamiento de datos MS/MS con la base de datos SwissProt (descargada en marzo de 2011) restringida a entradas humanas (22.000 entradas de proteínas) usando el motor de búsqueda Mascot (Perkins et al, Electrophoresis 20, 3551 (diciembre, 1999)). Las tolerancias de masa se establecieron en 5 ppm y 600 unidades de milimasa para los iones progenitor y fragmentos, respectivamente. Las modificaciones variables permitidas fueron fosforilación en Ser, Thr y Tyr, PyroGlu en N-terminal glutamina y oxidación de metionina. Los fosfopéptidos que tenían una expectativa de Mascot <0.05 (aproximadamente 2% de FDR) se incluyeron en una base de datos de sitios cuantificables por MS. El software Pescal (Casado and Cutillas, Mol Cell Proteomics 10, M110 003079 (Jan, 2011) y Cutillas and Vanhaesebroeck, Mol Cell Proteomics 6, 1560 (Sep, 2007)) se usó para obtener alturas de pico y áreas de cromatogramas de iones extraídos (XIC) de fosfopéptidos en esta base de datos en todas las muestras que se compararon. Los tiempos de retención alineados con Pescal usan los de los péptidos derivados de la enolasa previamente añadidos a las muestras como puntos de referencia a lo largo de los cromatogramas. Las ventanas de XIC fueron de 7 ppm y 2 minutos.

Los valores de intensidad de XIC se normalizaron a la suma de todos los valores en una muestra y a la media de las intensidades de fosfopéptidos en las muestras. La importancia de que las medias de los datos transformados con log<sub>2</sub> fueran diferentes en todas las muestras se evaluó mediante la prueba t de Student seguida por la corrección de pruebas múltiples de Benjamini Hochberg.

#### Análisis de enriquecimiento de sustrato de quinasa

Se agruparon, en conjuntos de sustrato, fosfopéptidos con un valor p (mediante prueba t de datos transformados con log<sub>2</sub>) de menos de 0.05. La característica común de los fosfopéptidos en estos grupos de sustratos era que estos tenían sitios que se sabe que son fosforilados por una quinasa específica o que el residuo fosforilado estaba presente en el contexto de motivos de fosforilación predefinidos. La información sobre las relaciones quinasa-sustrato se obtuvo de bases de datos disponibles públicamente, concretamente PhosphoSite (Hornbeck et al., Proteomics 4, 1551 (Jun, 2004)) and PhosphoElm (Dinkel et al., Nucleic Acids Res 39, D261 (Jan, 2011)), mientras que la lista de motivos se obtuvo de la literatura y de un análisis de nuestro conjunto de datos con Motif-X (Schwartz and Gygi, Nat Biotechnol 23, 1391 (noviembre de 2005)).

Se usaron tres metodologías para inferir las diferencias globales en la abundancia de grupos de sustrato en las muestras. El primero implicaba contar el número de fosfopéptidos en el grupo de sustrato cuyas intensidades aumentaban o disminuían con relación al control. El enriquecimiento de los sustratos de quinasa se calculó mediante un parámetro que denominamos "recuentos delta", que se definió como el número de fosfopéptidos en un grupo de sustrato que aumenta significativamente su intensidad en relación con el control menos aquellos que disminuyen su intensidad. La ventaja de este enfoque es que no implica divisiones y, por lo tanto, es aplicable a situaciones en las que el número de sustratos para un grupo de sustrato particular es cero en algunas de las muestras que se comparan. El montante del enriquecimiento se calculó mediante la prueba hipergeométrica seguida por la corrección de pruebas múltiples de Benjamini Hochberg. El segundo enfoque investigado aquí para evaluar el enriquecimiento de los grupos de sustratos implica la comparación de los medios (promedio aritmético) de las intensidades de todos los fosfopéptidos en un grupo de sustrato dado en una muestra en relación con el control. Para esto, las intensidades de fosfopéptidos se normalizaron primero para controlar y log2 se transformaron de modo que todas estas intensidades se expresaron en relación con el control. Además de derivar un valor de diferencia de pliegue de un grupo de sustrato dado a través de muestras de control y prueba, también calculamos valores de p usando la prueba t pareada seguida por la corrección de prueba múltiple de Benjamini Hochberg. El tercer enfoque implicaba calcular la relación de medias de las intensidades de fosfopéptidos en los grupos de sustratos con respecto a las intensidades de fosfopéptidos en el conjunto de datos completo (es decir, enriquecimiento = mS/mP, donde mS: Log2 significa intensidades de grupo de sustratos;) El montante del enriquecimiento fue calculada usando un puntaje  $Z = (mS - mP) * m^{1/2} / \delta$  (donde m: tamaño del grupo de sustrato;  $\delta$ : desviación estándar de las intensidades medias del conjunto de datos completo). El puntaje z se convirtió a un valor p en Microsoft Excel 2007. Se escribió un script en Visual Basic para Aplicaciones para automatizar la aplicación de algoritmos KSEA.

