

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 971 182**

21 Número de solicitud: 202230937

51 Int. Cl.:

C12Q 1/6886 (2008.01)

G16B 20/20 (2009.01)

G16H 50/20 (2008.01)

G01N 33/574 (2006.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

28.10.2022

43 Fecha de publicación de la solicitud:

03.06.2024

71 Solicitantes:

SERVICIO ANDALUZ DE SALUD (53.0%)

Avenida de la Constitución 18

41071 Sevilla (Sevilla) ES;

ROCHE FARMA, S.A. (20.0%);

UNIVERSIDAD DE MÁLAGA (17.0%) y

CONSORCIO CENTRO DE INVESTIGACIÓN

BIOMÉDICA EN RED (10.0%)

72 Inventor/es:

COMINO MÉNDEZ, Iñaki;

ALBA CONEJO, Emilio;

PASCUAL LÓPEZ, Francisco Javier;

ALBA BERNAL, Alfonso;

JIMÉNEZ RODRÍGUEZ, Begoña y

QUIRÓS ORTEGA, María Elena

74 Agente/Representante:

SAN MARTÍN ALARCIA, Esther

54 Título: **Panel de secuenciación para biopsia líquida de pacientes con cáncer de mama**

57 Resumen:

Panel de secuenciación masiva compuesto por 37 genes para biopsia líquida que permite el diagnóstico precoz del cáncer de mama y métodos de obtención de datos útiles para el diagnóstico y seguimiento de la enfermedad.

ES 2 971 182 A1

DESCRIPCIÓN

Panel de secuenciación para biopsia líquida de pacientes con cáncer de mama

CAMPO DE LA INVENCIÓN

La presente invención se encuentra dentro del campo de la medicina y la oncología, y se
5 refiere a un panel de secuenciación para tejidos tumorales y biopsia líquida de pacientes con
cáncer de mama.

ESTADO DEL ARTE

Denominamos cáncer de mama a los tumores que se desarrollan en los tejidos mamarios.
Existen dos tipos principales de cáncer de mama: el carcinoma ductal que comienza en los
10 conductos que llevan leche desde la mama hasta el pezón y que es el mayoritario, y el
carcinoma lobular, que comienza en los lóbulos que producen leche. Para detectar el cáncer
de mama, se utilizan diferentes pruebas como la mamografía, ultrasonido mamario con
transductores de alta resolución (ecografías) o imágenes por resonancia magnética. Hasta
la fecha, el diagnóstico de cáncer de mama solo se podía confirmar por medio de una
15 biopsia del tumor.

La biopsia líquida es una estrategia con un gran potencial en el manejo del cáncer,
incluyendo el diagnóstico, la detección de recaídas de forma temprana o permitiendo la
monitorización del estado de la enfermedad a lo largo del tratamiento para ver su evolución.
Se trata de una prueba que se realiza empleando biofluidos corporales, más comúnmente
20 sangre, aunque también se pueden analizar en otros líquidos del organismo como son el
que recubre las meninges, el que recubre el pulmón, la saliva o incluso la orina, con el fin de
buscar células cancerosas de un tumor o fragmentos de ADN de células tumorales, el
llamado ADN tumoral circulante (ctDNA). A diferencia de la biopsia de tejido, permite realizar
el análisis molecular de una forma mucho más sencilla y rápida, ya que se puede hacer una
25 analítica de sangre en cualquier momento, y además se puede repetir para monitorizar la
evolución del tumor.

Lo que se pretende identificar en las muestras son las diferentes mutaciones propias del
tumor, ya que estas mutaciones no están presentes en el resto de células normales que hay
en la sangre.

30 Los paneles de secuenciación son tests genéticos que engloban genes implicados en una
misma patología, en este caso, orientados al cáncer de mama.

La detección y caracterización de los materiales tumorales circulantes presenta dificultades en un escenario temprano debido a sus bajas cantidades en los biofluidos de los pacientes. Sin embargo, el avance en las técnicas de detección actualmente utilizadas, muy sensibles, permiten que, por muy poca cantidad de ctDNA que exista en la sangre o en otros biofluidos, se puede detectar y analizar pudiendo caracterizar la genética de los tumores.

La utilización del ctDNA junto con tecnologías NGS (*next generation sequencing*) ultrasensibles de última generación, también empleando paneles específicos de paciente, han servido para caracterizar los tumores, evaluar las respuestas a las quimioterapias neoadyuvantes y detectar la MRD (*minimal residual disease*) después de la cirugía.

10 NGS dirigidos como SafeSEQ - Safe-sequencing system o sistema de secuenciación segura han surgido recientemente como una nueva herramienta para aumentar la sensibilidad de los instrumentos del sistema de secuenciación paralela masiva para la identificación de variantes raras en el ADN plasmático. Esta estrategia de secuenciación utiliza códigos de barras de una sola molécula antes de la amplificación por PCR para reducir el error de
15 secuenciación y aumentar la precisión. El sistema SafeSEQ, o de secuenciación segura, se puede usar para distinguir entre mutaciones raras y errores técnicos y es capaz de detectar mutaciones que otras tecnologías pasarían por alto. SafeSEQ se ha aplicado hasta ahora a muestras de plasma de cáncer colorrectal metastásico y a pacientes con tumores del estroma gastrointestinal.

20 Otras tecnologías NGS son las sondas de enriquecimiento de regiones génicas concretas como son las de Agilent SureSelect, que se pueden utilizar para diseñar paneles de captura híbridos personalizados y secuenciar masivamente cualquier región codificante o no codificante del genoma o del transcriptoma.

Todo ello ha contribuido al desarrollo de las técnicas de biopsia líquida. Actualmente se
25 utilizan procedimientos de secuenciación masiva (NGS) y diferentes tecnologías para detectar ctDNA en dilución muy baja (1). Sin embargo, aún no está claro si las alteraciones moleculares en ctDNA pueden detectarse mediante ensayos basados en NGS de ctDNA en plasma en pacientes con cáncer de mama primario en el cribado y cómo se relacionan con la mutación del tejido tumoral encontrada en el mismo paciente.

30 Las ventajas que presenta este tipo de biopsia líquida en el seguimiento y la predicción del cáncer de mama metastásico (CMM) han sido comprobadas por diversos estudios (2), resaltando la importancia del ctDNA como potencial biomarcador para la evaluación, predicción y manejo clínico en los pacientes con CMM tratados con quimioterapia. En este

sentido, la monitorización de ctDNA durante el tratamiento puede mejorar significativamente los resultados al predecir la progresión de la enfermedad entre cuatro y seis meses antes que los métodos convencionales (3).

Se han llevado a cabo varios estudios que tratan de desarrollar un método para la monitorización de cáncer de mama utilizando la secuenciación de moléculas de ctDNA derivadas de una muestra obtenida del paciente y posterior análisis de dichas secuencias de ctDNA para identificar sus mutaciones. Un ejemplo de estos paneles de secuenciación lo tenemos en **WO2017181146A1** donde se incluyen los siguientes genes: *AKT1*, *ALK*, *APC*, *ATM*, *BRAF*, *CTNNB1*, *EGFR*, *ERBB2*, *ESR1*, *FGFR2*, *GATA3*, *GNAS*, *IDH1*, *IDH2*, *KIT*, *KRAS*, *MET*, *NRAS*, *PDGFRA*, *PIK3CA*, *PTEN*, *RB1*, *SMAD4*, *STK11* y *TP53* y que permite detectar mutaciones en una frecuencia alélica de hasta 0,025 %. Otro ejemplo en este mismo sentido se describe en el documento **US2021343372A1** donde en una muestra de biopsia líquida en pacientes con cáncer de mama se identifican mutaciones en *TP53* para el 51,1% de los pacientes, y con menor incidencia, en *PIK3CA*, *ESR1*, *BRCA2*, *NF1*, *ATM* y *APC*. Las frecuencias alélicas detectadas son del orden de 0,1 %, 0,25 % y 0,5 %.

Otros paneles de genes para el análisis de ctDNA y la identificación de cáncer de mama metastásico (CMM) incluyen los siguientes genes: *AKT1*, *AR*, *BRCA1*, *BRCA2*, *EGFR*, *ERCC4*, *ERBB2*, *ERBB3*, *ESR1*, *FGFR1*, *KRAS*, *MUC16*, *PIK3CA*, *PIK3R1*, *PTEN*, *PTGFR* y *TGFB1* (4).

También se ha valorado la idoneidad del análisis de ctDNA en el diagnóstico precoz de cáncer de mama, comparando mutaciones en *PIK3CA* y *TP53* entre biopsias tumorales y DNA plasmático pretratamiento (5), revelando cuatro mutaciones adicionales, tres en *TP53* y una en *PIK3CA*, no identificadas previamente en la biopsia de tejido de tres pacientes.

Teniendo en consideración los documentos citados en el estado de la técnica queda claro el gran interés que existe en el desarrollo de las técnicas de biopsia líquida y el análisis del ctDNA, con el fin de construir paneles de secuenciación personalizados para el diagnóstico precoz del cáncer de mama y que resulten útiles en la evaluación y seguimiento de la enfermedad, incluyendo la posibilidad de obtener predicciones sobre la progresión de la enfermedad y tomar decisiones terapéuticas personalizadas.

30 DESCRIPCIÓN DE LA INVENCION

En la presente invención se **propone un panel de secuenciación masiva** para el diagnóstico precoz del cáncer de mama, útiles en la evaluación y seguimiento de la enfermedad, compuesto por **37 genes** (ver **Tabla 1**).

Estos genes han sido seleccionados siguiendo los criterios que se exponen a continuación:

1. Genes con mutaciones en $\geq 1\%$ de las muestras de tumores de mama incluidas en la base de datos TCGA (<https://www.cbioportal.org>).
2. Genes secuenciados en un artículo seminal en tumores de mama y que presentaron 5 mutaciones en este tipo tumoral (6).
3. Genes con relevancia biológica en el cáncer de mama y/o cánceres en general y no basados en su frecuencia mutacional.

Tabla 1: Genes incluidos en el panel de secuenciación masiva customizado empleando la tecnología de Agilent SureSelect XT HS.

<i>AKT1</i>	<i>CREBBP</i>	<i>KRAS</i>	<i>RB1</i>
<i>ARID1A</i>	<i>CTCF</i>	<i>MAP2K4</i>	<i>RUNX1</i>
<i>ATM</i>	<i>ERBB2</i>	<i>MAP3K1</i>	<i>SF3B1</i>
<i>BAP1</i>	<i>ESR1</i>	<i>MEN1</i>	<i>SMAD4</i>
<i>BRAF</i>	<i>FAT1</i>	<i>NCOR1</i>	<i>TBX3</i>
<i>BRCA1</i>	<i>GATA3</i>	<i>NF1</i>	<i>TP53</i>
<i>BRCA2</i>	<i>HRAS</i>	<i>PBRM1</i>	<i>USP9X</i>
<i>CBFB</i>	<i>KDM6A</i>	<i>PIK3CA</i>	
<i>CDH1</i>	<i>KMT2C</i>	<i>PIK3R1</i>	
<i>CDKN1B</i>	<i>KMT2D</i>	<i>PTEN</i>	

10

Todos los genes de la tabla tienen secuencias conocidas, las cuales son públicas y accesibles a través de bases de datos públicas, por lo que no requieren mayor caracterización en la presente memoria para que un experto medio en la materia pueda reproducir la invención.

- 15 Las coordenadas genómicas de las regiones que han sido incluidas en el panel son las que se muestran en la siguiente Tabla 2.

Tabla 2: Coordenadas genómicas (versión del genoma humano=hg38):

Cromo soma	Coordenada inicial	Coordenada final	Gen
chr1	26696371	26697571	ARID1A
chr1	26729624	26729888	ARID1A
chr1	26731125	26731629	ARID1A

chr1	26732601	26732865	ARID1A
chr1	26760819	26761131	ARID1A
chr1	26761308	26761548	ARID1A
chr1	26762115	26762355	ARID1A
chr1	26762936	26763320	ARID1A

ES 2 971 182 A1

chr1	26766137	26766643	ARID1A
chr1	26767762	26768026	ARID1A
chr1	26771090	26771354	ARID1A
chr1	26772421	26772709	ARID1A
chr1	26772779	26773019	ARID1A
chr1	26773264	26773981	ARID1A
chr1	26774294	26775254	ARID1A
chr1	26775497	26775978	ARID1A
chr1	26778989	26780789	ARID1A
chr2	197392225	197392537	SF3B1
chr2	197392935	197393223	SF3B1
chr2	197396023	197396359	SF3B1
chr2	197397906	197398194	SF3B1
chr2	197398376	197398664	SF3B1
chr2	197398890	197399106	SF3B1
chr2	197399978	197400462	SF3B1
chr2	197400681	197400969	SF3B1
chr2	197401318	197401606	SF3B1
chr2	197401658	197402213	SF3B1
chr2	197402522	197402858	SF3B1
chr2	197402871	197403111	SF3B1
chr2	197403554	197403794	SF3B1
chr2	197404994	197405505	SF3B1
chr2	197407914	197408202	SF3B1
chr2	197408342	197408606	SF3B1
chr2	197409744	197410032	SF3B1
chr2	197416705	197416945	SF3B1
chr2	197418428	197418668	SF3B1
chr2	197418857	197419025	SF3B1
chr2	197420303	197420615	SF3B1
chr2	197420948	197421212	SF3B1
chr2	197423770	197424010	SF3B1
chr2	197434889	197435081	SF3B1
chr3	52401640	52402192	BAP1
chr3	52402210	52402498	BAP1
chr3	52402517	52403015	BAP1
chr3	52403061	52403373	BAP1
chr3	52403378	52403930	BAP1
chr3	52404375	52404663	BAP1
chr3	52405081	52405321	BAP1
chr3	52405682	52405994	BAP1
chr3	52406170	52406458	BAP1
chr3	52406747	52406987	BAP1
chr3	52407088	52407537	BAP1

chr3	52407873	52408161	BAP1
chr3	52408395	52408683	BAP1
chr3	52409472	52409955	BAP1
chr3	52548028	52548268	PBRM1
chr3	52550384	52550901	PBRM1
chr3	52554645	52554957	PBRM1
chr3	52558174	52558486	PBRM1
chr3	52561735	52561999	PBRM1
chr3	52563255	52563519	PBRM1
chr3	52564021	52564261	PBRM1
chr3	52576463	52576775	PBRM1
chr3	52578970	52579282	PBRM1
chr3	52586388	52586724	PBRM1
chr3	52587275	52587587	PBRM1
chr3	52589042	52589282	PBRM1
chr3	52603494	52603758	PBRM1
chr3	52609285	52609981	PBRM1
chr3	52615271	52615535	PBRM1
chr3	52617231	52617567	PBRM1
chr3	52624822	52625014	PBRM1
chr3	52627189	52627453	PBRM1
chr3	52628820	52629108	PBRM1
chr3	52634576	52634840	PBRM1
chr3	52641879	52642119	PBRM1
chr3	52643163	52643427	PBRM1
chr3	52644626	52644866	PBRM1
chr3	52648260	52648524	PBRM1
chr3	52651667	52651883	PBRM1
chr3	52658124	52658388	PBRM1
chr3	52662048	52662360	PBRM1
chr3	52668415	52668727	PBRM1
chr3	52678416	52678680	PBRM1
chr3	52679498	52679786	PBRM1
chr3	179198798	179199206	PIK3CA
chr3	179199662	179199926	PIK3CA
chr3	179201258	179201570	PIK3CA
chr3	179203510	179203822	PIK3CA
chr3	179204425	179204665	PIK3CA
chr3	179209515	179209779	PIK3CA
chr3	179210105	179210641	PIK3CA
chr3	179218127	179218415	PIK3CA
chr3	179219116	179219356	PIK3CA
chr3	179219496	179219808	PIK3CA
chr3	179219868	179220132	PIK3CA

