



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
MAESTRÍA EN CIENCIAS BIOQUÍMICAS

IDENTIFICACIÓN DE ALTERACIONES EN EL NÚMERO DE COPIAS
DE DNA EN TUMORES DE MAMA MEDIANTE SECUENCIACIÓN
GENÓMICA CON NANOPOROS

TESIS

QUE PARA OPTAR POR EL GRADO DE
Maestra en Ciencias

PRESENTA

Q.F.B. ILSE ALEJANDRA COLORADO GONZÁLEZ

Tutor principal

DR. LUIS ALONSO HERRERA MONTALVO
Instituto de Investigaciones Biomédicas, UNAM

Comité tutor

DR. OSBALDO RESENDIS ANTONIO
Instituto Nacional de Medicina Genómica
DR. ALEJANDRO MANUEL GARCÍA CARRANCÁ
Instituto de Investigaciones Biomédicas, UNAM

Ciudad de México. Enero, 2024



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

COLORADO GONZÁLEZ ILSE ALEJANDRA
Estudiante de Maestría en Ciencias Bioquímicas
Presente

Los miembros del Subcomité Académico en reunión ordinaria del **14 de noviembre de 2023**, conocieron su solicitud de asignación de **JURADO DE EXAMEN** para optar por el grado de **Maestra en Ciencias**, con la réplica de la tesis "**IDENTIFICACIÓN DE ALTERACIONES EN EL NÚMERO DE COPIAS DE DNA EN TUMORES DE MAMA MEDIANTE SECUENCIACIÓN GENÓMICA CON NANOPOROS**", dirigida por el/la Dr(a). **HERRERA MONTALVO LUIS ALONSO**.

De su análisis se acordó nombrar el siguiente jurado integrado por los doctores:

CANIZALES QUINTEROS SAMUEL	PMDCBQ	PRESIDENTE
OSTROSKY-WEGMAN SHEJET PATRICIA	PMDCBQ	SECRETARIO
JIMENEZ MORALES SILVIA	PMDCBQ	VOCAL
RODRÍGUEZ DORANTES MAURICIO	PMDCBQ	VOCAL
ROMERO CORDOBA SANDRA LORENA	PMDCBQ	VOCAL

Sin otro particular por el momento, aprovecho la ocasión para enviarle un cordial saludo.

Atentamente
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cd. Universitaria, Cd. Mx., a 14 de noviembre de 2023



Coordinadora
Dra. Claudia Lydia Treviño Santa Cruz



Instituto Nacional de Medicina Genómica MÉXICO

El presente trabajo fue realizado en el Laboratorio de Genómica del Cáncer (LGC) del Instituto Nacional de Medicina Genómica (INMEGEN), bajo la dirección del Dr. Luis Alonso Herrera Montalvo, y la tutela y revisión del Dr. Alfredo Hidalgo Miranda. Durante la realización de este proyecto de investigación se contó con el apoyo de la beca de maestría otorgada por el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT, CVU: 1083037).

AGRADECIMIENTOS INSTITUCIONALES Y ACADÉMICOS

A la *Universidad Nacional Autónoma de México* y al *Instituto Nacional de Medicina Genómica*, por su contribución fundamental a mi formación académica y profesional. Agradezco profundamente por haberme brindado las bases que sustentan mi presente y futuro académico.

Al *Dr. Luis Alonso Herrera Montalvo*, por su apoyo, respaldo y confianza durante la realización de mi maestría.

Al *Dr. Alfredo Hidalgo Miranda*, por su continua guía en el desarrollo de este trabajo y por fomentar siempre en mí el pensamiento crítico. Gracias por todo el apoyo que me ha brindado desde la licenciatura y por seguirme permitiendo aprender de sus conocimientos.

A los integrantes del comité tutor, el *Dr. Osbaldo Resendis Antonio* y el *Dr. Alejandro Manuel García Carrancá*, por sus valiosas observaciones y aportaciones durante mi formación que contribuyeron a la mejora de este proyecto a lo largo de cada semestre.

A la *Dra. Laura Lucila Gómez Romero*, por todas las asesorías y enseñanzas que recibí para poder llevar a cabo el análisis bioinformático de mis datos, destacando siempre su calidez humana. Sin duda, su apoyo fue fundamental para construir y darle dirección al trabajo.

A la *Dra. Lizbett Hidalgo Pérez*, cuyo trabajo fue fundamental para realizar el procesamiento de las biopsias de tejido tumoral empleadas en este estudio.

A la *Dra. Mireya Cisneros Villanueva*, por su apoyo técnico durante la ejecución de los experimentos de secuenciación.

Al *Dr. Javier Ordoñez Hernández*, por todos los consejos y el acompañamiento que me brindó para poder culminar mi maestría de la mejor forma posible.

Agradezco a todos los miembros del jurado de examen por su tiempo en la revisión y su valiosa retroalimentación que me permitió enriquecer y mejorar este trabajo.

DEDICATORIA

A mis padres

Gracias por estar siempre presentes en cada etapa de mi vida y nunca soltar mi mano.

Mi manera de demostrar el agradecimiento y amor que les tengo es tratar de hacerlos sentir orgullosos. Espero haberlo logrado. Este trabajo es para ustedes.

*En la travesía académica,
las dificultades no son obstáculos,
sino oportunidades de redescubrimiento y redefinición.*

Autor desconocido

CONTENIDO

ABREVIATURAS Y SIGLAS	1
RESUMEN	3
ABSTRACT	4
1 MARCO TEÓRICO	5
1.1 INESTABILIDAD GENÓMICA	5
1.1.1 Tipos de inestabilidad genómica	6
1.2 ALTERACIONES EN EL NÚMERO DE COPIAS DE DNA.....	8
1.2.1 Clasificación de las alteraciones en el número de copias de DNA.....	9
1.2.2 Panorama genómico de las alteraciones en el número de copias de DNA en tumores humanos.....	11
1.2.2.1 Cromosomas afectados con mayor frecuencia por alteraciones en el número de copias de DNA.....	13
1.3 MÉTODOS DE DETECCIÓN DE ALTERACIONES EN EL NÚMERO DE COPIAS DE DNA.....	13
1.3.1 Métodos citogenéticos.....	15
1.3.2 Métodos basados en biología molecular	16
1.3.2.1 Técnicas dirigidas.....	16
1.3.2.2 Técnicas genómicas.....	17
1.4 SECUENCIACION CON NANOPOROS	22
1.4.1 Fundamento de la secuenciación con nanoporos.....	23
1.4.2 Aplicación de la tecnología en la detección de de alteraciones en el número de copias de DNA	25
1.5 ESTRATEGIA BIOINFORMÁTICA PARA LA DETECCIÓN DE ALTERACIONES EN EL NÚMERO DE COPIAS DE DNA.....	26
1.6 LAS BIOPSIAS LÍQUIDAS COMO HERRAMIENTA CLÍNICA PARA LA DETECCIÓN DE ALTERACIONES EN EL NÚMERO DE COPIAS DE DNA	28
1.6.1 Limitaciones de las biopsias de tejido convencionales.....	28
1.6.2 Biopsias líquidas	29
1.6.2.1 DNA libre circulante.....	30
2 ANTecedentes	32

3	JUSTIFICACIÓN	34
4	PREGUNTA DE INVESTIGACIÓN	35
5	HIPÓTESIS	36
6	OBJETIVOS	37
6.1	OBJETIVO GENERAL	37
6.2	OBJETIVOS PARTICULARES	37
7	ESTRATEGIA EXPERIMENTAL	38
8	METODOLOGÍA	40
8.1	ESTANDARIZACIÓN DE LA TÉCNICA DE SECUENCIACIÓN GENÓMICA CON NANOPOROS DE FRAGMENTOS CORTOS DE DNA.....	40
8.1.1	Cultivo de líneas celulares	40
8.1.2	Extracción, cuantificación y fragmentación de DNA genómico	40
8.1.3	Preparación de librerías para secuenciación.....	41
8.2	ANÁLISIS BIOINFORMÁTICO	42
8.2.1	Procesamiento de datos crudos de secuenciación.....	42
8.2.2	Alineamiento de lecturas al genoma de referencia	42
8.2.3	Detección de alteraciones en el números de copias de DNA.....	42
8.2.3.1	Archivos de salida	44
8.3	VALIDACIÓN DE LOS RESULTADOS OBTENIDOS EN LÍNEAS CELULARES TUMORALES.....	45
8.3.1	Validación de los resultados mediante datos de secuenciación Illumina	45
8.3.1.1	Análisis bioinformático	46
8.3.1.2	Análisis de correlación de datos de segmentación entre tecnologías de secuenciación.....	46
8.3.1.3	Comparación de regiones genómicas alteradas a nivel citogenético entre tecnologías de secuenciación	47
8.4	IMPLEMENTACIÓN DEL FLUJO DE TRABAJO ESTANDARIZADO EN EL ANÁLISIS DE MUESTRAS DE BIOPSIA LÍQUIDA	48
8.4.1	Selección de muestras de pacientes	48
8.4.1.1	Criterios de inclusión, exclusión y eliminación	48
	Criterios de inclusión.....	48

8.4.2	Obtención y manejo de muestras.....	49
8.4.2.1	Obtención y manejo de muestras de sangre total.....	49
8.4.2.2	Obtención y manejo de muestras de tejido tumoral.....	49
8.4.3	Detección de alteraciones en el número de copias en biopsias de tejido tumoral mediante microarreglos de alta resolución.....	50
8.4.3.1	Extracción y cuantificación de DNA genómico.....	50
8.4.3.2	Procesamiento de muestras e hibridación al microarreglo de alta resolución.....	50
8.4.3.3	Análisis de datos.....	51
8.4.4	Detección de alteraciones en el número de copias en biopsias líquidas mediante secuenciación con nanoporos.....	51
8.4.4.1	Extracción y cuantificación de DNA libre circulante.....	51
8.4.4.2	Preparación de librerías para secuenciación y análisis bioinformático.....	52
8.4.5	Comparación de regiones genómicas alteradas a nivel citogenético entre ambos tipos de biopsias.....	52
9	RESULTADOS Y DISCUSIÓN.....	53
9.1	RENDIMIENTO Y ESTADÍSTICAS DE LA SECUENCIACIÓN GENÓMICA CON NANOPOROS DE FRAGMENTOS CORTOS EN LÍNEAS CELULARES TUMORALES.....	53
9.2	DETECCIÓN DE ALTERACIONES EN EL NÚMERO DE COPIAS DE DNA EN LÍNEAS CELULARES TUMORALES MEDIANTE SECUENCIACIÓN GENÓMICA CON NANOPOROS.....	55
9.3	VALIDACIÓN DE LOS RESULTADOS OBTENIDOS MEDIANTE SECUENCIACIÓN CON NANOPOROS CON LA TECNOLOGÍA DE SECUENCIACIÓN ILLUMINA.....	57
9.3.1	Análisis de correlación de datos de segmentación entre tecnologías de secuenciación.....	57
9.3.2	Comparación de regiones genómicas alteradas a nivel citogenético entre tecnologías de secuenciación.....	60
9.3.2.1	Línea celular MCF7.....	62
9.3.2.2	Línea celular MDA-MB-361.....	65
9.3.2.3	Línea celular SK-BR-3.....	69
9.3.2.4	Línea celular MDA-MD-231.....	73
9.4	DETECCIÓN DE ALTERACIONES EN EL NÚMERO DE COPIAS DE DNA EN MUESTRAS DE PACIENTES CON CÁNCER DE MAMA.....	78

9.4.1	Características clínico-patológicas de la cohorte de pacientes con cáncer de mama incluidas en el estudio	78
9.4.2	Detección de alteraciones en el número de copias de DNA en biopsias sólidas	79
9.4.2.1	Análisis de detección de alteraciones en el número de copias de DNA en la biopsia de tejido tumoral BL-147	80
9.4.2.2	Análisis de detección de alteraciones en el número de copias de DNA en la biopsia de tejido tumoral BL-158	81
9.4.3	Detección de alteraciones en el número de copias de DNA en biopsias líquidas	82
9.4.3.1	Aislamiento de DNA libre circulante a partir de muestras de plasma de pacientes con cáncer de mama.....	83
9.4.3.2	Secuenciación genómica con nanoporos de DNA libre circulante.....	85
9.4.3.3	Análisis de detección de alteraciones en el número de copias de DNA en las biopsias líquidas BL-147 y BL-158.....	86
10	CONCLUSIONES	91
11	PERSPECTIVAS	92
12	REFERENCIAS	93
13	APÉNDICES	104
13.1	LÍNEAS CELULARES.....	104
13.1.1	Reportes de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos de líneas celulares tumorales empleando el software NanoGLADIATOR.....	104
13.1.2	Reportes de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina de líneas celulares tumorales empleando el software XCAVATOR	114
13.2	MUESTRAS DE PACIENTES CON CÁNCER DE MAMA.....	125
13.2.1	Reportes de alteraciones en el número de copias de DNA detectadas mediante microarreglos de alta resolución en muestras de biopsia de tejido tumoral empleando el software Chromosome Analysis Suite	125

13.2.2	Reportes de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en muestras de biopsia líquida empleando el software NanoGLADIATOR.....	128
--------	---	-----

TABLAS

TABLA 1.	Selección de líneas celulares tumorales y sus características de origen	40
TABLA 2.	Claves de acceso de los datos de secuenciación de las líneas celulares tumorales.	45
TABLA 3.	Rendimiento y estadísticas de la secuenciación genómica con nanoporos de líneas celulares tumorales realizada en la plataforma de secuenciación GridION X5.....	54
TABLA 4.	Resumen de alteraciones en el número de copias de DNA detectadas mediante las tecnologías de secuenciación genómica con nanoporos y secuenciación Illumina en líneas celulares tumorales	61
TABLA 5.	Regiones genómicas alteradas en el número de copias que son concordantes entre ambas tecnologías de secuenciación (nanoporos e Illumina) en la línea celular MCF7	63
TABLA 6.	Regiones genómicas alteradas en el número de copias que son concordantes entre ambas tecnologías de secuenciación (nanoporos e Illumina) en la línea celular MDA-MB-361	66
TABLA 7.	Alteraciones en el número de copias de DNA detectadas mediante ambas tecnologías de secuenciación (nanoporos e Illumina) en la línea celular SK-BR-3	71
TABLA 8.	Alteraciones en el número de copias de DNA detectadas mediante ambas tecnologías de secuenciación (nanoporos e Illumina) en la línea celular MDA-MB-231.....	76
TABLA 9.	Características clínico-patológicas de la cohorte de pacientes con cáncer de mama.....	78
TABLA 10.	Características clínico-patológicas de las pacientes con cáncer de mama seleccionadas para evaluación en el estudio.	80
TABLA 11.	Rendimiento y estadísticas de la secuenciación genómica con nanoporos llevada a cabo en muestras de biopsia líquida utilizando la plataforma de secuenciación GridION X5	86
TABLA SUPLEMENTARIA 1.	Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular HeLa.....	104
TABLA SUPLEMENTARIA 2.	Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular MCF7.....	105
TABLA SUPLEMENTARIA 3.	Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular MDA-MB-361.	107
TABLA SUPLEMENTARIA 4.	Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular SK-BR-3	109

TABLA SUPLEMENTARIA 5. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular MDA-MB-231.	112
TABLA SUPLEMENTARIA 6. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular Hs578T.....	113
TABLA SUPLEMENTARIA 7. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina en la línea celular MCF7.....	114
TABLA SUPLEMENTARIA 8. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina en la línea celular MDA-MB-361.	116
TABLA SUPLEMENTARIA 9. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina en la línea celular SK-BR-3	118
TABLA SUPLEMENTARIA 10. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina en la línea celular MDA-MB-231.	123
TABLA SUPLEMENTARIA 11. Reporte de alteraciones en el número de copias de DNA detectadas mediante microarreglos de alta resolución en la muestra de biopsia de tejido tumoral BL-147.	125
TABLA SUPLEMENTARIA 12. Reporte de alteraciones en el número de copias de DNA detectadas mediante microarreglos de alta resolución en la muestra de biopsia de tejido tumoral BL-158.	126
TABLA SUPLEMENTARIA 13. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la muestra de biopsia líquida BL-147	128
TABLA SUPLEMENTARIA 14. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la muestra de biopsia líquida BL-158.....	129

FIGURAS

FIGURA 1. Tipos de inestabilidad genómica y mecanismos asociados.....	7
FIGURA 2. Clasificación de las alteraciones en el número de copias de DNA con base en su extensión genómica	10
FIGURA 3. Los tumores se subdividen en dos clases principales con base en las alteraciones genómicas que predominan en ellos	12
FIGURA 4. Clasificación de métodos para la detección de alteraciones en el número de copias de DNA	14
FIGURA 5. Fundamento de la secuenciación con nanoporos	24
FIGURA 6. La profundidad de lectura como método para la detección de alteraciones en el número de copias de DNA	27

FIGURA 7. Las biopsias líquidas son una herramienta clínica que permite obtiene información molecular clínicamente relevante sobre los tumores, de forma no invasiva	29
FIGURA 8. Estrategia experimental del proyecto	39
FIGURA 9. El análisis de fragmentos cortos de DNA mediante la tecnología de secuenciación genómica con nanoporos es útil para la detección de alteraciones en el número de copias de DNA	56
FIGURA 10. Los resultados de detección de alteraciones en el número de copias de DNA mediante las tecnologías de secuenciación con nanoporos y de secuenciación Illumina están altamente correlacionados.....	59
FIGURA 11. Perfiles de segmentación de los cromosomas 11 y 14 obtenido con secuenciación con nanoporos e Illumina en la línea celular MCF7.....	65
FIGURA 12. Perfiles de segmentación de los cromosomas 8 y 17 obtenido con secuenciación con nanoporos e Illumina en la línea celular MDA-MB-361	68
FIGURA 13. Perfiles de segmentación de los cromosomas 8 y 17 obtenido con secuenciación con nanoporos e Illumina en la línea celular SK-BR-3.....	73
FIGURA 14. Perfiles de segmentación de los cromosomas 6 y 13 obtenido con secuenciación con nanoporos e Illumina en la línea celular MDA-MB-231	76
FIGURA 15. Análisis del tamaño de los fragmentos de DNA aislados en muestras de biopsia líquida de pacientes con cáncer de mama.....	84
FIGURA 16. Análisis comparativo de alteraciones en el número de copias de DNA presentes en biopsias de tejido tumoral y biopsias líquidas en pacientes con cáncer de mama.....	87

ABREVIATURAS Y SIGLAS

BL	Biopsia líquida
CaMa	Cáncer de mama
cfDNA	DNA libre circulante (del inglés <i>cell-free DNA</i>)
CGH	Hibridación genómica comparativa (del inglés <i>Comparative Genomic Hybridization</i>)
ChAS	Chromosome Analysis Suite
chr	Cromosoma
CIN	Inestabilidad cromosómica (del inglés <i>Chromosomal Instability</i>)
CN	Número de copias (del inglés <i>Copy number</i>)
CNA	Alteración en el número de copias (del inglés <i>Copy Number Alteration</i>)
CNV	Variación en el número de copias (del inglés <i>Copy Number Variation</i>)
ctDNA	DNA tumoral circulante (del inglés <i>circulating-tumor DNA</i>)
DNA	Ácido desoxirribonucleico (del inglés <i>Desoxiribonucleic Acid</i>)
dPCR	PCR digital (del inglés <i>digital PCR</i>)
ER	Receptor de Estrógenos (del inglés <i>Estrogen Receptor</i>)
FFPE	Fijado con formaldehído y embebido en parafina (del inglés <i>Formalin-Fixed Paraffin-Embedded</i>)
FISH	Hibridación <i>in situ</i> fluorescente (del inglés <i>Fluorescence In Situ Hybridization</i>)
GC	Guanina y citosina
gDNA	DNA genómico (del inglés <i>genomic DNA</i>)
HER2	Receptor 2 del Factor de Crecimiento Epidérmico Humano (del inglés <i>Human Epidermal Growth Factor Receptor 2</i>)
IHQ	Inmunohistoquímica
LGC	Laboratorio de Genómica del Cáncer (INMEGEN)
Mb	Megabase
MIP	Sonda de inversión molecular (del inglés <i>Molecular Inversion Probe</i>)
MLPA	Amplificación múltiple de sondas dependiente de ligación (del inglés <i>Multiplex Ligation-dependent Probe Amplification</i>)
MSI	Inestabilidad microsatelital (del inglés <i>Microsatellite Instability</i>)
NGS	Secuenciación de nueva generación (del inglés <i>Next Generation Sequencing</i>)
NIN	Inestabilidad a nivel de nucleótidos (del inglés <i>Nucleotide Instability</i>)
p	Braza pequeño del cromosoma
pb	Pares de bases
PCR	Reacción en Cadena de la Polimerasa (del inglés <i>Polymerase Chain Reaction</i>)

PR	Receptor de Progesterona (del inglés <i>Progesterone Receptor</i>)
q	Brazo largo del cromosoma
qPCR	Reacción en Cadena de la Polimerasa Cuantitativa (del inglés <i>Quantitative Polymerase Chain Reaction</i>)
SLM	Algoritmo <i>Shifting Level Model</i>
SNP	Polimorfismo de un solo nucleótido (del inglés <i>Single Nucleotide Polymorphism</i>)
sWGS	Secuenciación a nivel de genoma completo de baja profundidad (del inglés <i>shallow Whole Genome Sequencing</i>)
WES	Secuenciación de exoma completo (del inglés <i>Whole Exome Sequencing</i>)
WGS	Secuenciación a nivel de genoma completo (del inglés <i>Whole Genome Sequencing</i>)

RESUMEN

Las alteraciones cromosómicas constituyen una característica distintiva en diversos tipos tumorales, incluido el cáncer de mama. Estas alteraciones pueden manifestarse a través de alteraciones en el número de copias de DNA, las cuales desempeñan un papel crucial en la progresión y el desarrollo de la enfermedad. Las tecnologías de secuenciación de nueva generación se han consolidado como herramientas fundamentales para la identificación precisa de estas alteraciones cromosómicas. A pesar de sus beneficios en el diagnóstico clínico, la implementación de estas tecnologías aún enfrenta diversos desafíos. La secuenciación genómica con nanoporos ha emergido como una alternativa prometedora a las técnicas de secuenciación tradicionales que la hacen potencialmente aplicable en un entorno clínico. Este trabajo se centró en explorar y evaluar el uso de la tecnología de secuenciación genómica con nanoporos para la detección de alteraciones en el número de copias de DNA, con especial énfasis en su implementación en el análisis de muestras de biopsias líquidas. Estas muestras no invasivas, han adquirido un interés creciente en el ámbito clínico puesto que se ha demostrado que proporcionan información valiosa sobre las características tumorales a partir del análisis de componentes circulantes derivados del tejido tumoral, como el DNA tumoral circulante. La investigación se dividió en dos secciones fundamentales: la primera etapa se centró en estandarizar un flujo de trabajo utilizando secuenciación con nanoporos de fragmentos cortos de DNA de líneas celulares tumorales de mama. Esto se realizó con el fin de evaluar la concordancia en la detección de las alteraciones en comparación con los resultados obtenidos mediante la tecnología Illumina, una plataforma líder de las tecnologías de secuenciación de nueva generación. La segunda etapa se enfocó en aplicar esta metodología al análisis de muestras de biopsias líquidas obtenidas de pacientes con cáncer de mama. Esto se llevó a cabo con el propósito de evaluar la concordancia en la detección de alteraciones en el número de copias de DNA con las biopsias sólidas correspondientes a cada paciente, analizadas mediante microarreglos de alta resolución. Los resultados obtenidos revelaron que, aunque la tecnología de secuenciación con nanoporos mostró una alta concordancia con la tecnología Illumina en el análisis de líneas celulares tumorales, su sensibilidad disminuyó al aplicarla al análisis de muestras de biopsias líquidas. Este descenso en la sensibilidad afectó la capacidad de detección de alteraciones en el número de copias de DNA, lo que se atribuyó a la baja cantidad de DNA tumoral circulante presente en las muestras de biopsias líquidas, posiblemente asociado a los estadios clínicos tempranos de la enfermedad en las pacientes evaluadas. Este estudio sentó las bases para comprender los desafíos asociados con la detección de alteraciones en el número de copias de DNA en biopsias líquidas utilizando secuenciación con nanoporos de baja profundidad, particularmente en las

etapas tempranas del cáncer de mama. Destaca la necesidad de desarrollar estrategias mejoradas para aumentar la sensibilidad en estas muestras o, alternativamente, de aplicar este enfoque en el análisis de biopsias líquidas provenientes de pacientes con estadios avanzados de la enfermedad, lo que podría impulsar su utilidad como potencial herramienta clínica.

ABSTRACT

Chromosome abnormalities are a distinctive feature in diverse cancer types, including breast cancer. These abnormalities can be manifested through copy number alterations, which significantly contribute to the progression and development of the disease. Next-generation sequencing methodologies have established their pivotal role in precisely detecting these chromosomal abnormalities. Despite their benefits in clinical diagnosis, the implementation of these technologies still faces several challenges. Nanopore sequencing has emerged as a promising alternative to traditional sequencing technologies, potentially applicable in a clinical setting. This study evaluated the efficacy of nanopore sequencing in detecting copy number alterations, with a particular focus on its application in analyzing liquid biopsy samples. These non-invasive samples have gained significant attention within the clinical field due to their ability to provide valuable insights into tumor characteristics. This is achieved through the analysis of circulating components derived from tumor tissue, notably circulating tumor DNA. The research comprised two phases. The first phase aimed to standardize a workflow using nanopore sequencing of short DNA fragments from breast cancer cell lines, to assess the concordance in detecting chromosome alterations compared to results obtained using Illumina technology, a leading next-generation sequencing platform. The second phase focused on applying this methodology to analyze liquid biopsy samples obtained from breast cancer patients, aiming to evaluate the concordance in detecting copy number alterations with corresponding solid biopsies for each patient, analyzed using high-resolution microarray technology. Findings revealed that although nanopore sequencing technology showed high concordance with Illumina technology in the analysis of cancer cell lines, its sensitivity decreased when applied to liquid biopsy samples. This decrease in sensitivity was attributed to the low amount of circulating tumor DNA present in the samples, possibly associated with the early clinical stages of the disease in the evaluated patients. This work provided insights into the complexities involved in identifying copy number alterations within liquid biopsies using shallow whole genome nanopore sequencing, particularly within the early stages of breast cancer. It highlights the need to develop enhanced strategies to increase sensitivity in these samples or, alternatively, to apply this approach in analyzing liquid biopsies from patients with advanced disease stages, potentially boosting its utility as a clinical tool.

1 MARCO TEÓRICO

1.1 Inestabilidad genómica

Un sello distintivo de casi todos los tipos tumorales humanos está representado por aberraciones en su arquitectura genómica (Negrini et al., 2010). La inestabilidad genómica se refiere al conjunto de eventos capaces de causar alteraciones de carácter temporal o permanente dentro del genoma (Jefford & Irminger-Finger, 2006). Estos cambios en la arquitectura genómica comprenden mutaciones puntuales, inserciones y deleciones cortas, expansión y contracción de microsatélites, así como alteraciones cromosómicas estructurales y numéricas, dentro de las cuales se incluyen: reordenamientos cromosómicos (translocaciones, inversiones, amplificaciones y deleciones), variaciones en el número de copias y pérdida de la heterocigosidad (Lee et al., 2016; Negrini et al., 2010).

La integridad del genoma es constantemente afectada por factores exógenos y endógenos que inducen lesiones en el DNA y generan daño a nivel cromosómico (Gerlach & Herranz, 2020). Dentro de los factores exógenos se incluyen los productos químicos como el humo del tabaco y el alcohol, las sustancias químicas naturales como las aflatoxinas producidas por hongos o diversas formas de radiación, por ejemplo, la luz ultravioleta (UV) y la radiación ionizante (Katerji & Duerksen-Hughes, 2021). Los factores endógenos involucrados en el daño al DNA son las especies reactivas de oxígeno (ROS), los aldehídos, los errores ocurridos en el proceso mitótico o errores debido a enzimas involucradas en los procesos de reparación del DNA y en la edición del genoma (Chatterjee & Walker, 2017). Así mismo, una célula puede adquirir secuencias de DNA a partir de fuentes exógenas, en particular de virus. Estos virus pueden contribuir a la carcinogénesis de uno o más tipos tumorales (Katerji & Duerksen-Hughes, 2021). Ejemplos de estos virus son el virus de la hepatitis B (cáncer de hígado), el virus del papiloma humano (cáncer cervical y cáncer anal), el virus de Epstein-Barr (linfoma de Burkitt y cáncer de nasofaringe) y el herpesvirus humano tipo 8 (sarcoma de Kaposi) (White et al., 2014).

Durante la división celular, se verifica la integridad del genoma mediante cuatro puntos de control principales: el punto de control G1/S, en el cual se identifican daños en el DNA previo a su replicación; el punto de control S/G2, dentro del que se verifica si el proceso de replicación se llevó a cabo correctamente; el punto de control G2/M, en el cual se monitorea la integridad del DNA previo a la mitosis; y el punto de control del ensamblaje del huso (SAC, *Spindle Assembly Checkpoint*), en el que se regula la transición metafase-anafase para garantizar la segregación

correcta de las cromátidas hermanas (Cunningham et al., 2019). Estos puntos de control bloquean la progresión del ciclo celular si la integridad del genoma se encuentra comprometida, lo que garantiza la euploidía. El mal funcionamiento de ellos conduce a la acumulación de rupturas de DNA de doble cadena (DSBs), errores en la segregación cromosómica y, finalmente, inestabilidad genómica (Bartek & Lukas, 2007).

1.1.1 Tipos de inestabilidad genómica

Las alteraciones en el DNA que ocurren como resultado de la inestabilidad genómica pueden abarcar desde un nucleótido hasta cromosomas completos (Lee et al., 2016). En función del tipo de alteraciones genómicas observadas, se han identificado tres clases principales de inestabilidad genómica (Figura 1) (Cunningham et al., 2019).

La mayoría de los tipos tumorales presentan una forma denominada inestabilidad cromosómica (CIN), en la que existe una alta tasa de cambios en la estructura y el número de cromosomas de las células tumorales a lo largo del tiempo con referencia a las células normales (Negrini et al., 2010). Generalmente, los tumores de origen epitelial como cáncer de pulmón, cáncer de mama, cáncer de ovario y melanoma suelen mostrar un mayor grado de inestabilidad cromosómica respecto a las neoplasias hematológicas, por ejemplo, leucemia linfoblástica aguda y leucemia mieloide aguda (Pikor et al., 2013). Se han identificado numerosos genes que contribuyen a la CIN, incluidos aquellos involucrados en la condensación y segregación cromosómica (Solomon et al., 2011), la cohesión de las cromátidas hermanas (Jallepalli et al., 2001), la función de los telómeros (O'Sullivan & Karlseder, 2010), el daño al DNA (Levitt & Hickson, 2002) y el punto de control del ensamblaje del huso (Manic et al., 2017).

Las alteraciones cromosómicas se clasifican en dos categorías principales: numéricas y estructurales (Kou et al., 2020). Las alteraciones cromosómicas numéricas, principalmente las aneuploidías y la CIN, se caracterizan por la ganancia o pérdida de cromosomas (van Jaarsveld & Kops, 2016). Se estima que aproximadamente el 90% de los tumores humanos presentan este tipo de alteraciones (Knouse et al., 2017). Por otra parte, las alteraciones cromosómicas estructurales engloban eventos de diversa complejidad, desde deleciones o amplificaciones a nivel del brazo cromosómico, hasta alteraciones de múltiples cromosomas (Kou et al., 2020). Entre estas, las amplificaciones y las deleciones cromosómicas son los eventos más comunes, encontrándose presentes en el 88% de los tumores humanos (Li et al., 2020).

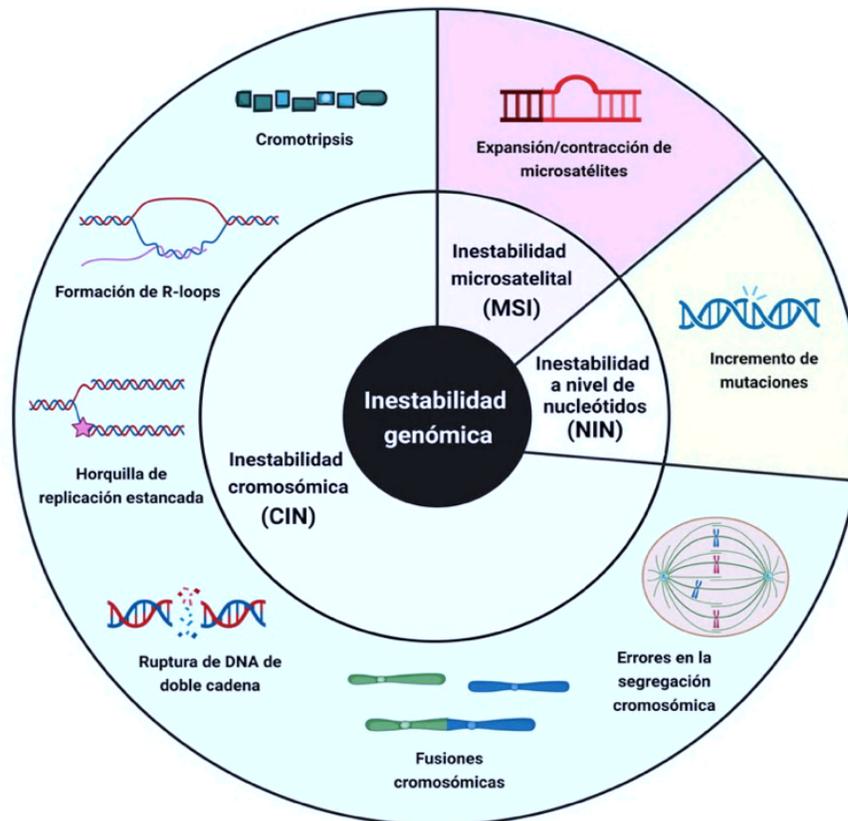


Figura 1. Tipos de inestabilidad genómica y mecanismos asociados. La inestabilidad genómica se agrupa en tres tipos distintos. 1) La inestabilidad a nivel de nucleótidos (NIN) generalmente se desarrolla por errores durante el proceso de replicación y en algunos mecanismos de reparación de daño al DNA, como el sistema de reparación por escisión de bases (BER, *Base Excision Repair*) y el sistema de reparación por escisión de nucleótidos (NER, *Nucleotide Excision Repair*), lo que conduce a cambios dentro de la secuencia en uno o pocos nucleótidos (sustituciones, deleciones e inserciones). 2) La inestabilidad microsatelital (MSI) surge por alteraciones en los genes que codifican proteínas del sistema de reparación de nucleótidos mal apareados (MMR, *Mismatch Repair*). Esto ocasiona cambios en la longitud de secuencias de DNA repetitivas denominadas microsatélites, las cuáles pueden contraerse o expandirse aleatoriamente y generar un fenotipo hipermutador. 3) La inestabilidad cromosómica (CIN) se distingue por el aumento en la tasa de ganancia o pérdida de segmentos cromosómicos o cromosomas completos y es la forma principal de inestabilidad presente en tumores sólidos. Los defectos en proteínas que participan en la reparación de rupturas de DNA de doble cadena (DSBs, *Double-Strand Breaks*), la función del cinetocoro y la segregación cromosómica contribuyen a la CIN. Durante el estrés replicativo prolongado, las horquillas de replicación estancadas pueden resultar en DSBs irreversibles que conducen a la inestabilidad del genoma. También, la ruptura de cromosomas puede generar ciclos de "ruptura-fusión-puente", en los que los extremos rotos de diferentes cromosomas se fusionan produciendo un cromosoma dicéntrico, el cual posteriormente puede romperse durante la mitosis. La reparación de este daño puede conducir a ciclos adicionales de ruptura-fusión, generando duplicaciones, deleciones o la muerte celular. Otro mecanismo asociado con esta clase de inestabilidad es la formación de R-loops, descritos como híbridos de DNA-RNA que se forman cuando el DNA de cadena sencilla se encuentra expuesto durante los procesos de replicación y reparación. Trabajos recientes han descrito un fenómeno llamado cromotripsis, en el que uno o varios cromosomas se fragmentan en múltiples pedazos y estos son reparados en un orden y orientación antinatural. Creada con BioRender.com

Si bien la CIN es la principal clase de inestabilidad genómica observada en los tumores humanos, también se han descrito otras formas (Negrini et al., 2010). Una de ellas, la inestabilidad microsatelital (MSI; también conocida como MIN), se caracteriza por la expansión o contracción del número de repeticiones de oligonucleótidos presentes en las secuencias de los microsatélites. Esta clase de inestabilidad es característica en varios tipos tumorales como cáncer gástrico, cáncer de endometrio, cáncer de ovario, cáncer de pulmón y cáncer colorrectal (Pikor et al., 2013). Así mismo, otra forma reconocida es la inestabilidad a nivel de nucleótidos (NIN), la cual se manifiesta como una tasa elevada de alteraciones de uno o pocos nucleótidos y surge debido a defectos en vías de reparación del DNA (Cunningham et al., 2019).

Todos los niveles de inestabilidad genómica contribuyen de forma significativa a la patogénesis del cáncer, a los paisajes genómicos y epigenómicos aberrantes de las células tumorales y pueden ocurrir simultáneamente dentro de la misma célula para afectar la función de un gen, un complejo proteico o toda una vía de señalización (Hanahan & Weinberg, 2011; Pikor et al., 2013). Sin embargo, la CIN tiene un mayor impacto en la estabilidad del genoma ya que induce la alteración del número de copias de grandes grupos de genes que promueven la carcinogénesis (Michor et al., 2005). Las alteraciones cromosómicas estructurales o numéricas pueden generar poblaciones de células genotípicamente distintas, lo que incrementa el fenómeno de heterogeneidad intratumoral (Gagos & Irminger-Finger, 2005). Además, la CIN se ha observado en muchos de los tipos tumorales con mayor mortalidad, entre los que se encuentran el cáncer colorrectal, el **cáncer de mama** y el cáncer de pulmón, de modo que se asocia con un mal pronóstico y resistencia terapéutica (Raynes & Weinreich, 2018).

1.2 Alteraciones en el número de copias de DNA

Las alteraciones en el número de copias (CNA, *Copy Number Alteration*) de DNA son cambios somáticos que afectan la estructura cromosómica, conduciendo a la ganancia o pérdida de copias de segmentos de DNA (Krijgsman et al., 2014). Estas secuencias de DNA afectadas por las CNAs exhiben un número de copias diferente entre el DNA de línea germinal de un individuo y el DNA de una subpoblación clonal de células tumorales (Beroukhim et al., 2010). Dichos cambios impactan en la activación de oncogenes y en la inactivación de genes supresores de tumores, contribuyendo así a la carcinogénesis y la progresión del cáncer (Zack et al., 2013).

Cabe aclarar que el término CNA no es sinónimo del de variación en el número de copias (CNV, *Copy Number Variation*). Las CNVs son aquellas secuencias de DNA que se encuentran en un número de copias diferente entre el DNA de línea germinal de dos individuos diferentes, es decir, son variantes heredadas desde un progenitor (Fitzgerald & Birney, 2022).