## Resultados

### 25 Principios del análisis de enriquecimiento del sustrato de quinasa (KSEA)

Para inferir la actividad de la quinasa a partir de los datos de la fosfoproteómica, se agruparon los fosfopéptidos identificados en la fosfoproteómica a gran escala en grupos de sustrato; estos contienen sitios de fosforilación que se sabe que son sustratos de quinasas específicas o comparten motivos de fosforilación específicos. Con el fin de definir grupos de sustrato, se obtuvo información sobre las relaciones entre sustratos de quinasas de bases de datos públicamente disponibles de sitios de fosforilación, concretamente PhosphoSite (Hornbeck et al, Proteomics 4, 1551 (junio de 2004)) y PhosphoElm (Dinkel et al, Nucleic Acids Res 39, D261 (enero de 2011)), mientras que los motivos de fosforilación se obtuvieron de la bibliografía y de Motif-X (Schwartz y Gygi, Nat Biotechnol 23, 1391 (noviembre de 2005)). Un total de 293 y 298 quinasas estuvieron representadas en las bases de datos PhosphoSite y PhosphoElm, respectivamente. Cada grupo de sustrato tenía un promedio de ~17 sustratos y una media de 5 sustratos por quinasa. El número de motivos analizados fue de 109. El grado de enriquecimiento de las actividades de quinasa se calculó considerando las intensidades de los fosfopéptidos que mostraban diferencias estadísticamente significativas entre las muestras experimentales. El montante del enriquecimiento de las actividades de quinasa se estimó mediante la prueba hipergeométrica, mientras que el significado estadístico de las diferencias en las intensidades medias de fosfopéptidos se evaluó mediante la prueba t pareada.

Aplicación de KSEA para la comparación de datos de fosfoproteómica de 2 líneas celulares de leucemia mieloide aguda (AML) diferentes

45 El KSEA se puede explicar mejor ilustrando su uso en un experimento en el que se compararon los fosfoproteomas de dos líneas celulares AML, a saber, P31/Fuj y Kasumi-1. Estábamos interesados en comparar estas líneas celulares debido a su trasfondo genético en términos de activación de PI3K (P31/Fuj y Kasumi-1 son PTEN negativas y positivas respectivamente) y su sensibilidad a la inhibición de la proliferación por inhibidores de señalización también es muy diferente, con P31/Fuj es resistente a múltiples fármacos en comparación con Kasumi-1 (Casado and Cutillas, Mol Cell Proteomics 10, M110 003079 (Jan, 2011). El fosfoproteoma de ambas líneas celulares de AML se analizó utilizando la metodología LC-MS/MS descrita anteriormente (Casado and Cutillas, Mol Cell Proteomics 10, M110 003079 (Jan, 2011). Se analizaron tres repeticiones biológicas por línea celular en cada experimento y cada experimento se repitió 3 veces, lo que condujo a un total de 9 repeticiones biológicas por línea celular. Se obtuvieron un total de 4300 fosfopéptidos cuantificados a través de las líneas celulares, con 431 y 306 de estos fosforilados preferentemente en células P31/Fuj o Kasumi-1, respectivamente con un valor p de corte <0.05; de estos, 204 y 134 fosfopéptidos tenían un valor de p <0.01 (Fig. 1A y B). Las intensidades de la mayoría de los fosfopéptidos no fueron significativamente diferentes entre las líneas celulares (Fig. 1B) y los cambios en los pliegues log2 centrados alrededor de cero (sin cambios, Fig. 1C). Sin embargo, las diferencias observadas fueron suficientes para separar las muestras P31/Fuj y Kasumi-1 basadas en el análisis del componente principal (PCA, Fig 60 ID).

Luego se aplicó KSEA a los datos de fosfoproteómica conectando sitios de fosforilación cuantificados a quinasas usando las dos estrategias diferentes descritas anteriormente: i) coincidencia de sitios de fosforilación con bases de datos de sustrato y ii) agrupación de sitios de fosforilación en motivos que luego pueden asignarse a diferentes grupos de quinasa.



La Figura 2 muestra los resultados obtenidos cuando se utilizaron dos métodos diferentes para cuantificar el enriquecimiento tomando la base de datos phosphoSite como la fuente de las relaciones quinasa-sustrato. Como medida del enriquecimiento del grupo de sustrato, contamos los sustratos significativamente aumentados en Kasumi-1 y los restamos a aquellos significativamente aumentados en P31/Fuj para cada quinasa; estos valores, denominados recuentos delta, se muestran en la Fig. 2A para todas las quinasas apareados con al menos tres sustratos del sitio de fosforilación. La prueba hipergeométrica se utilizó luego para evaluar el montante del enriquecimiento de los sitios de fosforilación que aumentan o disminuyen significativamente en relación con el número total de sitios para esa quinasa en particular (Fig. 2A). Como una prueba adicional para evaluar el enriquecimiento de los grupos de sustratos, también se compararon las medias de las intensidades del sustrato entre las dos líneas celulares para cada quinasa (Fig. 2B).

Se evaluó la significancia de las diferencias entre los medios mediante la prueba t pareada. Ambos tipos de análisis indicaron que la fosforilación de sustratos para CDK, PKA, PKB/AKT, PKC, P90RSK, DYRK2, PAK y ROCK era significativamente enriquecida en células P31/Fuj (Fig. 2), mientras que los sustratos para CK2, MEK1 y ERK1 eran significativamente diferentes. Enriquecido en células Kasumi-1 (Fig 2). El mismo análisis se realizó utilizando la base de datos phosphoElm (Fig 3) como fuente de las relaciones quinasa-sustrato. Estos análisis mostraron un acuerdo notable en los resultados independientemente de qué base de datos o prueba estadística se utilizara como base para KSEA. Los sustratos de quinasas cuya fosforilación se enriqueció incluyeron aquellos para PKC, PKB/AKT, RSK y PAK en P31/Fuj y un enriquecimiento de CK2 y aquellos para actividades MEK y ERK en células Kasumi-1, sugiriendo que estas quinasas eran más activas en la línea celular respectiva.

