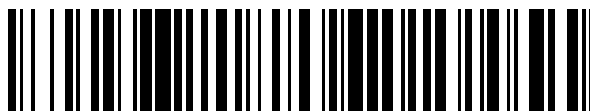


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 808 569**

51 Int. Cl.:

G16H 50/20

(2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **07.04.2008** **PCT/US2008/004523**

87 Fecha y número de publicación internacional: **16.10.2008** **WO08124138**

96 Fecha de presentación y número de la solicitud europea: **07.04.2008** **E 08742637 (5)**

97 Fecha y número de publicación de la concesión europea: **10.06.2020** **EP 2145276**

54 Título: **Sistema y método para manejar, diagnosticar y predecir la aparición de una afección médica**

30 Prioridad:

05.04.2007 US 922163 P
05.04.2007 US 922149 P
13.04.2007 US 923447 P
09.01.2008 US 10598

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
01.03.2021

73 Titular/es:

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

72 Inventor/es:

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

74 Agente/Representante:

SÁNCHEZ SILVA, Jesús Eladio

ES 2 808 569 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

Sistema y método para manejar, diagnosticar y predecir la aparición de una afección médica

5 Campo de la invención

La presente descripción se refiere a métodos y sistemas que usan información clínica, información molecular e información morfométrica 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). Por ejemplo, en un ejemplo, la descripción se refiere a métodos y sistemas que usan información clínica, molecular y morfométrica para tratar, diagnosticar y predecir la recurrencia del cáncer de próstata.

Antecedentes

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 los individuos 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.

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 prostático 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.

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 con respecto al tratamiento del cáncer de próstata después de que un paciente se haya sometido a una prostatectomía radical se limitan a los análisis de cribado de PSA en suero y los nomogramas generalizados. Un nomograma postoperatorio, desarrollado por Kattan y otros, patente de EE.UU. nº 6.409.664, es ampliamente usado 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 usadas 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 fallo 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. 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 exacta que un lanzamiento de moneda.

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 automáticamente grandes cantidades de características cuantitativas de valor continuo de una sola imagen se ha convertido en una realidad. Se dice que una característica X tiene un valor continuo si, para cierto $A < B$, 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 desempeña un papel más importante en el diagnóstico (por ejemplo, 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 tejido son típicamente más complejas que las imágenes celulares, y requieren un conocimiento experto completo de los dominios para comprenderlas.

El documento US 2005/0262031 A1 describe métodos de sistemas que usan información clínica, información molecular e información morfométrica generada por ordenador en un modelo predictivo para predecir la aparición o recurrencia de una afección médica tal como el cáncer.

5 En vista de lo anterior, sería conveniente proporcionar sistemas y métodos para tratar, diagnosticar y predecir la aparición de afecciones médicas, respuestas y otros fenómenos médicos con un mayor poder predictivo. También sería conveniente proporcionar sistemas y métodos implementados por ordenador que utilicen información a nivel de tejido para tratar, diagnosticar y predecir la aparición de afecciones médicas.

10 Resumen de la invención

La invención se define de acuerdo con las reivindicaciones adjuntas.

Breve descripción de los dibujos

15 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:

20 Las Figuras 1A y 1B son diagramas de bloques de sistemas que usan 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;

25 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;

30 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 las imágenes;

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

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;

40 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;

45 La Figura 9 muestra varias características clínicas, moleculares y morfométricas generadas por ordenador usadas 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;

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

55 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;

60 La Figura 13 muestra varias características morfométricas clínicas y generadas por ordenador usadas 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 clínicas, moleculares y morfométricas generadas por ordenador usadas por un modelo para predecir la recurrencia del cáncer de próstata;

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

La Figura 17 muestra varias características clínicas y de imagen de tejido generadas por ordenador usadas por un modelo para predecir la recurrencia del cáncer de próstata;

- 5 Las Figuras 18a-e ilustran la segmentación de imágenes múltiples para cuantificar AR nuclear en tejido prostático;

La Figura 19 muestra varias características clínicas y de imagen de tejido generadas por ordenador usadas por otro modelo para predecir el fallo clínico en un paciente posterior a la prostatectomía radical; y

- 10 La Figura 20 es un gráfico de línea que muestra las curvas de supervivencia para los 51 pacientes en un modelo para predecir la supervivencia de individuos con cáncer de pulmón de células no pequeñas (NSCLC) que se han tratado con gefitinib, separados en 2 grupos por puntuación del modelo por encima o por debajo del punto de corte 39, donde una puntuación de 39,5 o más predijo un tiempo de supervivencia general más corto.

- 15 Descripción detallada de las realizaciones preferidas

La descripción en el presente documento 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 clínica, molecular y morfométrica generada por ordenador puede usarse para predecir la recurrencia del cáncer de próstata. Las enseñanzas proporcionadas en el presente documento pueden usarse además para predecir la aparición de otras afecciones médicas tales como, por ejemplo, otros tipos de enfermedades (por ejemplo, neoplasias epiteliales y mixtas, que incluyen 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 otros individuos pueden usar estas predicciones para, por ejemplo, seleccionar un curso de tratamiento apropiado para un paciente y/o diagnosticar una afección médica en el paciente.

- En el presente documento se describe 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 clínicas, moleculares y morfométricas 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 usar para evaluar los datos de un nuevo paciente para predecir la aparición de la afección médica para el nuevo paciente. En algunas situaciones, la herramienta analítica solo puede usar un subconjunto de los tres tipos de datos (por ejemplo, datos clínicos y morfométricos solamente) para generar el modelo predictivo.

- Los datos clínicos, moleculares y/o morfométricos usados en el presente documento 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 para generar modelos predictivos se describen a continuación en relación con, por ejemplo, las Tablas 1, 2, 4 y/o 6. Se comprenderá 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 introducirlas 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. Mediante 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, otros individuos y/o equipos de procesamiento automatizado (por ejemplo, un sistema de análisis de imágenes de tejido) 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 comprenderá 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 varias propiedades estructurales y/o espectrales, por ejemplo, de muestras de tejido. En una realización, los datos morfométricos pueden incluir datos de características morfométricas de estroma, citoplasma, núcleos epiteliales, núcleos estromales, lumen, glóbulos rojos, artefactos de tejidos, fondo de tejidos o una combinación de los mismos. En un aspecto de la presente invención, 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 usa el programa informático Definens Cellenger. Tal sistema puede recibir una imagen teñida con H&E 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, tales 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 (por ejemplo, carcinoma ductal in situ para mama), tamaño, forma y patrón/distribución regional de células epiteliales (por ejemplo, mama, pulmón), grado de diferenciación (por ejemplo, diferenciación escamosa con cáncer de pulmón de células no microcíticas (NSCLC, producción de mucina como se observa con varios adenocarcinomas de mama y colon)), distribución morfológica/microscópica de las células (por ejemplo, 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).

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 mediante el uso de 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.

Las Figuras 1A y 1B muestran sistemas ilustrativos que usan 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 otro individuo asociado 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 el uso en una instalación tal como un hospital, otra instalación médica u otra ubicación adecuada.

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. En otra realización, 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. En algunas realizaciones, los datos para el paciente pueden generarse 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 H&E 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 tejido y/o muestras en relación con la Figura 3.

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, tal 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 (por ejemplo, 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. Un médico u otro individuo puede usar tal informe para ayudar a determinar un grado de tumor clínico-diagnóstico más refinado, desarrollar un medio eficaz para subclasificar a los pacientes y finalmente generar opciones de tratamiento más exactas (y apropiadas) para el paciente individual. El informe también puede ser útil porque puede ayudar al médico o individuo a explicar el riesgo al paciente.

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 tal 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 otros individuos 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. Tal servidor puede incluir un componente distinto de equipo y/o almacenamiento informático, pero también puede ser un programa informático o una combinación de equipo y programa informático. El servidor puede implementarse mediante el uso de uno o más ordenadores.

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.

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 (por ejemplo, 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). En algunas realizaciones, 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. En otras realizaciones, el kit de análisis 122 puede recibir datos preempaquetados de 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, tales 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, por ejemplo, cuando 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.

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 H&E, 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.

La base de datos 134 puede incluir cualquier información adecuada del paciente, tal 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), tales 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), tales 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 usar datos censurados y no censurados.

En una realización, la herramienta analítica 132 puede incluir una máquina de vectores de soporte (SVM). En una realización de este tipo, la herramienta 132 incluye preferiblemente una SVM capaz de realizar una regresión de vectores de soporte en datos censurados (SVRC). Como se describe en la solicitud de patente de EE.UU. en trámite nº 10/991,240, en la SVRC se proporciona una nueva función modificada de pérdida/penalización para el uso dentro de una SVM que puede permitir que la 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 con respecto al 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 a partir 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 mediante el uso de 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 usada 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 típicamente solo indica que el evento no ha sucedido dentro de un tiempo determinado, y no hay indicación de cuándo ocurrirá después del tiempo determinado, si es que ocurre.

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

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

donde $e = f(\mathbf{x}) - y$; y

$$f(\mathbf{X}) = \mathbf{W}^T \Phi(\mathbf{X}) + b$$

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

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

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

donde $e = f(\mathbf{x}) - y$

y $\varepsilon_n^* \leq \varepsilon_n$ y $C_n^* \geq C_n$.

En la descripción anterior, la $\mathbf{W}$ y b se obtienen al resolver un problema de optimización, cuya forma general es:

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{W}^T \mathbf{W}$$

$$\text{s.t. } y_i - (W^T \Phi(x_i) + b) \leq \varepsilon \\ (W^T \Phi(x_i) + b) - y_i \leq \varepsilon$$

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 usa 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 exactitud, lo que no es factible. La función de pérdida permite un rango de error respecto al 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 C mitiga su contribución. 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 el 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:

$$\min_{w, b} P = \frac{1}{2} W^T W + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$

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

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:

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

$$\text{s.t. } y_i - (W^T \Phi(x_i) + b) \leq \varepsilon_i + \xi_i \\ (W^T \Phi(x_i) + b) - y_i \leq \varepsilon_i^* + \xi_i^* \\ \xi_i^{(*)} \geq 0, i=1 \dots l \\ \text{donde } C_i^{(*)} = s_i C_s^{(*)} + (1 - s_i) C_n^{(*)} \\ \varepsilon_i^{(*)} = s_i \varepsilon_s^{(*)} + (1 - s_i) \varepsilon_n^{(*)}$$

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 .

Los detalles adicionales con respecto a los sistemas y métodos para realizar la regresión de vectores de soporte en datos censurados (SVRc) se describen en la solicitud de patente de EE.UU. nº 10/991,240 incorporada anteriormente, presentada el 17 de noviembre de 2004, y en la solicitud de patente provisional de EE.UU. nº 60/520,939, presentada el 18 de noviembre de 2003.

En otra realización, la herramienta analítica 132 puede incluir una red neuronal. En una realización de este tipo, la herramienta 132 incluye preferiblemente una red neuronal que es capaz de utilizar datos censurados. Además, la red neuronal usa 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 usado 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 la solicitud de patente de EE.UU. en trámite nº 11/067,066, presentada el 25 de febrero de 2005 (ahora patente de EE.UU. nº 7,321,881), este obstáculo puede superarse mediante el uso de 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 de la SVRc descrita anteriormente (por ejemplo, implementada con una máquina de vectores de soporte de la

herramienta analítica 132) o mediante el uso de 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 con respecto a 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:

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

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 t_j del paciente 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 exacto. Por lo tanto, al maximizar el CI de manera sustancial y preferida, o una aproximación del CI, se mejora el rendimiento de un modelo. Una realización de la presente invención proporciona una aproximación del CI de la siguiente manera:

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

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

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

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

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 $\hat{t}_i \cdot \hat{t}_j > \gamma$ y así hacer $I(\hat{t}_i, \hat{t}_j) = 1$.

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 eficazmente el sobreajuste de los datos durante el entrenamiento del modelo y hace que la optimización se centre preferiblemente en solo mover más pares de muestras en Ω para satisfacer $\hat{t}_i \cdot \hat{t}_j > \gamma$. 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 y en R es preferible para un mejor 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 conducir fácilmente a predicciones que están mucho más allá del rango de los datos de entrenamiento.

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 clasificar objetos patológicos en las imágenes (por ejemplo, clasificar objetos como citoplasma, lumen, 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 tejido con un aumento de 20X mediante el uso de 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 píxel en formato Tiff. Tal equipo es solo ilustrativo, y se puede usar cualquier otro 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. En una realización, la herramienta de procesamiento de imágenes puede incluir el Definiens Cellenger Developer Studio disponible comercialmente (v. 4.0) adaptado para realizar la segmentación y clasificación, por ejemplo, de los diversos objetos patológicos descritos anteriormente y para medir diversas características morfométricas de estos objetos. Se describen detalles adicionales con respecto al 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 de los canales rojo, verde, azul (RGB), tales 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). 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 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.

Segmentación inicial. En una primera etapa, la herramienta de procesamiento de imágenes puede segmentar una imagen (por ejemplo, una imagen de micromatriz de tejido (TMA) teñido con H&E o una tinción de H&E de un corte de tejido completo) en pequeños grupos de píxeles contiguos conocidos como objetos. Estos objetos pueden obtenerse mediante un método de crecimiento de regiones que encuentra regiones contiguas sobre la base de 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 morfométricas 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. En una realización, el tamaño de los objetos puede controlarse para que sea de 10-20 píxeles en el nivel más preciso. 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.

Extracción del fondo. Después de la segmentación inicial, la herramienta de procesamiento de imágenes puede segmentar la muestra de punción de tejido de la imagen respecto al fondo (región transparente del portaobjetos) mediante el uso del 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: "muestra de punción de tejido" y "fondo". Cualquier píxel con un valor de intensidad mayor o igual al umbral se clasifica como un píxel de "muestra de punción 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 .

Segmentación gruesa. En una etapa siguiente, la herramienta de procesamiento de imágenes puede volver a segmentar el primer plano (por ejemplo, muestra de punción 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 H&E 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 de intensidad promedio de píxeles rojos y azules. La relación longitud/ancho también se puede usar para determinar si un objeto debe clasificarse como área de núcleos. Por ejemplo, los objetos que caen por debajo de un umbral de característica ($R-B$) y por debajo de un umbral de longitud/ancho pueden clasificarse como área de núcleos. De manera

similar, se puede usar un umbral del canal verde para clasificar los objetos en la muestra de punción de tejido como espacios en blanco. El estroma del tejido está dominado por el color rojo. La diferencia de intensidad d , la "relación de rojo" $r = R/(R + G + B)$ y la desviación estándar del canal rojo σ_R de los objetos de imagen pueden usarse para clasificar los objetos de estroma.

Clasificación de espacios en blanco. En la etapa de segmentación gruesa, las regiones del espacio en blanco pueden corresponder tanto al lumen (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.

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.

La herramienta de procesamiento de imágenes puede usar dos enfoques diferentes para desfusionar los núcleos. El primer enfoque puede basarse en un método de crecimiento de regiones que fusiona los objetos de 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.

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 imagen que pertenecen a los núcleos etiquetados pueden usarse para diseñar clasificadores estadísticos.

En una realización, para reducir el número de dimensiones del espacio de características, la selección de características puede realizarse en el conjunto de entrenamiento mediante el uso de 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 método 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.

En algunas realizaciones, 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 base a 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 (por ejemplo, valores medios del canal de color, desviaciones estándar y brillo) y propiedades estructurales/de forma (por ejemplo, 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 el presente documento) muestran varios ejemplos de características morfométricas que pueden medirse de acuerdo con la presente invención. Las características morfométricas de estas tablas se nombran mediante el uso de 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 comprenderá un experto en la técnica.

Se comprenderá 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.StdDevAreaPxI", "Lumen" indica un tipo de objeto de imagen, "StdDev" indica una estadística (desviación estándar) que se calculará mediante el uso de todas las instancias del Lumen identificadas, y "AreaPxI" 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 morfológicas para todos los objetos previamente segmentados y clasificados en la imagen. Por ejemplo, la herramienta de procesamiento de imágenes puede medir características morfológicas para objetos que incluyen "Fondo", "Citoplasma", "Núcleos Epiteliales", "Lumen", "Estroma", "Núcleos Estromales" 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 (por ejemplo, área rosa que rodea un núcleo epitelial en una imagen, por ejemplo, de tejido teñido con H&E). "Núcleos epiteliales" se refiere al núcleo presente dentro de las células epiteliales/células luminales y basales de la unidad glandular, que aparecen como objetos "redondos" rodeados de citoplasma. "Lumen" se refiere al espacio glandular central donde las células epiteliales depositan las secreciones, que aparecen como áreas blancas cerradas rodeadas de células epiteliales. Ocasionalmente, el lumen puede llenarse con líquido prostático (que típicamente aparece rosado en el tejido teñido con H&E) u otros "restos" (por ejemplo, macrófagos, células muertas, etc.). Juntos, el lumen 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 H&E. "Núcleos estromales" 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 a continuación, es una relación relativa del área del núcleo respecto al citoplasma. Cuanto más anaplásica/maligna es la célula epitelial, más área está ocupada por el núcleo y mayor es la relación. "EN2SN" es el porcentaje o la cantidad relativa de células epiteliales respecto al estroma presentes en la imagen digital del tejido. "L2Core" es el número o área de lumen 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 lumen. En general, esto se debe a que las células epiteliales se replican de manera incontrolada cuando se produce el cáncer, lo que hace que el lumen se llene con las células epiteliales.

En un aspecto de la presente invención, se proporcionan sistemas y métodos para el cribado de un compuesto inhibidor de una afección médica (por ejemplo, enfermedad). La Figura 4 es un diagrama de flujo de las etapas ilustrativas involucradas en el cribado de un compuesto inhibidor de acuerdo con una realización de la presente invención. En la etapa 402, se puede obtener un primer conjunto de datos para un paciente que incluye uno o más datos clínicos, datos morfológicos y datos moleculares. Se puede administrar un compuesto de análisis 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 análisis indica que el compuesto de análisis es un compuesto inhibidor. La etapa 408 de comparación de los conjuntos de datos puede incluir, por ejemplo, comparar un resultado generado por un modelo predictivo de la presente invención que responde a una entrada del primer conjunto de datos con un resultado generado 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 varias aplicaciones ilustrativas de realizaciones de la presente invención para la predicción de afecciones médicas. En un primer ejemplo, una realización de la presente invención usó datos clínicos y morfológicos para predecir la recurrencia del cáncer de próstata. En un segundo ejemplo, una realización de la presente invención usó datos clínicos, morfológicos y moleculares para predecir la recurrencia del cáncer de próstata y la supervivencia general. En un tercer ejemplo, se usó una realización de la presente invención para predecir la aparición de una enfermedad agresiva posterior a una prostatectomía del paciente. En un cuarto ejemplo, se usó una realización de la presente invención para predecir la toxicología hepática. En los ejemplos quinto, sexto y octavo, se usaron realizaciones de la presente invención para predecir la recurrencia del cáncer de próstata. En los ejemplos séptimo y noveno, se usaron realizaciones de la presente invención para predecir el fallo clínico después de la prostatectomía. En un décimo ejemplo, se usó una realización de la presente invención para predecir la supervivencia de individuos con cáncer de pulmón de células no pequeñas (NSCLC) que se han tratado con gefitinib.

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 una etapa 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 exacto es fundamental al seleccionar la intervención terapéutica inicial, ya que la mayoría de los tumores son de progresión lenta 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 etapa temprana es curable con terapia local, aproximadamente del 25 al 40 % de los hombres desarrollará una recurrencia del PSA / 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 exactitud, 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 del 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 general.

Los nomogramas de pronóstico basados solo en datos de características clínicas, de hecho, proporcionan predicciones útiles de estados y resultados clínicos, pero necesitan una mejora tanto en la exactitud como en la universalidad [16]. Las realizaciones de la presente descripción proporcionan un enfoque de 'Patología de sistemas' para mejorar con éxito la exactitud de un modelo predictivo para PSA / BCR después de la prostatectomía. Esto representa una visión 'individualizada' de la propia muestra tumoral del paciente, que incluye la evaluación cuantitativa de las características morfológicas celulares y microanatómicas, perfiles clínicos y marcadores moleculares para crear un modelo de predicción muy exacto e integrador. Al utilizar la experiencia de los dominios, se han desarrollado modelos muy exactos para predecir la recurrencia del 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 morfológicas tisulares se pueden derivar e incluir en un análisis predictivo usado 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 del 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 morfológicas que representan en parte una puntuación de Gleason muy exacta, no subjetiva y cuantitativa, además de varios descriptores de tejidos novedosos que fueron importantes en el desarrollo y la exactitud del modelo. Las características morfológicas 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 receptora de andrógenos (AR) recibe hormonas androgénicas naturales (testosterona y su metabolito 5 .alpha. reducido, dihidrotestosterona) después de la síntesis de estas hormonas en las células de Leydig de los testículos. 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 de acción en trans 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 y otros, patente de EE.UU. nº 6,472,415, propone que el crecimiento del cáncer de próstata en etapas tempranas está controlado por andrógenos y se puede detener, al menos temporalmente, por la privación de andrógenos. French y otros, patente de EE.UU. nº 6,821,767, propone varias formas de medir 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 el presente documento.

Ejemplo 1: Predicción de la recurrencia del cáncer de próstata con datos clínicos y morfométricos

Se extrajeron varias características morfométricas en bruto, inicialmente hasta quinientas, de cada imagen de tejido prostático mediante el uso del sistema de análisis de imágenes de tejido 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 usó 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 mediante el uso de 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 del 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 exactitud 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 la 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 potencial en diferentes tipos de datos, se usaron dos métodos de análisis de supervivencia: 1) un modelo de regresión de Cox [17]; y 2) SVRc que se describió anteriormente y se aplica a una máquina de vectores de soporte. El índice de concordancia estimado mediante el uso de la validación cruzada 5 veces se usó para medir la exactitud predictiva de los modelos [13] [18].

Ambos métodos se aplicaron a tres conjuntos de datos: (1) las características clínicas de puntuación de Gleason solas; (2) las características morfométricas seleccionadas solas; y (3) la combinación de las características morfométricas y las características clínicas de 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.StdDevAreaPxl) y lumen (Lumen.MaxBorderLengthPxl, Lumen.MinRadiusofsmallestenclosing, Lumen.StdDevBorderLengthPxl, Lumen.SumBorderlengthPxl, 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) del lumen se correlacionan con la progresión de la enfermedad. Cuanto más pequeño y compacto es el lumen, más avanzado es el cáncer. De hecho, con un cáncer más agresivo (grado de Gleason 4 y 5), se puede esperar que el lumen 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 relación del producto de longitud y ancho de los núcleos epiteliales con respecto al área de 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 típicamente 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 de la muestra de punción de tejido usada en el análisis.

Tabla 3 - Comparación de la exactitud de predicción

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

De acuerdo con 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 morfológicas 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 morfológicas puede mejorar la predicción general de la recurrencia.

- 5 Ejemplo 2: Predicción de la recurrencia del cáncer de próstata y de la supervivencia general con datos clínicos, morfológicos y moleculares

10 Se realizaron dos estudios que predijeron con éxito la recurrencia del antígeno prostático específico (PSA) con un 88 % y un 87 % de exactitud predictiva, respectivamente. Al combinar las características clínicas, moleculares y morfológicas 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.

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

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

30 El presente estudio revela una tendencia progresiva de predicción mejorada de la 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.

35 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 mediante el uso de información anónima de los pacientes, que incluía la edad del paciente, el PSA preoperatorio y el grado de Gleason.

40

Tabla 4. Características clínicas recopiladas

Característica	Descripción
edad	Edad (en años)
raza	Raza
prepsa	Antígeno prostático específico (ng/dl)
tnm	Etapas clínicas TNM
uicc	Etapas clínicas 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
bxggl	Grado de Gleason dominante de la biopsia
bxggtot	Puntuación de Gleason de la biopsia
ggl	Grado de Gleason dominante postoperatorio
ggtot	Puntuación de Gleason postoperatoria
prsltd	Diploide, Tetraploide, Aneuploide

65

Característica	Descripción
pp_sphas	Porcentaje de células en ploidía en fase S
pp_frac	Fracción de proliferación en ploidía

Se construyeron micromatrices de tejido (TMA) a partir de bloques seleccionados de las muestras de prostatectomía. Las muestras de punción 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 con polilisina cargada y se usaron para análisis morfológicos e inmunohistoquímicos (IHC) (véase a continuación).

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 eficaz 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 del PSA (bioquímica).

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 muestras de punción de TMA, se analizaron cortes completos de prostatectomías radicales. El estudio 3 examinó la misma cohorte de 268 pacientes, pero se usó para predecir la supervivencia general, donde el resultado de interés fue la muerte por cualquier causa.

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

Análisis Molecular. Se aplicó un panel de 12 biomarcadores que incluyen Citoqueratina 18 (células luminales), 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 anticuerpos primarios (mostrados 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)
Citoqueratina 18	Clon DC-10 (Novocastra)
CD45	Clon X16/99
CD68	Clon 514H2 (Novocastra UK)
CD34	Clon QBEnd 101 (DAKO)
AR	Clon AR27 (Novocastra)
Citoqueratina 14	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

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 respectivas muestras de punción triplicadas 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 desempeña 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 inmunitarias tales 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)
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	Citoqueratina 18 (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)

	Característica	Descripción
5	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)
10	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)
15	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)
20	arasi	Índice de tinción de AR (glándula)
	c14asi	Índice de tinción de citoqueratina 14 (glándula)
	cdlasi	Índice de tinción de ciclina D1 (glándula)
	psaasi	Índice de tinción de PSA (glándula)
	psmaasi	Índice de tinción de PSMA (glándula)
25	p27asi	Índice de tinción de p27 (glándula)
	her2asi	Índice de tinción de Her-2/neu (glándula)

30 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 del PSA y desarrollar un modelo basado en esas características, con el objetivo final de usar 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 usó 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.

40 La exactitud predictiva de un modelo se evaluó mediante el uso del 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 aleatoriamente 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.

55 Estudio 1. En este análisis, el modelo de SVRc descrito anteriormente se aplicó secuencialmente a los datos clínicos, moleculares y morfométricos, y las características clínicas sirvieron primero como ancla para un método de selección "voraz directo" ("FS") de características a través de la ejecución de SVRc en los datos moleculares. Después de este paso, se ejecutó un segundo método de selección voraz directo de características por SVRc en los datos morfométricos, mediante el uso de la combinación de las características clínicas y moleculares seleccionadas como ancla. El último paso involucró la ejecución de un método de selección voraz inverso 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.

65 Se evaluó la exactitud predictiva del modelo mediante el uso de una validación interna y externa. La validación interna se realizó mediante el uso de una validación cruzada quíntuple. 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 aleatoriamente un subconjunto de pacientes del conjunto completo de registros de pacientes y solo se usaron los pacientes restantes para construir el modelo predictivo mediante el uso del procedimiento que se acaba de describir. Los registros retenidos se usaron para aplicarlos al modelo entrenado para obtener una exactitud 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.

Mediante el uso del 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 mediante el uso de una función objetivo sustancialmente de acuerdo con una aproximación del índice de concordancia. El resultado de este modelo final se usó para estimar el riesgo individual futuro del paciente de recurrencia del PSA.

Estudio 2. Los objetivos de este estudio fueron idénticos al Estudio 1; sin embargo, se usaron diferentes procedimientos de selección y validación de características. En lugar de usar el enfoque de anclaje, todas las características se clasificaron según su asociación con el tiempo hasta la recurrencia del 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 imagen, y estas características se evaluaron luego en una serie de modelos de n características (por ejemplo, 1 característica, 2 características, 3 características, etc.). Mediante el uso de un proceso de selección de características directo, 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, mediante el uso de un proceso de selección de características inverso, 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 remuestreo 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 método de selección de características se ejecutó mediante el uso solamente de las características que se encuentran en el nomograma postoperatorio de Kattan, que se describe en Kattan y otros, patente de EE.UU. nº 6,409,664, que se incorpora en el presente documento como referencia en su totalidad. El resultado del modelo final se usó para estimar el riesgo individual futuro de recurrencia del PSA del paciente.

Estudio 3. El objetivo de este estudio fue identificar características predictivas de supervivencia general mediante el uso de la misma cohorte y conjunto de características analizadas en el Estudio 2, así como el mismo método 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 morfológicos, firmas moleculares y perfiles clínicos de pacientes) para desarrollar modelos predictivos para la recurrencia del 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 clínico-patológicas del Estudio 1 en un análisis estándar del Modelo de Cox, la exactitud para predecir la recurrencia del PSA fue solo del 59 %. Solamente después de la integración de las características morfológicas y moleculares con SVRc, el nivel de exactitud predictiva aumentó al 88 %. Las siguientes secciones describen cómo se logró esta mejora.

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

ES 2 808 569 T3

	Estudio 1	Estudio 2 y 3
5	Mediana	63
	Rango	40 - 81
	Estudio 1	Estudio 2 y 3
	Raza	
	Caucásico	120 (90,9 %)
10	Hispano	241 (89,9 %)
	Afroamericano	8(6,1 %)
	Desconocido	12 (4,5 %)
	PSA preoperatorio (ng/dl)	2 (1,5 %)
15	Media	9 (3,4 %)
	Mediana	6 (2,2 %)
	Rango	12,2
	Etapa TNM	10,8
20	pT2N0	8,2
	pT3aN0	1,1 - 81,9
	pT3bN0	0,9 - 81,9
	pT1-3N+	69 (52,3 %)
25	Etapa UICC	157 (58,6 %)
	T1a < 5 %	40 (30,3 %)
	T1b ≥ 5 %	72 (26,9 %)
	T1c no palpable ni visible	13 (9,8 %)
30	T2a ≤ ½ lóbulo	22 (8,2 %)
	Etapa UICC	17 (6,3 %)
	T2b ≤ 1 lóbulo	
	T2c ambos lóbulos	27 (20,5 %)
35	T3a ECE unilateral	45 (16,8 %)
	T3c SV+	34 (12,7 %)
	Resultado de DRE	8 (6,1 %)
	No palpable	15 (5,6 %)
40	Palpable	2 (0,8 %)
	Afectación de ganglios linfáticos	
	Negativo	56 (42,4 %)
	Positivo	118 (44,0 %)
45	Afectación de las vesículas seminales	
	No	76 (57,6 %)
	Sí	150 (56,0 %)
	Márgenes quirúrgicos	
50	Negativo	113 (85,6 %)
	Positivo	236 (88,0 %)
	Afectación extracapsular	
	No	19 (14,4 %)
	Sí	32 (12,0 %)
55	Grado de Gleason dominante de la biopsia	
	1	108 (81,8 %)
	2	217 (81,0 %)
	3	24 (18,2 %)
60	4	85 (64,4 %)
	5	184 (68,7 %)
	Puntuación de Gleason de la biopsia	
	2	22 (16,7 %)
	3	38 (14,2 %)
65	0 (0,0 %)	1 (0,4 %)
	0 (0,0 %)	43 (16,0 %)
	0 (0,0 %)	184 (68,7 %)
	0 (0,0 %)	38 (14,2 %)
	0 (0,0 %)	2 (0,8 %)
	0 (0,0 %)	1 (0,4 %)
	0 (0,0 %)	0 (0,0 %)

	Estudio 1	Estudio 2 y 3
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 %)
10	Estudio 1	Estudio 2 y 3
Grado de Gleason dominante postoperatorio		
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 %)
7	79 (59,9 %)	148 (55,2 %)
8	10 (7,6 %)	12 (4,5 %)
9	1 (0,8 %)	4 (0,4 %)
25	Ploidía	
	Diploide	74 (56,1 %) 145 (54,1 %)
	Tetraploide	54 (40,9 %) 115 (42,9 %)
30	Aneuploide	4 (3,0 %) 8 (3,0 %)
Porcentaje de ploidía en fase S (%)		
	Media	2,3 2,4
	Mediana	1,1 1,1
35	Rango	0,0 - 63,8 0,0 - 66,4
Porcentaje de fracción de ploidía		
	Media	3,4 3,5
	Mediana	2,6 2,4
40	Rango	0,0 - 20,0 0,0 - 20,0

Veinte (15 %) pacientes experimentaron recurrencia del 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 del PSA. Las diecisiete características clínicas se seleccionaron como predictivas de la recurrencia del 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 la biopsia (112), raza (112), etapa clínica 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 del 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, lumen, 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), ancho mínimo 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 el lumen, longitud del borde para el citoplasma, densidad para el estroma (por ejemplo, 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 morfológicas de compacidad, ancho, valor del canal verde y radio del recinto más pequeño para los núcleos epiteliales (por ejemplo, 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 relación 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 morfológica 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 del lumen y, por lo tanto, sería consistente con un mayor grado de Gleason (es decir, mayor progresión de la enfermedad). Además, la característica morfológica del radio del recinto más pequeño del lumen se relaciona con el tamaño general del lumen, 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 (por ejemplo, más redonda con menos variaciones) y tamaño (por ejemplo, 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 morfológicas 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 del PSA / BCR.