En la mayoría de los casos, las células humanas contienen 2 copias de cada uno de sus genes. En función de esto, se distinguen dos tipos principales de CNAs (Krijgsman et al., 2014):

- **Ganancia o amplificación.** Es la presencia de múltiples copias de un segmento cromosómico, de forma que el número de copias totales de DNA correspondiente a esa región del genoma es mayor que 2. La "ganancia" implica un aumento de una o varias copias de un segmento de DNA, generalmente abarcando una región genómica amplia. Por el contrario, la "amplificación" se refiere a eventos más pequeños que pueden alcanzar un número de copias mucho más alto.
- **Pérdida o delección.** Es la disminución en el número de copias de un segmento cromosómico, lo que reduce el número total de copias de DNA en esa región a 1 o incluso 0, dependiendo de si se ha perdido una o ambas copias. Similar a la descripción anterior, la "pérdida" suele afectar una región genómica extensa, mientras que la "delección" se refiere a eventos de menor tamaño.

1.2.1 Clasificación de las alteraciones en el número de copias de DNA

El incremento o la disminución patológica de una parte del genoma puede abarcar desde unos cientos o miles de pares de bases (pb) hasta cromosomas completos (Beroukhim et al., 2010). De acuerdo con la extensión de la región alterada, las CNAs se clasifican en dos subtipos diferentes (Harbers et al., 2021). La ganancia o pérdida en el número de copias se asocia con CNAs a gran escala, también conocidas como "a nivel de brazo", en las que se incluyen las alteraciones que abarcan todo el brazo cromosómico o aberraciones en cromosomas completos (aneuploidías). Por el contrario, dentro de las CNAs focales se agrupan las amplificaciones y las delecciones, las cuáles poseen una extensión más limitada (Figura 2) (Krijgsman et al., 2014).

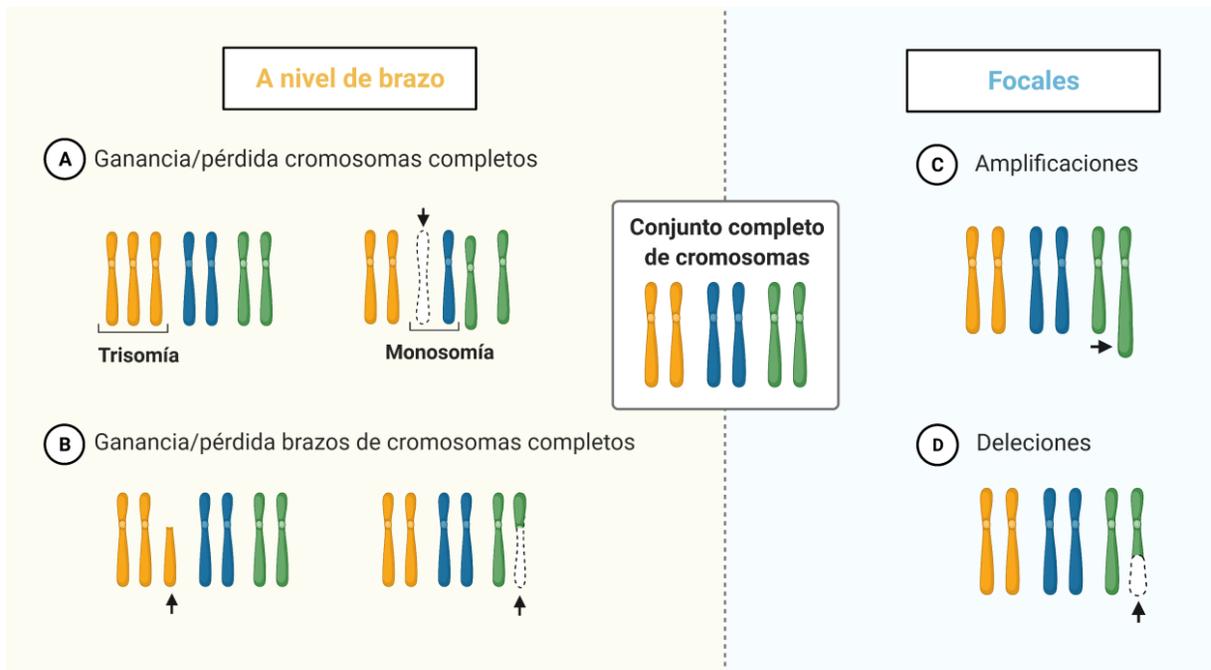


Figura 2. Clasificación de las alteraciones en el número de copias de DNA con base en su extensión genómica. Las alteraciones en el número de copias (CNAs) son alteraciones somáticas comunes en muchos tipos tumorales y el tamaño de las regiones genómicas afectadas por ellas varía ampliamente. De acuerdo con su extensión, se pueden distinguir dos subtipos principales de CNAs: las alteraciones de mayor extensión se denominan CNAs a nivel de brazo, en las que se incluyen la ganancia o la pérdida de cromosomas completos (A) y de brazos de cromosomas completos (B). Por otra parte, las CNAs focales son alteraciones que abarcan únicamente un segmento cromosómico; dentro de estas alteraciones se ubican las amplificaciones (C) y las delecciones (D), las cuáles van desde una región de un brazo cromosómico hasta unas pocas kb. Creada con BioRender.com

Si bien al momento no existe un consenso estricto para definir la extensión de las CNAs focales, en la literatura pueden encontrarse algunas definiciones dadas por diferentes autores. Una de ellas indica que las CNAs focales son alteraciones somáticas que generalmente afectan regiones de un tamaño no mayor a 3 Mb, dentro de las cuales se encuentran albergados un pequeño número de genes (Bierkens et al., 2013). En otro trabajo, las CNAs focales se definen como aberraciones con una longitud menor al 98% del brazo de un cromosoma (Beroukhim et al., 2010; Mermel et al., 2011). Un abordaje alternativo, desde el punto de vista bioinformático, señala que las CNAs focales se reconocen como picos estadísticamente identificados por un algoritmo particular, por ejemplo GISTIC ó CGHcall. Estos picos representan regiones en el genoma que están significativamente amplificadas o deletadas en un conjunto de muestras biológicas (Mermel et al., 2011).

1.2.2 Panorama genómico de las alteraciones en el número de copias de DNA en tumores humanos

Durante las últimas dos décadas, los avances en tecnologías de alto rendimiento han abierto la puerta a una exploración más profunda de las bases genómicas y epigenómicas del cáncer (Vogelstein et al., 2013). Los estudios genómicos realizados en diferentes tipos tumorales han favorecido la comprensión de los genomas del cáncer y su complejidad (Garraway & Lander, 2013).

El Atlas del Genoma del Cáncer (TCGA, *The Cancer Genome Atlas*) es un proyecto iniciado en 2006, creado con el objetivo de coleccionar y perfilar más de 10,000 muestras de tumores en al menos 20 tipos tumorales (Ciriello et al., 2013). Hasta la fecha, el consorcio ha llevado a cabo la caracterización molecular de más de 20,000 tumores primarios y sus respectivas muestras normales pareadas en 33 tipos de cáncer (más información disponible en <https://www.cancer.gov/tcga>). Además, el Consorcio Internacional del Genoma del Cáncer (ICGC, *International Cancer Genome Consortium*), del cual TCGA es miembro, ha contribuido sustancialmente al agregar miles de muestras adicionales y abordar otros tipos tumorales (Hudson et al., 2010).

El amplio conjunto de muestras analizadas por estos consorcios ha generado una gran cantidad de datos que está enriqueciendo constantemente nuestro conocimiento sobre las alteraciones moleculares presentes en los tumores humanos (Ciriello et al., 2013). Las CNAs han sido perfiladas en miles de estas muestras mediante tecnologías como los microarreglos de alta resolución y la secuenciación de nueva generación (NGS, *Next Generation Sequencing*). Esto ha permitido un estudio detallado de estas alteraciones, ampliando nuestro conocimiento sobre su prevalencia, tipología y distribución en diferentes tipos de tumores, así como la identificación de los genes asociados al cáncer que se ven más afectados por ellas (Harbers et al., 2021). De acuerdo con el trabajo de Ciriello et al. (2013), en algunos tipos tumorales, las CNAs desempeñan un papel más relevante que otra clase de alteraciones (Figura 3).

La carga de CNAs (del término en inglés *CNA burden*) de un tumor se define como el porcentaje del genoma tumoral que es afectado por CNAs (Zhang et al., 2018). Los tipos tumorales con mayor carga de CNAs son cáncer de ovario, sarcomas, carcinosarcoma uterino, cáncer de esófago, cáncer de vejiga y cáncer de mama (Harbers et al., 2021; Steele et al., 2022). El número de CNAs observadas, así como el porcentaje del genoma alterado por CNAs pueden

presentar gran variabilidad de un tumor a otro dentro del mismo tipo tumoral. Por ejemplo, en cáncer de ovario, sarcomas y cáncer de mama existe alta heterogeneidad en el número eventos de CNAs presentes entre muestras del mismo tipo tumoral. En la mayoría de los tipos tumorales, se ha observado que las amplificaciones son significativamente más abundantes y afectan a una mayor fracción del genoma, en comparación con las deleciones (Harbers et al., 2021).

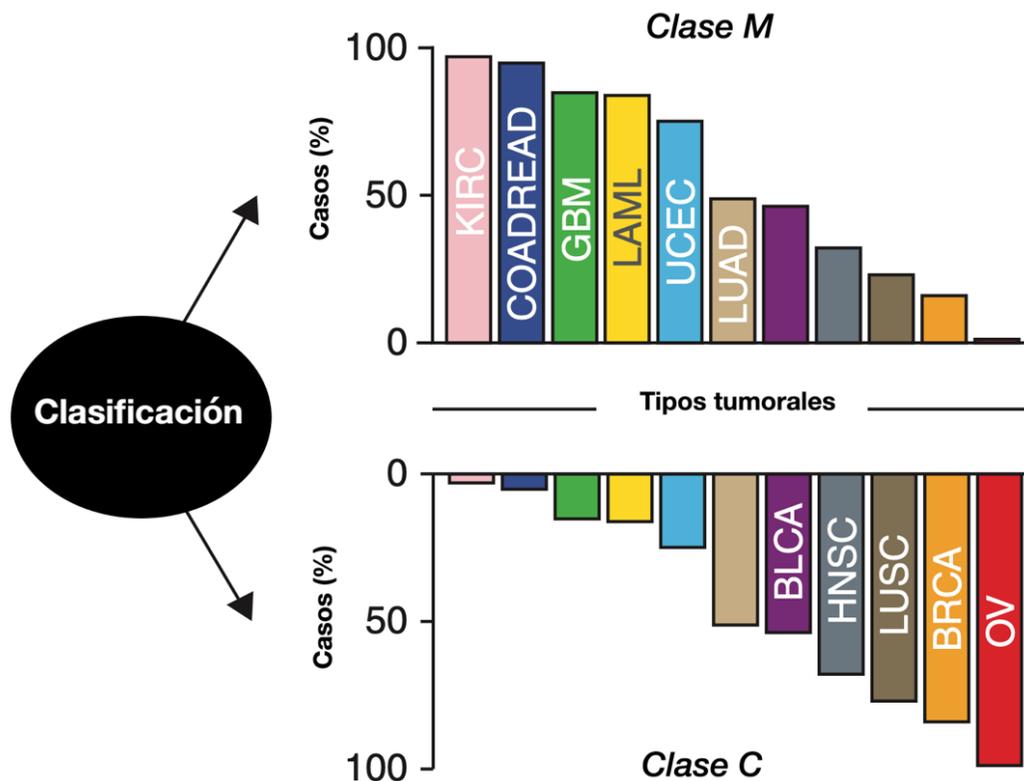


Figura 3. Los tumores se subdividen en dos clases principales con base en las alteraciones genómicas que predominan en ellos. Existen dos clases principales de tumores caracterizados por patrones distintos de alteraciones genómicas y epigenéticas: la clase M, que presenta una mayor prevalencia de mutaciones somáticas y la clase C, caracterizado por tener predominantemente alteraciones en el número de copias de DNA. En la clase M de tumores se incluye el cáncer de riñón, cáncer colorrectal, glioblastoma multiforme, leucemia mieloide aguda y carcinoma uterino. Por su parte, la clase C de tumores está caracterizada principalmente por la presencia de múltiples ganancias y pérdidas de regiones cromosómicas. En esta clase se incluye el cáncer de ovario, cáncer de mama, carcinomas de células escamosas de pulmón y cabeza/cuello y cáncer de vejiga. KIRC: Kidney renal cell carcinoma, COADREAD: Colon and rectum adenocarcinoma, GBM: Glioblastoma multiforme, LAML: Acute myeloid leukemia, UCEC: Uterine corpus endometrial carcinoma, LUAD: Lung adenocarcinoma, BLCA: Bladder cancer, HNSC: Head and neck squamous cell carcinoma, LUSC: Lung squamous cell carcinoma, BRCA: Breast cancer, OV: Ovarian cancer. Modificada de Ciriello et al. (2013).

1.2.2.1 Cromosomas afectados con mayor frecuencia por alteraciones en el número de copias de DNA

Se ha observado que existe una tendencia de las CNAs a ocurrir dentro de cromosomas específicos. En muchos tipos tumorales, ciertos cromosomas, como el 7 y el 20, se ganan con frecuencia; por el contrario, otros cromosomas, como el 13 y el 22, se suelen perder frecuentemente. Otro subconjunto de cromosomas, en el que se incluyen los cromosomas 1, 2 y 3, raramente se observan alterados por aneuploidías (Knouse et al., 2017). En el caso de las CNAs de tipo focal, en muchos tipos tumorales se ha reportado que una proporción relativamente grande de eventos ocurren dentro del cromosoma 8, mientras que otros cromosomas, como el 15, no se suelen observar afectados (Harbers et al., 2021).

También, se ha reportado que algunos tipos tumorales se caracterizan por patrones de CNAs en ciertos cromosomas. Por ejemplo, la mayoría de las CNAs focales detectadas en cáncer colorrectal se ubican en los cromosomas 13 y 20, en tanto que en tipos tumorales como cáncer de mama y colangiocarcinomas estas alteraciones afectan predominantemente al cromosoma 17 (Harbers et al., 2021). En particular, el cromosoma 8 alberga al gen *MYC*, el cual es uno de los genes más frecuentemente amplificados en los tumores humanos (Schaub et al., 2018). Por otra parte, en el cromosoma 17 se localiza el gen *ERBB2*, que se encuentra amplificado en aproximadamente el 20% de los casos de cáncer de mama invasivo (Arteaga et al., 2011). El comportamiento observado sobre que ciertos cromosomas o regiones cromosómicas tienden a ser más susceptibles a las CNAs en los tumores humanos podría ser dependiente de la extensión del daño al DNA y el tipo o la eficiencia de las vías de reparación del daño al DNA presentes a lo largo de cada uno de los cromosomas individuales (Harbers et al., 2021).

1.3 Métodos de detección de alteraciones en el número de copias de DNA

La frecuencia de las CNAs en casi todos los tipos tumorales indica que son eventos genómicos importantes involucrados en el inicio, desarrollo y progresión del cáncer (Jang et al., 2016). La amplificación de una región genómica que codifica un oncogén o la delección de una región que codifica un gen supresor tumoral puede contribuir a la transformación de una célula normal en una célula tumoral (Michels et al., 2007). De este modo, mediante el estudio de las CNAs es posible identificar aquellas regiones del genoma cuya alteración juega un papel fundamental en el cáncer. La evolución de los métodos de detección de CNAs ha influido significativamente en el progreso de la investigación de estas alteraciones (Pös et al., 2021). Estos

métodos se dividen principalmente en dos enfoques: los métodos citogenéticos y los métodos basados en biología molecular, los cuales se presentan de manera concisa en la Figura 4. El fin último de la detección de CNAs es identificar a las regiones genómicas asociadas a la progresión tumoral, a la supervivencia de los pacientes y/o a las posibilidades de éxito de determinados tratamientos (Kallioniemi, 2008).

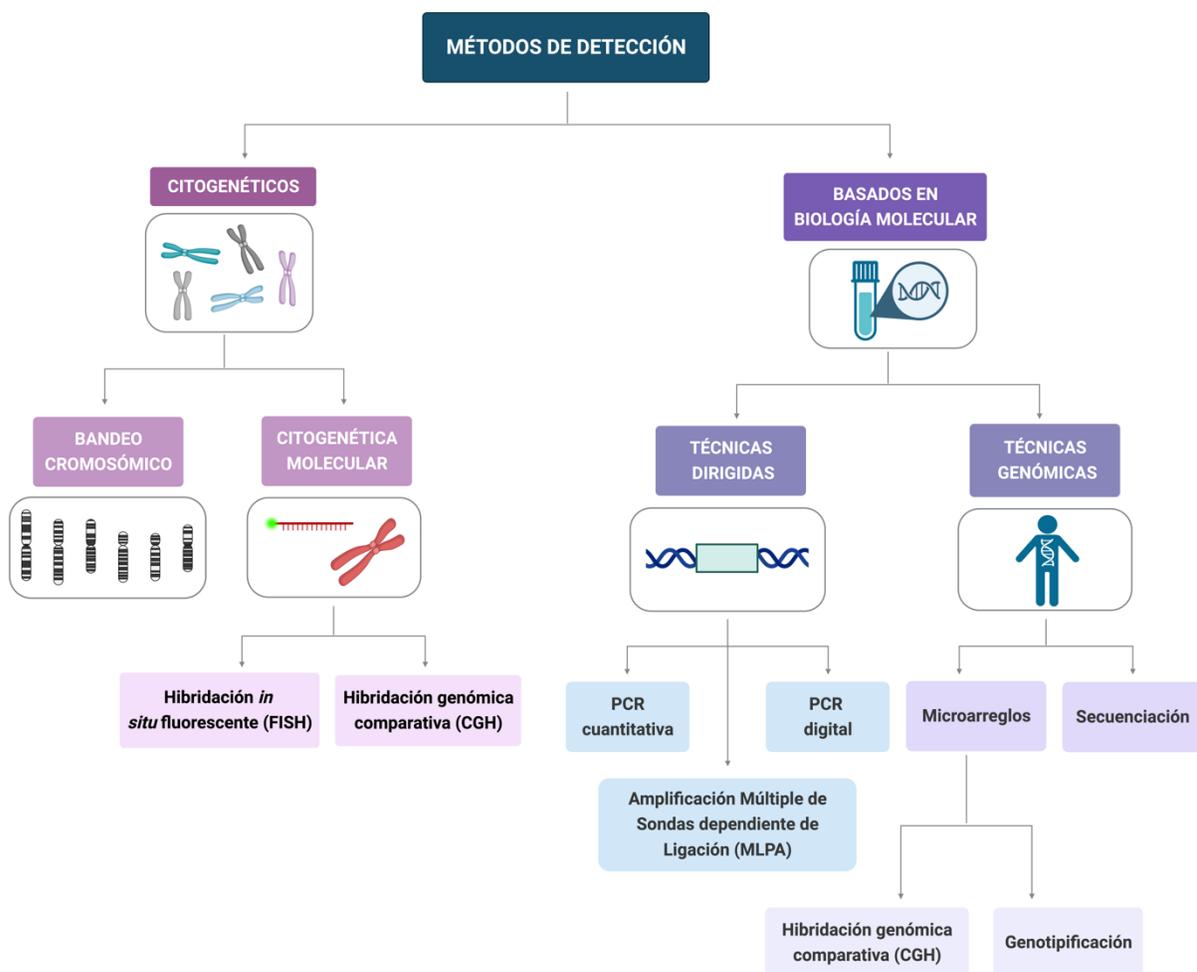


Figura 4. Clasificación de métodos para la detección de alteraciones en el número de copias de DNA. Los métodos empleados para la identificación de alteraciones en el número de copias de DNA (CNAs) se categorizan en dos principales enfoques: los métodos citogenéticos y los métodos basados en biología molecular. Los métodos citogenéticos comprenden el bandeo cromosómico, la Hibridación in situ Fluorescente (FISH) y la Hibridación Genómica Comparativa (CGH), que se basan en la observación directa de las estructuras cromosómicas. Por otro lado, los métodos basados en biología molecular abarcan técnicas como la PCR cuantitativa, la PCR digital, la Amplificación Múltiple de Sondas dependiente de Ligación (MLPA), los microarreglos de CGH, los microarreglos de genotipificación, así como las tecnologías de secuenciación, que se centran en la investigación de las alteraciones en regiones genómicas específicas o en el análisis del genoma completo. Esta clasificación proporciona una visión completa de las diversas aproximaciones utilizadas en la detección de CNAs. Creada con BioRender.com

1.3.1 Métodos citogenéticos

Los métodos citogenéticos, cuyo estudio se basa en la inspección visual de la estructura y la organización de los cromosomas, fueron los primeros empleados en la detección de cambios en el número de copias de DNA (Balajee & Hande, 2018). Los métodos citogenéticos convencionales incluyen el uso de células cultivadas a partir del tejido a estudiar y el arresto del ciclo celular en metafase empleando colchicina (Vermeesch et al., 2007). El desarrollo y mejoría de estas técnicas permitió la reducción gradual de los límites de detección, con una capacidad inicial para detectar alteraciones numéricas de cromosomas completos hasta alcanzar una resolución de unas pocas megabases (Mb) de tamaño (Pös et al., 2021).

El estudio detallado de los cromosomas humanos se realizó por primera vez en 1952, cuando se descubrió que el tratamiento hipotónico de las células podía dispersar los cromosomas en metafase para favorecer la visualización individual de las características de cada cromosoma (Hsu, 1952). Durante la década de 1970, los avances en la citogenética humana permitieron la introducción de diferentes técnicas de **bandeo cromosómico**, caracterizadas por revelar patrones de bandeo específicos para cada cromosoma (Martin & Warburton, 2015). Los métodos de bandeo más ampliamente utilizados son el bandeo Q con quinacrina basado en fluorescencia y el bandeo G obtenido mediante una digestión enzimática con tripsina y la posterior tinción con el colorante Giemsa (Schreck & Distèche, 2001).

La combinación de los métodos citogenéticos convencionales con técnicas moleculares como la hibridación condujeron al desarrollo de la citogenética molecular. Los principales métodos son la **Hibridación *in situ* Fluorescente** (FISH, *Fluorescence In Situ Hybridization*) y la **Hibridación Genómica Comparativa** (CGH, *Comparative Genomic Hybridization*), los cuales requieren de microscopía de fluorescencia (Pös et al., 2021). La técnica de FISH emplea fragmentos de DNA incorporados con nucleótidos acoplados a fluoróforos para detectar la presencia o ausencia de secuencias complementarias en células o tejidos fijados al observarse con un microscopio de fluorescencia (Martin & Warburton, 2015). La resolución de la técnica se encuentra en el rango de 100 a 200 kpb, la cual está determinada por la longitud de las sondas empleadas y que es hasta 50 veces mejor que la resolución de 5 a 10 Mb de un cariotipo obtenido por bandeo cromosómico (Cui et al., 2016). Por su parte, la técnica de CGH permite la comparación de las señales fluorescentes generadas a partir del DNA de una muestra de prueba y una muestra de referencia, a lo largo de los cromosomas con los que se hibridaron provenientes de células normales. Mediante esta técnica es posible identificar el aumento o la disminución en el número de copias de secuencias de 3 a 10 Mb de tamaño (Pös et al., 2021).

1.3.2 Métodos basados en biología molecular

Los métodos moleculares para el estudio de CNAs ofrecen una resolución significativamente superior en comparación con los métodos citogenéticos. Estos métodos se pueden clasificar en dos categorías principales: en primer lugar, las técnicas dirigidas, que se centran en la investigación de las alteraciones en regiones genómicas específicas; en segundo lugar, las técnicas genómicas, que posibilitan el análisis de las alteraciones a nivel de genoma completo, proporcionando una visión global de los cambios genómicos (Pös et al., 2021).

1.3.2.1 Técnicas dirigidas

La **Reacción en Cadena de la Polimerasa Cuantitativa** (qPCR, *quantitative Polymerase Chain Reaction*), también conocida como Reacción en Cadena de la Polimerasa en tiempo real (*real-time PCR*), es la técnica molecular basada en PCR más simple para el estudio de CNAs. Esta técnica se basa en la amplificación de un locus de prueba con número de copias desconocido y un locus de referencia con número de copias conocido, según la ploidía esperada. La fluorescencia se obtiene empleando intercalantes de DNA de doble cadena o sondas con secuencia complementaria específica cuya fluorescencia es activada mediante la amplificación del DNA templado. En cualquiera de los dos casos, la fluorescencia se duplica con cada ciclo de PCR, y la cantidad de DNA templado inicial se puede determinar a partir del número de ciclos necesarios para alcanzar un nivel umbral de fluorescencia (Ma & Chung, 2014). Los métodos basados en PCR pueden aplicarse a ensayos múltiplex para la detección simultánea de un número limitado de CNAs en un rango de longitud relativamente amplio, desde decenas de pb hasta Mb, a un bajo costo y tiempo de análisis corto (Pös et al., 2021).

Una alternativa a la qPCR tradicional para la detección de CNAs es la **PCR digital** (dPCR, *digital Polymerase Chain Reaction*). Esta técnica tiene como principal diferencia la presencia de miles de microreactores, algunos de los cuales contienen la molécula de DNA blanco mientras que otros carecen de ella (Quan et al., 2018). Posterior a la amplificación, se mide la fluorescencia en cada microreactor, considerándose como reacciones positivas aquellas en las que se alcanza un nivel umbral de fluorescencia. A diferencia de la qPCR en donde se obtiene un resultado con un valor continuo, en esta técnica el resultado es binario (sí/no). El conteo de las reacciones positivas y negativas se relaciona con las concentración de la secuencia de DNA genómico objetivo dado que el número total de microambientes de reacción con fluorescencia positiva representa el número total de moléculas blanco presentes en la muestra. Estadísticamente, cada

microambiente de reacción puede contener más de una molécula de DNA blanco, por lo que el conteo debe corregirse mediante una distribución de Poisson (Mao et al., 2019). El número de copias se calcula comparando el número de moléculas que surgen del segmento de interés con el número de moléculas que surgen de una región genómica de referencia (Bell et al., 2018).

Otra técnica basada en PCR es la **Amplificación Múltiple de Sondas dependiente de Ligación** (MLPA, *Multiplex Ligation-dependent Probe Amplification*), considerada un estándar de oro en el diagnóstico de diversas enfermedades humanas relacionadas con alteraciones en el número de copias de DNA (Hömig-Hölzel & Savola, 2012). En esta técnica se emplea un conjunto de hasta 50 sondas dirigidas a regiones específicas del DNA. Cada una de las sondas se compone de dos secuencias de nucleótidos en 5' y 3', físicamente separadas. La hibridación de ambos fragmentos con las regiones genómicas de interés posiciona los extremos internos de cada segmento de la sonda de forma adyacente de modo que permite su unión mediante una reacción de ligación. Además, las sondas contienen sitios de unión para primers para su posterior amplificación por PCR y en la región 3' poseen una secuencia adaptadora de longitud variable para obtener amplicones de PCR de diferente tamaño. Los productos de amplificación se analizan mediante electroforesis capilar, la cual detecta la intensidad de fluorescencia. La cuantificación relativa de la señal obtenida de fragmentos de tamaño único proporciona información sobre el número de copias de la regiones de DNA de interés (Schouten et al., 2019; Stuppia et al., 2012). La técnica de MLPA representa un mejor abordaje cuando se desea analizar un conjunto limitado de regiones del genoma. Por el contrario, los enfoques basados en PCR emergen como una elección más acertada al analizar regiones genómicas individuales (van Eijk et al., 2010).

1.3.2.2 Técnicas genómicas

En un esfuerzo por superar las limitaciones de resolución asociadas a la técnica de CGH tradicional, surgió el método de **CGH basado en microarreglos** (aCGH, *Microarray-based Comparative Genomic Hybridization*), caracterizado por su capacidad mejorada para el análisis de alta precisión de regiones específicas del genoma. Este enfoque marcó una transición desde la observación de cromosomas completos hacia el análisis simultáneo de miles de sondas de DNA específicas que representan diversas regiones en el genoma, las cuales se encuentran inmovilizadas y ordenadas en la superficie de vidrio de un chip (Redon & Carter, 2009). Estas sondas varían en tamaño, desde oligonucleótidos diseñados para representar regiones con longitudes de 25 a 85 pb, hasta cromosomas artificiales bacterianos, que pueden comprender de 80,000 a 200,000 pb. Dada la diferencia de tamaño entre las sondas y los cromosomas en

metafase, esta técnica ofrece una resolución considerablemente mayor que la técnica de CGH tradicional (Oostlander et al., 2004). La resolución está en función del tamaño de las sondas y la distancia genómica entre ellas, la cual es generalmente alrededor de 120 kpb (Ahn et al., 2015).

Para realizar el análisis, se preparan bibliotecas de DNA a partir de una muestra de interés y una muestra de referencia, ambas marcadas con diferentes fluoróforos, las cuales se hibridan al microarreglo. El microarreglo se lava y se lleva a cabo la detección de la fluorescencia en cada región genómica, la cual está en proporción con la cantidad de DNA que se ha unido (Kim & Pollack, 2009). Cuando el número de copias de un segmento de DNA en la muestra de interés se encuentra dentro de los límites normales, se observa una cantidad igual de DNA unido a la sonda correspondiente en el microarreglo proveniente de la muestra de interés y de la muestra de referencia. Esto genera una señal uniforme en el microarreglo. Sin embargo, en casos de ganancia o pérdida del número de copias en la muestra, la proporción de DNA marcado de la muestra y de la referencia se verá alterada, lo que se manifestará como un cambio en la señal detectada (Ahn et al., 2015; Holcomb & Trask, 2011).

Por otra parte, los **microarreglos de genotipificación** centran su atención en la identificación de Polimorfismos de un Solo Nucleótido (SNP, *Single Nucleotide Polimorfism*). Si bien su enfoque principal recae en la caracterización de estas variantes, la información acerca del estado de número de copias de DNA se obtiene de manera secundaria durante el proceso de análisis. El enfoque dual de genotipificación y detección de CNAs permite que esta tecnología sea ampliamente empleada en la investigación genómica (LaFramboise, 2009).

El funcionamiento de la tecnología se basa en la hibridación del DNA con sondas para marcadores de SNPs diseñadas para reconocer regiones genómicas específicas. Los SNPs a los cuales se dirigen estas sondas generalmente son bialélicos, lo que significa que para cada SNP se incluyen dos tipos de sondas diseñadas para reconocer entre dos alelos conocidos (Lin et al., 2013). Para llevar a cabo la hibridación del DNA al microarreglo, se procesa el material genético con enzimas de restricción, se realiza la unión de adaptadores y una reacción de PCR para obtener un conjunto de productos de PCR con un rango de tamaño específico. Los productos de PCR se fragmentan, se marcan con un fluoróforo y se hibridan al microarreglo (Nowak et al., 2009). El genotipo del SNP se determina mediante la detección de la señal de fluorescencia para cada una de las sondas diseñadas para reconocer los dos alelos conocidos de un SNP particular (Lin et al., 2013). Dado que las CNAs pueden generar un aumento o disminución en la intensidad de fluorescencia, durante el análisis se obtiene adicionalmente la información sobre el estado del

número de copias de DNA (LaFramboise, 2009). Las regiones genómicas alteradas en el número de copias que abarcan múltiples SNPs, muestran patrones distintivos en sus relaciones de intensidad de fluorescencia, en comparación con las regiones genómicas diploides (Lin et al., 2013). Algunos loci genómicos resultan notoriamente desafiantes de analizar para la detección de SNPs, de modo que dichas regiones son excluidas de las sondas que componen el microarreglo, lo que impide la detección de CNAs en esas regiones (Szu Hai & Vermeer, 2015).

Parte de este trabajo consiste en analizar biopsias de tejido tumoral fijadas con formaldehído y embebidas en parafina (FFPE, *Formalin-Fixed Paraffin-Embedded*) de pacientes con cáncer de mama mediante microarreglos de genotipificación para la detección de CNAs. Dado que los tejidos FFPE suelen presentar desafíos técnicos para el análisis molecular debido a la degradación y fragmentación del DNA causadas por los métodos de fijación y almacenamiento (Guyard et al., 2017), es fundamental el empleo de tecnologías que puedan abordar estas limitaciones. Un ejemplo de microarreglo que puede manejar este tipo de muestras es el microarreglo de genotipificación OncoScan™ CNV Plus Assay. Este utiliza sondas de inversión molecular (MIP, *Molecular Inversion Probe*), cada una con dos regiones de homología para su hibridación con la región genómica de interés. Con segmentos de homología de 20 pb, estas sondas pueden unirse a fragmentos de DNA tan cortos como 40 pb. Esta capacidad hace que los microarreglos basados en MIP sean menos dependientes de la calidad del DNA de partida en comparación con otras tecnologías genómicas que requieren fragmentos DNA más largos e íntegros (Jung et al., 2017).

Desde finales de los años 1990, las tecnologías basadas en microarreglos se han empleado extensamente como una opción factible y de resolución relativamente alta para la detección de CNAs en una amplia variedad de regiones genómicas (Carter, 2007). Las plataformas más recientes, que incluyen más de un millón de sondas, han ampliado significativamente su capacidad de detección, permitiendo un límite de detección de hasta 5 a 25 kpb (Magi et al., 2019). Esto ha favorecido su utilidad en la investigación genómica y en la identificación de CNAs en tumores humanos, contribuyendo al descubrimiento de oncogenes y genes supresores de tumores relevantes para comprender la biología del cáncer (Szu Hai & Vermeer, 2015). En el ámbito clínico, los microarreglos han sido la prueba más ampliamente utilizada para la detección de CNAs. No obstante, estos aún presentan limitaciones asociadas con el proceso de hibridación, lo que conlleva una baja sensibilidad y precisión, además de desafíos de resolución asociados a la cobertura y densidad de las sondas utilizadas en ellos (Nabavi & Zare, 2022).

En la actualidad, las **tecnologías de Secuenciación de Nueva Generación** (NGS), también conocidas como tecnologías de secuenciación de segunda generación, representan herramientas valiosas para el diagnóstico clínico, puesto que ofrecen un enfoque sensible y preciso para la detección de los principales tipos de alteraciones genómicas, que incluyen las CNAs (Wang et al., 2014). El análisis de CNAs basado en estas tecnologías se basa en tres estrategias principales: la secuenciación a nivel de genoma completo (WGS, *Whole Genome Sequencing*), la secuenciación de exoma completo (WES, *Whole Exome Sequencing*) y la secuenciación dirigida (Pös et al., 2021). Particularmente, la estrategia de secuenciación de WGS, en combinación con algoritmos computacionales, ha mejorado sustancialmente la detección de CNAs, permitiendo la identificación precisa de los límites de las regiones genómicas a nivel de pares de bases (Magi et al., 2019).

El enfoque de **secuenciación a nivel de genoma completo de alta profundidad** (*High-depth Whole Genome Sequencing*) es considerado el "estándar de oro" para la detección de todos los tipos de alteraciones presentes en el DNA que afectan la arquitectura genómica, incluidas las CNAs (Sims et al., 2014). Este enfoque permite aumentar la sensibilidad, detectar los límites de las CNAs con alta precisión y estimar el número de copias absoluto gracias a una mayor profundidad de lectura (Medvedev et al., 2010).

Aunque los costos de secuenciación han disminuido en los últimos años, estos siguen siendo significativos y varían según la estrategia de secuenciación utilizada. En el ámbito clínico, el diseño experimental para la detección de CNAs busca generar resultados sólidos al menor costo de secuenciación posible (Qin, 2019). En respuesta a esta necesidad, un enfoque de secuenciación que ha ido ganando rápida aceptación en los últimos años como alternativa para el análisis de CNAs en diversos centros clínicos es la **secuenciación a nivel de genoma completo de baja profundidad** (sWGS, *shallow Whole Genome Sequencing*, ó también conocida como *low pass Whole Genome Sequencing*) (Mazzonetto et al., 2024; Wragg et al., 2024; Zheng et al., 2022).

El enfoque de sWGS, se basa en un experimento de secuenciación a nivel de genoma completo con una profundidad de cobertura que varía entre 0.1x y 1x, lo que lo convierte en una alternativa de bajo costo que sigue proporcionando resultados altamente precisos (Raman et al., 2018). Esto contrasta con la secuenciación a nivel de genoma completo de alta profundidad, la cual generalmente se caracteriza por tener una profundidad de cobertura entre 30x y 50x, pero con un costo más elevado (Homburger et al., 2019; L. Raman et al., 2020).

El enfoque de sWGS busca maximizar la información obtenida a partir de los datos de secuenciación, a expensas de reducir la profundidad de cobertura del análisis. Sin embargo, este sacrificio se compensa al lograr una cobertura más extensa y, potencialmente, obtener información relevante de más de una muestra simultáneamente (Chat et al., 2022; Lou et al., 2021). Las aplicaciones más destacadas de la estrategia sWGS en la actualidad incluyen pruebas prenatales no invasivas, diagnóstico de cáncer mediante el análisis de biopsias líquidas, tejido tumoral fresco o tejido tumoral FFPE, así como diagnóstico genético preimplantacional y estudio de aberraciones cromosómicas (Chin et al., 2018; Deleye et al., 2015; Dheedene et al., 2016; Euskirchen et al., 2017; Lennart Raman et al., 2020).

1.3.2.2.1 Secuenciación por síntesis de Illumina

Introducida en 2006, la secuenciación por síntesis de Illumina ha destacado como la tecnología de NGS más robusta y ampliamente utilizada, revolucionando la secuenciación masiva y en paralelo de múltiples segmentos de DNA (Adewale, 2020). Este enfoque de secuenciación se basa en la generación de lecturas cortas, generalmente con un tamaño de alrededor de 150 pb (Oehler et al., 2023). La secuenciación Illumina ha desempeñado un papel importante en la reducción significativa de los costos y tiempos de procesamiento para la secuenciación de DNA, lo que ha impulsado su implementación en la investigación genómica. Su capacidad para generar datos con alto rendimiento y de forma asequible la han convertido en una herramienta fundamental para una amplia gama de aplicaciones, proporcionándole una ventaja significativa sobre los microarreglos (Hurd & Nelson, 2009; Wang et al., 2014).

Empleando una variedad de enfoques experimentales y estrategias computacionales, la secuenciación Illumina ha contribuido en el área de la genómica del cáncer a la obtención de datos de alta resolución para la identificación de CNAs relevantes a partir de datos de secuenciación de lecturas cortas (Mardis, 2019; McDermott, 2015). Estos esfuerzos han culminado en la implementación exitosa de la secuenciación de lecturas cortas en el entorno clínico (Frampton et al., 2013; Zehir et al., 2017). No obstante, esta aplicación no está exenta de desafíos, siendo su limitación más notable aquella relacionada con los costos de la implementación clínica, lo cual ha restringido su utilización a centros clínicos grandes con recursos sustanciales (Baslan et al., 2021).

En los últimos años, los avances tecnológicos sentaron las bases para una nueva forma de secuenciación denominada **secuenciación con nanoporos**. Esta tecnología ofrece diversas ventajas en relación con las tecnologías tradicionales de secuenciación de lecturas cortas, incluyendo la detección de molécula única, la portabilidad, los costos de instrumentación bajos, así como la capacidad para obtener información a partir de fragmentos de DNA largos (Baslan et al., 2021; Reuter et al., 2015).

Dado que la tecnología de secuenciación con nanoporos constituye un aspecto primordial para esta tesis, se abordará con mayor detalle en la siguiente sección.