Para validar los resultados de KSEA, se usó Inmunoprecipitación Western para medir los sitios de fosforilación que se correlacionan con la actividad de las quinasas seleccionadas que se identifican como reguladas diferencialmente o sobre proteínas que se sabe que son fosforiladas por algunas de estas quinasas (Fig. 4). Por lo tanto, la fosforilación de Ser-473 AKT, Ser-380 P90RSK, Thr-505 PKC $\delta$  y Thr-538 PKC $\theta$  (que se correlacionan con su estado de actividad) se incrementaron en las células P31/Fuj relativas a Kasumi-1, mientras que la fosforilación de ERK en Thr202/Tyr204 fue más prominente en las células Kasumi-1 (Fig 4). La fosforilación de GSK3 $\beta$  en Ser-9 y 6SRP en Ser-236/237 que están catalizadas por AKT y RSK, respectivamente, también se incrementaron en P31/Fuj (Fig. 4). Estos resultados son consistentes con, y por lo tanto validan, los obtenidos por KSEA de los datos de fosforilación basados en MS a gran escala que se muestran en las Figuras 2 y 3.

Los algoritmos KSEA también se aplicaron para encontrar motivos de fosforilación fosforilados diferencialmente en células P31/Fuj y Kasumi-1. El análisis basado tanto en el recuento delta como en la comparación de los medios indicó que los motivos ácidos tendían a estar más fosforilados en las células Kasumi-1, mientras que los motivos basófilos estaban más fosforilados en las células P31/Fuj (Fig. 5A). En ambos casos, la fosforilación de los motivos RxRxxS y KxRxxS, que están asociados con las isoformas PKB/AKT y RSK, estaban más fosforilados en las células P31/Fuj relativas a Kasumi-1, mientras que el motivo SDxExE, asociado a CK2, estaba predominantemente más fosforilado en Kasumi-1 células (Fig 5B). Estos resultados son consecuentes con los obtenidos de KSEA basados en bases de datos de relaciones quinasa-sustrato (Fig. 2 y 3).

El presente inventor, por lo tanto, ha descubierto que cuantificar el enriquecimiento de sustratos (que previamente se ha encontrado que está regulado diferencialmente mediante fosfoproteómica a gran escala) que pertenecen a grupos predefinidos puede predecir la activación de la ruta de quinasa con un buen grado de precisión.

#### 45 **Ejemplo 2 - Heterogeneidad de la fosforilación del grupo de sustrato a través de la AML primaria**

Para caracterizar la heterogeneidad de señalización de quinasas en AML primaria, se midió el enriquecimiento del grupo de sustrato a través de los 28 casos de AML mostrados en la Figura 7. Los datos de enriquecimiento se agruparon en base a cuatro fuentes diferentes de relaciones quinasa-sustrato, concretamente phosphoElm, phosphoPoint, phosphoSite, y la colección del inventor de motivos de fosforilación. La agrupación jerárquica no supervisada de los datos de enriquecimiento mostró doce grupos principales.

Las bases de datos phosphoElm, phosphoSite y phosphoPoint usan una nomenclatura inconsistente para nombrar proteínas quinasas, y éstas enumeran diferentes conjuntos de sustratos por quinasas (con solapamiento solo parcial). Sin embargo, una inspección minuciosa de los grupos de enriquecimiento de sustrato indicó que los datos de enriquecimiento fueron en general consistentes en la LMA primaria, independientemente de la fuente de los grupos de sustrato utilizados para KSEA. Por ejemplo, los grupos de sustrato llamados grupo MAPK\_, MAPK3 y ERK1 en phosphoElm, phosphoPoint y phosphoSite, respectivamente, todos los cuales consisten en sustratos de proteína quinasas activadas por mitógeno, agrupados en el Grupo 1. De manera similar, el Grupo 1 también contenía sustratos para grupo CDK\_ (phosphoElm), CDC2 (phosphoPoint) y CDK1 (phosphoSite), todos los cuales son nombres diferentes para la proteína quinasa 1 dependiente de ciclina. Estos datos son consistentes con MAPK y CDK que tienen un espectro similar de sustratos motivo y con su corregulación, y sugieren que estas dos actividades de quinasa se expresan conjuntamente en el panel AML probado. También fue interesante observar que DNA-PK (denominado como PRKDC, su nombre de gen, en phosphoPoint) y los grupos de sustrato ATM agrupados con el grupo definido por el motivo xSQx; estos grupos de sustrato se muestran en el Grupo 11. Estos datos son consistentes con la especificidad de sustrato conocida de DNA-PK y ATM. Otras asociaciones representativas

incluyen el enriquecimiento de sustratos de proteína tirosina quinasa en el Grupo 7, en el que los sustratos Btk, Syk y Lyn mostraron patrones de enriquecimiento similares independientemente de la base de datos de grupo de sustrato utilizada para el análisis (Fig. 7). El motivo fosfotirosina se agrupó con estos grupos de sustrato, aunque otras tirosina quinasa se agruparon con otros grupos. También fue interesante observar que los sustratos de caseína quinasa 2 $\alpha$  (denominados como CSNK2A1, CSNK2A2 por phosphoPoint y CK2 por phosphoElm y fosfoSite) se agruparon en el grupo 5, que también contenía varios motivos de fosforilación ricos en residuos ácidos a temperatura ambiente. Del mismo modo, el enriquecimiento de la mayoría de los grupos de sustratos definidos por motivos básicos agrupados con los definidos por la proteína quinasa A y la proteína quinasa C (Grupo 2). Estos datos son consistentes con las especificidades de sustrato conocidas de PKA, PKC y caseína quinasa e indican además que las actividades de CK2 $\alpha$  y quinasa basófila se expresan diferencialmente en nuestro panel de AML primaria.

Además de las correlaciones positivas de grupos de sustratos definidos por quinasa y motivos, también se observaron correlaciones negativas en la expresión de ciertos grupos de quinasa y sustratos. Por ejemplo, los grupos sustrato enriquecidos en el grupo 5a que comprende motivos ácidos y sustratos CK2 $\alpha$  correlacionados negativamente con los de los grupos 9, 6 y 7 que comprenden sustratos PKB/PKC, PAK/MAPAPK2 y Lyn/Syk/Try, respectivamente (Fig. 7). Estos resultados indican que la actividad de CK2 $\alpha$  quinasa tiende a ser mutuamente excluyente con actividades basófilas y de tirosina quinasa en la LMA primaria.

