

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 745 811**

51 Int. Cl.:

G16H 50/30 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.10.2006 PCT/US2006/040294**

87 Fecha y número de publicación internacional: **19.04.2007 WO07044944**

96 Fecha de presentación y número de la solicitud europea: **13.10.2006 E 06825991 (0)**

97 Fecha y número de publicación de la concesión europea: **24.07.2019 EP 1949285**

54 Título: **Sistemas y métodos para tratar, diagnosticar y predecir la aparición de una afección médica**

30 Prioridad:

13.10.2005 US 726809 P

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

03.03.2020

73 Titular/es:

**FUNDAÇÃO D. ANNA SOMMER CHAMPALIMAUD
E DR. CARLOS MONTEZ CHAMPALIMAUD
(100.0%)
Avenida Brasilia
1400-038 Lisboa, PT**

72 Inventor/es:

**SAIDI, OLIVIER;
VERBEL, DAVID, A. y
TEVEROVSKIY, MIKHAIL**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 745 811 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

Sistemas y métodos para tratar, diagnosticar y predecir la aparición de una afección médica

Campo de la invención

5 La invención se refiere a métodos y sistemas que usan información clínica, información molecular e información morfológica generada por ordenador en un modelo predictivo para predecir la aparición de una afección médica (por ejemplo, una enfermedad o sensibilidad o insensibilidad al tratamiento).

Antecedentes

10 Se requiere que los médicos tomen muchas decisiones médicas que van, por ejemplo, desde si un paciente puede experimentar una afección médica y cuándo, hasta cómo se debe tratar al paciente una vez que se le ha diagnosticado la afección. Determinar un curso de tratamiento apropiado para un paciente puede aumentar sus posibilidades, por ejemplo, de supervivencia y/o recuperación. Del mismo modo, predecir la aparición de un evento permite ventajosamente a las personas planificar el evento. Por ejemplo, predecir si es probable que un paciente experimente la aparición (por ejemplo, recurrencia) de una enfermedad puede permitir a un médico recomendar un curso de tratamiento apropiado para ese paciente.

15 Tradicionalmente, los médicos dependen en gran medida de su experiencia y entrenamiento para tratar, diagnosticar y predecir la aparición de afecciones médicas. Por ejemplo, los patólogos usan el sistema de puntuación de Gleason para evaluar el nivel de avance y agresión del cáncer de próstata, en el que el cáncer se clasifica según la apariencia del tejido de la próstata al microscopio según lo percibe un médico. Se otorgan puntuaciones de Gleason más altas a las muestras de tejido prostático que están más indiferenciadas [1]. Aunque los patólogos consideran que la clasificación de Gleason es fiable, es un sistema de puntuación subjetivo. Particularmente, diferentes patólogos que ven las mismas muestras de tejido pueden hacer interpretaciones contradictorias.

20 Las herramientas convencionales para ayudar a los médicos en el diagnóstico médico tienen un alcance y una aplicación limitados. Por ejemplo, las herramientas para ayudar a los médicos con las decisiones sobre el tratamiento del cáncer de próstata después de que un paciente se haya sometido a una prostatectomía radical se limitan a los ensayos de cribado de PSA en suero y los nomogramas generalizados. Un nomograma postoperatorio, desarrollado por Kattan et al., patente de EE.UU. nº 6.409.664, es ampliamente utilizado por los urólogos y permite predecir la probabilidad de recurrencia de la enfermedad a los 7 años para pacientes tratados con prostatectomía radical. Este nomograma proporciona información sobre la probabilidad de fallo bioquímico solamente (es decir, un aumento del nivel de PSA) y no predice el fallo clínico (muerte). Además, este nomograma solo predice si es probable que la afección de un paciente se repita dentro de los 7 años, y no predice cuándo podría reaparecer la afección del paciente en ese intervalo. Las variables de pronóstico utilizadas en este nomograma incluyen los niveles de PSA en suero previos al tratamiento, la puntuación de Gleason y la evaluación microscópica por un patólogo de la invasión capsular de la próstata, los márgenes quirúrgicos, la invasión de las vesículas seminales y el estado de los ganglios linfáticos. El fracaso del tratamiento se registra cuando hay evidencia clínica de recurrencia de la enfermedad, un aumento del PSA en suero o el inicio de la terapia adyuvante. Sin embargo, estos nomogramas tienen varias limitaciones, incluido que cada uno de ellos solo utiliza variables clínicas disponibles de forma rutinaria. Una de las limitaciones más notables es que incluso el mejor de estos nomogramas funciona solo un poco mejor que el punto intermedio entre un modelo con una discriminación perfecta (índice de concordancia = 1,0) y un modelo sin capacidad de discriminación (índice de concordancia = 0,5). Además, el resultado para aproximadamente el 30% de los pacientes que tienen predicciones de nomograma en el rango medio (supervivencia sin progresión de 7 años, 30-70%) es incierto, ya que la predicción no es más precisa que un lanzamiento de moneda.

45 Han surgido técnicas en el procesamiento y análisis de imágenes implementados por ordenador que proporcionan una potencia computacional significativamente mayor. En muchas aplicaciones, la capacidad de extraer grandes cantidades de características cuantitativas de valor continuo automáticamente de una sola imagen se ha convertido en una realidad. Se dice que una característica X tiene un valor continuo si, para algunos $A < B$, el conjunto de valores para la característica incluye todos los números x entre A y B. Se han desarrollado sistemas de análisis de imágenes de cáncer para imágenes tomadas de muestras citológicas [2] [3]. Sin embargo, tales sistemas solo capturan células y, por lo tanto, no utilizan toda la información arquitectónica observable a nivel tisular, y mucho menos combinan esa información con información clínica y molecular. No se han proporcionado sistemas de análisis de imágenes del cáncer para analizar la estructura de diferentes elementos patológicos a nivel tisular, que a menudo juega un papel más importante en el diagnóstico (p. ej., en el análisis de Gleason) que la apariencia de células individuales. Por lo tanto, los patólogos han recurrido a técnicas manuales para analizar la forma y el tamaño de la glándula prostática para determinar el grado patológico del cáncer [4]. La deficiencia de los sistemas convencionales de análisis de imágenes de cáncer se ve exacerbada por el hecho de que las imágenes de tejidos son típicamente más complejas que las imágenes celulares, y requieren un conocimiento experto de dominio completo para entenderlos.

55 El documento US 6.409.664 se refiere a métodos y aparatos para predecir la probabilidad de recurrencia de la enfermedad después de una prostatectomía radical usando factores clínicos y patológicos predeterminados. Revela nomogramas que pueden usarse antes y después de la operación en pacientes diagnosticados con adenocarcinoma

prostático para ayudar en la selección de un curso de terapia apropiado.

El documento WO 2005/091203 describe métodos y sistemas que utilizan información clínica, información molecular e información morfométrica generada por ordenador en un modelo predictivo para predecir la aparición (por ejemplo, recurrencia) de una afección médica, por ejemplo, cáncer.

5 **Compendio de la invención**

De acuerdo con la invención, se proporciona un aparato para evaluar un riesgo de fallo clínico en un paciente posterior a que el paciente se haya sometido a una prostatectomía radical como se establece en la reivindicación 1, y un método para evaluar un riesgo de fallo clínico en un paciente posterior a que el paciente se haya sometido a una prostatectomía radical como se establece en la reivindicación 4. Otras características opcionales se exponen en las reivindicaciones dependientes.

Breve descripción de los dibujos

Para una mejor comprensión de las realizaciones de la presente invención, se hace referencia a la siguiente descripción, tomada junto con los dibujos adjuntos, en los que los caracteres de referencia similares se refieren a partes similares, y en los que:

15 Las Figuras 1A y 1B son diagramas de bloques de sistemas que utilizan un modelo predictivo para tratar, diagnosticar o predecir la aparición de una afección médica;

La Figura 1C es un diagrama de bloques de un sistema para generar un modelo predictivo;

La Figura 2 muestra resultados ilustrativos para un paciente que se pueden generar mediante un modelo predictivo;

La Figura 3 es un diagrama de flujo de las etapas ilustrativas involucradas en el procesamiento de imágenes de tejido;

20 La Figura 4 es un diagrama de flujo de las etapas ilustrativas involucradas en el cribado de un compuesto inhibidor de una afección médica; Las Figuras 5a y 5b muestran imágenes digitales en escala de grises de muestras de tejido prostático sano y anormal, respectivamente, después de la segmentación y clasificación de imágenes;

La Figura 6 muestra varias características morfométricas clínicas, moleculares y generadas por ordenador utilizadas por un modelo para predecir la recurrencia del cáncer de próstata;

25 Las Figuras 7a y 7b muestran muestras de tejido teñidas que demuestran la presencia de dos características moleculares, particularmente el receptor de andrógenos (AR) y CD34;

La Figura 8 es un gráfico de una curva de Kaplan-Meier que demuestra una clasificación de pacientes con riesgo bajo, riesgo intermedio o riesgo alto de experimentar recurrencia del cáncer de próstata según lo predicho por un modelo basado en las características de la Figura 6;

30 La Figura 9 muestra varias características morfométricas clínicas, moleculares y generadas por ordenador utilizadas por un modelo para predecir la recurrencia del cáncer de próstata;

La Figura 10 es un gráfico de una curva de Kaplan-Meier que demuestra una clasificación de pacientes con riesgo bajo, riesgo intermedio o riesgo alto de experimentar recurrencia del cáncer de próstata según lo predicho por un modelo basado en las características de la Figura 9;

35 La Figura 11 muestra varias características morfométricas clínicas, moleculares y generadas por ordenador utilizadas por un modelo para predecir la supervivencia global del cáncer de próstata;

La Figura 12 es un gráfico de una curva de Kaplan-Meier que demuestra una clasificación de pacientes con riesgo bajo, riesgo intermedio o riesgo alto de muerte debido a cualquier causa, según lo predicho por un modelo basado en las características de la Figura 11;

40 La Figura 13 muestra varias características morfométricas clínicas y generadas por ordenador utilizadas por un modelo para predecir una enfermedad agresiva posterior a que un paciente se haya sometido a una prostatectomía;

Las Figuras 14 y 15 muestran varias características morfométricas clínicas, moleculares y generadas por ordenador utilizadas por un modelo para predecir la recurrencia del cáncer de próstata; y

45 La Figura 16 muestra varias características clínicas y de imagen de tejido generadas por ordenador utilizadas por un modelo para predecir el fallo clínico en un paciente posterior a la prostatectomía radical.

Descripción detallada de las realizaciones preferidas

La presente descripción se refiere a métodos y sistemas que usan información morfométrica generada por ordenador sola o en combinación con información clínica y/o información molecular en un modelo predictivo para predecir la

aparición de una afección médica. Por ejemplo, la información morfométrica clínica, molecular y generada por ordenador se usa para predecir la recurrencia del cáncer de próstata. Las enseñanzas proporcionadas en la presente memoria se usan para predecir la aparición de otras afecciones médicas como, por ejemplo, otros tipos de enfermedades (por ejemplo, neoplasias epiteliales y mixtas, que incluyen neoplasias de mama, colon, pulmón, vejiga, hígado, páncreas, células renales y tejido blando) y la sensibilidad o insensibilidad de un paciente a una o más terapias (por ejemplo, fármacos). Los médicos u otras personas pueden utilizar estas predicciones para, por ejemplo, seleccionar un curso de tratamiento apropiado para un paciente y/o diagnosticar una afección médica en el paciente.

En la presente memoria se divulga una herramienta analítica que incluye una máquina de vectores de soporte (SVM) y/o se puede proporcionar una red neuronal que determina las correlaciones entre las características morfométricas clínicas, moleculares y generadas por ordenador y una afección médica. Las características correlacionadas pueden formar un modelo que puede usarse para predecir la aparición o recurrencia de la afección. Por ejemplo, se puede usar una herramienta analítica para generar un modelo predictivo basado en datos para una cohorte de pacientes cuyos resultados con respecto a una afección médica (por ejemplo, tiempo hasta la recurrencia del cáncer) se conocen al menos parcialmente. El modelo se puede utilizar para evaluar los datos de un nuevo paciente con el fin de predecir la aparición de la afección médica para el nuevo paciente. La herramienta analítica solo puede utilizar un subconjunto de los tres tipos de datos (p. ej., datos clínicos y morfométricos únicamente) para generar el modelo predictivo.

Los datos clínicos, moleculares y/o morfométricos utilizados en la presente memoria pueden incluir cualquier dato clínico, molecular y/o morfométrico que sea relevante para el diagnóstico, tratamiento y/o predicción de una afección médica. Las características analizadas para las correlaciones con la recurrencia y supervivencia del cáncer de próstata con el fin de generar modelos predictivos se describen a continuación en relación con, por ejemplo, las Tablas 1, 2, 4 y/o 6. Se entenderá que al menos algunas de estas características (por ejemplo, neoplasias epiteliales y mixtas) pueden proporcionar una base para desarrollar modelos predictivos para otras afecciones médicas (por ejemplo, de mama, colon, pulmón, vejiga, hígado, páncreas, células renales, y tejidos blandos). Por ejemplo, una o más de las características de las Tablas 1, 2, 4 y/o 6 pueden evaluarse para pacientes que tienen alguna otra afección médica y luego introducirse en una herramienta analítica que determina si las características se correlacionan con la afección médica. Las características que aumentan la capacidad del modelo de predecir la aparición de la afección médica pueden incluirse en el modelo final, mientras que las características que no aumentan (por ejemplo, o disminuyen) el poder predictivo del modelo pueden eliminarse de la consideración. El uso de las características de las Tablas 1, 2, 4 y/o 6 como base para desarrollar un modelo predictivo puede enfocar los recursos de los médicos, otras personas y/o equipos de procesamiento automatizado (por ejemplo, un sistema de análisis de imágenes de tejidos) en la obtención de los datos del paciente que tienen más probabilidades de estar correlacionados con el resultado y, por lo tanto, ser útiles en el modelo predictivo final. Además, las características determinadas para correlacionarse con la recurrencia y supervivencia del cáncer de próstata se muestran en las Figuras 6, 9 y 11. Se entenderá que estas características pueden incluirse directamente en modelos finales predictivos de recurrencia y/o supervivencia del cáncer de próstata, y/o usarse para desarrollar modelos predictivos para otras afecciones médicas.

Los datos morfométricos pueden incluir datos generados por ordenador que indican diversas propiedades estructurales y/o espectrales, por ejemplo, de muestras de tejido.

Los datos morfométricos pueden incluir datos de características morfométricas del estroma, citoplasma, núcleos epiteliales, núcleos de estroma, luz, glóbulos rojos, artefactos de tejidos, fondo de tejidos o una combinación de los mismos. Se proporciona un sistema de análisis de imágenes de tejido para obtener mediciones de las características morfométricas de una imagen de tejido. Tal sistema puede ser el sistema MAGIC™ que utiliza el programa informático Definiens Cellenger. Tal sistema puede recibir una imagen teñida con hematoxilina y eosina como entrada, y puede generar varias mediciones de características morfométricas para objetos patológicos de la imagen. Los detalles adicionales con respecto a los sistemas y métodos para obtener características morfométricas de una imagen se describen a continuación en relación con la Figura 3.

Las características clínicas pueden incluir o basarse en datos de uno o más pacientes, como la edad, la raza, el peso, la altura, el historial médico, el genotipo y el estado de la enfermedad, donde el estado de la enfermedad se refiere a las características de estadificación clínica y patológica y a cualquier otra característica clínica recopilada a mano específicamente para el proceso de la enfermedad. En general, los datos clínicos los recopila un médico durante el curso del examen de un paciente y/o el tejido o las células del paciente. Los datos clínicos también pueden incluir datos clínicos que pueden ser más específicos para un contexto médico particular. Por ejemplo, en el contexto del cáncer de próstata, los datos clínicos pueden incluir datos que indiquen la concentración en sangre del antígeno prostático específico (PSA), el resultado de un tacto rectal, la puntuación de Gleason y/u otros datos clínicos que pueden ser más específicos para el cáncer de próstata. En general, cuando cualquier característica (es decir, clínica, morfométrica y/o molecular) de las Tablas 1, 2, 4 y/o 6 y/o las Figuras 6, 9 y/o 11 se aplica a contextos médicos distintos de la próstata, pueden no considerarse las características de estas tablas y/o figuras que son más específicas para la próstata. Opcionalmente, las características más específicas del contexto médico en cuestión pueden sustituirse por las características específicas de la próstata. Por ejemplo, otras características/manifestaciones histológicas específicas de la enfermedad pueden incluir regiones de necrosis (p. ej., carcinoma ductal in situ para mama), tamaño, forma y patrón/distribución regional de células epiteliales (p. ej., mama, pulmón), grado de diferenciación (p. ej., diferenciación escamosa con cáncer de pulmón de células no microcíticas (CPNM, producción de mucina como se observa con varios adenocarcinomas de mama y colon)), distribución morfológica/microscópica

de las células (p. ej., revestimiento de conductos en cáncer de mama, revestimiento de bronquiolos en NSCLC), y el grado y tipo de inflamación (por ejemplo, que tienen características diferentes para mama y NSCLC en comparación con la próstata).

5 Las características moleculares pueden incluir o basarse en datos que indican la presencia, ausencia, aumento o disminución relativa o ubicación relativa de moléculas biológicas que incluyen ácidos nucleicos, polipéptidos, sacáridos, esteroides y otras moléculas pequeñas o combinaciones de las anteriores, por ejemplo, glicoproteínas y complejos proteína-ARN. Las ubicaciones en las que se miden estas moléculas pueden incluir glándulas, tumores, estroma y/u otras ubicaciones, y pueden depender del contexto médico particular. En general, los datos moleculares se recopilan utilizando técnicas biológicas y bioquímicas moleculares comunes que incluyen transferencias de Southern, Western y Northern, reacción en cadena de la polimerasa (PCR), inmunohistoquímica e inmunofluorescencia. Además, la hibridación in situ puede usarse para mostrar tanto la abundancia relativa como la ubicación de las características biológicas moleculares. Los métodos y sistemas ilustrativos para la hibridación in situ de tejidos se describen en el documento US 2005/0019779 A1.

15 Las Figuras 1A y 1B muestran sistemas ilustrativos que utilizan un modelo predictivo para predecir la aparición de una afección médica en un paciente. La disposición en la Figura 1A puede usarse cuando, por ejemplo, un laboratorio de diagnóstico médico proporciona soporte para una decisión médica a un médico u otra persona asociada con un dispositivo de acceso remoto. La disposición en la Figura 1B puede usarse cuando, por ejemplo, se proporciona un kit de análisis que incluye el modelo predictivo para su uso en una instalación como un hospital, otra instalación médica u otra ubicación adecuada.

20 Con referencia a la Figura 1A, el modelo predictivo 102 se encuentra en la instalación de diagnóstico 104. El modelo predictivo 102 puede incluir cualquier equipo, programa informático o combinación de los mismos adecuados para recibir datos de un paciente, evaluar los datos para predecir la aparición (por ejemplo, recurrencia) de una afección médica para el paciente y generar los resultados de la evaluación.

25 El modelo 102 puede usarse para predecir la sensibilidad de un paciente a una o más terapias particulares. La instalación de diagnóstico 104 puede recibir datos para un paciente desde el dispositivo de acceso remoto 106 a través del proveedor de servicios de Internet (ISP) 108 y las redes de comunicaciones 110 y 112, y puede introducir los datos en el modelo predictivo 102 para su evaluación. Por supuesto, son posibles otras disposiciones para recibir y evaluar datos para un paciente desde una ubicación remota (por ejemplo, a través de otra conexión, como una línea telefónica o el correo físico). El médico o individuo ubicado a distancia puede adquirir los datos para el paciente de cualquier manera adecuada, y puede usar el dispositivo de acceso remoto 106 para transmitir los datos al centro de diagnóstico 104. Los datos para el paciente se pueden generar al menos parcialmente en la instalación de diagnóstico 104 u otra instalación. Por ejemplo, la instalación de diagnóstico 104 puede recibir una versión digitalizada de una imagen teñida con hematoxilina y eosina del dispositivo de acceso remoto 106 u otro dispositivo y puede generar datos morfométricos para el paciente basados en la imagen. En otro ejemplo, las muestras de tejido reales se pueden recibir y procesar en la instalación de diagnóstico 104 para generar los datos morfométricos. En otros ejemplos, un tercero puede recibir una imagen o tejido para un nuevo paciente, generar datos morfométricos basados en la imagen o tejido y proporcionar los datos morfométricos a la instalación de diagnóstico 104. A continuación se describe una herramienta de procesamiento de imágenes adecuada para generar datos morfométricos a partir de imágenes de tejidos y/o muestras en relación con la Figura 3.

40 La instalación de diagnóstico 104 puede proporcionar los resultados de la evaluación a un médico o individuo asociado con el dispositivo de acceso remoto 106 a través, por ejemplo, de una transmisión al dispositivo de acceso remoto 106 a través del ISP 108 y las redes de comunicaciones 110 y 112 o de otra manera, tal como el correo físico o una llamada telefónica. Los resultados pueden incluir una "puntuación" diagnóstica (por ejemplo, una indicación de la probabilidad de que el paciente experimente uno o más resultados relacionados con la afección médica, como el tiempo previsto para la recurrencia del evento), información que indica una o más características analizadas mediante el modelo predictivo 102 como correlacionado con la afección médica, información que indica la sensibilidad y/o especificidad del modelo predictivo u otra información de diagnóstico adecuada, o una combinación de los mismos. Por ejemplo, la Figura 2 muestra un ejemplo de un informe para un paciente ficticio que el modelo predictivo puede generar. Como se muestra, el informe asigna la probabilidad de resultado del paciente (p. ej., recurrencia del cáncer de próstata; es decir, eje y) respecto del tiempo en meses (eje x). En este ejemplo, el paciente tiene una puntuación de "520" que lo ubica en una categoría de riesgo alto. Tal informe lo puede utilizar un médico u otra persona para ayudar a determinar un grado de tumor clínico-diagnóstico más refinado, desarrollar un medio eficaz para subclasificar a los pacientes y finalmente generar algoritmos de opciones de tratamiento más precisos (y apropiados) para el paciente individual. El informe también puede ser útil porque puede ayudar al médico o al individuo a explicar el riesgo al paciente.

55 El dispositivo de acceso remoto 106 puede ser cualquier dispositivo remoto capaz de transmitir y/o recibir datos de la instalación de diagnóstico 104, como, por ejemplo, un ordenador personal, un dispositivo inalámbrico como un ordenador portátil, un teléfono celular o un asistente digital personal (PDA), o cualquier otro dispositivo de acceso remoto adecuado. Se pueden incluir múltiples dispositivos de acceso remoto 106 en el sistema de la Figura 1A (por ejemplo, para permitir que una pluralidad de médicos u otras personas en una pluralidad correspondiente de ubicaciones remotas comuniquen datos con la instalación de diagnóstico 104), aunque solo se ha incluido un dispositivo de acceso remoto 106 en la Figura 1A para evitar complicar demasiado el dibujo. La instalación de

diagnóstico 104 puede incluir un servidor capaz de recibir y procesar comunicaciones hacia y/o desde el dispositivo de acceso remoto 106. Dicho servidor puede incluir un componente distinto de equipo informático y/o almacenamiento, pero también puede ser un programa informático o una combinación de equipo y programa informático. El servidor puede implementarse usando uno o más ordenadores.

- 5 Cada uno de los enlaces de comunicaciones 110 y 112 puede ser cualquier ruta de comunicaciones cableada o inalámbrica adecuada o una combinación de rutas tales como, por ejemplo, una red de área local, red de área extensa, red telefónica, red de televisión por cable, intranet o Internet. Algunas redes de comunicaciones inalámbricas adecuadas pueden ser un sistema global para redes de comunicaciones móviles (GSM), una red de acceso múltiple por división de tiempo (TDMA), una red de acceso múltiple por división de código (CDMA), una red Bluetooth o cualquier otra red inalámbrica adecuada.

10 La Figura 1B muestra un sistema en el que se proporciona un kit de análisis 122 que incluye el modelo predictivo de la presente invención para el uso en la instalación 124, que puede ser un hospital, un consultorio médico u otra ubicación adecuada. El kit de análisis 122 puede incluir cualquier equipo, programa informático o combinación de los mismos (por ejemplo, un ordenador personal) que esté adaptado para recibir datos de un paciente (por ejemplo, al menos uno de datos clínicos, morfométricos y moleculares), evaluar los datos del paciente con un modelo predictivo (p. ej., programado en la memoria del kit de análisis) y generar los resultados de la evaluación. Por ejemplo, el kit de análisis 122 puede incluir un medio legible por ordenador codificado con instrucciones ejecutables por ordenador para realizar las funciones del modelo predictivo. El modelo predictivo puede ser un modelo predeterminado previamente generado (por ejemplo, por otro sistema o aplicación como el sistema de la Figura 1C). El kit de análisis 122 puede incluir opcionalmente una herramienta de procesamiento de imágenes capaz de generar datos correspondientes a características morfométricas, por ejemplo, de una muestra o imagen de tejido. A continuación se describe una herramienta de procesamiento de imágenes adecuada en relación con la Figura 3. Alternativamente, el kit de análisis 122 puede recibir datos preempaquetados para las características morfométricas como entrada, por ejemplo, desde un dispositivo de entrada (por ejemplo, teclado) u otro dispositivo o ubicación. El kit de análisis 122 puede incluir opcionalmente una entrada para recibir, por ejemplo, actualizaciones para el modelo predictivo. El kit de análisis también puede incluir opcionalmente una salida para transmitir datos, como datos útiles para la facturación del paciente y/o el seguimiento del uso, a una instalación principal u otro dispositivo o ubicación adecuada. Los datos de facturación pueden incluir, por ejemplo, información del seguro médico para un paciente evaluado mediante el kit de análisis (por ejemplo, nombre, proveedor del seguro y número de cuenta). Dicha información puede ser útil cuando, por ejemplo, un proveedor del kit de análisis cobra por el kit por uso y/o cuando el proveedor necesita la información del seguro de los pacientes para presentar reclamaciones a los proveedores de seguros.

15 La Figura 1C muestra un sistema ilustrativo para generar un modelo predictivo. El sistema incluye la herramienta analítica 132 (por ejemplo, que incluye una máquina de vectores de soporte (SVM) y/o una red neuronal) y una base de datos 134 de pacientes cuyos resultados se conocen al menos parcialmente. La herramienta analítica 132 puede incluir cualquier equipo, programa informático o combinación de los mismos adecuados para determinar las correlaciones entre los datos de la base de datos 134 y una afección médica. El sistema de la Figura 1C también puede incluir la herramienta de procesamiento de imágenes 136 capaz de generar datos morfométricos basados, por ejemplo, en una versión digitalizada de una imagen de tejido teñida con hematoxilina y eosina, una muestra de tejido real o ambas. La herramienta 136 puede generar datos morfométricos para, por ejemplo, los pacientes conocidos cuyos datos están incluidos en la base de datos 134. A continuación se describe una herramienta de procesamiento de imágenes 136 adecuada en relación con la Figura 3.

20 La base de datos 134 puede incluir cualquier información adecuada del paciente, como datos de características clínicas, características morfométricas, características moleculares o una combinación de las mismas. La base de datos 134 también puede incluir datos que indiquen los resultados de los pacientes, tales como si los pacientes experimentaron una recurrencia de la enfermedad y cuándo. Por ejemplo, la base de datos 134 puede incluir datos sin censura para pacientes (es decir, datos para pacientes cuyos resultados se conocen completamente), como datos para pacientes que han experimentado una recurrencia de una afección médica. La base de datos 134 puede incluir de manera alternativa o adicional datos censurados para pacientes (es decir, datos para pacientes cuyos resultados no se conocen completamente), como datos para pacientes que no han mostrado signos de recurrencia de la enfermedad en una o más visitas de seguimiento a un médico. El uso de datos censurados en la herramienta analítica 132 puede aumentar la cantidad de datos disponibles para generar el modelo predictivo y, por lo tanto, puede mejorar ventajosamente la fiabilidad y el poder predictivo del modelo. A continuación se describen ejemplos de máquinas de vectores de soporte (SVM) y redes neuronales (NNci) que pueden utilizar datos censurados y no censurados.

25 La herramienta analítica 132 puede incluir una máquina de vectores de soporte (SVM) capaz de realizar la regresión de vectores de soporte en datos censurados (SVRc). Como se describe en el documento US 2005/0108753, en SVRc se proporciona una nueva función modificada de pérdida/penalización para su uso dentro de un SVM que puede permitir que el SVM utilice datos censurados. Los datos que incluyen características clínicas, moleculares y/o morfométricas de pacientes conocidos de la base de datos 134 pueden introducirse en la SVM para determinar los parámetros para un modelo predictivo. Los parámetros pueden indicar la importancia relativa de las características de entrada, y pueden ajustarse para maximizar la capacidad de la SVM de predecir los resultados de los pacientes conocidos. Se describen detalles adicionales sobre el uso de SVM para determinar las correlaciones de las características con una afección médica en [5] y [6].

El uso de SVRc en la herramienta analítica 132 puede incluir la obtención de la base de datos 134 de vectores no lineales multidimensionales de información indicativa del estado de los pacientes, donde al menos uno de los vectores carece de una indicación de un momento de aparición de un evento con respecto a un paciente correspondiente. La herramienta analítica 132 puede entonces realizar una regresión usando los vectores para producir un modelo basado en el núcleo que proporciona un valor de salida relacionado con una predicción del tiempo hasta el evento basado en al menos parte de la información contenida en los vectores de información. La herramienta analítica 132 puede usar una función de pérdida para cada vector que contiene datos censurados que es diferente de una función de pérdida utilizada por la herramienta 132 para vectores que comprenden datos sin censurar. Una muestra de datos censurados puede manejarse de manera diferente porque puede proporcionar solo "información unilateral". Por ejemplo, en el caso de la predicción del tiempo de supervivencia, una muestra de datos censurados generalmente solo indica que el evento no ha sucedido dentro de un tiempo determinado, y no hay indicación de cuándo ocurrirá después de ese tiempo, si es que ocurre.

La función de pérdida utilizada por la herramienta analítica 132 para datos censurados puede ser la siguiente:

$$Pérdida(f(x), y, s = 1) = \begin{cases} C_S^*(e - \varepsilon_S^*) & e > \varepsilon_S^* \\ 0 & -\varepsilon_S \leq e \leq \varepsilon_S^* \\ C_S(\varepsilon_S - e) & e < -\varepsilon_S \end{cases}$$

donde $e = f(x) - y$; y

$$f(x) = W^T \Phi(X) + b$$

es una función de regresión lineal en un espacio de características F . Aquí, W es un vector en F , y $\Phi(x)$ asigna la entrada x a un vector en F .

Por el contrario, la función de pérdida utilizada por la herramienta 132 para datos sin censurar puede ser:

$$Pérdida(f(x), y, s = 0) = \begin{cases} C_n^*(e - \varepsilon_n^*) & e > \varepsilon_n^* \\ 0 & -\varepsilon_n \leq e \leq \varepsilon_n^* \\ C_n(\varepsilon_n - e) & e < -\varepsilon_n \end{cases}$$

donde $e = f(x) - y$ y $\varepsilon_n^* \leq \varepsilon_n$ y $C_n^* \geq C_n$.