1.4 Secuenciación con nanoporos

En 2022, la secuenciación de lecturas largas se destacó como la metodología más relevante en el campo de las ciencias genómicas. Las dos tecnologías comerciales líderes en este ámbito son la secuenciación de molécula única en tiempo real (SMRT, *Single Molecule Real-Time*) de Pacific Biosciences (PacBio®) y la secuenciación con nanoporos de Oxford Nanopore Technologies (ONT®) ("Method of the Year 2022: long-read sequencing," 2023).

La secuenciación con nanoporos, una tecnología de tercera generación, ha experimentado un rápido avance desde su introducción, permitiendo su implementación en una diversidad de aplicaciones (Zeng et al., 2020). El MinION (ONT®) fue el primer sistema de secuenciación con nanoporos introducido en 2014. Este secuenciador portátil se distingue por generar lecturas en tiempo real a partir de moléculas individuales, lo que, en combinación con procedimientos ágiles de preparación de librerías, reduce significativamente el intervalo entre el procesamiento de muestras y el análisis de datos (Deamer et al., 2016; Jain et al., 2016). Así mismo, mediante esta tecnología ha sido posible secuenciar moléculas de DNA de cualquier longitud, reportándose lecturas ultra largas de hasta 4 Mpb (Romagnoli et al., 2023). El MinION también se caracteriza por su tamaño compacto y su costo inferior en comparación con las plataformas convencionales de secuenciación de lecturas cortas, lo que lo hace adecuado para aplicaciones fuera del entorno de laboratorio tradicional, incluyendo su uso en entornos clínicos (Baslan et al., 2021). Posteriormente, en 2017, se comercializó el sistema de secuenciación GridION (ONT®), un instrumento que posibilita la realización de cinco experimentos de secuenciación con nanoporos de forma simultánea e independiente al permitir la introducción de cinco celdas de flujo; este secuenciador integra una infraestructura informática que facilita el proceso de llamado de bases (Cozzuto et al., 2020; Tshiabuila et al., 2022).

Una de las aplicaciones más importantes de la secuenciación con nanoporos fue la implementación de la tecnología en la obtención de la versión más completa del genoma humano, que incluye 3.055 billones de pares de bases, a través del Consorcio Telómero a Telómero (*T2T, Telomere-to-Telomere Consortium*) (Nurk et al., 2022). Utilizando esta tecnología, se superaron limitaciones previas al posibilitar la secuenciación del 8% del genoma faltante, anteriormente inaccesible debido a la presencia de regiones altamente repetitivas que escapaban a las capacidades de las tecnologías de secuenciación convencionales (Eichler et al., 2004). Esta tecnología no solo ha contribuido a completar integralmente el genoma humano, sino que también ha sentado las bases para una comprensión más profunda de la variabilidad genómica en humanos, impulsando así futuros avances en la investigación genómica y la comprensión de las bases moleculares de diversas enfermedades (Nurk et al., 2022). Además, mediante esta tecnología fue posible obtener información de forma directa sobre el estado de metilación del DNA, vinculado comúnmente a enfermedades como el cáncer, enriqueciendo aún más nuestra comprensión del genoma humano (Gershman et al., 2022).

1.4.1 Fundamento de la secuenciación con nanoporos

El término "nanoporo" hace referencia a poros biológicos de naturaleza proteica y escala nanométrica, los cuales se encuentran insertados en una membrana polimérica dieléctrica dentro de una celda de flujo y son empleados para llevar a cabo el proceso de secuenciación (Zeng et al., 2020). Cuando se aplica un potencial eléctrico, se genera una corriente de iones y un flujo del material genético a través de los nanoporos. Los nucleótidos en el poro alteran el flujo de iones, provocando la disrupción de la corriente eléctrica que se utiliza para inferir la secuencia de las moléculas de ácidos nucleicos (Wang et al., 2021). Los detalles del funcionamiento de la secuenciación con nanoporos se ilustran en la Figura 5.

El proceso de llamado de bases o *basecalling*, un paso fundamental en la secuenciación con nanoporos, consiste en la conversión de señales eléctricas en secuencias de ácidos nucleicos. La herramienta que realiza este proceso está diseñada para identificar las secuencias a partir de mediciones crudas de corriente eléctrica, lo que implica una notable complejidad en su ejecución. Esto se debe a que las señales eléctricas provienen de moléculas individuales, lo que origina datos con ruido y propensos a la estocasticidad (Wick et al., 2019). Adicionalmente, durante la secuenciación, el valor de la señal eléctrica está determinado por el conjunto de nucleótidos que residen en un momento específico en el punto más estrecho del nanoporo. Estos nanoporos tienen la capacidad de albergar k nucleótidos simultáneamente, lo que se denomina

k-mers. Por ejemplo, en las celdas de flujo R9.4.1 utilizadas en la química de secuenciación R9, *k* equivale a 5 nucleótidos, lo que genera una gran cantidad de combinaciones posibles en el caso de un modelo estándar con cuatro nucleótidos distintos ($4^5 = 1024$ combinaciones) (Zeng et al., 2020).

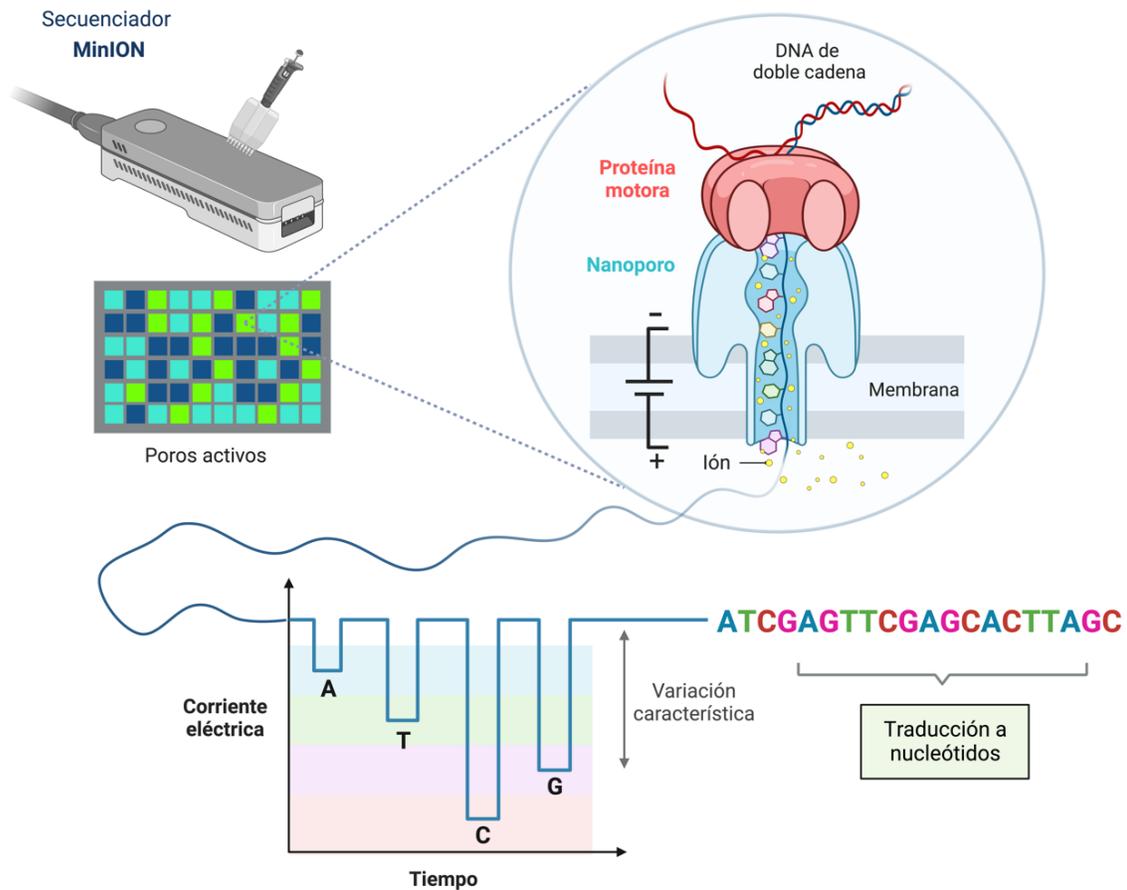


Figura 5. Fundamento de la secuenciación con nanoporos. El sistema de secuenciación con nanoporos (ej. MinION) utiliza poros biológicos insertados en una membrana sintética que se encuentra sumergida en un fluido conductor. Para llevar a cabo la secuenciación, las moléculas de DNA se preparan siguiendo uno de los protocolos de preparación de librerías, lo cual implica la unión de un adaptador precargado con una proteína motora a una de las hebras del DNA. La proteína motora realiza la apertura de las hebras del DNA de doble cadena. Al aplicar una diferencia de potencial eléctrico, se genera la conducción de iones a través del nanoporo y debido a la carga negativa de la molécula de DNA, la proteína motora permite que una sola hebra atraviese el nanoporo hacia el lado positivo de la membrana. Mientras pasa por la sección más estrecha del nanoporo, cada nucleótido genera una disrupción característica en la corriente eléctrica. Estas señales eléctricas se traducen en nucleótidos en un proceso denominado llamado de bases, lo cual permite conocer la secuencia de las moléculas de DNA. Creada con BioRender.com

Anteriormente, la secuenciación con nanoporos presentaba como principal desafío la baja precisión de lectura en comparación con las tecnologías de secuenciación de lecturas cortas, lo que se reflejaba en sus tasas de error de hasta 15% (Laver et al., 2015; Sahlin & Medvedev, 2021). Esta baja precisión complicaba el análisis de muestras complejas para la detección de alteraciones de uno o pocos nucleótidos, tales como mutaciones puntuales, inserciones y deleciones. A pesar de esto, en ciertas aplicaciones como la detección de alteraciones cromosómicas numéricas y estructurales, los beneficios de las lecturas largas superaban dicha limitación (Cretu Stancu et al., 2017). Como resultado, en sus etapas iniciales, la tecnología de secuenciación con nanoporos no pudo igualar el desempeño y el rendimiento alcanzado por las tecnologías de secuenciación tradicionales (Chen & Xu, 2023).

En desarrollos recientes, la tecnología ha experimentado mejoras significativas logrando una precisión de lectura superior al 99% (Cuber et al., 2023). Estas mejoras incluyen la introducción de la nueva química de secuenciación R10, diseñada específicamente para mejorar la precisión de lectura en comparación con su predecesora, la química de secuenciación R9 (Sanderson et al., 2023; Wagner et al., 2023). Además, se ha implementado el llamado de bases dúplex, que aprovecha la información de ambas cadenas de las moléculas de DNA para aumentar la precisión de lectura en una segunda ronda de llamado de bases. Este último avance permite alcanzar una precisión mayor a 99.9%, equivalente a un valor de calidad Q30, la cual es equiparable a la precisión de lectura obtenida mediante la tecnología de secuenciación Illumina (Bogaerts et al., 2024).

1.4.2 Aplicación de la tecnología en la detección de alteraciones en el número de copias de DNA

Aunque la capacidad de secuenciar moléculas de DNA largas ha recibido la mayor atención dentro de la investigación biomédica, es importante destacar las ventajas más relevantes de la tecnología de secuenciación para la aplicación de detección de CNAs (Baslan et al., 2021; Bogaerts et al., 2024; Martignano et al., 2021), que es el punto medular de este trabajo.

- **Secuenciación en tiempo real:** La tecnología permite realizar el llamado de bases en tiempo real mientras se lleva a cabo la secuenciación, lo que posibilita analizar los datos a medida que se generan. Esta capacidad acelera significativamente los tiempos de detección de CNAs, lo cual es especialmente relevante dado el interés de que esta aplicación tenga implementación clínica.

- **Costos de instrumentación bajos:** La tecnología ofrece como alternativa de menor costo el dispositivo de secuenciación MinION (ONT®), el cual requiere de una inversión inicial baja (aproximadamente 1,000 €). Este costo es significativamente menor en comparación con los secuenciadores ofrecidos por otras tecnologías de secuenciación, cuyos precios suelen estar en el rango de decenas de miles de euros. Los costos bajos favorecen que la tecnología sea accesible a la mayoría de centros clínicos en los que potencialmente se pueda implementar esta aplicación.
- **Reutilización de celdas de flujo:** Después de la generación de un número suficiente de lecturas para la detección de CNAs en una muestra, la corrida de secuenciación puede detenerse y explotar cualquier capacidad de secuenciación residual de la misma celda de flujo para realizar el análisis genómico de nuevas muestras.
- **Flujo de trabajo sin PCR:** Esta tecnología se distingue por prescindir de la amplificación mediante la técnica de Reacción en Cadena de la Polimerasa (PCR, *Polymerase Chain Reaction*), evitando así la introducción de sesgos en el proceso de amplificación que podrían influir en la detección de CNAs.

1.5 Estrategia bioinformática para la detección de alteraciones en el número de copias de DNA

La detección de CNAs a través de datos de secuenciación a nivel de genoma completo ha involucrado una variedad de enfoques metodológicos (Zhao et al., 2013). Entre ellos, el análisis basado en la profundidad de lectura representa uno de los abordajes más utilizados, distinguiéndose por ser menos restrictivo en términos del tamaño de las lecturas empleadas. Además, los datos de profundidad de lectura pueden provenir tanto de experimentos de secuenciación *single-end* como de *paired-end*, en tanto se implementen procedimientos adecuados de mapeo y normalización (Teo et al., 2012; Wei & Huang, 2020).

Los métodos basados en la profundidad de lectura tienen como hipótesis que la cantidad de lecturas secuenciadas en una región genómica específica está directamente relacionada con el número de copias presentes en dicha región (Wei & Huang, 2020). Por ejemplo, en comparación con las regiones que poseen un número de copias diploide, se espera que la cantidad relativa de lecturas generadas durante un experimento de secuenciación sea mayor en las regiones genómicas con ganancia en el número de copias y menor en aquellas con pérdida en el número de copias (Figura 6).

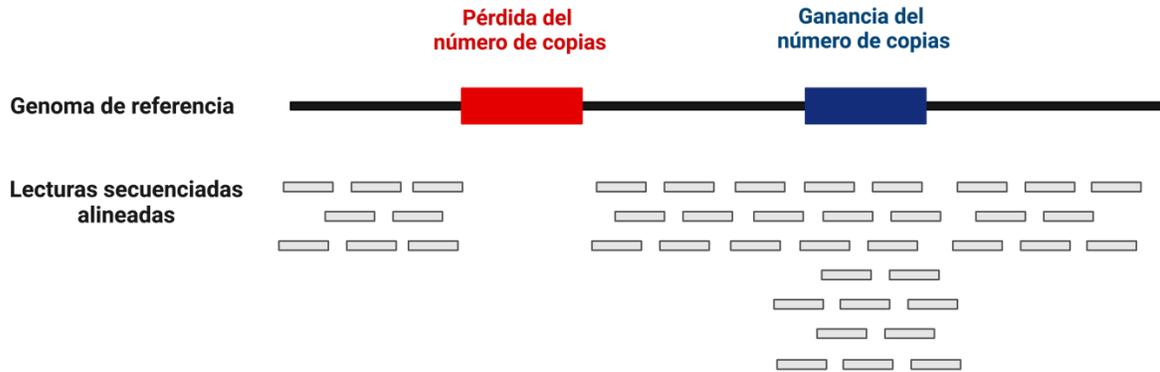


Figura 6. La profundidad de lectura como método para la detección de alteraciones en el número de copias de DNA. El enfoque basado en la profundidad de lectura detecta CNAs mediante el conteo del número de lecturas secuenciadas que se alinearon a distintas ventanas genómicas de un genoma de referencia. Asumiendo que el proceso de secuenciación es uniforme, se espera que el número de lecturas asignadas a una región genómica específica sea proporcional a su presencia en una muestra de DNA. Con base en ello, se puede estimar el número de copias de cualquier región genómica al realizar el conteo del número de lecturas alineadas a dicha región. En particular, en regiones con pérdida o ganancia del número de copias se espera una disminución o aumento en el número de lecturas alineadas, respectivamente. Creada con BioRender.com

Bajo esta estrategia, el genoma de referencia se divide en ventanas genómicas de un tamaño definido, generalmente entre 100 kpb y 2 Mpb. Después, se calcula el número de lecturas secuenciadas que se alinean a cada una de las ventanas genómicas, lo que constituye la profundidad de lectura. El conteo de lecturas se normaliza considerando la mapeabilidad y el contenido de GC, los cuales representan dos factores principales que pueden generar ruido durante el análisis (Magi et al., 2019).

El conteo de lecturas para cada ventana genómica se expresa como valores de \log_2 ratio, de acuerdo con la siguiente fórmula:

$$\log_2ratio = \log_2 \left(\frac{RC}{EXP} \right)$$

donde RC es el número de lecturas alineadas a una ventana genómica particular y EXP representa el número de lecturas esperado para una región genómica con un número de copias diploide. Este valor se calcula de acuerdo a la siguiente fórmula:

$$EXP = \frac{N * L}{G}$$

donde N es el número total de lecturas generadas durante el experimento de secuenciación, L es el tamaño de la ventana genómica y G es el tamaño del genoma de referencia.

Posteriormente, se utiliza un algoritmo de segmentación para identificar ventanas genómicas contiguas con valores de \log_2 ratio semejantes y ventanas genómicas con valores de \log_2 ratio significativamente diferentes respecto a las regiones genómicas adyacentes, lo que permite establecer los límites de las CNAs (Janevski et al., 2012).

A pesar de las diferencias en los algoritmos de segmentación y los métodos de normalización entre las distintas herramientas desarrolladas para la detección de CNAs a partir de lecturas de secuenciación a nivel de genoma completo, el principio general subyacente en su funcionamiento es muy semejante entre ellas (Magi et al., 2012). En la actualidad, existen diversas herramientas bioinformáticas basadas en la profundidad de lectura que facilitan la detección de CNAs a partir de datos de sWGS, superando incluso a los métodos basados en microarreglos (Smolander et al., 2021; Zhou et al., 2018).

1.6 Las biopsias líquidas como herramienta clínica para la detección de alteraciones en el número de copias de DNA

1.6.1 Limitaciones de las biopsias de tejido convencionales

Las biopsias de tejido tumoral, también llamadas biopsias sólidas, representan el estándar de oro actual en el manejo clínico del cáncer (Mannelli, 2019). Este tipo de muestra es relevante en la clínica dado que permite estudiar la histología de los tumores, identificar biomarcadores que guían el diagnóstico y tratamiento, así como obtener información sobre las alteraciones moleculares presentes en el tumor que influyen en la predicción y el pronóstico de la enfermedad (Alba-Bernal et al., 2020). A pesar de su relevancia, estas biopsias presentan diversas limitaciones técnicas que restringen su aplicación. Por ejemplo, el procedimiento para obtenerlas puede ser altamente invasivo y doloroso para los pacientes. En algunos casos, la ubicación del tumor dificulta su acceso, complicando la obtención de la muestra necesaria. Además, hay situaciones en las que el estado de salud de los pacientes no permite someterse a este procedimiento, lo que limita el acceso a información esencial para un diagnóstico y tratamiento adecuado del cáncer. Otra limitación importante es que este tipo de muestra no puede ser repetida múltiples veces a lo largo del tiempo, lo que restringe su utilidad en el seguimiento de la enfermedad (Alanazi et al., 2022; Kilgour et al., 2020).

Este conjunto de limitaciones muestra la necesidad de explorar nuevas herramientas clínicas alternativas que se caractericen por ser menos invasivas y que permitan apoyar el diagnóstico y seguimiento del cáncer. En respuesta a ello, las biopsias no invasivas, llamadas

biopsias líquidas, han adquirido una mayor relevancia en los últimos años como una alternativa para detectar las características de los tumores a partir de diversos fluidos biológicos (Connal et al., 2023; Heidrich et al., 2021).

1.6.2 Biopsias líquidas

Las biopsias líquidas, que comprenden el análisis no invasivo de componentes circulantes derivados del tejido tumoral, representan una herramienta clínica innovadora para la medicina de precisión (Figura 7) (Rolfo & Russo, 2020). Aunque la biopsia líquida generalmente se asocia con la obtención de muestras sanguíneas, prácticamente todos los fluidos corporales pueden emplearse en este tipo de análisis. Entre los más comunes se encuentran la orina, la saliva, las heces, el líquido cefalorraquídeo y el líquido pleural (De Rubis et al., 2019).

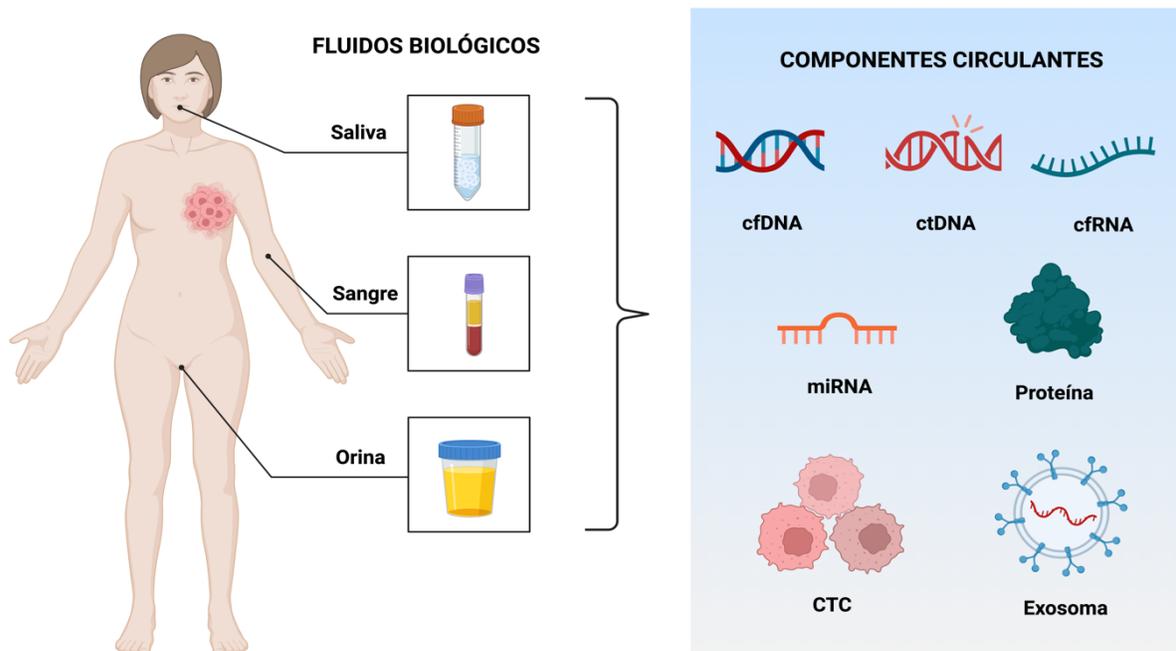


Figura 7. Las biopsias líquidas son una herramienta clínica que permite obtener información molecular clínicamente relevante sobre los tumores, de forma no invasiva. Las biopsias líquidas representan un abordaje clínico alternativo que busca superar las limitaciones asociadas a las biopsias de tejido tumoral. Estas biopsias permiten la recolección y el estudio de diversos componentes circulantes presentes en la saliva, sangre, orina y otros fluidos biológicos, que provienen tanto de tumores primarios como metastásicos. Estos componentes circulantes, que constituyen el “circuloma tumoral”, incluyen a las células tumorales circulantes (CTCs) y a los productos de células tumorales. Los productos de células tumorales son los ácidos nucleicos circulantes, como DNA libre circulante (cfDNA), DNA tumoral circulante (ctDNA) y RNA libre circulante (cfRNA); microRNAs (miRNAs), proteínas, exosomas, entre otros. Con base en estos componentes, las biopsias líquidas tienen diferentes aplicaciones en la práctica clínica, tales como el diagnóstico, la detección de recurrencia, la elección de tratamiento, la predicción de la respuesta terapéutica y el seguimiento clínico (Kilgour et al., 2020; Lone et al., 2022). Creada con BioRender.com

Durante la última década, los avances en la investigación de las biopsias líquidas han evidenciado su capacidad para proporcionar información clínicamente relevante en diversos tipos tumorales (Connal et al., 2023). El análisis de componentes circulantes como las células tumorales circulantes (CTCs), el DNA libre circulante (cfDNA) y los exosomas permite identificar una amplia gama de características de los tumores. Estas van desde alteraciones genéticas y epigenéticas, incluyendo mutaciones, CNAs, rearrreglos cromosómicos, hipometilación e hipermetilación, hasta información clínica crucial como el estadio tumoral, el pronóstico de la enfermedad y la respuesta al tratamiento (Cisneros-Villanueva et al., 2022; Gao et al., 2022; Lone et al., 2022).

1.6.2.1 DNA libre circulante

El término cfDNA hace referencia a los fragmentos de DNA extracelulares presentes en diversos fluidos corporales, liberados mediante procesos de apoptosis, necrosis y secreción activa a través de vesículas extracelulares (Aucamp et al., 2018). Dado que estos fragmentos suelen tener una longitud de 166 pb, coincidiendo con la extensión de una región de DNA envuelta alrededor de un nucleosoma, se sugiere que la apoptosis es el principal mecanismo de liberación (Bronkhorst et al., 2019). El cfDNA está presente tanto en individuos sanos como en enfermos, pero en los primeros se ha observado que los niveles en plasma son generalmente bajos, aproximadamente 10–15 ng/mL. Sin embargo, esta concentración puede incrementarse debido a procesos inflamatorios, cirugías, ejercicio, entre otros, sin necesariamente estar asociado con un proceso maligno (Connal et al., 2023; Gao et al., 2022). No obstante, en pacientes con cáncer, una fracción pequeña del cfDNA, generalmente entre el 0.1% y el 10%, se origina en el tejido tumoral y se conoce como **DNA tumoral circulante** (ctDNA) (Alimirzaie et al., 2019).

1.6.2.1.1 DNA tumoral circulante

El ctDNA presenta características distintivas que lo diferencian del cfDNA. Por ejemplo, tiene una longitud promedio más corta, alrededor de 144 pb. Además, se distingue por exhibir alteraciones genéticas y epigenéticas específicas del tumor de procedencia, incluyendo mutaciones, CNAs y patrones de metilación (Connal et al., 2023; Gao et al., 2022).

La proporción de ctDNA, respecto al cfDNA total, presente en los fluidos corporales muestra una gran variabilidad que responde a diversos factores específicos del tumor. Estos incluyen la carga mutacional tumoral (TMB, *Tumor Mutational Burden*), la tasa de proliferación y de muerte celular, el tamaño, la vascularización, la localización y el estadio clínico del tumor, así como la respuesta al tratamiento (Bettegowda et al., 2014). Por ejemplo, los niveles de ctDNA

varían según el estadio de la enfermedad. Se ha reportado que en pacientes con tumores en estadios tempranos, la proporción de ctDNA tiende a ser menor al 0.1%, mientras que en aquellos con tumores no metastásicos en estadios avanzados, representa alrededor del 1%. Sin embargo, en pacientes con tumores metastásicos en estadios avanzados, esta fracción puede alcanzar hasta un 10%, equivalente a una concentración de hasta 1,000 ng/mL (Diehl et al., 2008; Tomar et al., 2023). Así mismo, el tamaño del tumor también juega un papel crucial en la liberación de ctDNA. De acuerdo con Pinzani et al. (2021), en pacientes con tumores mayores a 100 g, el tumor puede liberar hasta un 3.3% del material genético hacia la circulación diariamente.

A través del análisis del ctDNA, es posible obtener información tanto cualitativa como cuantitativa (Wan et al., 2017). Desde una perspectiva cualitativa, se puede acceder al perfil de mutaciones, amplificaciones, deleciones, translocaciones y del estado de metilación, lo que permite la identificación de alteraciones genómicas asociadas al diagnóstico, a subtipos moleculares específicos, la respuesta terapéutica, entre otros (De Rubis et al., 2019). Por otro lado, la información cuantitativa se puede obtener mediante la determinación de la fracción alélica mutada, la cual se relaciona clínicamente con la detección de enfermedad mínima residual y metástasis ocultas, así como con la predicción de recurrencia y la evaluación de la efectividad de la terapia (Coombes et al., 2019; Misale et al., 2012; Thierry et al., 2014).

Si bien algunos trabajos han destacado la utilidad del ctDNA en la detección de cáncer en estadios tempranos (Beaver et al., 2014; Cohen et al., 2018; Phallen et al., 2017), en algunos tipos tumorales, como cáncer tiroideo y cáncer renal, la detección de ctDNA puede resultar limitada y menos sensible (Zhang et al., 2021). Además, en pacientes tratados con quimioterapia neoadyuvante, se ha observado que en aproximadamente el 90% de los casos, el ctDNA se vuelve indetectable, ya que los niveles descienden significativamente como respuesta al tratamiento (Diehl et al., 2008).

En este sentido, la detección de ctDNA puede resultar técnicamente desafiante, particularmente en aplicaciones en estadios tempranos, de modo que se requiere de tecnologías altamente sensibles capaces de extraer información molecular relevante a partir de cantidades bajas de DNA (Alba-Bernal et al., 2020). En la actualidad, las técnicas más utilizadas para el análisis de ctDNA se basan en la PCR, por ejemplo, la PCR digital, o en las tecnologías NGS (De Rubis et al., 2019). La secuenciación con nanoporos emerge como una tecnología prometedora debido a su bajo requisito de material genético de partida, lo que la convierte en una alternativa factible para el análisis de muestras con baja concentración de ctDNA (Oehler et al., 2023).

2 ANTECEDENTES

El antecedente principal de esta investigación es el trabajo desarrollado por Martignano et al. (2021), el cual fue el primero reportado en la literatura en abordar la identificación de CNAs a nivel de genoma completo, a partir de cfDNA, mediante la tecnología de secuenciación genómica con nanoporos. Dado que los protocolos de preparación de librerías específicos de esta tecnología están diseñados para el enriquecimiento y la secuenciación de lecturas largas, una de las aportaciones de su estudio es la modificación de la metodología para su aplicación en la secuenciación de moléculas cortas de cfDNA. A partir de dichas modificaciones, el flujo de trabajo estandarizado fue implementado en muestras de plasma de pacientes con cáncer de pulmón en estadio metastásico, con el objetivo de detectar CNAs asociadas a dicho tipo tumoral, empleando el enfoque de sWGS. Para llevar a cabo la detección de CNAs, emplearon la herramienta NanoGLADIATOR (Magi et al., 2019), la cual es específica para el análisis de este tipo de alteraciones a partir de datos generados en una plataforma de secuenciación con nanoporos. La principal aportación de su estudio es demostrar que el análisis de biopsias líquidas mediante la tecnología de secuenciación genómica con nanoporos es factible para generar cariotipos moleculares, útiles como herramienta en el monitoreo del cáncer. Además, demuestran que la información de CNAs obtenida mediante este abordaje permite generar resultados comparables con los obtenidos mediante la tecnología de secuenciación más utilizada en la actualidad, Illumina.

En nuestro laboratorio, el Laboratorio de Genómica del Cáncer (LGC), se han realizado trabajos previos en los que se aborda el estudio de las CNAs en pacientes con cáncer de mama. En 2011, como parte de la línea de investigación “Análisis de alta resolución de alteraciones somáticas en el genoma y transcriptoma de tumores de mama en pacientes mexicanas”, se realizó el trabajo titulado “Perfiles de alteraciones cromosómicas en cáncer de mama de pacientes mexicanas mediante análisis con microarreglos de alta resolución”. En este trabajo, se analizaron muestras pareadas de tejido normal y tejido tumoral proveniente de 100 pacientes con cáncer de mama mediante la tecnología de microarreglos de genotipificación, con el objetivo de identificar las CNAs más prevalentes en la población mexicana asociadas a cáncer de mama. Los resultados de este trabajo forman parte de un estudio reciente (Romero-Cordoba et al., 2021), el cual representa un esfuerzo por caracterizar molecularmente los tumores de mama en poblaciones hispano-mexicanas para comprender mejor la heterogeneidad asociada a este tipo tumoral y dar mayor representación de esta poblaciones dentro de las bases de datos públicas. Un hallazgo relevante de este trabajo fue la identificación de CNAs presentes en pacientes con cáncer de mama de ancestría hispano-mexicana, que no habían sido previamente reportadas en la literatura.

Por ejemplo, se observó la amplificación de las regiones 16p y 17p. Además, al comparar los perfiles de CNAs de estas pacientes con los de pacientes de diferentes ancestrías (caucásica, asiática, afroamericana), se observó que algunas amplificaciones y deleciones frecuentemente presentes en estas poblaciones no estaban presentes en las mujeres hispano-mexicanas.

Actualmente, el LGC trabaja en una nueva línea de investigación titulada “Análisis de DNA tumoral libre en circulación para la identificación de alteraciones clínicamente accionables en pacientes con cáncer de mama”, de la cual forma parte el presente trabajo. Dentro de ella, también se incluye el proyecto “Implementación de una prueba de biopsia líquida para el manejo del cáncer de mama en México”. Este permitió la ejecución exitosa del análisis de cfDNA en muestras plasmáticas de pacientes con tumores de mama, a través de un panel de secuenciación dirigida compuesto por 12 genes de relevancia clínica en el cáncer de mama. Dicho panel fue diseñado con el propósito de identificar mutaciones y CNAs clínicamente relevantes. Para ello, se empleó una plataforma de secuenciación de la tecnología Ion Torrent, la cual pertenece a las tecnologías de NGS.

Hasta la fecha, no se han reportado trabajos en tumores de mama en los que se integre la tecnología de secuenciación con nanoporos como parte del flujo de trabajo para el análisis de cfDNA con el propósito de identificar CNAs clínicamente relevantes. Aunque algunos estudios han analizado el perfil de alteraciones cromosómicas obtenido del análisis del cfDNA de pacientes con cáncer de mama, tanto en estadios no metastásicos como metastásicos, los procesos de secuenciación se han realizado utilizando las tecnologías de NGS tradicionales (Ambriz-Barrera et al., 2023; Kim et al., 2023; Prat et al., 2023; Stover et al., 2018).

Bajo este contexto, el presente proyecto pretende explorar la factibilidad y utilidad de la tecnología de secuenciación con nanoporos como una metodología alternativa para obtener información con potencial clínico, centrándose particularmente en el análisis detallado de CNAs, en muestras de biopsia líquida de pacientes con cáncer de mama.

3 JUSTIFICACIÓN

La realización de este proyecto se centra en desarrollar y validar una metodología efectiva para detectar CNAs en pacientes con cáncer de mama. Dado que la detección de CNAs es relevante para identificar aquellas regiones genómicas cuya alteración juega un papel fundamental en el inicio, desarrollo y progresión del cáncer, existe una demanda creciente de herramientas precisas y novedosas para su detección en biopsias líquidas. Este proyecto busca aprovechar los avances tecnológicos, como la secuenciación con nanoporos, y la experiencia previa del Laboratorio de Genómica del Cáncer (LGC) del INMEGEN en el análisis de CNAs para ofrecer una metodología alternativa y comparable a las tecnologías tradicionales, como la secuenciación Illumina y los microarreglos de alta resolución. Su realización permitirá avanzar en el conocimiento y la aplicación de la tecnología de secuenciación con nanoporos en el contexto de la detección de CNAs clínicamente relevantes en cáncer de mama.

4 PREGUNTA DE INVESTIGACIÓN

¿La secuenciación genómica con nanoporos en biopsias líquidas de cáncer de mama permite identificar CNAs de forma comparable a los microarreglos de alta resolución en biopsias de tejido tumoral pareadas?

5 HIPÓTESIS

La implementación de la tecnología de secuenciación genómica con nanoporos en el análisis de biopsias líquidas de pacientes con cáncer de mama permitirá identificar CNAs de manera precisa y comparable a los resultados obtenidos mediante microarreglos de alta resolución en biopsias de tejido tumoral pareadas.

6 OBJETIVOS

6.1 Objetivo general

Evaluar el desempeño de la tecnología de secuenciación genómica con nanoporos en la identificación de CNAs en muestras de biopsia líquida de cáncer de mama y comparar los resultados con los obtenidos mediante el análisis de biopsias de tejido tumoral pareadas utilizando microarreglos de alta resolución.

6.2 Objetivos particulares

- Establecer un protocolo estandarizado para la extracción de cfDNA en muestras de biopsia líquida de cáncer de mama para su análisis mediante secuenciación genómica con nanoporos.
- Analizar muestras de líneas celulares tumorales con el objetivo de optimizar el flujo de trabajo de secuenciación genómica con nanoporos para la detección de CNAs.
- Aplicar la tecnología de secuenciación genómica con nanoporos en un conjunto de muestras de biopsia líquida de pacientes con cáncer de mama y analizar los datos resultantes para la detección de CNAs.
- Comparar los resultados de la detección de CNAs obtenidos mediante la secuenciación genómica con nanoporos en biopsias líquidas con los obtenidos a través del análisis de microarreglos de alta resolución en biopsias de tejido tumoral pareadas.

7 ESTRATEGIA EXPERIMENTAL

El presente trabajo representa el primer esfuerzo en el LGC por implementar la tecnología de secuenciación genómica con nanoporos en la detección de alteraciones genómicas relevantes en tumores de mama, tales como las CNAs. Para cumplir los objetivos planteados, este se llevó a cabo en dos fases de acuerdo con la Figura 8.

Previo a responder la pregunta de investigación del trabajo, se llevó a cabo una primera fase en la que se estandarizó el flujo de trabajo en líneas celulares tumorales empleando la tecnología de secuenciación con nanoporos para detectar las CNAs que son representativas de cada una de ellas (Figura 8A). Los objetivos de esta fase fueron: 1) replicar las condiciones de la técnica de secuenciación con nanoporos de fragmentos cortos de DNA para comprobar que se obtiene un rendimiento semejante a lo reportado en la literatura bajo dichas condiciones; además, que el rendimiento, en términos del número de lecturas, fue suficiente para realizar la detección de CNAs; 2) simular condiciones experimentales en líneas celulares tumorales que fueran extrapolables al análisis de moléculas de DNA libre circulante para la detección de CNAs en muestras de biopsia líquida; 3) validar que las regiones genómicas con ganancia y pérdida de material genético detectadas fueran similares a aquellas que se identifican al emplear otras tecnologías genómicas, como la secuenciación Illumina. La finalidad de esta fase fue demostrar, mediante modelos experimentales, que la tecnología de secuenciación con nanoporos permite generar resultados comparables a los obtenidos con tecnologías genómicas que generalmente se emplean para detectar CNAs.

La segunda fase del proyecto consistió en implementar el flujo de trabajo anterior en el análisis de muestras de pacientes, particularmente, muestras de biopsia líquida (Figura 8B). Para ello, se realizó la identificación de CNAs mediante microarreglos de alta resolución en muestras de biopsia de tejido tumoral. Posteriormente, estos resultados se compararon con los generados al analizar la biopsia líquida pareada para cada paciente. Esta fase tiene como finalidad determinar si el análisis de biopsias líquidas mediante la tecnología de secuenciación con nanoporos representa un abordaje alternativo factible al análisis de biopsias de tejido para generar perfiles de CNAs como herramienta diagnóstica, pronóstica y terapéutica en pacientes con cáncer de mama.