La Fig. 7e muestra que los grupos de sustrato más frecuentemente enriquecidos en AML primaria incluyen aquellos para las tirosina quinasa Syk, Btk y Lyn. En cuanto a las serina/treonina quinasa, también se encontró que los sustratos de caseína quinasa y MAPKAPK2 se enriquecen consistentemente. Los motivos que mostraron un aumento más frecuente incluyeron aquellos en secuencias ácidas (motivos CK2, 11 casos), el motivo RxRxxS (10 casos), el motivo SQ (8 casos) y el motivo tirosina (6 casos).

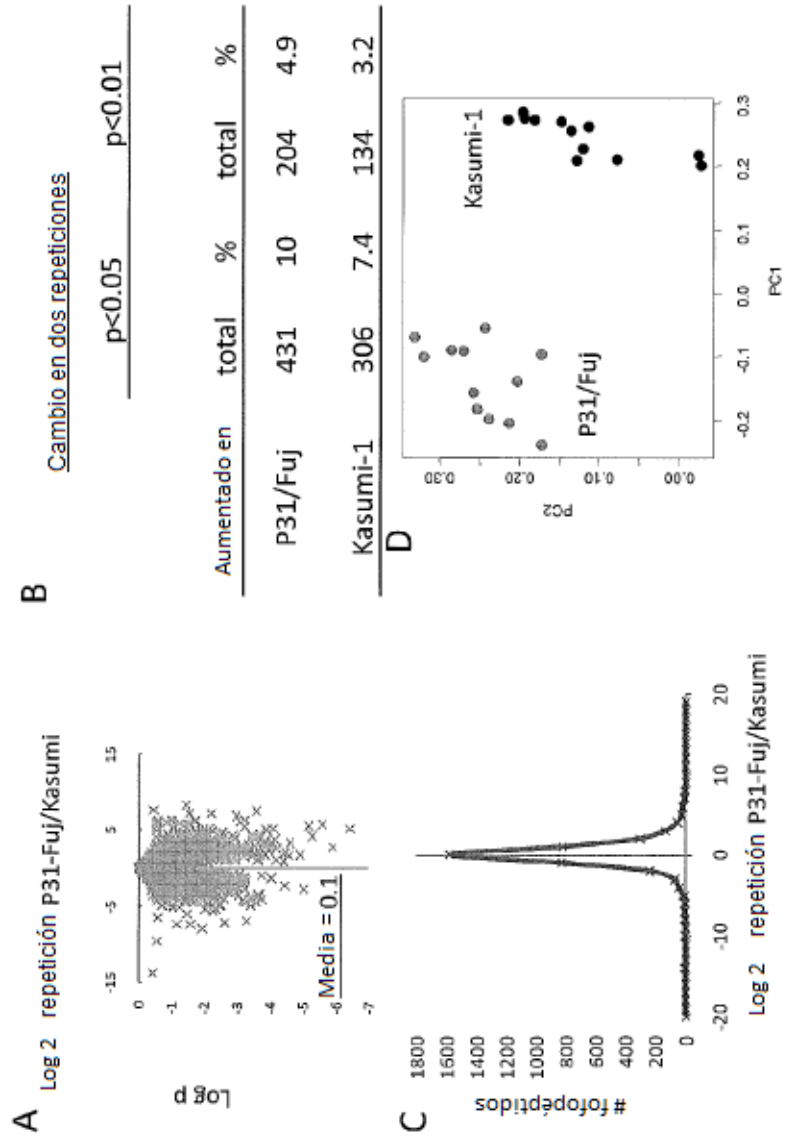
## REIVINDICACIONES

1. Un método in vitro para determinar la activación diferencial de una enzima en una primera muestra biológica con respecto a una segunda muestra biológica, en donde la enzima cataliza una reacción que implica la adición de un grupo funcional a una proteína o péptido, comprendiendo el método:
- (1) agrupar los péptidos modificados endógenos de la primera muestra biológica y los péptidos modificados endógenos de la segunda muestra biológica en un único grupo de acuerdo con uno de los siguientes parámetros:
- péptidos modificados endógenos que tienen un sitio de modificación que es modificado por la misma enzima; o
  - péptidos modificados endógenos que tienen un sitio de modificación que es parte de la misma secuencia específica de aminoácidos que se modifica en la misma posición por la misma enzima;
- (2) calcular el enriquecimiento de los péptidos modificados endógenos de la primera muestra biológica en comparación con los péptidos modificados endógenos de la segunda muestra biológica en el grupo, en el que se compara el aumento en abundancia o frecuencia de los péptidos modificados endógenos de la primera muestra biológica el aumento o disminución en la abundancia o frecuencia de los péptidos modificados endógenos de la segunda muestra biológica; y
- (3) calcular el significado estadístico de dicho enriquecimiento;
- en el que un enriquecimiento estadísticamente significativo es indicativo de que la enzima se activa en la primera muestra biológica en comparación con la segunda muestra biológica, y en el que el método comprende además identificar péptidos modificados endógenos en la primera muestra biológica y la segunda muestra biológica usando espectrometría de masas (MS) antes de la etapa (1), en el que la identificación de péptidos modificados endógenos en cada una de la primera muestra biológica y la segunda muestra biológica se lleva a cabo usando un método que comprende las siguientes etapas:
- (a) obtener péptidos endógenos a partir de la muestra biológica;
  - (b) añadir péptidos modificados de referencia en una cantidad conocida a los péptidos obtenidos en la etapa (a) para producir una mezcla de péptidos endógenos y péptidos modificados de referencia;
  - (c) llevar a cabo espectrometría de masas (MS) en dicha mezcla de péptidos endógenos y péptidos modificados de referencia y normalizar la señal de los péptidos modificados endógenos a la señal de los péptidos modificados de referencia para obtener datos relativos a los péptidos endógenos en la muestra biológica; y
  - (d) comparar los datos relacionados con los péptidos endógenos en la muestra biológica con datos en una base de datos de péptidos modificados utilizando un programa informático;
- en el que la base de datos de péptidos modificados se compila por un método que comprende:
- i obtener péptidos a partir de una muestra biológica;
  - ii enriquecer péptidos modificados a partir de los péptidos obtenidos en el paso i;
  - iii realizar cromatografía líquida-espectrometría de masas en tándem (LC-MS/MS) sobre los péptidos modificados enriquecidos obtenidos en la etapa ii;
  - iv comparar los péptidos modificados detectados en la etapa iii con una base de datos de referencia conocida para identificar los péptidos modificados; y
  - v compilar datos relacionados con los péptidos modificados identificados en el paso iv en una base de datos.
2. El método de acuerdo con la reivindicación 1, en el que el enriquecimiento se calcula por medio de:
- contar el número de péptidos modificados endógenos en el grupo que son más abundantes en la primera muestra biológica que en la segunda muestra biológica y el número de péptidos modificados endógenos en el grupo que son menos abundantes en la primera muestra biológica que en la segunda muestra biológica muestra; y restando el número de péptidos modificados endógenos en el grupo que son menos abundantes en la primera muestra biológica que en la segunda muestra biológica a partir del número de péptidos modificados endógenos en el grupo que son más abundantes en la primera muestra biológica que en la segunda muestra biológica muestra, o