Como se indicó anteriormente, se comprenderá que al menos algunas de las características morfológicas particulares determinadas por las enseñanzas proporcionadas en el presente documento 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 morfológicas. El programa informático Definens Cellenger y las características morfológicas particulares medidas por el programa informático descrito en el presente documento 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 ± DE	23,9 ± 31,38	9,8 ± 21,32	2,4 ± 4,64	25,3 ± 32,50	10,3 ± 21,51	2,6 ± 3,29	1,8 ± 9,96	0,0 ± 0,36	0,1 ± 0,63
Mediana	4,7	0,0	0,0	4,8	0,0	0,0	0,0	0,0	0,0
Rango	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 ± DE	NA	NA	100,0 ± 0,00	NA	NA	100,0 ± 0,00	NA	NA	100,0 ± 0,00
Mediana	NA	NA	100,0	NA	NA	100,0	NA	NA	100,0
Rango	NA	NA	100,0 - 100,0	NA	NA	100,0 - 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
Rango	NA	NA	0,0 - 0,4	NA	NA	0,0 - 0,1	NA	NA	0,0 - 0,0

Marcador	Tumor			PIN			Glándula		
	1+	2+	3+	1+	2+	3+	1+	2+	3+
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
Rango	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
Rango	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
AR			
Rango	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
Rango	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
Rango	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
Rango	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
Rango	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
Rango	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
Rango	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 del 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 atroficas (para AR) y el tumor / PIN (para CD34) sugieren una importancia biológica y funcional que afecta al resultado clínico de la recurrencia del 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. Mediante el uso de la experiencia de los dominios y el procedimiento de selección de características específicas de 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 la biopsia: los grados de Gleason resumidos (dominantes y secundarios) que se asignan a las múltiples muestras de tejido de biopsia con 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 (por ejemplo, afroamericana, caucásica, etc.)

3. Etapa 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 ganglios linfáticos y "M" representa la metástasis en un sitio distante.

4. Resultado de ploidía: contenido de ADN, 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 consiste en 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 colorante que se une al material genético. La concentración y distribución de este colorante (tinción de Fielgen) 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 mediante el uso de técnicas estandarizadas, y que se puede usar para realizar una verificación rápida de la exactitud 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 (por ejemplo, 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 ganglios linfáticos: una medida del grado en que los ganglios linfáticos contienen células tumorales (por ejemplo, células epiteliales del cáncer de próstata), que puede evaluarse mediante inspección clínica / quirúrgica o en el momento de una prostatectomía.

7. Grado de Gleason dominante de la biopsia: véase la descripción anterior de la puntuación de Gleason de la 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.
- 5 10. Etapa TNM: tumor, ganglio y metástasis según los criterios UICC después de la prostatectomía y según el examen patológico de muestras de tejido.
11. Grado de Gleason dominante postoperatorio: el grado de Gleason dominante que representa la característica histológica más predominante presente en la muestra de prostatectomía.
12. Edad
- 10 13. Afectación de las vesículas seminales: invasión de la vesícula seminal por el tumor.
14. PSA preoperatorio: nivel de PSA observado antes de la cirugía
15. Porcentaje de fracción de ploidía: véase la descripción anterior del resultado de ploidía.
16. Afectación de márgenes quirúrgicos: 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.
- 15 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.
- 25 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 fCD45 en asociación con PIN.
6. CD34 - tumor/estroma: La localización de los vasos CD34 que están asociados con el tumor.
- 30 7. Ki-67 - tumor 3: véase anteriormente.
8. p27 - tumor: La identificación de p27 en los núcleos de las células epiteliales tumorales.
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 lumbales de la unidad de la glándula.
- 35 12. PSMA - PIN: La identificación de PSMA en las células glandulares / lumbales 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.
- 40 Como cada dominio de datos se analizó durante este proceso mediante el uso de SVRc, la exactitud predictiva de los modelos aumentó. Mediante el uso de 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 morfológicas, alcanzó un índice de concordancia de 0,84. Cada uno de estos modelos validados internamente también se validó externamente (como se describió
- 45 anteriormente en Materiales y Métodos) con la misma tendencia observada. Mediante el uso de NNci en el conjunto final de características seleccionadas, el índice de concordancia alcanzó 0,88.

El resultado producto de los modelos NNci y SVRc puede interpretarse como una estimación del riesgo relativo de recurrencia del PSA para un paciente individual. Mediante el uso de 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 de acuerdo con el modelo NNci se presentan en la Figura 8. Los grupos mostraron una diferencia estadísticamente significativa en el tiempo hasta la recurrencia del PSA (prueba del 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, *supra* para los detalles de todas las características clínico-patológicas analizadas para esta cohorte). Treinta y ocho (14 %) pacientes experimentaron recurrencia del 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 del PSA. Se seleccionaron tres características clínicas como predictivas de la recurrencia del PSA: etapa clínica TNM, márgenes quirúrgicos y ganglios linfáticos.

- 5 Análisis de imágenes y estudios de morfometría. Mediante el uso de una versión actualizada del programa informático de análisis de imágenes, pero analizando los mismos portaobjetos teñidos con H&E, se generaron un total de 350 características morfométricas (que se muestran en la Tabla 2, anteriormente).

10 La Figura 9 muestra que, de las 350 características, se seleccionaron 6 características morfométricas como predictivas de la recurrencia del PSA, donde estas características morfométricas se relacionaron con los objetos patológicos de núcleos epiteliales, estroma, citoplasma, glóbulos rojos y lumen (es decir, EpithelialNucleiMinCompactne0215, StromaMaxStddevChannel30569, CytoplasmStddevMaxDiff0148, RedBloodCellMeanAreaPxI0386, RedBloodCellStddevAreaPxI0388 y LumenMinAsymmetry0295). Más particularmente, en este estudio, las

15 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 del lumen se seleccionaron como correlacionadas con el resultado.

20 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 del lumen proporcionarían 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 del PSA después de la cirugía. Además, la

25 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.

30 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

- 35 1. Etapa TNM
2. Afectación de márgenes quirúrgicos
3. Afectación de ganglios linfáticos

Característica molecular

- 40 1. Índice de tinción de AR (tumor)

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 de Etapa Clínica TNM, Márgenes Quirúrgicos, EpithelialNucleiMinCompactne0215, Ganglios Linfáticos y StromaMaxStddevChannel30569. El CI de un modelo

45 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.

50 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)

55

Marcador	Tumor			PIN			Glándula		
	1+	2+	3+	1+	2+	3+	1+	2+	3+
Ki-67	22,1 ±	7,3 ±	1,9 ±	23,2 ±	7,9 ±		1,3 ±	1,2 ±	
Media ± DE	30,30	17,04	4,01	31,36	18,16	2,0 ± 4,46	7,96	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
Rango	0,0 -	0,0 -	0,0 -	0,0 -	0,0 -	0,0-39,5	0,0 -	0,0 -	0,0 - 13,0
	100,0	100,0	26,3	100,0	100,0		96,0	96,5	

60

65

Marcador	Tumor			PIN			Glándula		
	1+	2+	3+	1+	2+	3+	1+	2+	3+
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
Rango	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
Rango	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
Rango	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
Rango	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
Rango	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
Rango	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
Rango	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
Rango	0 - 300	0 - 300	0 - 300

	Marcador	Tumor	PIN	Glándula
5	PSMA			
	Media $\pm$ DE	0,2 $\pm$ 2,09	6,4 $\pm$ 21,02	2,9 $\pm$ 22,94
	Mediana	0	0	0
	Rango	0 - 21	0 - 154	0 - 300
10	p27			
	Media $\pm$ DE	3,9 $\pm$ 8,20	6,4 $\pm$ 18,83	1,3 $\pm$ 8,65
	Mediana	0	0	0
	Rango	0 - 48	0 - 140	0 - 120
15	Her-2/neu			
	Media $\pm$ DE	3,4 $\pm$ 16,69	0,2 $\pm$ 1,12	0,0 $\pm$ 0,00
	Mediana	0	0	0
20	Rango	0 - 150	0 - 10	0 - 0

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

25 Estudios analíticos y estadísticos. Mediante el uso de la experiencia de los dominios y remuestreo simple, el método encontró un subconjunto de 10 características (3 clínico-patológicas, 6 morfométricas y 1 molecular) que tenían un índice de concordancia (CI) de 0,87 (la Figura 9 muestra la lista completa de las características seleccionadas). El resultado producto del modelo SVRc también se puede interpretar como una estimación del riesgo relativo de recurrencia del PSA para un paciente individual. Mediante el uso de 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 de SVRc se presentan en la Figura 10. Los grupos mostraron una diferencia estadísticamente significativa en el tiempo hasta la recurrencia del PSA (prueba del orden logarítmico, valor $p < 0,0001$).

35 Estudio 3. Este estudio usó la misma cohorte que la del Estudio 2 para que las características clínico-patológicas 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 del tiempo total hasta la muerte. Se seleccionaron dos características clínicas como predictivas de muerte por cualquier causa: etapa clínica TNM y edad del paciente.

45 Análisis de imágenes y estudios de morfometría. En este estudio se usó 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, RedBloodCellStddevStddevChann20472, StromaMaxMaxDiff0529, EpitheNucleMeanBordeLengtPx10206, EpithelialNucleiMeanAreaPx10194, EpithelNucleiStddevElliptFit0228, RedBloodCellStddevStddevChann30476, y RedBloodCellStddevElliptFit0420, 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 se correlacionan 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.

60 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 usó 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 atrofica.

Estudios analíticos y estadísticos. En esta cohorte, se seleccionaron un total de 14 características (2 clínico-patológicas, 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. Etapa 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 producto del modelo SVRc también se puede interpretar como una estimación del riesgo relativo de muerte para un paciente individual. Mediante el uso de 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. Mediante el uso de la prueba del 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 (combinadas) del Estudio 1 (41) al Estudio 2 (10) mientras se conserva la exactitud predictiva del modelo enfatizó la precisión y los atributos de filtrado que se lograron a través de diferentes métodos de aprendizaje automático. El índice de concordancia del modelo que se desarrolló en la cohorte de 268 pacientes fue de 0,87; en comparación, cuando se aplica el nomograma Kattan [20] a esta cohorte, 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 del 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.

En el Estudio 3, el objetivo fue utilizar el conocimiento existente de los dominios derivado del Estudio 2 y desarrollar un modelo predictivo para la supervivencia general. El resultado final exitoso fue la capacidad de predecir con un 80 % de exactitud la supervivencia general y el tiempo hasta la muerte de un individuo utilizando un total de 14 características de dominio combinadas. 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.

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 para tener un impacto en las cuestiones de tratamiento antes de la cirugía.

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. Mediante el uso de la experiencia de los dominios y la regresión de vectores de soporte para datos censurados (SVRc), se seleccionaron características de los tres dominios y se usaron para desarrollar un modelo predictivo para la recurrencia del PSA y la supervivencia general. Se comprenderá 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

Este estudio se realizó para predecir una enfermedad agresiva (es decir, un fallo clínico como lo demuestra una exploración ósea positiva 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 exactas para proporcionar tal predicción. Como se describió anteriormente, se ha demostrado que el enfoque de patología de sistemas de la presente invención predice con exactitud la recurrencia del PSA. Este estudio demuestra que la presente invención también se puede usar para predecir con exactitud la metástasis ósea distante después de la prostatectomía.

Se estudió una cohorte de 119 pacientes que se sometieron a prostatectomía radical incorporando micromatrices de tejido (TMA) construidas a partir de muestras de prostatectomía. Se realizaron estudios morfométricos (es decir, análisis de imágenes) mediante el uso de cortes de tejido teñidos con hematoxilina y eosina (H&E), 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 se evaluaron con una máquina de vectores de soporte para la regresión desarrollada para manejar datos censurados (SVRc). El rendimiento predictivo del modelo se estimó mediante el uso del índice de concordancia (CI) con puntuaciones generadas usadas para definir los grupos de riesgo.

De los 116 pacientes, se seleccionó un subconjunto de 61 pacientes en base a sus características clínicas, incluidos 20 individuos con fallo clínico identificados por metástasis ósea. Esta cohorte se usó 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 exactitud 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 lumen. Más particularmente, las características morfométricas seleccionadas fueron el área de citoplasma dividida por el área de tejido total, el área de lumen 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

1. Extensión Extracapsular (ECE)
2. Invasión de las vesículas seminales (SVI)
3. Grado de Gleason dominante de la prostatectomía (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 exacto 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 exactitud 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 la adición de las características moleculares a las características clínicas y morfométricas del modelo están actualmente pendientes.

Ejemplo 4: Toxicología hepática

Datos morfométricos

Este estudio se realizó para demostrar la capacidad del análisis de imágenes y de la modelación 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 exactitud de clasificación podría permitir la creación de una plataforma de alta productividad usada para cribar objetivamente las toxicidades en estudios preclínicos.

El estudio se dividió en dos fases. La fase inicial usó 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 usó para desarrollar una aplicación de análisis de imágenes mediante el uso del sistema de análisis de imágenes de tejido 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 analizaron los modelos estadísticos diseñados en la fase de entrenamiento.

Exactitud de la segmentación

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

5 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 usadas para la clasificación puede mejorar la solidez y la fiabilidad de la clasificación de los cortes. El segundo paso consistió en entrenar un modelo mediante el uso del conjunto de características y etiquetas seleccionadas para cada corte (anormal, normal) y luego analizar el modelo prediciendo la clasificación de un conjunto independiente de cortes de hígado de rata donde las etiquetas se desconocían.

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)
- 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.

La selección de características también implicó dos pasos. El primer paso utilizó la experiencia de los dominios. 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 mediante el uso del conocimiento de los dominios.

Estas características se examinaron mediante el uso de un análisis discriminante gradual 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 significación 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 usó 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 introducirla, el proceso de selección gradual se detuvo.

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. El resultado del modelo se corrigió con respecto a un sesgo potencial a través de validación cruzada.

Las redes neuronales también se exploraron como clasificador. Las características seleccionadas se usaron 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 exactitud que un modelo MLP entrenado mediante los criterios típicos, *por ejemplo*, error cuadrático medio y entropía cruzada.

El resultado de ambos modelos se usó para crear una curva característica operativa del receptor (ROC) eligiendo un valor diferente del resultado 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 (AUC) ROC usa ambas medidas para evaluar la exactitud 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 de verdaderos positivos y la tasa de verdaderos negativos, 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 de verdaderos positivos, es decir, la proporción de cortes de hígado clasificados correctamente como anormales; la especificidad, por otro lado, es la tasa de verdaderos negativos, es decir, la proporción de cortes de hígado clasificados correctamente 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

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 mediante el uso de un punto de corte individual estimado mediante el uso de los resultados de validación cruzada de cada uno de los resultados del modelo, respectivamente. Los puntos de corte correspondieron a una sensibilidad de 100 % y una especificidad de 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 realizó la evaluación de las exactitudes de los modelos sin ocultación sobre la verdadera clasificación de los cortes de hígado. Esta otra parte también proporcionó la clave de análisis para verificar los resultados.

Resultados

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 mediante el uso de LDA tiene un AUC de 0,99; la función derivada mediante el uso de 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

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

$$\text{Sensibilidad} = TP/(TP+FN) \times 100 = 42/(42+7) \times 100 = (42/49) \times 100 = \mathbf{86 \%}$$

$$\text{Especificidad} = TN/(FP+TN) \times 100 = 32/(19+32) \times 100 = (32/51) \times 100 = \mathbf{63 \%}$$

Tabla 11b

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

$$\text{Sensibilidad} = TP/(TP+FN) \times 100 = 36/(36+13) \times 100 = (36/49) \times 100 = \mathbf{73 \%}$$

$$\text{Especificidad} = TN/(FP+TN) \times 100 = 32/(19+32) \times 100 = (32/51) \times 100 = \mathbf{63 \%}$$

El punto de corte usado para el clasificador LDA fue igual a 0,0031; el punto de corte usado 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

Según 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, a diferencia del 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 previamente, pero esto no es sorprendente, ya que generalizar cualquier clasificador a un conjunto externo a menudo conduce a una caída en su exactitud. Este estudio demostró la aplicación eficaz de las tecnologías para la imagenología y la modelación estadística.

Ejemplo 5: Predicción de la recurrencia del cáncer de próstata

Datos clínicos, morfométricos y moleculares

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 del PSA (bioquímica). Los cortes de tejido prostático usados en este estudio estaban compuestos predominantemente de tumor, pero también incluían elementos benignos.

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 17 (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. En este estudio se evaluó la información clínica, molecular y morfométrica de 262 pacientes de la cohorte de 539 pacientes descrita anteriormente en relación con el Ejemplo 2. 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 usó un método de reducción de características por SVRC para la selección de características. La reducción de características por SVRC se describe en la solicitud de patente de EE.UU. de propiedad compartida nº 11/438,789, presentada el 22 de mayo de 2006, que se incorpora en el presente documento como referencia en su totalidad.

Resultados

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 del PSA fueron la afectación de las vesículas seminales (contribución de la característica = -5,2103), afectación de márgenes quirúrgicos (-7,3159) y afectación de 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 las 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, núcleos epiteliales, lumen, glóbulos rojos, núcleos estromales, estroma y artefactos. Las posibles razones para la

selección de las características morfométricas relacionadas con los núcleos epiteliales y las células estromales 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 producto del modelo SVRc también se puede interpretar como una estimación del riesgo relativo de recurrencia del PSA para un paciente individual. Mediante el uso de 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 del PSA (prueba del orden logarítmico, valor $p < 0,0001$).

Estudio de validación

El modelo final se validó con una cohorte externa que consistía en 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.

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 las 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 exactitud predictiva del modelo.

Ejemplo 6: Predicción de la recurrencia del cáncer de próstata

Datos clínicos, morfométricos y moleculares

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. Al igual que en los Ejemplos 2 y 5, el tiempo hasta la recurrencia se definió como el tiempo (en meses) desde la prostatectomía radical hasta la recurrencia del PSA (bioquímica). Los cortes de tejido prostático usados en este estudio estaban compuestos predominantemente de tumor, pero también incluían elementos benignos.

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 18 (adjunta en el presente documento). La información clínica, molecular y morfométrica para los mismos 262 pacientes del Ejemplo 5 se evaluó en este estudio. Este estudio usó el mismo método de reducción de características por SVRc mencionado anteriormente en el Ejemplo 5.

Resultados

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 morfométrica 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 la biopsia (-10,60), descrita anteriormente en relación con el Ejemplo 2. La característica morfométrica adicional seleccionada en este estudio fue la variación en la 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 base a sus propiedades de tinción, probablemente refleja los atributos bioquímicos del estroma asociado con el tumor en lugar de los elementos benignos.

Estudio de validación

El modelo final se validó con una cohorte externa que consistía en 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
5	N	262	61
	Edad (años)		
	Media	62	61
10	Mediana	63	62
	Rango	38-81	42-74
	Raza		
	Caucásico	235 (89,7 %)	58 (95,1 %)
15	Afroamericano (hispano y no hispano)	21 (8,0 %)	2 (3,3 %)
	Otro/Desconocido	6 (2,3 %)	1 (1,6 %)
	PSA preoperatorio (ng/ml)		
20	Media	10,7	12,9
	Mediana	7,8	10,0
	Rango	0,9-81,9	2,0-69,5
	Etapa patológica TNM		
25	T2N0	158 (60,3 %)	No recopilado
	T3aN0	70 (26,7 %)	No recopilado
	T3bN0	17 (6,5 %)	No recopilado
	T1-3N+	17 (6,5 %)	No recopilado
30	Ausente	0	No recopilado
	Etapa UICC		
	T1a < 5 %	1 (0,4 %)	No recopilado
	T1b ≥ 5 %	1 (0,4 %)	No recopilado
35	T1c no palpable ni visible	113 (43,1 %)	No recopilado
	T2a ≤ ½ lóbulo	54 (20,7 %)	No recopilado
	T2b ≤ 1 lóbulo	43 (16,4 %)	No recopilado
	T2c ambos lóbulos	33 (12,6 %)	No recopilado
40	T3a ECE unilateral	15 (5,7 %)	No recopilado
	T3c SV+	2 (0,8 %)	No recopilado
	Ausente	0	No recopilado
45	Resultado del tacto rectal		
	No palpable	122 (46,6 %)	32 (52,5 %)
	Palpable	140 (53,4 %)	29 (47,5 %)
	Ausente	0	0
50	Afectación de ganglios linfáticos		
	Negativo	246 (93,9 %)	56 (91,8 %)
	Positivo	16 (6,1 %)	5 (8,2 %)
	Ausente	0	0
55	Afectación de las vesículas seminales		
	No	233 (88,9 %)	51 (83,6 %)
	Sí	29 (11,1 %)	10 (16,4 %)
	Ausente	0	0
60	2	1 (0,4 %)	0 (0,0 %)
	3	0 (0,0 %)	0 (0,0 %)
	4	7 (2,7 %)	0 (0,0 %)
65	5	56 (21,4 %)	3 (4,9 %)
	6	97 (37,0 %)	27 (44,3 %)

	Característica	Entrenamiento	Validación
5	7	85 (32,4 %)	20 (32,8 %)
	8	13 (5,0 %)	8 (13,1 %)
	9	3 (1,2 %)	3 (4,9 %)
	Ausente	0	0
10	Grado de Gleason dominante de la muestra	20 (7,6 %)	0 (0,0 %)
	2	198 (75,6 %)	34 (55,7 %)
	3	44 (16,8 %)	23 (37,7 %)
	4	0 (0,0 %)	4 (6,6%)
15	5		
	Puntuación de Gleason de la muestra		
	5	21 (8,0 %)	1 (1,6 %)
20	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 %)
25	Ploidía		
	Diploide	141 (53,8 %)	No recopilado
	Tetraploide	113 (43,1 %)	No recopilado
	Aneuploide	8 (3,1 %)	No recopilado
30	Ausente	0 (0,0 %)	No recopilado
	Porcentaje de ploidía en fase S (%)		
	Media	2,4	No recopilado
	Mediana	1,2	No recopilado
35	Porcentaje de ploidía en fase S (%)		
	Rango	0,0 - 66,4	No recopilado
	Porcentaje de fracción de ploidía		
40	Media	3,4	No recopilado
	Mediana	2,4	No recopilado
	Rango	0,0-20,0	No recopilado
45			
50			
55			
60			
65			

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
Rango	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
Rango	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
Rango	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
Rango	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 ± DE	0,0 ± 0,04	0,0 ± 0,11	0,1 ± 0,18	0,0 ± 0,07	0,0 ± 0,08	0,0 ± 0,04
Mediana	0,0	0,0	0,0	0,0	0,0	0,0
Rango	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 ± DE	171,8 ± 75,9	79,9 ± 83,3	29,5 ± 67,9
Mediana	200	66,0	0
Rango	0-300	0-300	0-300
CK14			
Media ± DE	2,2 ± 6,4	35,2 ± 62,0	8,3 ± 32,6
Mediana	0	0	0
Rango	0-69	0-300	0-300
Ciclina D1			
Media ± DE	1,4 ± 7,1	0,0 ± 0,21	0,0 ± 0,0
Mediana	0	0	0
Rango	0-90	0-3	0-0
PSA			
Media ± DE	117,9 ± 71,2	140,5 ± 97,4	22,4 ± 54,9
Mediana	100	134	0
Rango	0-300	0-300	0-300
PSMA			
Media ± DE	0,3 ± 2,1	5,8 ± 19,8	3,0 ± 23,2
Mediana	0	0	0
Rango	0-21	0-154	0-300
p27Kipl			
Media ± DE	3,9 ± 8,2	6,6 ± 19,0	1,3 ± 8,7
Mediana	0	0	0
Rango	0-48	0-140	0-120
Her-2/neu			
Media ± DE	3,5 ± 16,9	0,2 ± 1,1	0,0 ± 0,0
Mediana	0	0	0
Rango	0-150	0-10	0-0

Ejemplo 7: Predicción del fallo clínico

posterior a la prostatectomía radical

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 (por ejemplo, exploración ósea positiva 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 etiquetados 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 'quintuplicado'. 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 lumbinales, 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 las 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 base a 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 (por ejemplo, intensidad promedio, desviación estándar, cuantiles). Cada secuencia de comandos/función lineal fue específica para un biomarcador concreto, y se diseñó mediante el uso del 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 mediante el uso de colocalizaciones 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 usada 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 sus muestras de punción faltaban en la matriz o si la muestra contenía estroma solamente.

Debido a que la mayoría de los pacientes tuvieron múltiples muestras de punción de las cuales se extrajeron las características de IF, se ideó un procedimiento para agregar los valores de las características en múltiples muestras de punción, para cada paciente y característica. Se consideraron cuatro funciones candidatas (mín, máx, mediana y media) (por ejemplo, si un paciente tuvo 3 muestras de punción, la función mín devuelve el valor más bajo de las 3 muestras de punción). Para una característica dada, cada una de estas funciones se aplicó para agregar los valores de las muestras de punción 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 a continuación 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.

Tabla 14.

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

Tabla 15

Característica	Descripción
5 EpitarposamaMeanMeanChannel30017	Intensidad media de los objetos DAPI clasificados como Núcleos Epiteliales con expresión positiva de biomarcador AR y ubicados en el citoplasma negativos con respecto al biomarcador AMACR
10 EpitarposamaMeanMeanChannel30021	Intensidad media de los objetos DAPI clasificados como Núcleos Epiteliales expresados negativamente con biomarcador AR y ubicados en el citoplasma positivos con respecto al biomarcador AMACR
15 Mamacrpostotalarea0045	Área total de los objetos AMACR con la intensidad por encima del umbral de ruido/señal
20 Mcytoplasmamacrpostotalarea0050	Área total de los objetos de citoplasma (CK18) con expresión positiva de AMACR
25 SumMen02_10totalarea	En cada objeto DAPI clasificado como núcleo epitelial y con expresión positiva del 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 relación de la intensidad 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 $M = 11$ grupos. Cada grupo caracteriza una cierta cantidad del AR en los núcleos epiteliales. 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
30 Mepithnucleiarposamacrpostot0085	Área total de los núcleos epiteliales con expresión positiva de AR y positivos con respecto a AMACR
40 RelAEpithNucARpos	Área relativa de los núcleos epiteliales con AR+ con respecto al área total de núcleos epiteliales
45 RelAEpithNucARpos	Fracción (área relativa) de los núcleos epiteliales AR+ con AMACR+ con respecto a los núcleos epiteliales AR+
50 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)
55 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 positivos con respecto al marcador AMACR.
60 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 a través de los valores de luminancia.
65	

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 $CI \geq 0,6$).

5 Resultados

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 la biopsia, la Afectación de ganglios linfáticos y la Puntuación de Gleason de la muestra (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 (EpiNucleiStddevMeanChannel40157). 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 lugar de 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 nucléolos), el desarrollo del cáncer típicamente se caracteriza por un mayor número de núcleos epiteliales con textura de cromatina desplegada, así como un mayor número de nucléolos, 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 para AMACR.

25 Validación

El modelo final se validó con una cohorte independiente que consistía en 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.

Tabla 16. Información clínica

	Característica	Entrenamiento	Validación
35	Raza		
	Blanco (hispano y no hispano)	328 (95,1)	301 (94,4)
	Afroamericano (hispano y no hispano)	11 (3,2)	9 (2,8)
40	Otro/Desconocido	6 (1,7)	9 (2,8)
	PSA preoperatorio (ng/ml)		
	Media	11,4	11,1
	Mediana	7,9	8,2
45	Rango	0,5-100,0	1,1-56,2
	Etapas clínicas TNM		
	T1a/b	4 (1,2)	4 (1,3)
	T1c	172 (49,9)	148 (46,4)
50	T2a	64 (18,6)	59 (18,5)
	T2b	29 (8,4)	26 (8,2)
	T2c	67 (19,4)	77 (24,1)
55	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)
60	Afectación de las vesículas seminales		
	No	316 (91,6)	291 (91,2)
	Sí	29 (8,4)	28 (8,8)
65	Márgenes quirúrgicos		
	Negativo	228 (66,1)	199 (62,4)
	Positivo	117 (33,9)	120 (37,6)

Característica	Entrenamiento	Validación
Afectación extracapsular		
No	242 (70,1)	222 (69,6)
Sí	103 (29,9)	97 (30,4)
Grado de Gleason dominante de la biopsia		
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)
Puntuación de Gleason de la 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)
Grado de Gleason dominante de la muestra		
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)
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)

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.

Ejemplo 8: Predicción de la recurrencia del cáncer de próstata

Datos clínicos, morfométricos y moleculares

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 en los Ejemplos 2, 5 y 6, el tiempo hasta la recurrencia se definió como el tiempo (en meses) desde la prostatectomía radical hasta la recurrencia del PSA (bioquímica).

Se evaluó una cohorte ampliada de 682 pacientes. Este estudio usó características clínicas comparables a las usadas en el Ejemplo 6 (véase la Tabla 19). 101 pacientes (14,8 %) experimentaron recurrencia del PSA, definida como dos mediciones consecutivas de PSA > 0,2 ng/ml. La mediana de seguimiento de los pacientes sin recurrencia observada fue de 74 meses. No se alcanzó la mediana del tiempo total hasta la recurrencia. Se evaluaron las mismas características morfométricas que se muestran en la Tabla 18, aunque en este estudio se usó Adobe Imageready 7.0 para delinear digitalmente y enmascarar con color las regiones de tumor infiltrante solamente dentro de las muestras de punción individuales para la segmentación de imágenes posterior. Las características moleculares se generaron

mediante inmunofluorescencia (IF) múltiple cuantitativa, generalmente como se describió anteriormente en relación con el Ejemplo 7. Detalles adicionales con respecto a la cuantificación de IF se proporcionan a continuación.

El modelo nuevamente predijo la recurrencia del PSA con alta exactitud, el índice de concordancia fue de 0,77, la sensibilidad de 72 % y la especificidad de 74 %. Como se muestra en la Figura 17, se seleccionó una característica molecular correspondiente al área relativa de núcleos epiteliales AR+. También se seleccionaron cinco características clínicas, que incluyen invasión de las vesículas seminales, puntuación de Gleason de la biopsia, extensión extracapsular, PSA preoperatorio y Grado de Gleason dominante de la prostatectomía, así como dos características morfológicas, textura de núcleos epiteliales y textura de citoplasma en regiones solo tumorales. En la Figura 17, un peso negativo indica que la presencia de cada característica (o un valor más alto de una característica continua) se asoció con un tiempo más corto hasta la recurrencia del PSA, mientras que un peso positivo indica lo contrario. Estos pesos ilustran la contribución respectiva de cada característica en el modelo completo.