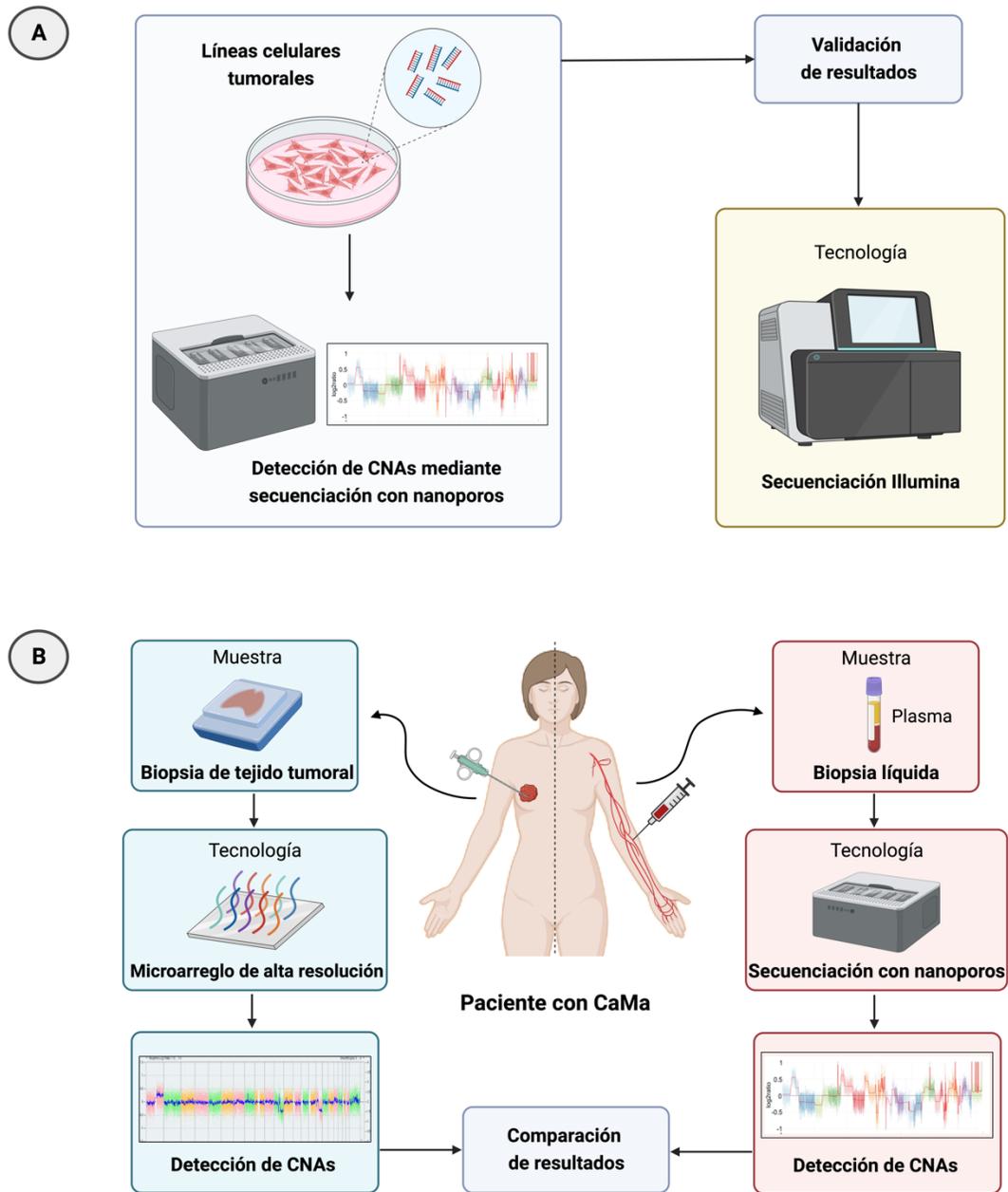


Figura 8. Estrategia experimental del proyecto. El presente trabajo se dividió en dos fases: A) Estandarización de las condiciones óptimas de la secuenciación con nanoporos de fragmentos cortos para la detección de CNAs en líneas celulares tumorales y validación de resultados con la tecnología de secuenciación Illumina. B) Implementación del flujo de trabajo estandarizado en la fase A) para el análisis de muestras de pacientes. Se compararon los resultados obtenidos de la detección de CNAs en biopsias de tejido tumoral mediante microarreglos de alta resolución con los generados en el análisis de biopsias líquidas mediante el flujo de trabajo estandarizado en la fase A), empleando secuenciación con nanoporos. Creada con BioRender.com. CaMa= Cáncer de mama.

8 METODOLOGÍA

8.1 Estandarización de la técnica de secuenciación genómica con nanoporos de fragmentos cortos de DNA

8.1.1 Cultivo de líneas celulares

Se seleccionaron 6 líneas celulares tumorales (Tabla 1), las cuáles fueron adquiridas de ATCC (American Type Culture Collection, Manassas, VA, USA). Las líneas celulares se cultivaron de acuerdo con las guías de ATCC, en medio suplementado con 10% de suero fetal bovino. Las células se mantuvieron incubadas a 37 °C en una atmósfera humidificada con 5% de CO₂.

Tabla 1. Selección de líneas celulares tumorales y sus características de origen.

Línea celular	Tejido de origen	Tipo de tumor	Estadio	Clasificación por subtipo molecular*
HeLa	Cérvix uterino de <i>H. sapiens</i>	Adenocarcinoma	Metastásico	No aplica
MCF7	Glándula mamaria de <i>H. sapiens</i>	Adenocarcinoma	Metastásico	Luminal A
MDA-MB-361	Glándula mamaria de <i>H. sapiens</i>	Adenocarcinoma	Metastásico	Luminal B
SK-BR-3	Glándula mamaria de <i>H. sapiens</i>	Adenocarcinoma	Metastásico	HER2 enriquecido
MDA-MB-231	Glándula mamaria de <i>H. sapiens</i>	Adenocarcinoma	Metastásico	Basal
Hs578T	Glándula mamaria de <i>H. sapiens</i>	Carcinoma	Primario	Basal

*La clasificación molecular del cáncer de mama estratifica a los tumores en cuatro subtipos principales: Luminal A, Luminal B, HER2 enriquecido y Basal, basándose en sus patrones de expresión génica. A nivel clínico, estos subtipos se distinguen por un diferente pronóstico (Harbeck et al., 2019).

8.1.2 Extracción, cuantificación y fragmentación de DNA genómico

Partiendo de 2-4 x 10⁶ células, se realizó la extracción de DNA genómico de cada línea celular empleando el kit de trabajo AllPrep® DNA/RNA Mini Kit (Qiagen® Inc, Valencia, CA), de acuerdo con las instrucciones del fabricante. La concentración de DNA genómico obtenido se

determinó mediante fluorimetría con el ensayo Qubit dsDNA High Sensitivity (ThermoFisher Scientific, Waltham, Massachusetts, USA), utilizando el instrumento Qubit 3.0 Fluorometer (Life Technologies, Thermo Fisher Scientific, Waltham, Massachusetts, USA); se evaluó la calidad del DNA extraído mediante espectrofotometría utilizando el instrumento NanoDrop™ One (ThermoFisher Scientific, Waltham, Massachusetts, USA) a través de los criterios de pureza dados por los ratios de absorbancia 260/280 y 260/230, útiles para detectar contaminación por proteína, y por sales y compuestos orgánicos, respectivamente.

Posteriormente, se realizó la fragmentación de 500 ng de DNA genómico por cada línea celular utilizando el equipo Bioruptor Pico (Diagenode, Belgium) durante 6 ciclos (30" ON; 90" OFF) a 4°C. El objetivo de este procedimiento fue simular condiciones experimentales en líneas celulares tumorales, con la finalidad de extrapolar el procedimiento al análisis de moléculas de DNA libre circulante. La cuantificación y la distribución de los fragmentos se evaluó mediante electroforesis automatizada con el ensayo Agilent High Sensitivity D1000 ScreenTape (Agilent Technologies, Santa Clara, CA) en el instrumento 4200 TapeStation (Agilent Technologies, Santa Clara, CA), obteniéndose fragmentos de aproximadamente 350 pb.

8.1.3 Preparación de librerías para secuenciación

La preparación de librerías específicas de la secuenciación con nanoporos se conforma de dos pasos principales: 1) la reparación de los extremos de los fragmentos de DNA con la posterior adición de la cola A en los extremos 3' y 2) la ligación de los adaptadores. Para ello se procesaron 200 fmol de DNA fragmentado con el kit de trabajo NEBNext® Companion Module for Oxford Nanopore Technologies® Ligation Sequencing (New England Biolabs, Ipswich, MA, United States) de acuerdo con las instrucciones del fabricante. La ligación de adaptadores se realizó siguiendo el protocolo SQK-LSK109 (Oxford Nanopore Technologies, United Kingdom). En el caso de la secuenciación con nanoporos múltiplex, el DNA se procesó como se describió anteriormente, añadiendo un paso intermedio en el que se ligan fragmentos cortos de DNA de secuencia conocida (*barcodes*) a los fragmentos previamente procesados, de acuerdo con el protocolo EXP-NBD104 (Oxford Nanopore Technologies, United Kingdom). Posterior al proceso de *barcoding*, se realizó un pool de las muestras y se purificó con una proporción de 2.5X de perlas magnéticas AMPure XP beads (Beckman Coulter, Brea, CA, USA). Para llevar a cabo la secuenciación, se cargaron 45 fmol de cada librería en celdas de flujo FLO-MIN106 (versión R9.4.1) utilizando la plataforma de secuenciación GridION X5 (Oxford Nanopore Technologies, United Kingdom).

8.2 Análisis bioinformático

8.2.1 Procesamiento de datos crudos de secuenciación

El llamado de bases se realizó en tiempo real empleando la herramienta Guppy, que forma parte del software MinKNOW, en el modo *High Accuracy basecalling* (HAC). Esta herramienta genera los archivos .fastq, realiza el corte de los adaptadores y la demultiplexación de las muestras en el caso de corridas múltiplex. La herramienta clasifica las lecturas crudas en aquellas que superan el valor umbral de Qscore, el cual se definió como 9.0 en todos los casos, y las que no lo alcanzan.

8.2.2 Alineamiento de lecturas al genoma de referencia

A partir de las lecturas que cumplen este control de calidad, se realizó el alineamiento de las lecturas sin procesar al genoma de referencia hg19 con el algoritmo BWA-MEM. Empleando la herramienta SAMtools se realizó la conversión de archivos .sam a .bam, el ordenamiento de los lecturas alineadas con respecto a la posición genómica y su indexación.

8.2.3 Detección de alteraciones en el números de copias de DNA

Con la finalidad de realizar la detección de CNAs en los datos generados mediante secuenciación genómica con nanoporos, se empleó el software NanoGLADIATOR (descargado en <https://sourceforge.net/projects/nanogladiator/>). NanoGLADIATOR permite detectar este tipo de alteraciones mediante el abordaje del conteo de lecturas; a partir de las regiones con incremento o disminución del conteo de lecturas en ventanas genómicas de tamaño definido se pueden detectar aquellas regiones del genoma con cambios en el número de copias.

El análisis se corrió en la modalidad “Off-line”, la cual permite ingresar varios archivos .bam para analizarlos simultáneamente dentro del mismo experimento. Esta modalidad se ejecuta mediante tres módulos o scripts de Perl, descritos brevemente a continuación.

1. ReferenceWindowInitialize.pl

Este script genera información sobre el contenido GC y valores de mapeabilidad para ventanas genómicas de tamaño específico en el genoma de referencia indicado por el usuario. Para su funcionamiento, el script requiere de 5 argumentos conforme al siguiente ejemplo:

```
perl ReferenceWindowInitialize.pl -M genome -S SourceTarget.txt -T
LabelName -W 50000 -A hg19
```

donde:

-M: Modalidad para crear información sobre el contenido GC y valores de mapeabilidad

-S: Ruta al archivo de texto delimitado por espacios cuyo contenido incluye dos columnas: 1) ruta absoluta al archivo bigWig (.bw) para el cálculo de mapeabilidad; 2) ruta absoluta al archivo .fasta del genoma de referencia definido para el cálculo del contenido GC

-T: Nombre del análisis

-W: Tamaño de las ventanas genómicas

-A: Ensamblaje del genoma de referencia (hg19/hg38)

Siguiendo lo reportado por Magi et al. (2019), para este análisis se seleccionó 100 kpb como tamaño de ventana genómica.

2. NanoGLADIATORPrepare.pl

Este script realiza el cálculo del conteo de lecturas en cada ventana genómica y la normalización del conteo de lecturas para corregir sesgos asociados al contenido GC y mapeabilidad. El comando requiere un argumento y tres parámetros de acuerdo con el siguiente ejemplo:

```
perl NanoGLADIATORPrepare.pl ExperimentalFilePrepare.txt --processors
6 --target LabelName --assembly hg19
```

donde:

ExperimentalFilePrepare.txt: Nombre del archivo de texto delimitado por espacios cuyo contenido incluye tres columnas: 1) ruta absoluta al archivo .bam a analizar; 2) ruta al directorio de salida en el que se almacenarán los resultados; 3) nombre de la muestra analizada. En este archivo, cada fila comprende la información de una muestra diferente.

--processors: Número de procesadores a emplear durante el análisis

--target: Nombre del análisis empleando en el script 1) en el argumento -T

--assembly: Ensamblaje del genoma de referencia empleado (hg19/hg38)

3. NanoGLADIATORAnalysis.pl

Empleando el algoritmo Shifting Level Model (SLM), este script ejecuta la segmentación de los datos de acuerdo con el aumento o disminución del número de lecturas detectado a través de las diversas ventanas genómicas. Adicionalmente, usando el algoritmo FractionPred, se lleva a cabo la predicción de la fracción alélica para cada alteración genómica detectada. Para ello, el comando requiere de un argumento y cinco parámetros de acuerdo con el siguiente ejemplo:

```
perl NanoGLADIATORAnalysis.pl ExperimentalFileAnalysis.txt --processors
1 --target LabelName --assembly hg19 --output PathWhereTosaveAnalyses
--mode nocontrol
```

donde:

ExperimentalFileAnalysis.txt: Nombre del archivo de texto delimitado por espacios cuyo contenido incluye tres columnas: 1) etiqueta que identifica la muestra, tomando en consideración si el análisis se llevará a cabo de forma pareada o no pareada (ej. T1, T2, C1, C2; donde T= tumor, C= control, n=número que identifica cada muestra a analizar); 2) ruta tal cual se indicó en el inciso 2) del archivo *ExperimentalFilePrepare.txt*; 3) nombre de la muestra tal cual se indicó en el inciso 3) del archivo *ExperimentalFilePrepare.txt*. Cada fila muestra la información de una muestra distinta.

--processors: Número de procesadores a emplear durante el análisis

--target: Nombre del análisis empleando en el script 1) en el argumento -T

--assembly: Ensamblaje del genoma de referencia empleado (hg19/hg38)

--output: Ruta al directorio de salida en el que se almacenarán los resultados

--mode: Modo en el que se analizarán las muestras (nocontrol/paired)

En este caso, el análisis se realizó en el modo “no control”, dado que no se cuenta con muestras control de individuos sanos.

8.2.3.1 Archivos de salida

La herramienta genera como resultado dos archivos de salida principales:

- **Reporte de texto de alteraciones en el número de copias de DNA detectadas**

Archivo de texto delimitado por tabulaciones, el cual incluye nueve columnas con información para cada región segmentada: 1) cromosoma; 2) posición de inicio (pb); 3) posición de fin (pb); 4) número de ventanas genómicas en el segmento; 5) estado de número de copias del segmento

(doble delección, delección, duplicación o duplicación de múltiples copias); 6) número de copias del segmento; 7) fracción alélica; 8) error de la fracción alélica; 9) mediana del valor \log_2 ratio del segmento.

- **Cariotipo molecular**

Archivo HTML interactivo, que muestra gráficamente el estado de número de copias de las regiones genómicas segmentadas con base en los resultados de todos los análisis previos: datos de \log_2 ratio, resultados de segmentación y de predicción de la fracción alélica.

8.3 Validación de los resultados obtenidos en líneas celulares tumorales

En esta sección se realizó la comparación de los resultados obtenidos mediante la tecnología de secuenciación con nanoporos con los generados mediante dos tecnologías genómicas que son rutinariamente empleadas para la detección de CNAs.

8.3.1 Validación de los resultados mediante datos de secuenciación Illumina

La validación de resultados mediante la tecnología de secuenciación Illumina se realizó a partir de los datos públicos reportados por el proyecto de caracterización genómica de la *Cancer Cell Line Encyclopedia* (CCLE) (Ghandi et al., 2019). Los datos crudos de secuenciación se descargaron de la base de datos Sequence Read Archive (SRA) mediante las claves de acceso mostradas en la Tabla 2 (clave de acceso del proyecto: PRJNA523380).

Tabla 2. Claves de acceso de los datos de secuenciación de las líneas celulares tumorales.

Línea celular	Clave SRA	ID de corrida	Cobertura	Tipo	Plataforma de secuenciación	Profundidad de cobertura
MCF7	SRX5449795	SRR8652105	WGS	Paired-end	Illumina HiSeq 2000	36x
MDA-MB-361	SRX5449775	SRR8652125	WGS	Paired-end	Illumina HiSeq 2000	37x
SK-BR-3	SRX5455745	SRR8657720	Targeted-capture sequencing	Paired-end	Illumina HiSeq 2000	1x
MDA-MB-231	SRX5449793	SRR8652107	WGS	Paired-end	Illumina HiSeq 2000	36x

WGS= Secuenciación a nivel de genoma completo; Targeted-capture sequencing: Secuenciación dirigida; Paired-end: Secuenciación de ambos extremos.

8.3.1.1 Análisis bioinformático

De acuerdo con Martignano et al. (2021), los resultados de detección de CNAs obtenidos mediante el análisis de lecturas que provienen de experimentos *single-end* y *paired-end* están altamente correlacionados. Siguiendo esto, únicamente se analizaron las lecturas R1 de los datos descargados con la finalidad de simplificar los pasos posteriores del análisis y la comparación de resultados entre las tecnologías de secuenciación.

Se alinearon las lecturas al genoma de referencia hg19 con el algoritmo BWA-MEM. Posteriormente, la conversión de archivos .sam a .bam, el ordenamiento de las lecturas alineadas con respecto a la posición genómica y su indexación se llevó a cabo utilizando la herramienta SAMtools. La detección de CNAs se realizó con XCAVATOR (descargado en <https://sourceforge.net/projects/xcavator/>); este es un software precedente a NanoGLADIATOR, desarrollado para la detección de CNAs en datos generados mediante plataformas de secuenciación NGS. Al igual que NanoGLADIATOR, XCAVATOR utiliza el abordaje del conteo de lecturas y realiza la segmentación de los datos empleando el algoritmo SLM (Magi et al., 2017). El análisis se ejecutó a través de tres pasos o scripts de Perl: 1) *ReferenceWindowInitialize.pl*, 2) *XCAVATORDataPrepare.pl* y 3) *XCAVATORDataAnalysis.pl*; estos siguen un funcionamiento semejante a lo descrito para el análisis con NanoGLADIATOR (apartado 10.2.2.). Se seleccionó 100 kpb como tamaño de ventana genómica y el modo “no control”.

8.3.1.2 Análisis de correlación de datos de segmentación entre tecnologías de secuenciación

Utilizando un script de R, se compararon los resultados de segmentación obtenidos por ambas herramientas: NanoGLADIATOR (para datos de secuenciación con nanoporos) y XCAVATOR (para datos de secuenciación Illumina). A partir de los reportes de segmentación generados por cada herramienta (*HSLM Results*), se extrajeron los valores promedio de segmentación (\log_2 ratio) para cada ventana genómica y se realizó el análisis de correlación entre estos datos (función= *stat_cor*, paquete= *ggpubr*, método= “*spearman*”). Se consideró como estadísticamente significativo el valor $p < 0.05$. Para visualizar la correlación, se realizaron diagramas de dispersión.

Aún cuando en ambos casos se seleccionó el mismo tamaño de ventana genómica (100 kpb), las posiciones físicas de las ventanas genómicas mostraron variaciones entre sí. Por tal motivo, para el análisis de correlación únicamente se consideraron las ventanas genómicas que tuvieran correspondencia entre las dos herramientas, siguiendo el siguiente supuesto: una ventana genómica de XCAVATOR corresponde a una ventana genómica de NanoGLADIATOR cuando su posición de inicio se encuentra entre la posición de inicio y la posición de fin de una ventana genómica de NanoGLADIATOR. El número total de ventanas genómicas fue de 26,927 y 28,451, para la tecnología de secuenciación con nanoporos y la tecnología de secuenciación Illumina, respectivamente. De estas, se seleccionaron 26,881 ventanas genómicas para realizar la comparación entre tecnologías (99.82% y 94.5% de ventanas genómicas correspondientes).

Posteriormente, empleando un script de R, se realizaron perfiles de segmentación para cada uno de los cromosomas del genoma humano, de acuerdo con los datos de segmentación de las dos tecnologías de secuenciación.

8.3.1.3 Comparación de regiones genómicas alteradas a nivel citogenético entre tecnologías de secuenciación

En esta sección se comparó, a nivel citogenético, el tamaño y el número de copias de las regiones genómicas con alteraciones cromosómicas detectadas en las líneas celulares tumorales mediante las dos tecnologías de secuenciación. Para relacionar la posición física con la posición citogenética de las regiones genómicas, se descargó el archivo de citobandas (hg19) del navegador genómico de la Universidad de California de Santa Cruz (UCSC Genome Browser, <https://genome.ucsc.edu/index.html>), el cual describe las posiciones físicas de las bandas citogenéticas en cada cromosoma. Mediante un script de python (librería= *pandas*), se utilizaron los datos de localización física (posición de inicio y posición de fin) de los reportes de CNAs generados por ambas herramientas para identificar sus bandas citogenéticas correspondientes. El cálculo del tamaño de la región genómica, expresado en kpb, se realizó siguiendo la siguiente fórmula:

$$\text{Tamaño (kpb)} = \text{posición de fin (pb)} - \text{posición de inicio (pb)} * \frac{1 \text{ kpb}}{1000 \text{ pb}}$$

Adicionalmente, en ambos conjuntos de datos, se realizó un filtrado de los segmentos utilizando un valor de $\log_2\text{ratio}$ más restrictivo para evitar resultados falsos positivos. Las regiones genómicas se definieron como alteradas, en cuanto al número de copias de DNA, cuando cumplían con los siguientes valores de corte:

$\log_2\text{ratio} < -0.40$: Deleción de una o ambas copias

$\log_2\text{ratio} > +0.40$: Duplicación de una o múltiples copias

8.4 Implementación del flujo de trabajo estandarizado en el análisis de muestras de biopsia líquida

8.4.1 Selección de muestras de pacientes

En esta fase del trabajo, se llevó a cabo la recolección de muestras biológicas procedentes de pacientes diagnosticadas con cáncer de mama. Estas pacientes fueron atendidas en la Fundación Mexicana de Fomento para la Prevención Oportuna del Cáncer de Mama A.C. (FUCAM), y se unieron al protocolo de investigación de manera voluntaria. Cada participante en el protocolo proporcionó dos tipos de muestras biológicas: una muestra de sangre periférica y una muestra de tejido tumoral. Cabe destacar que todas las muestras obtenidas en este proceso cuentan con un seguimiento clínico riguroso. Además, todas las pacientes brindaron su consentimiento informado de manera voluntaria y previa a su inclusión en el proyecto, el cual cuenta con la aprobación del Comité de Ética del INMEGEN, bajo el número de registro CE2009/11. Esta fase de selección y obtención de muestras se realizó en estricto cumplimiento de los estándares éticos y normativos establecidos para la investigación médica y genómica.

La selección de pacientes se realizó siguiendo los criterios de inclusión, exclusión y eliminación que se detallan a continuación.

8.4.1.1 Criterios de inclusión, exclusión y eliminación

Criterios de inclusión

- Pacientes con diagnóstico histológico confirmado de cáncer de mama
- Pacientes en todas las etapas clínicas (desde la etapa I hasta la IV)
- Pacientes que cuentan con muestra de tejido tumoral de diagnóstico
- Pacientes que hayan proporcionado su consentimiento informado por escrito

- Pacientes que acuden por primera vez al FUCAM y no hayan recibido tratamiento previo

Criterios de exclusión

- Pacientes con información clínica incompleta

Criterios de eliminación

- Pacientes que expresen su deseo de abandonar el protocolo en cualquier etapa del proceso
- La paciente no dispone del material biológico suficiente para llevar a cabo los estudios moleculares planificados

8.4.2 Obtención y manejo de muestras

8.4.2.1 Obtención y manejo de muestras de sangre total

Previo a someterse al procedimiento quirúrgico de mastectomía, se obtuvieron 12 mL de sangre periférica de cada paciente participante en el protocolo de investigación. Las muestras de sangre venosa fueron colectadas en tubos BD Vacutainer® con EDTA K2, mezcladas por inversión lenta y procesadas en las primeras dos horas siguientes a la venopunción. Como primer paso, las muestras se centrifugaron a 1000 g durante 10 minutos a 4 °C con desaceleración lenta. Después de obtener las fases separadas, se transfirió cuidadosamente la fracción plasmática a un nuevo tubo, evitando succionar la fracción leucocitaria. Posteriormente, el plasma se centrifugó a 2000 g durante 10 minutos a 4 °C y se transfirió el líquido sobrenadante a cuantos tubos de 2.0 mL fuesen necesarios, sin tocar el botón formado. El plasma aislado se almacenó a -80 °C hasta su uso. A partir de un volumen inicial de sangre total de 12 mL, se obtuvieron aproximadamente entre 5 y 6 mL de plasma.

8.4.2.2 Obtención y manejo de muestras de tejido tumoral

Después de la realización del procedimiento quirúrgico de mastectomía, se colectó la muestra de tejido tumoral de cada paciente participante en el protocolo de investigación. Para ello, se llevó a cabo la inspección macroscópica de la pieza quirúrgica en colaboración con un médico patólogo. El objetivo era identificar una región con un elevado contenido de células neoplásicas a partir de la cual poder colectar la muestra de tejido tumoral. La sección de tejido seleccionada se fijó en una solución de formaldehído al 4% durante 12 horas. Después, el tejido

se lavó en PBS durante 30 minutos, seguido de una deshidratación con concentraciones crecientes de etanol. Finalmente, se sumergió en xilol y se incubó a 60 °C durante 30 minutos. El tejido se incluyó en parafina durante 2 horas y se colocó en casetes específicos para su posterior análisis.

8.4.3 Detección de alteraciones en el número de copias en biopsias de tejido tumoral mediante microarreglos de alta resolución

8.4.3.1 Extracción y cuantificación de DNA genómico

Previo a realizar la extracción de DNA genómico, se llevó a cabo la desparafinación de los cortes histológicos siguiendo el siguiente procedimiento: 1) Incubación de las laminillas en xileno durante 5 minutos; 2) incubación en etanol al 100% durante 3 minutos; 3) incubación en etanol al 95% durante 3 minutos; 4) adición de agua desionizada durante 30 segundos. Únicamente se procesaron para la extracción de ácidos nucleicos aquellas muestras con un contenido tumoral superior al 60% y sin signos de necrosis.

Posteriormente, se realizó la extracción de material genético empleando el kit de trabajo AllPrep® DNA/RNA FFPE Kit (Qiagen® Inc, Valencia, CA) de acuerdo a las instrucciones del fabricante. La concentración de DNA genómico aislado se cuantificó mediante espectrofotometría, utilizando el instrumento NanoDrop™ One (ThermoFisher™ Scientific, Waltham, Massachusetts, USA). Se evaluó la calidad a través de los criterios de pureza dados por los ratios de absorbancia 260/280 y 260/230 para detectar contaminación por proteínas y por sales y compuestos orgánicos, respectivamente. Posteriormente, se almacenó a -80 °C hasta su procesamiento.

8.4.3.2 Procesamiento de muestras e hibridación al microarreglo de alta resolución

Para obtener información sobre el estado de número de copias de DNA, el DNA genómico aislado se analizó mediante el microarreglo de alta resolución OncoScan™ CNV Plus Assay (Affymetrix, Santa Clara, USA). Este microarreglo permite la detección de CNAs a nivel de genoma completo a partir del análisis de tejidos FFPE. El ensayo se llevó a cabo en la Unidad de Microarreglos del INMEGEN, empleando 80 ng de DNA genómico; el material de partida se procesó conforme al protocolo del fabricante.

8.4.3.3 Análisis de datos

Los datos obtenidos del escaneo de intensidad de fluorescencia en el microarreglo (archivos .cel), generados mediante el software Affymetrix® GeneChip® Command Console®, se analizaron mediante el software Chromosome Analysis Suite versión 4.3.0.7.1 (Affymetrix, Santa Clara, CA). Este consiste en dos módulos:

1. Analysis Workflow Module

Este módulo realiza el procesamiento de los archivos .cel para su normalización y corrección; después, utiliza algoritmos de segmentación para identificar aquellas regiones genómicas que presentan ganancias o pérdidas en el número de copias de DNA, en comparación con un panel de referencia compuesto por 400 muestras normales. Estos resultados son almacenados en archivos .oschp. Además de esto, el módulo genera un reporte de calidad que incorpora diversos parámetros, siendo el MAPD y WavinessSD los más relevantes para evaluar el rendimiento del ensayo, particularmente en lo que respecta al análisis de CNAs.

2. Browser Module

Los archivos .oschp son exportados a este módulo, el cual facilita la exploración detallada de los resultados mediante la visualización interactiva de los segmentos con CNAs. El software genera reportes que resumen las características genómicas de las CNAs detectadas, ofreciendo información acerca de su ubicación, tamaño, estado de número de copias, genes contenidos en estas regiones y su relevancia clínica.

8.4.4 Detección de alteraciones en el número de copias en biopsias líquidas mediante secuenciación con nanoporos

8.4.4.1 Extracción y cuantificación de DNA libre circulante

Se realizó la extracción del cfDNA a partir del plasma aislado de las muestras de sangre total empleando el kit de trabajo MagMAX™ Cell-Free DNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA, USA) de acuerdo con las instrucciones del fabricante. Para ello, se utilizó el instrumento KingFisher™ Duo Prime Magnetic Particle Processor (Thermo Fisher Scientific, Waltham, MA, USA) para ejecutar la extracción mediante un proceso automatizado.

Posteriormente, se procedió a realizar un análisis de calidad que permitiera la cuantificación y evaluación del tamaño de los fragmentos de DNA aislados. Así mismo, se determinó el porcentaje de cfDNA presente en las muestras con respecto al total del material

genético extraído. Para llevar a cabo esta evaluación, se empleó el ensayo Cell-free DNA ScreenTape Assay (Agilent Technologies, Santa Clara, CA, USA) y el bioanalizador Tape Station 4200 (Agilent Technologies, Santa Clara, CA, USA), ejecutándolo conforme a las indicaciones detalladas en el protocolo.

8.4.4.2 Preparación de librerías para secuenciación y análisis bioinformático

La preparación de librerías de secuenciación y el análisis bioinformático se realizaron conforme a lo descrito en las secciones 8.1.3 y 8.2 (que incluye las subsecciones 8.2.1, 8.2.2 y 8.2.3), respectivamente. Durante la preparación de librerías, se emplearon 15 ng de cfDNA como material de partida y se siguieron las modificaciones del protocolo estandarizadas por Martignano et al. (2021) para adaptarlo al análisis de muestras de biopsia líquida. Una de las modificaciones más importantes fue el aumento en la proporción de perlas magnéticas usadas durante los diferentes pasos de purificación a 1.8x.

8.4.5 Comparación de regiones genómicas alteradas a nivel citogenético entre ambos tipos de biopsias

Con el propósito de evaluar la concordancia entre las biopsias de tejido tumoral y las biopsias líquidas, se comparó la localización citogenética, el tamaño y el número de copias de las regiones genómicas con alteraciones cromosómicas detectadas en las muestras pareadas de cada paciente, de forma semejante a lo detallado en la sección 8.3.1.3.

9 RESULTADOS Y DISCUSIÓN

9.1 Rendimiento y estadísticas de la secuenciación genómica con nanoporos de fragmentos cortos en líneas celulares tumorales

Los protocolos de secuenciación genómica con nanoporos se han optimizado principalmente para analizar moléculas largas de ácidos nucleicos. Sin embargo, el objetivo principal de este trabajo se centra en la detección de CNAs en muestras de biopsias líquidas de pacientes con cáncer de mama. De este modo, este enfoque pretende explorar y validar una metodología eficaz para detectar estas alteraciones en fragmentos cortos de DNA, aprovechando las numerosas ventajas que ofrece la tecnología de secuenciación con nanoporos para su potencial aplicación clínica en el diagnóstico y seguimiento de la enfermedad. Por lo tanto, en esta sección inicial, se ha aplicado una metodología modificada, basada en el trabajo de Martignano et al. (2021), para analizar fragmentos cortos de DNA provenientes de líneas celulares tumorales con cariotipos previamente caracterizados en la literatura. Esto permitirá la comparación de nuestros resultados con datos previamente reportados en la literatura.

Dado que es el primer esfuerzo del LGC en el uso de esta tecnología de secuenciación para la detección de CNAs, se evaluó el rendimiento siguiendo la metodología. Esto se realizó mediante dos experimentos: 1) una corrida singleplex (secuenciación de una sola muestra en el experimento) y 2) una corrida múltiplex (secuenciación de dos o más muestras en el mismo experimento). La secuenciación de las librerías de líneas celulares tumorales produjo un rendimiento de 14,122,691 y 11,713,790 lecturas crudas para los experimentos singleplex (24 horas) y múltiplex (40 horas), respectivamente. De forma particular, en el experimento múltiplex se generó un rendimiento de lecturas crudas semejante entre las muestras y que, en todos los casos, fue mayor a dos millones de lecturas. El Qscore promedio fue igual o mayor a 12.0 en todas las muestras secuenciadas, valor que es superior al umbral definido en los experimentos ($Qscore_{umbral} = 9.0$) (Tabla 3).

Con base en el valor de Qscore, el 83% (11,713,790) de las lecturas del experimento singleplex cumplieron con la métrica del control de calidad para poder emplearse en el análisis de detección de CNAs. Por otra parte, en el experimento múltiplex, el 78% (2,053,916), 76% (2,069,406), 80% (1,616,190), 79% (1,709,931) y 77% (1,684,347) de las lecturas de las líneas celulares MCF7, MDA-MB-361, SK-BR-3, MDA-MB-231 y Hs578T, respectivamente, superaron el umbral de calidad (Tabla 3).

La mediana de la longitud de las lecturas secuenciadas varió entre 364 y 428 pb (Tabla 3), lo cual es consistente con el tamaño de las moléculas de DNA obtenidas posterior a la fragmentación del DNA genómico.

Tabla 3. Rendimiento y estadísticas de la secuenciación genómica con nanoporos de líneas celulares tumorales realizada en la plataforma de secuenciación GridION X5.

Línea celular	Experimento	Tiempo (h)	Lecturas totales (M)	Lecturas p/muestra (M), (%)	Qscore	Lecturas cumplen QC (M), (%)	Mediana tamaño de lectura (pb)	Profundidad de cobertura
HeLa	Singleplex	24	14.12	14.12 (100.0)	13.7	11.71 (83.0)	428	2x
MCF7	Múltiplex	40	11.71	2.63 (22.5)	12.2	2.05 (78.0)	405	0.3x
MDA-MB-361				2.72 (23.2)	12.0	2.07 (76.0)	420	0.4x
SK-BR-3				2.02 (17.2)	12.3	1.62 (80.2)	396	0.3x
MDA-MB-231				2.16 (18.5)	12.1	1.71 (79.0)	391	0.3x
Hs578T				2.18 (18.6)	12.0	1.68 (77.4)	364	0.2x

M= millones; QC= Control de calidad; pb= pares de bases

En el experimento singleplex, se alcanzó una profundidad de cobertura igual a 2x, mientras que en el experimento múltiplex, cada muestra obtuvo una profundidad de cobertura que osciló entre 0.2x y 0.4x (Tabla 3). Particularmente, la profundidad obtenida en el experimento múltiplex se clasifica como secuenciación a nivel de genoma completo de baja profundidad (sWGS, *shallow Whole Genome Sequencing*), dado que se sitúa por debajo del valor de 1x, representativo de este tipo de abordaje de secuenciación. Esta variación en la profundidad de secuenciación es relevante para la identificación precisa de CNAs a nivel de genoma completo, como lo han sugerido estudios previos (Baslan et al., 2015; Baslan et al., 2021), donde se establece que se necesitan al menos 500,000 lecturas de fragmentos cortos, equivalentes a 0.1x, para una detección precisa de CNAs.

Si bien los resultados mostraron mejor rendimiento y calidad en el experimento singleplex con relación al experimento múltiplex, mediante ambos enfoques se logró obtener un número significativamente alto de lecturas, con una profundidad de secuenciación suficiente, para realizar un análisis de detección de CNAs de alta resolución. Esto sugiere que este abordaje metodológico

tiene potencial aplicación en condiciones clínicas donde se requiere un procesamiento simultáneo de múltiples muestras, lo que simplificaría el análisis genómico e impactaría en el diagnóstico y seguimiento de algunas enfermedades, como el cáncer de mama.

Adicionalmente, la comparación de los datos de secuenciación generados en este trabajo con los reportados por Baslan et al. (2021) revela una diferencia notable en el rendimiento de lecturas obtenido. Mientras que en su investigación se obtuvieron aproximadamente 6 millones de lecturas, en el presente trabajo se lograron más de 14 millones en la corrida singleplex y más de 11 millones en la corrida múltiplex, considerando condiciones de secuenciación semejantes y un tamaño de fragmento similar, alrededor de 400-500 pb. Así mismo, nuestros datos mostraron un valor de Qscore promedio más elevado en todas las muestras (Qscore reportado por Baslan et al. (2021) =10.5). El valor de Qscore, ó *phred score*, es una medida de la calidad que indica la probabilidad de que una base haya sido llamada incorrectamente. En nuestros datos, el valor mínimo promedio de Qscore reportado fue de 12.0, lo que significa que existe una probabilidad promedio del 6.3% de que las bases hayan sido llamadas incorrectamente, o una precisión de lectura equivalente a 93.7%.

9.2 Detección de alteraciones en el número de copias de DNA en líneas celulares tumorales mediante secuenciación genómica con nanoporos

A partir de las lecturas que cumplieron con el control de calidad, se realizó el procesamiento de los datos para inferir el número de copias. El algoritmo utilizado relaciona la profundidad de lectura, es decir, la cantidad de moléculas secuenciadas en regiones genómicas específicas, con el número de copias del genoma. Se considera que el número de lecturas correspondiente a un número normal o diploide de copias ($CN= 2$) tiene un valor de $\log_2\text{ratio}$ igual a 0. De esta manera, valores superiores o inferiores en el $\log_2\text{ratio}$ indican una mayor o menor profundidad de lectura en el segmento analizado, lo que refleja ganancias ($CN > 2$) o pérdidas ($CN < 2$) en el número de copias, respectivamente.

Utilizando ventanas genómicas de un tamaño de 100 kpb y el modo “no control” del software NanoGLADIATOR, se generaron seis cariotipos moleculares para cada una de las líneas celulares secuenciadas (Figura 9). Los resultados revelaron que los cromosomas 3, 6, 8, 11, 13, 16, 18 y 20 mostraron alteraciones significativas en el número de copias en estas líneas celulares tumorales. Estos hallazgos son consistentes con la literatura, ya que se ha reportado que estos cromosomas están comúnmente afectados por esta clase de alteraciones en diversos tipos tumorales (Harbers et al., 2021; Knouse et al., 2017).