  - calcular la abundancia media de todos los péptidos modificados endógenos en el grupo de la primera muestra biológica y la abundancia media de todos los péptidos modificados endógenos en el grupo de la segunda muestra

biológica; y dividir la abundancia media de todos los péptidos modificados endógenos en el grupo de la primera muestra biológica por la abundancia media de todos los péptidos modificados endógenos en el grupo de la segunda muestra biológica, o

- 5 • calcular la media de todos los cambios de repetición de los péptidos modificados endógenos en el grupo y la media de todos los cambios de repetición de los péptidos modificados endógenos en todo el experimento; y luego dividiendo la media de todos los cambios de repetición de los péptidos modificados endógenos en el grupo por cambio de repetición medio de todos los péptidos modificados endógenos en todo el experimento.
- 10 3. El método de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que el significado estadístico de dicho enriquecimiento se calcula mediante la prueba hipergeométrica, la prueba t pareada, la prueba Z, la prueba Kolmogorov-Smirnov, la prueba Chi-cuadrado, la prueba exacta de Fisher, o la prueba de rango con signo de Wilcoxon y opcionalmente comprende realizar una corrección de prueba múltiple, opcionalmente en el que la corrección de prueba múltiple es el método de Tasa de Descubrimiento Falso (FDR) de Benjamini Hochberg.
- 15 4. El método de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que la enzima es una proteína quinasa.
- 20 5. Un método como se reivindica en una cualquiera de las reivindicaciones precedentes, en el que la etapa (b) comprende además enriquecer péptidos modificados de dicha mezcla de péptidos endógenos y péptidos modificados de referencia para producir una mezcla de péptidos modificados enriquecidos y la etapa (c) comprende llevar a cabo espectrometría de masas (MS) en dicha mezcla de péptidos modificados enriquecidos para obtener datos relativos a los péptidos modificados endógenos en la muestra.
- 25 6. Un método como se reivindica en la reivindicación 5, en el que:
- la etapa de enriquecimiento de péptidos modificados se lleva a cabo usando cromatografía, opcionalmente en la que la cromatografía se selecciona del grupo que consiste en cromatografía de afinidad de iones metálicos inmovilizados (IMAC), cromatografía en dióxido de titanio (TiO<sub>2</sub>) y cromatografía en dióxido de zirconio (ZrO<sub>2</sub>); o
- 30 • la etapa de enriquecimiento de péptidos modificados se lleva a cabo utilizando métodos basados en anticuerpos.
7. Un método como se reivindica en una cualquiera de las reivindicaciones precedentes, en el que:
- 35 • los datos relacionados con los péptidos en la muestra biológica comprenden la relación de masa a carga (m/z), carga (z) y tiempo de retención relativo de los péptidos; y/o
- dicha espectrometría de masas (MS) en la etapa (c) es cromatografía líquida-espectrometría de masas (LC-MS).
- 40 8. Un método como se reivindica en una cualquiera de las reivindicaciones precedentes, en el que:
- la etapa ii se lleva a cabo usando cromatografía multidimensional, opcionalmente en la que la cromatografía multidimensional se lleva a cabo usando cromatografía líquida de alta resolución de intercambio catiónico fuerte (SCX-HPLC), cromatografía de afinidad de iones metálicos inmovilizados (IMAC) y cromatografía en dióxido de titanio (TiO<sub>2</sub>) o cromatografía líquida de alta resolución de intercambio de aniones (SAX-HPLC), cromatografía de afinidad de iones metálicos inmovilizados (IMAC) y cromatografía en dióxido de titanio (TiO<sub>2</sub>); o
- 45 • el paso ii se lleva a cabo utilizando métodos basados en anticuerpos.
- 50 9. Un método como se reivindica en una cualquiera de las reivindicaciones precedentes, en el que la etapa iv se lleva a cabo usando el motor de búsqueda MASCOT.
10. Un método como se reivindica en una cualquiera de las reivindicaciones precedentes, en el que los datos relativos a los péptidos modificados identificados en la etapa iv se seleccionan del grupo que consiste en la identidad del péptido modificado, relación masa a carga (m/z), carga (z) y el tiempo de retención relativo del péptido modificado.
- 55 11. El método de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que la técnica MS usa marcadores de isótopos para la cuantificación.
- 60 12. El método de acuerdo con la reivindicación 11, en el que la técnica MS usa marcación metabólica o derivación química.
- 65 13. El método de acuerdo con la reivindicación 12, en el que la marcación metabólica comprende usar aminoácidos marcados con isótopos estables en cultivo (SILAC) o en el que la derivación química es iTRAQ, ICAT o TMT.