chr3	179220951	179221191	PIK3CA
chr3	179224001	179224265	PIK3CA
chr3	179224616	179224904	PIK3CA
chr3	179225880	179226120	PIK3CA
chr3	179229236	179229476	PIK3CA
chr3	179229930	179230456	PIK3CA
chr3	179234060	179234396	PIK3CA
chr4	186588557	186589253	FAT1
chr4	186590289	186590776	FAT1
chr4	186592614	186592830	FAT1
chr4	186595613	186595901	FAT1
chr4	186596507	186597203	FAT1
chr4	186597604	186597868	FAT1
chr4	186597892	186598204	FAT1
chr4	186599864	186600392	FAT1
chr4	186601191	186601503	FAT1
chr4	186602824	186603112	FAT1
chr4	186603144	186604008	FAT1
chr4	186604343	186604607	FAT1
chr4	186605985	186606297	FAT1
chr4	186609107	186609395	FAT1
chr4	186609763	186610051	FAT1
chr4	186611352	186611808	FAT1
chr4	186613081	186613369	FAT1
chr4	186614111	186614423	FAT1
chr4	186616970	186617234	FAT1
chr4	186617677	186621805	FAT1
chr4	186628110	186628398	FAT1
chr4	186628457	186628793	FAT1
chr4	186633609	186633897	FAT1
chr4	186635997	186636261	FAT1
chr4	186636557	186636941	FAT1
chr4	186639644	186639860	FAT1
chr4	186663263	186663647	FAT1
chr4	186706526	186709862	FAT1
chr5	56815538	56816090	MAP3K1
chr5	56856518	56856830	MAP3K1
chr5	56859682	56859946	MAP3K1
chr5	56864701	56864965	MAP3K1
chr5	56865265	56865529	MAP3K1
chr5	56865746	56866058	MAP3K1
chr5	56871826	56872114	MAP3K1
chr5	56872561	56873034	MAP3K1
chr5	56875002	56875338	MAP3K1

chr5	56878896	56879184	MAP3K1
chr5	56880636	56880876	MAP3K1
chr5	56881057	56881297	MAP3K1
chr5	56881533	56882901	MAP3K1
chr5	56883446	56883758	MAP3K1
chr5	56884588	56884900	MAP3K1
chr5	56885853	56886141	MAP3K1
chr5	56887292	56887604	MAP3K1
chr5	56888147	56888435	MAP3K1
chr5	56893449	56893761	MAP3K1
chr5	68226650	68227034	PIK3R1
chr5	68239821	68240013	PIK3R1
chr5	68273315	68273555	PIK3R1
chr5	68273855	68274095	PIK3R1
chr5	68279523	68279811	PIK3R1
chr5	68280496	68280760	PIK3R1
chr5	68280846	68281086	PIK3R1
chr5	68288659	68288827	PIK3R1
chr5	68290649	68290913	PIK3R1
chr5	68292177	68292600	PIK3R1
chr5	68293017	68293512	PIK3R1
chr5	68293627	68293915	PIK3R1
chr5	68294450	68294762	PIK3R1
chr5	68295115	68295561	PIK3R1
chr5	68296135	68296375	PIK3R1
chr5	68297386	68297626	PIK3R1
chr6	151807886	151808390	ESR1
chr6	151842562	151842826	ESR1
chr6	151880580	151880844	ESR1
chr6	151944136	151944544	ESR1
chr6	152011580	152011868	ESR1
chr6	152060913	152061201	ESR1
chr6	152094356	152094596	ESR1
chr6	152098704	152098992	ESR1
chr6	152125186	152125426	ESR1
chr7	140726408	140726600	BRAF
chr7	140734562	140734802	BRAF
chr7	140739734	140740022	BRAF
chr7	140747334	140747526	BRAF
chr7	140749208	140749496	BRAF
chr7	140753189	140753477	BRAF
chr7	140754101	140754317	BRAF
chr7	140776879	140777119	BRAF
chr7	140777912	140778152	BRAF

chr7	140781502	140781766	BRAF
chr7	140782944	140783232	BRAF
chr7	140787469	140787661	BRAF
chr7	140794231	140794543	BRAF
chr7	140800277	140800565	BRAF
chr7	140801329	140801641	BRAF
chr7	140807878	140808142	BRAF
chr7	140808221	140808389	BRAF
chr7	140808811	140809075	BRAF
chr7	140834572	140834908	BRAF
chr7	140850029	140850293	BRAF
chr7	140924490	140924778	BRAF
chr7	152136757	152136997	KMT2C
chr7	152138699	152138987	KMT2C
chr7	152139102	152139342	KMT2C
chr7	152139600	152139864	KMT2C
chr7	152144676	152144916	KMT2C
chr7	152145067	152145379	KMT2C
chr7	152146522	152146810	KMT2C
chr7	152148004	152149180	KMT2C
chr7	152150821	152151085	KMT2C
chr7	152151367	152151655	KMT2C
chr7	152152673	152152985	KMT2C
chr7	152153933	152154221	KMT2C
chr7	152154235	152154475	KMT2C
chr7	152155827	152156419	KMT2C
chr7	152157686	152157998	KMT2C
chr7	152158835	152159099	KMT2C
chr7	152162083	152163859	KMT2C
chr7	152167117	152167405	KMT2C
chr7	152169109	152169325	KMT2C
chr7	152171182	152171422	KMT2C
chr7	152174054	152174318	KMT2C
chr7	152176164	152178036	KMT2C
chr7	152179799	152180159	KMT2C
chr7	152180680	152182624	KMT2C
chr7	152182944	152183184	KMT2C
chr7	152185474	152185714	KMT2C
chr7	152187224	152187512	KMT2C
chr7	152187636	152187924	KMT2C
chr7	152193924	152194340	KMT2C
chr7	152194359	152194647	KMT2C
chr7	152195826	152196090	KMT2C
chr7	152199248	152199488	KMT2C

chr7	152202854	152203142	KMT2C
chr7	152205021	152205309	KMT2C
chr7	152207219	152207507	KMT2C
chr7	152220496	152220760	KMT2C
chr7	152221925	152222141	KMT2C
chr7	152222495	152222759	KMT2C
chr7	152223940	152224252	KMT2C
chr7	152224405	152224645	KMT2C
chr7	152229842	152230106	KMT2C
chr7	152230138	152230402	KMT2C
chr7	152235742	152236006	KMT2C
chr7	152238622	152238910	KMT2C
chr7	152247864	152248656	KMT2C
chr7	152249794	152250034	KMT2C
chr7	152250777	152251041	KMT2C
chr7	152251858	152252170	KMT2C
chr7	152252510	152252750	KMT2C
chr7	152262940	152263204	KMT2C
chr7	152265003	152265243	KMT2C
chr7	152273629	152273941	KMT2C
chr7	152309888	152310152	KMT2C
chr7	152311715	152312027	KMT2C
chr7	152315105	152315369	KMT2C
chr7	152330527	152330815	KMT2C
chr7	152358510	152358750	KMT2C
chr7	152435549	152435861	KMT2C
chr10	8055619	8055931	GATA3
chr10	8058272	8058872	GATA3
chr10	8063909	8064221	GATA3
chr10	8069391	8069679	GATA3
chr10	8073712	8074048	GATA3
chr10	87863912	87864584	PTEN
chr10	87893946	87894186	PTEN
chr10	87925438	87925630	PTEN
chr10	87930971	87931163	PTEN
chr10	87932975	87933287	PTEN
chr10	87942530	87942770	PTEN
chr10	87952044	87952332	PTEN
chr10	87957815	87958055	PTEN
chr10	87960866	87961154	PTEN
chr10	87965259	87965499	PTEN
chr11	532551	532839	HRAS
chr11	533197	533688	HRAS
chr11	533734	533974	HRAS

chr11	534134	534398	HRAS
chr11	64804298	64804850	MEN1
chr11	64804959	64805271	MEN1
chr11	64805558	64805846	MEN1
chr11	64806155	64806443	MEN1
chr11	64806934	64807294	MEN1
chr11	64807471	64807759	MEN1
chr11	64807862	64808126	MEN1
chr11	64809615	64810143	MEN1
chr11	108227540	108227963	ATM
chr11	108229098	108229410	ATM
chr11	108235637	108235877	ATM
chr11	108243879	108244191	ATM
chr11	108244750	108245062	ATM
chr11	108246889	108247201	ATM
chr11	108248897	108249137	ATM
chr11	108250670	108251102	ATM
chr11	108251801	108252065	ATM
chr11	108252732	108252996	ATM
chr11	108253782	108254070	ATM
chr11	108256133	108256421	ATM
chr11	108257399	108257687	ATM
chr11	108258910	108259150	ATM
chr11	108267136	108267376	ATM
chr11	108268377	108268641	ATM
chr11	108270984	108271484	ATM
chr11	108272449	108272930	ATM
chr11	108279417	108279681	ATM
chr11	108280961	108281201	ATM
chr11	108282674	108282914	ATM
chr11	108284193	108284505	ATM
chr11	108287525	108287789	ATM
chr11	108288895	108289183	ATM
chr11	108289569	108289833	ATM
chr11	108292585	108292825	ATM
chr11	108293238	108293550	ATM
chr11	108294848	108295136	ATM
chr11	108297202	108297466	ATM
chr11	108299679	108299919	ATM
chr11	108301574	108301862	ATM
chr11	108302820	108303060	ATM
chr11	108304643	108304883	ATM
chr11	108307820	108308060	ATM
chr11	108310081	108310393	ATM

chr11	108312334	108312574	ATM
chr11	108315746	108316193	ATM
chr11	108317290	108317602	ATM
chr11	108319873	108320137	ATM
chr11	108321216	108321504	ATM
chr11	108325282	108325570	ATM
chr11	108326021	108326261	ATM
chr11	108327569	108327833	ATM
chr11	108328985	108329273	ATM
chr11	108330185	108330449	ATM
chr11	108331368	108331632	ATM
chr11	108331801	108332113	ATM
chr11	108332686	108332974	ATM
chr11	108333806	108334046	ATM
chr11	108334894	108335182	ATM
chr11	108335770	108336034	ATM
chr11	108343140	108343452	ATM
chr11	108345669	108345981	ATM
chr11	108347201	108347441	ATM
chr11	108353690	108353954	ATM
chr11	108354734	108354950	ATM
chr11	108365005	108365293	ATM
chr11	108365296	108365536	ATM
chr12	12717812	12718340	CDKN1B
chr12	12718741	12719029	CDKN1B
chr12	12720979	12721291	CDKN1B
chr12	25209720	25209984	KRAS
chr12	25215356	25215644	KRAS
chr12	25225537	25225849	KRAS
chr12	25227202	25227442	KRAS
chr12	25235131	25235299	KRAS
chr12	25245196	25245460	KRAS
chr12	49021705	49021945	KMT2D
chr12	49021964	49022436	KMT2D
chr12	49022564	49022900	KMT2D
chr12	49023968	49024184	KMT2D
chr12	49024498	49025021	KMT2D
chr12	49026151	49027351	KMT2D
chr12	49027722	49028218	KMT2D
chr12	49028748	49029268	KMT2D
chr12	49029318	49029558	KMT2D
chr12	49030203	49030515	KMT2D
chr12	49030564	49030804	KMT2D
chr12	49030818	49031106	KMT2D

chr12	49031140	49033996	KMT2D
chr12	49034038	49034326	KMT2D
chr12	49034334	49035017	KMT2D
chr12	49037096	49039016	KMT2D
chr12	49039145	49039645	KMT2D
chr12	49039693	49041565	KMT2D
chr12	49041571	49041787	KMT2D
chr12	49041833	49042365	KMT2D
chr12	49042482	49042953	KMT2D
chr12	49042998	49043262	KMT2D
chr12	49043287	49043503	KMT2D
chr12	49043552	49044076	KMT2D
chr12	49044119	49044606	KMT2D
chr12	49044710	49044998	KMT2D
chr12	49045835	49046497	KMT2D
chr12	49046579	49046819	KMT2D
chr12	49047884	49048148	KMT2D
chr12	49048581	49048845	KMT2D
chr12	49049029	49049293	KMT2D
chr12	49049647	49050823	KMT2D
chr12	49050850	49052458	KMT2D
chr12	49052480	49052792	KMT2D
chr12	49052837	49053395	KMT2D
chr12	49053402	49053714	KMT2D
chr12	49053902	49054214	KMT2D
chr12	49054229	49054493	KMT2D
chr12	49054495	49054783	KMT2D
chr12	49054818	49055106	KMT2D
chr12	49055191	49055407	KMT2D
chr12	114671807	114672335	TBX3
chr12	114674127	114674871	TBX3
chr12	114676235	114676547	TBX3
chr12	114677497	114677737	TBX3
chr12	114679421	114679733	TBX3
chr12	114679834	114680050	TBX3
chr12	114680844	114681180	TBX3
chr12	114682777	114683233	TBX3
chr13	32316385	32316601	BRCA2
chr13	32319044	32319356	BRCA2
chr13	32324997	32325261	BRCA2
chr13	32326017	32326357	BRCA2
chr13	32326423	32326687	BRCA2
chr13	32329359	32329575	BRCA2
chr13	32330842	32331106	BRCA2

chr13	32332241	32333417	BRCA2
chr13	32336234	32341226	BRCA2
chr13	32344473	32344737	BRCA2
chr13	32346753	32346969	BRCA2
chr13	32354834	32355314	BRCA2
chr13	32356398	32356638	BRCA2
chr13	32357715	32357955	BRCA2
chr13	32362487	32362727	BRCA2
chr13	32363151	32363559	BRCA2
chr13	32370323	32370635	BRCA2
chr13	32370871	32371183	BRCA2
chr13	32375266	32375482	BRCA2
chr13	32376586	32376874	BRCA2
chr13	32379283	32379547	BRCA2
chr13	32379675	32380219	BRCA2
chr13	32394654	32394966	BRCA2
chr13	32395960	32396170	BRCA2
chr13	32396081	32396117	BRCA2
chr13	32396814	32397126	BRCA2
chr13	32398129	32398801	BRCA2
chr13	48303836	48304124	RB1
chr13	48307198	48307486	RB1
chr13	48342524	48342788	RB1
chr13	48344995	48345283	RB1
chr13	48347747	48347939	RB1
chr13	48348881	48349097	RB1
chr13	48359939	48360203	RB1
chr13	48362729	48363041	RB1
chr13	48364812	48365052	RB1
chr13	48367416	48367680	RB1
chr13	48368445	48368685	RB1
chr13	48373328	48373568	RB1
chr13	48376843	48377107	RB1
chr13	48379493	48379709	RB1
chr13	48379972	48380322	RB1
chr13	48381212	48381476	RB1
chr13	48452907	48453195	RB1
chr13	48456120	48456432	RB1
chr13	48459603	48459915	RB1
chr13	48463650	48463914	RB1
chr13	48464922	48465442	RB1
chr13	48473278	48473470	RB1
chr13	48476615	48476927	RB1
chr13	48477271	48477487	RB1

chr13	48479914	48480154	RB1
chr14	104769509	104769701	AKT1
chr14	104770260	104770500	AKT1
chr14	104770663	104770927	AKT1
chr14	104771727	104771895	AKT1
chr14	104772288	104772528	AKT1
chr14	104772840	104773128	AKT1
chr14	104773170	104773661	AKT1
chr14	104773837	104774053	AKT1
chr14	104774862	104775285	AKT1
chr14	104775569	104775881	AKT1
chr14	104776582	104776846	AKT1
chr14	104780007	104780295	AKT1
chr14	104792524	104792716	AKT1
chr16	3727691	3729899	CREBBP
chr16	3731164	3731500	CREBBP
chr16	3731700	3732012	CREBBP
chr16	3735999	3736239	CREBBP
chr16	3736576	3736888	CREBBP
chr16	3738483	3738747	CREBBP
chr16	3739492	3739804	CREBBP
chr16	3740317	3740629	CREBBP
chr16	3744819	3745035	CREBBP
chr16	3745195	3745435	CREBBP
chr16	3749546	3749762	CREBBP
chr16	3751648	3751912	CREBBP
chr16	3757211	3757451	CREBBP
chr16	3757772	3758084	CREBBP
chr16	3758768	3759056	CREBBP
chr16	3767694	3767934	CREBBP
chr16	3769143	3769383	CREBBP
chr16	3770537	3771017	CREBBP
chr16	3773691	3773955	CREBBP
chr16	3774486	3774774	CREBBP
chr16	3777538	3777730	CREBBP
chr16	3777976	3778216	CREBBP
chr16	3778626	3778890	CREBBP
chr16	3780648	3780960	CREBBP
chr16	3781122	3781386	CREBBP
chr16	3782648	3782960	CREBBP
chr16	3791905	3792169	CREBBP
chr16	3793349	3793661	CREBBP
chr16	3810570	3810810	CREBBP
chr16	3850268	3851036	CREBBP