Las características morfológicas seleccionadas para su inclusión en el modelo describen las propiedades de textura de las regiones de núcleos epiteliales y citoplasma mediante el uso de variaciones de color (desviación estándar de intensidad). Los procesos biológicos asociados con la progresión del cáncer, tales como la textura de cromatina desplegada y la apariencia de los nucléolos en los núcleos epiteliales, dan como resultado cambios en la textura de los objetos de imagen del núcleo. Esto último aumenta el valor de la desviación estándar del color. Las propiedades de textura de las áreas de citoplasma se deben a cambios estructurales en el tejido y a la invasión de las células epiteliales al estroma cuando el tejido progresa de etapas benignas a malignas.

Inmunofluorescencia (IF) cuantitativa

Las características cuantificadas por el análisis de imágenes inmunofluorescentes en este estudio se muestran y describen en las Tablas 14 y 15 (véase el Ejemplo 7). Es de destacar que una característica tal como la intensidad de AR se asocia con la concentración del antígeno y se expresó mediante el uso de un valor limitado de RGB-(rojo, verde, azul) de 12 bits, en lugar de 8 bits para (0-255), permitiendo así un rango dinámico expandido. Entre las mediciones de AR, una característica, la cantidad de AR en los núcleos de las células epiteliales (que incluyen elementos tumorales y no tumorales) en relación con el área total de DAPI de todos los núcleos epiteliales, fue muy concordante con el índice de tinción IHC para AR en el tumor (Coeficiente de correlación de rangos de Spearman=0,44; p=0,0011). Esta característica también se asoció independientemente con la recurrencia del PSA cuando se analizó de forma univariante. Por lo tanto, los datos de inmunofluorescencia cuantitativa respaldan los resultados de los biomarcadores IHC de AR y proporcionan un medio más objetivo para asignar un valor a la expresión de AR en los compartimentos celulares.

Las imágenes tiff en escala de grises de alta resolución (1280 x 1024 píxeles) se procesaron individualmente con un método orientado a objetos que identificó el nivel de intensidad para AR dentro de los núcleos epiteliales y estromales en las muestras de punción de TMA. Los elementos tumorales no se distinguieron de los elementos normales; más bien, solo se distinguieron las células estromales y epiteliales. La expresión de CK18 sirvió como marcador topográfico para todas las células epiteliales de próstata. Las capas 'morfológicas' individuales se superpusieron, es decir, DAPI para núcleos, CK18 para células epiteliales y AR, y, mediante el empleo de fluorocromos Alexa discretos y umbrales preestablecidos para eliminar el fondo y la unión inespecífica, la intensidad de AR se definió con exactitud. A continuación se proporciona una descripción adicional con respecto al establecimiento de umbrales para la localización y cuantificación automatizadas de múltiples proteínas a través de imágenes de fluorescencia múltiples. El análisis equiparó la concentración de proteínas con recuentos de píxeles positivos absolutos con características individuales que reflejan una serie de correlaciones morfológicas, que incluyen el área nuclear (teñida con DAPI) ocupada por AR y la cantidad relativa de proteína AR presente dentro de los núcleos epiteliales y estromales. Las células epiteliales positivas para AMACR se usaron como un descriptor adicional para el contenido de AR y las características derivadas se evaluaron como parte del desarrollo del modelo general.

Las Figuras 18a-e ilustran la segmentación de imágenes múltiples para cuantificar el AR nuclear en tejido prostático. La Figura 18a es una imagen estándar de múltiples fluorocromos que se deconstruyó para producir 3 imágenes en escala de grises (Figuras 18b-d) que representan mapas de AR y morfológicos (es decir, DAPI, CK18) discretos. La Figura 18e es una imagen segmentada que representa el AR nuclear dentro de las células epiteliales cuantificadas por la secuencia de comandos. Todas estas figuras se muestran en blanco y negro para su reproducibilidad. Más específicamente, la Figura 18a es una imagen inmunofluorescente múltiple de tejido prostático teñido con DAPI para núcleos (no se muestra), CK18 etiquetado con Alexa 488 para células epiteliales (típicamente se muestra en verde) y AR etiquetado con Alexa 568 en células estromales y epiteliales (típicamente se muestra en rojo). Mediante la aplicación de óptica espectral para separar los fluorocromos respectivos, se crearon imágenes individuales en escala de grises para núcleos (Figura 18b), células epiteliales (Figura 18c) y AR (Figura 18d). Mediante el uso de métodos basados en la clasificación de píxeles y objetos, se reprodujo una imagen a color combinada (Figura 18e; mostrada aquí en blanco y negro), que segmenta el AR presente dentro de los núcleos teñidos con DAPI que están dentro de las células epiteliales positivas para CK18, así como el AR que está presente dentro de los núcleos que son negativos para CK18 (estroma). La imagen segmentada fue la base para cuantificar el AR en los núcleos epiteliales, así como el AR dentro del estroma. Los parámetros calculados incluyeron el área nuclear general involucrada con AR, la intensidad media y máxima, y la distribución (% de células positivas).

Selección de características

En el estudio ampliado del Ejemplo 8, se empleó un método de SVRc de selección de características por remuestreo. Un paso de filtrado inicial eliminó las características que no se correlacionaban de forma univariante con el resultado de interés. A continuación, se hicieron N divisiones diferentes (en este estudio, N=25) de los datos de entrenamiento; en cada división, aproximadamente dos tercios del total de instancias de entrenamiento se asignaron aleatoriamente a un subconjunto de entrenamiento y aproximadamente un tercio del total de instancias de entrenamiento se asignó aleatoriamente a un subconjunto de análisis.

El método comenzó con un proceso de selección de características "voraz directo" que comienza con todas las características que pasaron el filtro inicial. Para cada característica, se construyeron N modelos en los subconjuntos de entrenamiento en todas las divisiones y se validaron en los N subconjuntos de análisis respectivos. El rendimiento general de cada característica se promedió en las N ejecuciones. Se seleccionó la característica con el mejor rendimiento general. En el siguiente paso, cada característica se añadió a la(s) característica(s) seleccionada(s) previamente y nuevamente se construyeron y analizaron N modelos en las divisiones. Se seleccionó la característica cuya adición dio como resultado el mejor rendimiento general. El método continuó de esta manera hasta que no hubo más funciones que mejoraran el rendimiento.

Posteriormente, se empleó un enfoque de selección de características "voraz inverso". Se eliminó cada característica y se construyeron y analizaron N modelos sin esa característica en las divisiones. La característica cuya eliminación dio como resultado el mejor rendimiento general se eliminó, y el procedimiento se repitió hasta que el rendimiento del modelo dejó de mejorar debido a la eliminación de las características. Este paso redujo la complejidad del modelo y eliminó las características que inicialmente pudieron haber sido significativas, pero su contribución de información se incluía dentro de una característica añadida posteriormente.

Finalmente, el modelo de SVRc completo se entrenó mediante el uso de todas las características seleccionadas en los datos de entrenamiento completos.

En otras realizaciones, se pueden proporcionar modelos para predecir la probabilidad de progresión del cáncer de próstata a metástasis óseas y/o aumento del PSA después de la terapia de privación de andrógenos, ya sea en el momento de la biopsia diagnóstica con aguja o después de la prostatectomía radical.

Ejemplo 9: Predicción del fallo clínico

posterior a la prostatectomía radical

Datos clínicos, morfométricos y moleculares

Se realizó otro estudio para generar un modelo para predecir el fallo clínico (CF) en pacientes con cáncer de próstata que se han sometido a prostatectomía radical. Al igual que con el Ejemplo 7, el fallo clínico se definió como el desarrollo de enfermedad metastásica y/o enfermedad independiente de andrógenos.

Se evaluó un conjunto de entrenamiento de 373 pacientes con 32 eventos de fallo clínico para generar el modelo a través del aprendizaje automático supervisado multivariante (SVRc con selección de características por remuestreo). El modelo tenía un CI de 0,92, sensibilidad de 90 % y especificidad de 91 % para predecir el fallo clínico dentro de los 5 años posteriores a la prostatectomía. El modelo se validó en una cohorte independiente de 385 pacientes con 29 eventos de fallo clínico, y produjo un CI de 0,84, sensibilidad de 84 % y especificidad de 85 %. Los altos niveles de receptor de andrógenos predijeron un tiempo más corto hasta el aumento del PSA en la castración después de la terapia de privación de andrógenos.

Como se muestra en la Figura 19, el modelo incluye las características clínicas de Grado de Gleason dominante de la prostatectomía y el estado de invasión de los ganglios linfáticos, y las características morfométricas de la longitud promedio del perímetro del lumen, el área relativa del lumen y la variación de la textura dentro del citoplasma en el canal verde. El modelo también incluye la característica molecular de la intensidad promedio del canal de AR en los núcleos epiteliales AR+/AMACR, que se generó mediante IF múltiple cuantitativa. En la Figura 19, un peso negativo indica que la presencia de cada característica (o un valor más alto de una característica continua) se asoció con un tiempo más corto hasta el fallo clínico, mientras que un peso positivo indica lo contrario. Estos pesos ilustran la contribución respectiva de cada variable en el modelo completo.

En el estudio, se excluyeron los pacientes que recibieron tratamiento antes de la prostatectomía o inmediatamente después, pero antes de la recurrencia bioquímica, lo que dejó a 881 pacientes en la cohorte completa. La cohorte se aleatorizó y se dividió equitativamente entre conjuntos de entrenamiento y de validación con números similares de eventos de fallo clínico. El fallo clínico (CF) se especificó previamente como evidencia radiográfica o patológica inequívoca de metástasis (que incluye enfermedad esquelética o enfermedad de tejidos blandos en ganglios linfáticos u órganos sólidos) o un aumento del PSA en estado de castración o muerte atribuida al cáncer de próstata. El tiempo

hasta CF se definió como el tiempo desde la prostatectomía radical hasta el primero de estos eventos. Si un paciente no experimentó un CF en su última visita, o el resultado del paciente en el momento de su visita más reciente era desconocido, entonces el resultado del paciente se consideró censurado.

Solo se estudiaron más a fondo los pacientes con datos clínicos, morfométricos y moleculares completos, así como sin falta de información de resultados. Como resultado, el número de pacientes evaluables fue de 373 en el conjunto de entrenamiento y 385 en el conjunto de validación (véanse las Tablas 20 y 21). Las características de estos 758 pacientes fueron similares a las de la cohorte completa de 881 pacientes. Las características morfométricas se extrajeron de imágenes tomadas de muestras de punción de pacientes, generalmente como se describió en relación con el Ejemplo 8 (por ejemplo, enmascaramiento digital de muestras de punción con una herramienta de etiquetado (Adobe 7.0) para incluir solo elementos tumorales para la cuantificación). De manera similar, las características moleculares se generaron mediante el uso de IF cuantitativa múltiple, generalmente como se describió en relación con el Ejemplo 8.

En el conjunto de entrenamiento de 373 pacientes, 33 (8 %) tenían CF después de la prostatectomía: 24 con una exploración ósea positiva antes de la posible terapia adyuvante, y 9 con un aumento del PSA en la castración. Todos menos uno de los pacientes con CF habían recibido terapia hormonal. El único paciente que no recibió terapia hormonal se sometió a radioterapia de rescate. Además, hubo un paciente que recibió radioterapia y terapia hormonal. A estos 373 pacientes se les realizó un seguimiento durante una mediana de 76 meses después de la prostatectomía; no se alcanzó la mediana del tiempo total hasta CF. Se incluyeron nueve características clínicas como variables para la selección por el modelo (Tabla 20). En la cohorte de validación que consistía en 385 pacientes con una mediana de seguimiento de 72 meses, veintinueve (7,5 %) pacientes tenían CF, 22 con una exploración ósea positiva antes de la posible terapia adyuvante y 7 con un aumento del PSA en la castración. Todos los pacientes con CF habían recibido terapia hormonal y 7 también recibieron radioterapia de rescate. El modelo de 6 características se aplicó con los datos de resultado ocultos.

A partir de las imágenes digitalizadas enmascaradas del tumor de las muestras de punción de tumor teñidas con H&E de cada paciente, se generaron 27 características morfométricas (Tabla 22), que reflejan el color y la textura de la tinción, así como el área total de las células epiteliales y estromales en la muestra de tejido prostático. En el análisis univariante, 12 características mostraron una asociación con CF en base al índice de concordancia de cada característica ($IC \leq 0,4$ o $IC \geq 0,6$) (Tabla 20).

AR y AMACR en tipos de células específicas (células tumorales epiteliales) se cuantificaron para generar características moleculares mediante la utilización de imágenes espectrales combinadas con la evaluación simultánea de múltiples antígenos y el análisis de imágenes. De las características de inmunofluorescencia relacionadas con los niveles de AR y AMACR, once de las características de inmunofluorescencia mostraron asociación con CF en el análisis univariante ($CI \leq 0,4$ o $CI \geq 0,6$) y se incluyeron en el modelo de selección (Tabla 20).

Con respecto al modelo final, cabe destacar el papel identificado para la diferenciación tumoral (es decir, el grado de Gleason) y la evidencia de metástasis (es decir, la afectación de ganglios linfáticos) en la predicción de la progresión del cáncer de próstata. Las tres características morfométricas del modelo ilustran la importancia de la tinción histoquímica como sustituto de las propiedades bioquímicas o metabólicas dentro de los diferentes tipos de células y la arquitectura tisular general asociada con el grado de Gleason. Los patólogos han reconocido tradicionalmente la importancia del color al distinguir los tipos de células benignas de las reactivas o malignas, pero no han podido estandarizarlo ni cuantificarlo objetivamente. Con el modelo actual, las propiedades de color de la textura del citoplasma epitelial se asociaron más fuertemente con CF que el grado de Gleason de la prostatectomía, lo que sugiere la importancia de las propiedades bioquímicas celulares en la progresión de la enfermedad. Las curvas individuales de Kaplan-Meier para las tres características de imagen ilustraron la capacidad de estratificar con exactitud a los pacientes. Con respecto a las características moleculares, solo se seleccionó el brillo/intensidad promedio normalizado de AR dentro de las células epiteliales negativas para AMACR para su inclusión en el modelo. El peso negativo para esta característica (Figura 19) indica que una cantidad creciente de AR se asoció con un tiempo más corto hasta CF. Es de destacar que la información del PSA preoperatorio y de la suma de Gleason no se seleccionaron para su inclusión en el modelo. A través del análisis de SVRc, estas variables se suplantaron con características de análisis de imágenes que reflejan las características de diferenciación y color vinculadas a las propiedades bioquímicas del cáncer de próstata.

Se cree que las predicciones muy exactas de este modelo de CF permitirán el desarrollo de planes de tratamiento más informados y apropiados. Esto incluiría la posibilidad de administrar ADT, que actualmente es la primera línea de terapia sistémica contra el cáncer de próstata avanzado, radioterapia y/o quimioterapia antes, donde existe un alto riesgo de recurrencia. Además, la identificación temprana de pacientes con alto riesgo crea la oportunidad de aumentar la vigilancia de la progresión. Las predicciones del modelo también pueden ser útiles para guiar las decisiones de tratamiento después de la recurrencia bioquímica. Es posible que algunos pacientes con recurrencia bioquímica no requieran intervención inmediata, ya que el intervalo entre la recurrencia bioquímica y los eventos clínicamente significativos es típicamente de años y también bastante variable. Además, es probable que muchos hombres diagnosticados hoy con recurrencia bioquímica tengan un curso de enfermedad de progresión lenta. El conocimiento de la probabilidad de enfermedad agresiva, por ejemplo, enfermedad que progresará a CF dentro de los 5 años

posteriores a la cirugía, puede ayudar al paciente y a su médico a decidir si la recurrencia bioquímica debería desencadenar una terapia agresiva.

Dada la importancia del AR en este modelo de progresión clínica, se realizó un estudio para determinar si el contenido de AR en la muestra de prostatectomía podría usarse para predecir el tiempo hasta el aumento del PSA en la castración después de la terapia de privación de andrógenos (ADT). Sesenta y tres de los 881 pacientes habían recibido ADT sin CF previo. Treinta y dos de estos pacientes (51 %) posteriormente tuvieron CF. Mediante el uso de los datos de resultados de los 63 pacientes, se investigó si alguna de las características de inmunofluorescencia asociadas con CF también se asoció con la respuesta a ADT. Las siguientes cuatro características se asociaron significativamente ($p < 0,05$ prueba de chi cuadrado) con un tiempo reducido hasta el aumento del PSA en la castración, todas ellas relacionadas con un aumento de los niveles nucleares de AR en las células epiteliales tumorales:

	Característica	Valor P
15	Intensidad de AR en células Epi AMACR(-)	0,0021
	Intensidad de AR en células Epi AMACR(+)	0,0003
	Intensidad media de AR en células Epi AMACR(+)	0,0285
20	Intensidad total de AR en todas las células Epi (normalizada)	0,018

Mediante el uso del contenido de AR de la muestra tumoral, el urólogo puede predecir la utilidad de la ADT e incluso para un paciente individual y ajustar los regímenes terapéuticos en consecuencia. Los pacientes con puntuaciones altas en el modelo actual de CF y altos niveles de AR enfrentan un dilema particular: una probabilidad elevada de progresión clínica acompañada de una disminución de la probabilidad de respuesta duradera a la ADT y posiblemente a la radioterapia de rescate. Estos pacientes pueden ser buenos candidatos para terapias, tales como los inhibidores de histona desacetilasa, que se dirigen al AR o sus cofactores. Por lo tanto, se cree que la cuantificación exacta del receptor de andrógenos ayudará a guiar la toma de decisiones terapéuticas. Tales enfoques podrían incluso mejorar la selección para ensayos clínicos y potencialmente servir como medidas del efecto del tratamiento, específicamente para terapias, tales como los inhibidores de histona desacetilasa, que se dirigen al AR o a cofactores de AR.

Localización y cuantificación automatizadas de múltiples proteínas

a través de imágenes de fluorescencia multiespectrales

Se ha desarrollado un sistema y un método para la localización y cuantificación automatizadas de la expresión de biomarcadores proteicos en imágenes microscópicas de inmunofluorescencia (IF). El sistema y el método discriminan la señal de biomarcador del fondo, donde la señal puede ser la expresión de cualquiera de los muchos biomarcadores o contraindicaciones usados en IF. El sistema y el método se basan en el aprendizaje supervisado y representan el umbral de intensidad del biomarcador en función de las características del fondo de la imagen. La utilidad del sistema y el método se demuestran al predecir la recurrencia del cáncer de próstata en pacientes sometidos a prostatectomía. Específicamente, se muestra que las características que representan la expresión del receptor de andrógenos (AR) se asocian de manera estadísticamente significativa con un mal resultado en el análisis univariante. Las características de AR también muestran ser valiosas para la predicción de la recurrencia multivariante.

Introducción

Los biomarcadores proteicos se usan ampliamente en histopatología para el diagnóstico, pronóstico y predicción de la respuesta terapéutica del cáncer. Estos proporcionan información sobre los niveles de expresión de proteínas en las células, lo que permite la detección de actividades celulares particulares asociadas con el estado de la enfermedad.

Un método para la cuantificación de la expresión de proteínas es IF, donde una proteína se localiza mediante la introducción de un anticuerpo etiquetado con un colorante fluorescente en el tejido que se une a la proteína objetivo [21]. El portaobjetos teñido se ilumina bajo un microscopio de fluorescencia con una fuente de luz que tiene una longitud de onda específica. El colorante fluorescente absorbe esta luz de excitación lo que hace que emita luz de una longitud de onda más larga. La intensidad de la luz emitida es una medida de la concentración de la proteína objetivo. En el análisis múltiple, el tejido se etiqueta con varios anticuerpos al mismo tiempo. Los anticuerpos se etiquetan con colorantes fluorescentes que tienen características espectrales distintas. La separación de múltiples biomarcadores se logra a través de imágenes multiespectrales del tejido seguido de separación de los espectros para obtener imágenes que representan la expresión de anticuerpos individuales.

Convencionalmente, los patólogos interpretan las imágenes de IF en base a sus niveles de intensidad percibidos de los objetos de interés (por ejemplo, núcleos) en el tejido. Esta práctica es laboriosa y sufre variaciones entre observadores y del mismo observador. El desarrollo de sistemas automatizados permite una alternativa objetiva y económica a la puntuación visual de las imágenes de IF.

La cuantificación de un biomarcador se logra en dos etapas. Primero, se detecta un compartimento relevante para el biomarcador. Luego, la señal se separa del fondo dentro del compartimento. Estas tareas a menudo se logran a través de la umbralización de la intensidad. Los umbrales interactivos [22]-[24] a menudo van seguidos de mediciones por ordenador para cuantificar la expresión de biomarcadores. Gordon y *otros* [23] identificaron proteínas en imágenes de células individuales como conjuntos de píxeles contiguos por encima de un umbral de tres desviaciones estándar del fondo. El modo del histograma de la imagen junto con la umbralización interactiva se usa para identificar los píxeles de fondo en el sistema AQUA [24]. Rao y *otros* [25] utilizan la delineación manual de núcleos epiteliales para marcar el compartimento de interés para la cuantificación. La reproducibilidad limitada de los métodos interactivos y el poco rendimiento de la umbralización con histograma en imágenes con relaciones señal/fondo bajas son las principales desventajas de las técnicas existentes.

Lo que se propone es un sistema y un método para la localización y cuantificación de biomarcadores que abordan las limitaciones anteriores. El sistema y el método están completamente automatizados y, por lo tanto, son reproducibles. El sistema y el método cuantifican una señal de biomarcador dentro de su compartimento subcelular relevante, lo que hace que el sistema y el método sean robustos frente a relaciones bajas de señal respecto al fondo.

El sistema y el método se basan en un enfoque para la discriminación de la señal de biomarcador del fondo. El aprendizaje supervisado se usa para modelar el umbral de intensidad para la discriminación de señales en función de las características del fondo de la imagen. La utilidad del sistema y el método propuestos se demuestra en el pronóstico del cáncer de próstata, aunque se comprenderá que el sistema y el método tienen muchos usos adicionales.

Análisis de imágenes múltiples

El sistema y el método propuestos se describen en relación con el ensayo de IF múltiple descrito anteriormente, que incluye la contratinción nuclear con 4'-6-diamidino-2-fenilindol (DAPI) junto con citoqueratina 18 (CK18), α -metilacil-CoA-racemasa (AMACR) y receptor de andrógenos (AR).

Plataforma de análisis de imágenes

El sistema de cuantificación se diseñó mediante el uso del paquete informático Image Intelligence Suite™ de Definiens Enterprise [26]. La segmentación de imágenes en objetos biológicos válidos (por ejemplo, núcleos) es un proceso de varias etapas basado en la *segmentación de imágenes orientada a objetos*. En este paradigma, los objetos en lugar de los píxeles son las unidades más pequeñas en las que se realizan las operaciones de procesamiento de imágenes y los cálculos de las características. Por ejemplo, cuando se aplica un umbral de intensidad a una imagen, las intensidades de los objetos se someten al umbral. La intensidad de un objeto es la intensidad promedio de todos los píxeles que pertenecen a ese objeto.

Se utilizan dos métodos para obtener las primitivas [26]. El método de segmentación multirresolución encuentra las primitivas en base a la similitud de color entre píxeles y la regularidad de la forma de los objetos. El método de segmentación de árbol cuaternario usa la similitud de color solamente. Un parámetro de escala controla el tamaño de los objetos en ambos métodos.

Localización y cuantificación de biomarcadores

En el enfoque desarrollado, el primer paso es la colocación espacial del biomarcador dentro del compartimento subcelular correspondiente. Por ejemplo, los núcleos epiteliales y el citoplasma son compartimentos subcelulares para AR y AMACR, respectivamente. El siguiente paso es discriminar la señal de biomarcador verdadera del fondo en las primitivas de los compartimentos correspondientes. El fondo consiste en autofluorescencia y unión inespecífica del colorante fluorescente al tejido. Los objetos del compartimento subcelular se clasifican como positivos con respecto al biomarcador si contienen una cierta cantidad de la señal verdadera. La cuantificación de los biomarcadores es el paso final del análisis donde se miden las propiedades (por ejemplo, el área) de los objetos clasificados.

Modelo de umbralización de biomarcadores

La siguiente etapa del análisis de imágenes IF es la discriminación de la señal de biomarcador verdadera del fondo mediante el uso de un umbral de intensidad. El umbral T se considera una función de las características del fondo del biomarcador. Para encontrar esta función, los objetos B se dividen en dos clases no superpuestas: colocados en un compartimento B_c y objetos B_b fuera del compartimento. Los objetos B_b se consideran el fondo, los cuales se describen mediante las siguientes características: media μ_b , desviación estándar σ_b y α -ésimo percentiles $q_{b,\alpha}$ de las intensidades de los objetos, donde $\alpha = 5, 10, \dots, 95$. Suponiendo una relación lineal entre el umbral T y las características del fondo, tenemos

$$T = \beta^T \mathbf{X}, \quad (1)$$

donde $\beta = [\beta_0, \beta_1, \dots, \beta_p]^T$ β son parámetros del modelo y $\mathbf{X} = [1, X_1, \dots, X_p]^T$ son características del fondo. Tenga en cuenta que en (1), T cambia de acuerdo con las propiedades de cada imagen.

Se usa regresión lineal multivariante para determinar β . Se selecciona un conjunto de imágenes para entrenar al modelo. Para cada imagen: a) se identifican los objetos B_c y B_b ; b) se extraen las características del fondo; y c) el patólogo determina el umbral de intensidad experto T . El conjunto de entrenamiento se usa para entrenar el modelo para predecir los umbrales expertos.

Para evitar el sobreajuste del modelo, se excluyen del entrenamiento las características redundantes y estadísticamente no significativas. Una de las dos características es redundante si su coeficiente de correlación por pares r satisface $r \geq 0,85$. La característica que tiene la correlación univariante más fuerte con los umbrales expertos se usa para la regresión multivariante. Se calculan los intervalos de confianza para los coeficientes de regresión y los coeficientes cuyos intervalos de confianza contienen cero se consideran estadísticamente no significativos y se eliminan. Los parámetros β se calculan con las características restantes.

Compartimentos subcelulares para la colocalización

Los compartimentos tisulares subcelulares básicos contienen señal verdadera que se origina a partir de los antígenos objetivo AR y AMACR. Los núcleos epiteliales y el citoplasma constituyen los compartimentos básicos mencionados anteriormente. Estos son objetos positivos segmentados de las imágenes de DAPI y CK18, respectivamente. El siguiente paso es dividir los objetos de núcleos en células epiteliales y estromales superponiéndolos con el citoplasma.

Es importante en la predicción de la recurrencia del cáncer distinguir los núcleos epiteliales ubicados en las áreas de cáncer invasivo. Estos últimos se reconocen como áreas de citoplasma que contienen expresión de AMACR. Esos núcleos epiteliales se denominan AMACR+, a diferencia de los AMACR-.

Los objetos primitivos *iniciales* se obtienen a través de la segmentación de árbol cuaternario del color de las imágenes de CK18 y AMACR. Estos objetos se vuelven a segmentar en objetos más gruesos mediante el uso de la segmentación multirresolución. Luego, se aplica el umbral (3) de la Tabla 23 para identificar los objetos de *citoplasma* o *CK18+* iniciales. Tenga en cuenta que la desviación estándar de la intensidad $b \sigma$ en (3) se calcula respecto a todos los objetos CK18 iniciales. Una vez que se identifican los objetos CK18+ iniciales, se usa el análisis de vecindad para refinar las etiquetas de clase. Los objetos de fondo pequeños cuya longitud del borde con el citoplasma en relación con su longitud total del borde es 0,6 o más, se reclasifican como citoplasma. Los objetos de citoplasma pequeños aislados se etiquetan como fondo.

Los objetos primitivos se superponen con la imagen de DAPI. El umbral (2) de la Tabla 23 se usa para clasificar las primitivas en objetos DAPI+ y de fondo. Esta clasificación se refina aún más mediante un análisis de vecindad similar al descrito anteriormente. Los objetos DAPI+ son pequeños fragmentos de núcleos reales representados en la imagen de DAPI.

Las primitivas DAPI+ se fusionan en núcleos mediante el uso de un método iterativo que consiste en el crecimiento de regiones y la clasificación de objetos. Finalmente, los objetos de núcleos se clasifican en núcleos epiteliales y estromales en base a su colocalización con objetos CK18+.

Expresiones de biomarcadores positivos y negativos

Las áreas de citoplasma AMACR+ se encuentran en dos pasos: a) superponer los objetos primitivos en la imagen de AMACR con las áreas CK18+; b) aplicar el umbral (5) en la Tabla 23 a los objetos colocalizados. Los objetos de citoplasma cuya intensidad excede a (5) se clasifican como AMACR+. Los núcleos epiteliales se etiquetan adicionalmente como AMACR+ y AMACR- dependiendo de su asociación con citoplasma positivo y negativo, respectivamente.

Las células epiteliales son los compartimentos subcelulares básicos para la expresión de AR. El modelo de umbral (4) en la Tabla 23 se aplica a objetos AR colocalizados en los núcleos epiteliales. Un núcleo epitelial se clasifica como AR positivo (AR+), si el área de sus objetos AR+ es más del 0,1 % de su área total. De lo contrario, se clasifica como AR negativo (AR-).

Cuantificación de los biomarcadores

Una vez que los biomarcadores se colocalizan y clasifican, a partir de los biomarcadores colocalizados se generan características cuantitativas que representan las características de intensidad y área basadas en objetos. Las características de intensidad representativas incluyen la intensidad de AR media y total dentro de las células epiteliales positivas/negativas para AMACR reconocidas en una imagen. Las características de área representativas incluyen el área de células epiteliales AR+ en relación con el área de todas las células epiteliales y la fracción de las células epiteliales AMACR+ entre todas las células AR+.

Resultados

La utilidad del sistema propuesto para el análisis de imágenes de IF se demostró al predecir la recurrencia del cáncer de próstata para pacientes que se habían sometido a prostatectomía. Se consideraron dos tipos de resultados adversos para la predicción. Un resultado fue la recurrencia del antígeno prostático específico (PSA) (PSAR) correspondiente a un aumento significativo en los niveles séricos de PSA del paciente. El otro resultado fue el fallo clínico (CF) correspondiente a una progresión significativa de la enfermedad medida por criterios tales como la metástasis a distancia.

Se obtuvieron muestras de punción de micromatrices de tejido con al menos 80 % de su área de tejido cubierta con tumor para 682 y 758 pacientes en las tareas de predicción de PSAR y CF, respectivamente. Las muestras de punción se etiquetaron con la contratinción con DAPI y los biomarcadores CK18, AR y AMACR, y se obtuvieron imágenes mediante el uso del sistema de imágenes multiespectrales CRI Nuance™ [27]. Para cada muestra de punción, se adquirió una sola imagen en escala de grises de 1280 X 1024 píxeles de 12 bits a la longitud de onda del pico de emisión de la contratinción con DAPI. Para cada uno de los biomarcadores CK18, AR y AMACR, se adquirió una pila de imágenes para un rango de longitudes de onda que abarca el espectro de emisión del colorante fluorescente correspondiente. Las pilas de imágenes resultantes se separaron mediante el uso del sistema Nuance™ para obtener tres imágenes, cada una correspondiente a uno de los biomarcadores.

En un conjunto de 60 imágenes múltiples del conjunto de entrenamiento, un patólogo certificado seleccionó los umbrales expertos para la contratinción con DAPI y los biomarcadores CK18, AR y AMACR y los modelos de umbralización se entrenaron como se describió anteriormente y se muestra en la Tabla 23. Los modelos se validaron mediante un examen visual de las imágenes con las mayores discrepancias entre sus umbrales expertos y predichos. Los biomarcadores AR y AMACR se cuantificaron en las imágenes y se obtuvieron las características que representan el área de los núcleos epiteliales AR+ en relación con el área total de los núcleos epiteliales, así como la intensidad de los núcleos epiteliales AR+. Características similares también se calcularon para núcleos epiteliales AR+ AMACR+ y AR+ AMACR-.