En la descripción anterior, las W y b se obtienen al resolver un problema de optimización, cuya forma general es:

$$\begin{aligned} \min_{w, b} & \frac{1}{2} W^T W \\ \text{s.t.} & y_i - (W^T \Phi(x_i) + b) \leq \varepsilon \\ & (W^T \Phi(x_i) + b) - y_i \leq \varepsilon \end{aligned}$$

Sin embargo, esta ecuación supone que el problema de optimización convexa siempre es factible, lo cual puede no ser el caso. Además, se desea permitir pequeños errores en la estimación de la regresión. Es por estas razones que se utiliza una función de pérdida para SVRc. La pérdida permite un margen de maniobra para la estimación de la regresión. Idealmente, el modelo construido calculará exactamente todos los resultados con precisión, lo que no es factible. La función de pérdida permite un rango de error desde el ideal, y este rango se controla mediante las variables de holgura ξ y ξ^* , y una penalización C . Los errores que se desvían del ideal, pero que están dentro del rango definido por ξ y ξ^* , se cuentan, pero su contribución es mitigada por C . Cuanto más errónea sea la instancia, mayor será la penalización. Cuanto menos errónea (más cercana al ideal) sea la instancia, menor será la penalización. Este concepto de aumentar la penalización con error da como resultado una pendiente, y C controla esta pendiente. Si bien se pueden usar varias funciones de pérdida, para una función de pérdida insensible a ε , la ecuación general se transforma en:

$$\begin{aligned} \min_{W, b} \quad & P = \frac{1}{2} W^T W + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & y_i - (W^T \Phi(x_i) + b) \leq \varepsilon + \xi_i \\ & (W^T \Phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, \quad i=1 \dots l \end{aligned}$$

Para una función de pérdida insensible a ε de acuerdo con la invención (con diferentes funciones de pérdida aplicadas a datos censurados y sin censurar), esta ecuación se convierte en:

$$\min_{W, b} \quad P_c = \frac{1}{2} W^T W + \sum_{i=1}^l (C_i \xi_i + C_i^* \xi_i^*)$$

$$\begin{aligned} \text{s.t. } y_i - (W^T \Phi(x_i) + b) &\leq \varepsilon_i + \xi_i \\ (W^T \Phi(x_i) + b) - y_i &\leq \varepsilon_i^* + \xi_i^* \\ \xi_i^{(*)} &\geq 0, \quad i=1 \dots l \end{aligned}$$

Donde

$$\begin{aligned} C_i^{(*)} &= s_i C_s^{(*)} + (1 - s_i) C_n^{(*)} \\ \varepsilon_i^{(*)} &= s_i \varepsilon_s^{(*)} + (1 - s_i) \varepsilon_n^{(*)} \end{aligned}$$

El criterio de optimización penaliza los puntos de datos cuyos valores y difieren de $f(x)$ en más de ε . Las variables de holgura, ξ , y ξ^* , corresponden a la magnitud de este exceso de desviación para desviaciones positivas y negativas, respectivamente. Este mecanismo de penalización tiene dos componentes, uno para datos sin censurar (es decir, sin censurar por la derecha) y otro para datos censurados. Aquí, ambos componentes están representados en forma de funciones de pérdida que se denominan funciones de pérdida insensibles a ε .

En el documento US 2005/0108753 se describen detalles adicionales con respecto a los sistemas y métodos para realizar la regresión de vectores de soporte con datos censurados (SVRc).

La herramienta analítica 132 puede incluir una red neuronal que sea capaz de utilizar datos censurados. Además, la red neuronal utiliza preferiblemente una función objetivo sustancialmente de acuerdo con una aproximación (por ejemplo, derivada) del índice de concordancia (CI) para entrenar un modelo asociado (NNci). Aunque el CI se ha utilizado durante mucho tiempo como un indicador de rendimiento para el análisis de supervivencia [7], el uso del CI para entrenar una red neuronal no se había propuesto previamente. La dificultad de usar el CI como una función objetivo de entrenamiento en el pasado es que el CI no es diferenciable, y no puede optimizarse mediante métodos basados en gradientes. Como se describe en el documento US2005/0197982, este obstáculo puede superarse utilizando una aproximación del CI como la función objetivo.

Por ejemplo, cuando la herramienta analítica 132 incluye una red neuronal que se usa para predecir la recurrencia del cáncer de próstata, la red neuronal puede procesar datos de entrada para una cohorte de pacientes cuyos resultados con respecto a la recurrencia del cáncer de próstata se conocen al menos parcialmente para producir un resultado. Las características particulares seleccionadas para la entrada a la red neuronal pueden seleccionarse mediante el uso del SVRc descrito anteriormente (por ejemplo, implementado con una máquina de vectores de soporte de la herramienta analítica 132) o utilizando otro proceso adecuado de selección de características. Un módulo de error de la herramienta 132 puede determinar un error entre el resultado y el resultado deseado correspondiente a los datos de entrada (por ejemplo, la diferencia entre un resultado predicho y el resultado conocido para un paciente). La herramienta analítica 132 puede usar una función objetivo sustancialmente de acuerdo con una aproximación del CI para calificar el rendimiento de la red neuronal. La herramienta analítica 132 puede adaptar las conexiones ponderadas (por ejemplo, la importancia relativa de las características) de la red neuronal basándose en los resultados de la función objetivo. Los detalles adicionales sobre la adaptación de las conexiones ponderadas de una red neuronal para ajustar las correlaciones de las características con un resultado previsto se describen en [8] y [9].

El índice de concordancia se puede expresar como:

$$CI = \frac{\sum_{(i,j) \in \Omega} I(\hat{t}_i, \hat{t}_j)}{|\Omega|}$$

en donde

$$I(\hat{t}_i, \hat{t}_j) = \begin{cases} 1 & \hat{t}_i > \hat{t}_j \\ 0 & \text{de lo contrario} \end{cases}$$

y puede basarse en comparaciones por pares entre las estimaciones pronósticas \hat{t}_i y \hat{t}_j para los pacientes i y j , respectivamente. En este ejemplo, Ω consiste en todos los pares de pacientes $\{i,j\}$ que cumplen las siguientes condiciones:

- ambos pacientes i y j experimentaron recurrencia, y el tiempo de recurrencia t_i del paciente i es más corto que el tiempo de recurrencia del paciente j t_j ; o
- solamente el paciente i experimentó recurrencia, y t_i es más corto que el tiempo de la visita de seguimiento t_j del paciente j .

El numerador del CI representa el número de veces que el paciente en el que se predijo que habría recurrencia antes mediante la red neuronal realmente experimentó recurrencia antes. El denominador es el número total de pares de pacientes que cumplen las condiciones predeterminadas.

En general, cuando se incrementa el CI, preferiblemente cuando se maximiza, el modelo es más preciso. Por lo tanto, al maximizar de manera sustancial el CI, o una aproximación del CI, se mejora el rendimiento de un modelo.

Una aproximación del CI es la siguiente:

$$C = \frac{\sum_{(i,j) \in \Omega} R(\hat{t}_i, \hat{t}_j)}{|\Omega|}$$

5 en donde

$$R(\hat{t}_i, \hat{t}_j) = \begin{cases} \left(-(\hat{t}_i - \hat{t}_j - \gamma) \right)^n : \hat{t}_i - \hat{t}_j < \gamma \\ 0 : \text{de lo contrario} \end{cases}$$

y donde $0 < \gamma \leq 1$ y $n > 1$. $R(\hat{t}_i, \hat{t}_j)$ puede considerarse como una aproximación a $I(-\hat{t}_i, -\hat{t}_j)$.

Otra aproximación del CI proporcionado por la presente invención que se ha demostrado empíricamente que logra mejores resultados es la siguiente:

10
$$C_\omega = \frac{\sum_{(i,j) \in \Omega} -(\hat{t}_i - \hat{t}_j) \cdot R(\hat{t}_i, \hat{t}_j)}{D}$$

en donde

$$D = \sum_{(i,j) \in \Omega} -(\hat{t}_i - \hat{t}_j)$$

es un factor de normalización. Aquí cada $R(\hat{t}_i, \hat{t}_j)$ está ponderado por la diferencia entre \hat{t}_i y \hat{t}_j . El proceso de minimizar el C_ω (o C) busca mover cada par de muestras en Ω para satisfacer $\hat{t}_i - \hat{t}_j > \gamma$ y así hacer $I(\hat{t}_i, \hat{t}_j) = 1$.

15 Cuando la diferencia entre los resultados de un par en Ω es mayor que el margen γ , este par de muestras dejará de contribuir a la función objetivo. Este mecanismo supera efectivamente el sobreajuste de los datos durante el entrenamiento del modelo y hace que la optimización se centre preferiblemente en mover solo más pares de muestras en Ω para satisfacer $\hat{t}_i - \hat{t}_j > \gamma$. La influencia de las muestras de entrenamiento se ajusta adaptativamente de acuerdo con las comparaciones por pares durante el entrenamiento. Obsérvese que el margen positivo γ en R es preferible
 20 para mejorar el rendimiento de la generalización. En otras palabras, los parámetros de la red neuronal se ajustan durante el entrenamiento calculando el CI después de que se hayan introducido todos los datos del paciente. La red neuronal luego ajusta los parámetros con el objetivo de minimizar la función objetivo y así maximizar el CI. Como se usó anteriormente, el sobreajuste generalmente se refiere a la complejidad de la red neuronal. Específicamente, si la red es demasiado compleja, la red reaccionará a los datos "ruidosos". El sobreajuste es arriesgado, ya que puede
 25 conducir fácilmente a predicciones que están mucho más allá del rango de los datos de entrenamiento.

En el documento US2005/0197982 se describen detalles adicionales con respecto a los sistemas y métodos para usar una función objetivo sustancialmente de acuerdo con una aproximación del CI para entrenar una red neuronal.

La Figura 3 es un diagrama de flujo de funciones ilustrativas de una herramienta de procesamiento de imágenes adecuada. Las funciones de la Figura 3 se relacionan principalmente con la segmentación de imágenes de tejido para
 30 clasificar objetos patológicos en las imágenes (por ejemplo, clasificar objetos como citoplasma, luz, núcleos, estroma, fondo, artefactos y glóbulos rojos). En un ejemplo, la herramienta de procesamiento de imágenes puede incluir un microscopio óptico que captura imágenes de tejidos con un aumento de 20X utilizando una cámara digital en color SPOT Insight QE (KAI2000) y que produce imágenes con 1600 x 1200 píxeles. Las imágenes pueden almacenarse como imágenes con 24 bits por pixel en formato Tiff. Tal equipo es solo ilustrativo, y se puede usar cualquier otro
 35 equipo adecuado para capturar imágenes sin apartarse del alcance de la presente invención. La herramienta de procesamiento de imágenes también puede incluir cualquier equipo, programa informático o combinación adecuada para segmentar y clasificar objetos en las imágenes capturadas, y luego medir las características morfométricas de los objetos. La herramienta de procesamiento de imágenes puede incluir el Definiens Cellenger Developer Studio (v. 4.0) adaptado para realizar la segmentación y clasificación, por ejemplo, de los diversos objetos patológicos descritos
 40 anteriormente y para medir diversas características morfométricas de estos objetos. Se describen detalles adicionales sobre el producto Definiens Cellenger en [10]. La herramienta de procesamiento de imágenes puede medir varias características morfométricas de los objetos, incluidas las características basadas en el espectro (características del canal rojo, verde, azul (RGB), como valores medios, desviaciones estándar, etc.), posición, tamaño, perímetro, forma (asimetría, compacidad, ajuste elíptico, etc.) y relaciones espaciales y de intensidad con objetos cercanos (contraste).
 45 La herramienta de procesamiento de imágenes puede medir estas características para cada instancia de cada objeto patológico identificado en la imagen, y puede generar estas características, por ejemplo, para la evaluación mediante el modelo predictivo 102 (Figura 1A), el kit de análisis 122 (Figura 1B) o la herramienta analítica 132 (Figura 1C). Opcionalmente, la herramienta de procesamiento de imágenes también puede generar un resumen estadístico general de la imagen para cada una de las características medidas. A continuación se describen detalles adicionales con
 50 respecto a la medición de las características morfométricas de los objetos patológicos clasificados en relación con las

Tablas 1 y 2. La siguiente es una descripción de las funciones que se muestran en la Figura 3 de la herramienta de procesamiento de imágenes.

5 *Segmentación inicial.* En una primera etapa, la herramienta de procesamiento de imágenes puede segmentar una imagen (por ejemplo, una imagen de micromatriz de tejido teñido con hematoxilina y eosina (TMA) o una tinción de hematoxilina y eosina de un corte de tejido completo) en pequeños grupos de píxeles contiguos conocidos como objetos. Estos objetos pueden obtenerse mediante un algoritmo de crecimiento de regiones que encuentra regiones contiguas basadas en la similitud de color y la regularidad de la forma. El tamaño de los objetos se puede variar ajustando algunos parámetros [11]. En este sistema, la unidad de procesamiento más pequeña suele ser un objeto, en lugar de un píxel. Por lo tanto, todos los cálculos y operaciones de características morfológicas pueden realizarse con respecto a los objetos. Por ejemplo, cuando se aplica un umbral a la imagen, los valores de las características del objeto están sujetos al umbral. Como resultado, todos los píxeles dentro de un objeto se asignan a la misma clase. El tamaño de los objetos puede controlarse para que sea de 10-20 píxeles en el nivel más fino. En base a este nivel, los niveles posteriores más altos y más gruesos se crean formando objetos más grandes a partir de los más pequeños en el nivel inferior.

15 *Extracción del fondo.* Después de la segmentación inicial, la herramienta de procesamiento de imágenes puede segmentar el núcleo del tejido de la imagen desde el fondo (región transparente del portaobjetos) utilizando el umbral de intensidad y la envolvente convexa. El umbral de intensidad es un valor de intensidad que separa los píxeles de la imagen en dos clases: "núcleo de tejido" y "fondo". Cualquier píxel con un valor de intensidad mayor o igual al umbral se clasifica como un píxel de "núcleo de tejido", de lo contrario, el píxel se clasifica como un píxel de "fondo". La envolvente convexa de un objeto geométrico es el conjunto convexo más pequeño (polígono) que contiene ese objeto. Un conjunto S es convexo si, cuando dos puntos P y Q están dentro de S , entonces todo el segmento de línea PQ también está en S .

25 *Segmentación gruesa.* En una etapa siguiente, la herramienta de procesamiento de imágenes puede volver a segmentar el primer plano (p. ej., núcleo de TMA) en regiones rugosas que corresponden a núcleos y espacios en blanco. Por ejemplo, la característica principal de los núcleos en las imágenes teñidas con hematoxilina y eosina es que están teñidas de azul en comparación con el resto de los objetos patológicos. Por lo tanto, la diferencia en los valores de intensidad de los canales rojo y azul ($R-B$) puede usarse como una característica distintiva. Particularmente, para cada objeto de imagen obtenido en el paso de segmentación inicial, se puede determinar la diferencia entre los valores promedio de intensidad de píxel rojo y azul. La relación longitud/anchura también se puede usar para determinar si un objeto debe clasificarse como área de núcleo. Por ejemplo, los objetos que caen por debajo de un umbral de característica ($R-B$) y por debajo de un umbral de longitud/anchura pueden clasificarse como área de núcleo. De manera similar, se puede usar un umbral del canal verde para clasificar los objetos en el núcleo de tejido como espacios en blanco. El estroma del tejido está dominado por el color rojo. La diferencia de intensidad d , la "proporción de rojo" $r = R/(R + G + B)$ y la desviación estándar del canal rojo σ_R de los objetos de imágenes se pueden usar para clasificar los objetos del estroma.

35 *Clasificación de espacios en blanco.* En la etapa de segmentación gruesa, las regiones del espacio en blanco pueden corresponder tanto a la luz (objeto patológico) como a artefactos (áreas de tejido roto) en la imagen. Los objetos de espacio en blanco más pequeños (área menor de 100 píxeles) suelen ser artefactos. Por lo tanto, la herramienta de procesamiento de imágenes puede aplicar un filtro de área para clasificarlos como artefactos.

40 *Desfusión y clasificación de núcleos.* En la etapa de segmentación gruesa, el área de núcleos a menudo se obtiene como regiones fusionadas contiguas que abarcan varios núcleos reales. Además, la región de núcleos también podría incluir un citoplasma circundante mal clasificado. Por lo tanto, puede ser necesario desfusionar estas áreas de núcleos fusionados para obtener núcleos individuales.

45 La herramienta de procesamiento de imágenes puede usar dos enfoques diferentes para desfusionar los núcleos. El primer enfoque puede basarse en un algoritmo de crecimiento de regiones que fusiona los objetos de la imagen que constituyen el área de núcleos bajo restricciones de forma (redondez). Se ha determinado que este enfoque funciona bien cuando la fusión no es grave.

50 En el caso de fusión grave, la herramienta de procesamiento de imágenes puede usar un enfoque diferente basado en el aprendizaje supervisado. Este enfoque implica el etiquetado manual de las áreas de núcleos por un experto (patólogo). Las características de los objetos de imágenes que pertenecen a los núcleos etiquetados pueden usarse para diseñar clasificadores estadísticos.

55 Para reducir el número de dimensiones del espacio de características, la selección de características se puede realizar en el conjunto de entrenamiento utilizando dos clasificadores diferentes: el clasificador bayesiano y el clasificador de los k vecinos más cercanos [12]. El método "dejando uno fuera" [13] puede usarse para la validación cruzada, y el algoritmo de búsqueda directa secuencial puede usarse para elegir las mejores características. Finalmente, pueden diseñarse dos clasificadores bayesianos con un número de características iguales a 1 y 5, respectivamente. Se puede suponer que las distribuciones condicionales de clases son gaussianas con matrices de covarianza diagonal.

La imagen de entrada puede incluir diferentes tipos de núcleos: núcleos epiteliales, fibroblastos, núcleos basales,

núcleos endoteliales, núcleos apoptóticos y glóbulos rojos. Dado que el número de núcleos epiteliales se considera típicamente como una característica importante en la clasificación de la extensión del tumor, puede ser importante distinguir los núcleos epiteliales de los demás. La herramienta de procesamiento de imágenes puede lograr esto clasificando los núcleos detectados en dos clases: núcleos epiteliales y "el resto" en función de las características de forma (excentricidad) y tamaño (área).

Como se describió anteriormente, la herramienta de procesamiento de imágenes puede medir varias características morfométricas posteriores a la segmentación y clasificación de objetos en la imagen mediante la herramienta. Estas características morfométricas pueden ser indicativas de una o más propiedades y/o estadísticas. Las propiedades del objeto pueden incluir propiedades espectrales (p. ej., valores medios del canal de color, desviaciones estándar y brillo) y propiedades estructurales/de forma (p. ej., área, longitud, ancho, compacidad, densidad). Las estadísticas pueden incluir desviación mínima, máxima, media y estándar, y pueden calcularse para cada propiedad de un objeto de imagen. Las Tablas 1 y 2 (adjuntas en la presente memoria) muestran varios ejemplos de características morfométricas que pueden medirse de acuerdo con la presente invención. Las características morfométricas de estas tablas se nombran usando una convención que indica las diversas propiedades y/o estadísticas medidas por estas características. La convención de nomenclatura particular mostrada en las Tablas 1 y 2 está adaptada del producto informático Definiens disponible comercialmente descrito anteriormente y, por lo tanto, lo entenderá un experto en la técnica.

Se entenderá que las características morfométricas generadas por ordenador que se muestran en las Tablas 1 y 2 son solo ilustrativas, y que se puede utilizar cualquier característica morfométrica generada por ordenador sin apartarse del alcance de la presente invención. Por ejemplo, las Tablas 1 y 2 incluyen diferentes conjuntos de características morfométricas. El conjunto reducido y modificado de características de la Tabla 2 (es decir, reducido y modificado en comparación con las características de la Tabla 1) resultó de una experimentación adicional en el campo de la recurrencia y supervivencia del cáncer de próstata desde el momento en que se realizó el estudio que involucra la Tabla 1. Particularmente, la experimentación adicional proporcionó información adicional con respecto a los tipos de características que pueden correlacionarse más con el resultado. Los inventores esperan que la experimentación continua y/o el uso de otro equipo, programa informático adecuado o una combinación de los mismos produzca otros diversos conjuntos de características generadas por ordenador (por ejemplo, un subconjunto de las características de la Tabla 1 (véanse las Tablas 10 y 11) o un subconjunto de las características de la Tabla 2) que pueden correlacionarse con estas y otras afecciones médicas.

En referencia a las Tablas 1 y 2, la característica "Lumen.StdDevAreaPxl", "Luz" indica un tipo de objeto de imagen, "StdDev" indica una estadística (desviación estándar) que se calculará utilizando todas las instancias de la "Luz" identificadas, y "AreaPxl" indica una característica de una instancia de objeto (área en forma de número de píxeles) a evaluar por la estadística. Una herramienta de procesamiento de imágenes puede medir características morfométricas para todos los objetos previamente segmentados y clasificados en la imagen. Por ejemplo, la herramienta de procesamiento de imágenes puede medir características morfométricas para objetos que incluyen "Fondo", "Citoplasma", "Núcleos Epiteliales", "Luz", "Estroma", "Núcleos de Estroma" y "Glóbulos Rojos". "Fondo" incluye partes de la imagen digital que no están ocupadas por el tejido. "Citoplasma" se refiere al citoplasma de una célula, que puede ser un área amorfa (p. ej., área rosa que rodea un núcleo epitelial en una imagen, por ejemplo, de tejido teñido con hematoxilina y eosina). "Núcleos Epiteliales" se refiere al núcleo presente dentro de las células epiteliales/células lumbales y basales de la unidad glandular, que aparecen como objetos "redondos" rodeados de citoplasma. "Luz" se refiere al espacio glandular central donde las células epiteliales depositan las secreciones, que aparecen como áreas blancas cerradas rodeadas de células epiteliales. Ocasionalmente, la luz puede llenarse con líquido prostático (que generalmente aparece rosado en el tejido teñido con hematoxilina y eosina) u otros "desechos" (por ejemplo, macrófagos, células muertas, etc.). Juntos, la luz y el citoplasma epitelial y los núcleos forman una unidad de glándula. "Estroma" se refiere a una forma de tejido conectivo con diferente densidad que mantiene la arquitectura del tejido prostático. El tejido de estroma está presente entre las unidades de glándula, y tiene una apariencia de rojo a rosa en el tejido teñido con hematoxilina y eosina. "Núcleos de Estroma" son células alargadas con cantidades nulas o mínimas de citoplasma (fibroblastos). Esta categoría también puede incluir células endoteliales y células inflamatorias, y los núcleos epiteliales también se pueden encontrar dispersos dentro del estroma si hay presente un cáncer. "Glóbulos Rojos" son pequeños objetos redondos y rojos generalmente ubicados dentro de los vasos (arterias o venas), pero también se pueden encontrar dispersos en todo el tejido.

"C2EN", en las tablas siguientes, es una proporción relativa del área del núcleo respecto del citoplasma. Cuanto más anaplásica/maligna es la célula epitelial, más área está ocupada por el núcleo y mayor es la proporción. "EN2SN" es el porcentaje o la cantidad relativa de células epiteliales respecto del estroma presentes en la imagen digital del tejido. "L2Core" es el número o área de luz presente dentro del tejido. Cuanto más alto es el grado de Gleason, más agresivo es el cáncer y, por lo tanto, hay presente menos cantidad de luz. En general, esto se debe a que las células epiteliales se replican de manera incontrolada cuando se produce el cáncer, lo que hace que la luz se llene con las células epiteliales.

La Figura 4 es un diagrama de flujo de etapas ilustrativas involucradas en el cribado de un compuesto inhibidor.

En la etapa 402, se puede obtener un primer conjunto de datos para un paciente que incluye uno o más datos clínicos, datos morfométricos y datos moleculares. Se puede administrar un compuesto de ensayo al paciente en la etapa 404.

Después de la etapa 404, se puede obtener un segundo conjunto de datos del paciente en la etapa 406. El segundo conjunto de datos puede incluir o no los mismos tipos de datos (es decir, características) incluidos en el primer conjunto de datos. En la etapa 408, el segundo conjunto de datos puede compararse con el primer conjunto de datos, donde un cambio en el segundo conjunto de datos después de la administración del compuesto de ensayo indica que el compuesto de ensayo es un compuesto inhibidor. La etapa 408 de comparación de los conjuntos de datos puede incluir, por ejemplo, comparar una salida generada por un modelo predictivo de la presente invención que responde a una entrada del primer conjunto de datos con una salida generada por el modelo predictivo que responde a una entrada del segundo conjunto de datos. Por ejemplo, el compuesto inhibidor puede ser un fármaco dado, y la presente invención puede determinar si el fármaco es eficaz como tratamiento médico para una afección médica.

Ahora se describirán varios ejemplos ilustrativos que demuestran la predicción de afecciones médicas. En un primer ejemplo, los datos clínicos y morfométricos predicen la recurrencia del cáncer de próstata. En un segundo ejemplo, los datos clínicos, morfométricos y moleculares predicen la recurrencia del cáncer de próstata y la supervivencia general. En un tercer ejemplo, se hace una predicción para la aparición de una enfermedad agresiva posterior a una prostatectomía del paciente. En un cuarto ejemplo, se hace una predicción para la toxicología hepática. En los ejemplos quinto y sexto, se hace una predicción para la recurrencia del cáncer de próstata. En un séptimo ejemplo, se usó una realización de la presente invención para predecir el fallo clínico después de la prostatectomía.

Descripción general del cáncer de próstata

El cáncer de próstata es una de las principales causas de muerte en hombres en los Estados Unidos, con 230.000 casos recién diagnosticados y casi 30.000 muertes en 2004. El uso extendido del cribado basado en suero con PSA ha ofrecido a los médicos la capacidad de detectar el cáncer de próstata en un estadio anterior (es decir, T1a-c, T2), ya sea localizado en la próstata o diseminado regionalmente, mientras que solo se detecta un pequeño porcentaje en la etapa metastásica. Los beneficios informados de la detección y el diagnóstico tempranos han ejercido una enorme presión tanto en el paciente como en el urólogo al seleccionar el curso del tratamiento. La necesidad de un pronóstico preciso es fundamental al seleccionar la intervención terapéutica inicial, ya que la mayoría de los tumores son poco activos y requieren una intervención mínima (es decir, "espera vigilante"), mientras que otros son más agresivos y se recomienda una intervención temprana (es decir, radioterapia/terapia hormonal/terapia adyuvante sistémica/colocación en ensayo clínico). Además, en un ensayo aleatorio que comparó la espera vigilante con la prostatectomía radical, solo se obtuvo un beneficio modesto de la cirugía (reducción del 6,6% en la mortalidad después de la prostatectomía), lo que sugiere que se necesitan mejores medidas de estratificación de pacientes para guiar la atención individualizada de los pacientes [14].

La historia natural del PCa vuelve a enfatizar los desafíos que enfrenta el paciente en el momento de su diagnóstico [15]. Aunque el cáncer de próstata de estadio temprano es curable con terapia local, aproximadamente del 25 al 40% de los hombres desarrollará PSA/recurrencia bioquímica (BCR). Para complicar aún más las cosas, un hombre con cáncer de próstata que haya tenido una recurrencia aún puede desarrollar una metástasis unos 8 años después del PSA/BCR (media de 8 años; mediana 5 años después de BCR), lo que sugiere que identificar a este grupo de pacientes en una etapa temprana en su régimen de tratamiento (tanto para predecir su tiempo hasta BCR como su propensión a desarrollar metástasis) es primordial para su supervivencia general. Desafortunadamente, los modelos predictivos existentes son limitados en su precisión, y no están individualizados para el paciente específico con respecto a su patología tumoral. Aunque una variedad de cambios genéticos, ambientales y de estilo de vida se han implicado en la patogénesis del PCa, en la actualidad no existe una vía bioquímica, mutación genética o biomarcador clínico que pueda predecir el resultado de un paciente dado. Veintiún años después de que la prostatectomía radical se volviera popular nuevamente y 15 años después del uso generalizado de PSA, los urólogos aún no pueden decir a los pacientes qué tratamiento para la enfermedad localizada da como resultado la mejor supervivencia clínica sin enfermedad o supervivencia global.

Los nomogramas de pronóstico basados solo en datos de características clínicas, de hecho, proporcionan predicciones útiles de estados clínicos y resultados, pero necesitan una mejora tanto en la precisión como en la universalidad [16]. La presente descripción proporciona un enfoque de 'Patología de sistemas' para mejorar con éxito la precisión de un modelo predictivo para PSA/BCR después de la prostatectomía. Esto representa una visión 'individualizada' de la propia muestra de tumor del paciente, que incluye la evaluación cuantitativa de las características morfométricas celulares y microanatómicas, perfiles clínicos y marcadores moleculares para crear un modelo de predicción altamente preciso e integrador. Al utilizar la experiencia en el dominio, se han desarrollado modelos altamente precisos para predecir la recurrencia de PSA. Estos esfuerzos han validado la utilidad de la patología de sistemas en la generación de modelos predictivos y pronósticos. Además, el análisis demuestra que un conjunto limitado de variables clínicas, biomarcadores moleculares y características morfométricas tisulares se pueden derivar e incluir en un análisis predictivo utilizado por urólogos/patólogos para construir planes óptimos de tratamiento del paciente basados en un resultado clínico designado. Las características moleculares seleccionadas que se asociaron con la recurrencia de PSA sugieren papeles convergentes para los mecanismos de señalización del factor de crecimiento (a través del receptor de andrógenos (en adelante "AR"), descrito a continuación) y la vascularización acoplada celular (a través de CD34). CD34 es una glucoproteína transmembrana que está presente en las células endoteliales que recubren los vasos del cuerpo humano. Se están realizando más estudios para comprender mejor estas observaciones y el impacto potencial en la predicción de la progresión del cáncer de próstata. También fue digna de mención la segmentación de imágenes seleccionadas y las características morfométricas que representan en parte

una puntuación de Gleason altamente precisa, no subjetiva y cuantitativa, además de varios descriptores de tejidos novedosos que fueron importantes en el desarrollo y la precisión del modelo. Las características morfométricas definidas relacionadas con el sistema de puntuación de Gleason incluyen en parte la apariencia general de las estructuras glandulares, la forma y el tamaño (composición citoplasmática) de las células epiteliales, los núcleos de las células epiteliales y la demostración de células epiteliales individuales mezcladas en el estroma.

La proteína del receptor de andrógenos (AR) recibe hormonas androgénicas naturales (testosterona y su metabolito 5-alfa reducido, dihidrotestosterona) después de que estas hormonas se sintetizan en las células de Leydig de los testículos masculinos. Particularmente, después de la síntesis, estas hormonas circulan por todo el cuerpo y se unen al AR. Los andrógenos, que actúan a través del receptor AR, estimulan el desarrollo de los genitales masculinos y las glándulas sexuales accesorias en el feto, la virilización y el crecimiento en la pubertad masculina, y el mantenimiento de la virilidad masculina y la función reproductiva en el adulto. El receptor de andrógenos, junto con otros receptores de hormonas esteroides, constituyen una familia de proteínas reguladoras de la transcripción trans-activas que controlan la transcripción génica a través de interacciones con secuencias genéticas específicas.

Los estudios sobre AR con respecto al cáncer de próstata han sugerido que puede existir una correlación positiva entre la presencia de receptores de andrógenos en las células cancerosas y su dependencia de la estimulación hormonal androgénica para el crecimiento. Por ejemplo, Sovak et al., patente de EE.UU. n° 6.472.415, propone que el crecimiento del cáncer de próstata en estadios tempranos está controlado por andrógenos y se puede detener, al menos temporalmente, por la privación de andrógenos. French y col., patente de EE.UU. n° 6.821.767, propone varias formas de medir la AR que pueden permitir el uso de ensayos de receptores de andrógenos en la evaluación diagnóstica del cáncer de próstata por parte de los médicos. Sin embargo, estos estudios no han propuesto el uso de mediciones de AR junto con modelos automatizados que predicen la aparición del cáncer de próstata, como se describe en la presente memoria.

Ejemplo 1: Predicción de la recurrencia del cáncer de próstata

Datos clínicos y morfométricos

Se extrajeron varias características morfométricas en bruto, inicialmente hasta quinientas, de cada imagen de tejido de próstata utilizando el sistema de análisis de imágenes de tejidos MAGIC que se basa en el programa informático Definiens Cellenger. El conjunto completo de características en bruto se eligió de manera escéptica para evitar ignorar las características potencialmente útiles. Sin embargo, no era probable que todas estas características morfométricas fueran igualmente informativas, y un modelo de predicción basado en el conjunto completo de características probablemente tendría un rendimiento predictivo deficiente debido a la "maldición de la dimensionalidad" [13]. Entonces se aplicó un procedimiento de reducción de dimensionalidad y finalmente se seleccionó un conjunto de ocho características morfométricas.

Se realizó un estudio basado en un subconjunto de 153 pacientes de una cohorte de pacientes con cáncer de próstata que se sometieron a prostatectomía radical. El antígeno prostático específico (PSA) medible después de la operación se utilizó para definir la recurrencia del cáncer de próstata (también conocida como recurrencia bioquímica (BCR)). Los pacientes se siguieron después de la operación. Se registró su estado de recurrencia en su última visita, así como su tiempo de seguimiento, lo que generó un conjunto de datos censurados por la derecha. Las puntuaciones de Gleason se midieron tanto antes de la operación de la muestra de biopsia como después de la operación utilizando la glándula prostática extirpada. Las cuatro medidas clínicas específicas, o características, consideradas en este estudio fueron (1) el grado de Gleason de la biopsia, (2) la puntuación de Gleason de la biopsia, (3) el grado de Gleason postoperatorio y (4) la puntuación de Gleason postoperatoria.

Las características morfométricas se analizaron por separado de la característica de puntuación de Gleason obtenida clínicamente para predecir tanto la probabilidad como el tiempo hasta la recurrencia de PSA/BCR. La imagen y la puntuación de Gleason (características) se combinaron para establecer una predicción de tiempo de recurrencia y tiempo hasta la recurrencia. La precisión mejorada de la predicción lograda por este conjunto colectivo de características indicó que las características de imágenes realmente proporcionaron información adicional y, por lo tanto, mejoraron la tasa de predicción de recurrencia y el modelo de predicción general.