chr16	3879753	3879993	CREBBP
chr16	67029326	67029566	CBFB
chr16	67029649	67029889	CBFB
chr16	67036564	67036828	CBFB
chr16	67066607	67066871	CBFB
chr16	67082131	67082419	CBFB
chr16	67098635	67098851	CBFB
chr16	67610802	67611642	CTCF
chr16	67611915	67612155	CTCF
chr16	67616667	67616955	CTCF
chr16	67620612	67620900	CTCF
chr16	67621360	67621672	CTCF
chr16	67626478	67626790	CTCF
chr16	67628340	67628580	CTCF
chr16	67629321	67629609	CTCF
chr16	67636614	67636926	CTCF
chr16	67637659	67637899	CTCF
chr16	68737331	68737547	CDH1
chr16	68738221	68738485	CDH1
chr16	68801637	68801925	CDH1
chr16	68808339	68808926	CDH1
chr16	68810112	68810424	CDH1
chr16	68811651	68811891	CDH1
chr16	68812054	68812342	CDH1
chr16	68813281	68813593	CDH1
chr16	68815480	68815792	CDH1
chr16	68819196	68819508	CDH1
chr16	68821968	68822256	CDH1
chr16	68823368	68823656	CDH1
chr16	68826446	68826638	CDH1
chr16	68828094	68828382	CDH1
chr16	68829569	68829881	CDH1
chr16	68833262	68833526	CDH1
chr16	68834151	68834391	CDH1
chr17	7661816	7662103	TP53
chr17	7666154	7666346	TP53
chr17	7661914	7661935	TP53
chr17	7667144	7667456	TP53
chr17	7669529	7669769	TP53
chr17	7670529	7670793	TP53
chr17	7673134	7673422	TP53
chr17	7673451	7673912	TP53
chr17	7674103	7674367	TP53
chr17	7674782	7675265	TP53

chr17	7675964	7676300	TP53
chr17	7676308	7676677	TP53
chr17	12020811	12021075	MAP2K4
chr17	12032185	12032377	MAP2K4
chr17	12054807	12055071	MAP2K4
chr17	12081322	12081562	MAP2K4
chr17	12095490	12095778	MAP2K4
chr17	12107705	12107993	MAP2K4
chr17	12110292	12110508	MAP2K4
chr17	12113152	12113440	MAP2K4
chr17	12125212	12125452	MAP2K4
chr17	12129056	12129368	MAP2K4
chr17	12139765	12139957	MAP2K4
chr17	12141071	12141335	MAP2K4
chr17	16032269	16032509	NCOR1
chr17	16034734	16034974	NCOR1
chr17	16039257	16039689	NCOR1
chr17	16040359	16040575	NCOR1
chr17	16046865	16047177	NCOR1
chr17	16048760	16049072	NCOR1
chr17	16057481	16057769	NCOR1
chr17	16057829	16058141	NCOR1
chr17	16058390	16058678	NCOR1
chr17	16061371	16061923	NCOR1
chr17	16062031	16062343	NCOR1
chr17	16063983	16064271	NCOR1
chr17	16064788	16065100	NCOR1
chr17	16065457	16065721	NCOR1
chr17	16067863	16068151	NCOR1
chr17	16070128	16070560	NCOR1
chr17	16071380	16071692	NCOR1
chr17	16072066	16072306	NCOR1
chr17	16073354	16073642	NCOR1
chr17	16075497	16075737	NCOR1
chr17	16079881	16080145	NCOR1
chr17	16080326	16080810	NCOR1
chr17	16086205	16086517	NCOR1
chr17	16087138	16087378	NCOR1
chr17	16091824	16092088	NCOR1
chr17	16098287	16098575	NCOR1
chr17	16101215	16101839	NCOR1
chr17	16108704	16108992	NCOR1
chr17	16117813	16118101	NCOR1
chr17	16119345	16119561	NCOR1

chr17	16121016	16121304	NCOR1
chr17	16126001	16126289	NCOR1
chr17	16137229	16137493	NCOR1
chr17	16138076	16138292	NCOR1
chr17	16138976	16139216	NCOR1
chr17	16143530	16143770	NCOR1
chr17	16146341	16146581	NCOR1
chr17	16149375	16149591	NCOR1
chr17	16151506	16151698	NCOR1
chr17	16151863	16152079	NCOR1
chr17	16153258	16153474	NCOR1
chr17	16158684	16158948	NCOR1
chr17	16164949	16165189	NCOR1
chr17	16171767	16172055	NCOR1
chr17	16186476	16186764	NCOR1
chr17	16194383	16194647	NCOR1
chr17	31095132	31095444	NF1
chr17	31155898	31156210	NF1
chr17	31158931	31159171	NF1
chr17	31163148	31163412	NF1
chr17	31169811	31170075	NF1
chr17	31181347	31181563	NF1
chr17	31181627	31181867	NF1
chr17	31182430	31182742	NF1
chr17	31200388	31200628	NF1
chr17	31200953	31201241	NF1
chr17	31201327	31201567	NF1
chr17	31206161	31206449	NF1
chr17	31214373	31214661	NF1
chr17	31218929	31219193	NF1
chr17	31221775	31222063	NF1
chr17	31223361	31223649	NF1
chr17	31225016	31225328	NF1
chr17	31226403	31226715	NF1
chr17	31227134	31227374	NF1
chr17	31227444	31227684	NF1
chr17	31228992	31229496	NF1
chr17	31229760	31230048	NF1
chr17	31230176	31230464	NF1
chr17	31230763	31231003	NF1
chr17	31231998	31232262	NF1
chr17	31232670	31232910	NF1
chr17	31232975	31233239	NF1
chr17	31235535	31236101	NF1

chr17	31248907	31249195	NF1
chr17	31252860	31253076	NF1
chr17	31257606	31257798	NF1
chr17	31258266	31258578	NF1
chr17	31258948	31259212	NF1
chr17	31260285	31260597	NF1
chr17	31261627	31261939	NF1
chr17	31265151	31265415	NF1
chr17	31299425	31299689	NF1
chr17	31325783	31326287	NF1
chr17	31327464	31327872	NF1
chr17	31330264	31330528	NF1
chr17	31334588	31335066	NF1
chr17	31336258	31336546	NF1
chr17	31336606	31336942	NF1
chr17	31337330	31337618	NF1
chr17	31337740	31338214	NF1
chr17	31338622	31338886	NF1
chr17	31340430	31340718	NF1
chr17	31342927	31343215	NF1
chr17	31349041	31349329	NF1
chr17	31350106	31350394	NF1
chr17	31352179	31352491	NF1
chr17	31356376	31356664	NF1
chr17	31356880	31357168	NF1
chr17	31357186	31357450	NF1
chr17	31358394	31358706	NF1
chr17	31358883	31359099	NF1
chr17	31360454	31360742	NF1
chr17	31367143	31367359	NF1
chr17	31373927	31374239	NF1
chr17	31378813	31379005	NF1
chr17	39699477	39699669	ERBB2
chr17	39700154	39700394	ERBB2
chr17	39706957	39707221	ERBB2
chr17	39708295	39708559	ERBB2
chr17	39709240	39709528	ERBB2
chr17	39709738	39709954	ERBB2
chr17	39710011	39710554	ERBB2
chr17	39711843	39712131	ERBB2
chr17	39712240	39712528	ERBB2
chr17	39713053	39713353	ERBB2
chr17	39713198	39713294	ERBB2
chr17	39715202	39715610	ERBB2

chr17	39715707	39715971	ERBB2
chr17	39716222	39716679	ERBB2
chr17	39717245	39717557	ERBB2
chr17	39719702	39719918	ERBB2
chr17	39723243	39723742	ERBB2
chr17	39723828	39724092	ERBB2
chr17	39724698	39724938	ERBB2
chr17	39724970	39725484	ERBB2
chr17	39725623	39725935	ERBB2
chr17	39726478	39726742	ERBB2
chr17	39726788	39727028	ERBB2
chr17	39727264	39727576	ERBB2
chr17	39727662	39728070	ERBB2
chr17	43045595	43045883	BRCA1
chr17	43047564	43047780	BRCA1
chr17	43049037	43049277	BRCA1
chr17	43049980	43050172	BRCA1
chr17	43050981	43051197	BRCA1
chr17	43056973	43057213	BRCA1
chr17	43063256	43063448	BRCA1
chr17	43063792	43064032	BRCA1
chr17	43067531	43067771	BRCA1
chr17	43070890	43071274	BRCA1
chr17	43074293	43074557	BRCA1
chr17	43076406	43076694	BRCA1
chr17	43079258	43079474	BRCA1
chr17	43082369	43082609	BRCA1
chr17	43090867	43091107	BRCA1
chr17	43091295	43094895	BRCA1
chr17	43095763	43096003	BRCA1
chr17	43097170	43097362	BRCA1
chr17	43099695	43099959	BRCA1
chr17	43104047	43104335	BRCA1
chr17	43104791	43105031	BRCA1
chr17	43106374	43106614	BRCA1
chr17	43110460	43110652	BRCA1
chr17	43115644	43115860	BRCA1
chr17	43123936	43124176	BRCA1
chr17	43125099	43125267	BRCA1
chr18	51047014	51047326	SMAD4
chr18	51048652	51048892	SMAD4
chr18	51049213	51049405	SMAD4
chr18	51054754	51055018	SMAD4
chr18	51058040	51058529	SMAD4

chr18	51059782	51059998	SMAD4
chr18	51065394	51065634	SMAD4
chr18	51066982	51067222	SMAD4
chr18	51076604	51076844	SMAD4
chr18	51077229	51077397	SMAD4
chr18	51078229	51078493	SMAD4
chr21	34792108	34792636	RUNX1
chr21	34799225	34799537	RUNX1
chr21	34821585	34821777	RUNX1
chr21	34834373	34834637	RUNX1
chr21	34859393	34859657	RUNX1
chr21	34880478	34880790	RUNX1
chr21	34886809	34887145	RUNX1
chr21	34889598	34889814	RUNX1
chr21	34892847	34893039	RUNX1
chr21	35048762	35048978	RUNX1
chrX	41123544	41123808	USP9X
chrX	41128916	41129228	USP9X
chrX	41131376	41131616	USP9X
chrX	41134648	41134912	USP9X
chrX	41136768	41137056	USP9X
chrX	41140581	41140845	USP9X
chrX	41140935	41141505	USP9X
chrX	41143210	41143522	USP9X
chrX	41144441	41144705	USP9X
chrX	41148339	41148603	USP9X
chrX	41150844	41151132	USP9X
chrX	41152870	41153158	USP9X
chrX	41162713	41162953	USP9X
chrX	41165838	41166246	USP9X
chrX	41167397	41167661	USP9X
chrX	41167980	41168244	USP9X
chrX	41169958	41170270	USP9X
chrX	41170388	41170700	USP9X
chrX	41171753	41172041	USP9X
chrX	41183918	41184206	USP9X
chrX	41184367	41184703	USP9X
chrX	41186435	41186723	USP9X
chrX	41187910	41188198	USP9X
chrX	41189271	41189511	USP9X
chrX	41195822	41196086	USP9X
chrX	41196172	41196436	USP9X
chrX	41196508	41196820	USP9X
chrX	41197280	41197592	USP9X

chrX	41198494	41198782	USP9X
chrX	41201025	41201313	USP9X
chrX	41205265	41205529	USP9X
chrX	41210475	41210715	USP9X
chrX	41214494	41214782	USP9X
chrX	41215867	41216683	USP9X
chrX	41217137	41217425	USP9X
chrX	41218340	41218628	USP9X
chrX	41219022	41219310	USP9X
chrX	41223189	41223429	USP9X
chrX	41224707	41225212	USP9X
chrX	41229174	41229486	USP9X
chrX	41229540	41229852	USP9X
chrX	41230416	41230680	USP9X
chrX	41232311	41232599	USP9X
chrX	44873475	44873787	KDM6A
chrX	44873847	44874063	KDM6A
chrX	44961205	44961469	KDM6A
chrX	44974582	44974798	KDM6A
chrX	45010881	45011097	KDM6A
chrX	45020525	45020813	KDM6A
chrX	45034849	45035065	KDM6A
chrX	45037575	45037767	KDM6A
chrX	45051635	45051875	KDM6A
chrX	45053747	45054035	KDM6A
chrX	45058922	45059186	KDM6A
chrX	45059212	45059500	KDM6A
chrX	45059944	45060232	KDM6A
chrX	45060530	45060842	KDM6A
chrX	45061239	45061503	KDM6A
chrX	45062565	45062829	KDM6A
chrX	45063391	45063847	KDM6A
chrX	45069547	45070387	KDM6A
chrX	45076617	45076905	KDM6A
chrX	45078320	45078584	KDM6A
chrX	45079116	45079380	KDM6A
chrX	45082499	45082871	KDM6A
chrX	45083377	45083689	KDM6A
chrX	45085789	45086053	KDM6A
chrX	45089716	45089956	KDM6A
chrX	45090649	45090937	KDM6A
chrX	45106463	45106727	KDM6A
chrX	45107328	45107616	KDM6A
chrX	45110043	45110283	KDM6A

chrX	45111300	45111492	KDM6A
------	----------	----------	-------

5 Gracias al empleo de códigos de barras moleculares, este panel permite la caracterización ultrasensible de aberraciones genéticas tumorales, por lo que puede ser empleado para la detección de mutaciones en frecuencias alélicas por debajo del 0,15%.

Esta alta sensibilidad permite la detección de mutaciones en biofluidos (biopsia líquida) de pacientes con tumores localizados/tempranos o aquellos cánceres con poca liberación de material tumoral. Este panel ha sido optimizado y validado usando cuatro muestras de tumores de mama con mutaciones conocidas (ver ejemplo 1) y sus resultados han sido comparados con otros paneles de secuenciación para este mismo tipo de cáncer, cáncer de mama, mostrando una mayor capacidad de detección de mutaciones (ver ejemplo 2). Su uso en biopsias líquidas ha sido respaldado en estudios comparativos con muestras tumorales, donde se demuestra la alta sensibilidad, especificidad y capacidad de valor predictivo del método de diagnóstico de cáncer de mama basado en este panel de secuenciación (ver ejemplo 3). También se han comprobado, mediante su correlación con variables clínico-patológicas, su utilidad como marcador de pronóstico o progresión de la enfermedad (ver ejemplo 4).

Por tanto, un **primer aspecto** de la invención se refiere al panel de secuenciación que comprende los genes de la **Tabla 1**.

En concreto y más preferiblemente, a su uso como biomarcadores para diagnosticar, diagnosticar de forma temprana, predecir o pronosticar la evolución el cáncer de mama. Y aún más preferiblemente, para su uso en biopsia líquida.

Métodos de la invención

25 Otro **aspecto** de la invención se refiere a un método *in vitro* de obtención de datos útiles para diagnosticar, diagnosticar de forma temprana, predecir o pronosticar la evolución del cáncer de mama en un individuo, de ahora en adelante primer método de la invención, que comprende:

- a) obtener una muestra biológica aislada del individuo,
- 30 b) identificar posibles mutaciones en los genes de la Tabla 1.

Otro **aspecto** de la invención se refiere a un método *in vitro* para diagnosticar o diagnosticar de forma temprana, el cáncer de mama, de ahora en adelante segundo método de la

invención, que comprende los pasos (a) - (b) según el primer método de la invención, y además comprende:

- c) asignar al individuo que presenta mutaciones en alguno de los genes de la Tabla 1 al grupo de pacientes que padece cáncer de mama.

- 5 En una realización preferida del segundo método de la invención además comprende asignar al individuo que presenta una mediana de frecuencia alélica de las mutaciones más alta, al grupo de pacientes con un grado tumoral más alto.

Otro aspecto de la invención se refiere a un método *in vitro* para predecir o pronosticar la evolución del cáncer de mama, que comprende el método de obtención de datos útiles de la
10 invención, y que además comprende asignar al individuo que presenta más de una mutación, al grupo de pacientes con probabilidad de recaída clínica.

Otro aspecto de la invención refiere al método *in vitro* para predecir o pronosticar la evolución del cáncer de mama, que comprende el método de obtención de datos útiles de la invención y que además comprende asignar al individuo que presenta mutaciones en *TP53*
15 al grupo de pacientes con tumor negativo a la expresión de receptores hormonales.

Una "muestra biológica" tal como se define aquí, es una pequeña parte de un sujeto, representativa del conjunto y puede estar constituida por una biopsia o una muestra de fluido corporal.