Los pacientes disponibles se agruparon en conjuntos de entrenamiento y validación de 342 y 340 pacientes para la predicción de PSAR y 373 y 385 pacientes para la predicción de CF. En la predicción de PSAR, el área relativa de los núcleos epiteliales AR+ mostró el índice de concordancia (CI) univariante más alto (entre todas las características de AR) de 0,37 (valor $p < 0,001$) en el conjunto de entrenamiento. En la predicción de CF, la intensidad de los núcleos epiteliales AR+ AMACR+ alcanzó el CI univariante más alto de 0,30 (valor $p < 0,001$).

Además, se entrenaron modelos multivariantes para las dos tareas de predicción en el contexto del paradigma de *patología de sistemas*. En este paradigma, la información dispar de los perfiles clínicos (por ejemplo, edad), histológicos (a través de características de imágenes extraídas de muestras de tejido teñidas con hematoxilina y eosina [28]) y moleculares (a través de características que miden la expresión de biomarcadores proteicos) del paciente se combinan en un marco de aprendizaje supervisado para predecir la recurrencia del cáncer. El modelo de PSAR resultante se describió anteriormente en el Ejemplo 8. El modelo de CF resultante se describió anteriormente en el Ejemplo 9.

Por consiguiente, en el presente documento se proporciona un sistema y un método para la localización y cuantificación automatizadas de la expresión de cualquier biomarcador proteico en imágenes microscópicas múltiples de IF. La señal de biomarcador verdadera se discrimina del fondo al abordar la umbralización de la imagen del biomarcador como un problema de aprendizaje supervisado, modelando el umbral en función de las características del fondo de la imagen. La aplicación de este sistema se demuestra, por ejemplo, para el componente múltiple de IF de un ensayo de pronóstico del cáncer de próstata. Se demostró que las características que representan el área y la intensidad de la expresión de AR tienen una correlación univariante significativa. Además, las características novedosas, reproducibles y robustas se seleccionaron como informativas en competencia con otros factores predictivos en un ambiente multivariante, para validar de esta manera la utilidad del sistema en aplicaciones prácticas.

Tabla 23. Modelos de umbralización para los biomarcadores múltiples

Biomarcador	Modelo	Eq.
DAPI	$T = 0,86\sigma_b + 1,1q_{b,5} - 32,0$	(2)
CK18	$T = 0,55 \sigma_b + 220$	(3)
AR	$T = 5,1q_{b,5} + 0,39q_{b,95} + 1100$	(4)
AMACR	$T = 2,1\mu_b - 0,82q_{b,25} + 1800$	(5)

Ejemplo 10: Predicción de la supervivencia de pacientes con cáncer de pulmón de células no pequeñas tratados con Gefitinib; datos clínicos, morfométricos y moleculares

Se realizó otro estudio para generar un modelo que predice la supervivencia de individuos con cáncer de pulmón de células no pequeñas (NSCLC) que se han tratado con gefitinib. Antes de la presente invención, no existían

herramientas analíticas exactas para proporcionar tal predicción. Como se describió anteriormente, se ha demostrado que el enfoque de patología de sistemas de la presente invención predice con exactitud la recurrencia del PSA y el fallo clínico después de la prostatectomía. Este estudio demuestra que la presente invención también puede usarse para predecir con exactitud la supervivencia después del tratamiento con gefitinib.

La abundante expresión del receptor del factor de crecimiento epidérmico (EGFR) en una variedad de tumores sólidos, que incluyen el cáncer de pulmón de células no pequeñas (NSCLC), lo ha convertido en un objetivo atractivo para la terapia molecular selectiva, específicamente los inhibidores de tirosina quinasa tales como el gefitinib (IRESSA). Aunque los resultados iniciales de gefitinib en pacientes tratados previamente fueron prometedores, en dos ensayos esenciales de Fase III de gefitinib en combinación con otros dos agentes, los pacientes con NSCLC tratados con gefitinib no demostraron una supervivencia general o sin progresión significativamente mejor en comparación con el grupo placebo. Además, en el ensayo de Fase III de evaluación de la supervivencia con IRESSA en cáncer de pulmón (ISEL) controlado con placebo, la monoterapia con gefitinib se asoció con cierta mejora en la supervivencia general; sin embargo, los datos no alcanzaron significación estadística. Las respuestas positivas a gefitinib se han asociado con mutaciones activadoras en el bolsillo de ATP de EGFR, al igual que características clínicas y demográficas tales como raza (asiática), género (femenino), antecedentes de tabaquismo (no fumador), buen estado funcional y patología tumoral (variante adenocarcinoma-broncoalveolar). Estos resultados han sugerido que subgrupos específicos de pacientes con NSCLC tienen más probabilidades de responder a gefitinib. Sin embargo, los datos se han complicado por una baja frecuencia de mutación de EGFR (específicamente en poblaciones caucásicas; EE.UU. ~10-15 %), lo que contrasta con la observación de que aproximadamente 45-50 % de los pacientes tratados con gefitinib (en estudios limitados) experimentaron alguna forma de beneficio clínico. Otro desafío es la falta de datos comparativos sobre mutaciones de EGFR entre pacientes asignados aleatoriamente a gefitinib frente a ningún tratamiento y frente a quimioterapia. Además, la tasa de respuesta parece depender mucho de las características demográficas del grupo en estudio (por ejemplo, las relaciones de género, el origen étnico y los antecedentes detallados de tabaquismo), lo que sugiere que los criterios de respuesta son más complicados de lo que se pensaba previamente. De importancia, la evidencia reciente del ensayo ISEL controlado con placebo, sugiere que un alto número de copias del gen de EGFR fue predictivo de beneficio clínico y supervivencia. Estos hallazgos ilustran parte de la complejidad observada en los intentos de comprender los criterios de respuesta.

Se evaluó una cohorte de 109 pacientes. Específicamente, se evaluaron muestras tumorales de diagnóstico de NSCLC refractario avanzado tratado con gefitinib. Las muestras tumorales se evaluaron mediante el análisis de mutaciones del ADN de EGFR, inmunohistoquímica de EGFR, morfometría histológica e inmunofluorescencia cuantitativa para 15 marcadores. Se utilizó un modelo matemático de regresión de vectores de soporte para integrar seis características clínicas (género, antecedentes de tabaquismo, edad al momento del diagnóstico, histología tumoral y estado funcional de ECOG) con histología y análisis cuantitativo de múltiples biomarcadores.

Pacientes y tejidos

La cohorte inicial consistió en 284 pacientes estadounidenses con NSCLC refractario avanzado tratados con 250 mg de gefitinib por vía oral cada día. Se analizaron cinco características clínicas: género, antecedentes de tabaquismo, edad al momento del diagnóstico, histología, estado funcional de ECOG, (la escala varía de 0 (saludable) a 5 (muerte por enfermedad)). (Tabla 24). Los portaobjetos sin teñir (que incluyen aspirados con aguja fina, sedimentos de células y citocentrifugaciones) y/o los bloques de parafina de la muestra de diagnóstico se evaluaron con hematoxilina y eosina (H&E) para determinar el contenido del tumor. Todos los biomarcadores se analizaron sin conocimiento del resultado clínico.

Análisis de mutaciones del dominio tirosina quinasa de EGFR

Se analizaron dos cortes secuenciales de 20 mm de grosor de cada bloque de parafina o un mínimo de 8 cortes sin teñir de portaobjetos de parafina. Se obtuvo el ADN genómico de muestras de tejido desparafinado, después de la incubación con proteinasa K y posterior extracción con cloroformo y precipitación con etanol. Las mutaciones de EGFR se analizaron en primer lugar mediante secuenciación del ADN de los exones 19, 20 y 21, y en segundo lugar mediante el uso del sistema de mutación refractario a la amplificación (ARMS), específicamente, la reacción en cadena de la polimerasa (PCR) específica de alelo para detectar la mutación puntual L858R del exón 21 y la delección más común del exón 19 (del G2235-A2249). Los pacientes se consideraron positivos para la mutación si se detectaba una mutación en el dominio tirosina quinasa mediante ARMS o secuenciación de genes tanto en dirección directa como inversa en al menos dos productos de PCR independientes del ADN tumoral.

Morfometría histológica

Los portaobjetos teñidos con H&E se prepararon a partir de los bloques originales o cortes sin teñir. Se adquirieron de una a seis imágenes de áreas representativas de tumor con un microscopio de campo brillante Olympus con un aumento de 20X mediante el uso de una cámara digital en color SPOT Insight QE (KAI2000). El programa informático de análisis de imágenes clasificó los objetos de imagen como elementos celulares histopatológicos que exhiben valores de canal de color rojo, verde y azul, características de forma genéricas (por ejemplo, área, longitud) y propiedades de relación espacial (por ejemplo, cantidades relativas de lumen con respecto al tejido total), de las cuales

se generaron estadísticas. Debido a las diferencias en la preparación de la muestra (es decir, citocentrifugación, biopsia con aguja y resección de tejido), la fijación, tinción y calidad del tejido, se desarrollaron varias secuencias de comandos diferentes para la segmentación de imágenes.

5 Inmunohistoquímica de EGFR

La proteína EGFR se analizó por inmunohistoquímica mediante el uso del kit de EGFR pharmDX (DAKO, Glostrup, Dinamarca). Las muestras se consideraron positivas si >10 % de las células tumorales tenían tinción. Se calculó un índice de tinción para cada muestra mediante la multiplicación de cada nivel de intensidad (0-3) por el porcentaje de células a ese nivel de intensidad; por lo tanto, el índice varía de 0 a 300.

Evaluación de múltiples biomarcadores (M-Plex™)

Se seleccionaron quince anticuerpos que incluyen: Citoqueratina 18, Ki67, Caspasa-3 activada, CD34, EGFR fosforilado, ERK fosforilada, AKT fosforilada, PTEN, ciclina D1, mTOR fosforilado, PI3-K, VEGF, VEGFR-2 y VEGFR-2 fosforilado y se agruparon en seis formatos múltiples. Cada anticuerpo representaba una evaluación de los componentes morfológicos (por ejemplo, citoqueratina 18 [CK18], CD34), vías de señalización o procesos biológicos de proliferación, apoptosis y angiogénesis.

Cada anticuerpo se evaluó en una serie de líneas celulares y/o muestras de tejido de cáncer de pulmón de control por inmunohistoquímica (IHC). Cada anticuerpo se tituló mediante el uso de procedimientos operativos estándar inmunohistoquímicos e inmunofluorescentes. Los anticuerpos progresaron a través de formatos inmunofluorescentes simples y múltiples. Los 15 anticuerpos se organizaron en 6 formatos múltiples basados en sistemas de recuperación de antígeno, isotipo y priorización con respecto a la aplicación biológica. Estos formatos de multiplicidad m se enumeran en la Tabla 27.

Tabla 27. Anticuerpos y agrupaciones de inmunofluorescencia múltiple.

Grupo de multiplicidad m	Anticuerpo	Vendedor	Cat. #	Isotipo	Dilución	Etiqueta fluorescente
1	Citoqueratina 18	Vector	VC-C414	IgG 1 de ratón	1:250	1° - 488
	Ciclina D1	BioCare Med.	CP236B	IgG de conejo	1:100	1° - 568
2	Citoqueratina 18	Vector	VC-C414	IgG1 de ratón	1:250	1° - 488
	Ki-67	Ventana	790-2910	IgG 1 de ratón	1:75 diluido previamente	de 2° - 555
	PI3 quinasa	Cell Signaling	3821	IgG de conejo	1:20	1° - 568
	Caspasa-3 activada	Chemicon	ab3623	IgG de conejo	1:5	2° - 594
3	Citoqueratina 18	Vector	VC-C414	IgG1 de ratón	1:200	1° - 488
	CD-34	Dako	M7165	IgG1 de ratón	1:50	2° - 555
	p-mTOR	Cell Signaling	ab2971	IgG de conejo	1:50	1° - 568
	pERK	Cell Signaling	ab4376	IgG de conejo	1:5	2° - 594

	Grupo de multiplicidad m	Anticuerpo	Vendedor	Cat. #	Isotipo	Dilución	Etiqueta fluorescente
5	4	Citoqueratina 18	Vector	VC-C414	IgG1 de ratón	1:200	1° - 488
		pAKT S473	Abeam	ab4802	IgG de conejo	1:20	1° - 555
10		PTEN	NeoMarkers	MS1797	IgG1 de ratón	1:50	2° - 568
	5	Citoqueratina 18	Vector	VC-C414	IgG 1 de ratón	1:250	1° - 488
15		VEGF	Abeam	ab1316	IgG1 de ratón	1:50	2° - 555
		pKDR	Upstate	07-374	IgG de conejo	1:150	2° - 594
20	6	Citoqueratina 18	Vector	VC-C414	IgG 1 de ratón	1:10	2° - 488
		EGFR	Dako	K1492	IgG 1 de ratón	diluido previamente	
25		pEGFR (Y1068)	Abeam	ab5644	IgG de conejo	1:250	1° - 594
30	La desparafinación y rehidratación de las muestras de tejido se realizó según los procedimientos operativos estándar. La recuperación de antígeno se realizó hirviendo los portaobjetos en un horno microondas durante 7,5 minutos en solución de Revelado IX (BioCare Medical). Los portaobjetos se dejaron enfriar durante 20 minutos a temperatura ambiente y luego se lavaron dos veces durante 3 minutos en solución salina tamponada con fosfato (PBS).						
35	Las muestras de tejido se sometieron a los siguientes pasos de tratamiento de prehibridación. Para ayudar a penetrar las estructuras celulares del tejido, las muestras se incubaron en PBT (PBS + Tritón X-100 al 0,2 %) a temperatura ambiente durante 30 minutos, seguido de tres enjuagues de 3 minutos cada uno en PBS. Para ayudar a reducir la autofluorescencia de los tejidos, las muestras se incubaron en alcohol ácido (HCl al 1 % en etanol al 70 %) a temperatura ambiente durante 20 minutos, seguido de tres enjuagues de 3 minutos cada uno en PBS. El bloqueo de						
40	los sitios de unión inespecífica se realizó mediante la incubación de los portaobjetos en reactivo de bloqueo de IF (BSA a 0,5 mg/ml en PBS) a temperatura ambiente durante 20 minutos. No se realizaron lavados entre el paso de bloqueo y el paso de hibridación posterior.						
45	Se preparó un cóctel de un anticuerpo IgG1 de ratón y un anticuerpo IgG de conejo mediante el uso de los títulos determinados previamente en el reactivo de bloqueo de IF. Se aplicó aproximadamente 100 ml de este cóctel de anticuerpos a la muestra de tejido, y se permitió que los anticuerpos y las muestras de tejido hibridaran en una cámara húmeda a temperatura ambiente durante 1 hora. La hibridación fue seguida de dos enjuagues de 5 minutos cada uno en PBT seguido de dos enjuagues de 3 minutos cada uno en PBS.						
50	Para el paso de etiquetado, se preparó un cóctel de la etiqueta de IgG de conejo Alexa Flúor Zenon apropiada y de la etiqueta de IgG1 de ratón Alexa Flúor Zenon apropiada (Invitrogen / Molecular Probes) para los anticuerpos hibridados anteriores en el reactivo de bloqueo de IF a las concentraciones recomendadas por el fabricante (dilución 1:50 para cada fragmento Fab). Se aplicó aproximadamente 100 ml de este cóctel de etiquetado a las muestras de tejido, y las muestras de tejido se incubaron en una cámara húmeda a temperatura ambiente durante 30 minutos. La reacción de						
55	etiquetado fue seguida de dos enjuagues. Los pasos de hibridación y etiquetado como se describieron anteriormente se repitieron para cada grupo posterior de anticuerpos incluidos en el análisis múltiple. Fueron posibles hasta dos rondas más de hibridación y etiquetado de anticuerpos. Cabe señalar que para obtener la discriminación entre los diferentes anticuerpos, no se usó más de una vez la etiqueta Alexa Flúor Zenon, ya sea específica de conejo o ratón.						
60	Se realizó un paso de fijación mediante la incubación de las muestras en formalina al 10 % a temperatura ambiente durante 10 minutos, seguido de dos enjuagues de 3 minutos cada uno en PBS. Se aplicó a las muestras aproximadamente 25,0 ml de reactivo protector de la fluorescencia <i>SlowFade</i> con solución de montaje de DAPI, las que luego se cubrieron con un cubreobjetos. Las muestras se almacenaron a -20 °C hasta que se pudo realizar el análisis.						
65							

Se adquirieron imágenes fluorescentes múltiples mediante el uso de una cámara multispectral CRI Nuance (Cambridge Research and Instrumentation, Inc.) montada en un microscopio de fluorescencia automatizado Nikon 90i y controlada por el programa informático en línea MetaMorph. Las imágenes resultantes se guardaron como imágenes TIFF en escala de grises cuantitativa (1280 por 1024 píxeles). Para las regiones de interés seleccionadas, se registró la contratinción nuclear con DAPI a una longitud de onda de 480 nm mediante el uso de un filtro pasabanda de DAPI (Chroma). Alexa 488 se capturó entre 520 y 560 nm en intervalos de 10 nm mediante el uso de un filtro de FITC (Chroma). Alexa 555, 568 y 594 se registraron entre 570 y 670 nm en intervalos de 10 nm mediante el uso de un filtro de paso largo personalizado (Chroma), mientras que Alexa 647 se registró entre 640 y 720 nm en intervalos de 10 nm mediante el uso de un segundo filtro de paso largo personalizado (Chroma). Se asignaron regiones representativas para cada colorante para crear una biblioteca de espectros para el proceso de separación de los espectros.

Las secuencias de comandos de análisis de imágenes superponen los atributos morfológicos (es decir, DAPI para núcleos, CK18 para células epiteliales), y usan la firma de fluorocromo Alexa para diferenciar los anticuerpos individuales mientras eliminan el fondo para optimizar las relaciones de señal respecto a ruido. Los perfiles de intensidad generan características, que incluyen la media, el máximo y la desviación estándar para cada anticuerpo etiquetado. Las relaciones específicas de los marcadores individuales con sus formas activadas (por ejemplo, PKDR:KDR, pEGFR:EGFR) y la intensidad de un marcador específico en las células que expresan otro marcador (por ejemplo, mTOR fosforilado en células epiteliales positivas para pERK) se generan a través del programa informático de análisis de características por secuencia de comandos.

Control de calidad de los anticuerpos

Para confirmar la especificidad de los anticuerpos contra pEGFR(Y1068) y pERK, la línea celular de NSCLC A431 se trató con factor de crecimiento epidérmico (EGF). Las células se lavaron dos veces con PBS frío y luego se lisaron con tampón de lisis (Tritón X-100 al 0,5 %, Tris 50 mmol/l, pH 7,6, NaCl 300 mmol/l, Na₃VO₄ 2 mmol/l) que contiene fluoruro de fenilmetilsulfonilo (1 mmol/l), leupeptina (25 µg/ml), aprotinina (25 µg/ml) y pepstatina A (3,5 µg/ml) durante 25 minutos a 4 C. Siguiendo protocolos estándar para la inmunoprecipitación, se realizaron transferencias Western quimioluminiscentes IP ECL Plus con líneas celulares tanto activadas como no activadas con EGF. Se realizaron análisis similares con pKDR en líneas celulares HUVEC activadas con VEGF. Además, tanto para EGF como para VEGF, las células tratadas y no tratadas se procesaron para IHC de rutina.

Generación de características de inmunofluorescencia

Se desarrollaron secuencias de comandos de inmunofluorescencia para cada par anticuerpo:antígeno mediante el uso de muestras de tejido de cáncer de pulmón de control. Las secuencias de comandos se diseñaron de acuerdo con la distribución celular esperada de las proteínas individuales y utilizan la imagen en escala de grises adquirida para identificar tanto el área como la intensidad de fluorescencia posterior para cada marcador. Las secuencias de comandos preliminares generaron >400 características (muchas de ellas redundantes), que se redujeron a 84 características para la entrada en los modelos.

Análisis estadístico

El tiempo hasta la muerte se definió como el tiempo desde el inicio del tratamiento con gefitinib hasta la fecha de la muerte o la fecha del último seguimiento y se informó en semanas. Los pacientes sin muerte registrada se censuraron en la fecha del último seguimiento. Se aplicó una regresión de vectores de soporte para datos censurados con reducción de características (SVRc-FR) a las variables clínicas y las características generadas por M-Plex™, morfometría histológica e inmunohistoquímica de EGFR. SVRc es una adaptación de la SVR tradicional para dar cabida a los datos censurados. Para lograr esto, se definió una función modificada de pérdida/penalización que permite el procesamiento de datos censurados por la derecha y no censurados. El algoritmo SVRc-FR tiene un paso de filtrado inicial para eliminar las características que no se correlacionan de forma univariante con el resultado de interés, en base al índice de concordancia (CI). El CI es la probabilidad de predecir, en el orden correcto, los tiempos para dos pacientes elegidos aleatoriamente, donde ambos murieron o uno murió antes del último seguimiento del paciente censurado. El CI varía de 0 a 1, con 0,5 que indica una asociación aleatoria. Los CI más alejados de 0,5 indican asociaciones más fuertes con el resultado. Solo se conservaron las características con CI <0,4 o >0,6. El modelo de Cox se usó como un comparador para características múltiples, que incluyen las seleccionadas por SVRc-FR. Mediante el uso de un enfoque gradual, el criterio de entrada para una característica fue un valor $p \leq 0,15$. Para permanecer en el modelo, una característica tenía que tener un valor $p \leq 0,05$.

Resultados

De los 284 pacientes, las muestras de tejido de 109 (38 %) tenían suficiente tumor (>50 %) para análisis adicionales. De las 109 muestras de pacientes, se obtuvieron resultados de mutación de EGFR de 87, mientras que 51 pacientes tenían datos completos (es decir, clínicos, inmunohistoquímicos y de inmunofluorescencia) para su inclusión en un modelo final. Las características demográficas de la cohorte total (284 pacientes) y el modelo final (51 pacientes) están contenidas en la Tabla 24.

Tabla 24. Características de los pacientes en el modelo predictivo final en comparación con la cohorte total de pacientes.

5	Característica	Pacientes en el modelo de 51 pacientes: N (%)	Cohorte completa de 284 pacientes: N (%)
10	Estado funcional		
	0	2 (3,9)	13 (4,58)
	1	39 (76,5)	203 (71,48)
	2	10 (19,6)	68 (23,94)
15	Género		
	Femenino	22 (43,1)	126 (44,37)
	Masculino	29 (56,9)	158 (55,63)
20	Histología		
	Adenocarcinoma	31 (60,8)	138 (48,59)
	Broncoalveolar	1 (2,0)	5 (1,76)
	Célula grande	5 (9,8)	36 (12,68)
	Mixto	2 (3,9)	20 (7,04)
25	NSCL	5 (9,8)	30 (10,56)
	Escamoso	7 (13,7)	55 (19,37)
30	Antecedentes de tabaquismo		
	No	8 (15,7)	45 (15,85)
	Sí	43 (84,3)	239 (84,15)
35	Regímenes de quimioterapia		
	0	1 (2,0)	19 (6,69)
	1	19 (37,3)	106 (37,32)
	2 o 3	25 (49,0)	135 (47,54)
	4, 5 o 6	6 (11,8)	24 (8,45)
40	Edad		
	Media +/- DE	60,8 +/- 12,83	63,02 +/- 11,29
	Mediana	62,0	64,0
	Rango	24,0 - 82,0	24,0 - 86,0
45	Supervivencia		
	Muertes	50 (98,0)	268 (94,37)
	Censurado	1 (2,0)	16 (5,63)
	Tiempo medio de supervivencia, 25,0 semanas		27,0
50	<i>Nota: los valores medios para las características continuas son comparables</i>		

Estado de mutación de EGFR

Se realizó un análisis de mutaciones de EGFR, en base a las secuencias publicadas para mutaciones en los exones 19-21, 4, 10 para los 87 pacientes que tenían datos de resultados clínicos y material suficiente para el análisis de ADN. Cuatro pacientes (5 %) tenían mutaciones en el dominio tirosina quinasa de EGFR en los exones 19, 20 o 21. Dos de estos pacientes exhibieron delecciones en el marco de lectura en el exón 19 (delL747-S752insV y delL747P753insS), uno tenía una mutación somática (T790M) en el exón 20 y el cuarto tenía una mutación L858R en el exón 21. Los 4 pacientes con mutación de EGFR eran mujeres y se habían diagnosticado con adenocarcinoma. Dos de estas pacientes habían logrado una respuesta parcial y 2 progresaron (Tabla 25). La mutación T790M, que estaba presente en un paciente que tuvo una supervivencia corta (4 semanas), se mostró previamente asociada con resistencia a los fármacos. La mediana de supervivencia general para todos los pacientes analizados en busca de mutaciones fue de 24 semanas, y la supervivencia general a 1 año fue de 26 %. Debido a la baja frecuencia de mutaciones de EGFR, el estado de mutación no se usó como variable en el desarrollo posterior del modelo.

Tabla 25. Estado de respuesta para pacientes con mutaciones de EGFR.

Número del paciente	Exón	Mutación	Respuesta objetiva	Tiempo hasta la muerte, semanas
46	19	del L747-S752insV	Respuesta parcial	65
127B	19	Del L747-P753insS	Respuesta parcial	92
69	21	L858R	Progresión	22
196	20	T790M	Progresión	4

Inmunohistoquímica de EGFR

La inmunohistoquímica se realizó en 60 muestras de pacientes para las cuales estaba presente > 50 % del tumor. En resumen, se evaluó la heterogeneidad de la proteína EGFR solo en regiones tumorales, con 14 muestras (23 %) que exhiben un índice de tinción ≥ 200 lo que sugiere un aumento de la expresión, mientras que 20 muestras (30 %) tenían un índice de tinción de 0. El índice de tinción tumoral fue la única variable incluida en el modelo predictivo. La inmunohistoquímica de la proteína EGFR se analizó mediante el uso del % de células positivas multiplicado por la intensidad (0-3+) para generar un índice de tinción (puntuación H) en pacientes con muestra disponible para el análisis (Tabla 28).

Tabla 28. Resultados detallados de la inmunohistoquímica de EGFR.

Número de protocolo DM05010						
Portaobjetos DM núm.	Total de células positivas, %	células +, %	células ++, %	células +++, %	Puntuación H*	
1	95	25	35	35	200	
2	80	30	30	20	150	
3	100	0	10	90	290	
4	10	10	0	0	10	
5	10	5	5	0	15	
6	0	0	0	0	0	
7	80	10	40	20	150	
8	60	20	40	0	100	
9	0	0	0	0	0	
10	100	0	20	80	280	
11	75	35	25	15	130	
12	100	0	10	90	290	
13	30	20	10	0	40	
14	40	10	20	10	80	
15	100	10	40	50	240	
16	0	0	0	0	0	
17	0	0	0	0	0	
18	90	10	40	40	210	
19	40	15	15	10	75	
20	90	20	30	40	200	
21	40	10	20	10	80	
22	0	0	0	0	0	
23	0	0	0	0	0	
24	0	0	0	0	0	
25	100	0	0	100	300	
26	0	0	0	0	0	
27	90	20	20	50	210	

Número de protocolo DM05010					
Portaobjetos DM núm.	Total de células positivas, %	células +, %	células ++, %	células +++, %	Puntuación H*
28	40	20	10	10	70
29	0	0	0	0	0
30	0	0	0	0	0
31	100	0	0	100	300
32	90	10	10	70	240
33	30	10	20	0	50
34	0	0	0	0	0
35	70	20	50	0	120
36	0	0	0	0	0
37	50	20	20	10	90
38	0	0	0	0	0
39	20	20	0	0	20
40	0	0	0	0	0
41	30	10	10	10	60
42	30	10	10	10	60
43	90	20	60	10	170
44	0	0	0	0	0
45	40	30	10	0	50
46	0	0	0	0	0
47	10	10	0	0	10
48	0	0	0	0	0
49	90	10	10	70	240
50	70	30	30	10	120
51	70	10	40	20	150
52	90	10	10	70	240
53	10	10	0	0	10
54	60	50	10	0	70
55	0	0	0	0	0
56	30	10	10	10	60
57	0	0	0	0	0
58	0	0	0	0	0
59	90	10	20	60	230
60	10	0	10	0	20
61	80	10	60	10	160
*Puntuación H = (% de células calificadas como +++) x 3 + (% de células ++) x 2 + (% de células +) x 1. La puntuación máxima de H es 300.					

Análisis de imágenes de H&E

Las imágenes digitalizadas seleccionadas de 109 muestras de pacientes teñidas con H&E se procesaron con un programa informático de imágenes. El programa informático segmenta y clasifica elementos tisulares - celulares individuales, núcleos epiteliales segregantes, citoplasma epitelial, estroma, así como lúmenes alveolares. Los elementos se clasifican mediante el uso de características espectrales de 'color', forma y relaciones espaciales entre

objetos de tejido. El lumen, por ejemplo, se clasifica a partir de la asociación general del espacio en blanco rodeado de células epiteliales alveolares 'tumores'. Treinta y nueve características de imagen se generaron e introdujeron en los modelos predictivos.

- 5 La Tabla 29 enumera las 39 características de H&E extraídas de las muestras de pacientes individuales, que incluyen cortes completos, biopsias con aguja y aspirados.

Tabla 29. Características de la morfometría histológica.