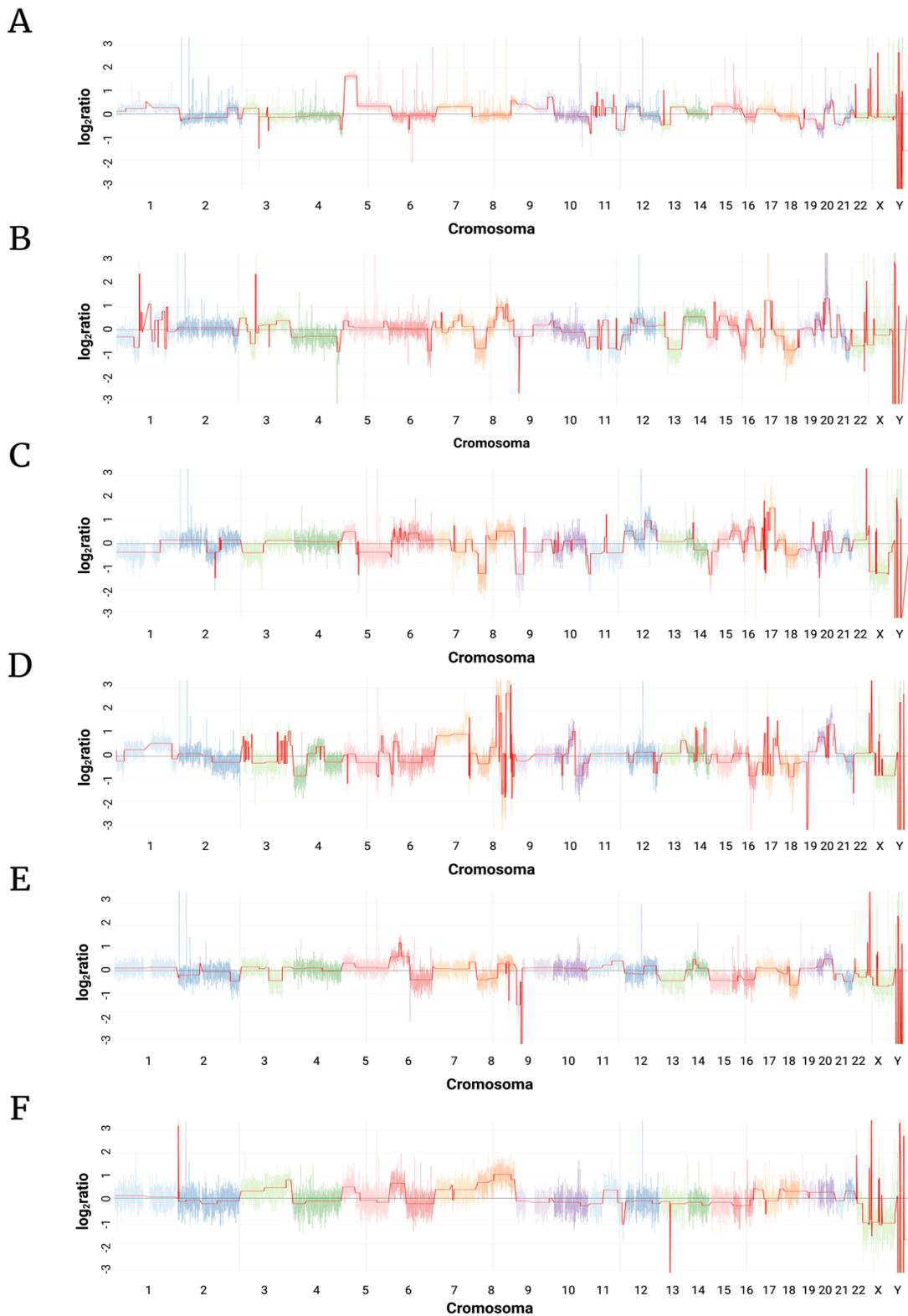


Figura 9. El análisis de fragmentos cortos de DNA mediante la tecnología de secuenciación genómica con nanoporos es útil para la detección de alteraciones en el número de copias de DNA. En esta figura se presentan los cariotipos moleculares generados con la herramienta NanoGLADIATOR para la detección de CNAs en

las líneas celulares: A) HeLa, B) MCF7, C) MDA-MB-361, D) SK-BR-3, E) MDA-MB-231 y F) Hs578T. El análisis se realizó con un tamaño de ventana genómica de 100 kpb y en el modo “no control”. En el eje de las abcisas se muestran los cromosomas en orden ascendente, donde cada cromosoma está representado por un color diferente. En el eje de las ordenadas, la línea roja continua representa el perfil de segmentación a través del genoma completo. Un valor de \log_2 ratio de 0 indica un número de copias normal o diploide. El aumento o la disminución del valor de \log_2 ratio corresponde a regiones genómicas con ganancia o pérdida del número de copias, respectivamente.

Específicamente en cáncer de mama, se han reportado con frecuencia alteraciones en los cromosomas 8, 11, 17 y 20 (Harbers et al., 2021; Sun & Liu, 2015), lo que respalda aún más nuestros hallazgos en estas líneas celulares tumorales, provenientes en su mayoría de tejido mamario (MCF7, MDA-MB-361, SK-BR-3, MDA-MB-231 y Hs578T). Este patrón de variación en diversas regiones genómicas resalta la complejidad y heterogeneidad de las alteraciones genómicas presentes en estos modelos de células tumorales. Importante destacar que, dada la amplia caracterización del genoma de estas líneas celulares, nuestros resultados validan la capacidad de la tecnología de secuenciación con nanoporos para reproducir de manera fiel las CNAs previamente reportadas.

Adicionalmente, la resolución de las alteraciones observadas en los cariotipos moleculares resalta la viabilidad del enfoque de secuenciación de baja profundidad (sWGS) en un experimento múltiplex, mostrando resultados precisos con una cobertura menor a 0.5x (Figura 7B-F). Estos hallazgos sugieren que este método ofrece una opción factible para identificar alteraciones cromosómicas, el cual puede ser relevante en contextos clínicos donde se pretende optimizar los costos sin comprometer la precisión en la detección de CNAs.

9.3 Validación de los resultados obtenidos mediante secuenciación con nanoporos con la tecnología de secuenciación Illumina

9.3.1 Análisis de correlación de datos de segmentación entre tecnologías de secuenciación

La evaluación de la correlación entre los resultados obtenidos a partir de la secuenciación genómica con nanoporos en líneas celulares tumorales y los datos provenientes de la secuenciación Illumina, disponibles en bases de datos públicas, constituyó un aspecto clave de esta trabajo. Se utilizó el software XCAVATOR para el análisis de los datos Illumina, el cual es equiparable al software NanoGLADIATOR empleado para la secuenciación con nanoporos, utilizando los mismos parámetros en ambos casos: modo 'no control' y una ventana genómica de 100 kpb.

Se compararon los datos de segmentación (valores de \log_2 ratio) de cuatro de las seis líneas celulares secuenciadas a través de 26,881 ventanas genómicas que eran correspondientes entre ambas tecnologías de secuenciación. Este análisis reveló una correlación muy alta entre los resultados obtenidos mediante secuenciación genómica con nanoporos y los datos Illumina en todas las líneas celulares estudiadas (coeficiente de correlación de Spearman = 0.90–0.94, valor $p < 2.2e-16$) (Figura 10).

Es importante destacar que la comparación se llevó a cabo considerando profundidades de secuenciación notoriamente diferentes. Los datos obtenidos mediante secuenciación con nanoporos reflejan una baja profundidad, menor a 0.5x. Por otro lado, los datos generados a través de secuenciación Illumina, según se indica en la Tabla 2, poseen una profundidad de secuenciación superior a 35x en tres de las cuatro líneas celulares analizadas (MCF7, MDA-MB-361, MDA-MB-231). Este contraste en las profundidades de secuenciación denota una diferencia considerable en la cantidad de datos generados por cada tecnología. A pesar de esta diferencia, los resultados revelaron una correlación significativamente alta entre ambos conjuntos de datos. Esto no solo evidencia la utilidad de la secuenciación con nanoporos para su empleo en entornos donde se busca una alternativa rentable y eficiente para la detección de alteraciones genómicas, sino que también señala su capacidad para producir resultados comparables con tecnologías de secuenciación de alto rendimiento, aún en condiciones de baja profundidad de secuenciación.

La correlación observada entre las tecnologías de secuenciación se ve respaldada por estudios previos. Baslan et al. (2021) reportó la correlación entre la secuenciación con nanoporos, utilizando el secuenciador MinION, y la secuenciación Illumina en la línea celular SK-BR-3, identificando una correlación de 0.98 en un tamaño de ventana genómica de 150 kpb y de 0.99 en una ventana genómica de 600 kpb. Estos hallazgos refuerzan la capacidad de la secuenciación con nanoporos para generar datos comparables a los obtenidos con tecnologías de secuenciación de lecturas cortas, incluso al utilizar ventanas genómicas de distintos tamaños.

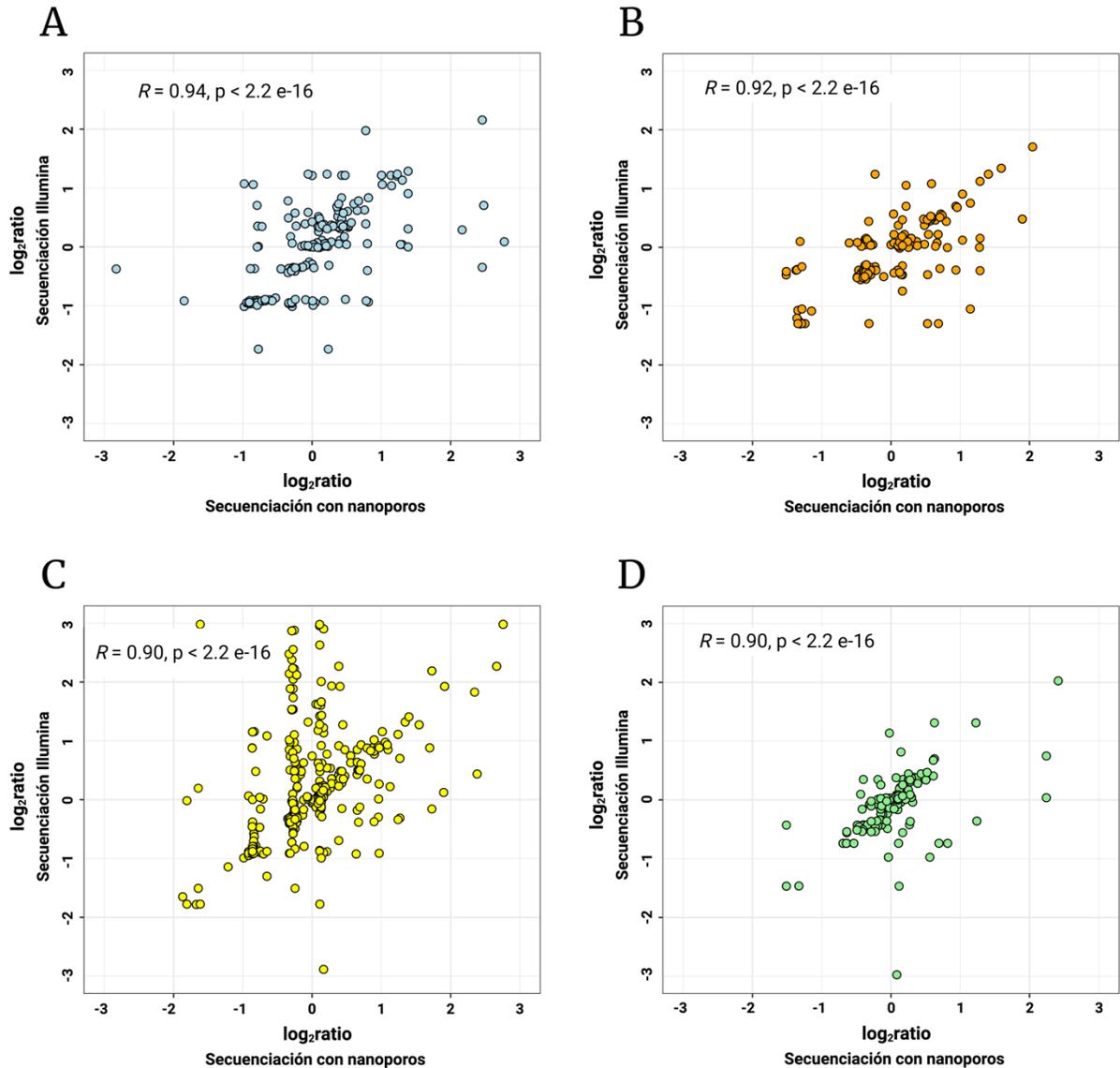


Figura 10. Los resultados de detección de alteraciones en el número de copias de DNA mediante las tecnologías de secuenciación con nanoporos y de secuenciación Illumina están altamente correlacionados.

En esta figura se presentan los gráficos de correlación entre los datos de secuenciación genómica con nanoporos analizados con el software NanoGLADIATOR y los datos de secuenciación Illumina analizados con el software XCAVATOR de las líneas celulares: A) MCF7 (Coeficiente de correlación de Spearman = 0.94, valor $p < 2.2e-16$, $n = 26,881$), B) MDAMB361 (Coeficiente de correlación de Spearman = 0.92, valor $p < 2.2e-16$, $n = 26,881$), C) SKBR3 (Coeficiente de correlación de Spearman = 0.90, valor $p < 2.2e-16$, $n = 26,881$) y D) MDAMB231 (Coeficiente de correlación de Spearman = 0.90, valor $p < 2.2e-16$, $n = 26,881$). En el eje de las abcisas se muestra el valor promedio de segmentación (\log_2 ratio) calculado para cada ventana genómica en los datos de secuenciación con nanoporos y en el eje de las abcisas los calculados para cada ventana genómica en los datos de secuenciación Illumina. En ambos casos, el análisis se realizó con un tamaño de ventana genómica de 100 kpb y en el modo “no control”. La correlación de los datos se evaluó en 26,881 ventanas genómicas.

9.3.2 Comparación de regiones genómicas alteradas a nivel citogenético entre tecnologías de secuenciación

Si bien el análisis de los valores de $\log_2\text{ratio}$ a través de ventanas de tamaño definido en un genoma proporciona una visión general del tipo de alteraciones presentes en una muestra, es decir, ganancia o pérdida en el número de copias, el objetivo final del análisis de las CNAs es obtener una comprensión más profunda de las características moleculares de las células tumorales. Esta comprensión detallada se logra al identificar la ubicación específica de cada alteración cromosómica detectada. Esto resulta crucial para identificar con precisión genes cruciales en la biología del cáncer de mama que residen dentro de las regiones alteradas, los cuales pueden ser clave para el diagnóstico, pronóstico y tratamiento de las pacientes.

Para llevar a cabo la comparación a nivel citogenético entre las tecnologías de secuenciación, se aplicaron criterios específicos para eliminar los valores de segmentación ($\log_2\text{ratio}$) de baja amplitud. Estos valores, originados por fluctuaciones aleatorias que pueden generar resultados falsos positivos, se filtraron mediante el establecimiento de valores de corte. Se clasificaron como ganancias (duplicaciones de una o múltiples copias) a las regiones genómicas con valores de $\log_2\text{ratio}$ mayores a 0.40, mientras que se consideraron como pérdidas (deleciones de una o ambas copias) aquellas con valores de $\log_2\text{ratio}$ menores a -0.40. Bajo estas condiciones, se detectaron 24, 60, 45, 66, 25 y 15 alteraciones cromosómicas en el genoma de las líneas celulares HeLa, MCF7, MDA-MB-361, SK-BR-3, MDA-MB-231 y Hs578T, respectivamente (Tablas Suplementarias 1-6).

El análisis comparativo se centró en las mismas cuatro líneas celulares estudiadas en la sección 9.3.1. La localización citogenética de las regiones genómicas alteradas se comparó con la localización de las regiones detectadas mediante secuenciación Illumina. Los valores de segmentación para los datos Illumina también se filtraron utilizando los mismos valores de corte. En este caso, el análisis de datos reveló la presencia de 58, 54, 110 y 56 CNAs en el genoma de las líneas celulares MCF7, MDA-MB-361, SK-BR-3 y MDA-MB-231, respectivamente (Tablas Suplementarias 7-10).

La Tabla 4 proporciona un resumen del conteo global de alteraciones cromosómicas identificadas en cada línea celular utilizando ambas tecnologías de secuenciación. Además, se presenta la categorización de estas alteraciones según su estado de número de copias, expresado en términos numéricos y porcentuales.

Tabla 4. Resumen de alteraciones en el número de copias de DNA detectadas mediante las tecnologías de secuenciación genómica con nanoporos y secuenciación Illumina en líneas celulares tumorales.

Tecnología	Estado de CN	Línea celular			
		MCF7 n (%)	MDA-MB-361 n (%)	SK-BR-3 n (%)	MDA-MB-231 n (%)
Secuenciación con nanoporos	Duplicación	25 (41.7)	17 (37.8)	21 (31.8)	8 (33.3)
	Duplicación de múltiples copias	11 (18.3)	7 (15.6)	20 (30.3)	1 (4.2)
	Delección	22 (36.7)	20 (44.4)	23 (34.9)	13 (54.2)
	Doble delección	2 (3.3)	1 (2.2)	2 (3.0)	2 (8.3)
	Total	60 (100.0)	45 (100.0)	66 (100.0)	24 (100.0)
Secuenciación Illumina	Duplicación	17 (29.3)	18 (33.3)	25 (22.7)	15 (26.8)
	Duplicación de múltiples copias	16 (27.6)	14 (25.9)	60 (54.6)	17 (30.3)
	Delección	25 (43.1)	22 (40.8)	24 (21.8)	23 (41.1)
	Doble delección	0 (0.0)	0 (0.0)	1 (0.9)	1 (1.8)
	Total	58 (100.0)	54 (100.0)	110 (100.0)	56 (100.0)

CN= Número de copias

Se observa una diferencia notable en la proporción de duplicaciones y duplicaciones de múltiples copias detectadas entre las tecnologías de secuenciación con nanoporos e Illumina en las líneas celulares. En general, la secuenciación con nanoporos muestra una mayor cantidad de duplicaciones en comparación con las duplicaciones de múltiples copias, mientras que con la secuenciación Illumina ocurre lo contrario.

Una posible explicación para esto podría deberse a las diferencias en la profundidad de secuenciación entre ambas tecnologías. Es importante tener en cuenta que los datos de secuenciación con nanoporos corresponden a una secuenciación de baja profundidad, mientras que los datos de secuenciación Illumina provienen de una secuenciación de alta profundidad. Esta disparidad en la profundidad de secuenciación puede afectar la capacidad de estimar el número de copias relativo para cada región genómica. En particular, la baja profundidad de secuenciación con nanoporos podría estar influyendo en la precisión de la detección de duplicaciones de múltiples copias, dificultando la distinción entre duplicaciones de una copia y duplicaciones de múltiples copias en el análisis de los datos.

Respecto a las deleciones detectadas en las líneas celulares, se observa en términos generales una cantidad semejante entre ambas tecnologías, aunque se detecta ligeramente un mayor número de deleciones dobles mediante la secuenciación con nanoporos. A diferencia de las duplicaciones, donde el número de copias ganadas puede ser alto y la distinción entre duplicaciones simples y de múltiples copias puede ser más complicada, en el caso de las deleciones, la distinción se reduce a la pérdida de una o dos copias de la región. Por lo tanto, en este caso, el enfoque de profundidad de secuenciación puede no estar afectando en gran medida la capacidad para distinguir entre una deleción de una copia y una de dos copias.

9.3.2.1 Línea celular MCF7

La secuenciación genómica con nanoporos reveló 60 CNAs en esta línea celular: 42% duplicaciones (800 - 86,925 kpb), 18% duplicaciones de múltiples copias (500 - 26,015 kpb), 37% deleciones (600 - 54,475 kpb) y solo 3% deleciones de ambas copias (200 - 4,885 kpb) (Tabla 4 y Tabla Suplementaria 2). Por otro lado, la secuenciación Illumina detectó 58 CNAs: 29% duplicaciones (998 - 88,200 kpb), 28% duplicaciones de múltiples copias (400 - 43,691 kpb) y 43% deleciones (500 - 54,720 kpb) (Tabla 4 y Tabla Suplementaria 7).

Los resultados de la secuenciación genómica con nanoporos y la secuenciación Illumina revelaron una serie de CNAs en esta línea celular. Estos cambios genómicos incluyen duplicaciones y deleciones que afectan diversas regiones, mostrando una marcada predominancia de duplicaciones de una y múltiples copias. Aunque se observaron diferencias en la detección de deleciones de ambas copias entre ambas tecnologías, la similitud en el recuento total de CNAs sugiere una coherencia en la identificación de cambios genómicos.

A partir de estos resultados, se lograron identificar 34 regiones genómicas alteradas consistentes entre ambas tecnologías. En la mayoría de los casos, estas regiones presentaron tamaños y números de copias similares, detallados en la Tabla 5. De estas regiones, se observó que los cromosomas con mayor frecuencia de CNAs son el 1, 3, 11 y 20 (Figura 11). En el cromosoma 14 se detectó la alteración con mayor extensión (secuenciación con nanoporos = 86,925 kpb; secuenciación Illumina = 88,200 kpb), la cual genera una ganancia que afecta todo el brazo largo del cromosoma (Figura 11). Además, mediante secuenciación con nanoporos se detectaron amplificaciones de 11 copias en los cromosomas 1 y 3. Estos hallazgos concuerdan con reportes previos sobre el perfil de CNAs que caracteriza a esta línea celular (Beroukhim et al., 2010; Jönsson et al., 2007; Mackay et al., 2009; Shadoo & Lam, 2006).

Tabla 5. Regiones genómicas alteradas en el número de copias que son concordantes entre ambas tecnologías de secuenciación (nanoporos e Illumina) en la línea celular MCF7.

Cromosoma	Citobanda de inicio	Citobanda de fin	Secuenciación con nanoporos		Secuenciación Illumina		Genes clínicamente relevantes
			CN	Tamaño (kpb)	CN	Tamaño (kpb)	
1	1p22.3	1p21.1	1	20,625	1	20,686	
1	1p21.1	1p13.3	11	1,900	3	2,300	
1	1q21.1(N) / 1q12(l)	1q21.3	4	6,970	4	9,974	
1	1q23.1	1q24.2	1	9,800	1	9,900	
1	1q31.3	1q32.2	3	9,430	4	9,473	
1	1q32.2	1q41	1	6,765	1	6,900	
1	1q41	1q41	4	4,800	4	5,000	
3	3p22.1	3p14.2	1	19,495	1	20,200	
3	3p14.2	3p14.1	11	3,400	9	2,400	
3	3q13.31	3q13.31	1	700	1	800	
4	4q34.3	4q35.2	1	9,180	1	9,000	
6	6q25.2	6q27	1	14,597	1	14,712	
7	7p14.3	7p14.3(N) / 7p14.2(l)	3	3,200	3	3,600	
7	7q22.1	7q31.1	3	13,710	3	13,840	
8	8p23.3	8p11.1	1	42,870	1	43,732	
11	11p15.5	11p15.3	1	11,650	1	11,853	
11	11q11	11q12.1	1	4,195	1	4,800	
11	11q14.2	11q23.3	1	30,470	1	30,563	<i>PGR, ATM</i>
11	11q23.3	11q23.3	3	2,100	3	2,200	
11	11q23.3	11q25	1	15,285	1	15,600	
13	13q21.2	13q34	1	54,475	1	54,720	
14	14q11.2(N) / 14q11.1(l)	14q32.33	3	86,925	3	88,200	<i>ESR2, AKT1</i>
15	15q21.2	15q21.3	5	4,800	5	4,900	

Cromosoma	Citobanda de inicio	Citobanda de fin	Secuenciación con nanoporos		Secuenciación Illumina		Genes clínicamente relevantes
			CN	Tamaño (kpb)	CN	Tamaño (kpb)	
15	15q23	15q26.3	3	31,460	3	39,425	
16	16q11.2	16q22.1	1	20,575	1	20,800	
17	17q11.2	17q21.1(N) / 17q21.2(I)	1	9,930	1	10,263	<i>NF1</i>
17	17q21.32	17q24.3	4	22,180	4	22,235	
18	18q12.2	18q23	1	44,032	1	44,161	
19	19p13.3	19p13.2	1	9,255	1	9,977	
20	20p12.3	20p11.21(N) / 20p11.1(I)	3	18,800	3	19,600	
20	20q12	20q12	5	1,600	5	1,900	
20	20q13.12(N) / 20q13.13(I)	20q13.33	5	17,316	4-5	14,716	
21	21q22.13	21q22.3	3	8,545	3	8,745	
22	22q12.3	22q13.33	1	15,725	1	16,105	

Para cada CNA detectada por medio de ambas tecnologías de secuenciación, se presenta la localización citogenética, el estado de número de copias y el tamaño de la alteración detectado con secuenciación con nanoporos (columna verde) y con secuenciación Illumina (columna azul). kpb=kilopares de bases; N= Secuenciación con nanoporos; I= Secuenciación Illumina

Estas alteraciones identificadas revelaron la delección de genes clínicamente relevantes para el cáncer de mama, como *PGR* (11q22.1), *ATM* (11q22.3), *ESR2* (14q23.2-q23.3), *AKT1* (14q32.33) y *NF1* (17q11.2). Además de los anteriores, los datos obtenidos mediante la tecnología de secuenciación genómica con nanoporos permitieron identificar otros genes relevantes que no fueron detectados en los datos de secuenciación Illumina. Destacan la amplificación de los genes *PIK3CA* (3q26.32), *CCND1* (11q13.3), *CDK4* (12q14.1) y *MDM2* (12q15), y la delección del gen *RB1* (13q14.2). Es relevante subrayar que la mayoría de estos genes y las regiones genómicas en las que se localizan ya habían sido previamente descritos para esta línea celular mediante diversos métodos de detección como qPCR, FISH, CGH y CGH basado en microarreglos (Nugoli et al., 2003; Ono et al., 2021; Shadeo & Lam, 2006; Wu et al., 2005).

Esta diferencia en la capacidad de detección entre las plataformas sugiere una ventaja de la secuenciación con nanoporos para obtener una perspectiva más completa y detallada de las alteraciones presentes en las células tumorales.

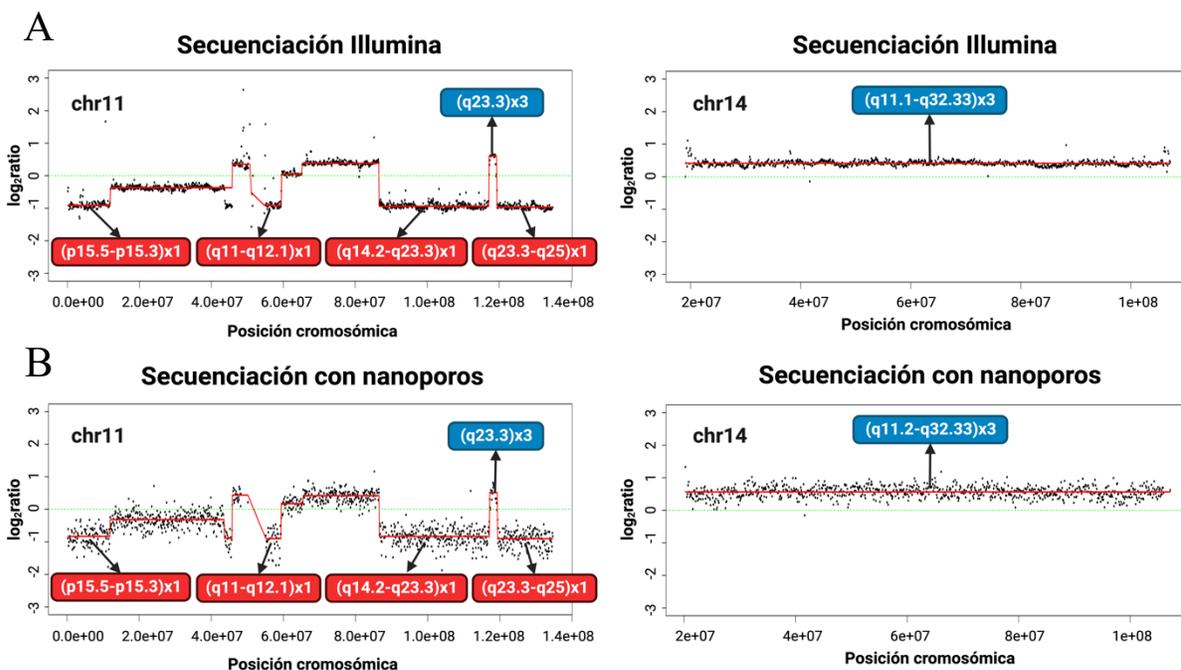


Figura 11. Perfiles de segmentación de los cromosomas 11 y 14 obtenido con secuenciación con nanoporos e Illumina en la línea celular MCF7. En esta figura se presentan dos ejemplos de perfiles de segmentación, a nivel cromosómico, de la línea celular MCF7. En el gráfico superior se muestran los resultados generados con los datos de secuenciación Illumina (A), mientras que en el gráfico inferior se presentan los resultados obtenidos mediante secuenciación con nanoporos (B). En el eje de las abcisas se muestra la posición cromosómica. En el eje de las ordenadas, la línea roja indica el perfil de segmentación en dicho cromosoma. Un valor de \log_2 ratio de 0 indica un número de copias normal o diploide. El aumento o la disminución del valor de \log_2 ratio corresponde a regiones genómicas con ganancia (recuadros azules) o pérdida del número de copias (recuadros rojos), respectivamente. En cada recuadro se muestra la localización citogenética de las regiones genómicas alteradas (dentro del paréntesis) y su número de copias (fuera del paréntesis). En el cromosoma 11, se detectaron cinco CNAs; cuatro de ellas se identificaron como deleciones de una copia, localizadas en las citobandas p15.5-p15.3, q11-q12.1, q14.2-q23.3 y q23.2-q25, y una como duplicación, ubicada en la región q23.3. En el cromosoma 14, se identificó la ganancia de 3 copias de la región que abarca las citobandas q11.1/11.2-q22.1, la cual fue la CNA de mayor extensión encontrada en esta línea celular.

9.3.2.2 Línea celular MDA-MB-361

En esta línea celular se detectaron 45 CNAs utilizando la tecnología de secuenciación con nanoporos. El 38% de ellas se identificaron como duplicaciones (1,000 – 70,310 kpb), el 16% como duplicaciones de múltiples copias (800 – 30,795 kpb), el 44% como deleciones (1,200 –

64,325 kpb) y el 2% como deleciones dobles (900 – 31,045 kpb) (Tabla 4 y Tabla Suplementaria 3). En contraposición, la secuenciación Illumina identificó 58 CNAs: 33% duplicaciones (600 – 70,497 kpb), 26% duplicaciones de múltiples copias (200 – 25,800 kpb) y 41% deleciones (701 – 48,364 kpb), sin encontrar deleciones de ambas copias (Tabla 4 y Tabla Suplementaria 8).

Las duplicaciones de una copia fueron la única clase de alteración que se detectó de forma semejante entre ambas tecnologías (secuenciación con nanoporos = 38%; secuenciación Illumina = 33%). Particularmente, con secuenciación con nanoporos se detectó 1 deleción de ambas copias, la cual no fue identificada mediante la otra tecnología (Tabla 4).

De estas alteraciones, se identificaron 30 regiones genómicas consistentemente alteradas entre ambas tecnologías, mostrando semejanzas en tamaños y números de copias (Tabla 6). Destacan los cromosomas 8, 11, 12 y 17 con mayor frecuencia de CNAs, siendo el cromosoma 8 el que mostró la alteración de mayor extensión (secuenciación con nanoporos = 70,310 kpb; secuenciación Illumina = 70,497 kpb). En cambio, el cromosoma 17 se caracterizó por estar mayormente afectado por alteraciones de aumento en el número de copias (Figura 12). El perfil de alteraciones observado en esta línea celular concuerda con lo reportado en trabajos anteriores (Jönsson et al., 2007; Mackay et al., 2009).

Tabla 6. Regiones genómicas alteradas en el número de copias que son concordantes entre ambas tecnologías de secuenciación (nanoporos e Illumina) en la línea celular MDA-MB-361.

Cromosoma	Citobanda de inicio	Citobanda de fin	Secuenciación con nanoporos		Secuenciación Illumina		Genes clínicamente relevantes
			CN	Tamaño (kpb)	CN	Tamaño (kpb)	
5	5p15.33	5p12(N) / 5p11(l)	3	45,740	3	46,371	
7	7q11.22	7q11.23	3	4,185	3	4,999	
8	8p23.3	8p23.2	1	2,450	1	3,000	
8	8p23.2	8p12	1	31,045	1	31,032	
8	8q21.11	8q24.3	3	70,310	3	70,497	<i>MYC</i>
9	9p24.3	9p21.1	1	30,090	1	30,400	<i>CDKN2A</i> , <i>CDKN2B</i>
9	9p21.1	9p13.2	3	5,910	3	6,900	
10	10p15.3	10p15.1	1	4,950	1	5,100	

Cromosoma	Citobanda de inicio	Citobanda de fin	Secuenciación con nanoporos		Secuenciación Illumina		Genes clínicamente relevantes
			CN	Tamaño (kpb)	CN	Tamaño (kpb)	
10	10q26.2	10q26.3	1	6,361	1	6,361	
11	11p15.5	11p15.4	1	6,585	1	6,853	
11	11p15.4	11p11.12	1	43,220	1	44,678	
11	11q13.3	11q13.3	5	1,680	4	1,580	<i>CCND1</i>
12	12p13.33	12p11.1	3	34,050	3	34,694	
12	12q21.2	12q23.3	4	25,730	4	25,800	
12	12q23.3	12q24.32	3	21,840	3	21,824	
14	14q21.3	14q22.1	4	3,600	3	3,700	
15	15q11.2	15q13.2	1	8,805	1	7,663	
16	16p13.3	16p11.1	3	34,935	3	35,213	
16	16q22.1	16q24.3	3	23,350	3	23,454	
17	17q11.1	17q11.2	3	3,840	3	4,200	
17	17q11.2	17q12	3	2,000	3	3,100	
17	17q12	17q12	1	1,200	1	1,663	
17	17q12	17q21.1	4	1,595	3	1,900	<i>ERBB2, GRB7</i>
17	17q21.1	17q21.31	1	2,825	1	2,900	<i>BRCA1</i>
17	17q21.32	17q25.3	5	30,795	5	31,035	
18	18q12.1	18q23	1	47,132	1	47,161	
19	19q13.42	19q13.43	4	3,345	3	3,500	
20	20q11.21	20q12	3	7,912	3	8,428	
20	20q13.13	20q13.33	3	13,318	3	13,516	
21	21q11.2	21q22.3	1	32,435	1	33,545	

Para cada CNA detectada por medio de ambas tecnologías de secuenciación, se presenta la localización citogenética, el estado de número de copias y el tamaño de la alteración detectado con secuenciación con nanoporos (columna verde) y con secuenciación Illumina (columna azul).kpb=kilopares de bases; N= Secuenciación con nanoporos; I= Secuenciación Illumina

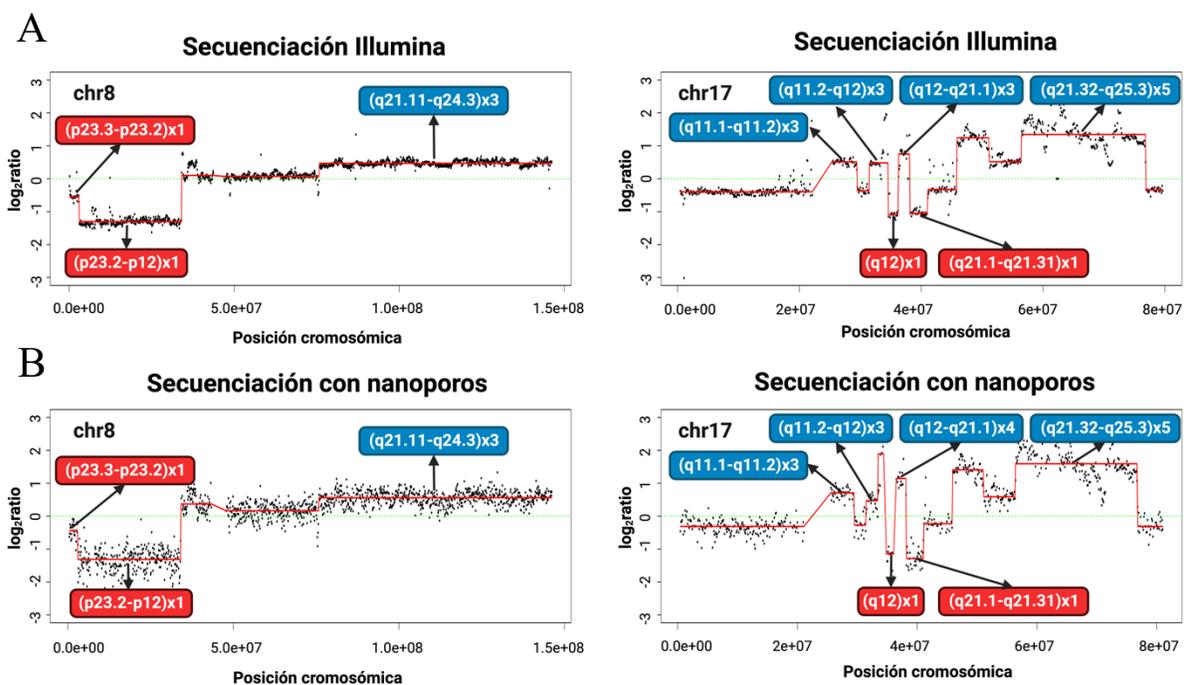


Figura 12. Perfiles de segmentación de los cromosomas 8 y 17 obtenido con secuenciación con nanoporos e Illumina en la línea celular MDA-MB-361. En esta figura se presentan dos ejemplos de perfiles de segmentación, a nivel cromosómico, de la línea celular MDA-MB-361. En el gráfico superior se muestran los resultados generados con los datos de secuenciación Illumina (A), mientras que en el gráfico inferior se presentan los resultados obtenidos mediante secuenciación con nanoporos (B). En el eje de las abcisas se muestra la posición cromosómica. En el eje de las ordenadas, la línea roja indica el perfil de segmentación en dicho cromosoma. Un valor de $\log_2\text{ratio}$ de 0 indica un número de copias normal o diploide. El aumento o la disminución del valor de $\log_2\text{ratio}$ corresponde a regiones genómicas con ganancia (recuadros azules) o pérdida del número de copias (recuadros rojos), respectivamente. En cada recuadro se muestra la localización citogenética de las regiones genómicas alteradas (dentro del paréntesis) y su número de copias (fuera del paréntesis). En el cromosoma 8, se detectaron tres CNAs; dos de ellas se identificaron como deleciones de una copia en el brazo corto del cromosoma, en las regiones p23.3-p23.2 y p23.2-p12, y una ganancia de 3 copias de la región que abarca las citobandas q21.11-q24.3; esta última fue la alteración de mayor extensión encontrada en la línea celular. En el cromosoma 17, se identificó el mayor número de CNAs. De estas, cuatro generaron un aumento en el número de copias en las regiones: q11.1-q11.2, q11.2-q12, q12-q21.1 y q21.32-q25.3; en el caso de las regiones q12 y q21.1-q21.1-q21.31, se detectaron deleciones de una copia.