Preferiblemente, la muestra biológica es plasma, proveniente de la separación física de la
20 muestra de sangre tomada del paciente.

La detección las mutaciones en los genes de la Tabla 1 puede realizarse por cualquier medio conocido en el estado de la técnica.

Preferiblemente se lleva a cabo mediante tecnologías de secuenciación de próxima generación o "next-generation sequencing" (NGS). Más preferiblemente la tecnología NGS
25 utilizada son sondas de enriquecimiento tipo Agilent SureSelect.

El método de la presente invención se puede aplicar con muestras de individuos de cualquier sexo, es decir, hombres o mujeres, y a cualquier edad. Preferiblemente se aplica a muestras obtenidas de mujeres.

En la presente invención se entiende por "pronóstico" la evolución esperada de una
30 enfermedad y se refiere a la valoración de la probabilidad según la cual un sujeto padece una enfermedad, así como a la valoración de su inicio, estado de desarrollo, evolución, o de

su regresión, y/o el pronóstico del curso de la enfermedad en el futuro. Como entenderán los expertos en la materia, tal valoración, aunque se prefiere que sea, normalmente puede no ser correcta para el 100% de los sujetos que se van a diagnosticar. El término, sin embargo, requiere que una parte estadísticamente significativa de los sujetos se pueda identificar como que padecen la enfermedad o que tienen predisposición a la misma. Si una parte es estadísticamente significativa se puede determinar sin más por el experto en la materia usando varias herramientas de evaluación estadística bien conocidas, por ejemplo, determinación de intervalos de confianza, determinación de valores p, prueba de la t de Student, prueba de Mann-Whitney, etc. Los intervalos de confianza preferidos son al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%. Los valores de p son, preferiblemente, 0,2, 0,1, 0,05. A su vez, atendiendo al método de la presente invención, se podrían establecer otras subclasificaciones dentro de esta principal, facilitando, por tanto, la elección y el establecimiento de regímenes terapéuticos o tratamiento adecuados. Esta discriminación tal y como es entendida por un experto en la materia no pretende ser correcta en un 100% de las muestras analizadas. Sin embargo, requiere que una cantidad estadísticamente significativa de las muestras analizadas sean clasificadas correctamente. La cantidad que es estadísticamente significativa puede ser establecida por un experto en la materia mediante el uso de diferentes herramientas estadísticas, por ejemplo, pero sin limitarse, mediante la determinación de intervalos de confianza, determinación del valor significación P, test de Student o funciones discriminantes de Fisher, medidas no paramétricas de Mann Whitney, correlación de Spearman, regresión logística, regresión lineal, área bajo la curva de ROC (AUC). Preferiblemente, los intervalos de confianza son al menos del 90%, al menos del 95%, al menos del 97%, al menos del 98% o al menos del 99%. Preferiblemente, el valor de p es menor de 0,1, de 0,05, de 0,01, de 0,005 o de 0,0001. Preferiblemente, la presente invención permite detectar correctamente la enfermedad de forma diferencial en al menos el 60%, más preferiblemente en al menos el 70%, mucho más preferiblemente en al menos el 80%, o aún mucho más preferiblemente en al menos el 90% de los sujetos de un determinado grupo o población analizada.

Kit y usos

Otro **aspecto** de la invención se refiere a un kit o dispositivo, de ahora en adelante kit o dispositivo de la invención, que comprende los elementos necesarios para detectar mutaciones en los genes de la Tabla 1 simultáneamente.

Automatización del método de la invención implementándolo en un programa de ordenador

Otro **aspecto** de la invención se refiere a un programa de ordenador que comprende instrucciones para realizar el procedimiento de acuerdo con cualquiera de los métodos de la invención.

En particular, la invención abarca programas de ordenador dispuestos sobre o dentro de una portadora. La portadora puede ser cualquier entidad o dispositivo capaz de soportar el programa. Como variante, la portadora podría ser un circuito integrado en el que va incluido el programa y que se haya adaptado para ejecutar, o para ser utilizado en la ejecución de los procesos correspondientes.

Por ejemplo, los programas podrían estar incorporados en un medio de almacenamiento, como una memoria ROM, una memoria CD ROM o una memoria ROM de semiconductor, una memoria USB, o un soporte de grabación magnética, por ejemplo, un disco flexible o un disco duro. Alternativamente, los programas podrían estar soportados en una señal portadora transmisible; por ejemplo, podría tratarse de una señal eléctrica u óptica que podría transportarse a través de cable eléctrico u óptico, por radio o por cualesquiera otros medios.

La invención se extiende también a programas de ordenador adaptados para que cualquier medio de procesamiento pueda llevar a la práctica los métodos de la invención. Los programas de ordenador también abarcan aplicaciones en la nube basadas en dicho procedimiento.

Otros aspectos de la invención se refieren al medio de almacenamiento legible y a la señal transmisible que comprende instrucciones de programa necesarias para la ejecución del método de invención por un ordenador.

A lo largo de la descripción y las reivindicaciones la palabra "comprende" y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Para los expertos en la materia, otros objetos, ventajas y características de la invención se desprenderán en parte de la descripción y en parte de la práctica de la invención. Los siguientes ejemplos y dibujos se proporcionan a modo de ilustración, y no se pretende que sean limitativos de la presente invención.

DESCRIPCIÓN DE LAS FIGURAS

Figura 1. Mutaciones encontradas en las muestras de plasma de pacientes con cáncer de mama localizado. **A:** mutaciones identificadas en las muestras de plasma tanto en los subestudios informado y no informado por la genética del tumor. **B:** características clínico-patológicas de los pacientes mostrando el estado clínico actual (recaída clínica).

- 5 **Figura 2.** Mutaciones identificadas en la secuenciación del ADN tumoral de las muestras incluidas en el estudio. Oncoplot. **A:** mutaciones observadas en las muestras tumorales indicando si se encontraron en plasma. **B:** características clínico-patológicas del paciente.

EJEMPLOS DE LA INVENCION

Ejemplo 1: Optimización y validación del panel de genes de la Tabla 1

- 10 El panel de 37 genes implicados en cáncer de mama de la Tabla 1 fue evaluado generando librerías de secuenciación con el ADN obtenido de 4 tumores de mama y comparando los resultados con otro panel, que emplea una tecnología diferente.

- Se utilizó para su secuenciación tecnologías ultrasensibles de última generación (NGS - next generation sequencing), en concreto SureSelect XT HS. Se comprobó que la cobertura de
15 las zonas genómicas fue correcta y se detectaron mutaciones conocidas en estos tumores, que habían sido secuenciados previamente con otra tecnología, llamada TruSeq de Illumina.

- Tabla 3:** Mutaciones observadas en los cuatro tumores testados usando el panel personalizado de 37 genes (Sure Select XT HS) y un panel validado que emplea una tecnología y set de genes diferente (Illumina sequencing Truseq), en concreto *TP53*, *CDH1*
20 y *PIK3CA*.

MUESTRA	SureSelect XT HS		Illumina sequencing (TruSeq)	
	Gen/Mutación	VAF	Gen/Mutación	VAF
09MS	TP53/P.R213T	44,10%	TP53/P.R213T	46,90%
	CDH1/p.E841*	5,85%	ND	ND
	FAT1/p.P147L	0,81%	NT	NT
	ESR1/p.R394C	0,58%	NT	NT
	ATM/p.D360Y	0,81%	NT	NT
	BRCA2/p.D878N	0,66%	NT	NT
	CREBBP/p.R1091C	0,68%	NT	NT
	ERBB2/p.S310F	0,76%	NT	NT
	ERBB2/p.R688Q	0,58%	NT	NT
13MS	TP53/p.M133K	78,40%	TP53/p.M133K	75,20%
	ESR1/p.R243H	0,62%	NT	NT
	CTCF/p.R29W	0,78%	NT	NT
14MS	PIK3CA/p.H1047R	25,30%	PIK3CA/p.H1047R	25,70%
	KMT2D/p.R845Q	0,67%	NT	NT
17MS	PIK3CA/p.H1047R	26,90%	PIK3CA/p.H1047R	30,10%
	NF1/p.R997M	0,71%	NT	NT

ND, No detectado; NT, No testado, el panel no incluye el gen, VAF- *Variant allele frequency* frecuencia alélica de una variante.

5 Comparando las mutaciones encontradas usando ambos paneles en los 4 tumores de prueba, se observaron frecuencias alélicas totalmente comparables para las mutaciones que fueron detectadas por ambos paneles, validando, por lo tanto, el panel de la invención.

Además, se detectó una mutación con baja frecuencia alélica y patogénica en el gen *CDH1* que no fue detectada en la secuenciación previa.

10 Hay que destacar que usando el panel de la invención se detectaron nuevas mutaciones debido a que este panel incluye la secuenciación de muchos más genes.

Ejemplo 2: Sensibilidad del panel de genes de la Tabla 1 para su uso como biomarcadores de cáncer de mama en biopsia líquida.

Diseño del estudio y población de pacientes

15 El estudio incluyó a 29 pacientes con hallazgos de mamografía BI-RADS 4C/5 y posterior diagnóstico de cáncer de mama primario en el Hospital Virgen de la Victoria de Málaga y el Hospital Clínico Universitario de Valencia. Estos pacientes fueron los mismos que los objeto de estudio en Jimenez Rodriguez, B. et al en "*Detection of TP53 and PIK3CA Mutations in*

Circulating Tumor DNA Using Next-Generation Sequencing in the Screening Process for Early Breast Cancer Diagnosis” (5)

Las muestras de sangre se recolectaron inmediatamente (menos de una hora) antes de la biopsia de tejido y antes de recibir cualquier tipo de tratamiento. Los tejidos tumorales coincidentes se recogieron a través de biopsias con aguja gruesa de tumores de mama, que posteriormente se congelaron en fresco.

Se realizó un análisis inmunohistoquímico (IHC) para cuantificar la expresión del receptor 2 del factor de crecimiento epidérmico humano (HER2), los receptores hormonales (HR) y Ki67.

El receptor de estrógeno (ER) y el receptor de progesterona (PR) se consideraron positivos si los tumores tenían más del 1 % de células con tinción nuclear.

La tinción de HER2 se calificó en una escala de 0 a 3+, de acuerdo con las pautas de HercepTest. El estado de HER2 se consideró positivo cuando se calificó como 3+, mientras que 0 a 1+ fue negativo y 2+ fue un resultado no concluyente y en esos casos se realizó hibridación *in situ* con plata (SISH) para confirmar la positividad.

El tumor con receptores hormonales positivos más un índice Ki 67 de <14 % se consideró un tumor luminal A, mientras que >14 % se consideró un tumor luminal B.

Los subtipos de cáncer de mama IHC se definieron mediante una combinación de estos marcadores IHC de la siguiente manera: luminal A (PR positivo, HER2 negativo y Ki-67 < 14 %), luminal B (ER positivo y/o PR positivo, HER2 negativo y Ki-67 > 14 %), HER2 positivo (ER negativo, PR negativo, HER2 positivo) y triple negativo (ER, PR y HER2 negativo).

El estudio fue aprobado por los comités de ética de investigación del Hospital Virgen de la Victoria y del Hospital Clínico de Valencia, todos los pacientes dieron su consentimiento informado por escrito. La investigación se llevó a cabo de acuerdo con las Buenas Prácticas Clínicas y la Declaración de Helsinki.

Las características clínico-patológicas de los tumores de las pacientes con cáncer de mama inscritas se muestran en la **Tabla 4**. La edad promedio de las pacientes fue de 64 años (rango 44–92 años). Los 29 tumores de mama primarios examinados incluyeron 24 carcinoma ductal invasivo (CDI) (82,7 %), dos carcinomas lobulares invasivos (CLI) (6,9 %), dos carcinomas papilares (6,9 %) y un carcinoma tubular (3,4 %). catorce tumores presentaron diámetros menores a 2 cm (48,3%), mientras que 15 presentaron diámetros

mayores a 2 cm (51,7%). Se detectó metástasis en los ganglios linfáticos axilares mediante biopsia con aguja gruesa en nueve pacientes (31,0 %).

- Diez pacientes sin metástasis en los ganglios linfáticos tenían tumores con diámetros de más de 2 cm (34,5%). Se utilizó inmunohistoquímica para determinar los subtipos de tumores IHC de la siguiente manera, 14 tumores luminales A (48,3 %), 11 tumores luminales B (37,9 %), tres tumores triple negativos (10,3 %) y uno tumor HER2 positivo (3,4 %). Con respecto al estado de los receptores hormonales, 25 de 29 pacientes con BC tenían receptores de estrógeno (ER) positivos (86,2 %) y 17 de 29 pacientes tenían receptores de progesterona (PR) negativos (58,6 %).
- 10 En cuanto al grado tumoral, la mayoría de los pacientes presentaron tumores de grado 2 (16 pacientes, 55%) mientras que siete pacientes tenían tumores de grado 3 (24,5%) y cinco pacientes tenían tumores de grado 1 (17%). Se desconocía el grado tumoral en un paciente (3,5%).

Tabla 4: Características clínico-patológicas de los pacientes.

Edad de diagnóstico (años)	64 años (44 - 92)
Media (rango)	
Tamaño del tumor mamográfico n (%)	
>2 cm	15 (51,7)
2 a 5 cm	13 (44,8)
>5 cm	1 (3,4)
Tipo de tumor	
CDI	24 (82,7)
CLI	2 (6,9)
Carcinoma papilar	2 (6,9)
Carcinoma tubular	1 (3,4)
Grado del tumor n (%)	
I	5 /15,2)
II	16 (55,2)
III	7 (24,13)
Desconocido	1 (3,4)
Nódulo linfático auxiliar	
Positivo	9 (31,3)
Negativo	20 (68,9)
Receptor progesterona n (%)	
Positivo	12 (41,4)
Negativo	17 (58,6)
Receptor estrógeno n (%)	
Positivo	25 (86,20)

Negativo	4 (13,79)
Estatus HER2 n (%)	
Positivo	1 (3,4)
Negativo	28 (96,6)
Subtipo IHC n (%)	
Tumor luminal A	14 (48,3)
Tumor luminal B	11 (37,9)
Tumor HER2 positivo	1 (3,4)
Tumor triple negativo	3 (10,3)
Categoría BIRADS	
4c	14 (48,3)
5	15 (51,7)

CDI: Carcinoma ductal invasivo; CLI: carcinoma lobulillar invasivo; HER2: receptor 2 del factor de crecimiento epidérmico humano; IHC: Inmunohistoquímica; BIRADS: Sistema de datos e informes de imágenes mamarias.

5 Extracción de ADN

El ADN tumoral se aisló de muestras de tejido fresco congelado utilizando el kit de sangre y tejido DNeasy (Qiagen, Valencia, CA, EE. UU.) siguiendo las instrucciones del fabricante.

Las muestras de ADN plasmático se recogieron a partir de muestras de plasma sanguíneo de 10 ml en tubos STRECK inmediatamente antes de la biopsia de tejido. Dentro de las 2h posteriores a la recolección, el plasma se separó de las muestras de sangre completa mediante centrifugación durante 10 min a 3000 rpm a temperatura ambiente y se almacenó a -80°C hasta su uso posterior. Las muestras de plasma se descongelaron y se centrifugaron nuevamente durante 10 minutos a 13.000 rpm a temperatura ambiente antes de la extracción de ADN para eliminar los desechos. El aislamiento de ADN de 2 ml de plasma se realizó con el kit de ácido nucleico circulante QIAamp (Qiagen) de acuerdo con las instrucciones del fabricante y se eluyó en 140 µL de tampón AVE (Qiagen). La cantidad total de ADN genómico humano amplificable aislado de muestras de plasma se cuantificó utilizando una versión modificada del ensayo de PCR en tiempo real del elemento 1 intercalado largo humano (LINE-1) y se notificó como equivalentes del genoma (GE).

20 Secuenciación dirigida en muestras de tejido y del ADN plasmático

El estudio NGS (next generation sequencing) de las muestras se realizó con un diseño personalizado de última generación (NGS - next generation sequencing), en concreto SureSelect XT HS) para los 37 genes de la Tabla 1.

La preparación de la librería de secuenciación y el procesamiento de datos de secuenciación se llevo a cabo siguiendo la metodología que se describe en los respectivos epígrafes del ejemplo 3.