10	Nombre de la característica	Descripción
	CytoplasmMeanMeanChannel40058	Media del valor de píxel de todos los objetos de citoplasma, canal rojo
	CytoplasmMeanMeanChannel50059	Media del valor de píxel de todos los objetos de citoplasma, canal verde
15	CytoplasmMeanMeanChannel60060	Media del valor de píxel de todos los objetos de citoplasma, canal azul
	CytoplasmMeanStddevChannel40066	Media del valor de píxel de todos los objetos de citoplasma estándar, canal rojo
	CytoplasmMeanStddevChannel50067	Media del valor de píxel de todos los objetos de citoplasma estándar, canal verde
20	CytoplasmMeanStddevChannel60068	Media del valor de píxel de todos los objetos de citoplasma estándar, canal azul
	CytoplasmStddevMeanChannel40081	DE la media del valor de píxel de todos los objetos de citoplasma, canal rojo
	CytoplasmStddevMeanChannel50082	DE de la media del valor de píxel de todos los objetos de citoplasma, canal verde
25	CytoplasmStddevMeanChannel60083	DE de la media del valor de píxel de todos los objetos de citoplasma, canal azul
	EpithelialNucleiMeanAreaPx10093	Media del área de todos los objetos de núcleos epiteliales
30	EpithelNucleiMeanMeanChanne40104	Media de la media del valor de píxel de todos los objetos de núcleos epiteliales, canal rojo
	EpithelNucleiMeanMeanChanne50105	Media de la media del valor de píxel de todos los objetos de núcleos epiteliales, canal verde
	EpithelNucleiMeanMeanChanne60106	Media de la media del valor de píxel de todos los objetos de núcleos epiteliales, canal azul
35	EpitheNucleiMeanStddevChann40112	Media de DE del valor de píxel de todos los objetos de núcleos epiteliales, canal rojo
	EpitheNucleiMeanStddevChann50113	Media de DE del valor de píxel de todos los objetos de núcleos epiteliales, canal verde
40	EpitheNucleiMeanStddevChann60114	Media de DE del valor de píxel de todos los objetos de núcleos epiteliales, canal azul
	EpitheliaNucleiStddevAreaPx10116	DE del área de todos los objetos de núcleos epiteliales
	EpitheNucleiStddevMeanChann40127	DE de la media del valor de píxel de todos los objetos de núcleos epiteliales, canal rojo
45	EpitheNucleiStddevMeanChann50128	DE de la media del valor de píxel de todos los objetos de núcleos epiteliales, canal verde
	EpitheNucleiStddevMeanChann60129	DE de la media del valor de píxel de todos los objetos de núcleos epiteliales, canal azul
50	StromaNucleiMeanAreaPx10300	Media del área de todos los objetos de núcleos estromales
	StromaNucleiMeanMeanChannel40311	Media de la media del valor de píxel de todos los objetos de núcleos estromales, canal rojo
55	StromaNucleiMeanMeanChannel50312	Media de la media del valor de píxel de todos los objetos de núcleos estromales, canal verde
	StromaNucleiMeanMeanChannel60313	Media de la media del valor de píxel de todos los objetos de núcleos estromales, canal azul
	StromaNucleiMeanStddevChann40319	Media del valor de píxel de todos los objetos de núcleos estromales estándar, canal rojo
60	StromaNucleiMeanStddevChann50320	Media de DE del valor de píxel de todos los objetos de núcleos estromales, canal verde
	StromaNucleiMeanStddevChann60321	Media de DE del valor de píxel de todos los objetos de núcleos estromales, canal azul
65		

Nombre de la característica	Descripción
StromaNucleiStddevAreaPx10323	DE del área de todos los objetos de núcleos estromales
StromaNucleiStddevMeanChann40334	DE de la media del valor de píxel de todos los objetos de núcleos
StromaNucleiStddevMeanChann50335	DE de la media del valor de píxel de todos los objetos de núcleos estromales, canal verde
StromaNucleiStddevMeanChann60336	DE de la media del valor de píxel de todos los objetos de núcleos estromales, canal azul
AreaCytopdivTotTissueArea	Área de citoplasma dividida por el área de tejido total
AreaEpitNucdivTotTissueArea	Área de núcleos epiteliales dividida por el área de tejido total
AreaLumendivTotTissueArea	Área de lumen dividida por el área de tejido total
AreaLymphdivTotTissueArea	Área de linfoides dividida por el área de tejido total
AreaStrNucdivTotTissueArea	Área de núcleos estromales dividida por el área de tejido total
AreaStromadivTotTissueArea	Área de estroma dividida por el área de tejido total
NormTotalNucleiArea	Área total de unidades de núcleos dividida por el área de tejido total
NormGlandTissueArea	Área total de unidades de glándulas dividida por el área de tejido total
DE, desviación estándar	

Inmunofluorescencia M-Plex™ cuantitativa

- Los 15 anticuerpos usados exhibieron la localización y distribución celular esperadas cuando se evaluaron mediante el uso de muestras de tejido de control. Las muestras de tumor de 59 pacientes se evaluaron mediante el uso de los 15 anticuerpos que se enumeran en la sección de métodos. Se seleccionaron regiones de interés representativas para cada anticuerpo en base a la cantidad y calidad del tumor dentro de la muestra, evitando regiones de necrosis y restos celulares. Se adquirió un mínimo de tres campos por muestra de paciente en cada análisis múltiple, y todas las imágenes se procesaron mediante el uso del programa informático de inmunofluorescencia para generar 84 características cuantitativas. La caspasa-3 activada y el mTOR fosforilado se encontraron focalmente dentro de linfocitos infiltrantes y células epiteliales tumorales. Por el contrario, la ciclina D1 estaba presente predominantemente dentro de los núcleos de las células epiteliales tumorales. De interés, pKDR se encontró difusamente dentro de islas de células epiteliales y endoteliales tumorales.
- Las 84 características (Tabla 30) representan el resultado del algoritmo (media, máximo y desviación estándar en relación con la intensidad y el área) de los biomarcadores individuales con respecto al compartimento en el que se encuentran (es decir, DAPI=nuclear; nuclear para ciclina D1, Ki67, PTEN y ERK fosforilada (pERK); citoplasmático para CK18, caspasa-3a y VEGF; nuclear/citoplasmático para AKT fosforilada; membrana para EGFR y KDR; y citoplasmático/membrana para KDR fosforilado (pKDR), mTOR fosforilado, PI3K y EGFR fosforilado), distribución de la señal (es decir píxeles como puntos geométricos frente a píxeles dentro de un objeto, tal como un núcleo, citoplasma) y relación entre sí (por ejemplo, EGFR respecto a pEGFR; p-mTOR respecto a pERK).

Tabla 30. Características de inmunofluorescencia cuantitativa.

Nombre de la característica
Dapiwithcyc10002
Dcyclkmean500007
Dcyclvar1000013
Ocyclkmean500027
Ocyclvar1000030
Ocyclvar500032
Oquantile95cyc10038
Ratio1 00cyc0041
Ratio50cyc0043
B1 EgfrEgfrmean500004
B1EgfrEgfrq950010
B1 EgfrEgfrvar1 000016
B2pEgfrmean500023
B2pEgfrvar1000035
CorrB1 EgfrEgfrB2pEgfr0040
OB1 Egfrmean500053
OB1Egfrq950059

	Nombre de la característica
5	OB1 Egfrvar1000065 OB2pEgfrmean500072 OB2pEgfrq950077 OB2pEgfrvar1000083 VaroverlapB 1Egfr0094
10	VaroverlapB2pEgfr0095 B1 Vegfmean500004 B1Vegfq950010 B1 Vegfvar1000016
15	B2pKdrmean500023 B2pKdrq950029 B2pKdrvar1000035 CorrB1 VegfB2pKdr0040
20	OB1 Vegfmean500053 OB1 Vegfvar1000065 OB2pKdrmean500072 OB2pKdrq950077
25	OB2pKdrvar1000083 VaroverlapB2pKdr0095 B3Caspasemean 1000040 B3Caspasemean500042
30	B3Caspaseq950048 B3Caspaseratio 1000049 B3Caspaseratio500051 B3Caspasevar1000054
35	B4Pi3kmean500061 B4Pi3kq950067 B4Pi3kratio1000068 B4Pi3kratio500070
40	B4Pi3kvar1000073 Corrb 1B2Ki670078 OB2Ki67mean500122 OB2Ki67q950127
45	OB2Ki67ratio1000128 OB2Ki67ratio500130 OB2Ki67var1000133 VarB3Caspase0144
50	VarB4Pi3k0145 B3KDRmean500042 B3KDRq950048 B3KDRratio 1000049
55	B3KDRratio500051 B3KDRvar1000054 B4pAKTmean500061 B4pAKTq950067
60	B4pAKTratio 1000068 B4pAKTratio500070 B4pAKTvar1000073 Corrb1 B2pTEN0078
65	OB2pTENmean500122

	Nombre de la característica
5	OB2pTENq950127
	OB2pTENratio1000128
	OB2pTENratio500130
	OB2pTENvar1000133
10	VarB3KDR0144
	VarB4pAKT0145
	MeanVesselArea0030
	StdVesselArea0034
	EpitInPmtoWithPerkMinMeaCha40026
15	Dmptormean500036
	Dperkmean500039
	Dperkvar1000042
	Dpmtormean0048
20	Dpmtorquantile950056
	Dpmtorvar1000058
	Dquantile95perk0065
	Dvarperk0066

Desarrollo de modelos predictivos

Mediante el uso de un enfoque integrador de 'patología de sistemas', se desarrollaron ocho modelos de supervivencia general, basados en variables clínicas más diferentes combinaciones de las siguientes características: inmunohistoquímica de EGFR (molecular), morfometría histológica (morfométrica) e inmunofluorescencia (molecular). Tenga en cuenta que debido a la ausencia de datos de algunos pacientes para algunos de los dominios de características, a medida que se añaden dominios adicionales, el número de pacientes incluidos se reduce de un máximo de 284 a un conjunto mínimo de 51. Debido a que ninguna de las 6 características clínicas pasó el filtro de CI, no pudo aplicarse SVRc-FR al conjunto de datos con características clínicas solamente (cohorte de 284 pacientes).

Tabla 31. Rendimiento y características elegidas en los 8 modelos construidos por SVRc-FR.

	Conjunto de datos	Número de pacientes	Número de pacientes vivos	CI del mejor modelo	Características elegidas (en orden de importancia en el modelo)	Contribución de la característica
40	Clínica	284	16	NA	ninguna característica pasó el filtro de CI	NA
45	Clínico + IHC	55	1	0,62	ps_m	-5,35
	Clínico + IF	56	2	0,72	ps_m	-3,85
					B3 Caspasemean 1000040	-3,22
					Ocyclvar1000030	-2,29
50					B1Vegfvar1000016	-2,24
					VaroverlapB2pKdr0095	-1,99
	Clínico + IHC + IF	52	1	0,73	ps_m	-4,75
55					VaroverlapB2pKdr0095	-2,63
					B3Caspasemean1000040	-2,52
					Dcyclkmean500007	-2,21
60	Clínico + H&E	53	2	0,67	ps_m	-4,72
					AreaLumendivTotTissueArea	2,46
	Clínico + IHC + H&E	51	1	0,68	ps_m	-4,77
65					AreaLumendivTotTissueArea	2,51

5	Conjunto de datos	Número de pacientes	Número de pacientes vivos	CI del mejor modelo	Características elegidas (en orden de importancia en el modelo)	Contribución de la característica
	Clínico + IF + H&E	53	2	0,75	ps_m	-3,50
10					Dcyclkmean500007	-2,49
					AreaLumendivTotTissueArea	2,36
15					B3Caspasemean1000040	-1,89
					VaroverlapB2pKdr0095	-1,86
					EpitInPmtoWithPerkMinMeaCha40026	-1,80
					B2pKdrvar1000035	-1,54
					Ocyclvar1000030	-1,54
					VarB3Caspase0144	-1,41
20	Clínico + IHC + IF + H&E	51	1	0,74	ps_m	-4,77
25					B3Caspasemean1000040	-2,51
					Ocyclvar1000030	-2,32
					B2pKdrvar1000035	-2,17
					Dcyclkmean500007	-2,08
					AreaLumendivTotTissueArea	1,91
30	Para evaluar la utilidad del enfoque de patología de sistemas, los ocho modelos anteriores se entrenaron mediante el uso de datos del subconjunto de pacientes con información completa para todos los dominios de características (n = 51). Dado que IHC no pasó el filtro de CI, la exclusión de esta característica no cambió el modelo (por ejemplo, el modelo clínico + IHC es equivalente al modelo clínico). La Tabla 32 presenta los resultados de SVRc-FR en la cohorte de 51 pacientes. El estado funcional fue la única variable clínica que pasó el filtro de CI. La única característica de H&E seleccionada fue el área relativa del lumen. La adición de características de IF a las características clínicas mejoró el CI. La adición de las características de H&E a los datos clínicos también mejoró el CI, pero un poco menos que IF.					
35						
40	Tabla 32. Modelos de SVRc-FR contruidos a partir de la cohorte de 51 pacientes con información completa.					
45	Conjunto de datos	CI del mejor modelo	Características elegidas (en orden de importancia en el modelo)		Contribución de la característica	
	Clínica	0,62	ps_m		-4,766	
	Clínico + IF	0,73	ps_m		-4,766	
50			B3 Caspasemean 1000040		-2,4688	
			Ocyctvar1000030		-2,2744	
			B2pKdrvar1000035		-2,1261	
			Dcyclkmean500007		-2,0558	
55	Clínico + H&E	0,68	ps_m		-4,766	
			AreaLumendivTotTissueArea		2,5081	
60	Clínico + IF + H&E	0,74	ps_m		-4,766	
			B3Caspasemean1000040		-2,5079	
			Ocyclvar1000030		-2,3232	
65			B2pKdrvar1000035		-2,1663	

Conjunto de datos	CI del mejor modelo	Características elegidas (en orden de importancia en el modelo)	Contribución de la característica
5		Dcyclkmean500007	-2,0788
		AreaLumendivTotTissueArea	1,9107

10 Los resultados de Cox para los diferentes modelos se presentan en la Tabla 33. Estos son análisis multivariantes basados en un modelo de Cox con selección gradual de variables. Todas las características que tenían un valor $p \leq 0,15$ se introdujeron como entrada en el modelo de Cox. Para permanecer en el modelo, una característica necesitaba un valor $p \leq 0,05$.

15 Una relación de riesgos mayor que 1 indica un efecto negativo sobre la supervivencia (un valor más alto de la característica predice un tiempo de supervivencia más corto); una relación de riesgos menor que 1 indica que la característica tiene un efecto protector (un valor más alto de la característica predice un mayor tiempo de supervivencia). En base a estos resultados, se encontró que el estado funcional es un predictor clínico significativo de supervivencia. De las características de IF, a menudo se seleccionó el mismo anticuerpo en los resultados de Cox y los modelos de SVRc; sin embargo, la característica seleccionada capturó un atributo diferente del atributo seleccionado mediante el uso de SVRc-FR. De las características de H&E, el modelo de Cox encontró que el área media de los núcleos epiteliales y el área relativa del citoplasma eran predictivas de supervivencia.

Tabla 33. Resultados de Cox.

Conjunto de datos	DF	Estimación de parámetros	Error estándar	Chi cuadrado	Pr > ChiCuad	HR
Clínico						
ps_m	1	0,73059	0,12598	33,6295	<0,0001	2,076
sex m	1	0,30755	0,12428	6,1234	0,0133	1,360
Clínico + IHC						
ps_m	1	1,11823	0,32541	11,8089	0,0006	3,059
Clínico + IF						
ps_m	1	2,76611	0,55914	24,4737	<0,0001	15,897
sex_m	1	-0,90167	0,39445	5,2254	0,0223	0,406
smk_hist_m	1	2,51627	0,55970	20,2116	<0,0001	12,382
Ocyclmean500027	1	0,00399	0,00125	10,1507	0,0014	1,004
B1EgfrEgfrq950010	1	-0,00143	0,0003890	13,5689	0,0002	0,999
B1Vegfmean500004	1	-0,0005077	0,0001250	16,4936	<0,0001	0,999
B1Vegfvar1000016	1	0,00108	0,0004208	6,5637	0,0104	1,001
OB2pKdrmean500072	1	0,0004915	0,0001789	7,5490	0,0060	1,000
B3Caspaseq950048	1	-0,00166	0,0005254	9,9609	0,0016	0,998
Corrb1B2Ki670078	1	-5,34023	1,10152	23,5036	<0,0001	0,005
VarB3Caspase0144	1	0,00643	0,00170	14,2993	0,0002	1,006
Corrb1B2pTEN0078	1	1,56428	0,76844	4,1439	0,0418	4,779
OB2pTENvar1000133	1	-0,01540	0,00423	13,2270	0,0003	0,985
Clínico + IF						
Dperkmean500039	1	0,00140	0,0003034	21,1775	<0,0001	1,001
Clínico + IHC + IF						
ps_m	1	1,38468	0,41388	11,1932	0,0008	3,994
Dcyclvar1000013	1	0,00164	0,0005454	9,0105	0,0027	1,002
B1Vegfmean500004	1	-0,0001949	0,0000782	6,2195	0,0126	1,000
EpitInPmtoWithPerkMinMeaCha40026	1	0,00378	0,00174	4,7192	0,0298	1,004

			Estimación de parámetros	Error estándar	Chi cuadrado	Pr > ChiCuad	HR
5	Conjunto de datos		DF				
	Clínico + H&E						
	ps_m	1	1,10707	0,33782	10,7395	0,0010	3,025
	smk_hist_m	1	1,70880	0,49665	11,8381	0,0006	5,522
	EpithelialNucleiMeanAreaPx10093	1	-0,00958	0,00329	8,4502	0,0037	0,990
10	Clínico + IHC + H&E						
	ps_m	1	1,25964	0,35019	12,9383	0,0003	3,524
	smk_hist_m	1	1,59729	0,49377	10,4646	0,0012	4,940
15	CytoplasmMeanMeanChannel40058	1	-0,01083	0,00550	3,8834	0,0488	0,989
	EpithelialNucleiMeanAreaPx10093	1	-0,01150	0,00349	10,8336	0,0010	0,989
	Clínico + IF + H&E						
20	Dcyclvar1000013	1	0,00200	0,0005022	15,8982	<0,0001	1,002
	OB2pKdrvar1000083	1	0,0006294	0,0002266	7,7146	0,0055	1,001
	OB2pTENratio500130	1	-21,08066	8,03902	6,8764	0,0087	0,000
	Clínico + IHC + IF + H&E						
25	ps_m	1	1,48368	0,41363	12,8661	0,0003	4,409
	smk_hist_m	1	1,04319	0,47848	4,7534	0,0292	2,838
	Dcyclvar1000013	1	0,00186	0,0006154	9,1377	0,0025	1,002
	B1Vegfmean500004	1	-0,0002506	0,0000807	9,6380	0,0019	1,000
30	EpitInPmtoWithPerkMinMeaCha40026	1	0,00398	0,00184	4,6619	0,0308	1,004
	AreaCytopdivTotTissueArea	1	2,55288	0,99480	6,5855	0,0103	12,844
	DF, grados de libertad; HR, relación de riesgos.						
35	A continuación se describe una descripción detallada del modelo basado en los cuatro dominios de características. Los datos completos para cada uno de los dominios de entrada estaban disponibles para 51 pacientes. Sus características demográficas fueron similares a las de la cohorte general (Tabla 24).						
40	Las características seleccionadas en el modelo final se enumeran en la Tabla 26, en orden de importancia en el modelo. Una contribución positiva para una característica significa que los valores más altos predicen un mejor resultado (mayor tiempo de supervivencia) y una contribución negativa significa que los valores más altos predicen un peor resultado. La característica cuya contribución está más alejada de cero se considera la más importante.						
45	El modelo final se realizó con un CI de 0,74. En comparación, el modelo que solo utiliza datos clínicos produjo un CI de 0,62; el estado funcional fue la única característica clínica seleccionada. Por lo tanto, la adición de características de inmunofluorescencia y H&E aumentó la capacidad de predecir la supervivencia en un 12 %.						
50	Las puntuaciones del modelo se analizaron mediante una estadística de chi-cuadrado basada en la prueba del orden logarítmico; este análisis demostró que un punto de corte de 39,5 separó más fuertemente a los pacientes en términos de su experiencia de supervivencia real (chi-cuadrado, 21,39; valor p ajustado, 0,0002). Las puntuaciones por encima de 39,5 tenían una relación de riesgos de 5,26 (intervalo de confianza de 95 %: 2,60, 10,62), es decir, para pacientes con puntuaciones por encima de 39,5 en comparación con aquellos con puntuaciones más bajas, el riesgo de muerte durante el período de estudio fue 5 veces mayor. Esta relación de riesgos para el riesgo de muerte necesitaría confirmación en una cohorte retrospectiva e independiente. Se generaron curvas de Kaplan-Meier para pacientes por encima y por debajo del punto de corte, como se muestra en la Figura 20.						
55							
60	El valor predictivo de las características individuales se analizó adicionalmente mediante análisis de Cox. Este análisis confirmó que el estado funcional (relación de riesgos: 4,4, intervalo de confianza de 95 % 1,96 - 9,92) fue un predictor significativo de supervivencia en todos los modelos. Para las características de IF, que se desempeñaron de manera menos robusta que el estado funcional, a menudo se seleccionó el mismo marcador (es decir, ciclina D1, pKDR, caspasa-3), aunque se seleccionaron diferentes características derivadas, que reflejan diferentes atributos del marcador entre y dentro de los diferentes modelos.						
65							

Tabla 26. Características seleccionadas en el modelo final, desarrollado a partir de dominios de características clínicas, morfometría histológica, inmunohistoquímica e inmunofluorescencia

	Característica	Contribución
5	Estado funcional	-4,766
	Nivel de intensidad media de caspasa-3	-2,5079
	Intensidad total de ciclina D1	-2,3232
10	Variabilidad de pKDR	-2,1663
	Intensidad media de Ciclina D1	-2,0788
	Área relativa del espacio luminal (alveolar)	1,9107

15 Discusión

Existe un consenso en la comunidad médica sobre la importancia crítica de identificar a los pacientes con NSCLC que pueden beneficiarse de gefitinib, incluso cuando no se observa una respuesta objetiva en los estudios de imágenes. Por esta razón, se investigó la respuesta a gefitinib mediante el uso de la supervivencia general como criterio de valoración. Se ha demostrado que un modelo predictivo para la supervivencia general de los pacientes tratados con gefitinib se mejora mediante la adición de información en las muestras tumorales de los pacientes. Al integrar las características cuantitativas de los biomarcadores y la morfometría histológica con los datos clínicos, la exactitud (CI) de tal modelo predictivo ha aumentado de 62 % (características clínicas solas) a 74 % (todos los dominios).

Se han propuesto numerosas variables que afectan la sensibilidad a gefitinib, que incluyen género, histología tumoral (por ejemplo, adenocarcinoma/carcinoma broncoalveolar), antecedentes de tabaquismo, origen étnico, número de copias del gen de EGFR, mutaciones en el dominio quinasa de EGFR e incluso el tipo de mutaciones de EGFR. Recientemente, se informó que los pacientes con una delección del exón 19 de EGFR tenían una mediana de supervivencia mayor que los pacientes con una mutación L858R. Además, los polimorfismos de la línea germinal en el promotor del EGFR se han propuesto como factores potencialmente relevantes para la respuesta a gefitinib en pacientes que no tienen mutaciones en el dominio tirosina quinasa. Además, se han observado respuestas a gefitinib en pacientes con NSCLC independientemente de la histología tumoral o las características demográficas del paciente.

En la cohorte EAP actual, se identificaron mutaciones de EGFR en los exones 19-21 en 4 de los 87 pacientes. Aunque la frecuencia de la mutación era demasiado baja para permitir la inclusión de estas mutaciones en un modelo predictivo, era consistente con estudios previos que involucraban pacientes predominantemente occidentales. En esta cohorte, como en estudios anteriores, la frecuencia de mutaciones de EGFR fue mucho menor que la frecuencia probable de beneficio clínico de gefitinib. En un análisis de 124 pacientes de la cohorte EAP actual, 45 % de los pacientes no tenían evidencia de progresión en la primera reevaluación y 29 % informó una mejoría en los síntomas relacionados con el cáncer de pulmón mientras recibían gefitinib.

La única variable clínica identificada de manera consistente como asociada con la supervivencia fue el estado funcional. Esto concuerda con un estudio anterior que demostró que un buen estado funcional (estado ECOG de 0-2) se asociaba con una mayor supervivencia. La histología broncoalveolar, que se ha asociado con un buen resultado clínico, no se seleccionó en nuestro modelo predictivo; sin embargo, la característica de imagen histológica seleccionada, 'área del lumen dividida por el área de tejido total', puede interpretarse como una característica morfológica sustituta de este subtipo de adenocarcinoma.

Los mecanismos moleculares que subyacen a las diferentes sensibilidades a gefitinib de las subpoblaciones de pacientes con NSCLC avanzado siguen sin comprenderse. Además de las mutaciones de EGFR y el número de copias del gen de EGFR, otras variables moleculares que se han examinado incluyen un EGFR intacto, EGFR fosforilado y moléculas corriente abajo tales como PI3K, pAkt y pERK.^{31,52-54} Desafortunadamente, hay desacuerdo sobre el papel, si lo hay, que cada uno de estos factores desempeña en la sensibilidad a gefitinib, y hay menos acuerdo sobre su importancia para la supervivencia general. Incluso EGFR, según lo evaluado por inmunohistoquímica, ha sido objeto de informes contradictorios de su asociación con la respuesta a gefitinib.

En el análisis actual, se identificaron varios marcadores por inmunofluorescencia cuantitativa que se asociaron con la supervivencia general. De importancia, el modelo final seleccionó dos características de ciclina D1, con mayores cantidades asociadas con una supervivencia general más corta.

Los biomarcadores restantes seleccionados por el modelo, que incluyen la caspasa-3 activada y el KDR fosforilado, no se han investigado ampliamente en muestras clínicas de cáncer de pulmón. La caspasa-3 ha sido objeto de informes contradictorios que vinculan la sobreexpresión con la mediana de supervivencia. Para pKDR; sin embargo, no hay estudios recientes publicados sobre inmunohistoquímica o inmunofluorescencia en muestras clínicas de NSCLC. Los modelos de Cox confirmaron que la sobreexpresión de VEGF (basados en características seleccionadas) se asoció con una supervivencia reducida, lo que concuerda con estudios anteriores que examinaron VEGF junto con

la densidad de microvasos y la invasión tumoral. El hecho de que ciclina D1, caspasa-3, pKDR y VEGF se seleccionaron repetidamente en varios submodelos sugiere la importancia de la regulación del ciclo celular y la angiogénesis para el crecimiento tumoral, la metástasis y la supervivencia.

- 5 Por lo tanto, se ha construido una serie de modelos integradores basados en sistemas que predicen con éxito la supervivencia general en pacientes con NSCLC avanzado tratados con gefitinib. Los biomarcadores seleccionados son provocativos porque representan manifestaciones de la regulación del ciclo celular, la apoptosis y la angiogénesis.

Realizaciones adicionales

- 10 Por lo tanto, se observa que se proporcionan métodos y sistemas para predecir la aparición de una afección médica. Aunque en el presente documento se han descrito realizaciones particulares en detalle, esto se ha hecho a modo de ejemplo únicamente con fines ilustrativos, y no pretende ser limitante con respecto al alcance de las reivindicaciones adjuntas, que siguen. En particular, los inventores contemplan que pueden realizarse varias sustituciones, alteraciones y modificaciones sin apartarse del alcance de la invención tal como se define en las reivindicaciones. Otros aspectos, ventajas y modificaciones se consideran dentro del alcance de las siguientes reivindicaciones. Las reivindicaciones presentadas son representativas de las invenciones descritas en el presente documento. También se contemplan otras invenciones no reivindicadas. Los solicitantes se reservan el derecho de luchar por tales invenciones en reivindicaciones posteriores.

- 20 En la medida en que las realizaciones de la invención descritas anteriormente sean implementables, al menos en parte, mediante el uso de 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.

- 30 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.

- 35 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, tal como un disco legible y/o grabable, por ejemplo, un disco compacto (CD) o un disco versátil digital (DVD), o una memoria magnética tal 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, algunas de las cuales se mencionaron anteriormente se incorporan en el presente documento como referencia en su totalidad:

- 45 [1] Scherr D., y otros, Urology. 61 (2 Supl. 1): 14-24, feb. de 2003, Swindle P.W., y otros, Urologic Clinics of North America. 30(2):377-401, mayo de 2003.
 [2] Wahlby C., y otros, Analytical Cellular Pathology 24, 101-111, 2002.
 50 [3] Street W.N., "Xcvt: 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
 [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.
 55 [5] Cristianni y otros, An Introduction to Support Vector Machines, Cambridge University Press (2000).
 [6] Hastie, The Elements of Statistical Learning, Springer (2001).
 [7] F.E. Harrell y otros, "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, April 2004.
 60 [11] Baatz M. y Schäpe 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. y otros, Pattern Classification, 2ª Edición, John Wiley & Sons Inc., 2001.
 65 [14] Holmberg L. y otros, A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer, N. Engl. M. Med., 347:781-789 (2002).

- [15] Pound CR y otros, Natural history of progression after PSA elevation following radical prostatectomy, JAMA 1999, 281:1591-1597.
- [16] Kumar-Sinha C. y otros, Molecular markers to identify patients at risk for recurrence after primary treatment for prostate cancer, Urology 2003; 62 Supl. 1:19-35.
- 5 [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 y otros, "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.
- [20] Kattan y otros, "Postoperative Nomogram for Disease Recurrence After Radical Prostatectomy for Prostate Cancer," Journal of Clinical Oncology, Vol. 17, N° 5 (Mayo), 1999: págs 1499-1507.
- 10 [21] C. Vonesch, F. Aquet, J.L. Vonesch y M. Unser, "The colored revolution of bioimaging," IEEE Signal Proc. Mag., vol. 23, núm. 3, pp. 20-31, mayo de 2006.
- [22] A. Krtolica, C. O. de Solorzano, S. Lockett y J. Campisi, "Quantification of epithelial cells in coculture with fibroblast by fluorescence image analysis," Cytometry, vol. 49, pp. 73-82, 2002.
- 15 [23] A. Gordon, A. Colman-Lerner, T. E. Chin, K. R. Benjamin, R. C. Yu, y R. Brent, "Single-cell quantification of molecules and rates using open-source microscope-based cytometry," Nature Methods, vol. 4, pp. 175-181, 2007.
- [24] R. Camp, G. G. Chung, y D. L. Rimm, "Automated subcellular localization and quantification of protein expression in tissue microarrays," Nature Medicine, vol. 8, pp. 1323-1327, 2002.
- [25] J. Y. Rao, D. Seligson, y G. P. Hemstreet, "Protein expression analysis using quantitative fluorescence image analysis on tissue microarray slides," BioTechniques, vol. 32, pp. 924-932, 2002.
- 20 [26] Definiens Understanding Images, Developer Version 6, 2007. <http://www.definiens.com>.
- [27] <http://www.cri-inc.com>.
- [28] A. Tabesh, M. Teverovskiy, H.-Y Pang, V. P. Kumar, D. Verbel, A. Kotsianti y O. Saidi, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," IEEE Trans. Medical Imag., vol. 26, pp. 1366-1378, 2007.
- 25 [29] Giaccone G, Herbst RS, Manegold C, y otros: Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial--INTACT 1. J Clin Oncol 22:777-84, 2004.
- [30] Herbst RS, Giaccone G, Schiller JH, y otros: Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial--INTACT 2. J Clin Oncol 22:785-94, 2004.
- 30 [31] Thatcher N, Chang A, Parikh P, y otros Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). Lancet 366:1527-37, 2005.
- [32] Lynch TJ, Bell DW, Sordella R, y otros: Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350:2129-39, 2004.
- 35 [33] Dudek AZ, Kmak KL, Koopmeiners J, y otros: Skin rash and bronchoalveolar histology correlates with clinical benefit in patients treated with gefitinib as a therapy for previously treated advanced or metastatic non-small cell lung cancer. Lung Cancer 51:89-96, 2006.
- [34] Cappuzzo F, Hirsch FR, Rossi E, y otros: Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. J Natl Cancer Inst 97:643-55, 2005.
- 40 [35] Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden y otros: Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 5:649-655, 1982. Cordon-Cardo C, Kotsianti A, Donovan M, y otros: Improved prediction of PSA recurrence through systems pathology. J Clin Oncol 22:4591, 2004 (resumen).
- [36] Zubek V, Verbel D, Saidi O: Censored time trees for predicting time to PSA recurrence, in Proceedings of the Fourth International Conference on Machine Learning and Applications (ICMLA 2005). Washington, DC, IEEE Computer Society, 2005, pp 221-226.
- 45 [37] Paez JG, Janne PA, Lee JC, y otros: EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304:1497-500, 2004.
- [38] Bell DW, Gore I, Okimoto RA, y otros: Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. Nat Genet 37:1315-6, 2005.
- 50 [39] Hirsch FR, Varella-Garcia M, Bunn PA, Jr., y otros: Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. J Clin Oncol 21:3798-807, 2003.
- [40] Chan SK, Gullick WJ, Hill ME: Mutations of the epidermal growth factor receptor in non-small cell lung cancer -- search and destroy. Eur J Cancer 42:17-23, 2006.
- 55 [41] Riely GJ, Pao W, Pham D, y otros: Clinical course of patients with non-small cell lung cancer and epidermal growth factor receptor exon 19 and exon 21 mutations treated with gefitinib or erlotinib. Clin Cancer Res 12:839-44, 2006.
- [42] Sequist LV, Bell DW, Lynch TJ, y otros: Molecular predictors of response to epidermal growth factor receptor antagonists in non-small-cell lung cancer. J Clin Oncol 25:587-95, 2007.
- 60 [43] Uchida A, Hirano S, Kitao H, y otros: Activation of downstream epidermal growth factor receptor (EGFR) signaling provides gefitinib-resistance in cells carrying EGFR mutation. Cancer Sci 98:357-63, 2007.
- [44] Sugio K, Uramoto H, Ono K, y otros: Mutations within the tyrosine kinase domain of EGFR gene specifically occur in lung adenocarcinoma patients with a low exposure of tobacco smoking. Br J Cancer 94:896-903, 2006.
- 65 [45] Edelman MJ: An update on the role of epidermal growth factor receptor inhibitors in non-small cell lung cancer. Semin Oncol 32:S3-8, 2005.