A partir de estas regiones genómicas alteradas, se identificaron siete genes clínicamente relevantes en el cáncer de mama. Entre estos genes, cuatro mostraron un aumento en el número de copias de DNA: *MYC* (8q24.21), *CCND1* (11q13.3), *ERBB2* (17q12) y *GRB7* (17q12). Por otro lado, tres genes presentaron una disminución en el número de copias de DNA: *CDKN2A* (9p21.3), *CDKN2B* (9p21.3) y *BRCA1* (17q21.31). Además de los anteriores, mediante los datos de secuenciación con nanoporos se logró la identificación de tres genes relevantes que no fueron

detectados al analizar los datos de secuenciación Illumina. Estos genes incluyen *PGR* (11q22.1) y *ATM* (11q22.3), ambos presentando una pérdida en el número de copias, cuya alteración ha sido previamente documentada en la literatura mediante CGH basado en microarreglos (Mackay et al., 2009). De forma semejante a lo observado en la línea celular MCF7, estos hallazgos destacan la capacidad de la secuenciación con nanoporos para proporcionar mayor información sobre las alteraciones genómicas presentes en las células tumorales, en comparación con la tecnología de secuenciación Illumina.

9.3.2.3 Línea celular SK-BR-3

Los resultados en esta línea celular mostraron la detección de 66 CNAs mediante secuenciación con nanoporos. El 32% se clasificó en duplicaciones (700 – 83,900 kpb), el 30% en duplicaciones de múltiples copias (300 - 65,135 kpb), el 35% en deleciones (400 - 48,822 kpb) y el 3% en deleciones de dos copias (500 – 6,000 kpb) (Tabla 4 y Tabla Suplementaria 4). Por otro lado, los datos de secuenciación Illumina permitieron la detección de 110 CNAs. De acuerdo con el estado de número de copias, se identificó que el 23% de estas corresponden a duplicaciones (1,200 - 70,600 kpb), el 54% a duplicaciones de múltiples copias (200 - 21,716 kpb), el 22% a deleciones (500 - 48,987 kpb) y sólo el 1% a deleciones dobles (4,236 kpb) (Tabla 4 y Tabla Suplementaria 9).

Destaca que el número total de CNAs fue notablemente mayor en los datos de secuenciación Illumina en comparación con los de secuenciación con nanoporos. Específicamente, se observó que las duplicaciones de múltiples copias fueron tres veces más frecuentes en los datos de Illumina. De este hallazgo es importante señalar que los datos de Illumina provienen de un experimento de secuenciación dirigida, mientras que los datos de secuenciación con nanoporos se obtuvieron de experimentos de WGS (Tabla 2), lo cual podría influir en las diferencias observadas. La secuenciación dirigida se enfoca en áreas específicas del genoma, lo que puede permitir una mayor profundidad de secuenciación en esas regiones específicas, lo que a su vez incrementa la probabilidad de detectar CNAs en ellas (Pei et al., 2023). Por el contrario, si bien la secuenciación a nivel de genoma completo ofrece una visión global, esta puede limitar la detección de CNAs en regiones genómicas con muy baja profundidad de secuenciación, especialmente en experimentos con una profundidad de cobertura inferior a 1x, como en este trabajo.

Dado que los datos provienen de experimentos de secuenciación diferentes, esto impide una comparación directa. Idealmente, sería más apropiado analizar los datos de secuenciación con nanoporos restringiendo la evaluación a los mismos genes que se analizaron en los datos de secuenciación dirigida, para así facilitar una comparación más directa y precisa. A pesar de esta limitación, se optó por realizar un análisis exploratorio para identificar posibles regiones genómicas alteradas que fueran concordantes entre ambas plataformas de secuenciación.

Se identificaron 42 regiones en el genoma cuyo estado alterado era concordante entre las tecnologías de secuenciación (Tabla 7). En estos resultados, se destacaron los cromosomas 3, 8, 10, 17 y 20 por presentar el mayor número de CNAs. Específicamente, en los cromosomas 8 y 17 se identificaron amplificaciones de hasta 18 y 9 copias, respectivamente (Figura 13). Aunque en el cromosoma 1 sólo se detectó una alteración, esta fue la de mayor extensión encontrada en el genoma de la línea celular (secuenciación con nanoporos = 83,900 kpb; secuenciación Illumina = 66,988 kpb). En el brazo corto del cromosoma 19 se halló la única doble delección reportada mediante ambas tecnologías. Al comparar este perfil de alteraciones con la literatura existente para esta línea celular, se observó una consistencia notable (Jönsson et al., 2007; Mackay et al., 2009; Shadeo & Lam, 2006).

Pese a la diferencia en la cantidad de CNAs detectadas entre las tecnologías de secuenciación, ambas permitieron la identificación de genes de relevancia clínica en el contexto del cáncer de mama. Las notables amplificaciones de 18 y 9 copias en los cromosomas 8 y 17, respectivamente, correspondieron a regiones genómicas que albergan a los genes *MYC* (8q24.21), *ERBB2* (17q12) y *GRB7* (17q12). Adicionalmente, se identificaron ganancias en el número de copias de *EGFR* (7p11.2) y *NF1* (17q11.2), mientras que se evidenció la pérdida en el número de copias en genes de importancia como *PTEN* (10q23.31) y *BRCA1* (17q21.31). La identificación de genes clínicamente relevantes, incluso en condiciones de menor resolución, destaca el potencial de el enfoque de sWGS, utilizando secuenciación genómica con nanoporos, para recuperar información crucial sobre las alteraciones genómicas más importantes presentes en las células tumorales.

Tabla 7. Alteraciones en el número de copias de DNA detectadas mediante ambas tecnologías de secuenciación (nanoporos e Illumina) en la línea celular SK-BR-3.

Cromosoma	Citobanda de inicio	Citobanda de fin	Secuenciación con nanoporos		Secuenciación Illumina		Genes clínicamente relevantes
			CN	Tamaño (kpb)	CN	Tamaño (kpb)	
1	1q21.1(N) / 1q23.3(I)	1q42.13	3	83,900	3	66,988	
3	3p26.1	3p26.1	3	3,600	4	3,700	
3	3p22.2	3p22.1	4	1,600	4	2,900	
3	3q26.2	3q26.31	3	3,585	4	3,900	
3	3q26.32	3q29	4	19,585	4	15,100	
4	4p16.3	4p11	1	48,822	1	48,987	
5	5p15.2	5p15.1	1	1,080	1	1,100	
5	5q23.3	5q31.2	1	8,295	1	9,700	
5	5q32	5q33.2	3	6,200	3	6,500	
6	6p25.3	6p24.3	1	7,000	1	7,300	
6	6p24.3	6p22.3	3	16,200	3	17,200	
7	7p22.3	7p11.2(N) / 7p11.1(I)	3	57,270	4	58,001	<i>EGFR</i>
7	7q11.21(N) / 7q11.1(I)	7q32.1(N) / 7q32.3(I)	4	65,135	4	70,600	
7	7q32.3	7q33	1	1,100	1	1,200	
8	8q13.3(N) / 8q13.1(I)	8q21.13	13	10,600	10	14,600	
8	8q21.3(N) / 8q21.2(I)	8q21.3	8	5,435	8	5,900	
8	8q21.3	8q22.1	1	6,000	1	6,100	
8	8q22.3	8q23.1	1	4,100	1	4,700	
8	8q23.1	8q24.13	14	16,800	16	17,000	
8	8q24.21	8q24.21	17	1,200	18	1,400	<i>MYC</i>
8	8q24.22	8q24.22	1	600	1	600	
8	8q24.23	8q24.3	1	4,800	1	4,300	
10	10p11.23	10p11.22	1	3,400	1	3,300	

Cromosoma	Citobanda de inicio	Citobanda de fin	Secuenciación con nanoporos		Secuenciación Illumina		Genes clínicamente relevantes
			CN	Tamaño (kpb)	CN	Tamaño (kpb)	
10	10q21.1	10q22.1(N) / 10q21.3(I)	3	16,465	3	14,000	
10	10q22.1(N) / 10q21.3(I)	10q23.1	4	8,580	4	12,700	
10	10q23.1	10q25.2	1	30,980	1	30,800	<i>PTEN</i>
12	12q23.3	12q24.11	1	1,200	1	1,300	
13	13q32.1	13q34	3	19,930	3	19,930	
14	14q31.3	14q32.12	5	6,980	4	7,000	
16	16q21	16q24.3	1	28,450	1	28,754	
17	17q11.1	17q11.2	4	3,585	4	4,500	<i>NF1</i>
17	17q11.2	17q12	1	7,945	1	7,463	
17	17q12	17q21.2	7	3,225	9	3,300	<i>ERBB2,</i> <i>GRB7</i>
17	17q21.2(N) / 17q21.31(I)	17q23.2	1	18,890	1	16,000	<i>BRCA1</i>
17	17q23.2	17q24.3	3	8,625	4	9,235	
17	17q25.3	17q25.3	6	1,455	5	2,936	
18	18q12.2	18q21.31	1	21,069	1	21,298	
19	19p12	19p12(N) / 19p11(I)	0	3,465	0	4,236	
20	20p13	20p11.21(N) / 20p11.1(I)	3	25,445	4	26,200	
20	20q11.22	20q11.23(N) / 20q12(I)	5	4,612	5	4,743	
20	20q12	20q13.33	5	21,516	5	21,716	
22	22q13.31	22q13.31	1	600	1	700	

Para cada CNA detectada por medio de ambas tecnologías de secuenciación, se presenta la localización citogenética, el estado de número de copias y el tamaño de la alteración detectado con secuenciación con nanoporos (columna verde) y con secuenciación Illumina (columna azul).kpb=kilopares de bases; N= Secuenciación con nanoporos; I= Secuenciación Illumina

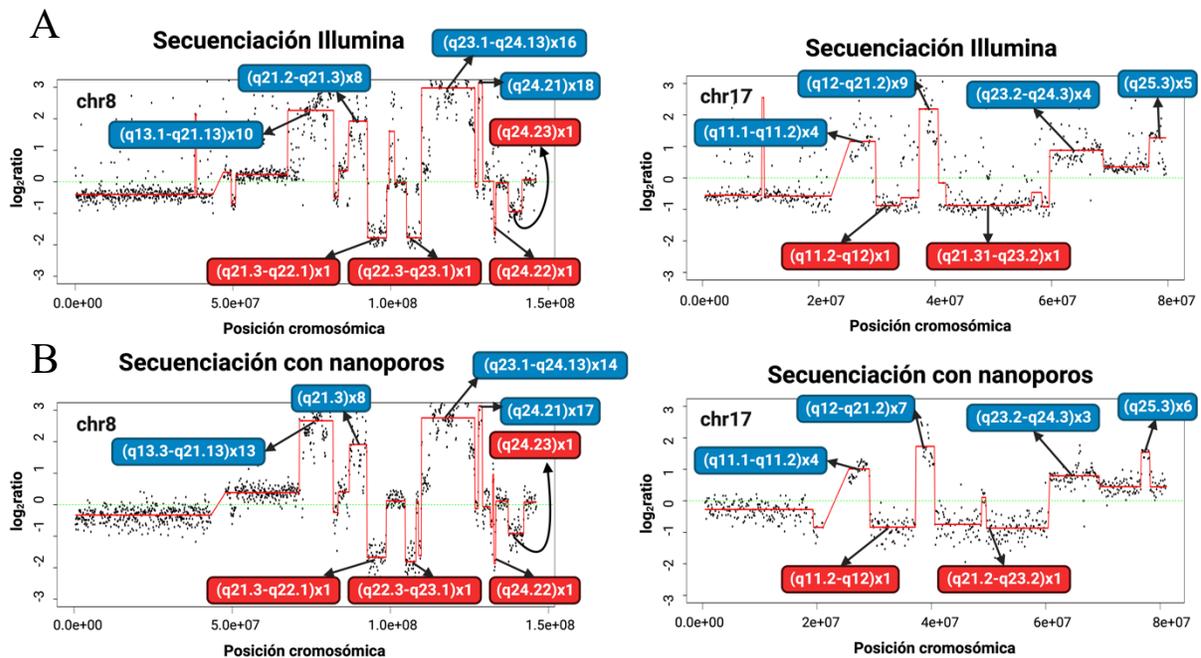


Figura 13. Perfiles de segmentación de los cromosomas 8 y 17 obtenido con secuenciación con nanoporos e Illumina en la línea celular SK-BR-3. En esta figura se presentan dos ejemplos de perfiles de segmentación, a nivel cromosómico, de la línea celular SK-BR-3. En el gráfico superior se muestran los resultados generados con los datos de secuenciación Illumina (A), mientras que en el gráfico inferior se presentan los resultados obtenidos mediante secuenciación con nanoporos (B). En el eje de las abcisas se muestra la posición cromosómica. En el eje de las ordenadas, la línea roja indica el perfil de segmentación en dicho cromosoma. Un valor de \log_2 ratio de 0 indica un número de copias normal o diploide. El aumento o la disminución del valor de \log_2 ratio corresponde a regiones genómicas con ganancia (recuadros azules) o pérdida del número de copias (recuadros rojos), respectivamente. En cada recuadro se muestra la localización citogenética de las regiones genómicas alteradas (dentro del paréntesis) y su número de copias (fuera del paréntesis). En el cromosoma 8, se detectó el mayor número de CNAs: cuatro duplicaciones de múltiples copias en las regiones: q13.1/q13.3-q21.13, q21.2/q21.3-q21.3, q23.1-q24.13 y q24.21; y cuatro deleciones de una copia en las regiones: q21.3-q22.1, q22.3-q23.1, q24.22 y q24.23. Se identificaron seis CNAs en el cromosoma 17. De estas, cuatro condujeron al incremento en la cantidad de copias presentes en las regiones cromosómicas: q11.1-q11.2, q12-q21.2, q23.2-q24.3 y q25.3. En lo que concierne a las regiones cromosómicas q11.2-q12 y q21.2-q23.2, se observaron deleciones que resultaron en la pérdida de una copia.

9.3.2.4 Línea celular MDA-MD-231

Los resultados de la secuenciación genómica con nanoporos revelaron 24 CNAs en esta línea celular. Se identificó que el 34% eran duplicaciones (1,300 - 44,685 kpb), el 4% eran duplicaciones de múltiples copias (8,200 kpb), el 54% eran deleciones (900 - 95,440 kpb) y solo el 8% eran deleciones de ambas copias (4,785 - 5,170 kpb). Por otra parte, la secuenciación Illumina detectó 56 CNAs, incluyendo 27% de duplicaciones (300 - 46,499 kpb), 30% de

duplicaciones de múltiples copias (200 – 8,400 kpb), 41% de deleciones (500 – 96,020 kpb) y 2% de deleciones dobles (5,300 kpb). Los detalles se encuentran en la Tabla 4 y Tabla Suplementaria 5 para la secuenciación con nanoporos, y en la Tabla 4 y Tabla Suplementaria 10 para la secuenciación Illumina.

Estos resultados muestran diferencias significativas en la detección de CNAs entre las tecnologías de secuenciación en esta línea celular. Mientras que la tecnología de secuenciación genómica con nanoporos reveló un número menor de CNAs global, la tecnología de secuenciación Illumina identificó un mayor número y diversidad de CNAs, incluyendo duplicaciones de múltiples copias que no fueron detectadas por la tecnología con nanoporos.

Se identificaron 21 regiones en el genoma donde el estado alterado coincidía entre las dos tecnologías de secuenciación (ver Tabla 8). Destacan los cromosomas 6 y 9 por exhibir la mayor cantidad de CNAs (14), seguidos por los cromosomas 8 y 16, que presentaron un menor número CNAs. Por otro lado, los cromosomas 13, 15 y 22 se caracterizaron por la deleción de una copia de sus brazos cromosómicos largos en su totalidad. En particular, el cromosoma 13 mostró la alteración de mayor extensión (secuenciación con nanoporos = 95,440 kpb; secuenciación Illumina = 96,020 kpb) (Figura 14). Además, en el brazo corto del cromosoma 9 se detectó la única deleción doble reportada por ambas tecnologías. Estos hallazgos son consistentes con lo reportado por Shadeo and Lam (2006), Jönsson et al. (2007) y Mackay et al. (2009) en esta línea celular.

Dentro de las regiones genómicas afectadas, se identificaron genes clínicamente accionables en cáncer de mama, detectados a través de ambas tecnologías de secuenciación. Estos hallazgos destacaron la presencia de amplificaciones en el número de copias significativas en genes como *CCND3* (6p21.1), *PGR* (11q22.1) y *ATM* (11q22.3). Asimismo, se observaron deleciones de una sola copia en los genes *ESR1* (6q25.1-q25.2), *CD274* (9p24.1), *JAK2* (9p24.1), *BRCA2* (13q13.1) y *RB1* (13q14.2) Además, se identificaron deleciones de ambas copias en los genes *CDKN2A* (9p21.3) y *CDKN2B* (9p21.3). Esta comparabilidad es los resultados entre tecnologías de secuenciación sugiere que la secuenciación con nanoporos puede ser una herramienta prometedora, aún bajo un enfoque de baja profundidad, para el estudio de alteraciones moleculares en el cáncer de mama y detección genes de importancia clínica.

Tabla 8. Alteraciones en el número de copias de DNA detectadas mediante ambas tecnologías de secuenciación (nanoporos e Illumina) en la línea celular MDA-MB-231.

Cromosoma	Citobanda de inicio	Citobanda de fin	Secuenciación con nanoporos		Secuenciación Illumina		Genes clínicamente relevantes
			CN	Tamaño (kpb)	CN	Tamaño (kpb)	
2	2q33.3	2q37.3	1	36,380	1	36,719	
3	3q13.13	3q26.1	1	53,910	1	53,800	
6	6p25.3	6p22.1(N) / 6p21.31(I)	3	28,080	3	34,800	
6	6p21.31	6p21.1	5	8,200	5	8,400	<i>CCND3</i>
6	6p21.1	6p11.2(N) / 6p11.1(I)	3	13,925	3	15,478	
6	6q11.1	6q13	3	12,910	3	13,098	
6	6q14.1	6q27	1	88,085	1	88,113	<i>ESR1</i>
8	8p23.3	8p11.1	1	42,870	1	43,732	
8	8q24.13	8q24.13	1	900	1	1,000	
9	9p24.2	9p22.2	1	14,400	1	14,500	<i>CD274, JAK2</i>
9	9p22.2	9p21.3	1	2,600	1	2,700	
9	9p21.3	9p21.2	0	5,170	0	5,300	<i>CDKN2A, CDKN2B</i>
11	11q14.3	11q25	3	44,685	3	46,499	<i>PGR, ATM</i>
13	13q12.11(N) / 13q11(I)	13q34	1	95,440	1	96,020	<i>BRCA2, RB1</i>
15	15q11.2	15q26.3	1	79,815	1	80,172	
16	16p13.3	16p12.2(N) / 16p11.2(I)	1	23,720	1	32,027	
16	16q11.2	16q24.3	1	43,525	1	44,554	
18	18q21.1	18q23	1	33,932	1	34,061	
20	20q11.21	20q13.33	3	32,928	3	33,059	
21	21q11.2	21q21.1	1	5,800	1	6,800	

Cromosoma	Citobanda de inicio	Citobanda de fin	Secuenciación con nanoporos		Secuenciación Illumina		Genes clínicamente relevantes
			CN	Tamaño (kpb)	CN	Tamaño (kpb)	
22	22q11.1	22q13.33	1	33,495	1	34,167	

Para cada CNA detectada por medio de ambas tecnologías de secuenciación, se presenta la localización citogenética, el estado de número de copias y el tamaño de la alteración detectado con secuenciación con nanoporos (columna verde) y con secuenciación Illumina (columna azul). kpb=kilopares de bases; N= Secuenciación con nanoporos; I= Secuenciación Illumina

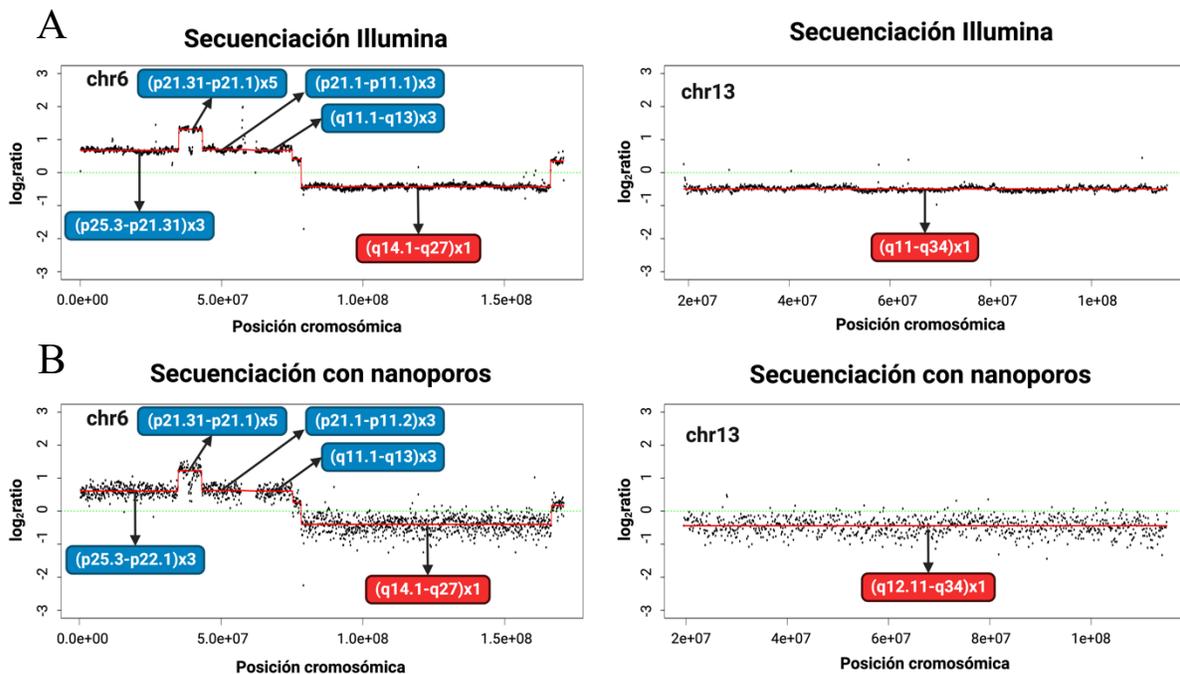


Figura 14. Perfiles de segmentación de los cromosomas 6 y 13 obtenido con secuenciación con nanoporos e Illumina en la línea celular MDA-MB-231. En esta figura se presentan dos ejemplos de perfiles de segmentación, a nivel cromosómico, de la línea celular MDA-MB-361. En el gráfico superior se muestran los resultados generados con los datos de secuenciación Illumina (A), mientras que en el gráfico inferior se presentan los resultados obtenidos mediante secuenciación con nanoporos (B). En el eje de las abcisas se muestra la posición cromosómica. En el eje de las ordenadas, la línea roja indica el perfil de segmentación en dicho cromosoma. Un valor de $\log_2\text{ratio}$ de 0 indica un número de copias normal o diploide. El aumento o la disminución del valor de $\log_2\text{ratio}$ corresponde a regiones genómicas con ganancia (recuadros azules) o pérdida del número de copias (recuadros rojos), respectivamente. En cada recuadro se muestra la localización citogenética de las regiones genómicas alteradas (dentro del paréntesis) y su número de copias (fuera del paréntesis). En el cromosoma 6, se identificó el mayor número de CNAs. Se detectaron tres duplicaciones simples y una duplicación de 5 copias, ubicadas en las regiones genómicas p25.3-p22.1/p21.31, p21.1-p11.2/p11.1, q11.1-q13, p21.31-p21.1, respectivamente. Además, en las región que abarca las citobandas q14.1-q27 se identificó una deleción de una copia. En el cromosoma 13, se identificó una alteración de ganancia con tres copias, en la región comprendida entre q12.11/q11-q34. Esta amplificación representó la CNA de mayor longitud detectada en la línea celular.

La fase inicial de este estudio se enfocó en establecer las condiciones óptimas para la secuenciación genómica con nanoporos de fragmentos cortos. Los resultados obtenidos demuestran que este enfoque tiene la capacidad de generar un alto número de lecturas con una profundidad adecuada para realizar un análisis de detección de CNAs, superando incluso el rendimiento reportado previamente en la literatura. Además, al identificar las regiones genómicas alteradas características en las líneas celulares tumorales, se evidenció que la secuenciación con nanoporos permitió detectar genes clínicamente relevantes en los tumores de mama, de manera comparable a la tecnología de secuenciación Illumina.

Un hallazgo significativo fue que en dos de las cuatro líneas celulares analizadas (MCF7 y MDA-MB-361), la secuenciación con nanoporos reveló alteraciones en el número de copias de genes, previamente identificadas en la literatura mediante otras técnicas de detección de CNAs, las cuales no fueron identificadas en los datos de secuenciación Illumina. Entre los genes albergados en estas regiones genómicas, *PIK3CA* (3q26.32), *CCND1* (11q13.3), *CDK4* (12q14.1) y *MDM2* (12q15), se encontraron afectados por una ganancia en el número de copias. Estos genes están ampliamente caracterizados en la literatura por desempeñar un papel oncogénico en el cáncer de mama (Baker et al., 2022; Haupt et al., 2017; Huw et al., 2013; Valla et al., 2022). En contraste, las regiones genómicas en las que se localizaban los genes *RB1* (13q14.2), *PGR* (11q22.1) y *ATM* (11q22.3) exhibieron pérdida en el número de copias y se distinguen por su función como supresores tumorales en el cáncer de mama (Bueno et al., 2014; Knudsen et al., 2020; Kunc et al., 2021). Esto resalta la capacidad de la tecnología para obtener información relevante sobre las alteraciones genómicas presentes en el cáncer de mama. Además, enfatiza la importancia de utilizar varios métodos de secuenciación para lograr una comprensión integral de las alteraciones genómicas del cáncer.

La potencial extrapolación de esta metodología hacia el análisis de muestras de biopsia líquida, especialmente a través de experimentos multiplex, resalta su utilidad en entornos clínicos que requieren el análisis simultáneo de múltiples muestras. Este enfoque metodológico podría simplificar el análisis genómico y tener un impacto significativo en el abordaje clínico del cáncer de mama. La combinación del rendimiento de los datos y la precisión en la detección de CNAs observada en esta fase permite sugerir la implementación de la secuenciación con nanoporos en aplicaciones clínicas.

9.4 Detección de alteraciones en el número de copias de DNA en muestras de pacientes con cáncer de mama

9.4.1 Características clínico-patológicas de la cohorte de pacientes con cáncer de mama incluidas en el estudio

Tras la estandarización del enfoque experimental utilizando líneas celulares tumorales, la segunda fase de este trabajo se orientó hacia la aplicación de esta metodología en muestras clínicas de pacientes con cáncer de mama. Esto implicó la adaptación del flujo de trabajo para el análisis de muestras de biopsia líquida, específicamente plasma, un tipo de muestra cada vez más relevante en la práctica clínica por su potencial para obtener información genómica de manera no invasiva (Connal et al., 2023).

Para ello, se colectaron muestras pareadas de biopsias líquidas y biopsias de tejido tumoral de 20 pacientes con cáncer de mama, con la finalidad de comparar si la información genómica obtenida a partir de ellas era semejante entre sí. Las características clínico-patológicas de las pacientes, a partir de las cuales se obtuvieron las muestras, se detallan en la Tabla 9. De esta cohorte de pacientes, destaca la predominancia de casos en etapas clínicas tempranas (I-II= 95%), es decir, sin presencia de metástasis; la alta frecuencia de receptores hormonales positivos (ER= 85% y PR= 85%) y la prevalencia del subtipo Luminal A (70%).

Tabla 9. Características clínico-patológicas de la cohorte de pacientes con cáncer de mama.

Características clínico-patológicas		n (%)
Total (n = 20)		
Edad (años)	≤ 50	7 (35)
	> 50	13 (65)
Etapas clínicas	I – II	19 (95)
	III – IV	1 (5)
Tamaño tumoral (mm)	≤ 20	0 (0)
	> 20	20 (100)
ER	Positivo	17 (85)
	Negativo	3 (15)
PR	Positivo	17 (85)

Características clínico-patológicas		n (%)
Total (n = 20)		
	Negativo	3 (15)
HER2	Positivo	1 (5)
	Negativo	19 (95)
Ki-67	≤ 20 %	12 (60)
	> 20 %	8 (40)
Subtipo por IHQ	Luminal A	14 (70)
	Luminal B	3 (15)
	HER2 amplificado	0 (0)
	Triple negativo	3 (15)
Metástasis	Si	0 (0)
	No	100 (0)

La clasificación de los tumores de mama por inmunohistoquímica (IHQ) emplea biomarcadores comunes en la práctica clínica, como el receptor de estrógeno (ER), el receptor de progesterona (PR) y el receptor 2 del factor de crecimiento epidérmico humano (HER2). Los tumores Luminal A y Luminal B se caracterizan por la expresión de los receptores hormonales ER y PR, y difieren en la expresión del marcador Ki-67 y la presencia o ausencia del receptor HER2. Por otro lado, el subtipo HER2 exhibe un aumento de la proteína HER2, pero carece de receptores hormonales. Por último, los tumores triples negativos carecen de la expresión de los receptores ER, PR y HER2 (Guiu et al., 2012).

9.4.2 Detección de alteraciones en el número de copias de DNA en biopsias sólidas

En el ámbito clínico, los microarreglos de alta resolución son reconocidos como la herramienta principal para el análisis molecular de las biopsias de tejido tumoral mamario. Por consiguiente, esta técnica fue empleada como método de referencia para identificar las alteraciones cromosómicas presentes en las muestras de biopsias sólidas de las pacientes de esta cohorte. A pesar de haber realizado el análisis por microarreglos en todas las muestras, por razones que se detallarán más adelante, este trabajo se centra en la presentación de resultados de únicamente dos muestras pareadas, las cuales se identifican por su ID: **BL-147** y **BL-158**. En la Tabla 10 se muestran las características clínico-patológicas que distinguen a las dos pacientes de las cuales se colectaron estas muestras.

Tabla 10. Características clínico-patológicas de las pacientes con cáncer de mama seleccionadas para evaluación en el estudio.

Características clínico-patológicas	ID paciente	
	BL-147	BL-158
Edad (años)	54	46
Etapa clínica	II	II
Tamaño tumoral (mm)	32	49
ER	Positivo	Positivo
PR	Positivo	Positivo
HER2	Negativo	Negativo
Ki-67 (%)	30	10
Subtipo por IHQ	Luminal B	Luminal A
Metástasis	No	No

La clasificación de los tumores de mama por inmunohistoquímica (IHQ) emplea biomarcadores comunes en la práctica clínica, como el receptor de estrógeno (ER), el receptor de progesterona (PR) y el receptor 2 del factor de crecimiento epidérmico humano (HER2). Los tumores Luminal A y Luminal B se caracterizan por la expresión de los receptores hormonales ER y PR, y difieren en la expresión del marcador Ki-67 y la presencia o ausencia del receptor HER2. Por otro lado, el subtipo HER2 exhibe un aumento de la proteína HER2, pero carece de receptores hormonales. Por último, los tumores triples negativos carecen de la expresión de los receptores ER, PR y HER2 (Guiu et al., 2012).

9.4.2.1 Análisis de detección de alteraciones en el número de copias de DNA en la biopsia de tejido tumoral BL-147

Los resultados obtenidos a través del análisis por microarreglos en esta muestra biológica revelaron un conjunto de alteraciones cromosómicas en la paciente, destacando la presencia de 11 CNAs en cuatro cromosomas distintos (1, 11, 16 y 20). Particularmente, el cromosoma 11 exhibió la mayor carga de CNAs, lo que indica un nivel significativo de inestabilidad genómica en este cromosoma. Del total de CNAs detectadas, el 55% de ellas se categorizaron como patogénicas, es decir, que su presencia se asocia con el desarrollo de tumores. Las alteraciones mostraron el aumento en el número de copias en las regiones 1q21.1-q44 y 11q13.2-q13.4, y la pérdida en el número de copias en las regiones 11p11.12-q12.3, 11q12.3-q13.2, 11q14.1-q25 y 16p11.2-q24.3 (Tabla Suplementaria 11).

La clasificación molecular del cáncer de mama identifica cuatro subtipos intrínsecos: Luminal A, Luminal B, HER2 y Basal, los cuales están asociados con distintos pronósticos clínicos y estrategias de tratamiento (Harbeck et al., 2019). En particular, los subtipos Luminales (Luminal A y Luminal B) se distinguen por la ganancia en las regiones 1q, 8q, 8p11, 11q13.3 y la pérdida en las regiones 6q, 8p, 16q y 22q (Kaur et al., 2021; Koboldt et al., 2012). Esta paciente, clasificada como Luminal B HER2- mediante inmunohistoquímica, exhibió algunas de estas alteraciones, específicamente la ganancia en 1q y la pérdida en 16q. Este análisis de la biopsia sólida permitió identificar CNAs específicas asociadas a los tumores Luminales, validando así la concordancia entre la clasificación por inmunohistoquímica, a nivel de proteínas, y los patrones genómicos. Además, otro hallazgo interesante fue la identificación de alteraciones relevantes a nivel clínico que albergan genes potencialmente accionables, como *AKT3* (1q43-q44) y *CCND1* (11q13.3), ambos presentando una duplicación de 3 copias, mientras que *PGR* (11q22.1) y *ATM* (11q22.3) exhibieron la delección de una de las copias.

9.4.2.2 Análisis de detección de alteraciones en el número de copias de DNA en la biopsia de tejido tumoral BL-158

El análisis de esta muestra biológica reveló un panorama genómico complejo, evidenciando un número significativo de CNAs, alcanzando un total de 48, distribuidas en múltiples cromosomas (1, 5, 6, 8, 11, 14, 15, 16, 22). Una observación interesante fue que el 54% de estas alteraciones se concentró en el cromosoma 1, mostrando tanto ganancias como pérdidas en regiones cromosómicas específicas, lo que sugiere la alta inestabilidad cromosómica presente en este. La mayoría de las alteraciones ubicadas en este cromosoma se clasificaron como duplicaciones de múltiples copias, presentando aumentos de 4 hasta 10 copias, aunque solo una de ellas mostró ser patogénica. Por otro lado, las alteraciones restantes se identificaron como delecciones que impactaron en ambos brazos cromosómicos (p y q), y todas ellas se categorizaron como patogénicas (Tabla Suplementaria 12).

Cabe destacar que otros cromosomas, como el 11 y el 22, también exhibieron una carga considerable de CNAs, las cuales generaron pérdidas en el número de copias de regiones localizadas en los brazos cromosómicos largos (q), todas identificadas como patogénicas (Tabla Suplementaria 12). De forma general, el 29% de las 48 CNAs que se detectaron se identificaron como patogénicas y, a excepción de una, todas correspondían a delecciones de una región cromosómica. Este hallazgo sugiere que las alteraciones de pérdida en el número de copias desempeñan un papel crucial en el proceso de carcinogénesis de esta paciente, posiblemente contribuyendo al fenotipo aberrante observado.

La paciente fue categorizada dentro del subtipo Luminal A mediante la evaluación por inmunohistoquímica. Las alteraciones genómicas halladas concordaron de manera consistente con el perfil de CNAs observado en los subtipos moleculares Luminales (Kaur et al., 2021; Koboldt et al., 2012), encontrándose un aumento en el número de copias en 1q, 8q, 8p11, y una disminución en 6q, 16q y 22q. Además, el análisis reveló una serie de genes con potencial relevancia clínica dentro de las alteraciones cromosómicas identificadas. Entre estos, se detectó la ganancia del gen *MYC* (8q24.21), con un número de copias igual a 4, el cual ha sido asociado con un papel crucial en la regulación de la proliferación celular y la progresión tumoral en varios tipos de cáncer, incluido el cáncer de mama (Schaub et al., 2018; Xu et al., 2010). Por otro lado, se encontraron deleciones en genes como *CCND1* (11q13.3), *PGR* (11q22.1), *ATM* (11q22.3) y *AKT1* (14q32.33), todos ellos mostrando la pérdida de una de sus copias. Es importante señalar que, si bien el gen *CCND1* suele estar amplificado en hasta un 35% de los tumores de mama, la presencia de deleciones significativas en este gen se ha descrito en algunos tumores del subtipo Luminal A (Kaur et al., 2021).

A pesar de que ambas pacientes se encuentran en estadios tempranos de la enfermedad, se evidenció una diferencia en la carga de alteraciones cromosómicas entre ellas. La paciente BL-158 exhibió un número notablemente mayor de CNAs en comparación con la paciente BL-147. Esta diferencia sugiere que las alteraciones cromosómicas pueden desempeñar un papel más relevante en el desarrollo y progresión del cáncer de mama en el caso de BL-158. Por otro lado, en el caso de la paciente BL-147, otro tipo de alteraciones moleculares podrían estar ejerciendo una influencia más destacada en el desarrollo de la enfermedad.

Además, el análisis mediante la tecnología de microarreglos permitió identificar CNAs asociadas a subtipos moleculares específicos del cáncer de mama, mostrando una notable concordancia con el diagnóstico inmunohistoquímico realizado en la clínica, y también reveló genes identificados como potencialmente accionables. Esto resalta la relevancia del análisis de las CNAs, ya que provee información crucial para el diagnóstico, la clasificación en subtipos moleculares, la predicción de la supervivencia y la posibilidad de establecer estrategias terapéuticas más precisas y personalizadas para las pacientes.

9.4.3 Detección de alteraciones en el número de copias de DNA en biopsias líquidas

Posterior a llevar a cabo el análisis de detección de CNAs en las biopsias de tejido tumoral, se procedió a realizar este mismo análisis en las biopsias líquidas pareadas de las pacientes. Esto con el objetivo de evaluar la concordancia entre ambos tipos de muestras, buscando no sólo

corroborar los hallazgos previos, sino también explorar la viabilidad de la tecnología de secuenciación con nanoporos, bajo un enfoque de baja profundidad, en el análisis de muestras menos invasivas con potencial relevancia clínica.

9.4.3.1 Aislamiento de DNA libre circulante a partir de muestras de plasma de pacientes con cáncer de mama

La extracción de cfDNA a partir de la fracción plasmática de las muestras de biopsia líquida, reveló patrones de fragmentación en las moléculas de DNA aisladas. Se observó consistentemente un pico predominante entre 180 y 200 pb en todas las muestras (Figura 15). Esto concuerda con la literatura, dado que se ha descrito que el tamaño promedio del cfDNA es alrededor de 166 pb, pero este puede variar desde 80 hasta 200 pb (De Rubis et al., 2019). Este hallazgo se relaciona con el proceso de degradación del DNA por apoptosis, que se considera el principal mecanismo de liberación de componentes hacia la circulación sanguínea (Gao et al., 2022). Además, en algunas muestras, se observaron picos correspondientes a fragmentos de DNA múltiplos del pico principal, sugiriendo la presencia de fragmentos asociados a estructuras como mononucleosomas, dinucleosomas y trinucleosomas (Ungerer et al., 2022). Estos hallazgos respaldan que el cfDNA liberado por apoptosis puede abarcar una variedad de tamaños relacionados con la organización estructural del DNA en las células.