Debido a que esta cohorte de pacientes tuvo datos de resultados censurados por la derecha, se tuvieron que construir modelos de análisis de supervivencia para la predicción de la recurrencia. Para evitar el sesgo algorítmico potencial en diferentes tipos de datos, se utilizaron dos algoritmos de análisis de supervivencia: 1) un modelo de regresión de Cox [17]; y 2) SVRc que se describe anteriormente y se aplica a una máquina de vectores de soporte. El índice de concordancia estimado mediante la validación cruzada 5 veces se utilizó para medir la precisión predictiva de los modelos [13] [18].

Ambos algoritmos se aplicaron a tres conjuntos de datos: (1) las características clínicas de la puntuación de Gleason solo; (2) las características morfométricas seleccionadas solas; y (3) la combinación de las características morfométricas y las características clínicas de la puntuación de Gleason. Los resultados experimentales se enumeran en la Tabla 3.

Las características clínicas seleccionadas en este ejemplo fueron BXGGTOT, BXGG1, GGTOT y GG1 y las

características morfométricas seleccionadas relacionadas con los núcleos epiteliales (Epithelial.Nuclei.MaxCompactness), fondo (Background.Std-DevAreaPxl) y luz (Lumen.MaxBorderLengthPxl, Lumen.MinRadiusofsmallestenclosing, Lumen.StdDevBorder-LengthPxl, Lumen.SumBorderlengthPx1, Lumen.StdDevAreaPxl y Lumen.MinCompactness). Más particularmente, en este ejemplo, se determinó que las características morfométricas relacionadas con el área, la longitud del borde y la forma (compacidad) de la luz se correlacionan con la progresión de la enfermedad. Cuanto más pequeña y compacta es la luz, más avanzado es el cáncer. De hecho, con un cáncer más agresivo (grado de Gleason 4 y 5), se puede esperar que la luz desaparezca casi o completamente del tejido. También se determinó que la característica morfométrica de la compacidad de los núcleos epiteliales se correlacionaba con la progresión del cáncer, donde el programa informático Definiens Cellenger calculó la compacidad como la proporción del producto de longitud y anchura de los núcleos epiteliales con respecto al área de los núcleos epiteliales. Esto puede deberse a que la invasión de los núcleos epiteliales en el estroma aumenta a medida que el cáncer progresa (es decir, el tejido con cáncer avanzado generalmente incluye una abundancia de núcleos epiteliales). La característica morfométrica basada en el fondo que se determinó que se correlaciona con el resultado en este ejemplo midió el tamaño real del núcleo del tejido utilizado en el análisis.

Tabla 3 - Comparación de la precisión de predicción

	Gleason	Imagen	Gleason + Imagen
Cox	0,6952	0,6373	0,7261
SVRc	0,6907	0,7269	0,7871

Según la Tabla 3, el rendimiento predictivo de las características morfométricas es comparable con el de las puntuaciones de Gleason, y la combinación de las características morfométricas y las puntuaciones de Gleason logra una tasa predictiva más alta, lo que confirma que las características morfométricas extraídas mediante el sistema de análisis de imágenes de tejido, de hecho, proporciona información adicional más allá de las puntuaciones de Gleason. Por lo tanto, el uso de las mediciones morfométricas puede mejorar la predicción general de la recurrencia.

Ejemplo 2: Predicción de la recurrencia del cáncer de próstata y de la supervivencia global con datos clínicos, morfométricos y moleculares

Se realizaron dos estudios que predijeron con éxito la recurrencia del antígeno prostático específico (PSA) con un 88% y un 87% de precisión predictiva, respectivamente. Al combinar las características clínicas, moleculares y morfométricas con el aprendizaje automático, se creó una plataforma robusta que tiene amplias aplicaciones en el diagnóstico de pacientes, la gestión del tratamiento y el pronóstico. Se realizó un tercer estudio para predecir la supervivencia general de los pacientes con cáncer de próstata, donde el resultado de interés fue la muerte por cualquier causa.

Se estudió una cohorte de 539 pacientes que se sometieron a prostatectomía radical incorporando micromatrices de tejido (TMAs) de alta densidad construidas a partir de muestras de prostatectomía. Se realizaron estudios morfométricos utilizando cortes de tejido teñidos con hematoxilina y eosina y se determinaron los determinantes biológicos moleculares con inmunohistoquímica (IHC). Un modelo predictivo tanto para la recurrencia de PSA como para la supervivencia general se obtuvo de un conjunto seleccionado de características a través del aprendizaje multivariante supervisado. Los pacientes con datos presentes completos en cada dominio fueron evaluados con una máquina de vectores de soporte para la regresión desarrollada para manejar datos censurados (SVRc). El rendimiento predictivo del modelo se estimó utilizando el índice de concordancia (CI) con puntuaciones generadas utilizadas para definir los grupos de riesgo.

Utilizando una cohorte de 132 pacientes, se seleccionaron 41 características (incluidas 17 clínicas, 14 moleculares y 10 morfométricas) que predecían la recurrencia de PSA con un 88% de precisión. En una cohorte de 268 pacientes, 10 características (3 clínicas, 1 molecular y 6 morfométricas) fueron predictivas de recurrencia de PSA con un 87% de precisión; además, se descubrió que 14 características (2 clínicas, 1 molecular y 11 morfométricas) predicen la supervivencia general con una precisión del 80%. Usando la prueba de orden logarítmico, se observaron diferencias significativas en la recurrencia del tumor y la muerte entre los grupos de riesgo ($p < 0,0001$).

El presente estudio revela una tendencia progresiva de predicción mejorada de recurrencia del cáncer de próstata mediante el uso de un nuevo enfoque de sistemas que combina variables clínicas, marcadores moleculares e histología de tejidos, analizados mediante aprendizaje automático.

Características clínicas del paciente. Se estudió una cohorte de 539 pacientes que se sometieron a prostatectomía radical. Diecisiete características clínicas (que se muestran a continuación en la Tabla 4) se recopilaron retrospectivamente utilizando información anónima de los pacientes, que incluía la edad del paciente, el PSA preoperatorio y el grado de Gleason.

Tabla 4. Características clínicas recogidas

Característica	Descripción
edad	Edad (en años)
raza	Raza
prepsa	Antígeno prostático específico (ng/dl)
tnm	Estadio clínico TNM
uicc	Estadio clínico UICC
dre	Palpable en tacto rectal
ln	Estado de los ganglios linfáticos
svi	Invasión de las vesículas seminales
márgenes	+/- márgenes quirúrgicos
ece	Tumor localizado fuera de la cápsula
bxgg1	Grado de Gleason de biopsia dominante
bxggtot	Puntuación de Gleason de biopsia
gg1	Grado de Gleason postoperatorio dominante
ggtot	Puntuación de Gleason postoperatoria
prs1tcd	Diploide, Tetraploide, Aneuploide
pp_sphas	Porcentaje de células en ploidía en fase S
pp_frac	Fracción de proliferación en ploidía

5 Se construyeron micromatrices de tejidos (TMA) a partir de bloques seleccionados de las muestras de prostatectomía. Los núcleos de tejido con un diámetro de 0,6 mm de cada muestra se distribuyeron aleatoriamente por triplicado para cada uno de los bloques de parafina receptores (Beecher Instruments, Silver Spring, MD). Se colocaron cortes (5 µm) de estos bloques de TMA en portaobjetos recubiertos de poli-lisina cargada y se usaron para análisis morfológicos e inmunohistoquímicos (IHC) (véase más adelante).

10 Los valores ausentes para las características clínicas se asignaron con modelos flexibles de regresión aditiva que contienen todas las características para estimar el valor de la característica ausente sin referencia al resultado, y solo aquellos pacientes con datos clínicos completos (después de la asignación), morfológicos y moleculares, así como la información de resultado presente, se estudiaron más a fondo. El tamaño efectivo de la muestra para el Estudio 1 (prueba de concepto) consistió en 132 pacientes. La clasificación principal de interés fue si un paciente mostró recurrencia o no después de la cirugía por cáncer de próstata. Se consideró que los pacientes que tuvieron dos elevaciones consecutivas observadas de PSA > 0,2 ng/mL tuvieron cáncer de próstata recurrente. Si un paciente no mostró recurrencia desde su última visita, o si el resultado del paciente se desconoció en su visita más reciente (es decir, debido a la pérdida de seguimiento), entonces el resultado del paciente se consideró censurado. El tiempo hasta la recurrencia se definió como el tiempo (en meses) desde la prostatectomía radical hasta la recurrencia de PSA (bioquímico).

20 El estudio 2 se realizó con 268 pacientes de la cohorte original de 539 pacientes, incluidos 129 de los 132 pacientes del estudio 1. En lugar de utilizar imágenes de tinciones de hematoxilina y eosina derivadas de núcleos de TMA, se analizaron cortes completos de prostatectomías radicales. El estudio 3 examinó la misma cohorte de 268 pacientes, pero se utilizó para predecir la supervivencia general, donde el resultado de interés fue la muerte por cualquier causa.

25 Análisis de imágenes y estudios de morfometría. Las áreas representativas del tejido tumoral original recuperadas de cada paciente, ya sea de un núcleo de tejido o una sección completa, se digitalizaron y se analizaron usando los portaobjetos teñidos con hematoxilina y eosina. Las imágenes se capturaron con un microscopio óptico a 20 aumentos utilizando una cámara digital en color SPOT Insight QE (KAI2000). Solo se seleccionaron áreas que contenían más del 80% de tumor para una segmentación de imagen óptima y un análisis cuantitativo.

30 Análisis Molecular. Se aplicó un panel de 12 biomarcadores que incluyen Citoqueratina 18 (células lumbales), Citoqueratina 14 (células basales), CD45 (linfocitos), CD34 (células endoteliales), CD68 (macrófagos), Ki67 (proliferación), PSA (hK-3, calicreína), PSMA (receptor de crecimiento), ciclina D1 (ciclo celular), p27 (ciclo celular), receptor de andrógenos (endocrino) y Her-2/neu (señalización) en los 7 bloques de TMA con inmunohistoquímica cromogénica estándar. La recuperación de antígeno se realizó con un tampón citrato 0,01 M (pH 6) durante 30 minutos en una olla a presión para todos los anticuerpos. Los métodos y sistemas ilustrativos relacionados con dicho proceso se describen en la solicitud de patente de EE.UU. n° 10/624.233, incorporada anteriormente, presentada el 21 de julio de 2003 y titulada "Métodos y composiciones para la preparación y uso de líneas celulares y tejidos tratados en forma fijada en hibridación fluorescente in situ". Los anticuerpos primarios (que se muestran en la Tabla 5) se diluyeron en solución salina tamponada con Tris con Tween al 0,1% y se aplicaron durante 16 h a 4 °C, seguido de anticuerpos

secundarios biotinilados (Vector) a una dilución 1:1000 durante 1 h.

Tabla 5. Lista de anticuerpos

Biomarcador	Clon
Ki-67	Clon ki-67 (DAKO)
Citoqueratina18	Clon DC-10 (Novocastra)
CD45	Clon X16/99
CD68	Clon 514H2 (Novocastra UK)
CD34	Clon QBEnd 101 (DAKO)
AR	Clon AR27 (Novocastra)
Citoqueratina14	Clon LL002 (Novocastra)
Ciclina D1	Clon P2D11F11
PSA	Clon PA05 (Neomarkers)
PSMA	Clon ZMD.80 (Zymed) ^P
p27	Clon DCS72 (Oncogene)
Her-2/neu	KIT DAKO ^P

^Ppoliclonal, el resto son monoclonales

5 Los portaobjetos de control negativo recibieron suero de ratón normal (DAKO) como anticuerpo primario. Los portaobjetos se contratiñeron con hematoxilina de Harris y los revisaron dos patólogos independientes, y todas las discrepancias las resolvió un tercer patólogo. Los datos registrados de IHC de los 539 pacientes y sus respectivos núcleos triplicados incluyeron el porcentaje y la intensidad (0-3+) de las células que se tiñeron para un antígeno particular en investigación. En su caso, estas dos medidas se combinaron para crear un índice de tinción para ese biomarcador particular (la Tabla 6, a continuación, muestra una lista ejemplar de características moleculares). Se calculó un índice de tinción para AR (receptor de andrógenos), CK14 (citoqueratina 14), ciclina D1, PSA (antígeno prostático específico), PSMA (antígeno prostático específico de membrana), p27 y Her2/neu, mientras que los marcadores restantes (es decir, Ki67, CK18 (citoqueratina 18), CD45, CD68) se evaluaron en base al porcentaje de células positivas con una intensidad dada. Estos biomarcadores se describen adicionalmente a continuación. El índice de tinción varió de 0 a 300, y se calculó de la siguiente manera: 1 * (porcentaje de células que se tiñen positivamente con intensidad 1+ para un biomarcador) + 2 * (porcentaje de células que se tiñen positivamente con intensidad 2+ para el biomarcador) + 3 * (porcentaje de células que se tiñen positivamente con intensidad 3+ para el biomarcador), donde el porcentaje de células que se tiñen positivamente se refiere al número de células positivas identificadas por cada 100 células contadas. Se describen detalles adicionales con respecto a este índice de tinción en [19]. Tal índice de tinción es solo ilustrativo, y se puede usar cualquier otra forma adecuada para medir las características moleculares sin apartarse del alcance de la presente invención.

En la discusión de biomarcadores anterior, p27 pertenece a la familia de reguladores del ciclo celular llamados inhibidores de la quinasa dependientes de ciclina, que se unen a los complejos de ciclina-CDK y causan la detención del ciclo celular en la fase G1. Se ha propuesto que el biomarcador p27 promueve la apoptosis, y desempeña un papel en la diferenciación terminal de algunos tejidos. Mediante inmunohistoquímica, la pérdida de la expresión nuclear de p27 se asocia con un fenotipo más agresivo. Her2/neu es miembro de la familia de receptores de tirosina quinasa EGFR y juega un papel importante en la patogénesis de ciertos cánceres humanos. La sobreexpresión de Her2/neu mediante inmunohistoquímica en membranas celulares se ha asociado con un tipo más agresivo de cáncer de mama. Ki67 es uno de los muchos marcadores proliferativos que tiñe el núcleo con diversos grados de intensidad, y se utiliza para evaluar un índice proliferativo o una medida de la actividad celular de la muestra tumoral en cuestión. CD45 es un antígeno de la superficie celular que se usa para identificar las células que están destinadas a convertirse en células inmunes como los linfocitos (células T, células B, células NK, etc.). Se cree que la intensidad no es tan importante como su distribución/presencia y asociación con otros elementos histológicos. CD68 es un antígeno citoplasmático estrechamente asociado con los lisosomas. Se expresa en toda la cascada de diferenciación de monocitos, pero generalmente es más intenso en los macrófagos que en los monocitos.

Tabla 6. Características moleculares

Característica	Descripción
atki67t1	Ki-67 en el área de intensidad 1 (tumor)
atki67t2	Ki-67 en el área de intensidad 2 (tumor)
atki67t3	Ki-67 en el área de intensidad 3 (tumor)
atki67p1	Ki-67 en el área de intensidad 1 (PIN)

ES 2 745 811 T3

atki67p2	Ki-67 en el área de intensidad 2 (PIN)
atki67p3	Ki-67 en el área de intensidad 3 (PIN)
atki67a1	Ki-67 en el área de intensidad 1 (glándula)
atki67a2	Ki-67 en el área de intensidad 2 (glándula)
atki67a3	Ki-67 en el área de intensidad 3 (glándula)
atc18t3	Citoqueratina18 (tumor)
atcd45t3	CD45 (tumor)
atcd68t3	CD68 (tumor)
atcd34p	CD34 (PIN)
atcd34s	CD34 (estroma)
atcd34t	CD34 (tumor)
atcd34tp	CD34 (tumor/PIN)
atcd34ts	CD34 (tumor/estroma)
atcd34ps	CD34 (PIN/estroma)
atc18p3	Citoqueratina 18 (PIN)
atcd45p3	CD45 (PIN)
atc18a3	Citoqueratina 18 (glándula)
atcd45a3	CD45 (glándula)
arsi	Índice de tinción de AR (tumor)
c14si	Índice de tinción de citoqueratina 14 (tumor)
cd1si	Índice de tinción de ciclina D1 (tumor)
psasi	Índice de tinción de PSA (tumor)
psmasi	Índice de tinción de PSMA (tumor)
p27si	Índice de tinción de p27 (tumor)
her2si	Índice de tinción de Her-2/neu (tumor)
arpsi	Índice de tinción de AR (PIN)
c14psi	Índice de tinción de citoqueratina 14 (PIN)
cdlpsi	Índice de tinción de ciclina D1 (PIN)
psapsi	Índice de tinción de PSA (PIN)
psmapsi	Índice de tinción de PSMA (PIN)
p27psi	Índice de tinción de p27 (PIN)
her2psi	Índice de tinción de Her-2/neu (PIN)
arasi	Índice de tinción de AR (glándula)
c14asi	Índice de tinción de citoqueratina 14 (glándula)
cd1asi	Índice de tinción de ciclina D1 (glándula)
psaasi	Índice de tinción de PSA (glándula)
psmaasi	Índice de tinción de PSMA (glándula)
p27asi	índice de tinción de p27 (glándula)
her2asi	Índice de tinción de Her-2/neu (glándula)

- Estudios analíticos y estadísticos. Se realizaron tres estudios: un análisis de prueba de concepto inicial (Estudio 1) con 132 pacientes y una investigación ampliada (Estudio 2 y Estudio 3) con 268 pacientes. Tanto en el Estudio 1 como en el Estudio 2, el análisis consistió en dos pasos: identificar las características predictivas de la recurrencia de PSA y desarrollar un modelo basado en esas características, con el objetivo final de utilizar el modelo para predecir la recurrencia bioquímica (PSA) en futuros pacientes con prostatectomía radical. Los objetivos del Estudio 3 fueron identificar características y desarrollar un modelo para predecir la supervivencia general después de la prostatectomía. Se utilizó la regresión de vectores de soporte para datos censurados (SVRc) del tipo descrito anteriormente para desarrollar los modelos resultantes en cada uno de estos estudios.
- 5
- 10 La precisión predictiva de un modelo se evaluó utilizando el índice de concordancia (CI). Al tratar con resultados censurados, esta suele ser con frecuencia la métrica de elección. El índice de concordancia se basa en comparaciones por pares entre las puntuaciones de pronóstico de dos pacientes seleccionados al azar que cumplen con cualquiera de los siguientes criterios: ambos pacientes experimentaron el evento y el tiempo del evento del primer paciente es

más corto que el del segundo paciente, o solo el primer paciente experimentó el evento y su tiempo de evento es más corto que el tiempo de seguimiento del segundo paciente. El CI estima la probabilidad de que un paciente con la puntuación de pronóstico más alta del modelo experimente el evento en un tiempo más corto que un paciente con una puntuación más baja y esté estrechamente asociado con el área bajo la curva ROC (AUC). También se pueden usar otras métricas para medir la capacidad de un modelo predictivo. Por ejemplo, la sensibilidad y la especificidad pueden usarse para evaluar el diagnóstico. Como otro ejemplo, se puede usar un "valor p" que representa la probabilidad de que el azar solo sea responsable, por ejemplo, de las diferencias observadas entre estratos (por ejemplo, véanse las Figuras 8, 10 y 12). Por lo tanto, cuanto menor sea el valor p, más probable es que haya una asociación estadística verdadera con el resultado. Típicamente, la norma es que cualquier valor p menor o igual a 0,05 es estadísticamente significativo.

Estudio 1. En este análisis, el modelo SVRc descrito anteriormente se aplicó secuencialmente a los datos clínicos, moleculares y morfométricos, y las características clínicas sirvieron primero como ancla para un algoritmo de selección "codicioso hacia delante" (FS) de características a través de la ejecución SVRc en los datos moleculares. Después de este paso, se ejecutó un segundo algoritmo de selección codicioso hacia delante de características SVRc con los datos morfométricos, utilizando la combinación de las características clínicas y moleculares seleccionadas como ancla. El último paso consistió en ejecutar un algoritmo de selección codicioso hacia atrás con la combinación de las características clínicas, moleculares seleccionadas y morfométricas seleccionadas para obtener un modelo final. Durante la selección de características, el criterio para determinar si una característica se introdujo (o mantuvo) en el modelo se basó en si la presencia (o ausencia) de esa característica aumentó el índice de concordancia, es decir, si añadió información predictiva.

Se evaluó la precisión predictiva del modelo utilizando una validación interna y externa. La validación interna se realizó mediante una validación cruzada cinco veces. Para realizar la validación externa, se creó una serie de conjuntos de análisis de pacientes a partir de la cohorte de pacientes, y el resultado previsto se comparó con el resultado real para estos pacientes a través del índice de concordancia. Al aplicar este diseño de validación de dos niveles, se seleccionó al azar un subconjunto de pacientes del conjunto completo de registros de pacientes y solo se usaron los pacientes restantes para construir el modelo predictivo utilizando el procedimiento que se acaba de describir. Los registros retenidos se utilizaron para aplicarlos al modelo entrenado con el fin de obtener una precisión predictiva. Estos dos pasos se repitieron B veces para obtener B tasas predictivas, donde la tasa predictiva final fue el promedio. Las características seleccionadas para el modelo final fueron aquellas que aparecieron una cantidad suficiente de veces en los B modelos distintos creados.

Usando el conjunto de características seleccionado, se desarrolló un modelo de red neuronal maximizando directamente el índice de concordancia. Particularmente, se usó una red neuronal (NNci) del tipo descrito anteriormente, en la cual la red se entrenó usando una función objetivo sustancialmente de acuerdo con una aproximación del índice de concordancia. El resultado de este modelo final se utilizó para estimar el riesgo individual futuro del paciente de recurrencia de PSA.

Estudio 2. Los objetivos de este estudio fueron idénticos al Estudio 1; sin embargo, se utilizaron diferentes procedimientos de selección y validación de características. En lugar de utilizar el enfoque de anclaje, todas las características se clasificaron según su asociación con el tiempo hasta la recurrencia de PSA (medido por el índice de concordancia), y se seleccionaron aquellas características que pasaron un cierto umbral predeterminado ($CI \geq 0,60$). Esto se realizó después de que los expertos en dominios redujeran el número de características de las imágenes, y estas características se evaluaron en una serie de modelos de n características (por ejemplo, 1 característica, 2 características, 3 características, etc.). Usando un proceso de selección de características hacia delante, las características que maximizaban el índice de concordancia de cada modelo de n características se usaron en el siguiente modelo de n + 1 características. Este proceso finalizó una vez que el CI no pudo mejorarse con un umbral predeterminado. Luego, utilizando un proceso de selección de características hacia atrás, las características se eliminaron en un esfuerzo por aumentar el CI. Este proceso finalizó cuando la eliminación de cualquier característica no mejoró el CI.

Se usó una técnica de arranque simple para la selección de las características. En este enfoque, se tomaron muestras de los pacientes con reemplazo y se usaron como conjunto de entrenamiento, mientras que el modelo se evaluó en los no seleccionados. A modo de comparación, este algoritmo de selección de características se ejecutó utilizando solo las características que se encuentran en el nomograma postoperatorio de Kattan, que se describe en Kattan et al., patente de EE.UU. n° 6.409.664, que se incorpora en la presente memoria como referencia en su totalidad. El resultado del modelo final se utilizó para estimar el riesgo individual futuro de recurrencia de PSA del paciente.

Estudio 3. El objetivo de este estudio fue identificar características predictivas de supervivencia global utilizando la misma cohorte y conjunto de características analizadas en el Estudio 2, así como el mismo algoritmo de selección de características. El resultado del modelo final se usó para estimar el riesgo individual futuro de muerte del paciente debido a cualquier causa.

RESULTADOS

El enfoque general fue aplicar la patología de sistemas (la combinación de análisis morfométricos, firmas moleculares

5 y perfiles clínicos de pacientes) para desarrollar modelos predictivos para la recurrencia de PSA y la supervivencia general en una cohorte de pacientes con cáncer de próstata después de la prostatectomía. Es importante tener en cuenta que cuando se utilizaron solamente las características clinicopatológicas del Estudio 1 en un análisis estándar del Modelo de Cox, la precisión para predecir la recurrencia de PSA fue solo del 59%. Solamente después de la integración de las características morfológicas y moleculares con SVRc, el nivel de precisión predictiva aumentó al 88%. Las siguientes secciones describen cómo se logró esta mejora.

10 *Estudio 1.* Para los 132 pacientes de esta cohorte, la mediana de edad en el momento del diagnóstico fue de 63 años (mínimo: 40, máximo: 81), y el PSA preoperatorio promedio fue de 8,2 ng/dl (mínimo: 1,1, máximo: 81,9). Según las muestras de prostatectomía, el 32% tuvo una puntuación de Gleason inferior a 7, el 60% tuvo una puntuación de Gleason 7 y el 8% restante tuvo una puntuación superior a 7. Sesenta y nueve pacientes (52%) fueron pT2N0M0, 40 pacientes (30%) pT3aN0M0 y los 23 pacientes restantes (18%) pT3bN0M0 o pT1-3N+. (La Tabla 7 contiene una lista resumida de las características clínicas de los tres estudios).

Tabla 7. Información clínica

	Estudio 1	Estudio 2 y 3
N	132	268
Edad (años)		
Media	62	62
Mediana	63	63
Intervalo	40 - 81	40 - 81
Raza		
Caucásico	120 (90,9%)	241 (89,9%)
Hispano	8 (6,1%)	12 (4,5%)
Afroamericano	2 (1,5%)	9 (3,4%)
Desconocido	2 (1,5%)	6 (2,2%)
PSA preoperatorio (ng/dl)		
Media	12,2	10,8
Mediana	8,2	7,8
Intervalo	1,1 - 81,9	0,9 - 81,9
Estadio TNM		
pT2N0	69 (52,3%)	157 (58,6%)
pT3aN0	40 (30,3%)	72 (26,9%)
pT3bN0	13 (9,8%)	22 (8,2%)
pT1-3N+	10 (7,6%)	17 (6,3%)
Estadio UICC		
T1a < 5%	0 (0,0%)	1 (0,3%)
T1b ≥ 5%	0 (0,0%)	1 (0,3%)
T1c no palpable ni visible	49 (37,1%)	112 (41,8%)
T2a ≤ ½ lóbulo	23 (17,4%)	58 (21,7%)
T2b ≤ 1 lóbulo	27 (20,5%)	45 (16,8%)
T2c ambos lóbulos	23 (17,4%)	34 (12,7%)
T3a ECE unilateral	8 (6,1%)	15 (5,6%)
T3c SV+	2 (1,5%)	2 (0,8%)
Resultado DRE		
No palpable	56 (42,4%)	118 (44,0%)
Palpable	76 (57,6%)	150 (56,0%)
Afectación de ganglios linfáticos		
Negativo	121 (91,7%)	250 (93,3%)
Positivo	11 (8,3%)	18 (6,7%)
Afectación de vesículas seminales		
No	113 (85,6%)	236 (88,0%)

ES 2 745 811 T3

Sí	19 (14,4%)	32 (12,0%)
Márgenes quirúrgicos		
Negativo	108 (81,8%)	217 (81,0%)
Positivo	24 (18,2%)	51 (19,0%)
Afectación extracapsular		
No	70 (53,0%)	159 (59,3%)
Sí	62 (47,0%)	109 (40,7%)
	Estudio 1	Estudio 2 y 3
Grado de Gleason de biopsia dominante		
1	0 (0,0%)	1 (0,4%)
2	24 (18,2%)	43 (16,0%)
3	85 (64,4%)	184 (68,7%)
4	22 (16,7%)	38 (14,2%)
5	1 (0,7%)	2 (0,8%)
Puntuación de Gleason de biopsia		
2	0 (0,0%)	1 (0,4%)
3	0 (0,0%)	0 (0,0%)
4	6 (4,6%)	7 (2,6%)
5	27 (20,5%)	56 (20,9%)
6	41 (31,1%)	97 (36,2%)
7	48 (36,4%)	90 (33,6%)
8	7 (5,3%)	13 (4,9%)
9	3 (2,3%)	4 (1,5%)
Grado de Gleason postoperatorio dominante		
2	3 (2,3%)	20 (7,5%)
3	98 (74,2%)	201 (75,0%)
4	31 (23,5%)	47 (17,5%)
Puntuación de Gleason postoperatoria		
5	6 (4,6%)	21 (7,8%)
6	36 (27,3%)	86 (32,1%)
Puntuación de Gleason postoperatoria		
7	79 (59,9%)	148 (55,2%)
8	10 (7,6%)	12 (4,5%)
9	1 (0,8%)	4 (0,4%)
Ploidía		
Diploide	74 (56,1%)	145 (54,1%)
Tetraploide	54 (40,9%)	115 (42,9%)
Aneuploide	4 (3,0%)	8 (3,0%)
Porcentaje de ploidía en fase S (%)		
Media	2,3	2,4
Mediana	1,1	1,1
Intervalo	0,0 - 63,8	0,0 - 66,4
Porcentaje de fracción de ploidía		
Media	3,4	3,5
Mediana	2,6	2,4
Intervalo	0,0 - 20,0	0,0 - 20,0

Veinte (15%) pacientes experimentaron recurrencia de PSA, mientras que los pacientes restantes (85%) se censuraron. Para los pacientes censurados, el tiempo medio de seguimiento fue de 60,8 meses, o poco más de 5 años. No se alcanzó la mediana del tiempo total hasta la recurrencia de PSA. Las diecisiete características clínicas se seleccionaron como predictivas de la recurrencia de PSA, y la más informativa se anotó de la siguiente manera (característica clínico-patológica y nº de veces seleccionada por el modelo): grado de Gleason de biopsia (112), raza (112), estadio clínico UICC (110), ploidía (110) y resultados DRE (109).

Análisis de imágenes y estudios de morfometría. Las Figuras 5a y 5b ilustran imágenes digitalizadas de tejido prostático sano y anormal, respectivamente, obtenidas después de la segmentación y clasificación de acuerdo con la presente invención. Se han etiquetado varios objetos patológicos en el tejido para ilustración. El programa informático de análisis de imágenes generó un total de 496 características morfométricas (que se muestran en la Tabla 1).

De las 496 características morfométricas, las 10 características morfométricas que se muestran en la Figura 6 se seleccionaron como predictivas de recurrencia de PSA. Las características morfométricas seleccionadas se relacionan con los siguientes objetos patológicos, donde los números entre paréntesis junto a las características indican cuántas veces se seleccionaron las características en relación con el resultado durante la generación del modelo final: glóbulos rojos, núcleos epiteliales, luz, estroma, citoplasma y fondo de tejido (longitud mínima de glóbulos rojos en píxeles (20), compactación máxima de núcleos epiteliales (17), radio mínimo del lumen del recinto más pequeño (14), anchura mínima de los núcleos epiteliales en píxeles (11), densidad máxima del estroma (10), longitud máxima del borde del lumen en píxeles (10), canal 2 de desviación estándar mínima de núcleos epiteliales (10), radio máximo de núcleos epiteliales del recinto más pequeño (10), desviación estándar del citoplasma de la longitud del borde en píxeles (10) y desviación estándar del fondo del área en píxeles (10)). Más particularmente, en este ejemplo, las características morfométricas de longitud para glóbulos rojos, radio del recinto más pequeño y longitud del borde para la luz, longitud del borde para el citoplasma, densidad para el estroma (p. ej., raíz cuadrada del área cubierta por un estroma dividido por su radio), y se determinó que el área de fondo se correlaciona con el resultado. También se determinaron las características morfométricas de compacidad, anchura, valor del canal verde y radio del recinto más pequeño para los núcleos epiteliales (p. ej., se crea una elipse con la misma área que el objeto y luego se amplía hasta que encierra completamente los núcleos epiteliales, y se calcula la proporción del radio de la elipse de cierre más pequeña respecto del radio de la elipse original) para correlacionarlas con el resultado.

Anteriormente se describen varios motivos posibles para al menos algunas de estas correlaciones en relación con el Ejemplo 1. Por ejemplo, la característica morfométrica de compacidad de los núcleos epiteliales puede ser un reflejo de la naturaleza 'adosada' de las células epiteliales en un patrón circunferencial que sugeriría una pérdida de formación/diferenciación glandular y de la luz y, por lo tanto, sería consistente con un mayor grado de Gleason (es decir, mayor progresión de la enfermedad). Además, la característica morfométrica del radio del recinto más pequeño de la luz se relaciona con el tamaño general de la luz, que se reduce y disminuye drásticamente a medida que aumenta el grado de Gleason.