- [46] Kris MG, Natale RB, Herbst RS, y otros: Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *Jama* 290:2149-58, 2003.
- [47] Al-Kuraya K, Siraj AK, Bavi P, y otros: High epidermal growth factor receptor amplification rate but low mutation frequency in Middle East lung cancer population. *Hum Pathol* 37:453-7, 2006.
- [48] Hainsworth JD, Mainwaring MG, Thomas M, y otros: Gefitinib in the treatment of advanced, refractory non-smallcell lung cancer: results in 124 patients. *Clin Lung Cancer* 4:347-55, 2003.
- [49] Su WP, Yang CH, Yu CJ, y otros: Gefitinib treatment for non-small cell lung cancer -- a study including patients with poor performance status. *J Formos Med Assoc* 104:557-62, 2005.
- [50] Ertan M, Grunenwald D, Penault-Llorca F, y otros: Epidermal growth factor receptor, HER-2/neu and related pathways in lung adenocarcinomas with bronchioloalveolar features. *Lung Cancer* 47:315-23, 2005.
- [51] Massion PP, Taflan PM, Shyr Y, y otros: Early involvement of the phosphatidylinositol 3-kinase/Akt pathway in lung cancer progression. *Am J Respir Crit Care Med* 170:1088-94, 2004.
- [52] Cappuzzo F, Magrini E, Ceresoli GL, y otros: Akt phosphorylation and gefitinib efficacy in patients with advanced non-small-cell lung cancer. *J Natl Cancer Inst* 96:1133-41, 2004.
- [53] Han SW, Hwang PG, Chung DH, y otros: Epidermal growth factor receptor (EGFR) downstream molecules as response predictive markers for gefitinib (Iressa, ZD1839) in chemotherapy-resistant non-small cell lung cancer. *Int J Cancer* 113:109-15, 2005.
- [54] Bailey L, Kris M, Wolf M, y otros: Tumor EGFR membrane staining is not clinically relevant for predicting response in patients receiving gefitinib (Iressa, ZD 1839) monotherapy for pretreated advanced non-small-cell lung cancer: IDEAL 1 and 2. *Proc Am Assoc Cancer Res* 44:1362, 2003 (resumen).
- [55] Bailey L, Janas M, Schmidt K, y otros: Evaluation of epidermal growth factor receptor (EGFR) as a predictive marker in patients with non-small-cell lung cancer (NSCLC) receiving first-line gefitinib combined with platinum-based chemotherapy. *J Clin Oncol* 22:7013, 2004 (resumen).
- [56] Jin M, Inoue S, Umemura T, y otros: Cyclin D1, p16 and retinoblastoma gene product expression as a predictor for prognosis in non-small cell lung cancer at stages I and II. *Lung Cancer* 34:207-18, 2001.
- [57] Keum JS, Kong G, Yang SC, y otros: Cyclin D1 overexpression is an indicator of poor prognosis in resectable non-small cell lung cancer. *Br J Cancer* 81:127-32, 1999.
- [58] Kalish LH, Kwong RA, Cole IE, y otros: Deregulated cyclin D1 expression is associated with decreased efficacy of the selective epidermal growth factor receptor tyrosine kinase inhibitor gefitinib in head and neck squamous cell carcinoma cell lines. *Clin Cancer Res* 10:7764-74, 2004.
- [59] Koomagi R, Volm M: Relationship between the expression of caspase-3 and the clinical outcome of patients with non-small cell lung cancer. *Anticancer Res* 20:493-6, 2000.
- [60] Takata T, Tanaka F, Yamada T, y otros: Clinical significance of caspase-3 expression in pathologic-stage I, nonsmall-cell lung cancer. *Int J Cancer* 96 Supl.:54-60, 2001.
- [61] Mineo TC, Ambrogi V, Baldi A, y otros: Prognostic impact of VEGF, CD31, CD34, and CD105 expression and tumour vessel invasion after radical surgery for IB-IIA non-small cell lung cancer. *J Clin Pathol* 57:591-7, 2004.

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

Característica	
5	Cytoplasm.SumBorderlengthPxI
	Cytoplasm.MaxBrightness
	Cytoplasm.MeanBrightness
	Cytoplasm.MinBrightness
	Cytoplasm.StdDevBrightness
10	Cytoplasm.MaxCompactness
	Cytoplasm. Mean Compactness
	Cytoplasm.MinCompactness
	Cytoplasm.StdDevCompactness
15	Cytoplasm.MaxDensity
	Cytoplasm.Mean Density
	Cytoplasm.MinDensity
	Cytoplasm.StdDevDensity
20	Cytoplasm.MaxDiff.ofenclosing.encl
	Cytoplasm.MeanDiff.ofenclosing.encl
	Cytoplasm.MinDiff.ofenclosing.encl
	Cytoplasm.StdDevDiff.ofenclosing.en
25	Cytoplasm.MaxEllipticFit
	Cytoplasm.MeanEllipticFit
	Cytoplasm.MinEllipticFit
	Cytoplasm.StdDevEllipticFit
30	Cytoplasm.MaxLengthPxI
	Cytoplasm.MeanLengthPxI
	Cytoplasm.MinLengthPxI
	Cytoplasm.StdDevLengthPxI
35	Cytoplasm.SumLengthPxI
	Cytoplasm.MaxMax.Diff.
	Cytoplasm.MeanMax.Diff.
	Cytoplasm.MinMax.Diff.
40	Cytoplasm.StdDevMax.Diff.
	Cytoplasm.MaxMeanChannel1
	Cytoplasm.MeanMeanChannel1
	Cytoplasm.MinMeanChannel1
45	Cytoplasm.StdDevMeanChannel1
	Cytoplasm.MaxMeanChannel2
	Cytoplasm.MeanMeanChannel2
	Cytoplasm.MinMeanChannel2
50	Cytoplasm.StdDevMeanChannel2
	Cytoplasm.MaxMeanChannel3
	Cytoplasm.MeanMeanChannel3
	Cytoplasm.MinMeanChannel3
55	Cytoplasm.StdDevMeanChannel3
	Cytoplasm.MaxRadiusoflargestenclose
	Cytoplasm.MeanRadiusoflargestenclos
	Cytoplasm.MinRadiusoflargestenclose
60	Cytoplasm.StdDevRadiusoflargestenclo
	Cytoplasm.MaxRadiusofsmallestenclos
	Cytoplasm.MeanRadiusofsmallestenclo
	Cytoplasm.MinRadiusofsmallestenclos
65	Cytoplasm.StdDevRadiusofsmallestenc
	Cytoplasm.MaxStdevChannel1
	Cytoplasm.MeanStdevChannel1

Característica	
5	Cytoplasm.MinStddevChannel1
	Cytoplasm.StdDevStddevChannel1
	Cytoplasm.MaxStddevChannel2
	Cytoplasm.MeanStddevChannel2
	Cytoplasm.MinStddevChannel2
10	Cytoplasm.StdDevStddevChannel2
	Cytoplasm.MaxStddevChannel3
	Cytoplasm.MeanStddevChannel3
	Cytoplasm.MinStddevChannel3
	Cytoplasm.StdDevStddevChannel3
15	Cytoplasm.MaxWidthPxI
	Cytoplasm.MeanWidthPxI
	Cytoplasm.MinWidthPxI
	Cytoplasm.StdDevWidthPxI
	Epithelial.Nuclei.Objects
20	Epithelial.Nuclei.ObjectsPct
	Epithelial.Nuclei.MaxAreaPxI
	Epithelial.Nuclei.MeanAreaPxI
	Epithelial.Nuclei.MinAreaPxI
	Epithelial.Nuclei.StdDevAreaPxI
25	Epithelial.Nuclei.SumAreaPxI
	Epithelial.Nuclei.MaxAsymmetry
	Epithelial.Nuclei.MeanAsymmetry
	Epithelial.Nuclei.MinAsymmetry
	Epithelial.Nuclei.StdDevAsymmetry
30	Epithelial.Nuclei.MaxBorderlengthPx
	Epithelial.Nuclei.MeanBorderlengthPx
	Epithelial.Nuclei.MinBorderlengthPx
	Epithelial.Nuclei.StdDevBorderlengt
	Epithelial.Nuclei.SumBorderlengthPx
40	Epithelial.Nuclei.MaxBrightness
	Epithelial.Nuclei.MeanBrightness
	Epithelial.Nuclei.MinBrightness
	Epithelial.Nuclei.StdDevBrightness
	Epithelial.Nuclei.MaxCompactness
45	Epithelial.Nuclei.MeanCompactness
	Epithelial.Nuclei.MinCompactness
	Epithelial.Nuclei.StdDevCompactness
	Epithelial.Nuclei.MaxDensity
	Epithelial.Nuclei.MeanDensity
50	Epithelial.Nuclei.MinDensity
	Epithelial.Nuclei.StdDevDensity
	Epithelial.Nuclei.MaxDiff.ofenclosi
	Epithelial.Nuclei.MeanDiff.ofenclosi
	Epithelial.Nuclei.MinDiff.ofenclosi
60	Epithelial.Nuclei.StdDevDiff.ofenclosi
	Epithelial.Nuclei.MaxEllipticFit
	Epithelial.Nuclei.MeanEllipticFit
	Epithelial.Nuclei.MinEllipticFit
	Epithelial.Nuclei.StdDevEllipticFit
65	Epithelial.Nuclei.MaxLengthPxI

Característica	
5	Epithelial.Nuclei.MeanLengthPxI
	Epithelial.Nuclei.MinLengthPxI
	Epithelial.Nuclei.StdDevLengthPxI
	Epithelial.Nuclei.SumLengthPxI
	Epithelial.Nuclei.MaxMax.Diff.
10	Epithelial.Nuclei .MeanMax.Diff.
	Epithelial.Nuclei.MinMax.Diff.
	Epithelial.Nuclei.StdDevMax.Diff.
	Epithelial.Nuclei.MaxMeanChannel1
	Epithelial.Nuclei.MeanMeanChannel1
15	Epithelial.Nuclei.MinMeanChannel1
	Epithelial.Nuclei.StdDevMeanChannel
	Epithelial.Nuclei.MaxMeanChannel2
	Epithelial.Nuclei.MeanMeanChannel2
	Epithelial.Nuclei.MinMeanChannel2
20	Epithelial.Nuclei.StdDevMeanChannel
	Epithelial.Nuclei.MaxMeanChannel3
	Epithelial.Nuclei.MeanMeanChannel3
	Epithelial.Nuclei.MinMeanChannel3
	Epithelial.Nuclei.StdDevMeanChannel
25	Epithelial.Nuclei.MaxRadiusoflarges
	Epithelial.Nuclei.MeanRadiusoflarge
	Epithelial.Nuclei.MinRadiusoflarges
	Epithelial.Nuclei.StdDevRadiusoflar
	Epithelial.Nuclei.MaxRadiusofsmalle
30	Epithelial.Nuclei.MeanRadiusofsmall
	Epithelial.Nuclei.MinRadiusofsmalle
	Epithelial.Nuclei.StdDevRadiusofsma
	Epithelial.Nuclei.MaxStdevChannel1
	Epithelial.Nuclei.MeanStdevChannel1
35	Epithelial.Nuclei.MinStdevChannell
	Epithelial.Nuclei.StdDevStdevChanne
	Epithelial.Nuclei.MaxStdevChannel2
	Epithelial.Nuclei.MeanStdevChannel2
	Epithelial.Nuclei.MinStdevChannel2
40	Epithelial.Nuclei.StdDevStdevChanne
	Epithelial.Nuclei.MaxStdevChannel3
	Epithelial.Nuclei.MeanStdevChannel3
	Epithelial.Nuclei.MinStdevChannel3
	Epithelial.Nuclei.StdDevStdevChanne
45	Epithelial.Nuclei.MaxWidthPx1
	Epithelial.Nuclei.MeanWidthPxI
	Epithelial.Nuclei.MinWidthPxI
	Epithelial.Nuclei.StdDevWidthPxI
	Lumen.Objects
50	Lumen. Obj ectsPct
	Lumen.MaxAreaPxI
	Lumen.MeanAreaPxI
	Lumen.MinAreaPxI
	Lumen.StdDevAreaPxI
55	
60	
65	

	Característica
	Lumen.SumAreaPxl
	Lumen.MaxAsymmetry
5	Lumen.MeanAsymmetry
	Lumen.MinAsymmetry
	Lumen.StdDevAsymmetry
	Lumen.MaxBorderlengthPxl
10	Lumen.MeanBorderlengthPxl
	Lumen.MinBorderlengthPxl
	Lumen.StdDevBorderlengthPxl
	Lumen.SumBorderlengthPxl
15	Lumen.MaxBrightness
	Lumen.MeanBrightness
	Lumen.MinBrightness
	Lumen.StdDevBrightness
20	Lumen.MaxCompactness
	Lumen.MeanCompactness
	Lumen.MinCompactness
	Lumen.StdDevCompactness
25	Lumen.MaxDensity
	Lumen.MeanDensity
	Lumen.MinDensity
	Lumen.StdDevDensity
30	Lumen.MaxDiff.ofenclosing.enclosede
	Lumen.MeanDiff.ofenclosing.enclosed
	Lumen.MinDiff.ofenclosing.enclosede
	Lumen.StdDevDiff.ofenclosing.enclos
35	Lumen.MaxEllipticFit
	Lumen.MeanEllipticFit
	Lumen.MinEllipticFit
	Lumen. StdDevEllipticFit
40	Lumen.MaxLengthPxl
	Lumen.MeanLengthPxl
	Lumen.MinLengthPxl
	Lumen.StdDevLengthPxl
45	Lumen.SumLengthPxl
	Lumen.MaxMax.Diff.
	Lumen.MeanMax.Diff.
	Lumen.MinMax.Diff.
50	Lumen.StdDevMax.Diff.
	Lumen.MaxMeanChannel1
	Lumen.MeanMeanChannel1
	Lumen.MinMeanChannel1
55	Lumen.StdDevMeanChannel1
	Lumen.MaxMeanChannel2
	Lumen.MeanMeanChannel2
	Lumen.MinMeanChannel2
60	Lumen.StdDevMeanChannel2
	Lumen.MaxMeanChannel3
	Lumen.MeanMeanChannel3
	Lumen.MinMeanChannel3
65	Lumen.StdDevMeanChannel3

Característica	
5	Lumen.MaxRadiusoflargestenclosedell
	Lumen.MeanRadiusoflargestenclosedel
	Lumen.MinRadiusoflargestenclosedell
10	Lumen. StdDevRadiusoflargestenclosed
	Lumen.MaxRadiusofsmallestenclosinge
	Lumen.MeanRadiusofsmallestenclosing
15	Lumen.MinRadiusofsmallestenclosinge
	Lumen. StdDevRadiusofsmallestenclosi
	Lumen.MaxStdevChannel1
20	Lumen.MeanStdevChannel1
	Lumen.MinStdevChannel1
	Lumen.StdDevStdevChannel1
25	Lumen.MaxStdevChannel2
	Lumen.MeanStdevChannel2
	Lumen.MinStdevChannel2
30	Lumen.StdDevStdevChannel2
	Lumen.MaxStdevChannel3
	Lumen.MeanStdevChannel3
35	Lumen.MinStdevChannel3
	Lumen.StdDevStdevChannel3
	Lumen.MaxWidthPxI
40	Lumen.MeanWidthPxI
	Lumen.MinWidthPxI
	Lumen.StdDevWidthPxI
45	Red.Blood. Cell. Objects
	Red.Blood.Cell.ObjectsPct
	Red.Blood.Cell.MaxAreaPxI
50	Red.Blood.Cell.MeanAreaPxI
	Red.Blood.Cell.MinAreaPxI
	Red.Blood.Cell.StdDevAreaPxI
55	Red.Blood.Cell.SumAreaPxI
	Red.Blood.Cell.MaxAsymmetry
	Red.Blood.Cell.MeanAsymmetry
60	Red.Blood.Cell.MinAsymmetry
	Red.Blood.Cell.StdDevAsymmetry
	Red.Blood.Cell.MaxBorderlengthPxI
65	Red.Blood.Cell.MeanBorderlengthPx
	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

Característica	
5	Red.Blood.Cell.MinDensity
	Red.Blood.Cell.StdDevDensity
	Red.Blood.Cell.MaxDiff.ofenclosing.
	Red.Blood.Cell.MeanDiff.ofenclosing
10	Red.Blood.Cell.MinDiff.ofenclosing.
	Red.Blood.Cell.StdDevDiff.ofenclosi
	Red.Blood.Cell.MaxEllipticFit
	Red.Blood.Cell.MeanEllipticFit
15	Red.Blood.Cell.MinEllipticFit
	Red.Blood.Cell.StdDevEllipticFit
	Red.Blood.Cell.MaxLengthPxI
	Red.Blood.Cell.MeanLengthPxI
20	Red.Blood.Cell.MinLengthPxI
	Red.Blood.Cell.StdDevLengthPxI
	Red.Blood.Cell.SumLengthPxI
	Red.Blood.Cell.MaxMax.Diff.
25	Red.Blood.Cell.MeanMax.Diff.
	Red.Blood.Cell.MinMax.Diff.
	Red.Blood.Cell.StdDevMax.Diff.
	Red.Blood.Cell.MaxMeanChannel1
30	Red.Blood.Cell.MeanMeanChannel1
	Red.Blood.Cell.MinMeanChannel1
	Red.Blood.Cell.StdDevMeanChannel1
	Red.Blood.Cell.MaxMeanChannel2
35	Red.Blood.Cell.MeanMeanChannel2
	Red.Blood.Cell.MinMeanChannel2
	Red.Blood.Cell.StdDevMeanChannel2
	Red.Blood.Cell.MaxMeanChannel3
40	Red.Blood.Cell.MeanMeanChannel3
	Red.Blood.Cell.MinMeanChannel3
	Red.Blood.Cell.StdDevMeanChannel3
	Red.Blood.Cell.MaxRadiusoflargesten
45	Red.Blood.Cell.MeanRadiusoflargeste
	Red.Blood.Cell.MinRadiusoflargesten
	Red.Blood.Cell.StdDevRadiusoflarges
	Red.Blood.Cell.MaxRadiusofsmalleste
50	Red.Blood.Cell.MeanRadiusofsmallest
	Red.Blood.Cell.MinRadiusofsmalleste
	Red.Blood.Cell.StdDevRadiusofsmalle
	Red.Blood.Cell.MaxStdevChannel1
55	Red.Blood.Cell.MeanStdevChannel1
	Red.Blood.Cell.MinStdevChannel1
	Red.Blood.Cell.StdDevStdevChannel1
	Red.Blood.Cell.MaxStdevChannel2
60	Red.Blood.Cell.MeanStdevChannel2
	Red.Blood.Cell.MinStdevChannel2
	Red.Blood.Cell.StdDevStdevChannel2
	Red.Blood.Cell.MaxStdevChannel3
65	Red.Blood.Cell.MeanStdevChannel3
	Red.Blood.Cell.MinStdevChannel3
	Red.Blood.Cell.StdDevStdevChannel3

Característica	
5	Red.Blood.Cell.MaxWidthPxI
	Red.Blood.Cell.MeanWidthPxI
	Red.Blood.Cell.MinWidthPxI
	Red.Blood.Cell.StdDevWidthPxI
10	Stroma.Objects
	Stroma.ObjectsPct
	Stroma.MaxAreaPxI
	Stroma.MeanAreaPxI
15	Stroma.MinAreaPxI
	Stroma.StdDevAreaPxI
	Stroma.SumAreaPxI
	Stroma.MaxAsymmetry
20	Stroma.MeanAsymmetry
	Stroma.MinAsymmetry
	Stroma.StdDevAsymmetry
	Stroma.MaxBorderlengthPxI
25	Stroma.MeanBorderlengthPxI
	Stroma.MinBorderlengthPxI
	Stroma.StdDevBorderlengthPxI
	Stroma.SumBorderlengthPxI
30	Stroma.MaxBrightness
	Stroma.MeanBrightness
	Stroma.MinBrightness
	Stroma.StdDevBrightness
35	Stroma.MaxCompactness
	Stroma.MeanCompactness
	Stroma.MinCompactness
	Stroma.StdDevCompactness
40	Stroma.MaxDensity
	Stroma.MeanDensity
	Stroma.MinDensity
	Stroma.StdDevDensity
45	Stroma.MaxDiff.ofenclosing.enclosed
	Stroma.MeanDiff.ofenclosing.enclose
	Stroma.MinDiff.ofenclosing.enclosed
	Stroma.StdDevDiff.ofenclosing.enclo
50	Stroma.MaxEllipticFit
	Stroma.MeanEllipticFit
	Stroma.MinEllipticFit
	Stroma.StdDevEllipticFit
55	Stroma.MaxLengthPxI
	Stroma.MeanLengthPxI
	Stroma.MinLengthPxI
	Stroma.StdDevLengthPxI
60	Stroma.SumLengthPxI
	Stroma.MaxMax.Diff.
	Stroma.MeanMax.Diff.
	Stroma.MinMax.Diff.
65	Stroma.StdDevMax.Diff.
	Stroma.MaxMeanChannel 1

Característica	
5	Stroma.MeanMeanChannel1
	Stroma.MinMeanChannel 1
	Stroma.StdDevMeanChannel1
	Stroma.MaxMeanChannel2
	Stroma.MeanMeanChannel2
10	Stroma.MinMeanChannel2
	Stroma.StdDevMeanChannel2
	Stroma.MaxMeanChannel3
	Stroma.MeanMeanChannel3
	Stroma.MinMeanChannel3
15	Stroma.StdDevMeanChannel3
	Stroma. MaxRadiusoflargestenclosedel
	Stroma.MeanRadiusoflargestenclosede
	Stroma.MinRadiusoflargestenclosedel
	Stroma. StdDevRadiusoflargestenclose
20	Stroma.MaxRadiusofsmallestenclosing
	Stroma.MeanRadiusofsmallestenclosin
	Stroma.MinRadiusofsmallestenclosing
	Stroma. StdDev Radiusofsmallestenclos
	Stroma.MaxStdevChannel 1
25	Stroma.MeanStdevChannel 1
	Stroma.MinStdevChannel 1
	Stroma.StdDevStdevChannel 1
	Stroma.MaxStdevChannel2
	Stroma.MeanStdevChannel2
30	Stroma.MinStdevChannel2
	Stroma.StdDevStdevChannel2
	Stroma.MaxStdevChannel3
	Stroma.MeanStdevChannel3
	Stroma.MinStdevChannel3
35	Stroma.StdDevStdevChannel3
	Stroma.MaxWidthPxI
	Stroma.MeanWidthPxI
	Stroma.MinWidthPxI
	Stroma.StdDevWidthPxI
40	Stroma.Nuclei .Objects
	Stroma.Nuclei.ObjectsPct
	Stroma.Nuclei.MaxAreaPxI
	Stroma.Nuclei .MeanAreaPxI
	Stroma.Nuclei.MinAreaPxI
45	Stroma.Nuclei.StdDevAreaPxI
	Stroma.Nuclei.SumAreaPxI
	Stroma.Nuclei.MaxAsymmetry
	Stroma.Nuclei .MeanAsymmetry
	Stroma.Nuclei.MinAsymmetry
50	Stroma.Nuclei.StdDevAsymmetry
	Stroma.Nuclei.MaxBorderlengthPxI
	Stroma.Nuclei.MeanBorderlengthPxI
	Stroma.Nuclei.MinBorderlengthPxI
	Stroma.Nuclei.StdDevBorderlengthPxI
55	Stroma.Nuclei.SumBorderlengthPxI
60	
65	

Característica	
5	Stroma.Nuclei .MaxBrightness
	Stroma.Nuclei .MeanBrightness
	Stroma.Nuclei .MinBrightness
	Stroma.Nuclei .StdDevBrightness
	Stroma.Nuclei .MaxCompactness
10	Stroma.Nuclei .MeanCompactness
	Stroma.Nuclei.MinCompactness
	Stroma.Nuclei .StdDevCompactness
	Stroma.Nuclei .MaxDensity
	Stroma.Nuclei .MeanDensity
15	Stroma.Nuclei.MinDensity
	Stroma. Nuclei .StdDevDensity
	Stroma.Nuclei.MaxDiff.ofenclosing.e
	Stroma.Nuclei .MeanDiff.ofenclosing.
	Stroma.Nuclei .MinDiff.ofenclosing.e
20	Stroma.Nuclei .StdDevDiff.ofenclosin
	Stroma.Nuclei.MaxEllipticFit
	Stroma.Nuclei.MeanEllipticFit
	Stroma.Nuclei .MinEllipticFit
	Stroma.Nuclei.StdDevEllipticFit
25	Stroma.Nuclei.MaxLengthPxl
	Stroma.Nuclei.MeanLengthPxl
	Stroma.Nuclei.MinLengthPxl
	Stroma.Nuclei.StdDevLengthPxl
	Stroma. Nuclei .SumLengthPxl
30	Stroma.Nuclei .MaxMax.Diff.
	Stroma.Nuclei.MeanMax.Diff.
	Stroma.Nuclei .MinMax.Diff.
	Stroma.Nuclei .StdDevMax.Diff.
	Stroma.Nuclei.MaxMeanChannel1
35	Stroma.Nuclei.MeanMeanChannel1
	Stroma.Nuclei.MinMeanChannel1
	Stroma.Nuclei.StdDevMeanChannel1
	Stroma.Nuclei.MaxMeanChannel2
	Stroma.Nuclei.MeanMeanChannel2
40	Stroma.Nuclei.MinMeanChannel2
	Stroma.Nuclei.StdDevMeanChannel2
	Stroma.Nuclei.MaxMeanChannel3
	Stroma.Nuclei.MeanMeanChannel3
	Stroma.Nuclei.MinMeanChannel3
45	Stroma.Nuclei.StdDevMeanChannel3
	Stroma.Nuclei.MaxRadiusoflargestenc
	Stroma.Nuclei.MeanRadiusoflargesten
	Stroma.Nuclei.MinRadiusoflargestenc
	Stroma.Nuclei.StdDevRadiusoflargest
50	Stroma.Nuclei.MaxRadiusofsmallesten
	Stroma.Nuclei.MeanRadiusofsmalleste
	Stroma.Nuclei .MinRadiusofsmallesten
	Stroma.Nuclei.StdDevRadiusofsmalles
	Stroma.Nuclei.MaxStdevChannel1
55	Stroma.Nuclei.MeanStdevChannel1
	Stroma.Nuclei.MinStdevChannel1
60	
65	

Característica	
5	Stroma.Nuclei.StdDevStddevChannel1
	Stroma.Nuclei.MaxStddevChannel2
	Stroma.Nuclei.MeanStddevChannel2
	Stroma.Nuclei.MinStddevChannel2
	Stroma.Nuclei.StdDevStddevChannel2
10	Stroma.Nuclei.MaxStddevChannel3
	Stroma.Nuclei.MeanStddevChannel3
	Stroma.Nuclei.MinStddevChannel3
	Stroma.Nuclei.StdDevStddevChannel3
	Stroma.Nuclei.MaxWidthPx1
15	Stroma.Nuclei.MeanWidthPx1
	Stroma.Nuclei.MinWidthPx1
	Stroma.Nuclei.StdDevWidthPx1
	C2EN
	EN2SN
20	L2Core
	C2L
	CEN2L
25	
Tabla 2. Características Morfométricas Secuencia de Comandos v2.0 (350 características)	
Característica	
30	Área media de artefacto en pxl
	StdDev de área de artefacto en pxl
	Asimetría media de artefacto
	StdDev de asimetría de artefacto
	Índice medio de borde de artefacto
35	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
40	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
45	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
50	Longitud/ancho media de artefacto
	StdDev de longitud/ancho de artefacto
	Dirección principal media de artefacto
	StdDev de dirección principal de artefacto
	Dif. máx. media de artefacto
55	StdDev de dif. máx. de artefacto
60	
65	

Característica	
5	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
10	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
15	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
20	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
25	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
30	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
35	Asimetría media de citoplasma
	StdDev de asimetría de citoplasma
	Índice de borde medio de citoplasma
	StdDev de índice de borde de citoplasma
40	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
45	Compacidad media de citoplasma
	StdDev de Compacidad de citoplasma
	Densidad media de citoplasma
	StdDev de densidad de citoplasma
50	Dif. media de citoplasma de elipse de cierre/cerrada
	StdDev de dif. de citoplasma de elipse de cierre/cerrada
	Ajuste elíptico medio de citoplasma
	StdDev de ajuste elíptico de citoplasma
55	Longitud media de citoplasma en pxl
	StdDev de longitud de citoplasma en pxl
	Longitud/ancho media de citoplasma
	StdDev de longitud/ancho de citoplasma
60	Dirección principal media de citoplasma
	StdDev de dirección principal de citoplasma
	Diferencia máx. media de citoplasma
	StdDev de diferencia máx. de citoplasma
65	Media de media de citoplasma del canal 1
	StdDev de media de citoplasma del canal 1
	Media de media de citoplasma del canal 2

Característica	
5	StdDev de media de citoplasma del canal 2
	Media de media de citoplasma del canal 3
	StdDev de media de citoplasma del canal 3
10	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
15	StdDev de radio de citoplasma de la elipse cerrada más pequeña
	Ajuste rectangular medio de citoplasma
	StdDev de ajuste rectangular de citoplasma
20	Índice de forma media de citoplasma
	StdDev de índice de forma de citoplasma
	Stddev de media de citoplasma del canal 1
25	Stddev de StdDev de citoplasma del canal 1
	Stddev de media de citoplasma del canal 2
	Stddev de StdDev de citoplasma del canal 2
30	Stddev de media de citoplasma del canal 3
	Stddev de StdDev de citoplasma del canal 3
	Ancho medio de citoplasma en pxl
35	StdDev de ancho de citoplasma en pxl
	Área media de núcleos epiteliales en pxl
	StdDev de área de núcleos epiteliales en pxl
40	Asimetría media de núcleos epiteliales
	StdDev de asimetría de núcleos epiteliales
	Índice de borde medio de núcleos epiteliales
45	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
50	Brillo medio de núcleos epiteliales
	StdDev de brillo de núcleos epiteliales
	Compacidad media de núcleos epiteliales
55	StdDev de compacidad de núcleos epiteliales
	Densidad media de núcleos epiteliales
	StdDev de densidad de núcleos epiteliales
60	Dif. media de núcleos epiteliales de elipse de cierre/cerrada
	StdDev de dif. de núcleos epiteliales de elipse de cierre/cerrada
	Ajuste elíptico medio de núcleos epiteliales
65	StdDev de ajuste elíptico de núcleos epiteliales
	Longitud media de núcleos epiteliales en pxl
	StdDev de longitud de núcleos epiteliales en pxl
70	Longitud/ancho media de núcleos epiteliales
	StdDev de longitud/ancho de núcleos epiteliales
	Dirección principal media de núcleos epiteliales
75	StdDev de dirección principal de núcleos epiteliales
	Dif. máx. media de núcleos epiteliales
	StdDev de dif. máx. de núcleos epiteliales
80	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
85	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