Análisis estadístico

- 5 El análisis estadístico se realizó con SPSS (19.0, SPSS Inc., Chicago, IL, EE. UU.). Para los análisis de asociación entre variables clínico-patológicas y mutaciones derivadas de los genes de la **Tabla 1** en sangre, se utilizaron la correlación de Spearman y la prueba exacta de Fisher (para variables categóricas). El umbral de significación estadística se fijó en $p < 0,05$.
- 10 Los resultados usando el panel de 37 genes tanto para biopsias tumorales como para muestras de plasma se muestran en la **Tabla 5**. Se presenta a su vez una comparativa con los resultados obtenidos para estos mismos 29 pacientes incluidos en el artículo de Jimenez Rodriguez, B. et al "*Detection of TP53 and PIK3CA Mutations in Circulating Tumor DNA Using Next-Generation Sequencing in the Screening Process for Early Breast Cancer* Diagnosis J. Clin Med. 2019 (5), donde se utilizó otra técnica de NGS, Plasma SafeSEQ (Sysmex Inostics S.A.).
- 15

Tabla 5: Resultados comparativos de las mutaciones detectadas en biopsia tumoral y en muestras de plasma utilizando el panel de genes de la Tabla 1; y comparación con las mutaciones en plasma localizadas por otros métodos del estado de la técnica.

Muestra	Biopsia Tumoral				Plasma		
	Gen	Variante / mutación detectada		VAF (%)	VAF (%)	Gen testado en ref. 5	Gen detectado en ref. 5
		Cambio nucleótido	Cambio proteína				
002MS	TP53	c.524G<A	p.R175H	54,5	0,2	SI	NO
007MS	PIK3CA	c.3140A>G	p.H1047R	31,2	0,5	SI	SI
009MS	TP53	c.637 C>T	p.R213*	45,2	0,8	SI	SI
013MS	TP53	c.398T>A	p.m133K	76,9	19,5	SI	SI
015MS	GATA3	c.922-3_922-2 delCA	.	22,4	0,1	NO	-
	TP53	c.377A>G	p.Y126C	58,2	0,2	SI	NO
019MS	PIK3CA	c.1624G>A	p.E542K	53,1	0,1	SI	NO
016MS	TP53	c.743G>T	p.R248L	37,9	4	SI	SI
	SMAD4	c.725C>G	p.S242*	23,7	3,2	NO	-
	PIK3CA	c.3140A>T	p.H1047L	35,1	4,6	SI	SI
021MS	TP53	c.376-2A>G	p.X126_splice	33,4	1,8	SI	NO
033MS	TP53	c.503A>T	p.H168L	37,6	0,4	SI	NO
044MS	KRAS	c.35G>C	p.G12A	29,3	1	NO	-

	<i>TP53</i>	c.587G>C	p.R196P	51,2	1,2	SI	SI
067MS	<i>TP53</i>	c.842A>C	p.D281A	51,8	0,3	SI	NO
	<i>PIK3CA</i>	c.3145G>C	p.G1049R	86,2	0,3	SI	SI
079MS	<i>TP53</i>	c.742C>T	p.R248W	9,1	0,1	SI	NO

La columna “VAF (%)” expresa las frecuencias alélicas observadas usando el panel de la Tabla 1, tanto en tumor sólido como en plasma, donde se llega a detectar frecuencias alélicas de 0,1% en biopsias líquidas.

- 5 El método de la invención es capaz de detectar las variantes o mutaciones concordantes entre plasma y tumor que se describen la referencia 5.

Además, gracias al panel de genes de la **Tabla 1**, es posible detectar 7 variantes en 7 muestras de plasma diferentes, en los genes *PIK3CA* y *TP53* (que son los testados con ambos paneles) y que no se detectaban en la referencia 5. Se señalan estos resultados en negrita, indicándose SI en testado/ NO en detectado en plasma.

10

Al incluir muchas más regiones genéticas en el panel de la Tabla 1, es posible detectar 3 mutaciones más en 3 muestras de plasma diferentes (015MS, 016MS, 044MS).

Ejemplo 3: Sensibilidad y capacidad de detección y capacidad de diagnóstico del panel de secuenciación personalizado.

- 15 Para analizar la sensibilidad del panel de secuenciación de la invención se ha realizado una comparativa entre muestras de pacientes sanos (control) y pacientes que desarrollaron cáncer de mama.

El estudio fue aprobado por el comité de ética local y se realizó de acuerdo con las directrices de Buenas Prácticas Clínicas y la Declaración de Helsinki. Se obtuvo el consentimiento informado de todas las mujeres que fueron reclutadas en el Hospital Virgen de la Victoria de Málaga y en el Hospital Clínico Universitario de Valencia antes de realizar cualquier evaluación requerida según el protocolo.

20

Se han utilizado las siguientes muestras:

- 75 muestras de plasma de pacientes con cáncer de mama, de las que analizables fueron 74. Estas muestras se obtuvieron de pacientes con hallazgos de mamografía BIRADS 4C/5 justo antes de la biopsia de tejido, antes del diagnóstico y de ser sometidas a tratamiento para el cáncer.

25

- 71 biopsias sólidas pre-tratamiento de los pacientes antes citados.
- 22 muestras de plasma de mujeres que fueron incluidas en el estudio por mostrar mamografía sospechosa pero que finalmente el diagnóstico fueron negativas para la presencia cáncer de mama, y que aquí servirán como control negativo.

5 **Tabla 6:** Características clínico-patológicas de las 74 pacientes de las que se extrajeron las muestras de plasma.

Características clínicas	Total
	n
Edad de diagnóstico (años)	
30-50	13
>50	61
Tipo de tumor	
CDI	59
CDIS	5
CLI	3
CP	1
TC	3
MC	3
Tamaño de tumor	
<2cm	32
2-5cm	37
>5cm	5
Grado del Tumor	
I	15
II	37
III	22
Nódulo linfático auxiliar	
Positivo	28
Negativo	38
Desconocido	8
Receptor estrógeno	
Positivo	66
Negativo	7
Desconocido	1
Receptor progesterona	
Positivo	56
Negativo	17
Desconocido	1
Estatus HER2	
Positivo	6
Negativo	66
Desconocido	1

Categoría BIRADS

4/B/C 40

5B/C 33

Desconocido 1

Recaída clínica

Yes 8

No 64

Desconocido 2

Ki67

<20 36

≥20 36

CDI: carcinoma ductal invasivo; CLI: carcinoma lobulillar invasivo; CDIS: carcinoma ductal *in situ*; CP: carcinoma papilar; TC: carcinoma tubular; MC: carcinoma mucinoso; HER2: receptor 2 del factor de crecimiento epidérmico humano; IHC: Inmunohistoquímica; BIRADS: Sistema de datos e informes de imágenes mamarias.

- 5 Las **biopsias tumorales** se extrajeron utilizando agujas gruesas de biopsia y posteriormente se congelaron. Se realizó un análisis inmunohistoquímico para cuantificar la expresión del receptor 2 del factor de crecimiento epidérmico humano (HER2), los receptores hormonales (HR) y Ki67. El receptor de estrógeno (RE) y el receptor de progesterona (PR) se consideraron positivos en tumores que presentaban más del 1% de células teñidas con
- 10 núcleo. La tinción de HER2 se calificó en una escala de 0 a 3+, de acuerdo con las pautas de HercepTest17. El estado de HER2 se consideró positivo cuando se calificó como 3+, mientras que 0 a 1+ fue negativo y 2+ fue un resultado no concluyente y en esos casos se realizó hibridación *in situ* con plata (SISH) para confirmar la positividad. El tumor con receptores hormonales positivos más un índice Ki67 de <14 % se consideró un tumor
- 15 luminal A, mientras que >14 % se consideró un tumor luminal B. Los subtipos IHC BC se definieron mediante una combinación de marcadores inmunohistoquímicos de la siguiente manera: Luminal A (PR positivo, HER2 negativo y Ki67 < 14 %), luminal B (ER positivo y/o PR positivo, HER2 negativo y Ki67 ≥ 14 %), HER2 positivo (ER positivo/negativo, PR positivo/negativo, HER2 positivo) y triple negativo (ER, PR y HER2 negativo).

20 **Procesamiento de muestras de sangre**

Se obtuvieron 10 ml de plasma de cada individuo en tubos STRECK. Dentro de las 2 h posteriores a la recolección, el plasma se aisló de la sangre mediante centrifugación durante 10 min a 3000 rpm a temperatura ambiente y se almacenó a -80 °C hasta la extracción del ADN libre circulante (cfDNA).

Extracción y cuantificación de ADN a partir de plasma y biopsias sólidas

El cfDNA se extrajo de muestras de plasma utilizando el kit de ácido nucleico circulante QIAamp (Qiagen) de acuerdo con las instrucciones del fabricante. El ADN tumoral se aisló de muestras de tejido fresco congelado utilizando el kit de sangre y tejido DNeasy (Qiagen) siguiendo las instrucciones del fabricante. El cfDNA y el ADN de tumores sólidos se cuantificaron mediante PCR digital en gota (ddPCR) y el ensayo RNaseP (Life Technologies) como se indica en Garcia-Murillas I. *et al.* (9).

Diseño del panel de secuenciación para cáncer de mama

El panel NGS personalizado para cáncer de mama se diseñó utilizando el software SureDesign (Agilent) con la siguiente configuración: 5x para “tiling”, menos estricto para enmascarar, XTHSBoosting para potenciar y un valor de 30 para extensión en repeticiones.

El panel incluyó las regiones codificantes de la lista de genes en la **Tabla 1**.

Preparación de la librería de secuenciación

Se empleó la metodología SureSelectXTHS (Agilent) para generar librerías de secuenciación. Se realizaron librerías utilizando una mediana de entrada de ADN plasmático de 39,78 ng (máx. 173,91 ng - mín. 5,01 ng) de pacientes con cáncer de mama y 21,78 ng (máx. 113,52 ng - mín. 1,71 ng) de individuos sanos y una mediana de ADN tisular de 199,5 ng (máx. 200 ng - mín. 6,95 ng) de tumores. El ADN del tejido se fragmentó con el kit de fragmentación enzimática SureSelect (Agilent) y las librerías se prepararon con los kits del sistema de enriquecimiento de diana SureSelectXT (Agilent) siguiendo las indicaciones del fabricante. Todos los pasos de la PCR se llevaron a cabo en el Termociclador C1000 Touch (Bio-Rad).

Los intervalos de fragmentos de las librerías se analizaron con chips de ADN de alta sensibilidad Bioanalyzer (Agilent) y se cuantificaron con el KAPA Library Quantification Kit (Roche). Para la secuenciación del ADN del tejido tumoral, se prepararon y secuenciaron 8 grupos que contenían de 8 a 9 muestras de librería por grupo. Para el ADN de plasma de pacientes con cáncer de mama, también se prepararon y secuenciaron 8 grupos que contenían de 9 a 10 muestras de librería por grupo y 3 grupos que contenían de 7 a 8 muestras de librería por grupo de ADN de plasma de controles sanos. Se emplearon 19 carriles (1 carril por grupo) para secuenciar las librerías con el objetivo de obtener una secuenciación ultra profunda de alrededor de 20,000X antes de la de-duplicación en la plataforma DNBseq-G400 (MGI) en 100 lecturas pareadas siguiendo las instrucciones del

fabricante para la secuenciación de códigos de barras moleculares, también llamados identificadores moleculares únicos (UMI).

Procesamiento de datos de secuenciación

Se creó un método personalizado para el procesamiento de los datos de secuenciación de
 5 SureSelectXT HS (Agilent). Inicialmente se realizó el control de calidad de los datos de
 secuenciación utilizando fastQC v0.11.9. A continuación, se recortaron las lecturas de los
 adaptadores y el filtrado de calidad mediante trim-galore v0.6.7. Para realizar los pasos de
 procesamiento que involucran datos con código de barras, se usó un subconjunto de
 herramientas fgbio v1.5.1. Se asignaron los datos al genoma de referencia GRCh38
 10 utilizando bwa v0.7.17. Luego se usó fgbio GroupReadsByUmi para agrupar por código de
 barras usando la opción Identity para tener en cuenta que los códigos de barras
 SureSelectXT-HS están degenerados. A continuación, se generaron lecturas de consenso
 mediante fgbio CallMolecularConsensusReads. Las lecturas de consenso generadas se
 mapearon nuevamente con bwa v0.7.17. Luego, se filtraron estas lecturas de consenso
 15 alineadas usando fgbio FilterConsensusReads requiriendo una calidad de base mínima de
 30 y manteniendo las lecturas de consenso respaldadas por al menos un número mínimo de
 lecturas. Entonces, se usó fgbio ClipBam para eliminar las regiones superpuestas de
 lecturas hacia adelante y hacia atrás.

Finalmente, se llevó a cabo la identificación de variantes con MuTect2 (gatk v4.2.2.0-1),
 20 incluido un panel de ADN no tumoral y un archivo de anotación de variantes de línea
 germinal para el genoma GRCh38, obtenido del paquete de recursos gatk, que se utiliza
 para anotar variantes para filtrar y solo considerar las regiones incluidas en el panel
 SureSelect. Se anotaron las variantes con ANNOVA v20200608 con información acerca de
 su presencia en las bases de datos públicas COSMIC versión 95 y “The Cancer Genome
 25 Atlas” (TCGA).

Filtración y análisis de variantes

Para el tumor, se utilizó un enfoque más estricto con el fin de generar una referencia sólida
 para comparar con los hallazgos de ctDNA. Se generaron lecturas de consenso que
 requirieron un mínimo de 3 lecturas por familia de código de barra molecular. Se aceptaron
 30 como llamadas válidas solo variantes con VAF > 0,05 que también estaban presentes en
 COSMIC o TCGA (aumentando el umbral de VAF a VAF > 0,2 para tejidos embebidos en
 parafina fijados con formalina (FFPE)).

En el caso de ctDNA, se identificaron mutaciones utilizando dos métodos: i) Estricto; usando el mismo enfoque que se describió anteriormente, pero filtrando para un mínimo de 1 lectura por código de barra molecular, sin aplicar un umbral de VAF. Para considerar las mutaciones no encontradas en el tumor como detectadas en el plasma, se requirió que

5 tuvieran una configuración dúplex, con al menos dos fragmentos mapeados en diferentes coordenadas y que estuvieran presentes tanto en COSMIC como en TCGA-BC. Se aplicó el mismo enfoque de procesamiento a las muestras de control. ii) Exploratorio; visualización de las alineaciones en el navegador del genoma IGV (Robinson, Thorvaldsdottir, Wenger, Zehir y Mesirov, 2017) para identificar mutaciones encontradas en los tumores correspondientes

10 pero que los códigos de barras moleculares automatizados no detectaron. Para estas mutaciones identificadas de manera manual, se contó el número de lecturas que portaban la mutación en dicha muestra y el número de lecturas que no portaban la mutación. Luego, se compararon con las mismas proporciones de lectura en controles y muestras de plasma de cáncer de mama sin la mutación correspondiente mediante una prueba de Fisher.

15 **Análisis estadísticos y visualización de datos**

Se realizó el análisis estadístico y los gráficos de los datos con R (<https://www.R-project.org/>) Se aplicó la prueba exacta de Fisher o la prueba de Chi-cuadrado cuando fue apropiado, tanto para probar la asociación entre las variables clínico-patológicas y los datos de secuenciación de plasma como también en los análisis de datos de secuenciación.

20 También se aplicó la prueba de Wilcoxon para probar las diferencias en la cobertura de secuenciación entre casos y controles. El umbral de significación estadística se estableció en $p < 0,05$.

Los valores de sensibilidad, especificidad y valor predictivo positivo (VPP) se calcularon utilizando el paquete caret v6.0.93. Se utilizó la función oncoplot del paquete maftools

25 (Mayakonda, Lin, Assenov, Plass, & Koeffler, 2018) v2.12.0 para trazar mutaciones y datos clínico-patológicos.

Resultados

En primer lugar, se secuenció el ADN del tumor utilizando la tecnología Agilent SureSelect XT HS, siguiendo las recomendaciones del protocolo, como se indicó anteriormente. La

30 secuenciación del tumor se realizó con una mediana de cobertura de 15,483X. El procesamiento bioinformático posterior que utiliza códigos de barras moleculares para minimizar los errores de secuenciación proporcionó una cobertura mediana final de 1,698X. Entre las regiones capturadas, solo 3 presentaron menos de 100X en más del 10% de las

bases secuenciadas. Además, todos los genes presentaron una cobertura homogénea en todas las muestras. A continuación, se realizó un filtrado personalizado utilizando información de bases de datos genómicas públicas para identificar mutaciones somáticas.