Además, las correlaciones determinadas en este estudio pueden explicarse, al menos parcialmente, por la hipótesis de que los núcleos epiteliales suelen tener una forma menos diversa (p. ej., más redonda con menos variaciones) y tamaño (p. ej., área y longitud del borde), y tienen menos variación de color a medida que los núcleos epiteliales invaden el estroma. Esta invasión del estroma también puede explicar por qué se ha determinado que las características morfométricas del estroma se correlacionan con la progresión de la enfermedad. Particularmente, las imágenes cancerosas se caracterizan típicamente por una pequeña cantidad de estroma porque el área del estroma se reemplaza por citoplasma de células epiteliales a medida que el cáncer progresa. Esto hace que los valores de densidad para el estroma sean más altos porque la compacidad del estroma se reduce y adquiere una forma más fractal (el radio del objeto aumenta más que el área a medida que los objetos se deforman y se vuelven más delgados). Un razonamiento adicional para las correlaciones determinadas en este estudio puede ser que una gran cantidad de glóbulos rojos que viajan a través del tejido pueden reflejar alguna medida de angiogénesis o formación de nuevos vasos sanguíneos, que pueden estar relacionados con la progresión de la enfermedad como medio para que las células salgan de la próstata y se diseminen externamente, lo que afecta al resultado clínico de la recurrencia de PSA/BCR.

Como se indicó anteriormente, se entenderá que al menos algunas de las características morfométricas particulares determinadas por las enseñanzas proporcionadas en la presente memoria para correlacionarse con el resultado pueden depender, por ejemplo, del equipo, programa informático o combinación particular de los mismos que se usa en la presente invención para calcular las características morfométricas. El programa informático Definens Cellenger y las características morfométricas particulares medidas por el programa informático descrito en la presente memoria son solo ilustrativos, y puede usarse cualquier otro equipo, programa informático o combinación de los mismos sin apartarse del alcance de la invención.

Análisis Molecular. De los 12 biomarcadores que se evaluaron mediante IHC, se registraron un total de 43 características únicas. (Las tablas 8a, 8b y 8c, a continuación, muestran un resumen del biomarcador observado - características moleculares).

Tabla 8a. Células (%) Teñidas (+) por Componente Histológico e Intensidad (Estudio 1)

Marcador	Tumor			PIN			Glándula		
	1+	2+	3+	1+	2+	3+	1+	2+	3+
Ki-67									
Media ±	23,9 ±	9,8 ±	2,4 ±	25,3 ±	10,3 ±	2,6 ±	1,8 ±	0,0 ±	0,1 ±
D.E.	31,38	21,32	4,64	32,50	21,51	3,29	9,96	0,36	0,63
Mediana	4,7	0,0	0,0	4,8	0,0	0,0	0,0	0,0	0,0
Intervalo	0,0 - 100,0	0,0 - -100,0	0,0 - 26,3	0,0 - 100,0	0,0 - 100,0	0,0 - 39,5	0,0 - 96,0	0,0 - 4,0	0,0 - 6,3
CK 18									
Media ±			100,0 ±			100,0 ±			100,0 ±
D.E.	NA	NA	0,00	NA	NA	0,00	NA	NA	0,00
Mediana	NA	NA	100,0	NA	NA	100,0	NA	NA	100,0
Intervalo	NA	NA	100,0 - 100,0	NA	NA	100,0 - 100,0	NA	NA	-100,0
CD45									
Media ±			0,0 ±			0,0 ±			0,0 ±
D.E.	NA	NA	0,04	NA	NA	0,01	NA	NA	0,00
Mediana	NA	NA	0,0	NA	NA	0,0	NA	NA	0,0
Intervalo	NA	NA	0,0 - 0,4	NA	NA	0,0 - 0,1	NA	NA	0,0 - 0,0
CD68									
Media ±			0,0 ±						
D.E.	NA	NA	0,01	NA	NA	NA	NA	NA	NA
Mediana	NA	NA	0,0	NA	NA	NA	NA	NA	NA
Intervalo	NA	NA	0,0 - 0,1	NA	NA	NA	NA	NA	NA

Tabla 8b. Células CD34 (%) Teñidas (+) por Componente Histológico (Estudio 1)

	PIN	Estroma	Tumor	Tumor/PIN	Tumor/estroma	PIN/estroma
Media ± DE	0,0 ± 0,05	0,0 ± 0,03	0,1 ± 0,21	0,0 ± 0,06	0,0 ± 0,08	0,0 ± 0,05
Mediana	0,0	0,0	0,0	0,0	0,0	0,0
Intervalo	0,0-0,4	0,0-0,2	0,0-0,9	0,0-0,5	0,0-0,4	0,0-0,3

Tabla 8c. Índice de Tinción (0 - 300) por Componente Histológico (Estudio 1)

Marcador	Tumor	PIN	Glándula
AR			
Media ± DE	179,8 ± 71,4	64,3 ± 75,10	22,6 ± 56,86
Mediana	200	36,5	0
Intervalo	0 - 300	0 - 300	0 - 300
CK14			
Media ± DE	2,6 ± 5,83	31,2 ± 57,35	4,7 ± 20,42
Mediana	0	0	0
Intervalo	0 - 42	0 - 285	0 - 150
Ciclina D1			
Media ± DE	1,5 ± 5,15	0,0 ± 0,27	0,0 ± 0,0
Mediana	0	0	0
Intervalo	0 - 33	0-3	0-0
PSA			

Media ± DE	128,0 ± 68,85	135,7 ± 97,88	13,9 ± 41,32
Mediana	100	111	0
Intervalo	0 - 300	0 - 300	0 - 201
PSMA			
Media ± DE	0,5 ± 2,97	9,5 ± 26,93	2,5 ± 15,00
Mediana	0	0	0
Intervalo	0 - 21	0 - 154	0 - 99
p27			
Media ± DE	4,3 ± 9,61	7,0 ± 19,49	2,1 ± 12,03
Mediana	0	0	0
Intervalo	0-80	0 - 140	0 - 120
Her-2/neu			
Media ± DE	4,1 ± 18,50	0,1 ± 1,00	0,0 ± 0,00
Mediana	0	0	0
Intervalo	0 - 146	0 - 10	0-0

De estos 12 anticuerpos, se seleccionaron 8 biomarcadores que abarcan 14 características moleculares específicas como asociadas con la recurrencia de PSA. Algunos ejemplos de las características moleculares más altamente seleccionadas se anotan de la siguiente manera (biomarcador - nº de veces seleccionado por el modelo) e incluyen: Índice de tinción de AR - tumor (93), Índice de tinción de AR - glándula atrófica (54), CD34 - tumor asociado/PIN (22), Ki-67 - tumor (18) y CD45 - asociado con PIN (17), donde PIN es una abreviatura de neoplasia intraepitelial prostática. Las Figuras 7a y 7b ilustran campos representativos que demuestran los perfiles de expresión para AR y CD34, respectivamente. El perfil de expresión de biomarcadores fue notable para los patrones de expresión altamente seleccionados y algo heterogéneos de AR y CD34. Estos marcadores y su relación con el tumor, las glándulas atróficas (para AR) y el tumor/PIN (para CD34) sugieren un significado biológico y funcional que afecta al resultado clínico de la recurrencia de PSA. El segundo grupo de marcadores seleccionados incluyó Ki-67 y CD45, los cuales tuvieron una frecuencia de selección prominente pero en general baja en comparación con AR y CD34.

Estudios analíticos y estadísticos. Utilizando la experiencia en el dominio y el procedimiento de selección de características específicas del dominio anterior donde se crearon 120 divisiones aleatorias para el entrenamiento (N = 100) y el análisis (N = 32) de los modelos, el conjunto de características finales se redujo a 41 características totales de las cuales 17 fueron clínicas, 10 morfométricas y 14 moleculares. La Figura 5 muestra una lista completa de las características seleccionadas. Las 10 características morfométricas se describieron anteriormente. Las características clínicas y moleculares se describen adicionalmente a continuación.

Características Clínicas

1. Puntuación de Gleason de biopsia: los grados de Gleason resumidos (dominante y secundario) que se asignan a las múltiples muestras de tejido de biopsia de aguja recibidas por un patólogo. El sistema de puntuación de Gleason se desarrolló para crear un medio estandarizado, algo subjetivo, de representar la arquitectura del adenocarcinoma prostático mediante histología con la producción de grados individuales. Los grados varían de 1 a 5 según el grado de diferenciación de las unidades glandulares y las células epiteliales. Los patrones dominantes (primarios) y subdominantes (secundarios) se suman para crear un Resumen de Gleason. Además, las características de la compacidad general del estroma, el tamaño de las células epiteliales y las características nucleares se consideran ocasionalmente en el sistema de clasificación general.
2. Raza (p. ej., afroamericana, caucásica, etc.)
3. Estadio UICC: uso del sistema de estadificación TNM de la Unión Internacional contra el Cáncer para definir la estadificación clínica del cáncer, donde "T" representa el tamaño del tumor, "N" representa la afectación de los ganglios linfáticos y "M" representa la metástasis en un sitio distante.
4. Resultado de ploidía: contenido de AD, que es un reflejo del contenido general de ADN dentro de las células epiteliales del cáncer de próstata. Las células benignas y las células tumorales con buen comportamiento crecen y se dividen de manera ordenada. En estado de reposo, contienen un conjunto completo de cromosomas (esta es la condición diploide). Este conjunto completo de cromosomas consta de 23 cromosomas (o N) de Ma y 23 (N nuevamente) cromosomas de Pa (lo que equivale a un total de 2N). Una célula debe duplicar el número de sus cromosomas antes de poder dividirse, creando dos conjuntos completos de cromosomas (esto es 4N, o el estado tetraploide). Una vez completada la

división, cada nueva célula recibe la mitad del material genético y, por lo tanto, se vuelve diploide (2N) una vez más. Si el análisis de la ploidía del ADN se realizara en un grupo de estas células, se vería que la mayoría de las células serían diploides y una pequeña fracción de ellas (las que se están preparando para dividirse) serían tetraploides. Además, al medir y crear un gráfico de la cantidad de material genético en cada célula, se vería un pico diploide dominante y un pico tetraploide menor. La cantidad de ADN en una célula se puede medir tiñéndola con un tinte que se une al material genético. La concentración y distribución de este colorante (tinción de Feulgen) se puede medir mediante microscopía de análisis de imágenes.

Cuando los tumores empeoran, tienden a no dividirse tan ordenadamente como antes. En lugar de tener en estado de reposo un conjunto completo de cromosomas, el estado de reposo puede tener solamente un conjunto y medio. Dichas células tendrían un contenido de ADN que no sería ni diploide ni tetraploide, sino intermedio. Al trazar estas células en el gráfico descrito anteriormente, se obtendría un pico aneuploide a mitad de camino entre los otros dos picos. Los estudios han demostrado que los tumores que tienen un pico aneuploide significativo no se comportan tan bien como los que no lo tienen. Esto no es sorprendente, porque existe una fuerte correlación entre el estado de la ploidía y el grado nuclear. Cualquier patólogo con suficiente experiencia en cáncer de próstata puede evaluar un grado nuclear. El valor que añade el análisis de la ploidía de ADN es que es una medición objetiva que se puede comparar entre laboratorios utilizando técnicas estandarizadas, y que se puede utilizar para realizar una verificación rápida de la precisión aproximada de la puntuación de Gleason. Por ejemplo, cualquier tumor con puntuación de Gleason $2 + 2 = 4$ o $2 + 3 = 5$ que tenga un pico aneuploide debe ser potencialmente reevaluado para un posible ajuste de la puntuación.

5. Resultado DRE: resultado de un tacto rectal (p. ej., negativo o positivo) que se utiliza para determinar el alcance de la enfermedad, tanto en la próstata como en la extensión extraprostática por palpación.
6. Afectación de los ganglios linfáticos: una medida del grado en que los ganglios linfáticos contienen células tumorales (p. ej., células epiteliales del cáncer de próstata), que puede evaluarse mediante inspección clínica/quirúrgica o en el momento de una prostatectomía.
7. Grado de Gleason de biopsia dominante: véase la descripción anterior de la puntuación de Gleason de biopsia. Esto refleja el patrón de clasificación de Gleason dominante observado en una biopsia o una muestra de prostatectomía.
8. Porcentaje de ploidía en fase S: representa una fracción del contenido celular que se encuentra en una fase proliferativa o S del ciclo celular y refleja el potencial de crecimiento del tumor.
9. Puntuación de Gleason postoperatoria: puntuación del tejido tomado después de la cirugía de varias regiones de la muestra de prostatectomía.
10. Estadío TNM: tumor, ganglio y metástasis según los criterios UICC posteriores a la prostatectomía y según el examen patológico de muestras de tejido.
11. Grado de Gleason postoperatorio dominante: el grado de Gleason dominante que representa la característica histológica más predominante presente en la muestra de prostatectomía.
12. Edad
13. Afectación de la vesícula seminal: invasión de la vesícula seminal por el tumor.
14. PSA preoperatorio: nivel de PSA observado antes de la cirugía
15. Fracción de ploidía en porcentaje: véase la descripción anterior del resultado de ploidía.
16. Afectación del margen quirúrgico: afectación de los márgenes quirúrgicos por el tumor, que refleja el grado en que el lecho del que se extrajo el tumor/próstata en el momento de la cirugía contenía células tumorales.
17. Afectación extracapsular: extensión del tumor más allá de la cápsula de la próstata.

Características moleculares

1. AR - tumor: Índice de tinción del receptor de andrógenos (AR) para un tumor, que es una medida del porcentaje y la intensidad de las células que se tiñen positivamente para AR. Con respecto al cáncer de próstata, el índice de tinción puede representar el grado de producto de reacción marrón que se detecta en los núcleos de las células epiteliales en las muestras de próstata evaluadas.
2. AR - glándula: Índice de tinción de AR para un tumor, que está presente dentro de una estructura

glandular.

3. CD34 - tumor/PIN: La localización de CD34 en las células endoteliales de los vasos que están asociadas con el tumor y la PIN.
- 5 4. Ki67 - tumor 2: La identificación de núcleos positivos para ki67 en los núcleos de las células epiteliales tumorales.
5. CD45 - PIN 3: La identificación de linfocitos positivos para CD45 en asociación con PIN.
6. CD34 - tumor/estroma: La localización de los vasos CD34 que están asociados con el tumor.
7. Ki-67 - tumor 3: véase anteriormente.
8. p27 - tumor: La identificación de p27 en los núcleos de las células epiteliales tumorales.
- 10 9. C14 - PIN: La identificación de la citoqueratina 14 en las células basales (epiteliales) de la unidad glandular.
10. CD34 - tumor: La localización de CD34 en los vasos que están asociados con el tumor.
11. PSA - glándula: La identificación de PSA en las células epiteliales lumbinales de la unidad de la glándula.
- 15 12. PSMA - PIN: La identificación de PSMA en las células glandulares/lumbinales de las regiones identificadas como PIN.
13. CD34 - PIN/estroma: La localización de CD34 en los vasos asociados con PIN.
14. CD45 - tumor 3: La identificación de linfocitos positivos para CD45 que están asociados con el tumor.

Como cada dominio de datos se analizó durante este proceso utilizando SVRc, la precisión predictiva de los modelos aumentó. Utilizando la validación interna, al observar solo los datos clínicos, el índice de concordancia fue de 0,79. Al añadir características del dominio molecular, el índice de concordancia aumentó a 0,81. El modelo final, formado por la adición de las características morfométricas, alcanzó un índice de concordancia de 0,84. Cada uno de estos modelos validados internamente también fue validado externamente (como se describió anteriormente en Materiales y Métodos) con la misma tendencia observada. Usando NNci en el conjunto final de características seleccionadas, el índice de concordancia alcanzó 0,88.

El resultado resultante de los modelos NNci y SVRc puede interpretarse como una estimación del riesgo relativo de recurrencia de PSA para un paciente individual. Usando los cuartiles de esta puntuación ($\leq 25\%$, $>25\%-75\%$, $>75\%$), se crearon grupos de pacientes de riesgo; las estimaciones de recurrencia de Kaplan-Meier para cada grupo de riesgo según el modelo NNci se presentan en la Figura 8. Los grupos mostraron una diferencia estadísticamente significativa en el tiempo hasta la recurrencia de PSA (prueba de orden logarítmico, valor $p < 0,0001$). El valor p representa la probabilidad de que solo el azar sea responsable de las diferencias observadas entre estratos (grupos de riesgo en estos ejemplos). Por lo tanto, cuanto menor sea el valor p , más probable es que se esté observando una verdadera asociación estadística. Generalmente, cualquier valor p menor o igual a 0,05 es estadísticamente significativo.

Estudio 2. Para los 268 pacientes en esta cohorte, que contiene 129 de los 132 pacientes analizados en el Estudio 1, la edad media en el momento del diagnóstico fue de 63 años (mínimo: 38, máximo: 81), y el PSA medio antes de la prostatectomía radical fue de 7,8 ng/dl (mínimo: 0,9, máximo: 81,9). Según las muestras de prostatectomía, el 40% de los tumores tuvieron una puntuación de Gleason menor a 7, mientras que el 55% de las prostatectomías tuvieron un Gleason 7. El 5% restante de las prostatectomías tuvieron una puntuación de Gleason mayor de 7. A ciento cincuenta y siete pacientes (59%) se les diagnosticó una enfermedad pT2N0M0, 72 pacientes (27%) pT3aN0M0 y los 39 pacientes restantes (14%) pT3bN0M0 o pT1-3N+. (Véase la Tabla 5, anteriormente, para los detalles de todas las características clínico-patológicas analizadas para esta cohorte). Treinta y ocho (14%) pacientes experimentaron recurrencia de PSA, mientras que los pacientes restantes (86%) se censuraron. Para los pacientes censurados, el tiempo medio de seguimiento fue de 58,7 meses, o poco menos de 5 años. No se alcanzó la mediana del tiempo total hasta la recurrencia de PSA. Se seleccionaron tres características clínicas como predictivas de la recurrencia de PSA: estadio clínico TNM, márgenes quirúrgicos y ganglios linfáticos.

Análisis de imágenes y estudios de morfometría. Usando una versión actualizada del programa informático de análisis de imágenes, pero analizando los mismos portaobjetos teñidos con hematoxilina y eosina, se generaron un total de 350 características morfométricas (que se muestran en la Tabla 2, anteriormente).

La Figura 9 muestra que, de las 350 características, se seleccionaron 6 características morfométricas como predictivas de la recurrencia de PSA, donde estas características morfométricas se relacionaron con los objetos patológicos de núcleos epiteliales, estroma, citoplasma, glóbulos rojos y luz (es decir, EpithelialNucleiMinCompactne0215, StromaMaxStddevChannel30569, CytoplasmStddevMaxDiff0148, RedBloodCellMeanAreaPxI0386,

5 RedBloodCellStddevAreaPx10388 y LumenMinAsymmetry0295). Más particularmente, en este estudio, las características morfométricas de la compacidad de los núcleos epiteliales, el valor del canal azul para el estroma, la diferencia máxima para el citoplasma (por ejemplo, el valor medio mínimo que pertenece al citoplasma se resta de su valor máximo en todos los canales de color para el citoplasma, donde el resultado se divide por el brillo del objeto), el área para los glóbulos rojos y la asimetría de la luz se seleccionaron como correlacionadas con el resultado.

10 Varios posibles motivos para al menos algunas de estas correlaciones se describieron anteriormente en relación con el Ejemplo 1 y/o el Estudio 1. Por ejemplo, las características morfométricas, que incluyen la compacidad de las células epiteliales, la variación e interrupción del estroma por infiltración de las células epiteliales, y la evidencia de un tamaño reducido de la luz proporcionaría una evidencia histológica de un mayor grado de Gleason (es decir, una mayor progresión de la enfermedad). Un mayor grado de Gleason sugiere un tumor de próstata más agresivo que apoyaría la metástasis y/o la extensión del tumor, que respalda la recurrencia de PSA después de la cirugía. Además, la identificación de glóbulos rojos en varios formatos sugeriría una abundancia de vasos. La evidencia de vasos adicionales crearía una posible ruta por la cual las células epiteliales podrían salir de la próstata y distribuirse en lugares externos, y se produciría PSA.

15 Las características clínicas y moleculares seleccionadas en el estudio 2 se muestran en la Figura 9 y se enumeran a continuación. Las descripciones de estas características clínicas y moleculares se proporcionaron anteriormente.

Características Clínicas

1. Estadio TNM
2. Afectación del margen quirúrgico
- 20 3. Afectación de ganglios linfáticos

Característica molecular

1. Índice de tinción de AR (tumor)

25 Cada número de la Figura 9 representa el índice de concordancia de un modelo predictivo basado en la característica correspondiente y todas las demás características de la Figura 9 que tienen números más pequeños. Por ejemplo, 0,8483 es el CI de un modelo basado en las características TNM Estadio Clínico, Márgenes Quirúrgicos, EpithelialNucleiMinCompactne0215, Ganglios Linfáticos y StromaMaxStddevChannel30569. El CI de un modelo basado en las mismas 5 características más el índice de tinción de AR (tumor) es 0,8528. En otras palabras, la adición de la característica molecular del Índice de tinción de AR al modelo aumenta el poder predictivo del modelo.

30 Análisis Molecular. No se necesitaron estudios inmunohistoquímicos adicionales. Los datos recopilados originalmente se usaron como se describe en Materiales y Métodos (véanse las Tablas 9a, 9b y 9c para un resumen completo de las características moleculares).

Tabla 9a. Células (%) Teñidas (+) por Componente Histológico e Intensidad (Estudio 2 y Estudio 3)

Marcador	Tumor			PIN			Glándula		
	1 +	2+	3+	1 +	2+	3+	1 +	2+	3+
Ki-67									
Media ± DE	22,1 ± 30,30	7,3 ± 17,04	1,9 ± 4,01	23,2 ± 31,36	7,9 ± 18,16	2,0 ± 4,46	1,3 ± 7,96	1,2 ± 9,78	0,3 ± 1,55
Mediana	1,3	0,0	0,0	1,0	0,0	0,0	0,0	0,0	0,0
Intervalo	0,0-100,0	0,0-100,0	0,0-26,3	0,0-100,0	0,0-100,0	0,0-39,5	0,0-96,0	0,0-96,5	0,0- 13,0
CK 18									
Media ± DE.	NA	NA	100,0 ± 0,00	NA	NA	1,0 ± 0,04	NA	NA	100,0 ± 0,00
Mediana	NA	NA	100,0	NA	NA	100,0	NA	NA	100,0
Intervalo	NA	NA	100,0-100,0	NA	NA	0,5-100,0	NA	NA	100,0-100,0
CD45									
Media ± DE	NA	NA	0,0 ± 0,04	NA	NA	0,0 ± 0,01	NA	NA	0,0 ± 0,00
Mediana	NA	NA	0,0	NA	NA	0,0	NA	NA	0,0
Intervalo	NA	NA	0,0-0,4	NA	NA	0,0-0,1	NA	NA	0,0-0,0
CD68									
Media ± DE	NA	NA	0,0 ± 0,01	NA	NA	NA	NA	NA	NA
Mediana	NA	NA	0,0	NA	NA	NA	NA	NA	NA
Intervalo	NA	NA	0,0-0,1	NA	NA	NA	NA	NA	NA

Tabla 9b. Células CD34 (%) Teñidas (+) por Componente Histológico (Estudio 2 y Estudio 3)

	PIN	Estroma	Tumor	Tumor/PIN	Tumor/estroma	PIN/estroma
Media ± DE	0,0 ± 0,04	0,0 ± 0,11	0,1 ± 0,18	0,0 ± 0,08	0,0 ± 0,08	0,0 ± 0,04
Mediana	0,0	0,0	0,0	0,0	0,0	0,0
Intervalo	0,0-0,4	0,0-1,7	0,0 - 0,9	0,0-0,6	0,0-0,4	0,0-0,3

Tabla 9c. Índice de Tinción (0 - 300) por Componente histológico (Estudio 2 y Estudio 3)

Marcador	Tumor	PIN	Glándula
AR			
Media ± DE	172.1 ± 75.3	79.6 ± 82.74	28.9 ± 67.25
Mediana	200	66,0	0
Intervalo	0 - 300	0 - 300	0 - 300
CK14			
Media ± DE	2.1 ± 6.32	34.4 ± 61.46	8.5 ± 32.62
Mediana	0	0	0
Intervalo	0 - 69	0 - 300	0 - 300
Ciclina D1			
Media ± DE	1.4 ± 6.99	0.0 ± 0.21	0.0 ± 0.0
Mediana	0	0	0
Intervalo	0-90	0-3	0-0
PSA			
Media ± DE	118.3 ± 71.10	139.4 ± 97.16	22.8 ± 55.14
Mediana	100	134	0
Intervalo	0 - 300	0 - 300	0 - 300
PSMA			
Media ± DE	0.2 ± 2.09	6.4 ± 21.02	2.9 ± 22.94
Mediana	0	0	0
Intervalo	0 - 21	0 - 154	0 - 300
p27			
Media ± DE	3.9 ± 8.20	6.4 ± 18.83	1.3 ± 8.65
Mediana	0	0	0
Intervalo	0-48	0 - 140	0-120
Her-2/neu			
Media ± DE	3.4 ± 16.69	0.2 ± 1.12	0.0 ± 0.00
Mediana	0	0	0
Intervalo	0 - 150	0 - 10	0-0

- 5 Se seleccionó una característica molecular única como predictiva de la recurrencia de PSA: Índice de tinción de AR - tumor.

Estudios analíticos y estadísticos. Utilizando la experiencia en el dominio y el arranque simple, el algoritmo encontró un subconjunto de 10 características (3 clinicopatológicas, 6 morfométricas y 1 molecular) que tenían un índice de concordancia (CI) de 0,87 (la Tabla 9, anteriormente, muestra la lista completa de las características seleccionadas). El resultado resultante del modelo SVRC también se puede interpretar como una estimación del riesgo relativo de recurrencia de PSA para un paciente individual. Usando los cuartiles de esta puntuación (<25%, >25%-75%, >75%), se crearon grupos de pacientes de riesgo; las estimaciones de recurrencia de Kaplan-Meier para cada grupo de riesgo según lo predicho por el modelo SVRC se presentan en la Figura 10. Los grupos mostraron una diferencia estadísticamente significativa en el tiempo hasta la recurrencia de PSA (prueba de orden logarítmico, valor p < 0,0001).

Estudio 3. Este estudio utilizó la misma cohorte que la del Estudio 2 para que las características clinicopatológicas de los pacientes fueran idénticas. En términos de resultados, diecinueve (7%) pacientes murieron por cualquier causa, mientras que los pacientes restantes (93%) estuvieron vivos desde su última visita y se censuraron. Para los pacientes censurados, el tiempo medio de seguimiento fue de 64,8 meses, o poco más de 5 años. No se alcanzó la mediana de tiempo total hasta la muerte. Se seleccionaron dos características clínicas como predictivas de muerte por cualquier causa: estadio clínico TNM y edad del paciente.

Análisis de imágenes y estudios de morfometría. En este estudio se utilizó el mismo conjunto de 350 características morfométricas del Estudio 2. La Figura 11 muestra que, de las 350 características, se seleccionaron 11 características morfométricas como predictivas de muerte por cualquier causa, donde estas características se relacionan con los objetos patológicos del estroma, glóbulos rojos y núcleos epiteliales (es decir, StromaMinMeanChannel10535, RedBloodCellMeanStddevChann30474, StromaMinMeanChannel20539, RedBloodCellMinMeanChannel20443, RedBloodCellStddevChann20472, StromaMaxMaxDiff0529, EpitheNucleMeanBordeLengtPx10206, EpithelialNucleiMeanAreaPx10194, EpithelNucleiStddevElliptFit0228, RedBloodCellStddevStddevChann30476, y RedBloodCellStddevElliptiFit0420, donde "canal" se refiere a los canales de colores rojo (R), verde (G) y azul (B) de una imagen). Más particularmente, en este estudio, se determinó que las características morfométricas del valor medio del canal de color rojo, el valor medio del canal de color azul y la diferencia máxima para el estroma se correlacionan con el resultado. Se determinó que las características morfométricas de la desviación media y estándar del canal rojo, la desviación media y estándar del canal verde y el ajuste elíptico para los glóbulos rojos estaban correlacionadas con el resultado. Para determinar la característica morfométrica del ajuste elíptico, se creó una elipse con la misma área que el glóbulo rojo, se comparó el área del glóbulo rojo fuera de la elipse con el área dentro de la elipse que no estaba llena con el glóbulo rojo, y se asignó un valor de 0 donde no había ajuste, mientras que se asignó un valor de 1 para un objeto de ajuste completo. Se determinó que las características morfométricas de la longitud del borde, el área y el ajuste elíptico para los núcleos epiteliales se correlacionan con el resultado.

Varios posibles motivos para al menos algunas de estas correlaciones se describieron anteriormente en relación con el Ejemplo 1 y/o el Estudio 1. Por ejemplo, la forma general de los núcleos epiteliales refleja una apariencia histológica de un mayor grado de Gleason. Además, en este estudio, la correlación con respecto al estroma puede explicarse por la comprensión de que el estroma exhibirá un contraste reducido (medido por la característica morfométrica de diferencia máxima) a medida que el cáncer progresa debido a su interrupción con las células epiteliales.

Análisis Molecular. Se utilizó el mismo conjunto de características moleculares del Estudio 2 en este estudio. Se seleccionó una característica única como predictiva de muerte por cualquier causa: Índice de tinción de PSA - glándula atrófica.

Estudios analíticos y estadísticos. En esta cohorte, se seleccionaron un total de 14 características (2 clinicopatológicas, 11 morfométricas y 1 molecular). El modelo final tuvo un índice de concordancia (CI) de 0,80. La lista completa de características seleccionadas se muestra en la Figura 11 y se enumera a continuación. Las características clínicas y moleculares seleccionadas se enumeran a continuación. Las descripciones de las características clínicas se proporcionan arriba.

Características Clínicas

1. Estadio TNM
2. edad

Característica molecular

1. psapsi: se refiere al índice de tinción del antígeno prostático específico (PSA) en la neoplasia intraepitelial prostática (PIN).

Cada número en la Figura 11 representa el índice de concordancia de un modelo predictivo basado en la característica correspondiente y todas las demás características de la Figura 11 que tienen números más pequeños. Por ejemplo, 0,6804 es el CI de un modelo basado en StromaMinMeanChannel10535 y 0,7362 es el CI cuando el modelo se basa en StromaMinMeanChannel10535 y TNM.

El resultado resultante del modelo SVRc también se puede interpretar como una estimación del riesgo relativo de muerte para un paciente individual. Usando los cuartiles de esta puntuación (<25%, >25%-75%, >75%), se crearon grupos de pacientes de riesgo; las estimaciones de recurrencia de Kaplan-Meier para cada grupo de riesgo según lo predicho por el modelo SVRc se presentan en la Figura 12. Usando la prueba de orden logarítmico, se observó una diferencia significativa en la supervivencia entre los grupos de riesgo ($p < 0,0001$).

Discusión de resultados (Ejemplo 2)

La reducción observada de las características seleccionadas (compuestas) del Estudio 1 (41) al Estudio 2 (10) mientras se conserva la precisión predictiva del modelo enfatizó la precisión y los atributos de filtrado que se lograron a través de diferentes algoritmos de aprendizaje automático. El índice de concordancia del modelo que se desarrolló en la

5 coorte de 268 pacientes fue de 0,87; en comparación, cuando se aplica el nomograma Kattan [20] a esta coorte, alcanzó un índice de concordancia de 0,78. Quizás lo más sorprendente es la capacidad del modelo anterior como se discutió en el Estudio 2 de clasificar correctamente a los pacientes con recurrencias tempranas de PSA (dentro de los 5 años) con una sensibilidad del 80%. En comparación, el nomograma de Kattan puede hacer la misma predicción con una sensibilidad de solo el 54%. Esto enfatiza aún más el papel que tal análisis predictivo proporcionaría en la toma de decisiones para la intervención temprana. Finalmente, el resultado del modelo presentado se puede usar para estimar la probabilidad de que un paciente presente recurrencia con el tiempo, en lugar de ofrecer una estimación única de la probabilidad de que un paciente presente recurrencia dentro de un número determinado de años, sin ninguna indicación de cuándo dentro de ese periodo de tiempo.

10 En el Estudio 3, el objetivo fue utilizar el conocimiento de dominio existente derivado del Estudio 2 y desarrollar un modelo predictivo para la supervivencia general. El resultado final eficaz fue la capacidad de predecir con un 80% de precisión la supervivencia general y el tiempo hasta la muerte de un individuo utilizando un total de 14 características de dominio combinado. Aunque limitado por el pequeño número de eventos (7% de muertes por cualquier causa) y la ausencia de un nomograma publicado comparable, los resultados respaldan aún más el uso de un enfoque de sistemas para desarrollar este tipo de análisis predictivos.