En la actualidad, existe evidencia de que el cfDNA derivado de pacientes con cáncer tiende a estar más fragmentado que el de individuos sanos, exhibiendo una proporción significativamente mayor de fragmentos inferiores a 145 pb (Jiang et al., 2015; Mouliere et al., 2018; Underhill et al., 2016). Sin embargo, nuestros datos no muestran la presencia de estos fragmentos más pequeños, comúnmente asociados a la presencia de ctDNA. Esto podría atribuirse a las características particulares de nuestra cohorte de estudio, donde predominan las pacientes en etapas tempranas de la enfermedad. De acuerdo con Bettegowda et al. (2014), la fracción de ctDNA varía dependiendo de las características específicas del tumor, como el estadio clínico, el tamaño y la vascularización, observándose que en etapas tempranas de la enfermedad la proporción puede ser menor al 0.1%, en relación con el cfDNA (Diehl et al., 2008). Esto podría justificar lo observado en estos resultados, sugiriendo que el material genético aislado de las muestras de biopsias líquidas probablemente corresponde en su mayoría a cfDNA originado, principalmente, de células normales.

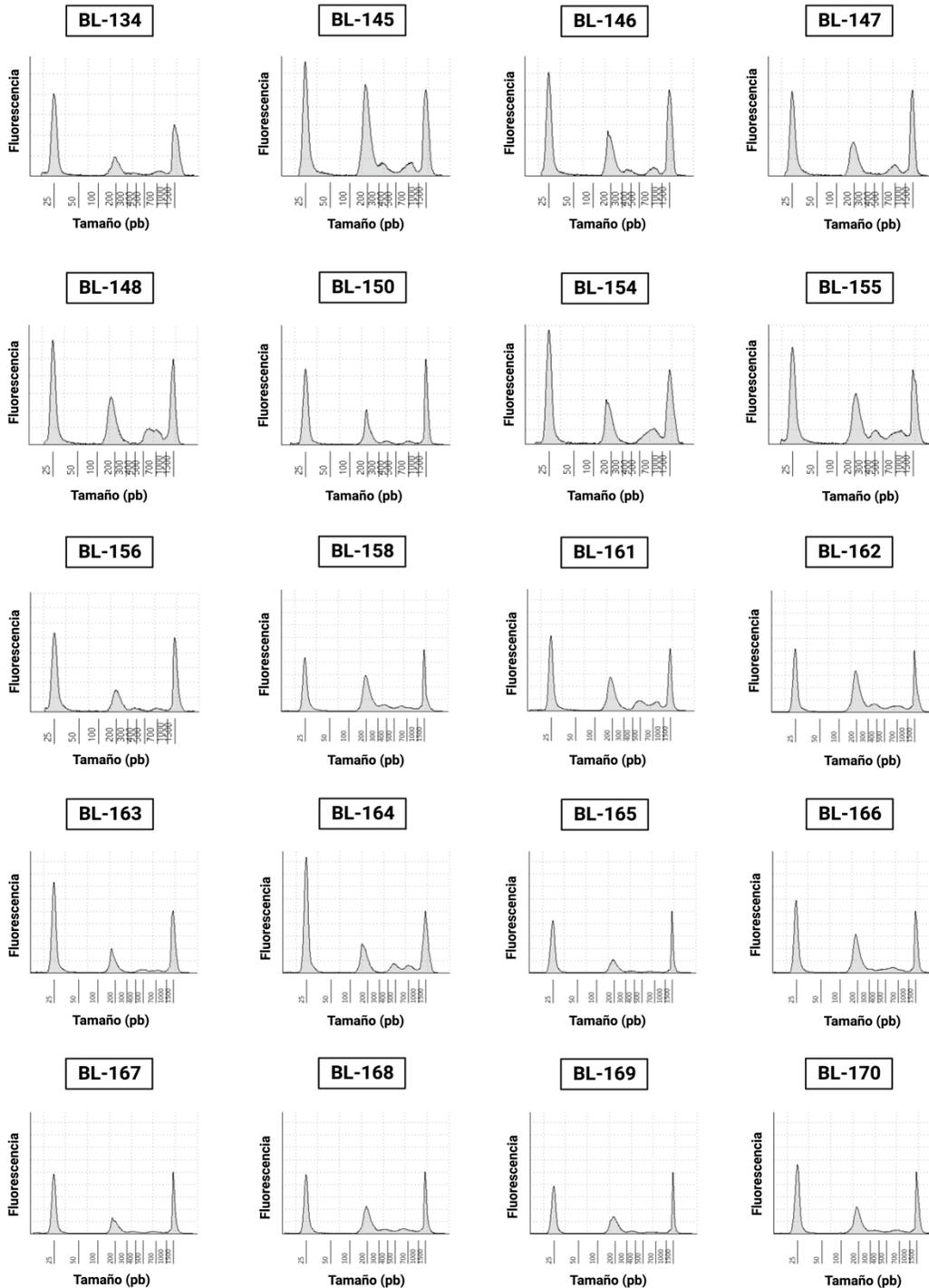


Figura 15. Análisis del tamaño de los fragmentos de DNA aislados en muestras de biopsia líquida de pacientes con cáncer de mama. En esta figura se presentan los electroferogramas que muestran el tamaño de los fragmentos de DNA obtenidos a partir de la extracción de cfDNA de las 20 muestras de pacientes. Para cada gráfica, se presenta el identificador de la paciente, en el eje de las abscisas se muestra el tamaño en pb y el eje de las ordenadas se muestra la intensidad de fluorescencia correspondiente a los diversos fragmentos de DNA. En los extremos se incluyen los picos correspondientes a los marcadores de peso molecular bajo y alto (25 y 1500 pb), que se utilizan como referencia para estimar el tamaño de los fragmentos.

En cuanto a la cantidad de cfDNA obtenido, para las 20 muestras se observó una variabilidad entre 4 y 40 ng totales, con un promedio de 19 ng (resultados no mostrados). La secuenciación de fragmentos cortos de DNA mediante la tecnología de secuenciación con nanoporos establece requerimientos específicos, puesto que sus protocolos requieren al menos 100-200 fmol de DNA de partida, equivalente a aproximadamente 30 a 60 ng de material genético. Sin embargo, el trabajo de Martignano et al. (2021) ofrece una perspectiva valiosa al demostrar que a partir de 15 ng de cfDNA se logra un rendimiento de secuenciación, en términos de la cantidad de lecturas obtenidas, similar al obtenido con 30 o 60 ng. Esto sugiere que la cantidad de cfDNA extraída podría ser suficiente para llevar a cabo el proceso de secuenciación en al menos el 77% de las muestras de las pacientes.

9.4.3.2 Secuenciación genómica con nanoporos de DNA libre circulante

Posteriormente, se realizó un análisis preliminar del rendimiento de la secuenciación del cfDNA extraído de muestras de biopsias líquidas de pacientes con cáncer de mama, con el objetivo de comparar dicho rendimiento con los experimentos previos llevados a cabo en líneas celulares. Para este propósito, se seleccionaron dos muestras: BL-147 y BL-158. Estas muestras presentaban una cantidad de 15 ng de cfDNA y contaban con información sobre las CNAs presentes en el tejido tumoral pareado a través de la tecnología de microarreglos de alta resolución. Tal como se mencionó anteriormente, esta cantidad (15 ng de cfDNA) ha sido sugerida por Martignano et al. (2021) como suficiente para aplicar la tecnología de secuenciación con nanoporos en este tipo de muestras. Ambas muestras fueron sometidas a un experimento de secuenciación singleplex con una duración de 24 horas.

De acuerdo con los datos presentados en la Tabla 11, se observó consistencia en el rendimiento de secuenciación entre ambas muestras, generando cada una más de 13 millones de lecturas crudas, con una mediana de tamaño menor a 300 pb. En ambas muestras, más del 80% de las lecturas cumplió con la métrica del control de calidad necesaria para el subsiguiente análisis de detección de CNAs. En particular, la muestra BL-147 generó un total de 11,710,459 lecturas, mientras que la muestra BL-158 generó 10,482,203 lecturas. Ambas muestras alcanzaron una profundidad de cobertura de 0.9x, correspondiente al enfoque sWGS.

Estos resultados indican un rendimiento de secuenciación adecuado para llevar a cabo el análisis de detección de CNAs, lo cual concuerda con los resultados de los experimentos previos con líneas celulares. Además, al comparar este rendimiento con el obtenido por Martignano et al.

(2021) en muestras de biopsia líquida de cáncer de pulmón, se observa una similitud notable. Estos hallazgos sugieren que la tecnología de secuenciación con nanoporos puede ser útil para identificar alteraciones genómicas relevantes en las muestras de las pacientes, tal como se logró realizar en el análisis de líneas celulares tumorales.

Tabla 11. Rendimiento y estadísticas de la secuenciación genómica con nanoporos llevada a cabo en muestras de biopsia líquida utilizando la plataforma de secuenciación GridION X5.

Muestra	Experimento	Tiempo (h)	Lecturas totales (M)	Qscore	Lecturas cumplen QC (M), (%)	Mediana tamaño de lectura (pb)	Profundidad de cobertura
BL-147	Singleplex	24	13.48	13.8	11.71 (86.9)	238	0.9x
BL-158	Singleplex	24	13.07	12.2	10.48 (80.2)	291	0.9x

M= millones; QC= Control de calidad; pb= pares de bases

9.4.3.3 Análisis de detección de alteraciones en el número de copias de DNA en las biopsias líquidas BL-147 y BL-158

Las muestras secuenciadas se analizaron para identificar su patrón de CNAs y evaluar la concordancia con las biopsias de tejido tumoral pareadas. En la muestra BL-147 se detectaron 25 CNAs localizadas en los cromosomas 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 16, 17 y 20 (Tabla Suplementaria 13), mientras que en la muestra BL-158 se identificaron 11 CNAs en los cromosomas 2, 4, 6, 7, 10, 12, 16 y 17 (Tabla Suplementaria 14). Sin embargo, al comparar las regiones genómicas alteradas en estas muestras con las biopsias de tejido tumoral correspondientes, se observó que ninguna de las alteraciones coincidía en ubicación, tamaño o estado de número de copias (Figura 16). Además, no se identificaron las alteraciones característicamente asociadas a los tumores clasificados dentro del subtipo Luminal, tal como se observó al analizar las muestras de biopsia sólida. Las representaciones gráficas de los cariotipos moleculares sugieren que en ambos casos, el genoma parece ser diploide (Figura 16).

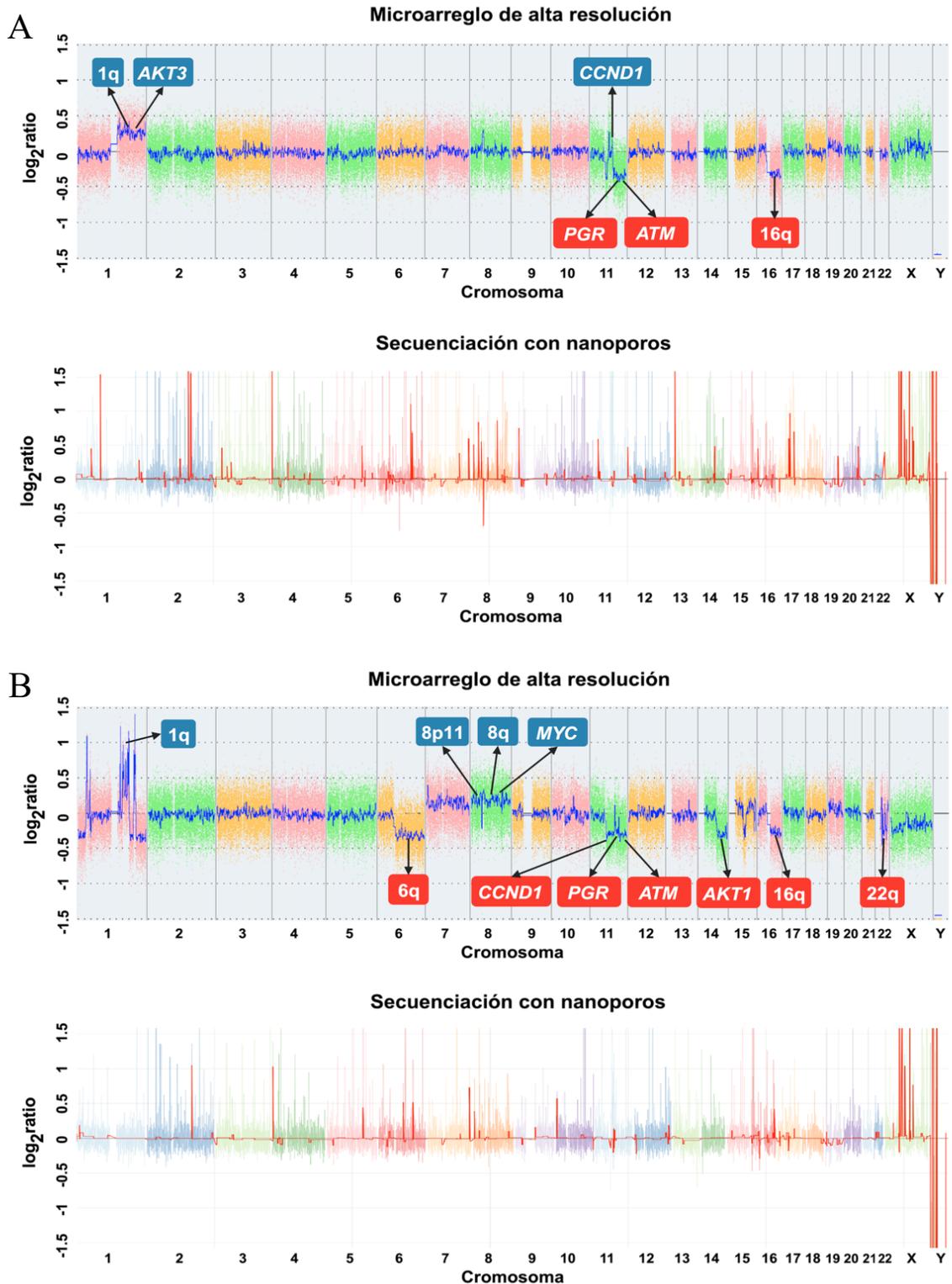


Figura 16. Análisis comparativo de alteraciones en el número de copias de DNA presentes en biopsias de tejido tumoral y biopsias líquidas en pacientes con cáncer de mama. En esta figura se presentan los cariotipos moleculares obtenidos a través del análisis de muestras biológicas de pacientes con cáncer de mama, donde A) representa a la paciente BL-147 y B) representa a la paciente BL-158. En el gráfico superior se muestran los resultados

generados con los datos de microarreglos de alta resolución al analizar la biopsia de tejido tumoral, mientras que en el gráfico inferior se presentan los resultados obtenidos mediante secuenciación con nanoporos al analizar la biopsia líquida. En el eje de las abscisas se muestra la posición cromosómica. En el eje de las ordenadas, la línea azul (microarreglos de alta resolución) o roja (secuenciación con nanoporos) indica el perfil de segmentación en dicho cromosoma. Un valor de \log_2 ratio de 0 indica un número de copias normal o diploide. El aumento o la disminución del valor de \log_2 ratio corresponde a regiones genómicas con ganancia o pérdida del número de copias, respectivamente. En los cariotipos moleculares generados a partir de datos de microarreglos, los recuadros azules destacan las regiones o genes con relevancia clínica con alteraciones positivas, mientras que los rojos señalan aquellas y aquellos con alteraciones negativas.

De forma interesante, algunas de las alteraciones detectadas se compartieron entre las dos muestras de biopsias líquidas, como las regiones citogenéticas 2q24.2, 4p16.3, 6q22.31, 7q36.1 y 17q21.31. Estos resultados, en conjunto con observaciones previas que sugieren que el material genético aislado de las muestras probablemente proviene en su mayoría de cfDNA de células normales, apuntan a la posibilidad de que las alteraciones detectadas sean variantes en el número de copias (CNVs), es decir, duplicaciones o deleciones en el genoma que están presentes en los individuos y que alteran el número de copias de las secuencias de DNA, pero que no necesariamente tienen un impacto en el fenotipo.

Para validar esta hipótesis, se buscaron todas estas regiones cromosómicas en la base de datos *Database of Genomic Variants* (DGV, <http://dgv.tcag.ca/dgv/app/home>), un esfuerzo genómico curado que cataloga las variaciones estructurales presentes en el genoma humano de individuos sanos y su relación con el fenotipo. La búsqueda reveló que todas las regiones genómicas detectadas en las muestras están descritas por poseer CNVs, ya sea de ganancia o pérdida de material genético, concordante con lo detectado en este trabajo. Además, en la mayoría de ellas, se encontró la presencia de uno o más genes dentro de la región. Si bien sólo algunas de estas regiones están asociadas con enfermedades, ninguna es relacionada con el cáncer de mama.

Por ejemplo, en la paciente BL-147, se identificó que los genes en la región cromosómica 4p16.3, como *TACC3* y *FGFR3*, tienen asociación con diversos tipos de cáncer, como vejiga, cervical, colorrectal y glioblastomas, así como con acondroplasia y camptodactilia, entre otras enfermedades (Deng et al., 1996; Frattini et al., 2018; Makrythanasis et al., 2014). Además, en la región cromosómica 8p12, se localizó el gen *NGR1*, asociado con la susceptibilidad a esquizofrenia (Stefansson et al., 2002), en tanto que en la región cromosómica 11q23.3, el gen *NECTIN1* está asociado con el síndrome de labio leporino y paladar hendido (Sözen et al., 2001).

En cuanto a la paciente BL-158, las CNVs identificadas con una asociación a enfermedades se localizaron en 4p16.3 (similar a BL-147), 6p22.1-p21.32 y 17q21.31. En la región 6p22.1-p21.32, se identificó un alto número de genes, los cuales están asociados a condiciones como narcolepsia, psoriasis, sordera, susceptibilidad a malaria, entre otras enfermedades (Chen et al., 2005; Delahaye et al., 2007; Hor et al., 2011; Nair et al., 2006); mientras que en la región 17q21.31 se identificó a los genes *PLEKHM1* y *MAPT*, los cuales están asociados con osteopetrosis y susceptibilidad a la enfermedad de Parkinson, respectivamente (Bo et al., 2016; Elbaz et al., 2011).

Estos hallazgos plantean la posibilidad de que las muestras de biopsias líquidas estén predominantemente compuestas por cfDNA de origen normal, lo que sugiere una presencia baja de cfDNA tumoral o ctDNA. Esta baja presencia podría ser el factor limitante que impide obtener perfiles de CNAs en estas muestras, dando lugar a resultados negativos. Además, la baja profundidad de los datos, inferior a 1x, podría justificar la dificultad para identificar estas alteraciones, especialmente en comparación con los experimentos realizados en las líneas celulares tumorales, donde se partía de células genéticamente homogéneas. En este sentido, resulta fundamental reconocer la complejidad de las muestras de biopsias líquidas, que contienen tanto cfDNA normal como ctDNA, lo cual añade desafíos para la detección precisa de CNAs, ya que las alteraciones de origen tumoral pueden perderse en el fondo del cfDNA normal. Frente a este escenario, una estrategia apropiada para eliminar las señales generadas por las posibles CNVs detectadas y facilitar la interpretación de resultados acerca de las alteraciones genómicas presentes en las pacientes, sería obtener las muestras normales pareadas de cada una de ellas. Estas muestras permitirían eliminar las variaciones genómicas comunes entre una muestra de referencia y una muestra problema, brindando una base más sólida para la interpretación del perfil de número de copias obtenido en cada paciente.

La decisión de no continuar con el análisis de las muestras de biopsias líquidas del resto de las pacientes en la cohorte se basa en la anticipación de resultados similares debido a la predominancia de pacientes en estadios clínicos tempranos al momento de la toma de las muestras. Esto subraya la importancia de considerar el estadio clínico al diseñar estudios de biopsias líquidas, particularmente en trabajos como este que tienen un enfoque de secuenciación de genoma completo de baja profundidad, ya que las características genómicas pueden variar significativamente entre estadios y afectar la capacidad de detección de CNAs.

El conjunto de resultados obtenidos en la segunda fase de este trabajo permiten rechazar la hipótesis planteada respecto a la capacidad de la tecnología de secuenciación genómica con nanoporos para identificar CNAs de manera comparable a los resultados obtenidos en biopsias sólidas pareadas, analizadas mediante microarreglos de alta resolución. No obstante, es importante recalcar que esta conclusión es únicamente aplicable bajo el contexto de este trabajo, el cual se centró en analizar muestras de pacientes con cáncer de mama en estadios tempranos de la enfermedad.

A pesar de los resultados desfavorables obtenidos en esta investigación, es importante destacar que la biopsia líquida sigue siendo una herramienta prometedora a nivel clínico, especialmente en pacientes metastásicos, donde su utilidad ya ha sido demostrada en trabajos anteriores como los de Martignano et al. (2021) y Kim et al. (2023). De este modo, la aplicación de la secuenciación con nanoporos podría ser viable para analizar muestras de biopsia líquida en pacientes con cáncer de mama en estadios metastásicos. Este enfoque podría permitir la detección de regiones cromosómicas relevantes que albergan genes con funciones oncogénicas o de supresión tumoral, proporcionando información crucial para la toma de decisiones a nivel clínico en estas pacientes.

10 CONCLUSIONES

- La aplicación de la tecnología de secuenciación genómica con nanoporos, a través de un enfoque de secuenciación de genoma completo de baja profundidad, demostró ser factible para la detección precisa de CNAs en líneas celulares tumorales de mama. Este abordaje exhibió concordancia con los resultados obtenidos mediante la tecnología de secuenciación Illumina, resaltando su capacidad para identificar de forma precisa genes con potencial relevancia clínica para el cáncer de mama.
- Al aplicar este enfoque en el análisis de muestras de biopsias líquidas de pacientes con cáncer de mama en estadios tempranos de la enfermedad, se evidenció una sensibilidad limitada para la detección de CNAs, posiblemente debido a la baja presencia de ctDNA en las muestras.
- A pesar de lograr la detección de alteraciones cromosómicas específicas del subtipo molecular de las pacientes en las biopsias de tejido tumoral, la baja presencia de material genético de origen tumoral en las biopsias líquidas influyó negativamente en la capacidad de detección de alteraciones cromosómicas, lo que resultó en la falta de concordancia entre ambos tipos de biopsias.

11 PERSPECTIVAS

A pesar de los desafíos encontrados en la detección de CNAs en biopsias líquidas de pacientes con cáncer de mama en estadios tempranos, este trabajo sienta una base sólida para continuar explorando la aplicación de la tecnología de secuenciación con nanoporos en la identificación de alteraciones cromosómicas clínicamente relevantes en este tipo de muestra. Para continuar el trabajo, se propone lo siguiente:

- Evaluar si el aumento de la profundidad de la secuenciación genómicas con nanoporos mejora la detección de ctDNA en muestras de biopsias líquidas de pacientes en estadios tempranos del cáncer de mama. Así mismo, aplicar estrategias de enriquecimiento específicas para ctDNA para mejorar la sensibilidad en la detección de alteraciones genómicas partiendo de muestras con baja concentración de ctDNA.
- Analizar muestras pareadas de biopsias sólidas y biopsias líquidas empleando la tecnología de secuenciación genómica con nanoporos para evaluar si la dificultad en la detección de CNAs radica en el enfoque de baja profundidad de la tecnología o si se debe a las características intrínsecas de las muestras de biopsias líquidas.
- Dirigir la investigación hacia el análisis de muestras de biopsias líquidas provenientes de pacientes con cáncer de mama en estadios avanzados de la enfermedad, con el objetivo de determinar la capacidad para detectar alteraciones cromosómicas clínicamente relevantes utilizando este enfoque.

12 REFERENCIAS

- Adewale, B. A. (2020). Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr J Lab Med*, 9(1), 1340. <https://doi.org/10.4102/ajlm.v9i1.1340>
- Ahn, J. W., Coldwell, M., Bint, S., & Mackie Ogilvie, C. (2015). Array comparative genomic hybridization (array CGH) for detection of genomic copy number variants. *J Vis Exp*(96), e51718. <https://doi.org/10.3791/51718>
- Alanazi, L., Alqahtani, R. N., Masud, N., Zuraie, M. M., Bin Aff, A. A., & Alanazi, S. H. (2022). The Role of Tissue and Liquid Biopsy in the Clinical Management of Adult Lung Cancer Patients in King Abdul-Aziz Medical City in Riyadh, Saudi Arabia. *Cureus*, 14(1), e20914. <https://doi.org/10.7759/cureus.20914>
- Alba-Bernal, A., Lavado-Valenzuela, R., Domínguez-Recio, M. E., Jiménez-Rodríguez, B., Queipo-Ortuño, M. I., Alba, E., & Comino-Méndez, I. (2020). Challenges and achievements of liquid biopsy technologies employed in early breast cancer. *EBioMedicine*, 62, 103100. <https://doi.org/10.1016/j.ebiom.2020.103100>
- Alimirzaie, S., Bagherzadeh, M., & Akbari, M. R. (2019). Liquid biopsy in breast cancer: A comprehensive review. *Clin Genet*, 95(6), 643-660. <https://doi.org/10.1111/cge.13514>
- Ambriz-Barrera, F., Rojas-Jimenez, E., Díaz-Velásquez, C. E., De-La-Cruz-Montoya, A. H., Martínez-Gregorio, H. C., Ruiz-De-La-Cruz, M., Huertas, A., Montealegre, A. L., Castro-Rojas, C., Acosta, G., Vaca-Paniagua, F., & Perdomo, S. (2023). Mutational spectrum of breast cancer by shallow whole-genome sequencing of cfDNA and tumor gene panel analysis. In: Research Square.
- Arteaga, C. L., Sliwkowski, M. X., Osborne, C. K., Perez, E. A., Puglisi, F., & Gianni, L. (2011). Treatment of HER2-positive breast cancer: current status and future perspectives. *Nat Rev Clin Oncol*, 9(1), 16-32. <https://doi.org/10.1038/nrclinonc.2011.177>
- Aucamp, J., Bronkhorst, A. J., Badenhorst, C. P. S., & Pretorius, P. J. (2018). The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature. *Biol Rev Camb Philos Soc*, 93(3), 1649-1683. <https://doi.org/10.1111/brv.12413>
- Baker, S. J., Poulikakos, P. I., Irie, H. Y., Parekh, S., & Reddy, E. P. (2022). CDK4: a master regulator of the cell cycle and its role in cancer. *Genes Cancer*, 13, 21-45. <https://doi.org/10.18632/genesandcancer.221>
- Balajee, A. S., & Hande, M. P. (2018). History and evolution of cytogenetic techniques: Current and future applications in basic and clinical research. *Mutat Res Genet Toxicol Environ Mutagen*, 836(Pt A), 3-12. <https://doi.org/10.1016/j.mrgentox.2018.08.008>
- Bartek, J., & Lukas, J. (2007). DNA damage checkpoints: from initiation to recovery or adaptation. *Curr Opin Cell Biol*, 19(2), 238-245. <https://doi.org/10.1016/j.ceb.2007.02.009>
- Baslan, T., Kovaka, S., Sedlazeck, F. J., Zhang, Y., Wappel, R., Tian, S., Lowe, Scott W., Goodwin, S., & Schatz, Michael C. (2021). High resolution copy number inference in cancer using short-molecule nanopore sequencing. *Nucleic Acids Research*, 49(21), e124-e124. <https://doi.org/10.1093/nar/gkab812>
- Beaver, J. A., Jelovac, D., Balukrishna, S., Cochran, R., Croessmann, S., Zabransky, D. J., Wong, H. Y., Toro, P. V., Cidado, J., Blair, B. G., Chu, D., Burns, T., Higgins, M. J., Stearns, V., Jacobs, L., Habibi, M., Lange, J., Hurley, P. J., Lauring, J., . . . Park, B. H. (2014). Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin Cancer Res*, 20(10), 2643-2650. <https://doi.org/10.1158/1078-0432.Ccr-13-2933>
- Bell, A. D., Usher, C. L., & McCarroll, S. A. (2018). Analyzing Copy Number Variation with Droplet Digital PCR. *Methods Mol Biol*, 1768, 143-160. https://doi.org/10.1007/978-1-4939-7778-9_9
- Beroukhi, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., Mc Henry, K. T., Pinchback, R. M., Ligon, A. H., Cho, Y.-J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M. S., . . . Meyerson, M. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, 463(7283), 899-905. <https://doi.org/10.1038/nature08822>
- Bettgowda, C., Sausen, M., Leary, R. J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B. R., Wang, H., Luber, B., Alani, R. M., Antonarakis, E. S., Azad, N. S., Bardelli, A., Brem, H., Cameron, J. L., Lee, C. C., Fecher, L. A., Gallia, G. L., Gibbs, P., . . . Diaz, L. A., Jr. (2014). Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*, 6(224), 224ra224. <https://doi.org/10.1126/scitranslmed.3007094>
- Bierkens, M., Krijgsman, O., Wilting, S. M., Bosch, L., Jaspers, A., Meijer, G. A., Meijer, C. J., Snijders, P. J., Ylstra, B., & Steenbergen, R. D. (2013). Focal aberrations indicate EYA2 and hsa-miR-375 as

- oncogene and tumor suppressor in cervical carcinogenesis. *Genes Chromosomes Cancer*, 52(1), 56-68. <https://doi.org/10.1002/gcc.22006>
- Bo, T., Yan, F., Guo, J., Lin, X., Zhang, H., Guan, Q., Wang, H., Fang, L., Gao, L., Zhao, J., & Xu, C. (2016). Characterization of a Relatively Malignant Form of Osteopetrosis Caused by a Novel Mutation in the PLEKHM1 Gene. *J Bone Miner Res*, 31(11), 1979-1987. <https://doi.org/10.1002/jbmr.2885>
- Bogaerts, B., Van den Bossche, A., Verhaegen, B., Delbrassinne, L., Mattheus, W., Nouws, S., Godfroid, M., Hoffman, S., Roosens, N. H. C., De Keersmaecker, S. C. J., & Vanneste, K. (2024). Closing the gap: Oxford Nanopore Technologies R10 sequencing allows comparable results to Illumina sequencing for SNP-based outbreak investigation of bacterial pathogens. *J Clin Microbiol*, e0157623. <https://doi.org/10.1128/jcm.01576-23>
- Bronkhorst, A. J., Ungerer, V., & Holdenrieder, S. (2019). The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol Detect Quantif*, 17, 100087. <https://doi.org/10.1016/j.bdq.2019.100087>
- Bueno, R. C., Canevari, R. A., Villacis, R. A. R., Domingues, M. A. C., Caldeira, J. R. F., Rocha, R. M., Drigo, S. A., & Rogatto, S. R. (2014). ATM down-regulation is associated with poor prognosis in sporadic breast carcinomas. *Annals of Oncology*, 25(1), 69-75. <https://doi.org/10.1093/annonc/mdt421>
- Carter, N. P. (2007). Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet*, 39(7 Suppl), S16-21. <https://doi.org/10.1038/ng2028>
- Chat, V., Ferguson, R., Morales, L., & Kirchhoff, T. (2022). Ultra Low-Coverage Whole-Genome Sequencing as an Alternative to Genotyping Arrays in Genome-Wide Association Studies [Original Research]. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.790445>
- Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen*, 58(5), 235-263. <https://doi.org/10.1002/em.22087>
- Chen, J., & Xu, F. (2023). Application of Nanopore Sequencing in the Diagnosis and Treatment of Pulmonary Infections. *Molecular Diagnosis & Therapy*, 27(6), 685-701. <https://doi.org/10.1007/s40291-023-00669-8>
- Chen, W., Kahrizi, K., Meyer, N. C., Riazalhosseini, Y., Van Camp, G., Najmabadi, H., & Smith, R. J. (2005). Mutation of COL11A2 causes autosomal recessive non-syndromic hearing loss at the DFNB53 locus. *J Med Genet*, 42(10), e61. <https://doi.org/10.1136/jmg.2005.032615>
- Chin, S. F., Santonja, A., Grzelak, M., Ahn, S., Sammut, S. J., Clifford, H., Rueda, O. M., Pugh, M., Goldgraben, M. A., Bardwell, H. A., Cho, E. Y., Provenzano, E., Rojo, F., Alba, E., & Caldas, C. (2018). Shallow whole genome sequencing for robust copy number profiling of formalin-fixed paraffin-embedded breast cancers. *Exp Mol Pathol*, 104(3), 161-169. <https://doi.org/10.1016/j.yexmp.2018.03.006>
- Ciriello, G., Miller, M. L., Aksoy, B. A., Senbabaoglu, Y., Schultz, N., & Sander, C. (2013). Emerging landscape of oncogenic signatures across human cancers. *Nature Genetics*, 45(10), 1127-1133. <https://doi.org/10.1038/ng.2762>
- Cisneros-Villanueva, M., Hidalgo-Pérez, L., Rios-Romero, M., Cedro-Tanda, A., Ruiz-Villavicencio, C. A., Page, K., Hastings, R., Fernandez-Garcia, D., Allsopp, R., Fonseca-Montaño, M. A., Jimenez-Morales, S., Padilla-Palma, V., Shaw, J. A., & Hidalgo-Miranda, A. (2022). Cell-free DNA analysis in current cancer clinical trials: a review. *Br J Cancer*, 126(3), 391-400. <https://doi.org/10.1038/s41416-021-01696-0>
- Cohen, J. D., Li, L., Wang, Y., Thoburn, C., Afsari, B., Danilova, L., Douville, C., Javed, A. A., Wong, F., Mattox, A., Hruban, R. H., Wolfgang, C. L., Goggins, M. G., Dal Molin, M., Wang, T. L., Roden, R., Klein, A. P., Ptak, J., Dobbyn, L., . . . Papadopoulos, N. (2018). Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*, 359(6378), 926-930. <https://doi.org/10.1126/science.aar3247>
- Connal, S., Cameron, J. M., Sala, A., Brennan, P. M., Palmer, D. S., Palmer, J. D., Perlow, H., & Baker, M. J. (2023). Liquid biopsies: the future of cancer early detection. *J Transl Med*, 21(1), 118. <https://doi.org/10.1186/s12967-023-03960-8>
- Coombes, R. C., Page, K., Salari, R., Hastings, R. K., Armstrong, A., Ahmed, S., Ali, S., Cleator, S., Kenny, L., Stebbing, J., Rutherford, M., Sethi, H., Boydell, A., Swenerton, R., Fernandez-Garcia, D., Gleason, K. L. T., Goddard, K., Guttery, D. S., Assaf, Z. J., . . . Shaw, J. A. (2019). Personalized Detection of Circulating Tumor DNA Antedates Breast Cancer Metastatic Recurrence. *Clin Cancer Res*, 25(14), 4255-4263. <https://doi.org/10.1158/1078-0432.Ccr-18-3663>
- Cozzuto, L., Liu, H., Prysycz, L. P., Pulido, T. H., Delgado-Tejedor, A., Ponomarenko, J., & Novoa, E. M. (2020). MasterOfPores: A Workflow for the Analysis of Oxford Nanopore Direct RNA Sequencing Datasets. *Front Genet*, 11, 211. <https://doi.org/10.3389/fgene.2020.00211>

- Cretu Stancu, M., van Roosmalen, M. J., Renkens, I., Nieboer, M. M., Middelkamp, S., de Ligt, J., Pregno, G., Giachino, D., Mandrile, G., Espejo Valle-Inclan, J., Korzelius, J., de Bruijn, E., Cuppen, E., Talkowski, M. E., Marschall, T., de Ridder, J., & Kloosterman, W. P. (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nature Communications*, 8(1), 1326. <https://doi.org/10.1038/s41467-017-01343-4>
- Cuber, P., Chooneea, D., Geeves, C., Salatino, S., Creedy, T. J., Griffin, C., Sivess, L., Barnes, I., Price, B., & Misra, R. (2023). Comparing the accuracy and efficiency of third generation sequencing technologies, Oxford Nanopore Technologies, and Pacific Biosciences, for DNA barcode sequencing applications. *Ecological Genetics and Genomics*, 28, 100181. <https://doi.org/https://doi.org/10.1016/j.egg.2023.100181>
- Cui, C., Shu, W., & Li, P. (2016). Fluorescence In situ Hybridization: Cell-Based Genetic Diagnostic and Research Applications [Review]. *Frontiers in Cell and Developmental Biology*, 4. <https://doi.org/10.3389/fcell.2016.00089>
- Cunningham, C. E., MacAuley, M. J., Yadav, G., Vizeacoumar, F. S., Freywald, A., & Vizeacoumar, F. J. (2019). Targeting the CINful genome: Strategies to overcome tumor heterogeneity. *Prog Biophys Mol Biol*, 147, 77-91. <https://doi.org/10.1016/j.pbiomolbio.2019.02.006>
- De Rubis, G., Rajeev Krishnan, S., & Bebawy, M. (2019). Liquid Biopsies in Cancer Diagnosis, Monitoring, and Prognosis. *Trends Pharmacol Sci*, 40(3), 172-186. <https://doi.org/10.1016/j.tips.2019.01.006>
- Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nat Biotechnol*, 34(5), 518-524. <https://doi.org/10.1038/nbt.3423>
- Delahaye, N. F., Barbier, M., Fumoux, F., & Rihet, P. (2007). Association analyses of NCR3 polymorphisms with *P. falciparum* mild malaria. *Microbes Infect*, 9(2), 160-166. <https://doi.org/10.1016/j.micinf.2006.11.002>
- Deleye, L., Dheedene, A., De Coninck, D., Sante, T., Christodoulou, C., Heindryckx, B., Van den Abbeel, E., De Sutter, P., Deforce, D., Menten, B., & Van Nieuwerburgh, F. (2015). Shallow whole genome sequencing is well suited for the detection of chromosomal aberrations in human blastocysts. *Fertil Steril*, 104(5), 1276-1285.e1271. <https://doi.org/10.1016/j.fertnstert.2015.07.1144>
- Deng, C., Wynshaw-Boris, A., Zhou, F., Kuo, A., & Leder, P. (1996). Fibroblast growth factor receptor 3 is a negative regulator of bone growth. *Cell*, 84(6), 911-921. [https://doi.org/10.1016/s0092-8674\(00\)81069-7](https://doi.org/10.1016/s0092-8674(00)81069-7)
- Dheedene, A., Sante, T., De Smet, M., Vanbellinghen, J. F., Grisart, B., Vergult, S., Janssens, S., & Menten, B. (2016). Implementation of non-invasive prenatal testing by semiconductor sequencing in a genetic laboratory. *Prenat Diagn*, 36(8), 699-707. <https://doi.org/10.1002/pd.4841>
- Diehl, F., Schmidt, K., Choti, M. A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S. A., Kinzler, K. W., Vogelstein, B., & Diaz, L. A., Jr. (2008). Circulating mutant DNA to assess tumor dynamics. *Nat Med*, 14(9), 985-990. <https://doi.org/10.1038/nm.1789>
- Eichler, E. E., Clark, R. A., & She, X. (2004). An assessment of the sequence gaps: Unfinished business in a finished human genome. *Nature Reviews Genetics*, 5(5), 345-354. <https://doi.org/10.1038/nrg1322>
- Elbaz, A., Ross, O. A., Ioannidis, J. P., Soto-Ortolaza, A. I., Moisan, F., Aasly, J., Annesi, G., Bozi, M., Brighina, L., Chartier-Harlin, M. C., Destée, A., Ferrarese, C., Ferraris, A., Gibson, J. M., Gispert, S., Hadjigeorgiou, G. M., Jasinska-Myga, B., Klein, C., Krüger, R., . . . Farrer, M. J. (2011). Independent and joint effects of the MAPT and SNCA genes in Parkinson disease. *Ann Neurol*, 69(5), 778-792. <https://doi.org/10.1002/ana.22321>
- Euskirchen, P., Bielle, F., Labreche, K., Kloosterman, W. P., Rosenberg, S., Daniau, M., Schmitt, C., Masliah-Planchon, J., Bourdeaut, F., Dehais, C., Marie, Y., Delattre, J. Y., & Idhahbi, A. (2017). Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing. *Acta Neuropathol*, 134(5), 691-703. <https://doi.org/10.1007/s00401-017-1743-5>
- Fitzgerald, T., & Birney, E. (2022). CNest: A novel copy number association discovery method uncovers 862 new associations from 200,629 whole-exome sequence datasets in the UK Biobank. *Cell genomics*, 2(8), 100167-100167. <https://doi.org/10.1016/i.xgen.2022.100167>
- Frampton, G. M., Fichtenholtz, A., Otto, G. A., Wang, K., Downing, S. R., He, J., Schnall-Levin, M., White, J., Sanford, E. M., An, P., Sun, J., Juhn, F., Brennan, K., Iwanik, K., Maillet, A., Buell, J., White, E., Zhao, M., Balasubramanian, S., . . . Yelensky, R. (2013). Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol*, 31(11), 1023-1031. <https://doi.org/10.1038/nbt.2696>
- Frattini, V., Pagnotta, S. M., Tala, Fan, J. J., Russo, M. V., Lee, S. B., Garofano, L., Zhang, J., Shi, P., Lewis, G., Sanson, H., Frederick, V., Castano, A. M., Cerulo, L., Rolland, D. C. M., Mall, R., Mokhtari, K.,