Se identificaron 61 mutaciones en 40/71 (56,33%) de las muestras de tumores. Entre ellos, 33 se ubicaron en el gen *PIK3CA* (54,09%), 12 en *TP53* (19,67%) y 4 en *GATA3* (6,55%) (**Figura 2, Tabla 7**), que representan los genes mutados con mayor frecuencia en nuestro conjunto de tumores

Para estudiar **la relación entre las mutaciones observadas en los tumores y las detectadas también en el plasma**, se aplicó el panel de secuenciación personalizado de la **Tabla 1** junto con una secuenciación de alta profundidad con una mediana de cobertura de 17,704X utilizando el ADN plasmático de los pacientes correspondientes. En total se secuenciaron 74 muestras de plasma de los pacientes. Después del procesamiento de los códigos de barras moleculares, la cobertura mediana fue de 2,525X. Entre las regiones de genes secuenciados, 3 presentaron una cobertura baja y todos los genes mostraron una cobertura homogénea. Después de los análisis bioinformáticos utilizando el identificador de mutaciones establecido, **se encontraron 13/61 (21,31%) mutaciones en plasma que también estaban presentes en los tumores correspondientes**, 7 mutaciones en el gen *TP53* (53,84%) y 3 en *PIK3CA* (27,27%) como los genes mutados con mayor frecuencia.

Además, todas las mutaciones observadas en los tumores se inspeccionaron manualmente en los datos sin procesar de secuenciación de plasma. Los datos alineados se usaron para identificar lecturas de apoyo para los alelos alternativos usando el software IGV. Las mutaciones que se encontraron en al menos 2 lecturas con diferentes coordenadas genómicas pasaron al siguiente paso de análisis.

Para considerar las variantes como válidas, se aplicó una prueba exacta de Fisher utilizando datos de secuenciación de 22 controles sanos de plasma y pacientes que no portaban dicha mutación en sus tumores. En total, **se rescataron 5 mutaciones** de 4 pacientes diferentes de la secuenciación de plasma mediante inspección manual. Entre ellas, 3 mutaciones se localizaron en el gen *TP53* y 2 en *GATA3*. Curiosamente, las 2 variantes estructurales en *GATA3* con estadísticas de secuenciación robustas se recuperaron mediante inspección manual.

Teniendo en cuenta el total de las variantes detectadas, 18 (13 + 5) en plasma y 61 en tumor, el 29,50% de las variantes encontradas en el tejido tumoral también se

localizaron en las muestras de plasma. Por tanto, la sensibilidad de este método sería de alrededor de un 30%.

Tabla 7: Mutaciones detectadas en muestras de tumor y plasma.

Muestra	Tumor				Plasma		
	Gen	Nucleotido cambiado	Aminoacido cambiado	VAF (%)	Detectado por llamador No(N)/ SI (Y)	Detectado manualmente No(N)/ SI(Y)	VAF (%)
001MS	PIK3CA	c.G3145C	p.G1049R	14,6	N	N	-
002MS	CDH1	c.C2245T	p.R749W	5,3	N	N	-
	TP53	c.G524A	p.R175H	54,5	N	Y	0,2
	PIK3CA	c.G1252A	p.E418K	35	N	N	-
	PIK3CA	c.A3140T	p.H1047L	34,4	N	N	-
007MS	PIK3CA	c.A3140G	p.H1047R	31,2	Y	N	0,5
009MS	TP53	c.C637T	p.R213X	45,2	Y	N	0,8
010MS	PIK3CA	c.A3140G	p.H1047R	15	N	N	-
013MS	TP53	c.T398A	p.M133K	76,9	Y	N	19,5
014MS	PIK3CA	c.A3140G	p.H1047R	25,9	N	N	-
015MS	GATA3	c.922-3_922-2delCA	p.X308_splice	22,4	N	Y	0,09
	TP53	c.A377G	p.Y126C	58,2	N	Y	0,24
016MS	TP53	c.G743T	p.R248L	37,9	Y	N	4,0
	SMAD4	c.C725G	p.S242X	23,7	Y	N	3,2
	PIK3CA	c.A3140T	p.H1047L	35,1	Y	N	4,6
017MS	PIK3CA	c.A3140G	p.H1047R	27,3	N	N	-
019MS	PIK3CA	c.G1624A	p.E542K	53,1	N	Y	0,07
021MS	TP53	c.376-2A>G	p.X126_splice	33,4	Y	N	1,8
022MS	KDM6A	c.C1747T	p.Q583X	12,5	N	N	-

023MS	TP53	c.G524A	p.R175H	63	N	N	-
	PIK3CA	c.G1633A	p.E545K	39,3	N	N	-
030MS	GATA3	c.922-3_922-2delCA	p.X308_splice	38,7	N	N	-
	PIK3CA	c.G1633A	p.E545K	77	N	N	-
031MS	PIK3CA	c.G1633A	p.E545K	15,7	N	N	-
032MS	PIK3CA	c.G353A	p.G118D	6,9	N	N	-
	PIK3CA	c.G2908A	p.E970K	14,7	N	N	-
	PIK3CA	c.A3140G	p.H1047R	11,4	N	N	-
	PIK3CA	c.A3140T	p.H1047L	3	N	N	-
033MS	TP53	c.A503T	p.H168L	37,6	Y	N	0,37
035MS	PIK3CA	c.G1633A	p.E545K	34,8	N	N	-
036MS	NF1	c.3478delG	p.G1160Vfs*6	5,5	N	N	-
	PIK3CA	c.A3140T	p.H1047L	31,5	N	N	-
039MS	AKT1	c.G49A	p.E17K	33,7	N	N	-
	NCOR1	c.G6751T	p.G2251C	10,3	N	N	-
040MS	PIK3CA	c.G1093A	p.E365K	21,7	N	N	-
	PIK3CA	c.G1624A	p.E542K	40	N	N	-
044MS	KRAS	c.G35C	p.G12A	29,3	Y	N	0,97
	TP53	c.G587C	p.R196P	51,2	Y	N	1,2
045MS	AKT1	c.G49A	p.E17K	6,9	N	N	-
047MS	PIK3CA	c.A3140G	p.H1047R	7,2	N	N	-
052MS	PIK3CA	c.A1637G	p.Q546R	19,8	N	N	-
	PIK3CA	c.A3073G	p.T1025A	21,6	N	N	-
056MS	PIK3CA	c.G1624A	p.E542K	17,9	N	N	-
057MS	PIK3CA	c.G1633A	p.E545K	32,1	N	N	-
060MS	MAP3K1	c.813_814del	p.R273Sfs*27	11,6	N	N	-
064MS	PIK3CA	c.T1035A	p.N345K	34,9	N	N	-
065MS	GATA3	c.922-3_922-2delCA	p.X308_splice	23,3	N	N	-
	PIK3CA	c.A3140G	p.H1047R	25,9	N	N	-
066MS	ERBB2	c.G2305T	p.D769Y	23,4	N	N	-
	PIK3CA	c.G1624A	p.E542K	27,5	N	N	-
067MS	TP53	c.A842C	p.D281A	51,8	Y	N	0,31
	PIK3CA	c.G3145C	p.G1049R	86,2	Y	N	0,32
079MS	TP53	c.C742T	p.R248W	9,1	N	Y	0,05
	PIK3CA	c.A1637G	p.Q546R	11,9	N	N	-
080MS	PIK3CA	c.G1633A	p.E545K	26,4	N	N	-
081MS	PTEN	c.T406C	p.C136R	54	Y	N	3,3
	TP53	c.G743A	p.R248Q	52,9	Y	N	1,5
093MS	PIK3CA	c.A1634G	p.E545G	29,1	N	N	-
095MS	PIK3CA	c.A3140T	p.H1047L	52	N	N	-
099MS	PIK3CA	c.A3140T	p.H1047L	28,6	N	N	-
101MS	SF3B1	c.A2098G	p.K700E	20,9	N	N	-
104MS	TP53	c.A715G	p.N239D	22,2	N	N	-
107MS	GATA3	c.922-3_922-2delCA	p.X308_splice	40,1	N	Y	0,03

El método de la invención presenta **una sensibilidad (30%) que supera con creces otros métodos conocidos hasta la fecha**, sobre todo si lo comparamos con los dos principales estudios donde se han desarrollado tecnologías/metodologías en las que emplean paneles NGS (amplificación y captura) en el screening del cáncer de mama, esto es, como método

5 diagnóstico. Por un lado, se han realizado estudios utilizando paneles PAN-cáncer como método de diagnóstico, donde se ha llegado a alcanzar una sensibilidad en cáncer de mama de un 33%, similar a la de la presente invención, pero para ello ha sido necesario incluir además otros analitos en circulación sanguínea, tales como proteínas, a modo de biomarcadores adicionales (7).

10 Por otro lado, cuando se han empleado únicamente paneles paciente-específicos ultrasensibles y de amplificación, la sensibilidad obtenida ha sido solamente del 14,3%, siendo además necesaria la secuenciación previa del tumor para poder diseñarlos, lo cual no es necesario según el método de la presente invención. Cabe destacar que esta baja sensibilidad se da a pesar de que se utiliza una de las tecnologías más novedosas en el

15 campo, Ampliseq HD (ThermoFisher).

Por tanto, podemos concluir que es la selección específica de los genes de la **Tabla 1** la que permite obtener una método de diagnóstico de cáncer de mama que supera las limitaciones en cuanto a sensibilidad que presentaban las metodologías conocidas hasta la fecha, basadas también en paneles de secuenciación.

20 **Ejemplo 4: Capacidad del panel de secuenciación de próxima generación (NGS) para detectar pacientes con cáncer de mama después de mamografías sospechosas usando biopsia líquida**

Para investigar la capacidad del panel de secuenciación de próxima generación (NGS) para detectar de forma no invasiva cáncer de mama después de mamografías sospechosas se

25 desarrolló un análisis bioinformático a partir de los datos obtenidos en las muestras del ejemplo 3. En esta ocasión el análisis está orientado a localizar todas las posibles mutaciones que se puedan presentar en plasma, a diferencia del ejemplo anterior donde se hace un análisis informado por la genética del tumor, es decir, donde únicamente se buscan en plasma las mutaciones que aparecen en el tumor.

30 Se empleó Mutect2 para detectar mutaciones en muestras de plasma usando 2 familias de códigos de barras moleculares y sin filtros en frecuencias alélicas variantes (VAF). Las variantes se consideraron asociadas al tumor si i) afectaban las regiones exónicas, ii) estaban anotadas en las bases de datos COSMIC, TCGA BC y TCGA que incluyen todos

los tipos de cáncer, y si iii) presentaban lecturas alineadas en 2 o más coordenadas genómicas diferentes visualizadas manualmente utilizando el software IGV.

En total, **25/74 (33,78 %) de las pacientes presentaron mutaciones tumorales detectadas en plasma (Tabla 8, Figura 1), 16 de las mutaciones no se observaron en el**

análisis anterior informado por la genética del tumor. Los genes más frecuentemente afectados por las 25 mutaciones diferentes detectadas fueron *TP53* (14 mutaciones, 48,27%), *PIK3CA* (4 mutaciones, 13,79%) y *GATA3* (3 mutaciones, 10,34%). Además, se observaron mutaciones de ctDNA en 8 muestras de plasma en cuyas biopsias tumorales correspondientes no se detectaron mutaciones. Además, se observó una mutación en una muestra de plasma sin tejido tumoral disponible.

Tabla 8: Resultados en muestras de plasma de 74 pacientes que han desarrollado cáncer de mama, utilizando el panel de genes de la Tabla 1.

Muestra	Gen	Cambio nucleótido	Aminoácido cambiado	VAF (%)
002MS	<i>TP53</i>	c.G524A	p.R175H	0,21
007MS	<i>PIK3CA</i>	c.A3140G	p.H1047R	0,45
009MS	<i>TP53</i>	c.C637T	p.R213*	0,80
011MS	<i>MAP3K1</i>	c.C1292A	p.S431X	0,24
013MS	<i>TP53</i>	c.T398A	p.M133K	19,5
015MS	<i>GATA3</i>	c.922-3_922-2delCA	p.X308_splice	0,09
	<i>TP53</i>	c.A377G	p.Y126C	0,24
016MS	<i>PIK3CA</i>	c.A3140T	p.H1047L	4,60
019MS	<i>PIK3CA</i>	c.G1624A	p.E542K	0,07
021MS	<i>TP53</i>	c.376-2A>G	p.X126_splice	1,80
030MS	<i>SF3B1</i>	c.C1898T	p.A633V	0,39
049MS	<i>NCOR1</i>	c.C3022T	p.Q1008X	0,36
050MS	<i>SF3B1</i>	c.A2098G	p.K700E	1,40
	<i>TP53</i>	c.G733A	p.G245S	0,32
053MS	<i>HRAS</i>	c.G34T	p.G12C	0,89
056MS	<i>USP9X</i>	c.C1795T	p.R599C	0,48
062MS	<i>BAP1</i>	c.C709T	p.R237C	0,41
067MS	<i>PIK3CA</i>	c.G3145C	p.G1049R	0,32
079MS	<i>TP53</i>	c.C742T	p.R248W	0,05
080MS	<i>PIK3R1</i>	c.C1669T	p.R557X	0,32
081MS	<i>TP53</i>	c.C637T	p.R213X	0,64

	<i>TP53</i>	c.G743A	p.R248Q	1,5
092MS	<i>ARID1A</i>	c.2879-1G>A	p.X960_splice	0,33
094MS	<i>TP53</i>	c.C528A	p.C176X	0,33
099MS	<i>BRCA2</i>	c.G1786A	p.D596N	0,45
101MS	<i>USP9X</i>	c.C1795T	p.R599C	0,21
	<i>CDH1</i>	c.C220T	p.R74X	0,28
105MS	<i>GATA3</i>	c.G917A	p.R306Q	0,28
107MS	<i>GATA3</i>	c.922-3_922-2delCA	p.X308_splice	0,03

En total, se han encontrado mutaciones en 25 de las 74 muestras, lo que supone un 33.78% del total.

En cuanto a los controles, se han localizado las siguientes mutaciones.

- 5 **Tabla 9:** Resultados en muestras de plasma de 22 sujetos sanos, utilizando el panel de genes de la Tabla 1.

Muestra	Gen	Cambio nucleótido	Aminoácido cambiado	VAF
042MS	<i>MAP3K1</i>	c.A3373G	p.N1125D	0,0031
077MS	<i>ERBB2</i>	c.G2524A	p.V842I	0,0027
086MS	<i>SMAD4</i>	c.G794A	p.R265H	0,0049

- Únicamente se han encontrado mutaciones en 3 de las 22 muestras control analizadas, es decir en un 13.63% de las muestras. Se trata de una mutación del gen *MAP3K1* (p.N1125D) descrito en la base de datos COSMIC en una muestra de tumor de cáncer de mama, una mutación en el gen *ERBB2* (p.V842I) que se ha observado con una frecuencia sustancialmente mayor en cánceres de colon y endometrio y otra mutación localizada en el gen *SMAD4* (p.R361H) también frecuente en adenocarcinoma de colon; y por otro lado, que el método de la invención tiene una **capacidad de detección de frecuencias alélicas** de hasta 0,03%, lo cual iguala, cuando no supera, otros estudio previos anteriormente citados en los antecedentes de la invención.

- Con estos mismos datos se ha estudiado la **capacidad de diagnóstico del panel de secuenciación de la Tabla 1** utilizando un análisis estadístico, en concreto: MedCalc Software Ltd. Diagnostic test evaluation calculator.
https://www.medcalc.org/calc/diagnostic_test.php.

Los resultados de dicho análisis se muestran a continuación.

Tabla 10: Resultados.

Estadística	Valor	95% IC – intervalo de confianza
Sensibilidad	31,08%	20.83% a 42.90%,
Especificidad	86.36%	65.09% a 97.09%
Razón de verosimilitud positiva	2.35	0.78 a 7.06
Razón de verosimilitud negativa	0.79	0.63 a 0.99
Valor predictivo positivo (*)	88.46%	71.75% a 95.86%.
Valor predictivo negativo (*)	26.39%	22.25% a 30.99%
Precisión (*)	44.00%	34.08% a 54.28%

(*) Valores dependientes de la prevalencia de la enfermedad.