15 Se están realizando esfuerzos adicionales con respecto a la expansión de este análisis de 'supervivencia general' para incluir medidas clínicas de malos resultados (es decir, metástasis y/o muerte por cáncer de próstata) utilizando una población retrospectiva multiinstitucional con un estudio de validación externa independiente. Además, recientemente se ha iniciado un enfoque de 'Patología de sistemas' para analizar las biopsias con aguja de diagnóstico a fin de tener un impacto en las cuestiones de tratamiento antes de la cirugía.

20 El ejemplo anterior demuestra que se ha desarrollado con éxito una plataforma de 'Patología de sistemas' que integra características clínicas, morfometría de tejido tumoral y análisis moleculares. Al utilizar la experiencia en el dominio y la regresión de vectores de soporte para datos censurados (SVRc), se seleccionaron características de los tres dominios y se utilizaron para desarrollar un modelo predictivo para la recurrencia de PSA y la supervivencia general. Se entenderá que este enfoque novedoso de 'Patología de sistemas' tiene una amplia aplicación en el campo de la medicina personalizada en lo que se refiere al diagnóstico de tumores, el pronóstico del paciente y como una herramienta para predecir la respuesta a terapias específicas.

Ejemplo 3: Predicción de enfermedad agresiva posterior a la prostatectomía con datos clínicos y morfométricos

30 Este estudio se realizó para predecir una enfermedad agresiva (es decir, un fallo clínico como lo demuestra un escáner óseo positivo que representa cáncer de próstata metastásico en hueso) después de que un paciente se someta a una prostatectomía. Antes de la presente invención, no existían herramientas analíticas precisas para proporcionar tal predicción. Como se describió anteriormente, se ha demostrado que el enfoque de patología de sistemas de la presente invención predice con precisión la recurrencia de PSA. Este estudio demuestra que la presente invención también se puede usar para predecir con precisión la metástasis ósea distante después de prostatectomía.

35 Se estudió una coorte de 119 pacientes que se sometieron a prostatectomía radical incorporando micromatrices de tejidos (TMA) construidas a partir de muestras de prostatectomía. Se realizaron estudios morfométricos (es decir, análisis de imágenes) utilizando cortes de tejido teñidos con hematoxilina y eosina, y los determinantes biológicos se evaluaron con inmunohistoquímica (IHC) utilizando una serie de biomarcadores seleccionados por su relevancia biológica potencial para la progresión del cáncer de próstata. Un modelo predictivo para el fallo clínico (es decir, exploración ósea positiva) se obtuvo de un conjunto seleccionado de características a través del aprendizaje multivariante supervisado. Los pacientes con datos completos presentes (n = 116) en cada dominio fueron evaluados con una máquina de vectores de soporte para la regresión desarrollada para manejar datos censurados (SVRc). El rendimiento predictivo del modelo se estimó utilizando el índice de concordancia (CI) con puntuaciones generadas utilizadas para definir los grupos de riesgo.

45 De los 116 pacientes, se seleccionó un subconjunto de 61 pacientes en función de sus características clínicas, incluidos 20 individuos con fallo clínico identificados por metástasis ósea. Esta coorte se utilizó para crear un modelo para predecir la probabilidad de una exploración ósea positiva dentro de los 5 años posteriores a la prostatectomía. Se seleccionaron las siete características que se muestran en la Figura 13 (incluidas cuatro características clínicas y tres morfométricas) que predijeron el fallo clínico con un 89 por ciento de precisión y una sensibilidad y especificidad del 86 y 85 por ciento, respectivamente. Las características morfométricas seleccionadas se relacionaron con los objetos patológicos de citoplasma y luz. Más particularmente, las características morfométricas seleccionadas fueron el área de citoplasma dividida por el área de tejido total, el área de luz dividida por el área de tejido total y la desviación estándar del citoplasma del canal rojo medio. Las características clínicas se enumeran a continuación.

Características Clínicas

- 55
1. Extensión Extracapsular (ECE)
 2. Invasión de vesículas seminales (SVI)
 3. Grado de Gleason de Prostatectomía Dominante (PGG1)

4. Invasión de Ganglios Linfáticos (LNI)

Conclusión

La integración de las características clínicas con las características morfométricas dio como resultado el primer análisis pronóstico preciso para predecir el fallo clínico dentro de los 5 años posteriores a la prostatectomía. Como se describe, el análisis puede predecir con un 89% de precisión qué pacientes tienen más probabilidades de tener un fallo clínico (y cuándo) dentro de un período de 5 años después de la prostatectomía. Los resultados de añadir características moleculares a las características clínicas y morfométricas del modelo están actualmente pendientes.

Ejemplo 4: Datos morfométricos de toxicología hepática

Este estudio se realizó para demostrar la capacidad del análisis de imágenes y de la modelización estadística en el área de la toxicología. Específicamente, el estudio requirió la adquisición y el análisis de cortes de hígado de rata con el objetivo general de clasificar los cortes como normales o anormales. Ser capaz de automatizar este proceso y, al mismo tiempo, lograr un alto nivel de precisión de clasificación podría permitir la creación de una plataforma de alto rendimiento utilizada para cribar objetivamente las toxicidades en estudios preclínicos.

El estudio se dividió en dos fases. La fase inicial utilizó un conjunto de 100 cortes de hígado de rata como conjunto de entrenamiento; 80 cortes de hígado normales y 20 anormales. Este conjunto de cortes se utilizó para desarrollar una aplicación de análisis de imágenes utilizando el sistema de análisis de imágenes de tejidos descrito anteriormente, así como para realizar la selección de características y modelos para clasificar los cortes. El proceso de análisis de imágenes establecido se aplicó luego a un conjunto sin etiquetar de 100 cortes de hígado de rata en la segunda fase del estudio, en la que se ensayaron los modelos estadísticos diseñados en la fase de entrenamiento.

20 Precisión de segmentación

La precisión de segmentación global para todos los objetos, medida por la evaluación de un patólogo, fue del 80% al 90%.

Estadística

El componente estadístico del estudio incluyó dos pasos. El primer paso consistió en seleccionar características a partir de los datos de imágenes generados por el análisis de imágenes de los cortes. La reducción del número de características utilizadas para la clasificación puede mejorar la solidez y la fiabilidad de la clasificación de los cortes. El segundo paso consistió en entrenar un modelo usando el conjunto de características y etiquetas seleccionadas para cada corte (anormal, normal) y luego probar el modelo prediciendo la clasificación de un conjunto independiente de cortes de hígado de rata donde las etiquetas se desconocían.

30 Selección de características

Las mediciones estadísticas generadas para cada uno de los objetos anteriores fueron:

- Numero de objetos
- Área relativa (porcentaje, en relación con el área total de la imagen)
- Tamaño mínimo (en píxeles)
- 35 - Tamaño máximo (en píxeles)
- Tamaño promedio (en píxeles)
- Desviación estándar del tamaño.

Dado que se analizaron múltiples imágenes por corte, estas medidas se promediaron en todas las imágenes para un corte individual de hígado de rata. El número total de características originales fue de 378.

40 La selección de características también implicó dos pasos. El primer paso utilizó la experiencia del dominio. Un patólogo seleccionó características de la lista de características original generada por el análisis de imágenes de los cortes. La decisión de incluir o excluir características se basó en la comprensión de la patología del hígado y las posibles anomalías/toxicidades que podrían encontrarse. Del conjunto original de 378 características, se seleccionaron 90 características utilizando el conocimiento del dominio.

45 Estas características fueron examinadas usando un análisis discriminante por pasos para reducir aún más el número de características para la clasificación. Se asumió que el conjunto de características que formaban cada clase era multivariante normal con una matriz de covarianza común. Las características se eligieron para entrar o salir del modelo de acuerdo con el nivel de significancia de una prueba F a partir de un análisis de covarianza, donde las características ya elegidas actúan como covariables y la característica en consideración es la variable dependiente.

Se utilizó un nivel de significación de 0,15.

- La selección gradual comenzó sin características en el modelo. En cada paso, se examinó el modelo.
- Si la característica en el modelo que contribuyó menos al poder discriminatorio del modelo medido por la lambda de Wilks (el criterio de probabilidad) no cumplió con el criterio para permanecer, entonces esa característica se eliminó.
- De lo contrario, se introdujo la característica que no estaba en el modelo que más contribuyó al poder discriminatorio del modelo.
- Cuando todas las características en el modelo cumplieron con el criterio para permanecer y ninguna de las otras características cumplió con el criterio para introducir las, el proceso de selección por pasos se detuvo.

10 Clasificación/Entrenamiento del Modelo

Las características seleccionadas se introdujeron luego en un análisis discriminante lineal (LDA) que clasificó cada uno de los cortes de hígado como anormales o normales. La salida del modelo se corrigió con respecto a un sesgo potencial mediante validación cruzada.

15 Las redes neuronales también se exploraron como clasificador. Las características seleccionadas se utilizaron como entradas para el modelo de red neuronal, que es una estructura estándar de perceptrón multicapa (MLP) con cero unidades ocultas y conexión directa entre las capas de entrada y salida. El modelo se entrenó tratando de maximizar directamente una aproximación al área bajo la curva ROC, que se explica a continuación. Se descubrió que el modelo MLP entrenado mediante este criterio logra una mejor precisión que un modelo MLP entrenado mediante los criterios típicos, por ejemplo, error cuadrático medio y entropía cruzada.

20 La salida de ambos modelos se utilizó para crear una curva característica operativa del receptor (ROC) eligiendo un valor diferente de la salida del modelo como punto de corte, calculando la sensibilidad y especificidad para cada punto de corte, y trazándolos en un diagrama bidimensional (sensibilidad en el eje y y especificidad en el eje x). El área bajo la curva ROC (AUC) utiliza ambas medidas para evaluar la precisión de cada modelo, y puede interpretarse como la capacidad del modelo de clasificar correctamente los cortes de hígado como anormales o normales. Típicamente, la sensibilidad y la especificidad se describen en términos de la tasa positiva verdadera y la tasa negativa verdadera, respectivamente. Por lo tanto, en el contexto de este estudio, la clase anormal se consideró un resultado "positivo", mientras que la clase normal se consideró un resultado "negativo". La sensibilidad, por lo tanto, es la tasa positiva verdadera, es decir, la proporción de cortes de hígado correctamente clasificados como anormales; la especificidad, por otro lado, es la tasa negativa verdadera, es decir, la proporción de cortes de hígado correctamente clasificados como normales.

A partir de las curvas ROC, las sensibilidades y especificidades seleccionadas del conjunto de entrenamiento se proporcionan en la sección de Resultados a continuación.

Análisis del modelo

35 Una vez desarrollados, los parámetros tanto de la función discriminante lineal como de la red neuronal se bloquearon. Al recibir las mediciones estadísticas del conjunto de análisis de imágenes de hígado de rata, ambos clasificadores se aplicaron usando un punto de corte individual estimado usando los resultados de validación cruzada de cada uno de los resultados del modelo, respectivamente. Los puntos de corte correspondieron a una sensibilidad del 100% y una especificidad del 90% (ambos basados en la validación cruzada) para una futura aplicación de grado industrial. Para la evaluación inicial de este conjunto de validación externa de hígados, otra parte llevó a cabo la evaluación de las precisiones de los modelos sin enmascaramiento sobre la verdadera clasificación de los cortes de hígado. Esta otra parte también proporcionó la clave de análisis para verificar los resultados.

Resultados

45 El área bajo la curva ROC para ambos modelos está muy cerca de 1, lo que indica una discriminación casi perfecta entre cortes de hígado anormales y normales. La función derivada usando LDA tiene un AUC de 0,99; la función derivada usando redes neuronales tiene un AUC de 0,98.

También se observó en las curvas ROC la sensibilidad y especificidad de cada modelo, dependiendo del punto de corte aplicado a los resultados del modelo para clasificar un corte de hígado como anormal o normal. La Tabla 10 resume una selección de pares de sensibilidad-especificidad.

Tabla 10

LDA		NN	
Especificidad	Sensibilidad	Especificidad	Sensibilidad
100%	65%	100%	65%
99%	75%	99%	70%
98%	100%	98%	85%

Análisis

5 Las etiquetas de las claves de análisis se compararon con las clasificaciones predichas de la función discriminante lineal y las de las redes neuronales. Según la clave, los resultados se resumen en las Tablas 11a y 11b de la siguiente manera:

Tabla 11a

Etiqueta de la clave de análisis			
		Anormal	Normal
Etiqueta LDA	Anormal	42 (TP)	19 (FP)
	Normal	7 (FN)	32 (TN)
		49	51 100

Sensibilidad = $TP / (TP + FN) \times 100 = 42 / (42 + 7) \times 100 = (42/49) \times 100 = 86\%$
 Especificidad = $TN / (FP + TN) \times 100 = 32 / (19 + 32) \times 100 = (32/51) \times 100 = 63\%$

Tabla 11b

Etiqueta de la clave de análisis			
		Anormal	Normal
Etiqueta NN	Anormal	36 (TP)	19 (FP)
	Normal	13 (FN)	32 (TN)
		49	51 100

Sensibilidad = $TP / (TP + FN) \times 100 = 36 / (36 + 13) \times 100 = (36/49) \times 100 = 73\%$
 Especificidad = $TN / (FP + TN) \times 100 = 32 / (19 + 32) \times 100 = (32/51) \times 100 = 63\%$

10

El punto de corte utilizado para el clasificador LDA fue igual a 0,0031; el punto de corte utilizado para el clasificador NN fue igual a 0,0002. Ambos corresponden a los requisitos del sistema de 100% de sensibilidad y 90% de especificidad.

Discusión

15 Basándose en la sensibilidad y especificidad de cada clasificador después de aplicarlos al conjunto de análisis, LDA superó a NN. El clasificador LDA logró una sensibilidad del 86%, lo que significa que este clasificador etiquetó correctamente los cortes anormales de hígado de rata como anormales el 86% de las veces, al contrario que el clasificador de la red neuronal que logró una sensibilidad del 73%. La especificidad para ambos clasificadores fue del 63%. Tanto la sensibilidad como la especificidad de cada modelo son inferiores a las observadas anteriormente, pero esto no es sorprendente, ya que generalizar cualquier clasificador a un conjunto externo a menudo conduce a una caída en su precisión. Este estudio demostró la aplicación eficaz de las tecnologías para la imagenología y la modelización estadística.

20

Ejemplo 5: Predicción de la recurrencia del cáncer de próstata con datos clínicos, morfométricos y moleculares

25 Se realizó otro estudio para generar un modelo que predice el tiempo hasta la recurrencia del cáncer de próstata en pacientes que se han sometido a una prostatectomía radical. Al igual que con el Ejemplo 2, el tiempo hasta la

recurrencia se definió como el tiempo (en meses) desde la prostatectomía radical hasta la recurrencia de PSA (bioquímico). Los cortes de tejido prostático utilizados en este estudio estaban compuestos predominantemente de tumor, pero también incluían elementos benignos.

5 Este estudio se basó en información para las mismas 17 características clínicas (Tabla 4) y 43 características moleculares (Tabla 6) evaluadas en el Ejemplo 2. El conjunto de 496 características morfométricas (Tabla 1) se redujo a las 38 características que se muestran en la Tabla 10 (adjunta en el presente documento) en base, por ejemplo, al conocimiento experto y la experimentación adicional en el campo de la recurrencia del cáncer de próstata. La información clínica, molecular y morfométrica de 262 pacientes de la cohorte de 539 pacientes descrita anteriormente en relación con el Ejemplo 2 se evaluó en este estudio. Además del filtrado que redujo el número de características morfométricas de 496 a 38, la principal diferencia entre este estudio y el Ejemplo 2 es que este estudio utilizó un método de reducción de características SVRC para la selección de características. La reducción de características de SVRC se describe en la solicitud de patente de EE.UU. de propiedad compartida n° 11/438.789, presentada el 22 de mayo de 2006, que se incorpora en la presente memoria como referencia en su totalidad.

RESULTADOS

15 Como resultado del estudio, se generó un modelo final basado en 6 características (3 clínico-patológicas, 1 molecular y 2 morfométricas) y que tenía un índice de concordancia (CI) de 0,83. Las 6 características incluidas en el modelo se muestran en la Figura 14, junto con sus respectivas contribuciones de características al modelo final. Las tres características clínicas seleccionadas como predictivas de recurrencia de PSA fueron la afectación de las vesículas seminales (contribución de la característica = -5,2103), afectación del margen quirúrgico (-7,3159) y afectación de los ganglios linfáticos (-9,3742). La única característica molecular seleccionada fue el índice de tinción del receptor de andrógenos (AR) presente en el tumor (-3,5404). Estas características clínicas y moleculares se describen anteriormente en relación con el Ejemplo 2. Las dos características morfométricas seleccionadas fueron el área ocupada por núcleos de células epiteliales dividida por el área de tejido total (3,2975) y el área ocupada por el estroma dividida por el área de tejido total (-0,34225). Por ejemplo, el área de tejido total puede incluir la suma (en píxeles, siendo 1920000 el máximo donde el tamaño de la imagen es 1200x1600 píxeles) de las áreas de citoplasma, núcleo epitelial, luz, glóbulos rojos, núcleo estromal, estroma y artefactos. Las posibles razones para la selección de las características morfométricas relacionadas con los núcleos epiteliales y las células del estroma también se describen a continuación en relación con el estudio de validación.

El modelo de entrenamiento final tenía una sensibilidad del 82% y una especificidad del 81% para predecir correctamente la recurrencia del cáncer de próstata antes de los 5 años. El resultado resultante del modelo SVRC también se puede interpretar como una estimación del riesgo relativo de recurrencia de PSA para un paciente individual. Usando los cuartiles de esta puntuación (<25%, >25% -75%, >75%), se crearon grupos de pacientes de riesgo y se generaron estimaciones de recurrencia de Kaplan Meier para cada grupo de riesgo. Los grupos mostraron una diferencia estadísticamente significativa en el tiempo hasta la recurrencia de PSA (prueba de orden logarítmico, valor $p < 0,0001$).

Estudio de validación

El modelo final fue validado con una cohorte externa compuesta por 61 pacientes. El modelo final produjo un CI de 0,80, sensibilidad del 91% y especificidad del 70% para la validación para identificar pacientes en riesgo de experimentar recurrencia del cáncer de próstata en los primeros 5 años.

40 Para comprender mejor la importancia de las dos características morfométricas seleccionadas en el modelo final, se generaron curvas de Kaplan-Meier para cada característica morfométrica. Se observó que cantidades crecientes de estroma ($p = 0,004$) y núcleos epiteliales (aunque no estadísticamente significativos, $p = 0,28$) se asociaron independientemente con un resultado favorable. Esto planteó la posibilidad de que las mediciones cuantitativas derivadas de estos patrones de imagen puedan representar determinantes más objetivos del sistema de clasificación de Gleason.

Este estudio demostró que solo se requiere un conjunto limitado de características clínicas, moleculares y morfométricas para crear un análisis predictivo clínicamente útil. Esta reducción de características se logró al tiempo que se conservaba la precisión predictiva del modelo.

Ejemplo 6: Predicción de recurrencia del cáncer de próstata con datos clínicos, morfométricos y moleculares

50 Sin embargo, se realizó otro estudio para generar un modelo que predice el tiempo hasta la recurrencia del cáncer de próstata en pacientes que se han sometido a una prostatectomía radical. Como en los ejemplos 2 y 6, el tiempo hasta la recurrencia se definió como el tiempo (en meses) desde la prostatectomía radical hasta la recurrencia de PSA (bioquímico). Los cortes de tejido prostático utilizados en este estudio estaban compuestos predominantemente de tumor, pero también incluían elementos benignos.

55 Este estudio se basó en información para las mismas 17 características clínicas (Tabla 4) y 43 características moleculares (Tabla 6) evaluadas en los Ejemplos 2 y 5. El conjunto de 496 características morfométricas (Tabla 1) se redujo a las 33 características que se muestran en la Tabla 11 (adjunta en la presente memoria). La información clínica,

molecular y morfológica para los mismos 262 pacientes del Ejemplo 5 se evaluó en este estudio. Este estudio utilizó el mismo método de reducción de características SVRc mencionado anteriormente en el ejemplo 5.

RESULTADOS

5 Como resultado del estudio, se generó un modelo final basado en 8 características (las mismas 6 características seleccionadas en el Ejemplo 5 más 1 característica clínico-patológica adicional y 1 característica morfológica adicional) y con un índice de concordancia (CI) de 0,86. Las 8 características incluidas en el modelo se muestran en la Figura 15, junto con sus respectivas contribuciones de características al modelo final. La característica clínica adicional seleccionada en este estudio fue la puntuación de Gleason de biopsia (-10,60), descrita anteriormente en relación con el Ejemplo 2. La característica morfológica adicional seleccionada en este estudio fue la variación en la
10 textura dentro del estroma como se expresa en el canal rojo (-11,26). Esta característica, que indica la variación en la textura del estroma en función de sus propiedades de tinción, probablemente refleja los atributos bioquímicos del estroma asociado con el tumor en oposición a los elementos benignos.

Estudio de validación

15 El modelo final fue validado con una cohorte externa que constaba de 366 pacientes. El modelo final produjo para la validación un CI de 0,82, sensibilidad del 96% y especificidad del 72% para identificar a los pacientes en riesgo de experimentar recurrencia del cáncer de próstata en los primeros 5 años. La Tabla 12 a continuación muestra las características clínicas observadas para las cohortes de entrenamiento y validación. Las tablas 13a-c muestran las características moleculares de biomarcadores observadas a partir de la cohorte de entrenamiento.

Tabla 12

Característica	Entrenamiento	Validación
N	262	61
Edad (años)		
Media	62	61
Mediana	63	62
Intervalo	38-81	42-74
Raza		
Caucásico	235 (89,7%)	58 (95,1%)
Afroamericanos (hispanos y no hispanos)	21 (8,0%)	2 (3,3%)
Otro/Desconocido	6 (2,3%)	1 (1,6%)
PSA preoperatorio (ng/mL)		
Media	10,7	12,9
Mediana	7,8	10,0
Intervalo	0,9-81,9	2,0-69,5
Estadio TNM patológico		
T2N0	158 (60,3%)	No recogido
T3aN0	70 (26,7%)	No recogido
T3bN0	17 (6,5%)	No recogido
T1-3N+	17 (6,5%)	No recogido
Ausente	0	No recogido
Estadio UICC		
T1a < 5%	1 (0,4%)	No recogido
T1b > 5%	1 (0,4%)	No recogido
T1c no palpable ni visible	113 (43,1%)	No recogido
T2a < © lóbulo	54 (20,7%)	No recogido
T2b < 1 lóbulo	43 (16,4%)	No recogido
T2c ambos lóbulos	33 (12,6%)	No recogido
T3a ECE unilateral	15 (5,7%)	No recogido
T3c SV+	2 (0,8%)	No recogido
Ausente	0	No recogido

ES 2 745 811 T3

Resultado del tacto rectal		
No palpable	122 (46,6%)	32 (52,5%)
Palpable	140 (53,4%)	29 (47,5%)
Ausente	0	0
Afectación de ganglios linfáticos		
Negativo	246 (93,9%)	56 (91,8%)
Positivo	16 (6,1%)	5 (8,2%)
Ausente	0	0
Afectación de vesículas seminales		
No	233 (88,9%)	51 (83,6%)
Sí	29 (11,1%)	10 (16,4%)
Ausente	0	0
Márgenes quirúrgicos		
Negativo	216 (82,4%)	36 (59,0%)
Positivo	46 (17,6%)	25 (41,0%)
Afectación extracapsular		
No	159 (60,7%)	43 (70,5%)
Sí	103 (39,3%)	18 (29,5%)
Ausente	0	0
Grado de Gleason de biopsia dominante		
1	1 (0,4%)	0 (0,0%)
2	43 (16,4%)	0 (0,0%)
3	181 (69,1%)	39 (63,9%)
4	36 (13,7%)	22 (36,1%)
5	1 (0,4%)	0 (0,0%)
Ausente	0	0
Puntuación de Gleason de biopsia		
2	1 (0,4%)	0 (0,0%)
3	0 (0,0%)	0 (0,0%)
4	7 (2,7%)	0 (0,0%)
5	56 (21,4%)	3 (4,9%)
6	97 (37,0%)	27 (44,3%)
7	85 (32,4%)	20 (32,8%)
8	13 (5,0%)	8 (13,1%)
9	3 (1,2%)	3 (4,9%)
Ausente	0	0
Grado de Gleason de muestra dominante		
2	20 (7,6%)	0 (0,0%)
3	198 (75,6%)	34 (55,7%)
4	44 (16,8%)	23 (37,7%)
5	0 (0,0%)	4 (6,6%)
Puntuación de Gleason de la muestra		
5	21 (8,0%)	1 (1,6%)
6	86 (32,8%)	8 (13,1%)
7	144 (55,0%)	37 (60,7%)
8	11 (4,2%)	7 (11,5%)
9	0 (0,0%)	8 (13,1%)

ES 2 745 811 T3

Ploidía		
Diploide	141 (53,8%)	No recogido
Tetraploide	113 (43,1%)	No recogido
Aneuploide	8 (3,1%)	No recogido
Ausente	0 (0,0%)	No recogido
Porcentaje de ploidía en fase S (%)		
Media	2,4	No recogido
Mediana	1,2	No recogido
Intervalo	0,0-66,4	No recogido
Porcentaje de fracción de ploidía		
Media	3,4	No recogido
Mediana	2,4	No recogido
Intervalo	0,0-20,0	No recogido

Tabla 13a. Porcentaje de tinción de células, por componente histológico e intensidad de tinción (conjunto de entrenamiento)

Marcador	Tumor			PIN			Glándula Atrófica		
	1 +	2+	3+	1 +	2+	3+	1 +	2+	3+
Ki-67									
Media ± DE	22,0 ± 30,4	7,2 ± 17,1	1,8 ± 4,0	23,0 ± 31,5	7,8 ± 18,3	2,0 ± 4,5	1,3 ± 8,05	1,2 ± 9,9	0,3 ± 1,6
Mediana	0,7	0,0	0,0	1,0	0,0	0,0	0,0	0,0	0,0
Intervalo	0,0-100,0	0,0-100,0	0,0-26,3	0,0-100,0	0,0-100,0	0,0-39,5	0,0-96,0	0,0-96,5	0,0-13,0
CK18									
Media ± DE	NA	NA	100,0 ± 0,04	NA	NA	100,0 ± 0,04	NA	NA	100,0 ± 0,00
Mediana	NA	NA	100,0	NA	NA	100,0	NA	NA	100,0
Intervalo	NA	NA	50,0-100,0	NA	NA	50,0-100,0	NA	NA	100,0-100,0
CD45									
Media ± DE	NA	NA	0,0 ± 0,04	NA	NA	0,0 ± 0,01	NA	NA	0,0 ± 0,00
Mediana	NA	NA	0,0	NA	NA	0,0	NA	NA	0,0
Intervalo	NA	NA	0,0-0,4	NA	NA	0,0-0,1	NA	NA	0,0-0,0
CD68									
Media ± DE	NA	NA	0,0 ± 0,01	NA	NA	NA	NA	NA	NA
Mediana	NA	NA	0,0	NA	NA	NA	NA	NA	NA
Intervalo	NA	NA	0,0-0,1	NA	NA	NA	NA	NA	NA

Tabla 13b. Porcentaje de células con tinción de CD34, por componente histológico (entrenamiento)

	PIN	Estroma	Tumor	Tumor/PIN	Tumor/estroma	PIN/estroma
Media \pm DE	0,0 \pm 0,04	0,0 \pm 0,11	0,1 \pm 0,18	0,0 \pm 0,07	0,0 \pm 0,08	0,0 \pm 0,04
Mediana	0,0	0,0	0,0	0,0	0,0	0,0
Intervalo	0,0-0,4	0,0-1,7	0,0-0,9	0,0-0,5	0,0-0,4	0,0-0,3

Tabla 13c. Índice de tinción por componente histológico (conjunto de entrenamiento)

Marcador	Tumor	PIN	Glándula
AR			
Media \pm DE	171,8 \pm 75,9	79,9 \pm 83,3	29,5 \pm 67,9
Mediana	200	66,0	0
Intervalo	0-300	0-300	0-300
CK14			
Media \pm DE	2,2 \pm 6,4	35,2 \pm 62,0	8,3 \pm 32,6
Mediana	0	0	0
Intervalo	0-69	0-300	0-300
Ciclina D1			
Media \pm DE	1,4 \pm 7,1	0,0 \pm 0,21	0,0 \pm 0,0
Mediana	0	0	0
Intervalo	0-90	0-3	0-0
PSA			
Media \pm DE	117,9 \pm 71,2	140,5 \pm 97,4	22,4 \pm 54,9
Mediana	100	134	0
Intervalo	0-300	0-300	0-300
PSMA			
Media \pm DE	0,3 \pm 2,1	5,8 \pm 19,8	3,0 \pm 23,2
Mediana	0	0	0
Intervalo	0-21	0-154	0-300
p27Kip1			
Media \pm DE	3,9 \pm 8,2	6,6 \pm 19,0	1,3 \pm 8,7
Mediana	0	0	0
Intervalo	0-48	0-140	0-120
Her-2/neu			
Media \pm DE	3,5 \pm 16,9	0,2 \pm 1,1	0,0 \pm 0,0
Mediana	0	0	0
Intervalo	0-150	0-10	0-0

- 5 Ejemplo 7: Predicción de fallo clínico posterior a la prostatectomía radical con datos clínicos, morfométricos y moleculares

Este estudio generó un modelo para predecir el fallo clínico en pacientes con cáncer de próstata que se han sometido a prostatectomía radical. En este estudio, el fallo clínico se definió como el desarrollo de enfermedad metastásica y/o enfermedad independiente de andrógenos (p. ej., escaneo óseo positivo para metástasis o aumento de PSA durante la ADT después de la cirugía). Esto contrasta con el estudio de fallo clínico descrito en el Ejemplo 2, en el que el fallo clínico se definió como la muerte por cualquier causa.

En este estudio se evaluó la información clínica y la información morfométrica para 345 pacientes. Se evaluaron diez (10) características clínicas y 27 características morfométricas similares a las características clínicas y morfométricas que se muestran en las Tablas 4 y 6. También se evaluaron once (11) características moleculares. Sin embargo, estas características moleculares se generaron en un ordenador a partir de imágenes de tejido sometido a detección de

inmunofluorescencia (IF), y no se basaron en IHC como con las características moleculares descritas en el Ejemplo 2.

Más particularmente, se usaron anticuerpos marcados con fluorocromo Alexa para el receptor de andrógenos (AR), racemasa (AMACR), citoqueratina 18 (CK18), TP73L (p63) y queratina de alto peso molecular, junto con DAPI, en un ensayo 'quint-plex'. Basándose en los perfiles espectrales distintivos de los fluorocromos, se adquirieron imágenes en escala de grises específicas de antígeno. Se generaron secuencias de comandos para el producto Definiens Cellenger para localizar los antígenos individuales. A partir de la distribución e intensidad del antígeno, las secuencias de comandos identificaron tipos de células y compartimentos celulares (por ejemplo, células epiteliales lumbales, núcleos epiteliales/estromales) y se cuantificó AR y AMACR en tumor de próstata, glándulas benignas y estroma. Es decir, se generaron secuencias de comandos para el producto Definiens Cellenger que segmentaron imágenes de DAPI y CK18 en núcleos y objetos de citoplasma válidos, respectivamente. Estas secuencias de comandos clasificaron un objeto de imagen como núcleos o citoplasma en función de un umbral de intensidad, donde el umbral de intensidad fue una función lineal de las características de imagen de la escala de grises del biomarcador (p. ej., intensidad media, desviación estándar, cuantiles). Cada secuencia de comandos/función lineal fue específica para un biomarcador concreto, y se diseñó utilizando el aprendizaje supervisado (configuración del umbral por parte de un experto) y la regresión lineal entre umbrales definidos por un experto y características de imagen. Los objetos de citoplasma identificados sirvieron como objetos de anclaje para los núcleos divididos en objetos epiteliales y estromales. Los biomarcadores AR y AMACR también se segmentaron usando colocaciones espaciales y de intensidad. La localización espacial identificó un biomarcador dentro del compartimento especificado: núcleos epiteliales y estromales para AR y células epiteliales (citoplasma) para AMACR. La función de umbral similar a la utilizada para DAPI y CK18 permitió una clasificación adicional de la señal de biomarcador dentro de un compartimento como válida o como ruido. Los pacientes se excluyeron si faltaban sus núcleos en la matriz o si la muestra contenía solo estroma.