Figura 1



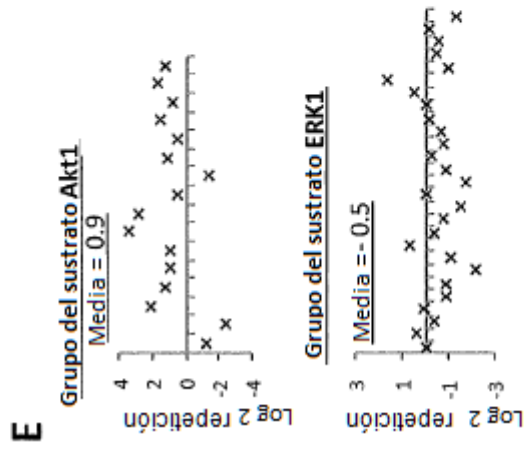
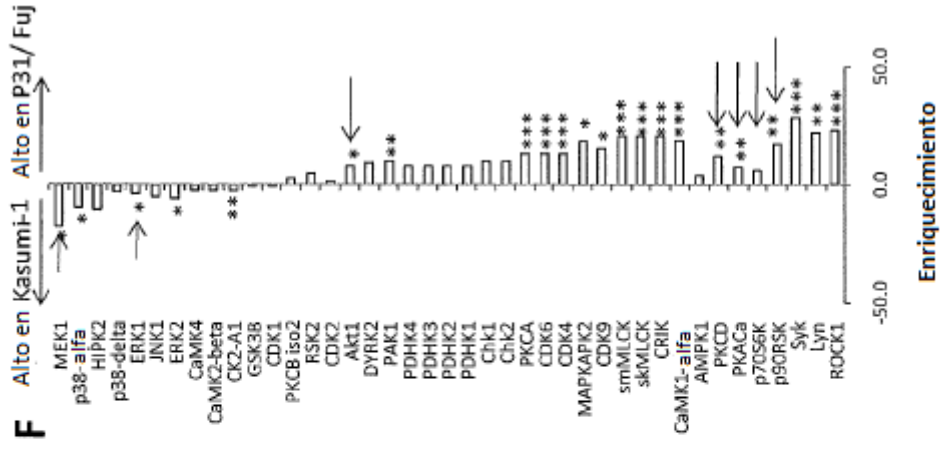