5 Sensibilidad: probabilidad de que el resultado de una prueba sea positivo cuando la enfermedad está presente (tasa de verdaderos positivos).

Especificidad: probabilidad de que el resultado de una prueba sea negativo cuando la enfermedad no está presente (tasa de verdaderos negativos).

10 Razón de verosimilitud positiva: relación entre la probabilidad de un resultado positivo de la prueba dada la presencia de la enfermedad y la probabilidad de un resultado positivo de la prueba dada la ausencia de la enfermedad.

Razón de verosimilitud negativa: relación entre la probabilidad de un resultado negativo de la prueba dada la presencia de la enfermedad y la probabilidad de un resultado negativo de la prueba dada la ausencia de la enfermedad.

15 Valor predictivo positivo: probabilidad de que la enfermedad esté presente cuando la prueba es positiva.

Valor predictivo negativo: probabilidad de que la enfermedad no esté presente cuando la prueba es negativa.

Precisión: probabilidad global de que un paciente esté correctamente clasificado.

20 Tal y como se desprende de este análisis, el empleo del panel personalizado junto con una secuenciación ultra profunda y un análisis bioinformático personalizado no informado por el tumor condujo a una sensibilidad del 31,08 %, **el método de la invención presenta una muy alta especificidad, de un 86,36% y un valor predictivo positivo de un 88,46%, por**

lo que, en caso de presentarse alguna mutación en alguno de los genes del panel de la tabla 1, existe una probabilidad de casi un 90% de presentar un cáncer de mama.

Con todo esto podemos concluir que la presente invención aporta, gracias al panel de secuenciación seleccionado, una metodología que permite el diagnóstico de cáncer de mama, incluso a partir de biopsias líquidas, con una sensibilidad, capacidad de detección de frecuencias alélicas bajas y una especificidad y valor predictivo positivo, que supera las limitaciones de otros métodos similares conocidos hasta la fecha.

También se investigó la **asociación de variables clínico-patológicas** con la detección de mutaciones en plasma. En concreto, se estudió la positividad de ctDNA en plasma, la frecuencia alélica mediana (FA) de las mutaciones, el número de mutaciones por muestra y las muestras con mutaciones en *TP53* para determinar su asociación con las características clínicas (**Tablas 11 y 12**). En general, **una mediana de FA más alta se asoció con un grado tumoral más alto ($p= 0,04639$), la presencia de más de 1 mutación en plasma con probabilidad de recaída clínica ($p= 0,02372$) y las mutaciones en *TP53* se observaron más frecuentemente en tumores negativos a la expresión de receptores hormonales (HR), en concreto receptor estrógeno (ER) negativos ($p= 0,0316$) y receptor progesterona (PR) negativos ($P=0,0257$)**. Además, la asociación de la recaída clínica y las mutaciones en plasma con una mediana de frecuencia alélica (FA) alta, definida como mutaciones con $> 0,05\%$ en la FA, fue interesantemente cercana a la significación ($p = 0,059$).

Cabe destacar que el 38,35% de los pacientes incluidos en este estudio eran asintomáticos y fueron diagnosticados por el programa de detección precoz de cáncer de mama. De ellos, el 28,57% presentaba mutaciones plasmáticas, porcentaje similar al 33,33% de mujeres sintomáticas con mutaciones.

Tabla 11: Relación entre características clínicas y ctDNA positivos y frecuencia alélica mediana (FA) en plasma.

Características clínicas	Total	ctDNA-positivo	p-valor	Total	Mediana FA >0,5%	p-valor
	n	n (%)		n	n (%)	
Edad de diagnóstico (años)			n.s.			n.s.
30-50	13	4 (30,76%)		4	2 (50%)	
>50	61	19 (31,14%)		19	4 (21,05%)	
Tamaño de tumor			n.s.			n.s.
<2cm	32	10 (31,25%)		10	2 (20%)	

2-5cm	37	11 (29,72%)		11	3 (27,27%)	
>5cm	5	2 (40%)		2	1 (50%)	
Grado del Tumor			n.s.			<0,05
I	15	2 (13,33%)		2	1 (50%)	
II	37	15 (40,54%)		15	1 (6,66%)	
III	22	6 (27,27%)		6	3 (50%)	
Nódulo linfático auxiliar			n.s.			n.s.
Positivo	36	9 (25%)		9	4 (44,44%)	
Negativo	30	11 (36,66%)		11	2 (18,18%)	
Receptor estrógeno			n.s.			n.s.
Positivo	66	20 (30,30%)		20	4 (20%)	
Negativo	7	3 (42,85%)		3	2 (66,66%)	
Receptor progesterona			n.s.			n.s.
Positivo	56	16 (28,57%)		16	3 (18,75%)	
Negativo	17	7 (41,17%)		7	3 (42,85%)	
Estatus HER2			n.s.			n.s.
Positivo	6	2 (33,33%)		1	1 (100%)	
Negativo	67	21 (31,34%)		22	5 (22,72%)	
ki67			n.s.			n.s.
<20	36	11 (30,55%)		11	2 (18,18%)	
>=20	36	12 (33,33%)		12	4 (33,33%)	
Categoría BIRADS			n.s.			n.s.
4/B/C	40	12 (30%)		12	2 (16,66%)	
5B/C	33	11 (33,33%)		11	4 (36,36%)	
Recaída clínica			n.s.			0,059
SI	8	2 (25%)		2	2 (100%)	
No	64	21 (32,81%)		21	4 (19,94%)	

Tabla 12: Relación entre características clínicas y 2 mutaciones en plasma y TP53 Mutado.

Características clínicas	Total	2 mutaciones en plasma	p-valor	Total	TP53 Mutado	p-valor
	n	n (%)		n	n (%)	
Edad de diagnóstico (años)			n.s.			n.s.
30-50	4	0 (0%)		4	3 (75%)	
>50	19	3 (15,78%)		19	5 (26,31%)	
Tamaño de tumor			n.s.			n.s.
<2cm	10	1 (10%)		10	4 (40%)	
2-5cm	11	3 (27,27%)		11	4 (36,36%)	
>5cm	2	0 (0%)		2	0 (0%)	
Grado del Tumor			n.s.			n.s.
I	2	0 (0%)		2	0 (0%)	
II	15	2 (13,33%)		15	4 (26,66%)	

III	6	2 (33,33%)		6	4 (66.66%)	
Nódulo linfático auxiliar			n.s.			n.s.
Positivo	9	2 (22,22%)		9	4 (44.44%)	
Negativo	11	1 (9,09%)		11	3 (27.27%)	
Receptor estrógeno			n.s.			<0.05
Positivo	20	4 (20%)		20	5 (25%)	
Negativo	3	0 (0%)		3	3 (100%)	
Receptor progesterona			n.s.			<0.05
Positivo	16	2 (12,5%)		16	3 (18.75%)	
Negativo	7	2 (28,57%)		7	5 (71.42%)	
Estatus HER2			n.s.			n.s.
Positivo	1	0 (0%)		1	1 (100%)	
Negativo	22	4 (18,18%)		22	7 (31.81%)	
ki67			n.s.			n.s.
<20	11	3 (27,27%)		11	3 (27.27%)	
>=20	12	1 (8,33%)		12	5 (41.66%)	
Categoría BIRADS			n.s.			n.s.
4/B/C	12	2 (16,66%)		12	4 (33.33%)	
5B/C	11	2 (18,18%)		11	4 (36.36%)	
Recaída clínica			<0,05			n.s.
SI	2	2 (100%)		2	2 (100%)	
No	21	2 (9,5%)		21	6 (28,57%)	

En conclusión, observamos una concordancia de mutaciones entre tumor y plasma del 29,50% con una sensibilidad por debajo del 0,075% en la frecuencia de alelos mutantes. En el enfoque no informado por el tumor, detectamos mutaciones en el 33,78% de las muestras de plasma observando 8 pacientes con mutaciones solo en plasma. Determinamos una especificidad del 86,36% y un valor predictivo positivo del 88,46%. Además, demostramos una asociación entre una mediana de FA más alta y un grado tumoral más alto, la presencia de múltiples mutaciones en plasma con probabilidad de recaída clínica y mutaciones en *TP53* con positividad del receptor de estrógeno. Se trata por tanto de un panel de secuenciación único con una tecnología que no se había empleado anteriormente en el cáncer de mama temprano, demostrando una sensibilidad ultra alta y allanando el camino para su uso en la detección del cáncer de mama.

Teniendo en cuenta lo anterior, el panel de secuenciación de la invención para plasma mostró capacidades mejoradas para detectar ctDNA en pacientes con cáncer de mama localizados en las primeras etapas de diagnóstico, mejorando la sensibilidad de detección de investigaciones anteriores y agregando evidencia de que ctDNA podría ayudar en el

proceso de diagnóstico de la población asintomática. Esto también está respaldado por los porcentajes similares en la identificación de mutaciones plasmáticas entre mujeres sintomáticas y asintomáticas de 33,33% y 28,57% respectivamente. En este sentido, desarrollamos una metodología bioinformática personalizada para identificar mutaciones plasmáticas sin información tumoral, demostrando un alto VPP y sugiriendo que enfoques similares podrían constituir una herramienta de detección para el cáncer de mama. También queda demostrado que mediante la secuenciación temprana del ADN plasmático de los pacientes con cáncer de mama, es factible obtener información importante sobre la enfermedad, así como predecir la evolución clínica en estas pacientes.

10 Referencias

1. Alfonso, Alba-Bernal *et al.*; Challenges and achievements of liquid biopsy technologies employed in early breast cancer. *eBioMedicine* – Volumen 62, 103100. <https://doi.org/10.1016/j.ebiom.2020.103100>.
- 15 2. Sun, M.-Y.; *et al.*. Targeted Next-Generation Sequencing of Circulating Tumor DNA Mutations among Metastatic Breast Cancer Patients. *Curr. Oncol.* 2021, 28, 2326-2336. <https://doi.org/10.3390/curroncol28040214>.
3. Priskin, K.; *et al.*, BC-Monitor: Towards a Routinely Accessible Circulating Tumor DNA-Based Tool for Real-Time Monitoring Breast Cancer Progression and Treatment Effectiveness. *Cancers* 2021, 13, 3489. <https://doi.org/10.3390/cancers13143489>.
- 20 4. Keup, C., Benyaa *et al.* Targeted deep sequencing revealed variants in cell-free DNA of hormone receptor-positive metastatic breast cancer patients. *Cell. Mol. Life Sci.* 77, 497–509 (2020). <https://doi.org/10.1007/s00018-019-03189-z>.
- 5- Jimenez Rodriguez, B.; *et al.*, Detection of *TP53* and *PIK3CA* Mutations in Circulating Tumor DNA Using Next-Generation Sequencing in the Screening Process for Early Breast Cancer Diagnosis. *J. Clin. Med.* 2019, 8, 1183. <https://doi.org/10.3390/jcm8081183>
- 25 identificadores moleculares únicos 6. Pereira, B., *et al.*, The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun* 7, 11479 (2016). <https://doi.org/10.1038/ncomms11479>.
- 30 7. Cohen JD *et al.*, Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018 Feb 23;359(6378):926-930. doi: 10.1126/science.aar3247.
8. Page K, *et al.*, Prevalence of ctDNA in early screen-detected breast cancers using highly sensitive and specific dual molecular barcoded personalised mutation assays. *Ann Oncol.*

2021 Aug;32(8):1057-1060. doi: 10.1016/j.annonc.2021.04.018. Epub 2021 Apr 29. PMID: 33932505.

9. Garcia-Murillas, I. *et al.* Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* **7**, 302ra133 (2015).

REIVINDICACIONES

1. Panel de secuenciación que comprende los genes de la Tabla 1
2. Uso del panel de secuenciación de la Tabla 1 para el diagnóstico o pronóstico del cáncer de mama.
- 5 3. Uso del panel de secuenciación según la reivindicación anterior en biopsias líquidas.
4. Un método *in vitro* de obtención de datos útiles para diagnosticar, predecir o pronosticar la evolución del cáncer de mama, a partir de una muestra biológica previamente aislada de un individuo, que comprende identificar posibles mutaciones en los genes de la Tabla 1.
- 10 5. Un método *in vitro* para diagnosticar o diagnosticar precozmente el cáncer de mama que comprende el método de obtención de datos útiles de la reivindicación 4 y que además comprende asignar al individuo que presenta mutaciones en alguno de los genes de la Tabla 1 al grupo de pacientes que padece cáncer de mama.
6. Un método *in vitro* para diagnosticar el cáncer de mama según la reivindicación anterior que además comprende asignar al individuo que presenta una mediana de frecuencia alélica de las mutaciones de los genes de la Tabla 1 más alta, al grupo de
15 pacientes con un grado tumoral más alto.
7. Un método *in vitro* para predecir o pronosticar la evolución del cáncer de mama que comprende el método de obtención de datos útiles de la reivindicación 4 y que además
20 comprende asignar al individuo que presenta más de una mutación en los genes de la Tabla 1, al grupo de pacientes con probabilidad de recaída clínica.
8. Un método *in vitro* para predecir o pronosticar la evolución del cáncer de mama que comprende el método de obtención de datos útiles de la reivindicación 4 y que además comprende asignar al individuo que presenta mutaciones en *TP53* al grupo de pacientes con
25 tumor negativo a la expresión de receptores hormonales.
9. El método según cualquiera de las reivindicaciones 4 a 8 donde la muestra biológica es plasma.
10. El método según cualquiera de las reivindicaciones 4 a 9, en el que las mutaciones en los genes de la Tabla 1 se detectan mediante tecnologías de secuenciación de última
30 generación o *next generation sequencing*.

11. El método según la reivindicación anterior donde la tecnología de secuenciación de última generación utilizada son sondas de enriquecimiento tipo Agilent Sure Select.
12. Un kit que comprende los elementos necesarios para la detección simultánea de mutaciones en cualquiera de los genes de la Tabla 1.
- 5 13. Uso del kit según la reivindicación 12 para diagnosticar o pronosticar la evolución del cáncer de mama.
14. Un programa de ordenador adaptado para que cualquier medio de procesamiento pueda llevar a la práctica el método según cualquiera las reivindicaciones 4 a 11.
- 10 15. Un medio de almacenamiento legible por un ordenador que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo los pasos del método según cualquiera las reivindicaciones 4 a 11.
16. Una señal transmisible que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo los pasos del método según cualquiera las reivindicaciones 4 a 11.