Debido a que la mayoría de los pacientes tuvieron múltiples núcleos de los cuales se extrajeron las características de IF, se ideó un procedimiento para agregar los valores de las características en múltiples núcleos, para cada paciente y característica. Se consideraron cuatro funciones candidatas (min, max, mediana y media) (por ejemplo, si un paciente tuvo 3 núcleos, la función min devuelve el valor más bajo de los 3 núcleos). Para una característica dada, cada una de estas funciones se aplicó para agregar los valores centrales de cada paciente en el conjunto de entrenamiento; luego se calculó el índice de concordancia para cada función de agregación para evaluarlo como un predictor de fallo clínico. Se consideró que la mejor función de agregación para una característica fue aquella cuyo índice de concordancia estuvo más alejado del aleatorio (0,5). La Tabla 14 siguiente enumera las 11 características de IF y las funciones de agregación seleccionadas correspondientes, y la Tabla 15 describe las características. La función de agregación seleccionada se usó para generar un valor único para cada característica de IF para cada paciente. Los detalles adicionales con respecto a la inmunofluorescencia múltiple se describen en la solicitud de patente de EE.UU. de propiedad compartida n° , titulada "Análisis inmunohistoquímico in situ multiplex", presentada el 13 de octubre de 2006, que se incorpora en la presente memoria como referencia en su totalidad.

Tabla 14.

Característica de IF	Función de agregación seleccionada
EpitarposamaMeanMeanChannel30017	media
EpitarposamaMeanMeanChannel30021	mediana
Mamacrpostotalarea0045	max
Mcytoplasmamacrpostotalarea0050	max
SumMen02_10totalarea	no aplicable; esta es una característica derivada cuyo valor es el promedio de los valores agregados de 10 características principales
Mepithnucleiarposamacrpostot0085	max
RelAEpithNucARpos	mediana
RelAEpithNucARposAMACRpos	max
MeanEpithNucARPosIntensity	min
MeanEpithNucARPosAMACRPosIntens	max
NormSimpleAverageTotalEpithLum	min

Tabla 15

Característica	Descripción
EpitarposamaMeanMeanChannel30017	Intensidad media de los objetos DAPI clasificados como Núcleos Epiteliales expresados positivamente con biomarcador AR y ubicados en el citoplasma negativo con respecto al biomarcador AMACR
EpitarposamaMeanMeanChannel30021	Intensidad media de los objetos DAPI clasificados como Núcleos Epiteliales expresados negativamente con biomarcador AR y ubicados en el citoplasma positivo con respecto al biomarcador AMACR
Mamacrpostotalarea0045	Área total de los objetos AMACR con la intensidad por encima del umbral de ruido/señal
Mcytoplasmamacrpostotalarea0050	Área total de los objetos de citoplasma (CK18) expresados positivamente con AMACR
SumMen02_10totalarea	<p>En cada objeto DAPI clasificado como núcleo epitelial y expresado positivamente con el biomarcador AR, se calcula la suma de los valores normalizados de intensidad del objeto AR $\tilde{Y}_k: I = \sum_k \tilde{Y}_k$ Donde $\tilde{Y}_k = Y_k/Y_*$ es la proporción de la intensidad k del objeto AR real Y_k y los valores Y_* de umbral de intensidad de ruido frente a señal locales (con respecto a una imagen de AR). Los valores I están asociados con la cantidad total de la proteína detectada por el biomarcador. Los valores I están sujetos a agrupación. Los núcleos epiteliales se clasifican en $M = 11$ agrupaciones. Cada agrupación caracteriza una cierta cantidad de AR en los núcleos epiteliales.</p> <p>La característica es el área de los núcleos epiteliales con $20 \leq I < 30$. Los valores del área se dividen en el área total de los núcleos epiteliales detectados en la imagen.</p>
Mepithnuclearposamacrpostot0085	Área total de los núcleos epiteliales expresada positivamente con AR y positiva con respecto a AMACR
RelAEpithNucARpos	Área relativa de los núcleos epiteliales con AR+ con respecto al área total de los núcleos epiteliales
RelAEpithNucARposAMACRpos	Fracción (área relativa) de los núcleos epiteliales AR+ con AMACR+ con respecto a los núcleos epiteliales AR+
MeanEpithNucARPosIntensity	Valor de la intensidad sumada (valores I) para los objetos AR en núcleos epiteliales calculado como promedio de los valores de agrupación de I . La característica se puede interpretar como la cantidad promedio de AR expresada en los núcleos epiteliales en una imagen (región de interés)
MeanEpithNucARPosAMACRPosIntens	Valor de la intensidad sumada (valores I) para los objetos AR en los núcleos epiteliales AR+ y AMACR+ estimado a través del valor medio de la intensidad. La característica es proporcional a la cantidad promedio de AR expresada en los núcleos epiteliales ubicados en el citoplasma positivo con respecto al marcador AMACR.
NormSimpleAverageTotalEpithLum	Se introdujo el concepto de luminancia (energía luminosa total emitida por el objeto). Para cada objeto DAPI clasificado como núcleo epitelial se calculó la luminancia (intensidad de AR media multiplicada en el área del objeto). La característica es una estimación de la cantidad de AR en los núcleos epiteliales calculada mediante los valores de luminancia.

Todas las características de IF, excepto 'AR nuclear presente dentro de las células epiteliales que son negativas para AMACR', mostraron asociación con el fallo clínico en el análisis univariante basado en el índice de concordancia ($CI \leq$

0,4 o CI \geq 0,6). La Tabla 15 a continuación describe varios

RESULTADOS

5 El modelo resultante de este estudio se basó en 7 características (3 clínicas, 1 molecular y 3 morfométricas) y tuvo un índice de concordancia (CI) de 0,91, una sensibilidad del 95% y una especificidad del 80% en la cohorte de entrenamiento. Las 7 características incluidas en el modelo se muestran en la Figura 16. Las 3 características clínicas seleccionadas como predictivas del fallo clínico después de la prostatectomía fueron la puntuación de Gleason de Biopsia, la Afectación de Ganglios Linfáticos y la Puntuación de Gleason de Muestras (prostatectomía). Las 3 características morfométricas fueron la intensidad media del citoplasma epitelial como se expresa en el canal azul (CytoplasmMeanMeanChannel60060), la variación en la textura dentro del estroma como se expresa en el canal rojo (StromaMeanStddevChannel40310) y la variación en la textura entre núcleos epiteliales como se expresa en el canal rojo (EpitheNucleiStddevMeanChann40157). Como se describió anteriormente, la característica relacionada con la textura del estroma se basa en sus propiedades de tinción, y lo más probable es que refleje los atributos bioquímicos del estroma asociado con el tumor en oposición a los elementos benignos. Además, típicamente cuando el color del citoplasma cambia de azul claro a azul oscuro, refleja cambios en el tejido debido al desarrollo del cáncer, es decir, invasión de células epiteliales en áreas del estroma. Con respecto a las variaciones de textura dentro de los núcleos epiteliales (textura de cromatina plegada/desplegada y nucleolos), el desarrollo del cáncer generalmente se caracteriza por un mayor número de núcleos epiteliales con textura de cromatina desplegada, así como un mayor número de nucleolos, lo que da como resultado valores más altos para esta característica. La única característica molecular seleccionada fue la intensidad de AR dentro de las células epiteliales que fueron positivas a AMACR.

20 Validación

El modelo final se validó con una cohorte independiente que constó de 319 pacientes. El modelo final produjo para la validación un CI de 0,85, sensibilidad del 89% y especificidad del 77% para predecir el fallo clínico después de la prostatectomía. La Tabla 16 a continuación muestra las características clínicas observadas para las cohortes de entrenamiento y validación.

25 Tabla 16. Información clínica

Característica	Entrenamiento	Validación
Raza		
Blanco (hispano y no hispano)	328 (95,1)	301 (94,4)
Afroamericano (hispano y no hispano)	11 (3,2)	9 (2,8)
Otro/Desconocido	6 (1,7)	9 (2,8)
PSA preoperatorio (ng/ml)		
Media	11,4	11,1
Mediana	7,9	8,2
Intervalo	0,5-100,0	1,1-56,2
Estadio TNM clínico		
T1a/b	4 (1,2)	4 (1,3)
T1c	172 (49,9)	148 (46,4)
T2a	64 (18,6)	59 (18,5)
T2b	29 (8,4)	26 (8,2)
T2c	67 (19,4)	77 (24,1)
T3	9 (2,6)	5 (1,6)
Afectación de ganglios linfáticos		
Negativo	332 (96,2)	308 (96,6)
Positivo	13 (3,8)	11 (3,4)
Afectación de vesículas seminales		
No	316 (91,6)	291 (91,2)
Sí	29 (8,4)	28 (8,8)
Márgenes quirúrgicos		
Negativo	228 (66,1)	199 (62,4)
Positivo	117 (33,9)	120 (37,6)
Afectación extracapsular		

No	242 (70,1)	222 (69,6)
Sí	103 (29,9)	97 (30,4)
<hr/>		
Grado de Gleason de biopsia dominante		
1	2 (0,6)	1 (0,3)
2	22 (6,4)	20 (6,3)
3	264 (76,5)	247 (77,4)
4	56 (19,4)	51 (16,0)
5	9 (2,6)	0 (0,0)
<hr/>		
Puntuación de Gleason de biopsia		
2	1 (0,3)	0 (0,0)
3	2 (0,6)	1 (0,3)
4	7 (2,0)	12 (3,8)
5	28 (8,1)	27 (8,5)
6	186 (53,9)	159 (49,8)
7	93 (27,0)	97 (30,4)
8	23 (6,7)	20 (6,3)
9	5 (1,5)	3 (0,9)
10	0 (0,0)	0 (0,0)
<hr/>		
Grado de Gleason de Muestra Dominante		
2	9 (2,6)	8 (2,5)
3	269 (78,0)	243 (76,2)
4	62 (18,0)	63 (19,8)
5	5 (1,5)	5 (1,6)
<hr/>		
Puntuación de Gleason de la muestra		
5	16 (4,6)	16 (5,0)
6	112 (32,5)	102 (32,0)
7	182 (52,8)	167 (52,4)
8	22 (6,4)	18 (5,6)
9	12 (3,5)	16 (5,0)
10	1 (0,3)	0 (0,0)

5 En otro aspecto, basado en una evaluación de un subconjunto de pacientes que se trataron con ADT después de la prostatectomía, se ha determinado que los niveles crecientes de AR pueden estar asociados con un tiempo más corto para el fallo clínico después del tratamiento con ADT. Por lo tanto, las mediciones del contenido de AR en la muestra de prostatectomía podrían usarse para predecir la respuesta a un tipo de terapia de supresión de andrógenos. En resumen, el contenido de AR en el momento de la muestra de prostatectomía y antes de cualquier forma de tratamiento puede usarse para predecir no solo la progresión de la enfermedad, sino también la posible respuesta al tratamiento.

10 En la medida en que las realizaciones de la invención descritas anteriormente sean implementables, al menos en parte, utilizando un sistema informático, se apreciará que un programa informático para implementar al menos parte de los métodos descritos y/o los sistemas descritos se contempla como un aspecto de la presente invención. El sistema informático puede ser cualquier aparato, sistema o dispositivo adecuado. Por ejemplo, el sistema informático puede ser un aparato de procesamiento de datos programable, un ordenador de uso general, un procesador de señales digitales o un microprocesador. El programa informático puede incorporarse como código fuente y someterse a compilación para su implementación en un ordenador, o puede incorporarse como código de objetos, por ejemplo.

15 También es concebible que parte o la totalidad de la funcionalidad atribuida al programa informático o sistema informático mencionado anteriormente se pueda implementar en un equipo informático, por ejemplo, por medio de uno o más circuitos integrados específicos de la aplicación.

20 Adecuadamente, el programa informático puede almacenarse en un medio portador en una forma utilizable en un ordenador, que también se considera un aspecto de la presente invención. Por ejemplo, el medio portador puede ser una memoria de estado sólido, memoria óptica o magnetoóptica, como un disco legible y/o grabable, por ejemplo, un disco compacto (CD) o un disco versátil digital (DVD), o una memoria magnética como un disco o una cinta, y el sistema informático puede utilizar el programa para configurarlo para el funcionamiento. El programa informático también puede suministrarse desde una fuente remota incorporada en un medio portador, tal como una señal

electrónica, que incluye una onda portadora de radiofrecuencia o una onda portadora óptica.

Referencias

Las siguientes referencias mencionadas anteriormente se incorporan en la presente memoria como referencia en su totalidad:

- 5 [1] Scherr D., et al., *Urology*. 61 (2 Supl. 1): 14-24, feb. de 2003, Swindle P.W., et al., *Urologic Clinics of North America*. 30 (2): 377-401, mayo de 2003.
- [2] Wahlby C., et al., *Analytical Cellular Pathology* 24, 101-111, 2002.
- [3] Street W.N., "Xcyt: A System for Remote Cytological Diagnosis and Prognosis of Breast Cancer," en *Soft Computing Techniques in Breast Cancer Prognosis and Diagnosis*, L.C. Jain (ed.), CRC Press, 1999
- 10 [4] Gleason D.F., "The Veteran's Administration Cooperative Urologic Research Group: Histologic Grading and Clinical Staging of Prostatic Carcinoma," en *Urologic Pathology: The Prostate*, Tannenbaum M. (ed.), 171-198, Lea y Febiger, Filadelfia, 1977.
- [5] Cristianni et al, *An Introduction to Support Vector Machines*, Cambridge University Press (2000).
- [6] Hastie, *The Elements of Statistical Learning*, Springer (2001).
- 15 [7] F.E. Harrell et al., "Evaluating the yield of medical tests", *JAMA*, 247(18):2543-2546, 1982.
- [8] Bishop, C., *Neural Networks for Pattern Recognition*, Oxford University Press (1995).
- [9] Fausett, L., *Fundamentals of Neural Networks*, Nueva York, Prentice Hall (1994).
- [10] Definiens Cellenger Architecture: A Technical Review, abril de 2004.
- 20 [11] Baatz M. y Schape A., "Multiresolution Segmentation - An Optimization Approach for High Quality Multi-scale Image Segmentation," en *Angewandte Geographische Informationsverarbeitung XII*, Strobl, J., Blaschke, T., Griesebner, G. (eds.), Wichmann-Verlag, Heidelberg, 12-23, 2000.
- [12] Fukunaga K., *Introduction to Statistical Pattern Recognition*, 2ª Edición, Boston: Academic Press, 1990.
- [13] Duda R.O. et al., *Pattern Classification*, 2ª Edición, John Wiley & Sons Inc., 2001.
- 25 [14] Holmberg L. et al., "A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer," *N. Engl. M. Med.*, 347:781-789 (2002).
- [15] Pound CR et al., "Natural history of progression after PSA elevation following radical prostatectomy," *JAMA* 1999, 281:1591-1597.
- [16] Kumar-Sinha C. et al., "Molecular markers to identify patients at risk for recurrence after primary treatment for prostate cancer," *Urology* 2003; 62 Supl. 1:19-35.
- 30 [17] Cox D.R., "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, B 34, 187-220, 1972.
- [18] Harrell F.E., *Regression Modeling Strategies*, Springer-Verlag 2001.
- [19] Tuxhorn et al., "Reactive Stroma in Human Prostate Cancer: Induction of Myofibroblast Phenotype and Extracellular Matrix Remodeling" *Clinical Cancer Research* 2912 Vol. 8, 2912-2923, septiembre de 2002.
- 35 [20] Kattan et al., "Postoperative Nomogram for Disease Recurrence After Radical Prostatectomy for Prostate Cancer," *Journal of Clinical Oncology*, Vol. 17, N° 5 (Mayo), 1999: págs 1499-1507.

Tabla 1. Características Morfométricas, Secuencia de Comandos v1.0 (496 Características)

Característica
Background.MaxAreaPxl
Background.MeanAreaPxl
Background.MinAreaPxl
Background.StdDevAreaPxl
Background.SumAreaPxl
Cytoplasm.Objects
Cytoplasm.ObjectsPct
Cytoplasm.MaxAreaPxl
Cytoplasm.MeanAreaPxl
Cytoplasm.MinAreaPxl
Cytoplasm.StdDevAreaPxl
Cytoplasm.SumAreaPxl
Cytoplasm.MaxAsymmetry
Cytoplasm.MeanAsymmetry
Cytoplasm.MinAsymmetry
Cytoplasm.StdDevAsymmetry
Cytoplasm.MaxBorderlengthPxl
Cytoplasm.MeanBorderlengthPxl
Cytoplasm.MinBorderlengthPxl
Cytoplasm.StdDevBorderlengthPxl
Cytoplasm.SumBorderlengthPxl
Cytoplasm.MaxBrightness
Cytoplasm.MeanBrightness
Cytoplasm.MinBrightness
Cytoplasm.StdDevBrightness
Cytoplasm.MaxCompactness
Cytoplasm.MeanCompactness
Cytoplasm.MinCompactness
Cytoplasm.StdDevCompactness
Cytoplasm.MaxDensity
Cytoplasm.MeanDensity
Cytoplasm.MinDensity
Cytoplasm.StdDevDensity
Cytoplasm.MaxDiff.ofenclosing.encl
Cytoplasm.MeanDiff.ofenclosing.encl
Cytoplasm.MinDiff.ofenclosing.encl
Cytoplasm.StdDevDiff.ofenclosing.en
Cytoplasm.MaxEllipticFit
Cytoplasm.MeanEllipticFit
Cytoplasm.MinEllipticFit
Cytoplasm.StdDevEllipticFit
Cytoplasm.MaxLengthPxl
Cytoplasm.MeanLengthPxl
Cytoplasm.MinLengthPxl
Cytoplasm.StdDevLengthPxl
Cytoplasm.SumLengthPxl
Cytoplasm.MaxMax.Diff.
Cytoplasm.MeanMax.Diff.
Cytoplasm.MinMax.Diff.

Cytoplasm.StdDevMax.Diff.
Cytoplasm.MaxMeanChannel1
Cytoplasm.MeanMeanChannel1
Cytoplasm.MinMeanChannel1
Cytoplasm.StdDevMeanChannel1
Cytoplasm.MaxMeanChannel2
Cytoplasm.MeanMeanChannel2
Cytoplasm.MinMeanChannel2
Cytoplasm.StdDevMeanChannel2
Cytoplasm.MaxMeanChannel3
Cytoplasm.MeanMeanChannel3
Cytoplasm.MinMeanChannel3
Cytoplasm.StdDevMeanChannel3
Cytoplasm.MaxRadiusoflargestenclose
Cytoplasm.MeanRadiusoflargestenclos
Cytoplasm.MinRadiusoflargestenclose
Cytoplasm. StdDevRadiusoflargestenclo
Cytoplasm.MaxRadiusofsmallestenclos
Cytoplasm.MeanRadiusofsmallestenclo
Cytoplasm.MinRadiusofsmallestenclos
Cytoplasm.StdDevRadiusofsmallestenc
Cytoplasm.MaxStdevChaniel1
Cytoplasm.MeanStdevChannel1
Cytoplasm.MinStdevChannel1
Cytoplasm.StdDevStdevChannel1
Cytoplasm.MaxStdevChannel2
Cytoplasm.MeanStdevChannel2
Cytoplasm.MinStdevChannel2
Cytoplasm.StdDevStdevChannel2
Cytoplasm.MaxStdevChannel3
Cytoplasm.MeanStdevChannel3
Cytoplasm.MinStdevChannel3
Cytoplasm.StdDevStdevChannel3
Cytoplasm.MaxWidthPxl
Cytoplasm.MeanWidthPxl
Cytoplasm.MinWidthPxl
Cytoplasm.StdDevWidthPxl
Epithelial.Nuclei.Objects
Epithelial.Nuclei.ObjectsPct
Epithelial.Nuclei.MaxAreaPxl
Epithelial.Nuclei.MeanAreaPxl
Epithelial.Nuclei.MinAreaPxl
Epithelial.Nuclei.StdDevAreaPxl
Epithelial.Nuclei.SumAreaPxl
Epithelial.Nuclei.MaxAsymmetry
Epithelial.Nuclei.MeanAsymmetry
Epithelial.Nuclei.MinAsymmetry
Epithelial.Nuclei.StdDevAsymmetry
Epithelial.Nuclei.MaxBorderlengthPx
Epithelial.Nuclei.MeanBorderlengthP
Epithelial.Nuclei.MinBorderlengthPx
Epithelial.Nuclei.StdDevBorderlengt

Epithelial.Nuclei.SumBorderlengthPx
 Epithelial.Nuclei.MaxBrightness
 Epithelial.Nuclei.MeanBrightness
 Epithelial.Nuclei.MinBrightness
 Epithelial.Nuclei.StdDevBrightness
 Epithelial.Nuclei.MaxCompactness
 Epithelial.Nuclei.MeanCompactness
 Epithelial.Nuclei.MinCompactness
 Epithelial.Nuclei.StdDevCompactness
 Epithelial.Nuclei.MaxDensity
 Epithelial.Nuclei.MeanDensity
 Epithelial.Nuclei.MinDensity
 Epithelial.Nuclei.StdDevDensity
 Epithelial.Nuclei.MaxDiff.ofenclosi
 Epithelial.Nuclei.MeanDiff.ofenclosi
 Epithelial.Nuclei.MinDiff.ofenclosi
 Epithelial.Nuclei.StdDevDiff.ofenclosi
 Epithelial.Nuclei.MaxEllipticFit
 Epithelial.Nuclei.MeanEllipticFit
 Epithelial.Nuclei.MinEllipticFit
 Epithelial.Nuclei.StdDevEllipticFit
 Epithelial.Nuclei.MaxLengthPx
 Epithelial.Nuclei.MeanLengthPx
 Epithelial.Nuclei.MinLengthPx
 Epithelial.Nuclei.StdDevLengthPx
 Epithelial.Nuclei.SumLengthPx
 Epithelial.Nuclei.MaxMax.Diff.
 Epithelial.Nuclei.MeanMax.Diff.
 Epithelial.Nuclei.MinMax.Diff.
 Epithelial.Nuclei.StdDevMax.Diff.
 Epithelial.Nuclei.MaxMeanChannel1
 Epithelial.Nuclei.MeanMeanChannel1
 Epithelial.Nuclei.MinMeanChannel1
 Epithelial.Nuclei.StdDevMeanChannel1
 Epithelial.Nuclei.MaxMeanChannel2
 Epithelial.Nuclei.MeanMeanChannel2
 Epithelial.Nuclei.MinMeanChannel2
 Epithelial.Nuclei.StdDevMeanChannel2
 Epithelial.Nuclei.MaxMeanChannel3
 Epithelial.Nuclei.MeanMeanChannel3
 Epithelial.Nuclei.MinMeanChannel3
 Epithelial.Nuclei.StdDevMeanChannel3
 Epithelial.Nuclei.MaxRadiusoflarges
 Epithelial.Nuclei.MeanRadiusoflarge
 Epithelial.Nuclei.MinRadiusoflarges
 Epithelial.Nuclei.StdDevRadiusoflar
 Epithelial.Nuclei.MaxRadiusofsmalle
 Epithelial.Nuclei.MeanRadiusofsmall
 Epithelial.Nuclei.MinRadiusofsmalle
 Epithelial.Nuclei.StdDevRadiusofsma
 Epithelial.Nuclei.MaxStdevChannel1
 Epithelial.Nuclei.MeanStdevChannel1

Epithelial.Nuclei.MinStdevChannel1
 Epithelial.Nuclei.StdDevStdevChanne
 Epithelial.Nuclei.MaxStdevChannel2
 Epithelial.Nuclei.MeanStdevChannel2
 Epithelial.Nuclei.MinStdevChannel2
 Epithelial.Nuclei.StdDevStdevChanne
 Epithelial.Nuclei.MaxStdevChannel3
 Epithelial.Nuclei.MeanStdevChannel3
 Epithelial.Nuclei.MinStdevChannel3
 Epithelial.Nuclei.StdDevStdevChanne
 Epithelial.Nuclei.MaxWidthPxl
 Epithelial.Nuclei.MeanWidthPxl
 Epithelial.Nuclei.MinWidthPxl
 Epithelial.Nuclei.StdDevWidthPxl
 Lumen.Objects
 Lumen.ObjectsPct
 Lumen.MaxAreaPxl
 Lumen.MeanAreaPxl
 Lumen.MinAreaPxl
 Lumen.StdDevAreaPxl
 Lumen.SumAreaPxl
 Lumen.MaxAsymmetry
 Lumen.MeanAsymmetry
 Lumen.MinAsymmetry
 Lumen.StdDevAsymmetry
 Lumen.MaxBorderlengthPxl
 Lumen.MeanBorderlengthPxl
 Lumen.MinBorderlengthPxl
 Lumen.StdDevBorderlengthPxl
 Lumen.SumBorderlengthPxl
 Lumen.MaxBrightness
 Lumen.MeanBrightness
 Lumen.MinBrightness
 Lumen.StdDevBrightness
 Lumen.MaxCompactness
 Lumen.MeanCompactness
 Lumen.MinCompactness
 Lumen.StdDevCompactness
 Lumen.MaxDensity
 Lumen.MeanDensity
 Lumen.MinDensity
 Lumen.StdDevDensity
 Lumen.MaxDiff.ofenclosing.enclosede
 Lumen.MeanDiff.ofenclosing.enclosed
 Lumen.MinDiff.ofenclosing.enclosede
 Lumen.StdDevDiff.ofenclosing.enclos
 Lumen.MaxEllipticFit
 Lumen.MeanEllipticFit
 Lumen.MinEllipticFit
 Lumen.StdDevEllipticFit
 Lumen.MaxLengthPxl
 Lumen.MeanLengthPxl

Lumen.MinLengthPxl
 Lumen.StdDevLengthPxl
 Lumen.SumLengthPxl
 Lumen.MaxMax.Diff.
 Lumen.MeanMax.Diff.
 Lumen.MinMax.Diff.
 Lumen.StdDevMax.Diff.
 Lumen.MaxMeanChannel1
 Lumen.MeanMeanChannel1
 Lumen.MinMeanChannel1
 Lumen.StdDevMeanChannel1
 Lumen.MaxMeanChannel2
 Lumen.MeanMeanChannel2
 Lumen.MinMeanChannel2
 Lumen.StdDevMeanChannel2
 Lumen.MaxMeanChannel3
 Lumen.MeanMeanChannel3
 Lumen.MinMeanChannel3
 Lumen.StdDevMeanChannel3
 Lumen.MaxRadiusoflargestenclosedell
 Lumen.MeanRadiusoflargestenclosedell
 Lumen.MinRadiusoflargestenclosedell
 Lumen.StdDevRadiusoflargestenclosed
 Lumen.MaxRadiusofsmallestenclosing
 Lumen.MeanRadiusofsmallestenclosing
 Lumen.MinRadiusofsmallestenclosing
 Lumen.StdDevRadiusofsmallestenclosi
 Lumen.MaxStdevChannel1
 Lumen.MeanStdevChannel1
 Lumen.MinStdevChannel1
 Lumen.StdDevStdevChannel1
 Lumen.MaxStdevChannel2
 Lumen.MeanStdevChannel2
 Lumen.MinStdevChannel2
 Lumen.StdDevStdevChannel2
 Lumen.MaxStdevChannel3
 Lumen.MeanStdevChannel3
 Lumen.MinStdevChannel3
 Lumen.StdDevStdevChannel3
 Lumen.MaxWidthPxl
 Lumen.MeanWidthPxl
 Lumen.MinWidthPxl
 Lumen.StdDevWidthPxl
 Red.Blood.Cell.Objects
 Red.Blood.Cell.ObjectsPct
 Red.Blood.Cell.MaxAreaPxl
 Red.Blood.Cell.MeanAreaPxl
 Red.Blood.Cell.MinAreaPxl
 Red.Blood.Cell.StdDevAreaPxl
 Red.Blood.Cell.SumAreaPxl
 Red.Blood.Cell.MaxAsymmetry
 Red.Blood.Cell.MeanAsymmetry

Red.Blood.Cell.MinAsymmetry
Red.Blood.Cell.StdDevAsymmetry
Red.Blood.Cell.MaxBorderlengthPxI
Red.Blood.Cell.MeanBorderlengthPxI
Red.Blood.Cell.MinBorderlengthPxI
Red.Blood.Cell.StdDevBorderlengthPx
Red.Blood.Cell.SumBorderlengthPxI
Red.Blood.Cell.MaxBrightness
Red.Blood.Cell.MeanBrightness
Red.Blood.Cell.MinBrightness
Red.Blood.Cell.StdDevBrightness
Red.Blood.Cell.MaxCompactness
Red.Blood.Cell.MeanCompactness
Red.Blood.Cell.MinCompactness
Red.Blood.Cell.StdDevCompactness
Red.Blood.Cell.MaxDensity
Red.Blood.Cell.MeanDensity
Red.Blood.Cell.MinDensity
Red.Blood.Cell.StdDevDensity
Red.Blood.Cell.MaxDiff.ofenclosing.
Red.Blood.Cell.MeanDiff.ofenclosing
Red.Blood.Cell.MinDiff.ofenclosing.
Red.Blood.Cell.StdDevDiff.ofenclosi
Red.Blood.Cell.MaxEllipticFit
Red.Blood.Cell.MeanEllipticFit
Red.Blood.Cell.MinEllipticFit
Red.Blood.Cell.StdDevEllipticFit
Red.Blood.Cell.MaxLengthPxI
Red.Blood.Cell.MeanLengthPxI
Red.Blood.Cell.MinLengthPxI
Red.Blood.Cell.StdDevLengthPxI
Red.Blood.Cell.SumLengthPxI
Red.Blood.Cell.MaxMax.Diff.
Red.Blood.Cell.MeanMax.Diff.
Red.Blood.Cell.MinMax.Diff.
Red.Blood.Cell.StdDevMax.Diff.
Red.Blood.Cell.MaxMeanChannel1
Red.Blood.Cell.MeanMeanChannel1
Red.Blood.Cell.MinMeanChannel1
Red.Blood.Cell.StdDevMeanChannel1
Red.Blood.Cell.MaxMeanChannel2
Red.Blood.Cell.MeanMeanChannel2
Red.Blood.Cell.MinMeanChannel2
Red.Blood.Cell.StdDevMeanChannel2
Red.Blood.Cell.MaxMeanChannel3
Red.Blood.Cell.MeanMeanChannel3
Red.Blood.Cell.MinMeanChannel3
Red.Blood.Cell. StdDevMeanChannel3
Red.Blood.Cell.MaxRadiusoflargesten
Red.Blood.Cell.MeanRadiusoflargeste
Red.Blood.Cell.MinRadiusoflargesten
Red.Blood.Cell.StdDevRadiusoflarges

Red.Blood.Cell.MaxRadiusofsmalleste
 Red.Blood.Cell.MeanRadiusofsmallest
 Red.Blood.Cell.MinRadiusofsmalleste
 Red.Blood.Cell.StdDevRadiusofsmalle
 Red.Blood.Cell.MaxStdevChannel1
 Red.Blood.Cell.MeanStdevChannel1
 Red.Blood.Cell.MinStdevChannel1
 Red.Blood.Cell.StdDevStdevChannel1
 Red.Blood.Cell.MaxStdevChannel2
 Red.Blood.Cell.MeanStdevChannel2
 Red.Blood.Cell.MinStdevChannel2
 Red.Blood.Cell.StdDevStdevChannel2
 Red.Blood.Cell.MaxStdevChannel3
 Red.Blood.Cell.MeanStdevChannel3
 Red.Blood.Cell.MinStdevChannel3
 Red.Blood.Cell.StdDevStdevChannel3
 Red.Blood.Cell.MaxWidthPxl
 Red.Blood.Cell.MeanWidthPxl
 Red.Blood.Cell.MinWidthPxl
 Red.Blood.Cell.StdDevWidthPxl
 Stroma.Objects
 Stroma.ObjectsPct
 Stroma.MaxAreaPxl
 Stroma.MeanAreaPxl
 Stroma.MinAreaPxl
 Stroma.StdDevAreaPxl
 Stroma.SumAreaPxl
 Stroma.MaxAsymmetry
 Stroma.MeanAsymmetry
 Stroma.MinAsymmetry
 Stroma.StdDevAsymmetry
 Stroma.MaxBorderlengthPxl
 Stroma.MeanBorderlengthPxl
 Stroma.MinBorderlengthPxl
 Stroma.StdDevBorderlengthPxl
 Stroma.SumBorderlengthPxl
 Stroma.MaxBrightness
 Stroma.MeanBrightness
 Stroma.MinBrightness
 Stroma.StdDevBrightness
 Stroma.MaxCompactness
 Stroma.MeanCompactness
 Stroma.MinCompactness
 Stroma.StdDevCompactness
 Stroma.MaxDensity
 Stroma.MeanDensity
 Stroma.MinDensity
 Stroma.StdDevDensity
 Stroma.MaxDiff.ofenclosing.encoded
 Stroma.MeanDiff.ofenclosing.encode
 Stroma.MinDiff.ofenclosing.encoded
 Stroma.StdDevDiff.ofenclosing.encl