Característica	
5	Radio medio de núcleos epiteliales de la elipse cerrada más grande
	StdDev de radio de núcleos epiteliales de la elipse cerrada más grande
	Radio medio de núcleos epiteliales de la elipse cerrada más pequeña
	StdDev de radio de núcleos epiteliales de la elipse cerrada más pequeña
10	Ajuste rectangular medio de núcleos epiteliales
	StdDev de ajuste rectangular de núcleos epiteliales
	Índice de forma media de núcleos epiteliales
	StdDev de índice de forma de núcleos epiteliales
15	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
20	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
25	Área media de lumen en pxl
	StdDev de área de lumen en pxl
	Asimetría media de lumen
	StdDev de asimetría de lumen
30	Índice de borde medio de lumen
	StdDev de índice de borde de lumen
	Longitud media de borde de lumen en pxl
	StdDev de longitud de borde de lumen en pxl
35	Brillo medio de lumen
	StdDev de brillo de lumen
	Compacidad media de lumen
	StdDev de compacidad de lumen
40	Densidad media de lumen
	StdDev de densidad de lumen
	Dif. media de lumen de elipse de cierre/cerrada
	StdDev de dif. de lumen de elipse de cierre/cerrada
45	Ajuste elíptico medio de lumen
	StdDev de ajuste elíptico de lumen
	Longitud media de lumen en pxl
	StdDev de longitud de lumen en pxl
50	Longitud/ancho medio de lumen
	StdDev de longitud/ancho de lumen
	Dirección principal media de lumen
	StdDev de dirección principal de lumen
55	Diferencia máx. media de lumen
	StdDev de diferencia máx. de lumen
	Media de media de lumen del canal 1
	StdDev de media de lumen del canal 1
60	Media de media de lumen del canal 2
	StdDev de media de lumen del canal 2
	Media de media de lumen del canal 3
	StdDev de media de lumen del canal 3
65	Radio medio de lumen de la elipse cerrada más grande
	StdDev de radio de lumen de la elipse cerrada más grande
	Radio medio de lumen de la elipse cerrada más pequeña

Característica	
5	StdDev de radio de lumen de la elipse cerrada más pequeña
	Ajuste Rectangular Medio de lumen
	StdDev de ajuste rectangular de lumen
	Índice de forma media de lumen
	StdDev de índice de forma de lumen
10	Stddev de media de lumen del canal 1
	Stddev de StdDev de lumen del canal 1
	Stddev de media de lumen del canal 2
	Stddev de StdDev de lumen del canal 2
	Stddev de media de lumen del canal 3
15	Stddev de StdDev de lumen del canal 3
	Ancho Medio de lumen en pxl
	StdDev de ancho de lumen en pxl
	Área media de estroma en pxl
	StdDev de área de estroma en pxl
20	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
25	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
30	Densidad media de estroma
	StdDev de densidad de estroma
	Dif. media de estroma de elipse de cierre/cerrada
	StdDev de dif. de estroma de elipse de cierre/cerrada
	Ajuste elíptico medio de estroma
35	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
40	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
45	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
50	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
55	StdDev de ajuste rectangular de estroma
60	
65	

Característica

	Índice de forma media de estroma
	StdDev de índice de forma de estroma
5	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
10	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
15	Área media de núcleos estromales en pxl
	StdDev de área de núcleos estromales en pxl
	Asimetría media de núcleos estromales
	StdDev de asimetría de núcleos estromales
20	Índice de borde medio de núcleos estromales
	StdDev de índice de borde de núcleos estromales
	Longitud de borde medio de núcleos estromales en pxl
	StdDev de longitud de borde de núcleos estromales en pxl
25	Brillo medio de núcleos estromales
	StdDev de brillo de núcleos estromales
	Compacidad media de núcleos estromales
	StdDev de compacidad de núcleos estromales
30	Densidad media de núcleos estromales
	StdDev de densidad de núcleos estromales
	Dif. media de núcleos estromales de elipse de cierre/cerrada
	StdDev de dif. de núcleos estromales de elipse de cierre/cerrada
35	Ajuste elíptico medio de núcleos estromales
	StdDev de ajuste elíptico de núcleos estromales
	Longitud media de núcleos estromales en pxl
	StdDev de longitud de núcleos estromales en pxl
40	Longitud/ancho medio de núcleos estromales
	StdDev de longitud/ancho de núcleos estromales
	Dirección principal media de núcleos estromales
	StdDev de dirección principal de núcleos estromales
45	Dif. máx. media de núcleos estromales
	StdDev de dif. máx. de núcleos estromales
	Media de media de núcleos estromales del canal 1
	StdDev de media de núcleos estromales del canal 1
50	Media de media de núcleos estromales del canal 2
	StdDev de media de núcleos estromales del canal 2
	Media de media de núcleos estromales del canal 3
	StdDev de media de núcleos estromales del canal 3
55	Radio medio de núcleos estromales de la elipse cerrada más grande
	StdDev de Radio de núcleos estromales de la elipse cerrada más grande
	Radio medio de núcleos estromales de la elipse cerrada más pequeña
	StdDev de Radio de núcleos estromales de la elipse cerrada más pequeña
60	Ajuste rectangular medio de núcleos estromales
	StdDev de ajuste rectangular de núcleos estromales
	Índice de forma media de núcleos estromales
	StdDev de índice de forma de núcleos estromales
65	Stddev de media de núcleos estromales del canal 1

Característica	
5	Stddev de StdDev de núcleos estromales del canal 1
	Stddev de media de núcleos estromales del canal 2
	Stddev de StdDev de núcleos estromales del canal 2
	Stddev de media de núcleos estromales del canal 3
	Stddev de StdDev de núcleos estromales del canal 3
10	Ancho Medio de núcleos estromales en pxl
	StdDev de ancho de núcleos estromales en pxl
	Área de artefactos en pxl
	Área de citoplasma en pxl
	Área de núcleos epiteliales en pxl
15	Área de lumen en pxl
	Área de glóbulos rojos en pxl
	Área de estroma en pxl
	Área de núcleos estromales en pxl
	Número de objetos de artefacto
20	Número de objetos de citoplasma
	Número de objetos de núcleos epiteliales
	Numero de objetos de lumen
	Número de objetos de glóbulos rojos
	Número de objetos de estroma
25	Número de objetos de núcleos estromales
	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
30	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
35	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
40	Dif. media de glóbulos rojos de elipse de cierre/cerrada
	StdDev de dif. media de glóbulos rojos de elipse de cierre/cerrada
	Ajuste elíptico medio de glóbulos rojos
	StdDev de ajuste elíptico de glóbulos rojos
	Longitud media de glóbulos rojos en pxl
45	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
50	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
55	StdDev de media de glóbulos rojos del canal 2
60	
65	

Característica

5	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
10	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
15	Ajuste rectangular medio de glóbulos rojos
	StdDev de ajuste rectangular de glóbulos rojos
	Índice de forma media de glóbulos rojos
20	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
25	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
30	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 17. Características morfométricas - Secuencia de Comandos v5.0 (38 características)

	Nombre de la Característica	Descripción
30	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
35	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
40	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
50	EpithelialNucleiMeanAreaPxI0101	Valor medio del área de núcleos epiteliales, 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
55	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
60	EpitheNucleiMeanStddevChann60122	Desviación estándar de la intensidad media del núcleo epitelial en el canal de color azul
	EpitheliaNucleiStddevAreaPxI0124	Desviación estándar del área de núcleos epiteliales
65		

	Nombre de la Característica	Descripción
5	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
10	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
15	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
20	StromaMeanStddevChannel40331	Desviación estándar de la intensidad media del estroma en el canal de color rojo
	StromaMeanStddevChannel50332	Desviación estándar de la intensidad del estroma en el canal de color verde
	StromaMeanStddevChannel60333	Desviación estándar de la intensidad del estroma en el canal de color azul
25	AreaOfCytoplasmPx10345	Área total de citoplasma, píxeles
	AreaOfEpithelialNucleiPx10350	Área total de núcleos epiteliales, píxeles
	AreaOfLumenPx10357	Área total de lúmenes, píxeles
30	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, %
35	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 18. Características morfométricas

	Nombre de la Característica	Descripción
50	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
55	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

	Nombre de la Característica	Descripción
	CytoplasmStddevMeanChannel40081	Desviación estándar de la intensidad media de citoplasma con el filtro rojo
5	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
10	EpithelNucleiMeanMeanChanne40112	Media de la intensidad de núcleo epitelial con el filtro rojo
	EpithelNucleiMeanMeanChanne50113	Media de la intensidad de núcleo epitelial con el filtro verde
15	EpithelNucleiMeanMeanChanne60114	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
20	EpitheNucleiMeanStddevChann60122	Media de la desviación estándar de la intensidad de núcleo epitelial con el filtro azul
	EpitheliaNucleiStddevAreaPxI0124	Desviación estándar del área de núcleos epiteliales
25	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
30	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
35	StromaMeanMeanChannel50263	Media de la intensidad de estroma con el filtro verde
	StromaMeanMeanChannel60264	Media de la intensidad de estroma con el filtro azul
40	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
45	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
50	StromaStddevMeanChannel60333	Desviación estándar de la intensidad de estroma con el filtro azul
	AreaCytopdivTotTissueArea	Área de citoplasma respecto de área de tejido, %
55	AreaEpitNucdivTotTissueArea	Área de núcleos epiteliales respecto de área de tejido, %
	AreaLumendivTotTissueArea	Área de lumen respecto al área de tejido, %
60	AreaRBCdivTotTissueArea	Área de glóbulos rojos respecto de área de tejido, %
	AreaStromadivTotTissueArea	Área de estroma respecto de área de tejido, %

Tabla 19. Información clínica del Ejemplo 8

	Característica	Entrenamiento	Validación
	N	342	340
5	PSA preoperatorio (ng/ml)		
	Media	10,3	10,8
	PSA preoperatorio (ng/ml)		
	Mediana	7,5	7,9
10	Rango	0,8-68,5	1,01-00,0
	Afectación de ganglios linfáticos		
	Negativo	339 (99,1 %)	335 (98,5 %)
	Positivo	3 (0,9 %)	5 (1,5 %)
15	Afectación de las vesículas seminales		
	No	320 (93,6 %)	321 (94,4 %)
	Sí	22 (6,4 %)	19 (5,6 %)
	Márgenes quirúrgicos		
20	Negativo	219 (64,0 %)	240 (71,0 %)
	Positivo	123 (36,0 %)	100 (29,0 %)
	Afectación extracapsular		
	No	245 (71,6,0 %)	254 (74,7 %)
25	Sí	97 (28,4 %)	86 (25,3 %)
	Grado de Gleason dominante de la biopsia		
	1	1 (0,3 %)	2 (0,6 %)
	2	17 (5,0 %)	23 (6,8 %)
30	3	279 (81,6 %)	273 (80,3 %)
	4	45 (13,2 %)	42 (12,4 %)
	5	0 (0,0 %)	0 (0,0 %)
	Puntuación de Gleason de la biopsia		
35	2	0 (0,0 %)	1 (0,3 %)
	3	1 (0,3 %)	2 (0,6 %)
	4	11 (3,2 %)	8 (2,4 %)
	5	23 (6,7 %)	30 (8,8 %)
40	6	190 (55,6 %)	187 (55,0 %)
	7	101 (29,5 %)	94 (27,7 %)
	8	15 (4,4 %)	13 (3,8 %)
	9	1 (0,3 %)	5 (1,5 %)
45	Grado de Gleason dominante de la muestra		
	2	8 (2,3 %)	5 (1,5 %)
	3	283 (82,8 %)	276 (81,2 %)
	4	49 (14,3 %)	55 (16,2 %)
50	5	2 (0,6 %)	4 (1,2 %)
	Puntuación de Gleason de la muestra		
	5	9 (2,6 %)	15 (4,4 %)
	6	116 (33,9 %)	110 (32,4 %)
55	7	193 (56,4 %)	187 (55,0 %)
	8	18 (5,3 %)	17 (5,0 %)
	9	6 (1,8 %)	10 (2,9 %)
	10	0 (0,0 %)	1 (0,3 %)
60			

Tabla 20. Características evaluadas y seleccionadas del Ejemplo 9

65

	Característica	Función de agregación	CI univariante*	Peso en el modelo de CF
<u>Características Clínicas</u>				
5	Etapa clínica TNM		032917	
	PSA preoperatorio		030974	
	Grado de Gleason dominante de la biopsia		03265	
10	Puntuación de Gleason de la biopsia		031104	
	Grado de Gleason dominante de la prostatectomía		0,26804	-16,4
	Suma de Gleason de la prostatectomía		031311	
	Extensión extracapsular		03358	
15	Estado de la vesícula seminal		037022	
	Afectación de ganglios linfáticos		0,42033	-25,938
	Estado de los márgenes quirúrgicos		037596	
<u>Características morfológicas</u>				
20	Área relativa de núcleos epi		03778	
	Característica de textura nuclear epitelial 1 norma		033503	
	Característica de textura nuclear epitelial 2		039107	
25	Característica de textura nuclear epitelial 3		036808	
	Característica de textura de citoplasma 1		0,72432	
	Característica de textura de citoplasma 2		0,74648	20,728
	Característica de textura de citoplasma 3		0,73593	
30	Área relativa de lumen		0,67302	8,9663
	Media de la longitud del borde del lumen		0,7646	17,217
	Densidad de núcleos epiteliales aislados		030897	
	Núcleos epiteliales bien definidos		034273	
35	Área nuclear epitelial aislada norma		0,3098	
			0,42957	
<u>Características moleculares</u>				
	Área relativa de núcleos epi	min	0,54152	
40	Área relativa de núcleos epi AR+	media	0,3778	
	Área relativa de células epi AMACR+	max	0,45243	
	Área relativa de núcleos epi AR+AMACR+/ AR+	max	0,50219	
	Área relativa de núcleos epi AR+AMACR+/-	max	0,4362	
45	Suma normalizada de intensidad de AR de Grupo 3-9/nuclear total	media		
	área		0,33266	
	Suma normalizada de intensidad de AR de Grupo 1-2/nuclear total	max		
50	área		0,60088	
	Suma normalizada de intensidad de AR de Grupo 3-5/nuclear total	media		
	área		0,38799	
55	Suma normalizada de intensidad de AR de Grupo 6-10/nuclear total	mín		
	área		0,45605	
	Intensidad media de AR en núcleos epi AMACR-	media	0,34475	-5,2075
60	Intensidad media de AR en núcleos epi AMACR+	media	0,30174	
	Intensidad total de AR en núcleos epi	media	0,32532	
	Intensidad de AR normalizada en núcleos epi	media	0,3406	
	Intensidad total de AR en núcleos epi AMACR+	max	0,42957	
65	Intensidad total de AR en núcleos epi AMACR-	max	0,44616	

Característica	Función de agregación CI univariante*		Peso en el modelo de CF
Características moleculares			
Intensidad total de AR x área, normalizada	media	0,34167	
Intensidad total de AR positivo x área, normalizada	media	0,33562	
Área total de intensidad de AR+ dentro de núcleos epi	media	0,37898	
* Índice de concordancia como un predictor de variable única de fallo clínico en el conjunto de entrenamiento.			

Tabla 21. Características clínicas de los pacientes del Ejemplo 9

		Entrenamiento		Validación	Validación
Número de pacientes PSA preoperatorio		373 Frecuencia	Porcentaje (%)	385 Frecuencia	Porcentaje (%)
	<4	26	6,97	30	7,79
	[4,10)	216	57,91	215	55,84
	[10,20)	92	24,66	101	26,23
	>20	39	10,46	39	10,13
Etapa clínica TNM		Frecuencia	Porcentaje	Frecuencia	Porcentaje
	1	3	0,8	4	1,04
	2	187	50,13	187	48,57
	3	69	18,5	61	15,84
	4	36	9,65	30	7,79
	5	72	19,3	98	25,45
	6	6	1,61	5	1,3
Puntuación de Gleason de la biopsia		Frecuencia	Porcentaje	Frecuencia	Porcentaje
	2	0	0	1	0,26
	3	1	0,27	2	0,52
	4	11	2,95	9	2,34
	5	23	6,17	31	8,05
	6	202	54,16	205	53,25
	7	113	30,29	108	28,05
	8	22	5,9	23	5,97
	9	1	0,27	6	1,56
Suma de Gleason de la prostatectomía		Frecuencia	Porcentaje	Frecuencia	Porcentaje
	5	9	2,41	16	4,16
	6	119	31,9	116	30,13
	7	214	57,37	212	55,06
	8	21	5,63	23	5,97
	9	10	2,68	17	4,42
	10	0	0	1	0,26
Estado de la vesícula seminal		Frecuencia	Porcentaje	Frecuencia	Porcentaje
	0	342	91,69	349	90,65
	1	31	8,31	36	9,35
Estado de los márgenes quirúrgicos		Frecuencia	Porcentaje	Frecuencia	Porcentaje
	0	234	62,73	258	67,01
	1	139	37,27	127	32,99
Extensión extracapsular		Frecuencia	Porcentaje	Frecuencia	Porcentaje

	Número de pacientes PSA preoperatorio	Entrenamiento	Entrenamiento	Validación	Validación
		373 Frecuencia	Porcentaje (%)	385 Frecuencia	Porcentaje (%)
5	0	254	68,1	276	71,69
	1	29	7,77	33	8,57
	2	90	24,13	76	19,74
10	Afectación de ganglios linfáticos	Frecuencia	Porcentaje	Frecuencia	Porcentaje
	0	364	97,59	373	96,88
	1	9	2,41	12	3,12
	censurado	Frecuencia	Porcentaje	Frecuencia	Porcentaje
15	0 (sin CF aún)	340	91,15	356	92,47
	1 (CF)	33	8,85	29	7,53

Tabla 22. Características morfométricas – Secuencia de comandos v6.0 (27 características)

	Nombre de la Característica	Descripción
20	Cytoplasm Mean MeanChannel40058	Media del valor medio de intensidad de citoplasma con el filtro rojo
	Cytoplasm Mean MeanChannel50059	Media del valor medio de intensidad de citoplasma con el filtro verde
25	Cytoplasm Mean MeanChannel60060	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
30	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
35	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
40	CytoplasmStddevMeanChannel60083	Desviación estándar de la intensidad media de citoplasma con el filtro azul
	EpitheNucleiMeanStddevChann40142	Media de la desviación estándar de la intensidad de núcleo epitelial con el filtro rojo
	EpitheNucleiMeanStddevChann50143	Media de la desviación estándar de la intensidad de núcleo epitelial con el filtro verde
45	EpitheNucleiMeanStddevChann60144	Media de la desviación estándar de la intensidad de núcleo epitelial con el filtro azul
	EpitheNucleiStddevMeanChann40157	Desviación estándar de la intensidad media de núcleo epitelial con el filtro rojo
50	EpitheNucleiStddevMeanChann50158	Desviación estándar de la intensidad media de núcleo epitelial con el filtro verde
	EpitheNucleiStddevMeanChann60159	Desviación estándar de la intensidad media de núcleo epitelial con el filtro azul
55	StromaMeanStddevChannel40310	Media de la desviación estándar de intensidad de estroma con el filtro rojo
	StromaMeanStddevChannel50311	Media de la desviación estándar de la intensidad de estroma con el filtro verde
60	StromaMeanStddevChannel60312	Media de la desviación estándar de la intensidad de estroma con el filtro azul
	AreaCytopdivTotTissueArea	Área de citoplasma respecto de área de tejido, %
	Area EpitNucdivTotTissueArea	Área de núcleos epiteliales respecto de área de tejido, %
65	AreaLumendivTotTissueArea	Área de lumen respecto al área de tejido, %

		Entrenamiento	Entrenamiento	Validación	Validación
Número de pacientes PSA preoperatorio		373		385	
5		Frecuencia	Porcentaje (%)	Frecuencia	Porcentaje (%)
	AreaStromadivTotTissueArea	Área de estroma respecto de área de tejido, %			
10	AreaOfLowNucleiDensityNorm	La densidad de los núcleos se define como el área relativa de los núcleos epiteliales dentro del círculo centrado en el objeto			
	AreaOfMedNucleiDensityNorm	La característica representa el área relativa de los objetos de imagen que tienen hasta 40 % del área de núcleos epiteliales			
15	AreaOfHighNucleiDensityNorm	La característica representa el área relativa de los objetos de imagen que tienen más de 40 % y menos de 60 % del área de			
	DensityPercent_Bins3_4	La característica representa el área relativa de los objetos de imagen que tienen más de 60 % del área de núcleos			
20	DensityPercent_Bin1	Área relativa del objeto de imagen que tiene más de 20 % y menos de 40 % de núcleos epiteliales dentro de un círculo de			
		Área relativa del objeto de imagen que tiene menos de 10 % de núcleos epiteliales dentro de un círculo de 27 píxeles de			

REIVINDICACIONES

1. Aparato para evaluar un riesgo de recurrencia del cáncer de próstata en un paciente, el aparato comprende:

(a) uno o más dispositivos de almacenamiento de datos configurados para almacenar un conjunto de datos del paciente y un conjunto de datos de la población, en donde el conjunto de datos de la población incluye datos relacionados con pacientes donde se conoció el resultado final de salud (sin censura) y datos relacionados con pacientes donde el resultado final de salud no se conoció (censurado); y tanto el conjunto de datos de la población como el conjunto de datos del paciente incluyen un conjunto común de valores de datos correspondientes al menos a una puntuación de Gleason de la biopsia; invasión de las vesículas seminales; extensión extracapsular; PSA preoperatorio; Grado de Gleason dominante de la prostatectomía; el área relativa de los núcleos epiteliales AR+; una medición morfométrica de la textura de los núcleos epiteliales tumorales y el citoplasma epitelial, en donde los datos en al menos el conjunto de datos del paciente se derivan en parte de una imagen de una muestra de tejido sometida a inmunofluorescencia; y
(b) un ordenador configurado para:

(i) generar, a partir del conjunto de datos de la población, una aplicación de pronóstico configurada para generar uno o más valores correspondientes al riesgo de recurrencia de cáncer de próstata en un paciente al realizar la regresión de vectores de soporte utilizando una primera función de pérdida en el conjunto común de valores de datos en el conjunto de datos censurado y la regresión de vectores de soporte utilizando una segunda función de pérdida en el conjunto común de valores de datos en el conjunto de datos sin censura, y almacenar la aplicación de pronóstico dentro de la memoria del ordenador;

(ii) acceder al conjunto de datos del paciente;

(iii) evaluar el conjunto común de valores de datos en el conjunto de datos del paciente con la aplicación de pronóstico para generar un valor correspondiente a la probabilidad de recurrencia del cáncer de próstata en el paciente, y

(iv) generar un archivo digital adecuado para la transmisión a un sistema informático remoto que incluya al menos los datos del sujeto y los uno o más valores indicativos del riesgo de recurrencia de cáncer de próstata,

caracterizado porque la imagen de la muestra de tejido se analiza mediante el uso de al menos una función de análisis de imágenes de árbol cuadrático y una resegmentación multirresolución para diferenciar la imagen de la muestra de tejido en objetos de fondo y que no son de fondo, donde los objetos de fondo tienen una intensidad promedio de píxeles por debajo de un umbral predeterminado, y en donde los objetos que no son de fondo comprenden compartimentos de biomarcadores dentro de los cuales las señales de los biomarcadores se discriminan del fondo mediante la umbralización de la intensidad, en donde los umbrales de intensidad para la discriminación de las señales son una función de las características de imagen de los objetos de fondo.

2. El aparato de la reivindicación 1, en donde el conjunto de datos de la población y el conjunto de datos del paciente incluyen al menos una medición morfométrica adicional.

3. El aparato de la reivindicación 2, en donde el conjunto de datos de la población y el conjunto de datos del paciente incluyen un valor correspondiente a la intensidad promedio de AR en los núcleos epiteliales AR+/AMACR-.