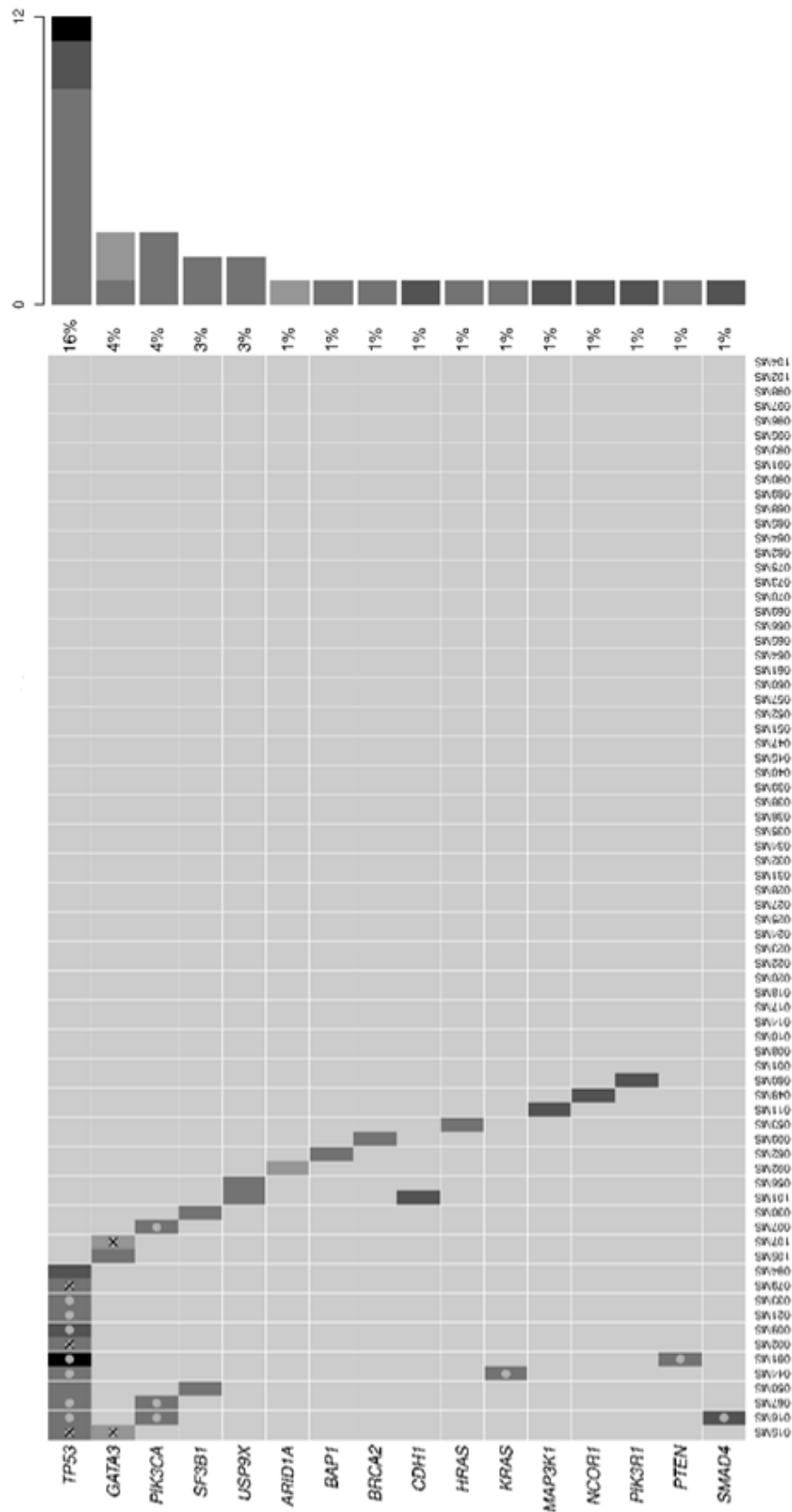


FIG 1.A

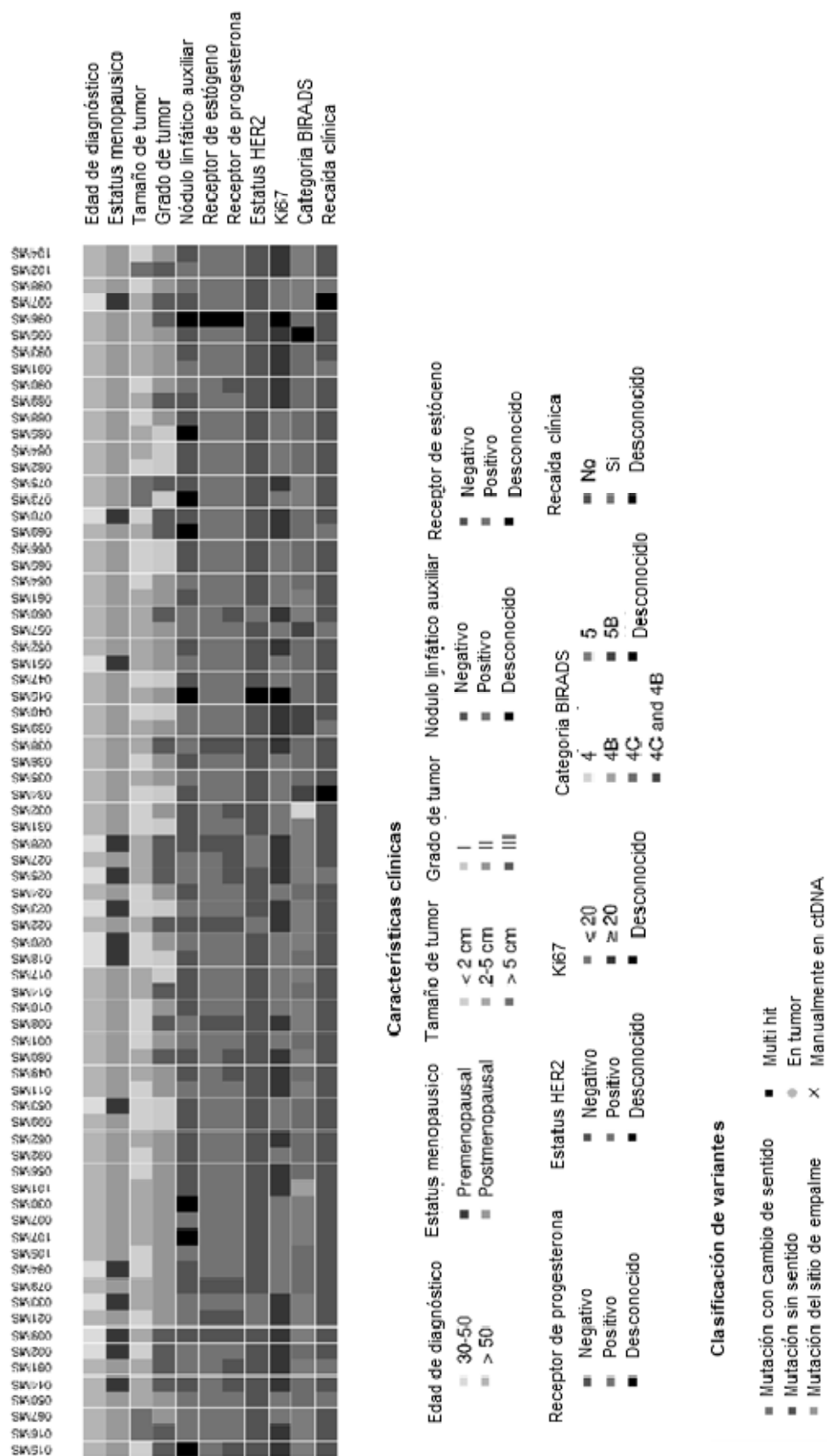


FIG 1.B

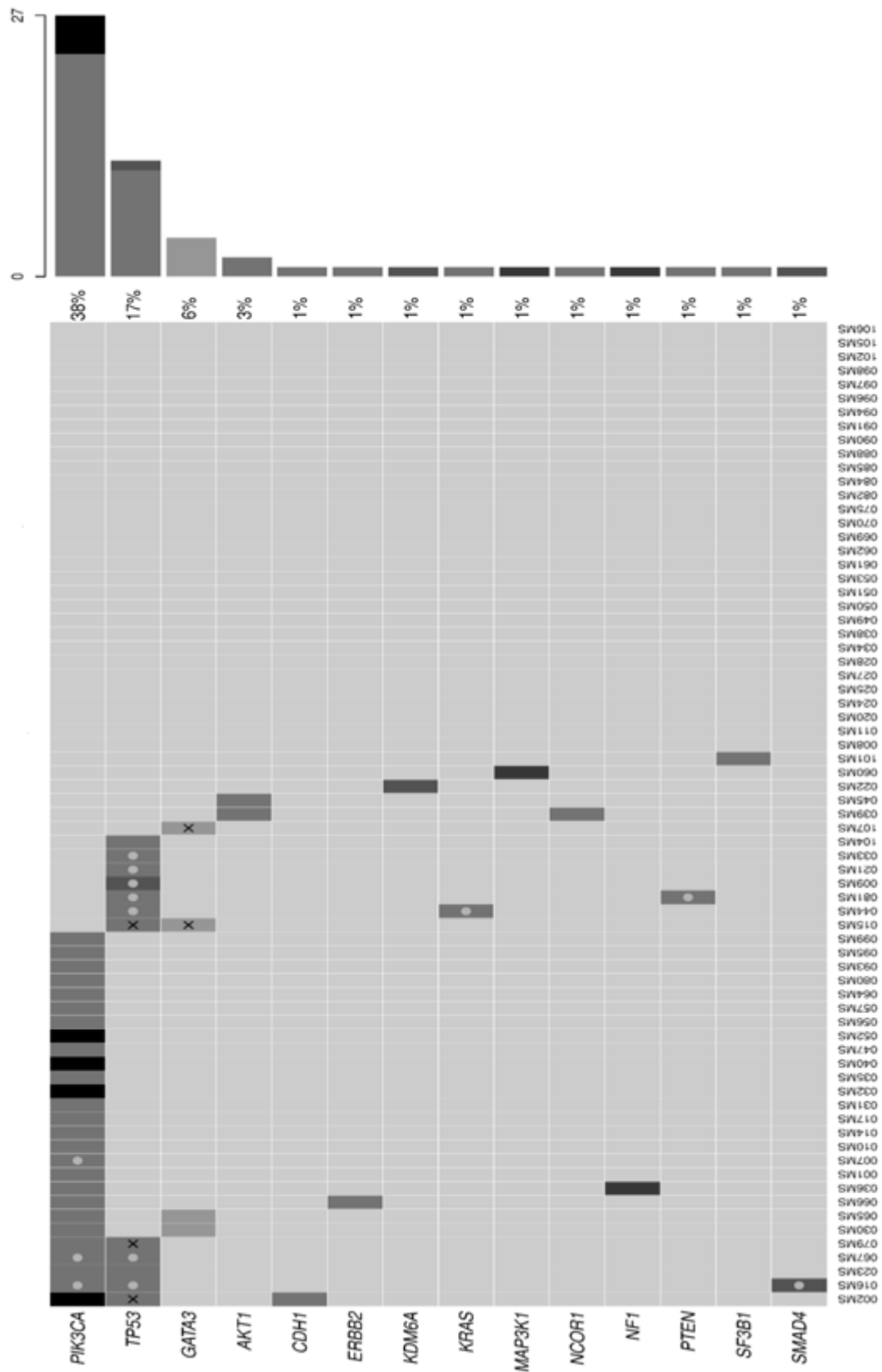


FIG. 2A

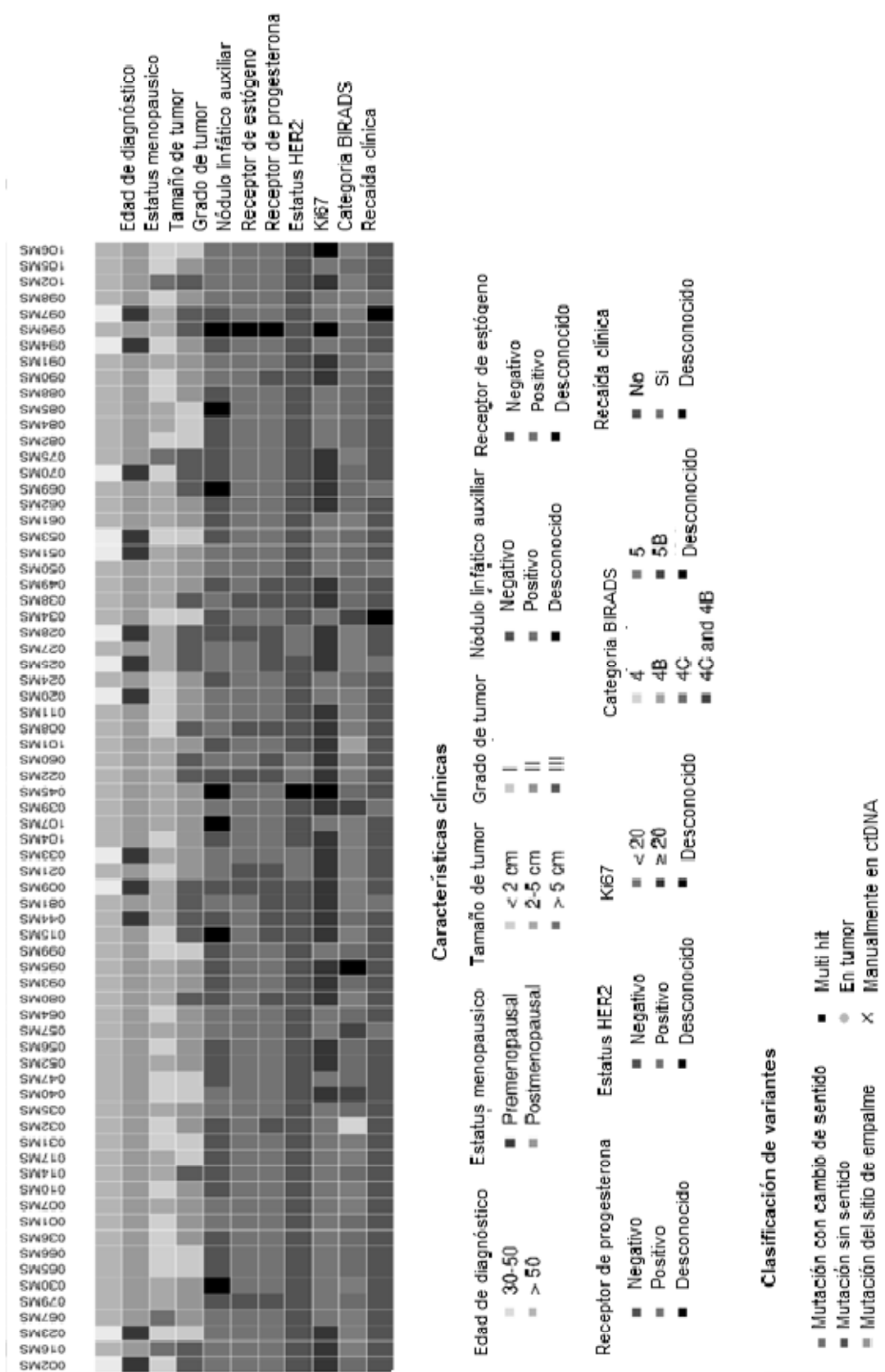


FIG. 2B



- ②① N.º solicitud: 202230937
②② Fecha de presentación de la solicitud: 28.10.2022
③② Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

⑤① Int. cl.: Ver Hoja Adicional

DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
X	US 2021343372 A1 (TELL ROBERT et al.) 04/11/2021. Párrafo [0224], lista 1; párrafo [0225], lista 2.	1
X	YOSHINAMI TETSUHIRO et al. Detection of ctDNA with Personalized Molecular Barcode NGS and Its Clinical Significance in Patients with Early Breast Cancer. Translational Oncology AUG 2020. 31/07/2020, Vol. 13, Nº 8, Páginas Article No.: 100787, ISSN 1936-5233(print) ISSN 1936-5233(electronic), <DOI: doi: 10.1016/j.tranon.2020.100787>. figura 13, fig 1B.	2-16
X	WO 2017181146 A1 (GUARDANT HEALTH INC) 19/10/2017. Párrafo [0159], reivindicaciones 1, 4, 11.	1
X	WO 2022225933 A1 (NATERO INC) 27/10/2022, párrafo [0571]; reivindicaciones 1, 5, 15.	1, 2
X	KEUP CORINNA et al. Targeted deep sequencing revealed variants in cell-free DNA of hormone receptor-positive metastatic breast cancer patients. Cellular and Molecular Life Sciences FEB 2020. 31/01/2020, Vol. 77, Nº 3, Páginas 497-509, ISSN 1420-682X(print) ISSN 1420-9071(electronic), <DOI: doi:10.1007/s00018-019-03189-z>. Fig. 2d.	2-16
X	PRISKIN KATALIN et al. BC-Monitor: Towards a Routinely Accessible Circulating Tumor DNA-Based Tool for Real-Time Monitoring Breast Cancer Progression and Treatment Effectiveness. Cancers JUL 2021. 30/06/2021, Vol. 13, Nº 14, Páginas Article No.: 3489, ISSN 2072-6694(electronic), <DOI: doi:10.3390/cancers13143489>. tabla 1.	2-16
X	SUN MIN-YING et al. Targeted Next-Generation Sequencing of Circulating Tumor DNA Mutations among Metastatic Breast Cancer Patients. Current oncology (Toronto, Ont.) Switzerland 24 06 2021. 24/06/2021, Vol. 28, Nº 4, Páginas 2326 - 2336, ISSN 1718-7729 (Electronic), <DOI: doi:10.3390/curroncol28040214 pubmed:34202466>. fig.1.	2-16

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones

☐ para las reivindicaciones nº:

Fecha de realización del informe
20.09.2023

Examinador
J. Manso Tomico

Página
1/2

CLASIFICACIÓN OBJETO DE LA SOLICITUD

C12Q1/6886 (2018.01)

G16B20/20 (2019.01)

G16H50/20 (2018.01)

G01N33/574 (2006.01)

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

C12Q, G16H, G01N, G16B

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, EMBASE, BIOSIS, MEDLINE, NPL