Stroma.MaxEllipticFit
 Stroma.MeanEllipticFit
 Stroma.MinEllipticFit
 Stroma.StdDevEllipticFit
 Stroma.MaxLengthPxl
 Stroma.MeanLengthPxl
 Stroma.MinLengthPxl
 Stroma.StdDevLengthPxl
 Stroma.SumLengthPxl
 Stroma.MaxMax.Diff.
 Stroma.MeanMax.Diff.
 Stroma.MinMax.Diff.
 Stroma.StdDevMax.Diff.
 Stroma.MaxMeanChannel1
 Stroma.MeanMeanChannel1
 Stroma.MinMeanChannel1
 Stroma.StdDevMeanChannel1
 Stroma.MaxMeanChannel2
 Stroma.MeanMeanChannel2
 Stroma.MinMeanChannel2
 Stroma.StdDevMeanChannel2
 Stroma.MaxMeanChannel3
 Stroma.MeanMeanChannel3
 Stroma.MinMeanChannel3
 Stroma.StdDevMeanChannel3
 Stroma.MaxRadiusoflargestenclosedel
 Stroma.MeanRadiusoflargestenclosede
 Stroma.MinRadiusoflargestenclosedel
 Stroma.StdDevRadiusoflargestenclose
 Stroma.MaxRadiusofsmallestenclosing
 Stroma.MeanRadiusofsmallestenclosin
 Stroma.MinRadiusofsmallestenclosing
 Stroma.StdDevRadiusofsmallestenclos
 Stroma.MaxStdevChannel1
 Stroma.MeanStdevChannel1
 Stroma.MinStdevChannel1
 Stroma.StdDevStdevChannel1
 Stroma.MaxStdevChannel2
 Stroma.MeanStdevChannel2
 Stroma.MinStdevChannel2
 Stroma.StdDevStdevChannel2
 Stroma.MaxStdevChannel3
 Stroma.MeanStdevChannel3
 Stroma.MinStdevChannel3
 Stroma.StdDevStdevChannel3
 Stroma.MaxWidthPxl
 Stroma.MeanWidthPxl
 Stroma.MinWidthPxl
 Stroma.StdDevWidthPxl
 Stroma.Nuclei.Objects
 Stroma.Nuclei.ObjectsPct
 Stroma.Nuclei.MaxAreaPxl

Stroma.Nuclei.MeanAreaPxl
 Stroma.Nuclei.MinAreaPxl
 Stroma.Nuclei.StdDevAreaPxl
 Stroma.Nuclei.SumAreaPxl
 Stroma.Nuclei.MaxAsymmetry
 Stroma.Nuclei.MeanAsymmetry
 Stroma.Nuclei.MinAsymmetry
 Stroma.Nuclei.StdDevAsymmetry
 Stroma.Nuclei.MaxBorderlengthPxl
 Stroma.Nuclei.MeanBorderlengthPxl
 Stroma.Nuclei.MinBorderlengthPxl
 Stroma.Nuclei.StdDevBorderlengthPxl
 Stroma.Nuclei.SumBorderlengthPxl
 Stroma.Nuclei.MaxBrightness
 Stroma.Nuclei.MeanBrightness
 Stroma.Nuclei.MinBrightness
 Stroma.Nuclei.StdDevBrightness
 Stroma.Nuclei.MaxCompactness
 Stroma.Nuclei.MeanCompactness
 Stroma.Nuclei.MinCompactness
 Stroma.Nuclei.StdDevCompactness
 Stroma.Nuclei.MaxDensity
 Stroma.Nuclei.MeanDensity
 Stroma.Nuclei.MinDensity
 Stroma.Nuclei.StdDevDensity
 Stroma.Nuclei.MaxDiff.ofenclosing.e
 Stroma.Nuclei.MeanDiff.ofenclosing.e
 Stroma.Nuclei.MinDiff.ofenclosing.e
 Stroma.Nuclei.StdDevDiff.ofenclosin
 Stroma.Nuclei.MaxEllipticFit
 Stroma.Nuclei.MeanEllipticFit
 Stroma.Nuclei.MinEllipticFit
 Stroma.Nuclei.StdDevEllipticFit
 Stroma.Nuclei.MaxLengthPxl
 Stroma.Nuclei.MeanLengthPxl
 Stroma.Nuclei.MinLengthPxl
 Stroma.Nuclei.StdDevLengthPxl
 Stroma.Nuclei.SumLengthPxl
 Stroma.Nuclei.MaxMax.Diff.
 Stroma.Nuclei.MeanMax.Diff.
 Stroma.Nuclei.MinMax.Diff.
 Stroma.Nuclei.StdDevMax.Diff.
 Stroma.Nuclei.MaxMeanChannel1
 Stroma.Nuclei.MeanMeanChannel1
 Stroma.Nuclei.MinMeanChannel1
 Stroma.Nuclei.StdDevMeanChannel1
 Stroma.Nuclei.MaxMeanChannel2
 Stroma.Nuclei.MeanMeanChannel2
 Stroma.Nuclei.MinMeanChannel2
 Stroma.Nuclei.StdDevMeanChannel2
 Stroma.Nuclei.MaxMeanChannel3
 Stroma.Nuclei.MeanMeanChannel3

Stroma.Nuclei.MinMeanChannel3
 Stroma.Nuclei.StdDevMeanChannel3
 Stroma.Nuclei.MaxRadiusoflargestenc
 Stroma.Nuclei.MeanRadiusoflargestenc
 Stroma.Nuclei.MinRadiusoflargestenc
 Stroma.Nuclei.StdDevRadiusoflargest
 Stroma.Nuclei.MaxRadiusofsmallestenc
 Stroma.Nuclei.MeanRadiusofsmallestenc
 Stroma.Nuclei.MinRadiusofsmallestenc
 Stroma.Nuclei.StdDevRadiusofsmalles
 Stroma.Nuclei.MaxStdevChannel1
 Stroma.Nuclei.MeanStdevChannel1
 Stroma.Nuclei.MinStdevChannel1
 Stroma.Nuclei.StdDevStdevChannel1
 Stroma.Nuclei.MaxStdevChannel2
 Stroma.Nuclei.MeanStdevChannel2
 Stroma.Nuclei.MinStdevChannel2
 Stroma.Nuclei.StdDevStdevChannel2
 Stroma.Nuclei.MaxStdevChannel3
 Stroma.Nuclei.MeanStdevChannel3
 Stroma.Nuclei.MinStdevChannel3
 Stroma.Nuclei.StdDevStdevChannel3
 Stroma.Nuclei.MaxWidthPxl
 Stroma.Nuclei.MeanWidthPxl
 Stroma.Nuclei.MinWidthPxl
 Stroma.Nuclei.StdDevWidthPxl
 C2EN
 EN2SN
 L2Core
 C2L
 CEN2L

Tabla 2. Características Morfométricas, Secuencia de Comandos v2.0 (350 características)

Característica
Área media de artefacto en pxl
StdDev de area de artefacto en pxl
Asimetría media de artefacto
StdDev de asimetría de artefacto
Índice medio de borde de artefacto
StdDev de índice de borde de artefacto
Longitud media de borde de artefacto en pxl
StdDev de longitud de borde de artefacto en pxl
Brillo medio de artefacto
StdDev de brillo de artefacto
Compacidad media de artefacto
StdDev de compacidad de artefacto
Densidad media de artefacto
StdDev de densidad de artefacto
Dif. media de artefacto de elipse de cierre/cerrada
StdDev de dif. de artefacto de elipse de cierre/cerrada
Ajuste elíptico medio de artefacto

StdDev de ajuste elíptico de artefacto
 Longitud media de artefacto en pxl
 StdDev de longitud de artefacto en pxl
 Longitud/anchura media de artefacto
 StdDev de longitud/anchura de artefacto
 Dirección principal media de artefacto
 StdDev de dirección principal de artefacto
 Dif. máx. media de artefacto
 StdDev de dif. máx. de artefacto
 Media de media de artefacto del canal 1
 StdDev de media de artefacto del canal 1
 Media de media de artefacto del canal 2
 StdDev de media de artefacto del canal 2
 Media de media de artefacto del canal 3
 StdDev de media de artefacto del canal 3
 Radio medio de artefacto de la elipse cerrada más grande
 StdDev de radio de artefacto de la elipse cerrada más grande
 Radio medio de artefacto de la elipse cerrada más pequeña
 StdDev de radio de artefacto de la elipse cerrada más pequeña
 Ajuste rectangular medio de artefacto
 StdDev de ajuste rectangular de artefacto
 Índice de forma media de artefacto
 StdDev de índice de forma de artefacto
 Stddev de media de artefacto del canal 1
 Stddev de StdDev de artefacto del canal 1
 Stddev de media de artefacto del canal 2
 Stddev de StdDev de artefacto del canal 2
 Stddev de media de artefacto del canal 3
 Stddev de StdDev de artefacto del canal 3
 Ancho medio de artefacto en pxl
 StdDev de ancho de artefacto en pxl
 Área media de citoplasma en pxl
 StdDev de área de citoplasma en pxl
 Asimetría media de citoplasma
 StdDev de asimetría de citoplasma
 Índice de borde medio de citoplasma
 StdDev de índice de borde de citoplasma
 Longitud media de borde de citoplasma en pxl
 StdDev de longitud de borde de citoplasma en pxl
 Brillo medio de citoplasma
 StdDev de brillo de citoplasma
 Compacidad media de citoplasma
 StdDev de Compacidad de citoplasma
 Densidad media de citoplasma
 StdDev de densidad de citoplasma
 Dif. media de citoplasma de elipse de cerramiento/cerrada
 StdDev de dif. de citoplasma de elipse de cerramiento/cerrada
 Ajuste elíptico medio de citoplasma
 StdDev de ajuste elíptico de citoplasma
 Longitud media de citoplasma en pxl
 StdDev de longitud de citoplasma en pxl
 Longitud/anchura media de citoplasma

StdDev de longitud/anchura de citoplasma
 Dirección principal media de citoplasma
 StdDev de dirección principal de citoplasma
 Diferencia máx. media de citoplasma
 StdDev de diferencia máx. de citoplasma
 Media de media de citoplasma del canal 1
 StdDev de media de citoplasma del canal 1
 Media de media de citoplasma del canal 2
 StdDev de media de citoplasma del canal 2
 Media de media de citoplasma del canal 3
 StdDev de media de citoplasma del canal 3
 Radio medio de citoplasma de la elipse cerrada más grande
 StdDev de radio de citoplasma de la elipse cerrada más grande
 Radio medio de citoplasma de la elipse cerrada más pequeña
 StdDev de radio de citoplasma de la elipse cerrada más pequeña
 Ajuste rectangular medio de citoplasma
 StdDev de ajuste rectangular de citoplasma
 Índice de forma media de citoplasma
 StdDev de índice de forma de citoplasma
 Stddev de media de citoplasma del canal 1
 Stddev de StdDev de citoplasma del canal 1
 Stddev de media de citoplasma del canal 2
 Stddev de StdDev de citoplasma del canal 2
 Stddev de media de citoplasma del canal 3
 Stddev de StdDev de citoplasma del canal 3
 Ancho medio de citoplasma en pxl
 StdDev de ancho de citoplasma en pxl
 Área media de núcleos epiteliales en pxl
 StdDev de área de núcleos epiteliales en pxl
 Asimetría media de núcleos epiteliales
 StdDev de asimetría de núcleos epiteliales
 Índice de borde medio de núcleos epiteliales
 StdDev de índice de borde de núcleos epiteliales
 Longitud media de borde de núcleos epiteliales en pxl
 StdDev de longitud de borde de núcleos epiteliales en pxl
 Brillo medio de núcleos epiteliales
 StdDev de brillo de núcleos epiteliales
 Compacidad media de núcleos epiteliales
 StdDev de compacidad de núcleos epiteliales
 Densidad media de núcleos epiteliales
 StdDev de densidad de núcleos epiteliales
 Dif. media de núcleos epiteliales de elipse de cerramiento/cerrada
 StdDev de dif. de núcleos epiteliales de elipse de cerramiento/cerrada
 Ajuste elíptico medio de núcleos epiteliales
 StdDev de ajuste elíptico de núcleos epiteliales
 Longitud media de núcleos epiteliales en pxl
 StdDev de longitud de núcleos epiteliales en pxl
 Longitud/anchura media de núcleos epiteliales
 StdDev de longitud/anchura de núcleos epiteliales
 Dirección principal media de núcleos epiteliales
 StdDev de dirección principal de núcleos epiteliales
 Dif. máx. media de núcleos epiteliales

StdDev de dif. máx. de núcleos epiteliales
 Media de media de núcleos epiteliales del canal 1
 StdDev de media de núcleos epiteliales del canal 1
 Media de media de núcleos epiteliales del canal 2
 StdDev de media de núcleos epiteliales del canal 2
 Media de media de núcleos epiteliales del canal 3
 StdDev de media de núcleos epiteliales del canal 3
 Radio medio de núcleos epiteliales de la elipse cerrada más grande
 StdDev de radio de núcleos epiteliales de la elipse cerrada más
 Radió medio de núcleos epiteliales de la elipse cerrada más pequeña
 StdDev de radio de núcleos epiteliales de la elipse cerrada más
 Ajuste rectangular medio de nucleos epiteliales
 StdDev de ajuste rectangular de nucleos epiteliales
 Índice de forma media de núcleos epiteliales
 StdDev de índice de forma de núcleos epiteliales
 Stddev de media de núcleos epiteliales del canal 1
 Stddev de StdDev de núcleos epiteliales del canal 1
 Stddev de media de núcleos epiteliales del canal 2
 Stddev de StdDev de núcleos epiteliales del canal 2
 Stddev de media de núcleos epiteliales del canal 3
 Stddev de StdDev de núcleos epiteliales del canal 3
 Ancho medio de núcleos epiteliales en pxl
 StdDev de ancho de núcleos epiteliales en pxl
 Área media de luz en pxl
 StdDev de área de luz en pxl
 Asimetría media de luz
 StdDev de asimetría de luz
 Índice de borde medio de luz
 StdDev de índice de borde de luz
 Longitud media de borde de luz en pxl
 StdDev de longitud de borde de luz en pxl
 Brillo medio de luz
 StdDev de brillo de luz
 Compacidad media de luz
 StdDev de compacidad de luz
 Densidad media de luz
 StdDev de densidad de luz
 Dif. media de luz de elipse de cerramiento/cerrada
 StdDev de dif. de luz de elipse de cerramiento/cerrada
 Ajuste elíptico medio de luz
 StdDev de ajuste elíptico de luz
 Longitud media de luz en pxl
 StdDev de longitud de luz en pxl
 Longitud/ancho medio de luz
 StdDev de longitud/ancho de luz
 Dirección principal media de luz
 StdDev de dirección principal de luz
 Diferencia máx. media de luz
 StdDev de diferencia máx. de luz
 Media de media de luz del canal 1
 StdDev de media de luz del canal 1
 Media de media de luz del canal 2

StdDev de media de luz del canal 2
 Media de media de luz del canal 3
 StdDev de media de luz del canal 3
 Radio medio de luz de la elipse cerrada más grande
 StdDev de radio de luz de la elipse cerrada más grande
 Radio medio de luz de la elipse cerrada más pequeña
 StdDev de radio de luz de la elipse cerrada más pequeña
 Ajuste Rectangular Medio de luz
 StdDev de ajuste rectangular de luz
 Índice de forma media de luz
 StdDev de índice de forma de luz
 Stddev de media de luz del canal 1
 Stddev de StdDev de luz del canal 1
 Stddev de media de luz del canal 2
 Stddev de StdDev de luz del canal 2
 Stddev de media de luz del canal 3
 Stddev de StdDev de luz del canal 3
 Ancho Medio de luz en pxl
 Luz de ancho medio de luz en pxl
 Área media de estroma en pxl
 StdDev de área de estroma en pxl
 Asimetría media de estroma
 StdDev de asimetría de estroma
 Índice de borde medio de estroma
 StdDev de índice de borde de estroma
 Longitud media de borde de estroma en pxl
 StdDev de longitud de borde de estroma en pxl
 Brillo medio de estroma
 StdDev de brillo de estroma
 Compacidad media de estroma
 StdDev de compacidad de estroma
 Densidad media de estroma
 StdDev de densidad de estroma
 Dif. media de estroma de elipse de cerramiento/cerrada
 StdDev de dif. de estroma de elipse de cerramiento/cerrada
 Ajuste elíptico medio de estroma
 StdDev de ajuste elíptico de estroma
 Longitud media de estroma en pxl
 StdDev de longitud de estroma en pxl
 Longitud/ancho medio de estroma
 StdDev de longitud/ancho de estroma
 Dirección principal media de estroma
 StdDev de dirección principal de estroma
 Dif. máx. media de estroma
 StdDev de dif. máx. de estroma
 Media de media de estroma del canal 1
 Media de StdDev de estroma del canal 1
 Media de media de estroma del canal 2
 Media de StdDev de estroma del canal 2
 Media de media de estroma del canal 3
 Media de StdDev de estroma del canal 3
 Radio medio de estroma de la elipse cerrada más grande

StdDev de Radio de estroma de la elipse cerrada más grande
 Radio medio de estroma de la elipse cerrada más pequeña
 StdDev de Radio de estroma de la elipse cerrada más pequeña
 Ajuste rectangular medio de estroma
 StdDev de ajuste rectangular de estroma
 Índice de forma media de estroma
 StdDev de índice de forma de estroma
 Stddev de media de estroma del canal 1
 Stddev de Stddev de estroma del canal 1
 Stddev de media de estroma del canal 2
 Stddev de Stddev de estroma del canal 2
 Stddev de media de estroma del canal 3
 Stddev de Stddev de estroma del canal 3
 Ancho medio de estroma en pxl
 StdDev de ancho de estroma en pxl
 Área media de núcleos de estroma en pxl
 StdDev de área de núcleos de estroma en pxl
 Asimetría media de núcleos de estroma
 StdDev de asimetría de núcleos de estroma
 Índice de borde medio de núcleos de estroma
 StdDev de índice de borde de núcleos de estroma
 Longitud de borde medio de núcleos de estroma en pxl
 StdDev de longitud de borde de núcleos de estroma en pxl
 Brillo medio de núcleos de estroma
 StdDev de brillo de núcleos de estroma
 Compacidad media de núcleos de estroma
 StdDev de compacidad de núcleos de estroma
 Densidad media de núcleos de estroma
 StdDev de densidad de núcleos de estroma
 Dif. media de núcleos de estroma de elipse de cerramiento/cerrada
 StdDev de dif. de núcleos de estroma de elipse de
 Ajuste elíptico medio de núcleos de estroma
 StdDev de ajuste elíptico de núcleos de estroma
 Longitud media de núcleos de estroma en pxl
 StdDev de longitud de núcleos de estroma en pxl
 Longitud/ancho medio de núcleos de estroma
 StdDev de longitud/ancho de núcleos de estroma
 Dirección principal media de núcleos de estroma
 StdDev de dirección principal de núcleos de estroma
 Dif. máx. media de núcleos de estroma
 StdDev de dif. máx. de núcleos de estroma
 Media de media de núcleos de estroma del canal 1
 StdDev de media de núcleos de estroma del canal 1
 Media de media de núcleos de estroma del canal 2
 StdDev de media de núcleos de estroma del canal 2
 Media de media de núcleos de estroma del canal 3
 StdDev de media de núcleos de estroma del canal 3
 Radio medio de núcleos de estroma de la elipse cerrada más grande
 StdDev de Radio de núcleos de estroma de la elipse cerrada más
 Radio medio de núcleos de estroma de la elipse cerrada más pequeña
 StdDev StdDev de Radio de núcleos de estroma de la elipse cerrada
 Ajuste rectangular medio de núcleos de estroma

StdDev de ajuste rectangular de núcleos de estroma
Índice de forma media de núcleos de estroma
StdDev de índice de forma de núcleos de estroma
Stddev de media de núcleos de estroma del canal 1
Stddev de StdDev de núcleos de estroma del canal 1
Stddev de media de núcleos de estroma del canal 2
Stddev de StdDev de núcleos de estroma del canal 2
Stddev de media de núcleos de estroma del canal 3
Stddev de StdDev de núcleos de estroma del canal 3
Ancho Medio de núcleos de estroma en pxl
StdDev de ancho de núcleos de estroma en pxl
Área de artefactos en pxl
Área de citoplasma en pxl
Área de núcleos epiteliales en pxl
Área de luz en pxl
Área de glóbulos rojos en pxl
Área de estroma en pxl
Área de núcleos de estroma en pxl
Número de objetos de artefacto
Número de objetos de citoplasma
Número de objetos de núcleos epiteliales
Numero de objetos de luz
Número de objetos de glóbulos rojos
Número de objetos de estroma
Número de objetos de núcleos de estroma
Área media de glóbulos rojos en pxl
StdDev de área de glóbulos rojos en pxl
Asimetría media de glóbulos rojos
StdDev de asimetría de glóbulos rojos
Índice de borde medio de glóbulos rojos
StdDev de índice de borde de glóbulos rojos
Longitud media de borde de glóbulos rojos en pxl
StdDev de longitud de borde de glóbulos rojos en pxl
Brillo medio de glóbulos rojos
StdDev de brillo de glóbulos rojos
Compacidad media de glóbulos rojos
StdDev de compacidad de glóbulos rojos
Densidad media de glóbulos rojos
StdDev de densidad de glóbulos rojos
Dif. media de glóbulos rojos de elipse de cerramiento/cerrada
StdDev de dif. media de glóbulos rojos de elipse de
Ajuste elíptico medio de glóbulos rojos
StdDev de ajuste elíptico de glóbulos rojos
Longitud media de glóbulos rojos en pxl
StdDev de longitud de glóbulos rojos en pxl
Longitud/ancho medio de glóbulos rojos
StdDev de longitud/ancho de glóbulos rojos
Dirección principal media de glóbulos rojos
StdDev de dirección principal de glóbulos rojos
Dif. máx. media de glóbulos rojos
StdDev de dif. máx. de glóbulos rojos
Media de media de glóbulos rojos del canal 1

StdDev de media de glóbulos rojos del canal 1
 Media de media de glóbulos rojos del canal 2
 StdDev de media de glóbulos rojos del canal 2
 Media de media de glóbulos rojos del canal 3
 StdDev de media de glóbulos rojos del canal 3
 Radio medio de glóbulos rojos de la elipse cerrada más grande
 StdDev de Radio de glóbulos rojos de la elipse cerrada más grande
 Radio medio de glóbulos rojos de la elipse cerrada más pequeña
 StdDev de Radio de glóbulos rojos de la elipse cerrada más pequeña
 Ajuste rectangular medio de glóbulos rojos
 StdDev de ajuste rectangular de glóbulos rojos
 Índice de forma media de glóbulos rojos
 StdDev de índice de forma de glóbulos rojos
 Stddev de media de glóbulos rojos del canal 1
 Stddev de StdDev de glóbulos rojos del canal 1
 Stddev de media de glóbulos rojos del canal 2
 Stddev de StdDev de glóbulos rojos del canal 2
 Stddev de media de glóbulos rojos del canal 3
 Stddev de StdDev de glóbulos rojos del canal 3
 Ancho medio de glóbulos rojos en pxl
 StdDev de ancho de glóbulos rojos en pxl

Tabla 10. Características morfométricas, Secuencia de Comandos v5.0 (38 características)

Nombre de la Característica	Descripción
CytoplasmMeanMeanChannel40058	Valor medio de intensidad de citoplasma en el canal de color rojo
CytoplasmMeanMeanChannel50059	Valor medio de intensidad de citoplasma en el canal de color verde
CytoplasmMeanMeanChannel60060	Valor medio de intensidad de citoplasma en el canal de color azul
CytoplasmMeanStddevChannel40066	Desviación estándar de la intensidad media de citoplasma en el canal de color rojo
CytoplasmMeanStddevChannel50067	Desviación estándar de la intensidad media de citoplasma en el canal de color verde
CytoplasmMeanStddevChannel60068	Desviación estándar de la intensidad media de citoplasma en el canal de color azul
CytoplasmStddevMeanChannel40081	Desviación estándar de la intensidad de citoplasma en el canal de color rojo.
CytoplasmStddevMeanChannel50082	Desviación estándar de la intensidad de citoplasma en el canal de color verde.
CytoplasmStddevMeanChannel60083	Desviación estándar de la intensidad de citoplasma en el canal de color azul.
EpithelialNucleiMeanAreaPxI0101	Valor medio del área del núcleo epitelial, píxeles
EpithelNucleiMeanMeanChanne40112	Intensidad media del núcleo epitelial en el canal de color rojo
EpithelNucleiMeanMeanChanne50113	Intensidad media del núcleo epitelial en el canal de color verde
EpithelNucleiMeanMeanChanne60114	Intensidad media del núcleo epitelial en el canal de color azul
EpitheNucleiMeanStddevChann40120	Desviación estándar de la intensidad media del núcleo epitelial en el canal de color rojo
EpitheNucleiMeanStddevChann50121	Desviación estándar de la intensidad media del núcleo epitelial en el canal de color verde
EpitheNucleiMeanStddevChann60122	Desviación estándar de la intensidad media del núcleo epitelial en el canal de color azul
EpitheliaNucleiStddevAreaPxI0124	Desviación estándar del área del núcleo epitelial
EpitheNucleiStddevMeanChann40135	Desviación estándar de la intensidad del núcleo epitelial en el canal de color rojo

EpitheNucleiStddevMeanChann50136	Desviación estándar de la intensidad del núcleo epitelial en el canal de color verde.
EpitheNucleiStddevMeanChann60137	Desviación estándar de la intensidad del núcleo epitelial en el canal de color azul
StromaMeanMeanChannel40262	Intensidad media del estroma en el canal de color rojo
StromaMeanMeanChannel50263	Intensidad media del estroma en el canal de color verde
StromaMeanMeanChannel60264	Intensidad media del estroma en el canal de color azul
StromaMeanStddevChannel40270	Desviación estándar de la intensidad media del estroma en el canal de color rojo
StromaMeanStddevChannel50271	Desviación estándar de la intensidad media del estroma en el canal de color verde
StromaMeanStddevChannel60272	Desviación estándar de la intensidad media del estroma en el canal de color azul
StromaStddevMeanChannel40331	Desviación estándar de la intensidad media del estroma en el canal de color rojo
StromaStddevMeanChannel50332	Desviación estándar de la intensidad del estroma en el canal de color verde
StromaStddevMeanChannel60333	Desviación estándar de la intensidad del estroma en el canal de color azul
AreaOfCytoplasmPxI0345	Área total de citoplasma, píxeles
AreaOfEpithelialNucleiPxI0350	Área total de núcleos epiteliales, píxeles
AreaOfLumenPxI0357	Área total de luces, píxeles
NumberOfObjectOfEpitheNuclei0364	Número total de objetos de núcleos epiteliales.
AreaCytodivTotTissueArea	Área relativa de citoplasma con respecto al área de tejido, %
AreaEpitNucdivTotTissueArea	Área relativa de núcleos epiteliales con respecto al área de tejido, %
AreaLumendivTotTissueArea	Área relativa de lúmenes con respecto al área de tejido, %
AreaStromadivTotTissueArea	Área relativa del estroma con respecto al área de tejido, %
NumObjEpitNucdivTotNumberNuc	Número relativo de núcleos epiteliales con respecto al número total de núcleos.

Tabla 11. Características morfométricas

Nombre de la Característica	Descripción
CytoplasmMeanMeanChannel40058	Media del valor medio de intensidad de citoplasma con el filtro rojo
CytoplasmMeanMeanChannel50059	Media del valor medio de intensidad de citoplasma con el filtro verde
CytoplasmMeanMeanChannel60060	Media del valor medio de intensidad de citoplasma con el filtro azul
CytoplasmMeanStddevChannel40066	Media de la desviación estándar de la intensidad de citoplasma con el filtro rojo
CytoplasmMeanStddevChannel50067	Media de la desviación estándar de la intensidad de citoplasma con el filtro verde
CytoplasmMeanStddevChannel60068	Media de la desviación estándar de la intensidad de citoplasma con el filtro azul
CytoplasmStddevMeanChannel40081	Desviación estándar de la intensidad media de citoplasma con el filtro rojo
CytoplasmStddevMeanChannel50082	Desviación estándar de la intensidad media de citoplasma con el filtro verde
CytoplasmStddevMeanChannel60083	Desviación estándar de la intensidad media de citoplasma con el filtro azul
EpitheNucleiMeanMeanChanne40112	Media de la intensidad de núcleo epitelial con el filtro rojo
EpitheNucleiMeanMeanChanne50113	Media de la intensidad de núcleo epitelial con el filtro verde
EpitheNucleiMeanMeanChanne60114	Media de la intensidad de núcleo epitelial con el filtro azul
EpitheNucleiMeanStddevChann40120	Media de la desviación estándar de la intensidad de núcleo epitelial con el filtro rojo
EpitheNucleiMeanStddevChann50121	Media de la desviación estándar de la intensidad de núcleo epitelial con el filtro verde

ES 2 745 811 T3

EpitheNucleiMeanStddevChann60122	Media de la desviación estándar de la intensidad de núcleo epitelial con el filtro azul
EpitheliaNucleiStddevAreaPx10124	Desviación estándar del área del núcleo epitelial
EpitheNucleiStddevMeanChann40135	Desviación estándar de la intensidad media de núcleo epitelial con el filtro rojo
EpitheNucleiStddevMeanChann50136	Desviación estándar de la intensidad media de núcleo epitelial con el filtro verde
EpitheNucleiStddevMeanChann60137	Desviación estándar de la intensidad media de núcleo epitelial con el filtro azul
StromaMeanMeanChannel40262	Media de la intensidad de estroma con el filtro rojo
StromaMeanMeanChannel50263	Media de la intensidad de estroma con el filtro verde
StromaMeanMeanChannel60264	Media de la intensidad de estroma con el filtro azul
StromaMeanStddevChannel40270	Media de la desviación estándar de intensidad de estroma con el filtro rojo
StromaMeanStddevChannel50271	Media de la desviación estándar de la intensidad de estroma con el filtro verde
StromaMeanStddevChannel60272	Media de la desviación estándar de la intensidad de estroma con el filtro azul
StromaStddevMeanChannel40331	Desviación estándar de la intensidad media de estroma con el filtro rojo
StromaStddevMeanChannel50332	Desviación estándar de la intensidad media de estroma con el filtro verde
StromaStddevMeanChannel60333	Desviación estándar de la intensidad de estroma con el filtro azul.
AreaCytopdivTotTissueArea	Área de citoplasma respecto de área de tejido, %
AreaEpitNucdivTotTissueArea	Área de núcleos epiteliales respecto de área de tejido, %
AreaLumendivTotTissueArea	Área de luz respecto de área de tejido, %
AreaRBCdivTotTissueArea	Área de glóbulos rojos respecto de área de tejido, %
AreaStromadivTotTissueArea	Área de estroma respecto de área de tejido, %

REIVINDICACIONES

1. Un aparato para evaluar un riesgo de fallo clínico en un paciente después de que el paciente se haya sometido a una prostatectomía radical, y el aparato comprende:
- 5 un modelo predictivo de fallo clínico configurado para evaluar un conjunto de datos para un paciente para así evaluar un riesgo de fallo clínico para el paciente, en el que el modelo se basa en:
- una puntuación de Gleason de biopsia;
 - la afectación de ganglios linfáticos;
 - puntuación de Gleason de una muestra (prostatectomía);
 - 10 intensidad del receptor de andrógenos (AR) en las células epiteliales positivas para racemasa (AMACR); y
 - tres características morfométricas derivadas de una imagen de tejido, y las tres características morfométricas comprenden:
 - 15 una primera medición morfométrica indicativa de la intensidad media del citoplasma epitelial como se expresa en el canal azul de la imagen de tejido,
 - una segunda medición morfométrica indicativa de la variación en la textura del estroma dentro del estroma como se expresa en un canal rojo de la imagen de tejido, y
 - una tercera medición morfométrica indicativa de la variación en la textura entre núcleos epiteliales como se expresa en el canal rojo de la imagen de tejido.
2. El aparato de la reivindicación 1, en donde la intensidad del receptor de andrógenos (AR) en las células epiteliales positivas para la racemasa (AMACR) se genera en base al análisis por ordenador de una imagen de tejido que muestra inmunofluorescencia.
- 20 3. El aparato de la reivindicación 2, en donde la intensidad del receptor de andrógenos (AR) comprende un área relativa de núcleos epiteliales con AR positivo con respecto al área total de núcleos epiteliales.
4. Un método para evaluar un riesgo de fallo clínico en un paciente posterior a que el paciente se haya sometido a una prostatectomía radical, y el método comprende:
- 25 evaluar un conjunto de datos para un paciente con un modelo predictivo de fallo clínico después de la prostatectomía, en donde el modelo se basa en:
- una puntuación de Gleason de biopsia;
 - la afectación de ganglios linfáticos;
 - 30 puntuación de Gleason de una muestra (prostatectomía);
 - intensidad del receptor de andrógenos (AR) en las células epiteliales positivas para racemasa (AMACR); y
 - tres características morfométricas derivadas de una imagen de tejido, y las tres características morfométricas comprenden:
 - 35 una primera medición morfométrica indicativa de la intensidad media del citoplasma epitelial como se expresa en el canal azul de la imagen del tejido,
 - una segunda medición morfométrica indicativa de la variación en la textura del estroma dentro del estroma como se expresa en el canal rojo de la imagen del tejido, y
 - 40 una tercera medición morfométrica indicativa de la variación en la textura entre los núcleos epiteliales como se expresa en el canal rojo de la imagen del tejido, evaluando así el riesgo de fallo clínico en el paciente.
5. El método de la reivindicación 4, en donde la intensidad del receptor de andrógenos (AR) en las células epiteliales positivas para la racemasa (AMACR) se genera en base al análisis por ordenador de una imagen de tejido que muestra inmunofluorescencia.
- 45 6. El método de la reivindicación 5, en donde la intensidad del receptor de andrógenos (AR) comprende un área relativa de núcleos epiteliales con AR positivo con respecto al área total de núcleos epiteliales.