Figura 2

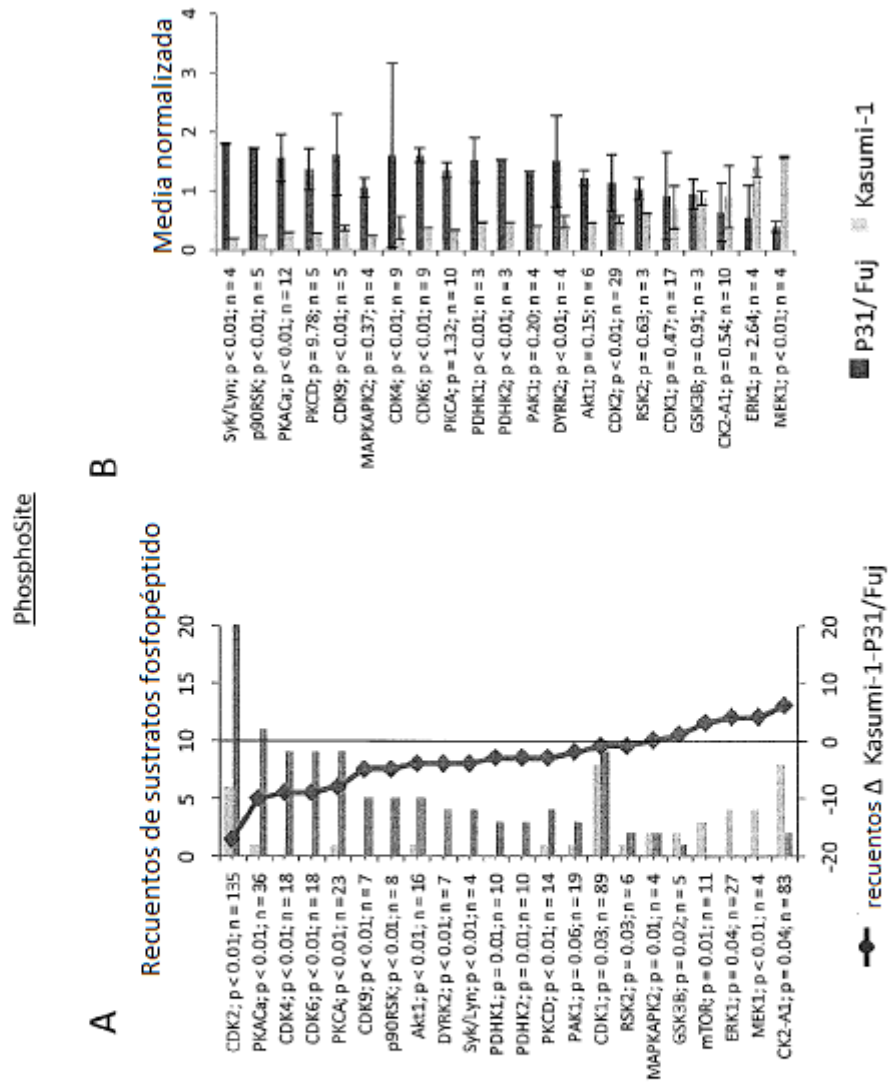


Figura 3

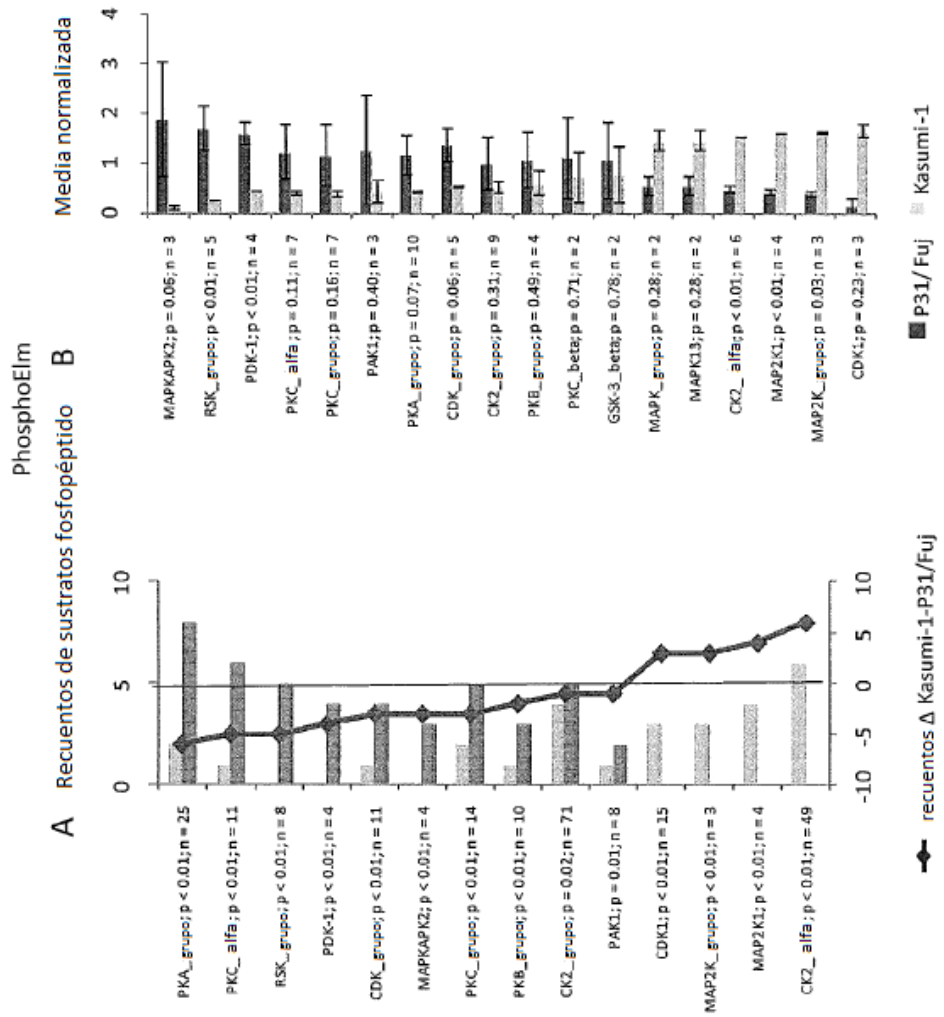




Figura 4

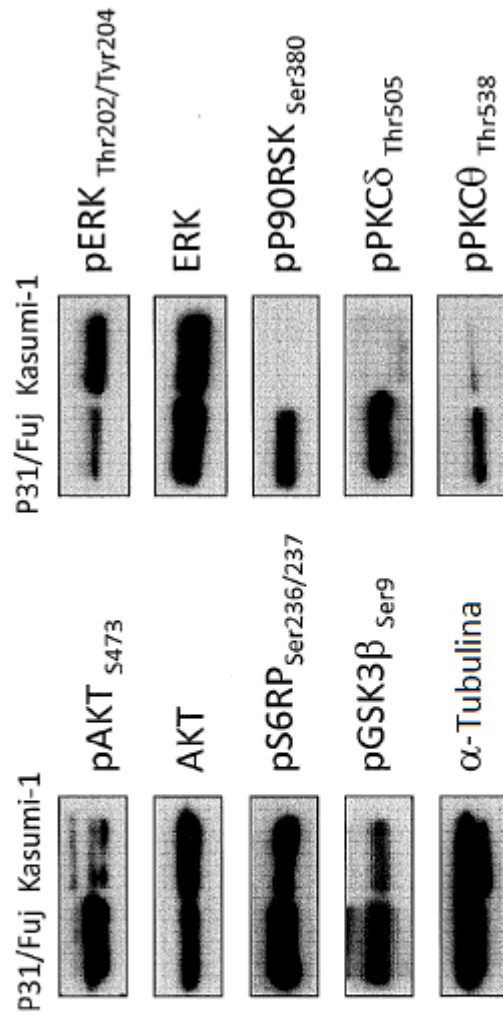
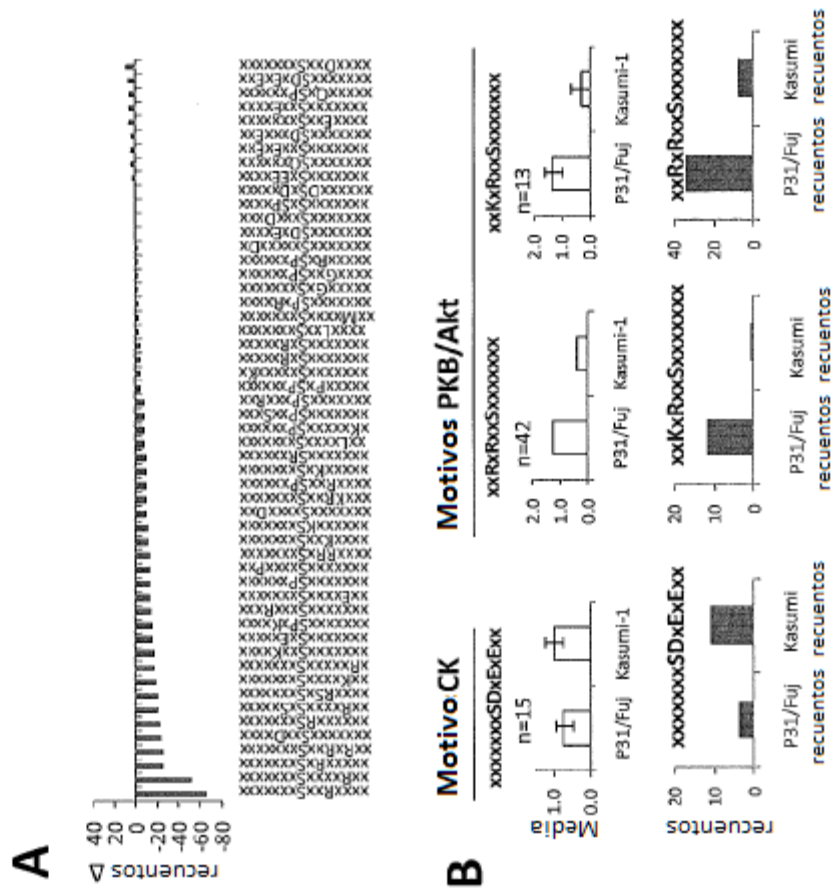