Figura 1A

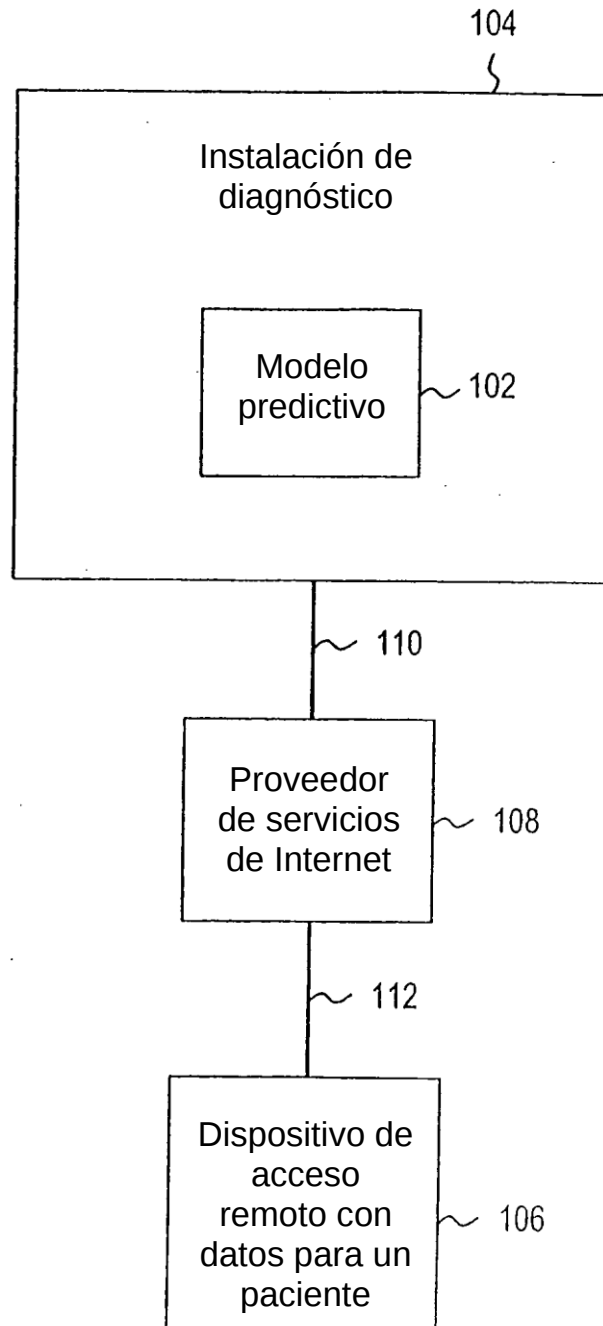


Figura 1B

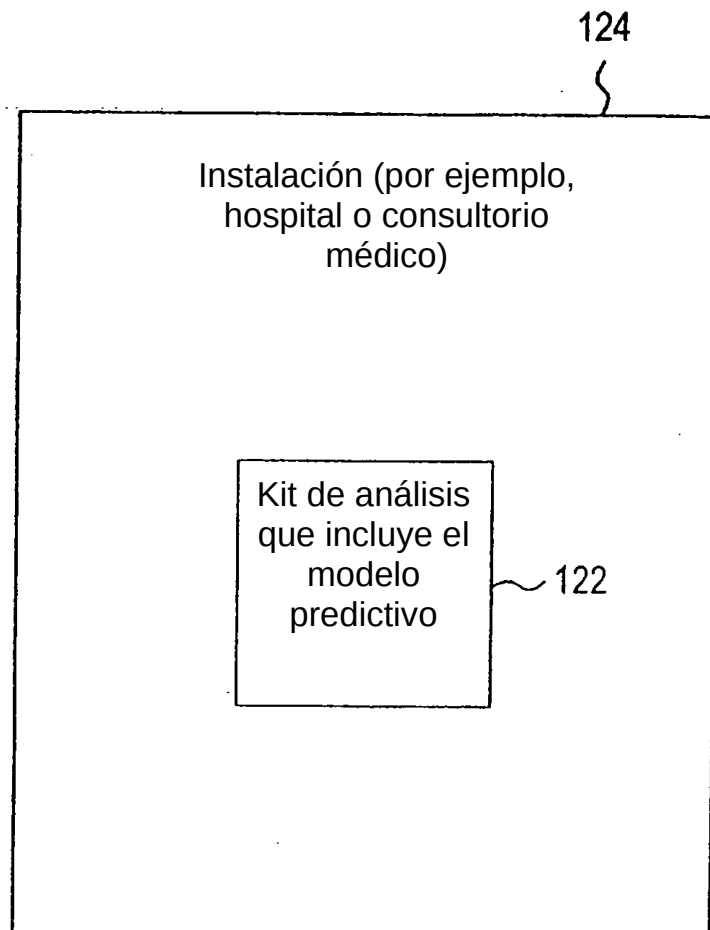


Figura 1C

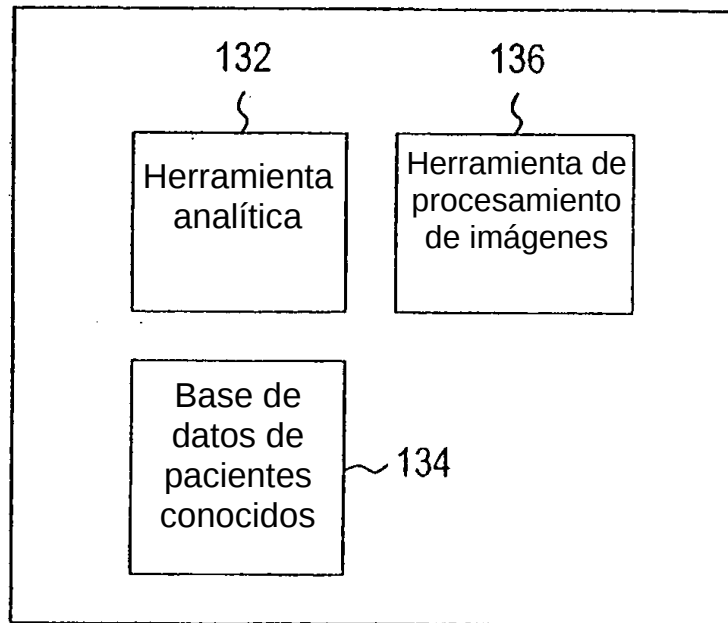


Figura 2

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

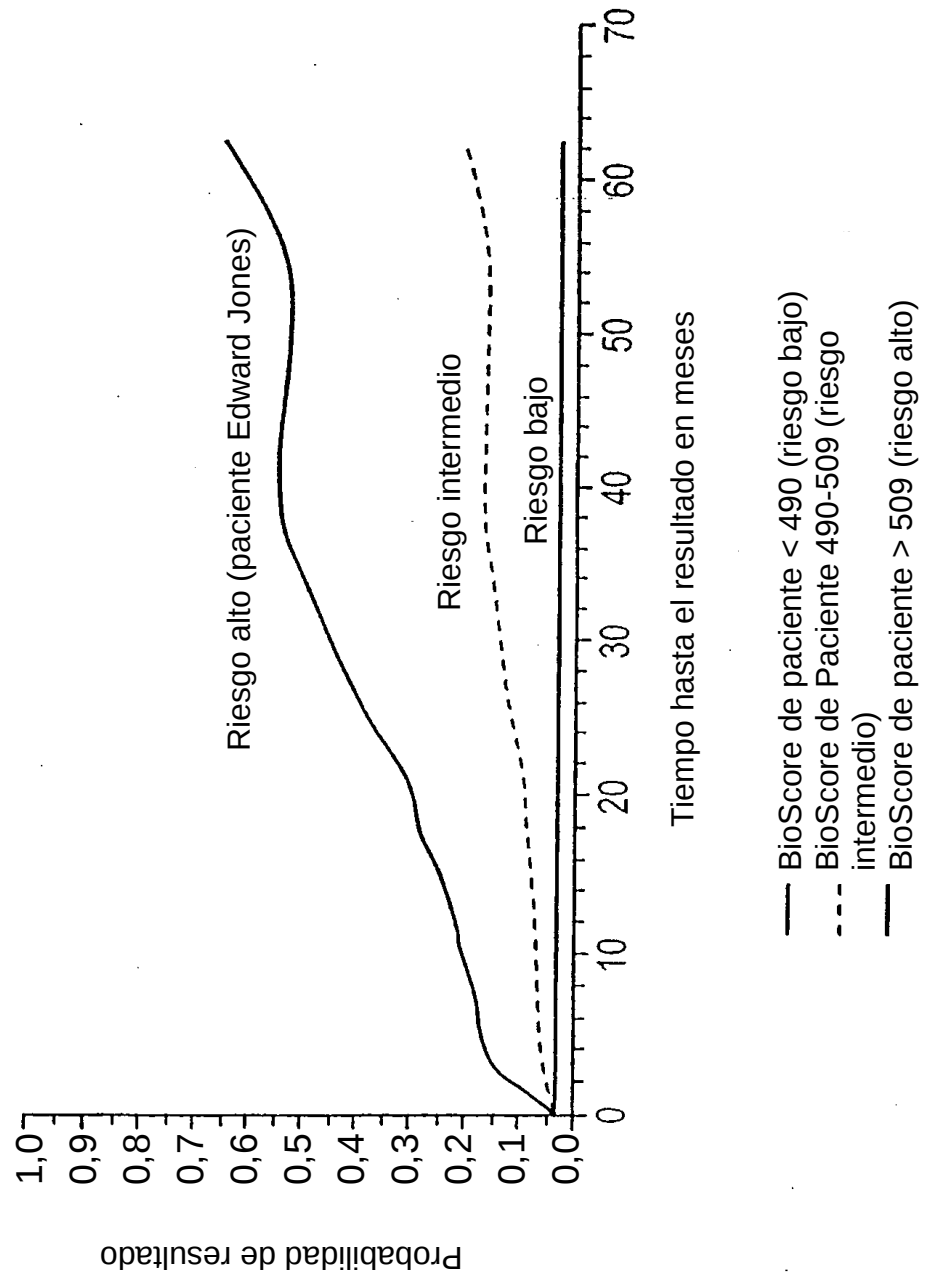


Figura 3

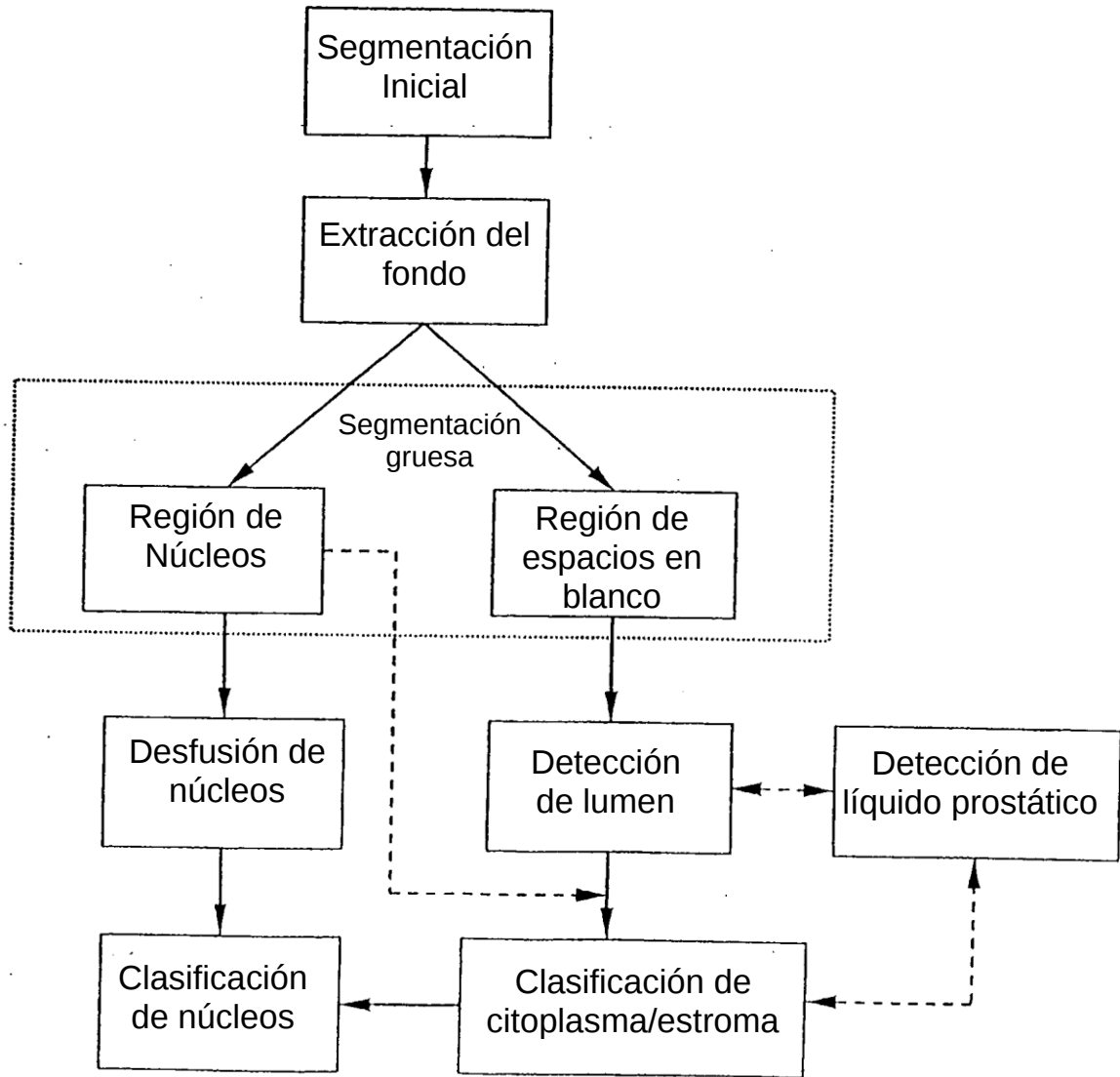


Figura 4

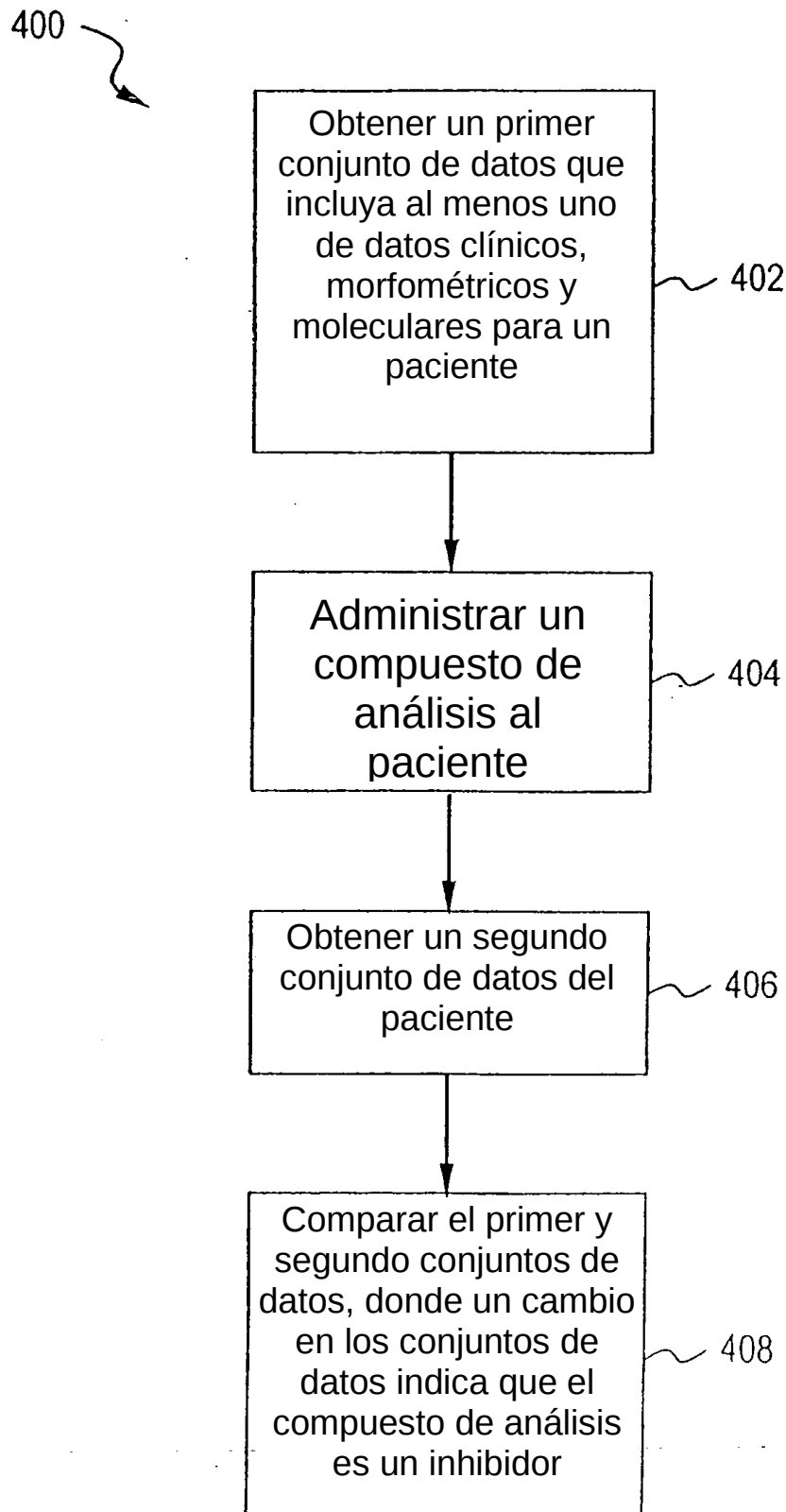


Figura 5A

Imagen de tejido prostático sano

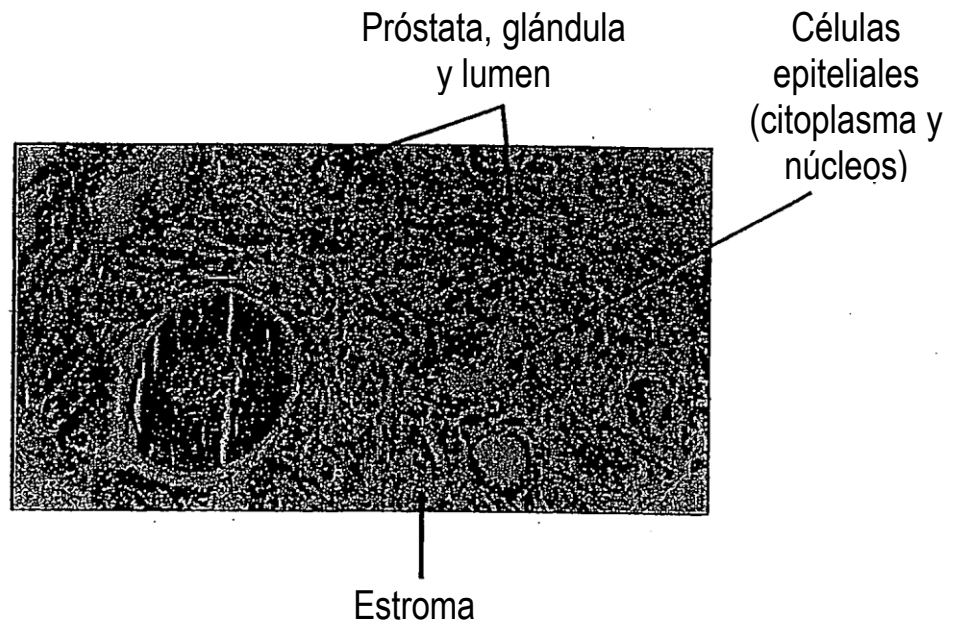


Figura 5B

Imagen de tejido prostático anormal

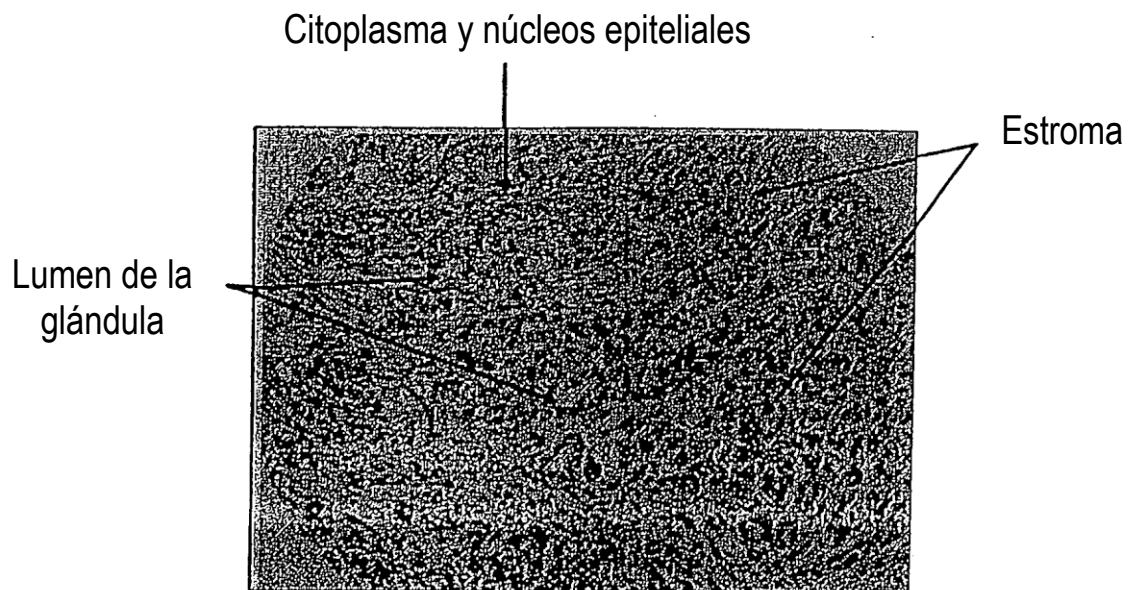
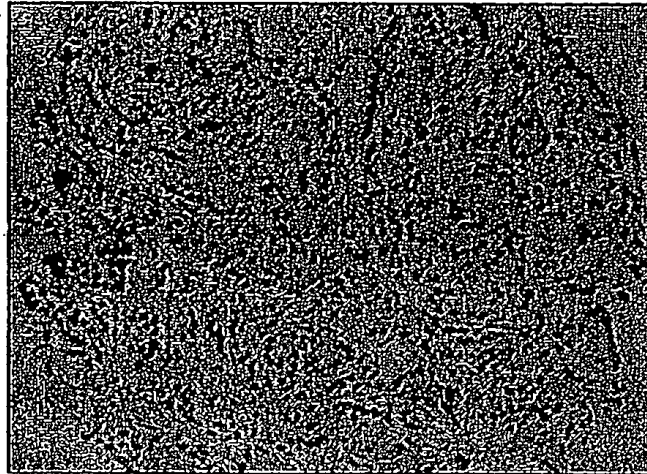


Figura 6

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

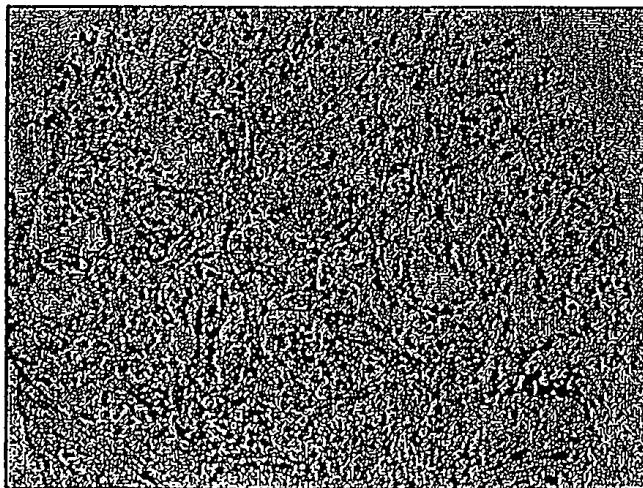
Característica	Descripción
Clínica	Puntuación de Gleason de la biopsia Raza Etapa UICC Resultado de ploidía Resultado de DRE Afectación de ganglios linfáticos Grado de Gleason dominante de la biopsia Porcentaje de ploidía en fase S Puntuación de Gleason postoperatoria Etapa TNM Grado de Gleason dominante postoperatorio Edad Afectación de las vesículas seminales PSA preoperatorio Porcentaje de fracción de ploidía Afectación de márgenes quirúrgicos Afectación extracapsular
Molecular	AR-tumor AR-glándula CD34-tumor/PIN Ki67-tumor 2 CD45-Pin 3 CD34-tumor/estroma Ki-67-tumor 3 p27-tumor C14-PIN CD34-tumor PSA-Glándula PSMA-PIN CD34-PIN/estroma CD45-tumor 3
Morfométrica	Glóbulo rojo. Longitud mínima en píxeles Compacidad máxima de núcleos epiteliales Radio mínimo de lumen 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 Radio máximo de núcleos epiteliales del recinto más Desviación estándar del citoplasma de la longitud del borde Desviación estándar del fondo del área en píxeles

Figura 7A



AR

Figura 7B



CD34

Figura 8

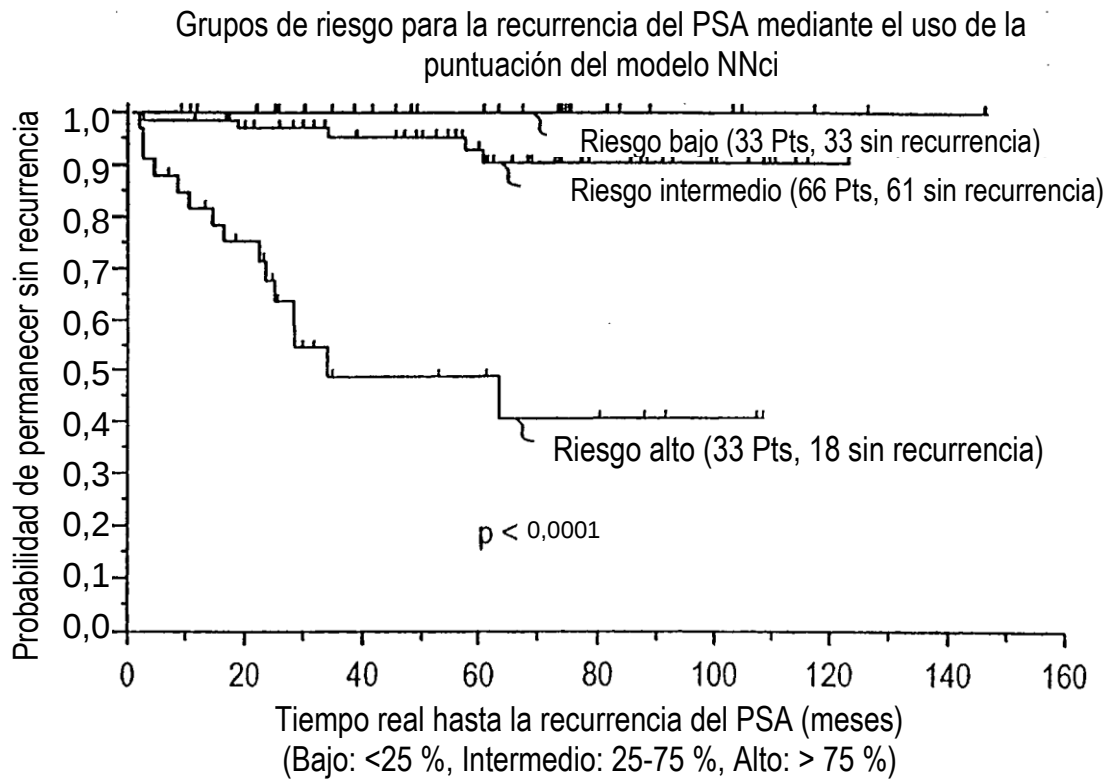


Figura 9

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

Clínica	Etapa clínica 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
	RedBloodCellMeanAreaPxI0386	0,8596
	RedBloodCellStddevAreaPxI0388	0,8621
	LumenMinAsymmetry0295	0,8635

Figura 10

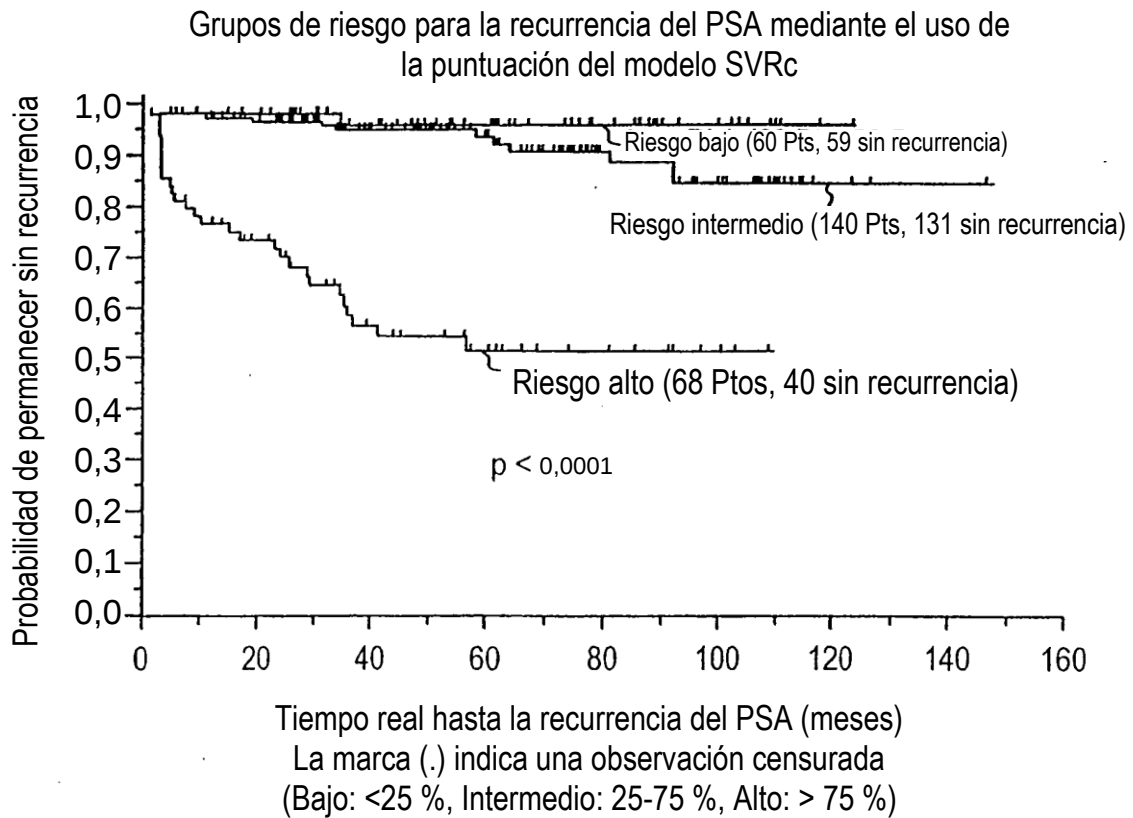


Figura 11

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
	RedBloodCellStddevStddevChann20472	0,7809
	StromaMaxMaxDiff0529	0,7852
	EpitheNucleMeanBordeLengtPx10206	0,7888
	EpithelialNucleMeanAreaPx10194	0,7921
	EpithelNucleiStddevElliptFit0228	0,7951
	RedBloodCellStddevStddevChann30476	0,7964
	RedBloodCellStddevElliptiFit0420	0,7976

Figura 12

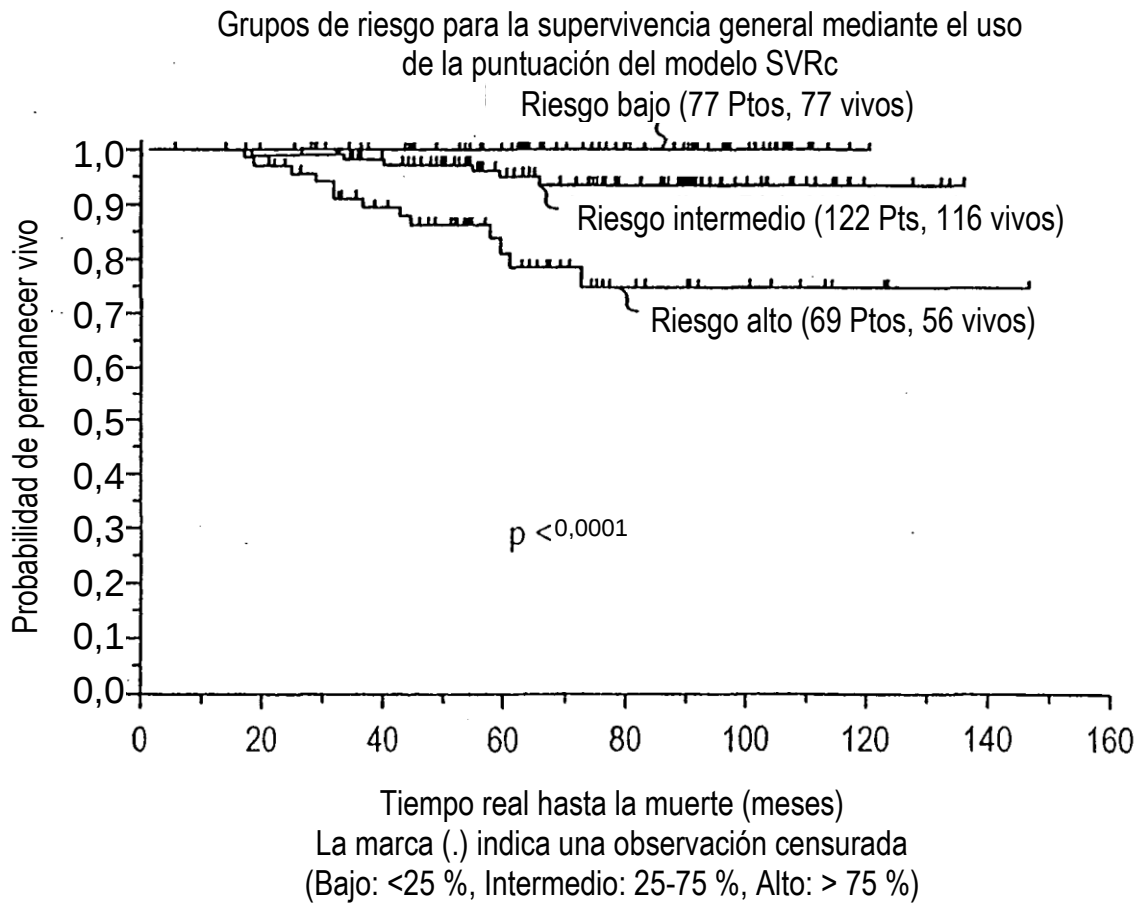


Figura 13

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

Clínica	Extensión extracapsular
	Invasión de las vesículas seminales
	Grado de Gleason dominante de la prostatectomía
	Invasión de ganglios linfáticos
Morfométrica	Área de citoplasma dividida por el área de tejido total
	Desviación estándar del citoplasma del canal rojo medio
	Área de lumen dividida por el área de tejido total

Figura 14

Características seleccionadas para la predicción de recurrencia del PSA (Ejemplo 5)		
Clínica	Ganglios linfáticos	-9,3742
	Afectación de márgenes quirúrgicos	-7,3159
	Afectación de las vesículas seminales	-5,2103
Molecular	Índice de tinción de AR (Tumor)	-3,5404
Morfométrica	AreaStromadivTotTissueArea	-.34225
	AreaEpitNuclidivTotTissueArea	3,2975

Figura 15

Características seleccionadas para la predicción de recurrencia del PSA (Ejemplo 6)		
Clínica	Ganglios linfáticos	-23,32
	Afectación de márgenes quirúrgicos	-11,73
	Puntuación de Gleason de la biopsia	-10,60
	Afectación de las vesículas seminales	-6,40
Molecular	Índice de tinción de AR (tumor)	-10,49
Morfométrica	AreaStromadivTotTissueArea	-16,15
	AreaEpitNucdivTotTissueArea	11,54
	StromaMeanStddevRedChann	-11,26

Figura 16

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

Clínica	Puntuación de Gleason de la 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	-6,20
	AMACR	
Morfométrica	CytoplasmMeanMeanChannel	19,68
	StromaMeanStddevChannel	-10,06
	EpitheNucleiStddevMeanChann	-8,87

Figura 17

Características seleccionadas para la predicción de la recurrencia del PSA (Ejemplo 8)		
Clínica	Puntuación de Gleason de la biopsia	-25,60
	Invasión de las vesículas seminales	-22,69
	Extensión extracapsular	-4,37
	PSA preoperatorio	-13,81
	Grado de Gleason dominante de la prostatectomía	-10,93
Molecular	Área relativa de núcleos epiteliales AR+	-10,64
Morfométrica	Textura de los núcleos epiteliales tumorales	-19,21
	Textura del citoplasma epitelial tumoral	18,19

Figura 18

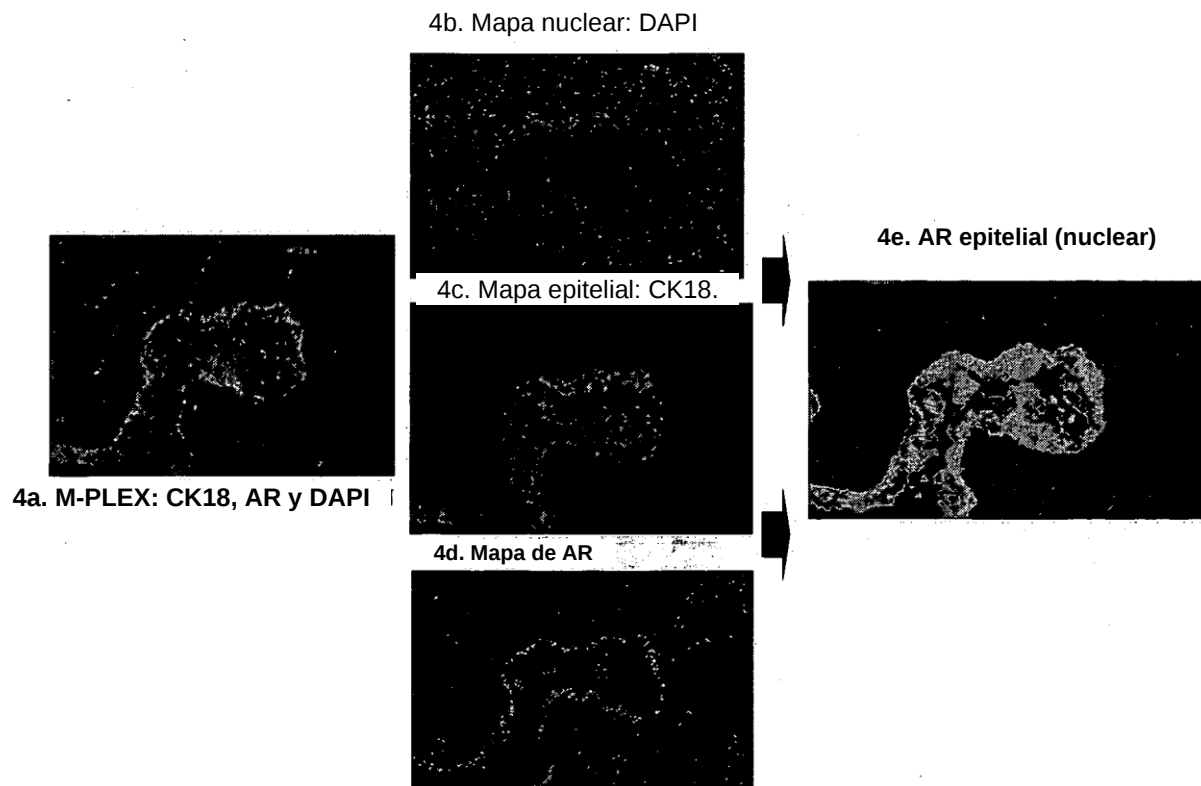


Figura 19

Características seleccionadas para la predicción del fallo clínico después de la prostatectomía (Ejemplo 9)

Clínica	Grado de Gleason dominante de la prostatectomía	-16,4
	Estado de la invasión de ganglios linfáticos	-25,938
Molecular	Intensidad promedio del canal de AR en núcleos epiteliales AR+/AMACR-	-5,2075
Morfométrica	Longitud promedio del perímetro del lumen	17,217
	Variación en la textura dentro del citoplasma en el canal verde	20,728
	Área relativa del lumen	8,9663

Figura 20