7. Un medio legible por ordenador que comprende instrucciones ejecutables por ordenador grabadas en el mismo, en el que, cuando se ejecuta mediante un ordenador, las instrucciones hacen que el ordenador lleve a cabo el método de cualquiera de las reivindicaciones 4 a 6.

FIG. 1A

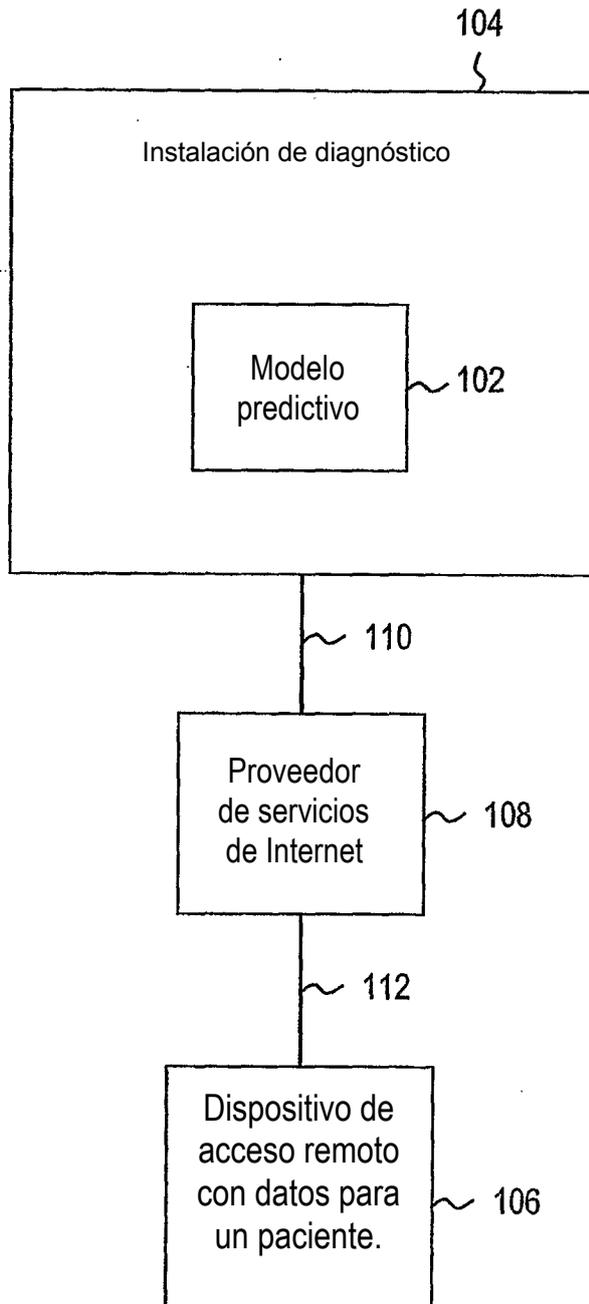


FIG. 1B

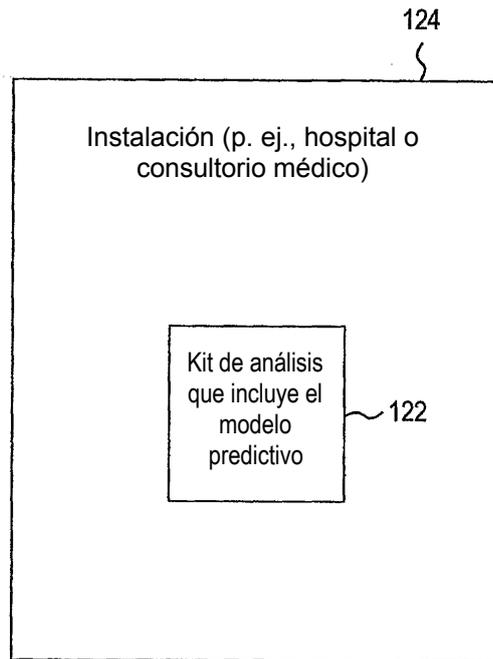


FIG. 1C

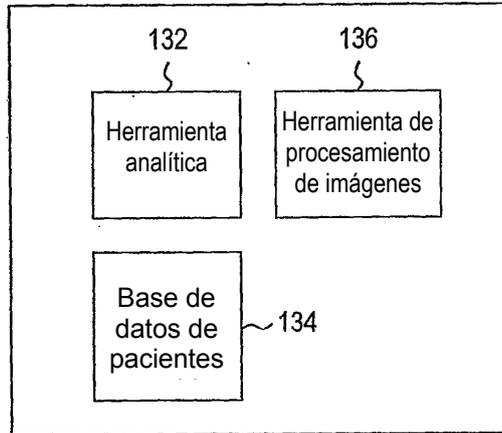


FIG. 2

Riesgo de resultado predicho para Edward Jones a lo largo del tiempo BioScore = 520

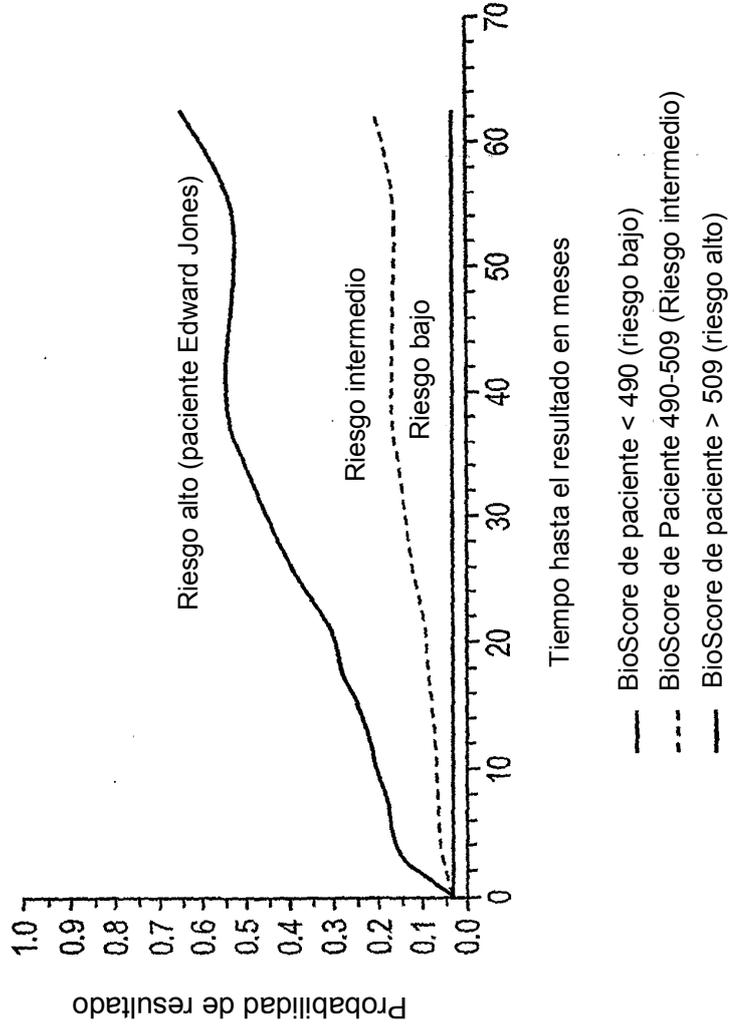


FIG. 3

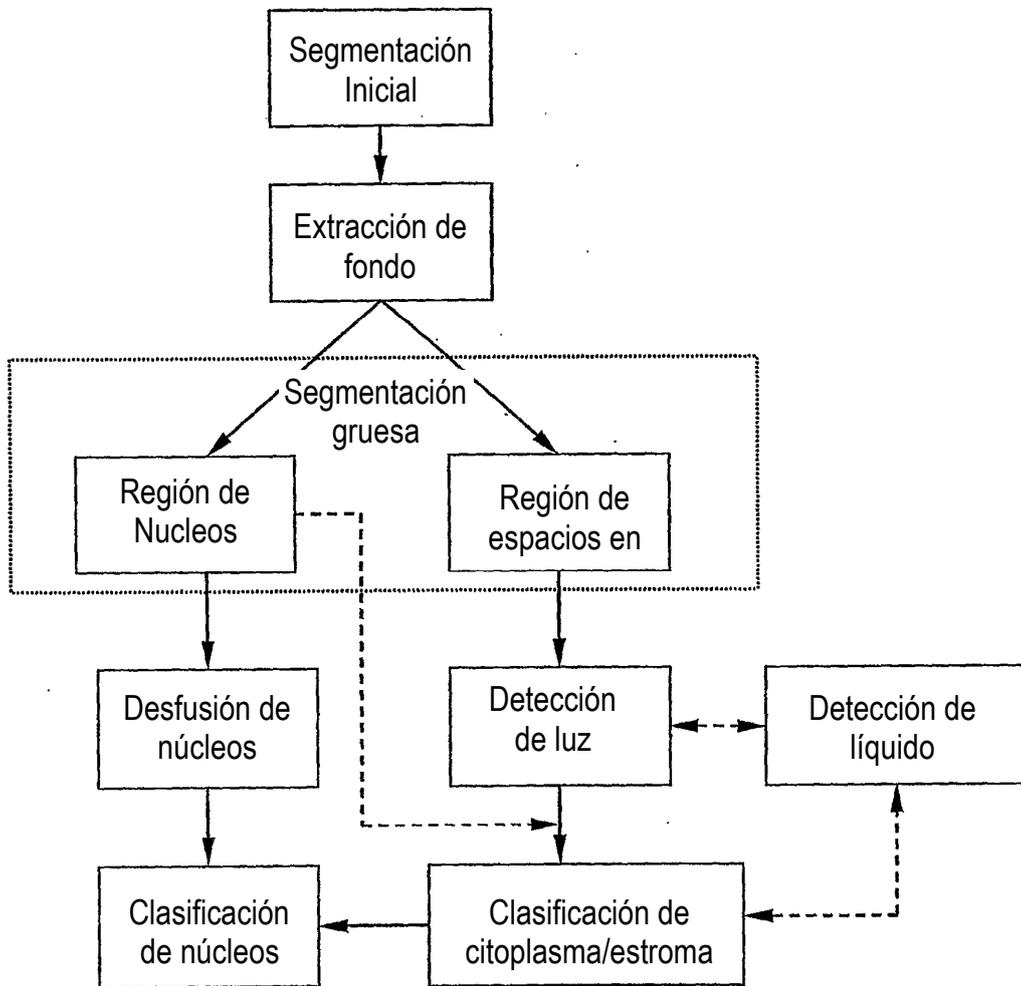


FIG. 4

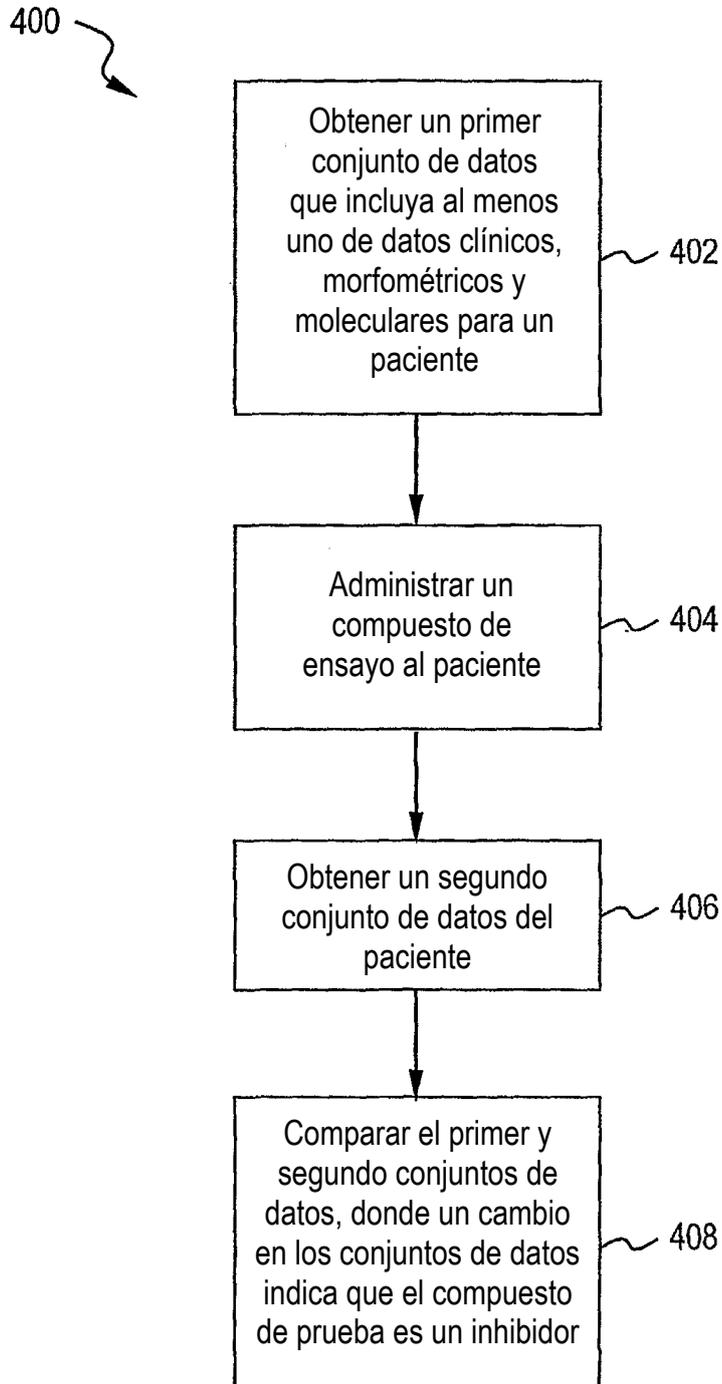


FIG. 5A

Imagen de tejido prostático sano

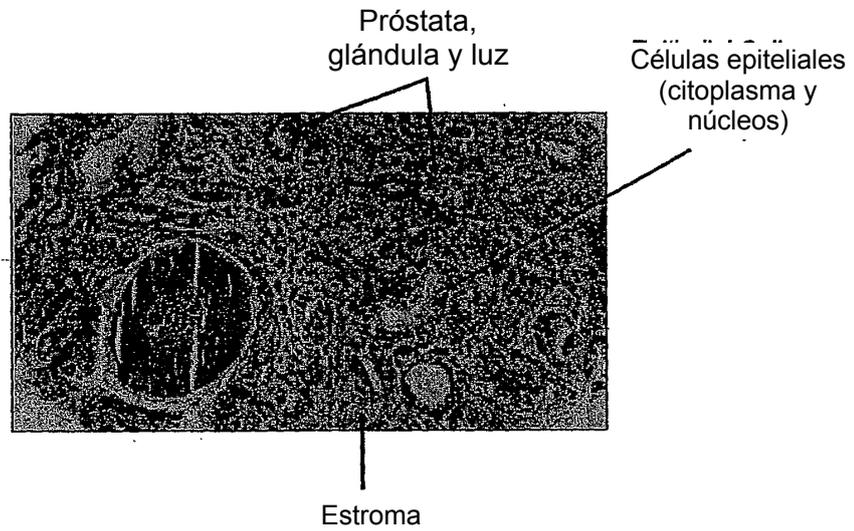


FIG. 5B

Imagen de tejido prostático

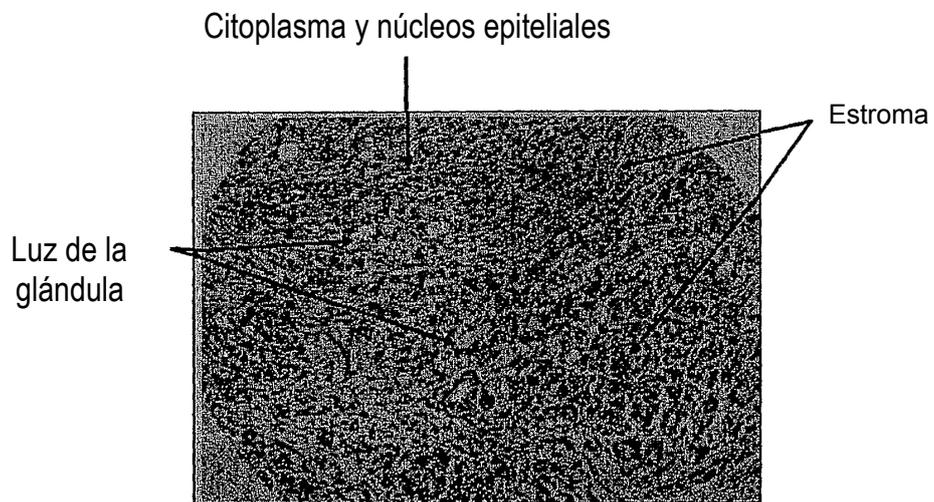
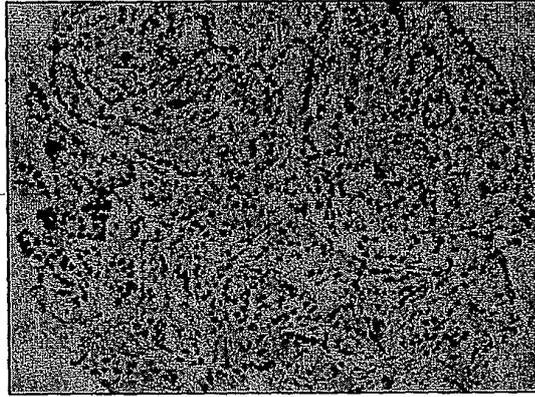


Fig. 6

Características seleccionadas para la predicción de recurrencia de PSA (Ejemplo 2, Estudio 1)

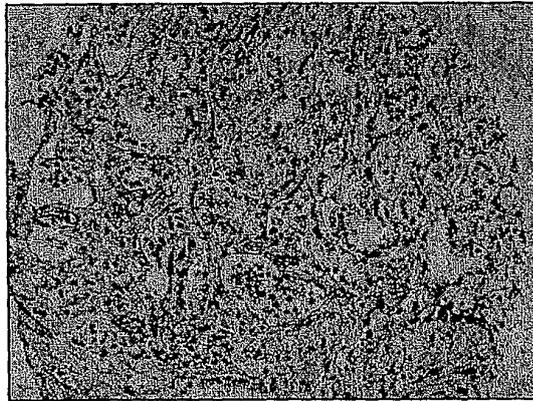
Característica	Descripción
Clínica	Puntuación de Gleason de biopsia
	Raza
	Estadio UICC
	Resultado de ploidía
	Resultado DRE
	Afectación de ganglios linfáticos
	Grado de Gleason de biopsia dominante
	Porcentaje de ploidía en fase S
	Puntuación de Gleason postoperatoria
	Estadio TNM
	Grado de Gleason postoperatorio dominante
	Edad
	Afectación de vesículas seminales
	PSA preoperatorio
	Porcentaje de fracción de ploidía
	Afectación del margen quirúrgico
Afectación extracapsular	
Molecular	AR-tumor
	AR-glándula
	CD34-tumor/PIN
	Ki67-tumor 2
	CD45-Pin3
	CD34-tumor/estroma
	Ki-67-tumor 3
	p27-tumor
	C14-PIN
	CD34-tumor
	PSA-Glándula
	PSMA-PIN
CD34-PIN/estroma	
CD45-tumor 3	
Morfométrica	Longitud mínima de glóbulos rojos en píxeles
	Compacidad máxima de núcleos epiteliales
	Radio mínimo de luz del recinto más pequeño
	Ancho mínimo de núcleos epiteliales en píxeles
	Densidad máxima de estroma
	Longitud máxima del borde de la luz en píxeles
	Desviación estándar mínima de núcleos epiteliales en el canal 2
	Radio máximo de núcleos epiteliales del recinto más pequeño
	Desviación estándar del citoplasma de la longitud del borde en píxeles
Desviación estándar del fondo del área en píxeles	

FIG. 7A



AR

FIG. 7B



CD34

FIG. 8

Grupos de riesgo para la recurrencia de PSA utilizando la puntuación del modelo NNci

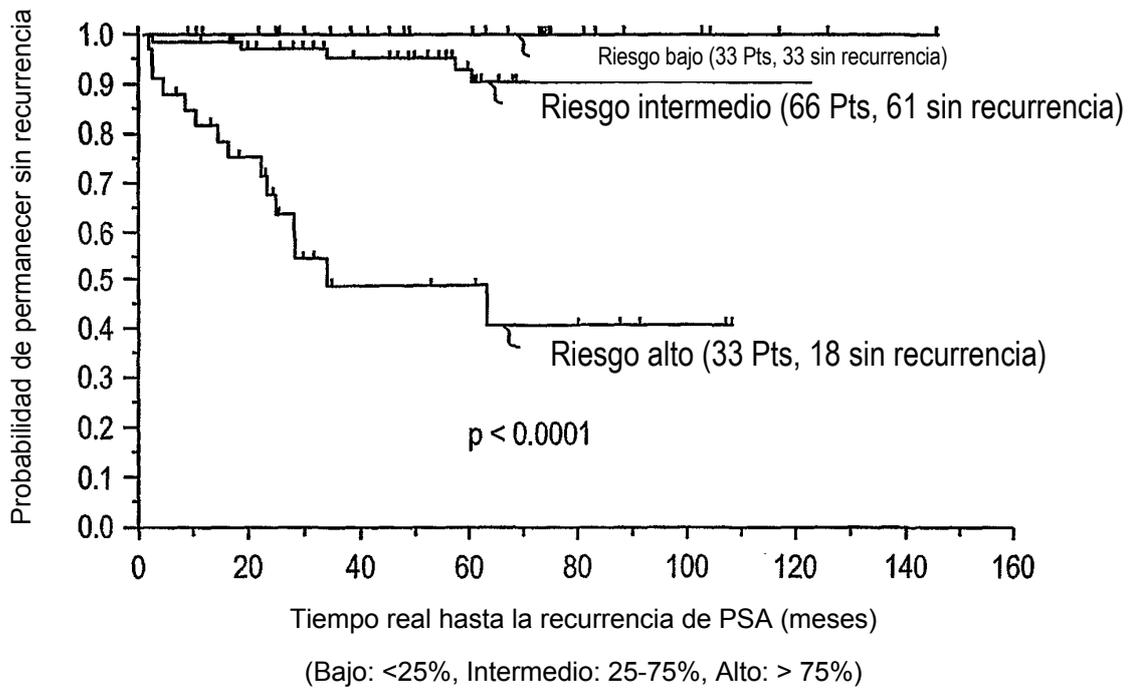


Fig. 9

Características seleccionadas para la predicción de la recurrencia de PSA (Ejemplo 2, Estudio 2)

Clínica	Estadio clínico TNM	0,7431
	Márgenes quirúrgicos	0,7937
	Ganglios linfáticos	0,8376
Molecular	Índice de tinción de AR (tumor)	0,8528
Morfométrica	EpithelialNucleiMinCompactne0215	0,8222
	StromaMaxStddevChannel30569	0,8483
	CytoplasmStddevMaxDiff0148	0,8569
	RedBloodCellMeanAreaPxl0386 '	0,8596
	RedBloodCellStddevAreaPxl0388	0,8621
	LumenMinAsymmetry0295	0,8635

FIG. 10

Grupos de riesgo para la recurrencia de PSA utilizando la puntuación del modelo SVRc

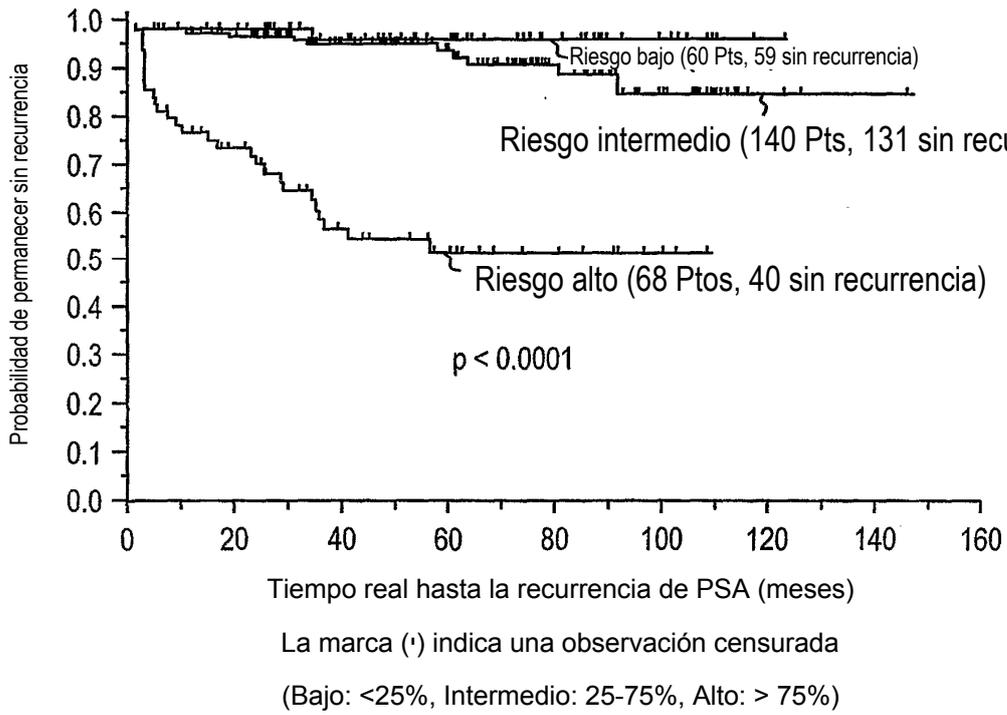


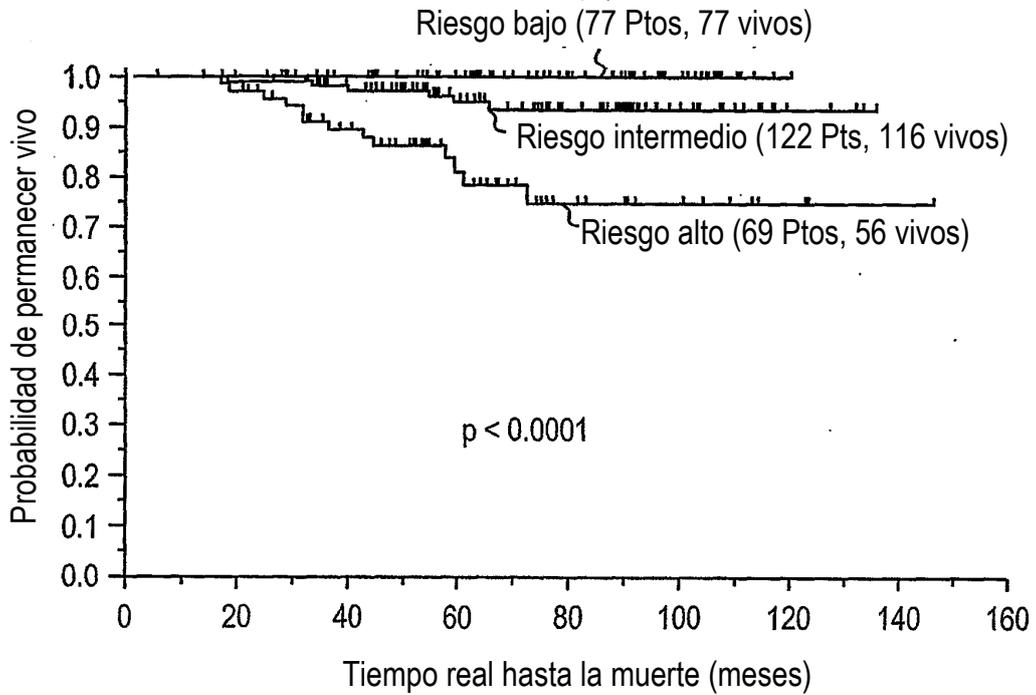
Fig. 11

Características seleccionadas para la predicción de la supervivencia general (Ejemplo 2, Estudio 3)

Clínica	tnm	0,7362
	edad	0,7906
Molecular	psapsi	0,7595
Morfométrica	StromaMinMeanChannel10535	0,6804
	RedBloodCellMeanStddevChann30474	0,7475
	StromaMinMeanChannel20539	0,7722
	RedBloodCellMinMeanChannel20443	0,7772
	RedBloodCellStddeStddeChann20472	0,7809
	StromaMaxMaxDiff0529	0,7852
	EpitheNucleMeanBordeLengtPx10206	0,7888
	EpithelialNucleMeanAreaPx10194	0,7921
	EpithelNucleiStddevElliptFit0228	0,7951
	RedBloodCellStddeStddeChann30476	0,7964
	RedBloodCellStddevElliptiFit0420	0,7976

FIG. 12

Grupos de riesgo para la supervivencia general utilizando la puntuación del modelo SVRc



La marca (*) indica una observación censurada
(Bajo: <25%, Intermedio: 25-75%, Alto: > 75%)

Fig. 13

Características seleccionadas para la predicción de fallo clínico (Ejemplo 3)

Clínica

Extensión extracapsular

Invasión de vesículas seminales

Grado de Gleason de prostatectomía dominante

Invasión de ganglios linfáticos

Morfométrico

Área de citoplasma dividida por el área total de tejido

Desviación estándar del citoplasma del canal rojo medio

Área de lumen dividida por el área total de tejido

Fig. 14

Características seleccionadas para la predicción de recurrencia de PSA (Ejemplo 5)

Clínica	Ganglios linfáticos	-9,3742
	Afectación del margen quirúrgico	-7,3159
	Afectación de vesículas seminales	-5,2103
Molecular	Índice de tinción de AR (tumor)	-3,5404
Morfométrica	AreaStromadivTotTissueArea	-0,34225
	AreaEpitNuclidivTotTissueArea	3,2975

Fig. 15

Características seleccionadas para la predicción de recurrencia de PSA (Ejemplo 6)

Clínica	Ganglios linfáticos	-23,32
	Afectación del margen quirúrgico	-11,73
	Puntuación de Gleason de biopsia	-10,60
	Afectación de vesículas seminales	-6,40
Molecular	Índice de tinción de AR (tumor)	-10,49
Morfométrica		-16,15
	AreaStromadivTotTissueArea	11,54
	AreaEpitNudivTotTissueArea StromaMeanStddevRedChann	-11,26

FIG. 16

Características seleccionadas para la predicción de fallo clínico tras prostatectomía (Ejemplo 7)

Clínica	Puntuación de Gleason de biopsia	-18,08
	Afectación de ganglios linfáticos	-11,87
	Puntuación de Gleason de la muestra (prostatectomía)	-6,36
Molecular	Intensidad de AR dentro de las células epiteliales AMACR	-6,20
Morfométrica	CytoplasmMeanMeanChannel	19,68
	StromaMeanStddevChannel	-10,06
	EpitheNucleiStddevMeanChann	-8,87