Figura 5



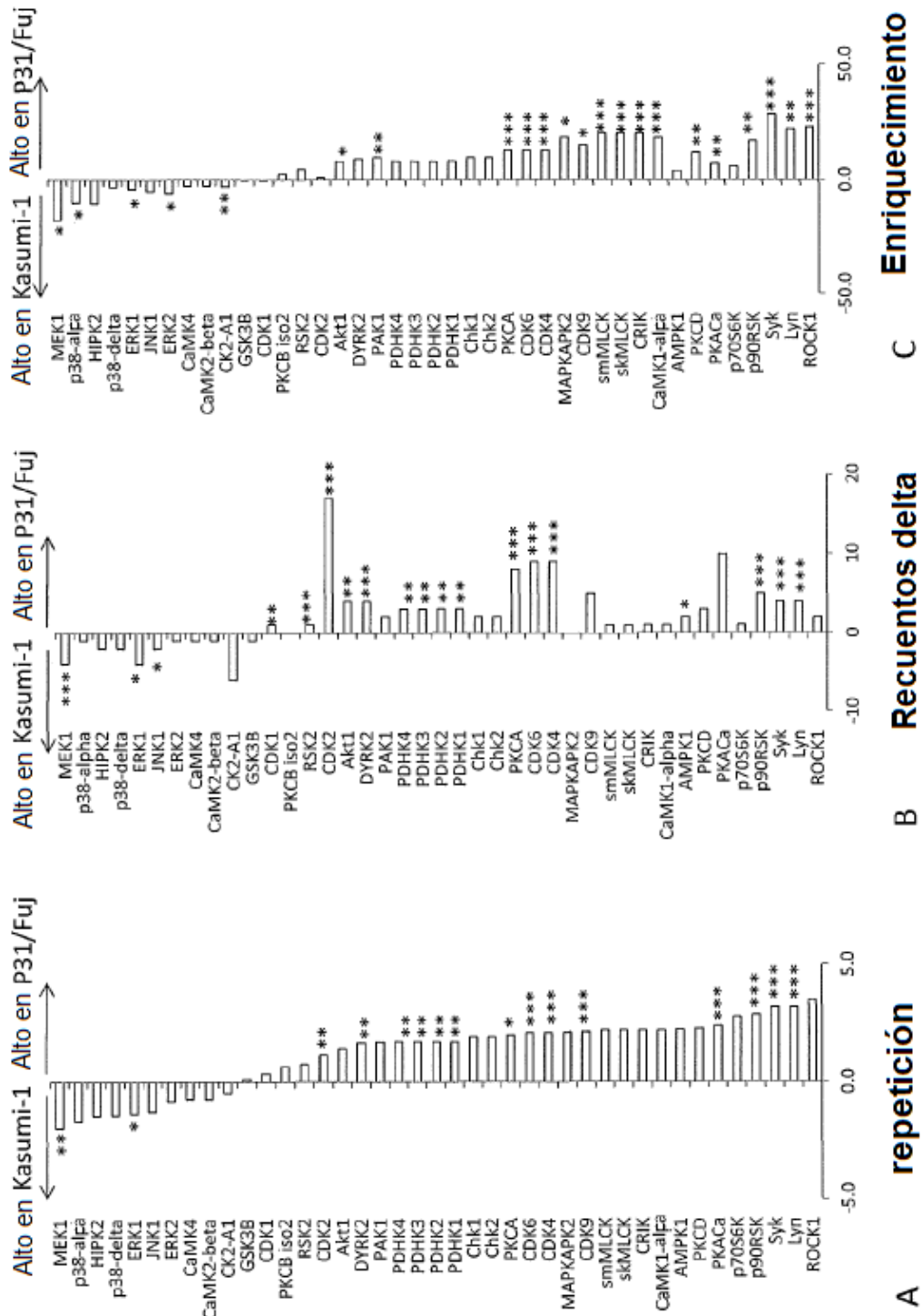


Figura 6

Figura 7