- Elenitoba-Johnson, K. S. J., Sanson, M., . . . Iavarone, A. (2018). A metabolic function of FGFR3-TACC3 gene fusions in cancer. *Nature*, 553(7687), 222-227. <https://doi.org/10.1038/nature25171>
- Gagos, S., & Irminger-Finger, I. (2005). Chromosome instability in neoplasia: chaotic roots to continuous growth. *Int J Biochem Cell Biol*, 37(5), 1014-1033. <https://doi.org/10.1016/j.biocel.2005.01.003>
- Gao, Q., Zeng, Q., Wang, Z., Li, C., Xu, Y., Cui, P., Zhu, X., Lu, H., Wang, G., Cai, S., Wang, J., & Fan, J. (2022). Circulating cell-free DNA for cancer early detection. *Innovation (Camb)*, 3(4), 100259. <https://doi.org/10.1016/j.xinn.2022.100259>
- Garraway, L. A., & Lander, E. S. (2013). Lessons from the cancer genome. *Cell*, 153(1), 17-37. <https://doi.org/10.1016/j.cell.2013.03.002>
- Gerlach, S. U., & Herranz, H. (2020). Genomic instability and cancer: lessons from *Drosophila*. *Open Biol*, 10(6), 200060. <https://doi.org/10.1098/rsob.200060>
- Gershman, A., Sauria, M. E. G., Guitart, X., Vollger, M. R., Hook, P. W., Hoyt, S. J., Jain, M., Shumate, A., Razaghi, R., Koren, S., Altemose, N., Caldas, G. V., Logsdon, G. A., Rhie, A., Eichler, E. E., Schatz, M. C., O'Neill, R. J., Phillippy, A. M., Miga, K. H., & Timp, W. (2022). Epigenetic patterns in a complete human genome. *Science*, 376(6588), eabj5089. <https://doi.org/10.1126/science.abj5089>
- Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., 3rd, Barretina, J., Gelfand, E. T., Bielski, C. M., Li, H., Hu, K., Andreev-Drakhlin, A. Y., Kim, J., Hess, J. M., Haas, B. J., Aguet, F., Weir, B. A., Rothberg, M. V., Paoletta, B. R., . . . Sellers, W. R. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, 569(7757), 503-508. <https://doi.org/10.1038/s41586-019-1186-3>
- Guiu, S., Michiels, S., André, F., Cortes, J., Denkert, C., Di Leo, A., Hennessy, B. T., Sorlie, T., Sotiriou, C., Turner, N., Van de Vijver, M., Viale, G., Loi, S., & Reis-Filho, J. S. (2012). Molecular subclasses of breast cancer: how do we define them? The IMPAKT 2012 Working Group Statement†. *Annals of Oncology*, 23(12), 2997-3006. <https://doi.org/https://doi.org/10.1093/annonc/mds586>
- Guyard, A., Boyez, A., Pujals, A., Robe, C., Tran Van Nhieu, J., Allory, Y., Moroch, J., Georges, O., Fournet, J. C., Zafrani, E. S., & Leroy, K. (2017). DNA degrades during storage in formalin-fixed and paraffin-embedded tissue blocks. *Virchows Arch*, 471(4), 491-500. <https://doi.org/10.1007/s00428-017-2213-0>
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5), 646-674. <https://doi.org/10.1016/j.cell.2011.02.013>
- Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J., & Cardoso, F. (2019). Breast cancer. *Nature Reviews Disease Primers*, 5(1), 66. <https://doi.org/10.1038/s41572-019-0111-2>
- Harbers, L., Agostini, F., Nicos, M., Poddighe, D., Bienko, M., & Crosetto, N. (2021). Somatic Copy Number Alterations in Human Cancers: An Analysis of Publicly Available Data From The Cancer Genome Atlas. *Front Oncol*, 11, 700568. <https://doi.org/10.3389/fonc.2021.700568>
- Haupt, S., Vijayakumaran, R., Miranda, P. J., Burgess, A., Lim, E., & Haupt, Y. (2017). The role of MDM2 and MDM4 in breast cancer development and prevention. *J Mol Cell Biol*, 9(1), 53-61. <https://doi.org/10.1093/jmcb/mjx007>
- Heidrich, I., Ačkar, L., Mossahebi Mohammadi, P., & Pantel, K. (2021). Liquid biopsies: Potential and challenges. *Int J Cancer*, 148(3), 528-545. <https://doi.org/10.1002/ijc.33217>
- Holcomb, I. N., & Trask, B. J. (2011). Comparative genomic hybridization to detect variation in the copy number of large DNA segments. *Cold Spring Harb Protoc*, 2011(11), 1323-1333. <https://doi.org/10.1101/pdb.top066589>
- Homburger, J. R., Neben, C. L., Mishne, G., Zhou, A. Y., Kathiresan, S., & Khera, A. V. (2019). Low coverage whole genome sequencing enables accurate assessment of common variants and calculation of genome-wide polygenic scores. *Genome Medicine*, 11(1), 74. <https://doi.org/10.1186/s13073-019-0682-2>
- Hömig-Hölzel, C., & Savola, S. (2012). Multiplex ligation-dependent probe amplification (MLPA) in tumor diagnostics and prognostics. *Diagn Mol Pathol*, 21(4), 189-206. <https://doi.org/10.1097/PDM.0b013e3182595516>
- Hor, H., Bartesaghi, L., Kutalik, Z., Vicário, J. L., de Andrés, C., Pfister, C., Lammers, G. J., Guex, N., Chrast, R., Tafti, M., & Peraita-Adrados, R. (2011). A missense mutation in myelin oligodendrocyte glycoprotein as a cause of familial narcolepsy with cataplexy. *Am J Hum Genet*, 89(3), 474-479. <https://doi.org/10.1016/j.ajhg.2011.08.007>
- Hsu, T.-C. (1952). Mammalian chromosomes in vitro: I. The karyotype of man. *Journal of Heredity*, 43(4), 167-172.

- Hudson, T. J., Anderson, W., Aretz, A., Barker, A. D., Bell, C., Bernabé, R. R., Bhan, M. K., Calvo, F., Eerola, I., Gerhard, D. S., Guttmacher, A., Guyer, M., Hemsley, F. M., Jennings, J. L., Kerr, D., Klatt, P., Kolar, P., Kusuda, J., Lane, D. P., . . . Liver, c. (2010). International network of cancer genome projects. *Nature*, 464(7291), 993-998. <https://doi.org/10.1038/nature08987>
- Hurd, P. J., & Nelson, C. J. (2009). Advantages of next-generation sequencing versus the microarray in epigenetic research. *Brief Funct Genomic Proteomic*, 8(3), 174-183. <https://doi.org/10.1093/bfgp/elp013>
- Huw, L. Y., O'Brien, C., Pandita, A., Mohan, S., Spoerke, J. M., Lu, S., Wang, Y., Hampton, G. M., Wilson, T. R., & Lackner, M. R. (2013). Acquired PIK3CA amplification causes resistance to selective phosphoinositide 3-kinase inhibitors in breast cancer. *Oncogenesis*, 2(12), e83-e83. <https://doi.org/10.1038/oncsis.2013.46>
- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*, 17(1), 239. <https://doi.org/10.1186/s13059-016-1103-0>
- Jallepalli, P. V., Waizenegger, I. C., Bunz, F., Langer, S., Speicher, M. R., Peters, J. M., Kinzler, K. W., Vogelstein, B., & Lengauer, C. (2001). Securin is required for chromosomal stability in human cells. *Cell*, 105(4), 445-457. [https://doi.org/10.1016/s0092-8674\(01\)00340-3](https://doi.org/10.1016/s0092-8674(01)00340-3)
- Janevski, A., Varadan, V., Kamalakaran, S., Banerjee, N., & Dimitrova, N. (2012). Effective normalization for copy number variation detection from whole genome sequencing. *BMC Genomics*, 13 Suppl 6(Suppl 6), S16. <https://doi.org/10.1186/1471-2164-13-s6-s16>
- Jang, H., Hur, Y., & Lee, H. (2016). Identification of cancer-driver genes in focal genomic alterations from whole genome sequencing data. *Sci Rep*, 6, 25582. <https://doi.org/10.1038/srep25582>
- Jefford, C. E., & Irminger-Finger, I. (2006). Mechanisms of chromosome instability in cancers. *Crit Rev Oncol Hematol*, 59(1), 1-14. <https://doi.org/10.1016/j.critrevonc.2006.02.005>
- Jiang, P., Chan, C. W., Chan, K. C., Cheng, S. H., Wong, J., Wong, V. W., Wong, G. L., Chan, S. L., Mok, T. S., Chan, H. L., Lai, P. B., Chiu, R. W., & Lo, Y. M. (2015). Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A*, 112(11), E1317-1325. <https://doi.org/10.1073/pnas.1500076112>
- Jönsson, G., Staaf, J., Olsson, E., Heidenblad, M., Vallon-Christersson, J., Osoegawa, K., de Jong, P., Oredsson, S., Ringné, M., Höglund, M., & Borg, A. (2007). High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. *Genes Chromosomes Cancer*, 46(6), 543-558. <https://doi.org/10.1002/gcc.20438>
- Jung, H.-S., Lefferts, J. A., & Tsongalis, G. J. (2017). Utilization of the oncoscan microarray assay in cancer diagnostics. *Applied Cancer Research*, 37(1), 1. <https://doi.org/10.1186/s41241-016-0007-3>
- Kallioniemi, A. (2008). CGH microarrays and cancer. *Curr Opin Biotechnol*, 19(1), 36-40. <https://doi.org/10.1016/j.copbio.2007.11.004>
- Katerji, M., & Duerksen-Hughes, P. J. (2021). DNA damage in cancer development: special implications in viral oncogenesis. *Am J Cancer Res*, 11(8), 3956-3979.
- Kaur, P., Porras, T. B., Colombo, A., Ring, A., Lu, J., Kang, I., & Lang, J. E. (2021). Identification of putative actionable alterations in clinically relevant genes in breast cancer. *British Journal of Cancer*, 125(9), 1270-1284. <https://doi.org/10.1038/s41416-021-01522-7>
- Kilgour, E., Rothwell, D. G., Brady, G., & Dive, C. (2020). Liquid Biopsy-Based Biomarkers of Treatment Response and Resistance. *Cancer Cell*, 37(4), 485-495. <https://doi.org/10.1016/j.ccell.2020.03.012>
- Kim, M. H., Kim, G. M., Ahn, J. M., Ryu, W. J., Kim, S. G., Kim, J. H., Kim, T. Y., Han, H. J., Kim, J. Y., Park, H. S., Park, S., Park, B. W., Kim, S. I., Jeong, J., Lee, J., Paik, S., Kim, S., Jung, K. H., Cho, E. H., & Sohn, J. (2023). Copy number aberrations in circulating tumor DNA enables prognosis prediction and molecular characterization of breast cancer. *J Natl Cancer Inst*, 115(9), 1036-1049. <https://doi.org/10.1093/jnci/djad080>
- Kim, Y. H., & Pollack, J. R. (2009). Comparative genomic hybridization on spotted oligonucleotide microarrays. *Methods Mol Biol*, 556, 21-32. https://doi.org/10.1007/978-1-60327-192-9_3
- Knouse, K. A., Davoli, T., Elledge, S. J., & Amon, A. (2017). Aneuploidy in Cancer: Seq-ing Answers to Old Questions.
- Knudsen, E. S., Nambiar, R., Rosario, S. R., Smiraglia, D. J., Goodrich, D. W., & Witkiewicz, A. K. (2020). Pan-cancer molecular analysis of the RB tumor suppressor pathway. *Communications Biology*, 3(1), 158. <https://doi.org/10.1038/s42003-020-0873-9>
- Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., McMichael, J. F., Fulton, L. L., Dooling, D. J., Ding, L., Mardis, E. R., Wilson, R. K., Ally, A., Balasundaram, M., Butterfield, Y. S. N., Carlsen, R., Carter, C., Chu, A., Chuah, E., Chun, H.-J. E., . . . Center, M. D. A. C. (2012).

- Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), 61-70. <https://doi.org/10.1038/nature11412>
- Kou, F., Wu, L., Ren, X., & Yang, L. (2020). Chromosome Abnormalities: New Insights into Their Clinical Significance in Cancer. *Molecular Therapy - Oncolytics*, 17, 562-570. <https://doi.org/https://doi.org/10.1016/j.omto.2020.05.010>
- Krijgsman, O., Carvalho, B., Meijer, G. A., Steenbergen, R. D., & Ylstra, B. (2014). Focal chromosomal copy number aberrations in cancer-Needles in a genome haystack. *Biochim Biophys Acta*, 1843(11), 2698-2704. <https://doi.org/10.1016/j.bbamcr.2014.08.001>
- Kunc, M., Popęda, M., Biernat, W., & Senkus, E. (2021). Lost but Not Least-Novel Insights into Progesterone Receptor Loss in Estrogen Receptor-Positive Breast Cancer. *Cancers (Basel)*, 13(19). <https://doi.org/10.3390/cancers13194755>
- LaFramboise, T. (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res*, 37(13), 4181-4193. <https://doi.org/10.1093/nar/gkp552>
- Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif*, 3, 1-8. <https://doi.org/10.1016/j.bdq.2015.02.001>
- Lee, J. K., Choi, Y. L., Kwon, M., & Park, P. J. (2016). Mechanisms and Consequences of Cancer Genome Instability: Lessons from Genome Sequencing Studies. *Annu Rev Pathol*, 11, 283-312. <https://doi.org/10.1146/annurev-pathol-012615-044446>
- Levitt, N. C., & Hickson, I. D. (2002). Caretaker tumour suppressor genes that defend genome integrity. *Trends Mol Med*, 8(4), 179-186. [https://doi.org/10.1016/s1471-4914\(02\)02298-0](https://doi.org/10.1016/s1471-4914(02)02298-0)
- Li, Y., Roberts, N. D., Wala, J. A., Shapira, O., Schumacher, S. E., Kumar, K., Khurana, E., Waszak, S., Korbel, J. O., Haber, J. E., Imielinski, M., Akdemir, K. C., Alvarez, E. G., Baez-Ortega, A., Beroukhi, R., Boutros, P. C., Bowtell, D. D. L., Brors, B., Burns, K. H., . . . Consortium, P. (2020). Patterns of somatic structural variation in human cancer genomes. *Nature*, 578(7793), 112-121. <https://doi.org/10.1038/s41586-019-1913-9>
- Lin, C. F., Naj, A. C., & Wang, L. S. (2013). Analyzing copy number variation using SNP array data: protocols for calling CNV and association tests. *Curr Protoc Hum Genet*, 79, 1.27.21-21.27.15. <https://doi.org/10.1002/0471142905.hg0127s79>
- Lone, S. N., Nisar, S., Masoodi, T., Singh, M., Rizwan, A., Hashem, S., El-Rifai, W., Bedognetti, D., Batra, S. K., Haris, M., Bhat, A. A., & Macha, M. A. (2022). Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. *Molecular Cancer*, 21(1), 79. <https://doi.org/10.1186/s12943-022-01543-7>
- Lou, R. N., Jacobs, A., Wilder, A. P., & Therikildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*, 30(23), 5966-5993. <https://doi.org/https://doi.org/10.1111/mec.16077>
- Ma, L., & Chung, W. K. (2014). Quantitative analysis of copy number variants based on real-time LightCycler PCR. *Curr Protoc Hum Genet*, 80, 7.21.21-27.21.28. <https://doi.org/10.1002/0471142905.hg0721s80>
- Mackay, A., Tamber, N., Fenwick, K., Iravani, M., Grigoriadis, A., Dexter, T., Lord, C. J., Reis-Filho, J. S., & Ashworth, A. (2009). A high-resolution integrated analysis of genetic and expression profiles of breast cancer cell lines. *Breast Cancer Res Treat*, 118(3), 481-498. <https://doi.org/10.1007/s10549-008-0296-7>
- Magi, A., Bolognini, D., Bartalucci, N., Mingrino, A., Semeraro, R., Giovannini, L., Bonifacio, S., Parrini, D., Pelo, E., Mannelli, F., Guglielmelli, P., & Maria Vannucchi, A. (2019). Nano-GLADIATOR: real-time detection of copy number alterations from nanopore sequencing data. *Bioinformatics*, 35(21), 4213-4221. <https://doi.org/10.1093/bioinformatics/btz241>
- Magi, A., Pippucci, T., & Sidore, C. (2017). XCAVATOR: accurate detection and genotyping of copy number variants from second and third generation whole-genome sequencing experiments. *BMC Genomics*, 18(1), 747. <https://doi.org/10.1186/s12864-017-4137-0>
- Magi, A., Tattini, L., Pippucci, T., Torricelli, F., & Benelli, M. (2012). Read count approach for DNA copy number variants detection. *Bioinformatics*, 28(4), 470-478. <https://doi.org/10.1093/bioinformatics/btr707>
- Makrythanasis, P., Temtamy, S., Aglan, M. S., Otaify, G. A., Hamamy, H., & Antonarakis, S. E. (2014). A novel homozygous mutation in FGFR3 causes tall stature, severe lateral tibial deviation, scoliosis, hearing impairment, camptodactyly, and arachnodactyly. *Hum Mutat*, 35(8), 959-963. <https://doi.org/10.1002/humu.22597>

- Manic, G., Corradi, F., Sistigu, A., Siteni, S., & Vitale, I. (2017). Molecular Regulation of the Spindle Assembly Checkpoint by Kinases and Phosphatases. *Int Rev Cell Mol Biol*, 328, 105-161. <https://doi.org/10.1016/bs.ircmb.2016.08.004>
- Mannelli, C. (2019). Tissue vs Liquid Biopsies for Cancer Detection: Ethical Issues. *Journal of Bioethical Inquiry*, 16(4), 551-557. <https://doi.org/10.1007/s11673-019-09944-y>
- Mao, X., Liu, C., Tong, H., Chen, Y., & Liu, K. (2019). Principles of digital PCR and its applications in current obstetrical and gynecological diseases. *Am J Transl Res*, 11(12), 7209-7222.
- Mardis, E. R. (2019). The Impact of Next-Generation Sequencing on Cancer Genomics: From Discovery to Clinic. *Cold Spring Harb Perspect Med*, 9(9). <https://doi.org/10.1101/cshperspect.a036269>
- Martignano, F., Munagala, U., Crucitta, S., Mingrino, A., Semeraro, R., Del Re, M., Petrini, I., Magi, A., & Conticello, S. G. (2021). Nanopore sequencing from liquid biopsy: analysis of copy number variations from cell-free DNA of lung cancer patients. *Mol Cancer*, 20(1), 32. <https://doi.org/10.1186/s12943-021-01327-5>
- Martin, C. L., & Warburton, D. (2015). Detection of Chromosomal Aberrations in Clinical Practice: From Karyotype to Genome Sequence. *Annual Review of Genomics and Human Genetics*, 16(1), 309-326. <https://doi.org/10.1146/annurev-genom-090413-025346>
- Mazzonetto, P. C., Villela, D., da Costa, S. S., Krepischi, A. C. V., Milanezi, F., Migliavacca, M. P., Pierry, P. M., Bonaldi, A., Almeida, L. G. D., De Souza, C. A., Kroll, J. E., Paula, M. G., Guarischi-Sousa, R., Scapulatempo-Neto, C., & Rosenberg, C. (2024). Low-pass whole genome sequencing is a reliable and cost-effective approach for copy number variant analysis in the clinical setting. *Ann Hum Genet*, 88(2), 113-125. <https://doi.org/10.1111/ahg.12532>
- McDermott, U. (2015). Next-generation sequencing and empowering personalised cancer medicine. *Drug Discovery Today*, 20(12), 1470-1475. <https://doi.org/https://doi.org/10.1016/j.drudis.2015.10.008>
- Medvedev, P., Fiume, M., Dzamba, M., Smith, T., & Brudno, M. (2010). Detecting copy number variation with mated short reads. *Genome Res*, 20(11), 1613-1622. <https://doi.org/10.1101/gr.106344.110>
- Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhi, R., & Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*, 12(4), R41. <https://doi.org/10.1186/gb-2011-12-4-r41>
- Method of the Year 2022: long-read sequencing. (2023). *Nature Methods*, 20(1), 1-1. <https://doi.org/10.1038/s41592-022-01759-x>
- Michels, E., De Preter, K., Van Roy, N., & Speleman, F. (2007). Detection of DNA copy number alterations in cancer by array comparative genomic hybridization. *Genet Med*, 9(9), 574-584. <https://doi.org/10.1097/gim.0b013e318145b25b>
- Michor, F., Iwasa, Y., Vogelstein, B., Lengauer, C., & Nowak, M. A. (2005). Can chromosomal instability initiate tumorigenesis? *Semin Cancer Biol*, 15(1), 43-49. <https://doi.org/10.1016/j.semcancer.2004.09.007>
- Misale, S., Yaeger, R., Hobor, S., Scala, E., Janakiraman, M., Liska, D., Valtorta, E., Schiavo, R., Buscarino, M., Siravegna, G., Bencardino, K., Cercek, A., Chen, C. T., Veronese, S., Zanon, C., Sartore-Bianchi, A., Gambacorta, M., Gallicchio, M., Vakiani, E., . . . Bardelli, A. (2012). Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature*, 486(7404), 532-536. <https://doi.org/10.1038/nature11156>
- Mouliere, F., Chandrananda, D., Piskorz, A. M., Moore, E. K., Morris, J., Ahlborn, L. B., Mair, R., Goranova, T., Marass, F., Heider, K., Wan, J. C. M., Supernat, A., Hudecova, I., Gounaris, I., Ros, S., Jimenez-Linan, M., Garcia-Corbacho, J., Patel, K., Østrup, O., . . . Rosenfeld, N. (2018). Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med*, 10(466). <https://doi.org/10.1126/scitranslmed.aat4921>
- Nabavi, S., & Zare, F. (2022). Identification of Copy Number Alterations from Next-Generation Sequencing Data. In A. Laganà (Ed.), *Computational Methods for Precision Oncology* (pp. 55-74). Springer International Publishing. https://doi.org/10.1007/978-3-030-91836-1_4
- Nair, R. P., Stuart, P. E., Nistor, I., Hiremagalore, R., Chia, N. V. C., Jenisch, S., Weichenthal, M., Abecasis, G. R., Lim, H. W., Christophers, E., Voorhees, J. J., & Elder, J. T. (2006). Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *Am J Hum Genet*, 78(5), 827-851. <https://doi.org/10.1086/503821>
- Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol*, 11(3), 220-228. <https://doi.org/10.1038/nrm2858>
- Nowak, D., Hofmann, W. K., & Koeffler, H. P. (2009). Genome-wide Mapping of Copy Number Variations Using SNP Arrays. *Transfus Med Hemother*, 36(4), 246-251. <https://doi.org/10.1159/000225372>
- Nugoli, M., Chuchana, P., Vendrell, J., Orsetti, B., Ursule, L., Nguyen, C., Birnbaum, D., Douzery, E. J., Cohen, P., & Theillet, C. (2003). Genetic variability in MCF-7 sublines: evidence of rapid genomic

- and RNA expression profile modifications. *BMC Cancer*, 3, 13. <https://doi.org/10.1186/1471-2407-3-13>
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bizakadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., . . . Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, 376(6588), 44-53. <https://doi.org/doi:10.1126/science.abj6987>
- O'Sullivan, R. J., & Karlseder, J. (2010). Telomeres: protecting chromosomes against genome instability. *Nature Reviews Molecular Cell Biology*, 11(3), 171-181. <https://doi.org/10.1038/nrm2848>
- Oehler, J. B., Wright, H., Stark, Z., Mallett, A. J., & Schmitz, U. (2023). The application of long-read sequencing in clinical settings. *Hum Genomics*, 17(1), 73. <https://doi.org/10.1186/s40246-023-00522-3>
- Ono, M., Oba, T., Shibata, T., & Ito, K. I. (2021). The mechanisms involved in the resistance of estrogen receptor-positive breast cancer cells to palbociclib are multiple and change over time. *J Cancer Res Clin Oncol*, 147(11), 3211-3224. <https://doi.org/10.1007/s00432-021-03722-3>
- Oostlander, A. E., Meijer, G. A., & Ylstra, B. (2004). Microarray-based comparative genomic hybridization and its applications in human genetics. *Clin Genet*, 66(6), 488-495. <https://doi.org/10.1111/j.1399-0004.2004.00322.x>
- Pei, X. M., Yeung, M. H. Y., Wong, A. N. N., Tsang, H. F., Yu, A. C. S., Yim, A. K. Y., & Wong, S. C. C. (2023). Targeted Sequencing Approach and Its Clinical Applications for the Molecular Diagnosis of Human Diseases. *Cells*, 12(3). <https://doi.org/10.3390/cells12030493>
- Phallen, J., Sausen, M., Adleff, V., Leal, A., Hruban, C., White, J., Anagnostou, V., Fiksel, J., Cristiano, S., Papp, E., Speir, S., Reinert, T., Orntoft, M. W., Woodward, B. D., Murphy, D., Parpart-Li, S., Riley, D., Nesselbush, M., Sengamalay, N., . . . Velculescu, V. E. (2017). Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med*, 9(403). <https://doi.org/10.1126/scitranslmed.aan2415>
- Pikor, L., Thu, K., Vucic, E., & Lam, W. (2013). The detection and implication of genome instability in cancer. *Cancer Metastasis Rev*, 32(3-4), 341-352. <https://doi.org/10.1007/s10555-013-9429-5>
- Pinzani, P., D'Argenio, V., Del Re, M., Pellegrini, C., Cucchiara, F., Salvianti, F., & Galbiati, S. (2021). Updates on liquid biopsy: current trends and future perspectives for clinical application in solid tumors. *Clin Chem Lab Med*, 59(7), 1181-1200. <https://doi.org/10.1515/cclm-2020-1685>
- Pös, O., Radvanszky, J., Styk, J., Pös, Z., Buglyó, G., Kajsik, M., Budis, J., Nagy, B., & Szemes, T. (2021). Copy Number Variation: Methods and Clinical Applications. *Applied Sciences*, 11(2), 819. <https://www.mdpi.com/2076-3417/11/2/819>
- Prat, A., Brasó-Maristany, F., Martínez-Sáez, O., Sanfeliu, E., Xia, Y., Bellet, M., Galván, P., Martínez, D., Pascual, T., Marín-Aguilera, M., Rodríguez, A., Chic, N., Adamo, B., Paré, L., Vidal, M., Margelí, M., Ballana, E., Gómez-Rey, M., Oliveira, M., . . . Vivancos, A. (2023). Circulating tumor DNA reveals complex biological features with clinical relevance in metastatic breast cancer. *Nature Communications*, 14(1), 1157. <https://doi.org/10.1038/s41467-023-36801-9>
- Qin, D. (2019). Next-generation sequencing and its clinical application. *Cancer Biol Med*, 16(1), 4-10. <https://doi.org/10.20892/j.issn.2095-3941.2018.0055>
- Quan, P. L., Sauzade, M., & Brouzes, E. (2018). dPCR: A Technology Review. *Sensors (Basel)*, 18(4). <https://doi.org/10.3390/s18041271>
- Raman, L., Dheedene, A., De Smet, M., Van Dorpe, J., & Menten, B. (2018). WisecondorX: improved copy number detection for routine shallow whole-genome sequencing. *Nucleic Acids Research*, 47(4), 1605-1614. <https://doi.org/10.1093/nar/gky1263>
- Raman, L., Van der Linden, M., Van der Eecken, K., Vermaelen, K., Demedts, I., Surmont, V., Himpe, U., Dedeurwaerdere, F., Ferdinande, L., Lievens, Y., Claes, K., Menten, B., & Van Dorpe, J. (2020). Shallow whole-genome sequencing of plasma cell-free DNA accurately differentiates small from non-small cell lung carcinoma. *Genome Med*, 12(1), 35. <https://doi.org/10.1186/s13073-020-00735-4>
- Raman, L., Van der Linden, M., Van der Eecken, K., Vermaelen, K., Demedts, I., Surmont, V., Himpe, U., Dedeurwaerdere, F., Ferdinande, L., Lievens, Y., Claes, K., Menten, B., & Van Dorpe, J. (2020). Shallow whole-genome sequencing of plasma cell-free DNA accurately differentiates small from non-small cell lung carcinoma. *Genome Medicine*, 12(1), 35. <https://doi.org/10.1186/s13073-020-00735-4>
- Raynes, Y., & Weinreich, D. M. (2018). Genomic clustering of fitness-affecting mutations favors the evolution of chromosomal instability. *Evolutionary applications*, 12(2), 301-313. <https://doi.org/10.1111/eva.12717>

- Redon, R., & Carter, N. P. (2009). Comparative genomic hybridization: microarray design and data interpretation. *Methods Mol Biol*, 529, 37-49. https://doi.org/10.1007/978-1-59745-538-1_3
- Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-throughput sequencing technologies. *Mol Cell*, 58(4), 586-597. <https://doi.org/10.1016/j.molcel.2015.05.004>
- Rolfo, C., & Russo, A. (2020). Liquid biopsy for early stage lung cancer moves ever closer. *Nature Reviews Clinical Oncology*, 17(9), 523-524. <https://doi.org/10.1038/s41571-020-0393-z>
- Romagnoli, S., Bartalucci, N., & Vannucchi, A. M. (2023). Resolving complex structural variants via nanopore sequencing. *Front Genet*, 14, 1213917. <https://doi.org/10.3389/fgene.2023.1213917>
- Romero-Cordoba, S. L., Salido-Guadarrama, I., Rebollar-Vega, R., Bautista-Piña, V., Dominguez-Reyes, C., Tenorio-Torres, A., Villegas-Carlos, F., Fernández-López, J. C., Uribe-Figueroa, L., Alfaro-Ruiz, L., & Hidalgo-Miranda, A. (2021). Comprehensive omic characterization of breast cancer in Mexican-Hispanic women. *Nature Communications*, 12(1), 2245. <https://doi.org/10.1038/s41467-021-22478-5>
- Sahlin, K., & Medvedev, P. (2021). Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. *Nat Commun*, 12(1), 2. <https://doi.org/10.1038/s41467-020-20340-8>
- Sanderson, N. D., Kapel, N., Rodger, G., Webster, H., Lipworth, S., Street, T. L., Peto, T., Crook, D., & Stoesser, N. (2023). Comparison of R9.4.1/Kit10 and R10/Kit12 Oxford Nanopore flowcells and chemistries in bacterial genome reconstruction. *Microb Genom*, 9(1). <https://doi.org/10.1099/mgen.0.000910>
- Schaub, F. X., Dhankani, V., Berger, A. C., Trivedi, M., Richardson, A. B., Shaw, R., Zhao, W., Zhang, X., Ventura, A., Liu, Y., Ayer, D. E., Hurlin, P. J., Cherniack, A. D., Eisenman, R. N., Bernard, B., & Grandori, C. (2018). Pan-cancer Alterations of the MYC Oncogene and Its Proximal Network across the Cancer Genome Atlas. *Cell Syst*, 6(3), 282-300.e282. <https://doi.org/10.1016/j.cels.2018.03.003>
- Schouten, J., van Vught, P., & Galjaard, R. J. (2019). Multiplex Ligation-Dependent Probe Amplification (MLPA) for Prenatal Diagnosis of Common Aneuploidies. *Methods Mol Biol*, 1885, 161-170. https://doi.org/10.1007/978-1-4939-8889-1_11
- Schreck, R. R., & Distèche, C. M. (2001). Chromosome banding techniques. *Curr Protoc Hum Genet*, Chapter 4, Unit4.2. <https://doi.org/10.1002/0471142905.hg0402s00>
- Shadeo, A., & Lam, W. L. (2006). Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Res*, 8(1), R9. <https://doi.org/10.1186/bcr1370>
- Sims, D., Sudbery, I., Illott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), 121-132. <https://doi.org/10.1038/nrg3642>
- Smolander, J., Khan, S., Singaravelu, K., Kauko, L., Lund, R. J., Laiho, A., & Elo, L. L. (2021). Evaluation of tools for identifying large copy number variations from ultra-low-coverage whole-genome sequencing data. *BMC Genomics*, 22(1), 357. <https://doi.org/10.1186/s12864-021-07686-z>
- Solomon, D. A., Kim, T., Diaz-Martinez, L. A., Fair, J., Elkahloun, A. G., Harris, B. T., Toretsky, J. A., Rosenberg, S. A., Shukla, N., Ladanyi, M., Samuels, Y., James, C. D., Yu, H., Kim, J. S., & Waldman, T. (2011). Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science*, 333(6045), 1039-1043. <https://doi.org/10.1126/science.1203619>
- Sözen, M. A., Suzuki, K., Tolarova, M. M., Bustos, T., Fernández Iglesias, J. E., & Spritz, R. A. (2001). Mutation of PVRL1 is associated with sporadic, non-syndromic cleft lip/palate in northern Venezuela. *Nat Genet*, 29(2), 141-142. <https://doi.org/10.1038/ng740>
- Steele, C. D., Abbasi, A., Islam, S. M. A., Bowes, A. L., Khandekar, A., Haase, K., Hames-Fathi, S., Ajayi, D., Verfaillie, A., Dhami, P., McLatchie, A., Lechner, M., Light, N., Shlien, A., Malkin, D., Feber, A., Proszek, P., Lesluyes, T., Mertens, F., . . . Pillay, N. (2022). Signatures of copy number alterations in human cancer. *Nature*, 606(7916), 984-991. <https://doi.org/10.1038/s41586-022-04738-6>
- Stefansson, H., Sigurdsson, E., Steinthorsdottir, V., Bjornsdottir, S., Sigmundsson, T., Ghosh, S., Brynjolfsson, J., Gunnarsdottir, S., Ivarsson, O., Chou, T. T., Hjaltason, O., Birgisdottir, B., Jonsson, H., Gudnadottir, V. G., Gudmundsdottir, E., Bjornsson, A., Ingvarsson, B., Ingason, A., Sigfusson, S., . . . Stefansson, K. (2002). Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet*, 71(4), 877-892. <https://doi.org/10.1086/342734>
- Stover, D. G., Parsons, H. A., Ha, G., Freeman, S. S., Barry, W. T., Guo, H., Choudhury, A. D., Gydush, G., Reed, S. C., Rhoades, J., Rotem, D., Hughes, M. E., Dillon, D. A., Partridge, A. H., Wagle, N., Krop, I. E., Getz, G., Golub, T. R., Love, J. C., . . . Adalsteinsson, V. A. (2018). Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-

- Negative Breast Cancer. *Journal of Clinical Oncology*, 36(6), 543-553. <https://doi.org/10.1200/jco.2017.76.0033>
- Stuppia, L., Antonucci, I., Palka, G., & Gatta, V. (2012). Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci*, 13(3), 3245-3276. <https://doi.org/10.3390/ijms13033245>
- Sun, Y., & Liu, Q. (2015). Deciphering the Correlation between Breast Tumor Samples and Cell Lines by Integrating Copy Number Changes and Gene Expression Profiles. *Biomed Res Int*, 2015, 901303. <https://doi.org/10.1155/2015/901303>
- Szuhai, K., & Vermeer, M. (2015). Microarray Techniques to Analyze Copy-Number Alterations in Genomic DNA: Array Comparative Genomic Hybridization and Single-Nucleotide Polymorphism Array. *Journal of Investigative Dermatology*, 135(10), 1-5. <https://doi.org/10.1038/jid.2015.308>
- Teo, S. M., Pawitan, Y., Ku, C. S., Chia, K. S., & Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics*, 28(21), 2711-2718. <https://doi.org/10.1093/bioinformatics/bts535>
- Thierry, A. R., Mouliere, F., El Messaoudi, S., Mollevi, C., Lopez-Crapez, E., Rolet, F., Gillet, B., Gongora, C., Dechelotte, P., Robert, B., Del Rio, M., Lamy, P. J., Bibeau, F., Nouaille, M., Lorient, V., Jarrousse, A. S., Molina, F., Mathonnet, M., Pezet, D., & Ychou, M. (2014). Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. *Nat Med*, 20(4), 430-435. <https://doi.org/10.1038/nm.3511>
- Tomar, U., Grover, N., Tomar, S., Bhalla, K., & Singh, S. (2023). Liquid biopsy and its significance in tumour - Detection in the field of pathology. *J Oral Maxillofac Pathol*, 27(1), 195-200. https://doi.org/10.4103/jomfp.jomfp_251_22
- Tshiabuila, D., Giandhari, J., Pillay, S., Ramphal, U., Ramphal, Y., Maharaj, A., Anyaneji, U. J., Naidoo, Y., Tegally, H., San, E. J., Wilkinson, E., Lessells, R. J., & de Oliveira, T. (2022). Comparison of SARS-CoV-2 sequencing using the ONT GridION and the Illumina MiSeq. *BMC Genomics*, 23(1), 319. <https://doi.org/10.1186/s12864-022-08541-5>
- Underhill, H. R., Kitzman, J. O., Hellwig, S., Welker, N. C., Daza, R., Baker, D. N., Gligorich, K. M., Rostomily, R. C., Bronner, M. P., & Shendure, J. (2016). Fragment Length of Circulating Tumor DNA. *PLoS Genet*, 12(7), e1006162. <https://doi.org/10.1371/journal.pgen.1006162>
- Ungerer, V., Bronkhorst, A. J., Uhlig, C., & Holdenrieder, S. (2022). Cell-Free DNA Fragmentation Patterns in a Cancer Cell Line. *Diagnostics (Basel)*, 12(8). <https://doi.org/10.3390/diagnostics12081896>
- Valla, M., Klæstad, E., Ytterhus, B., & Bofin, A. M. (2022). CCND1 Amplification in Breast Cancer - associations With Proliferation, Histopathological Grade, Molecular Subtype and Prognosis. *J Mammary Gland Biol Neoplasia*, 27(1), 67-77. <https://doi.org/10.1007/s10911-022-09516-8>
- van Eijk, R., Eilers, P. H., Natte, R., Cleton-Jansen, A. M., Morreau, H., van Wezel, T., & Oosting, J. (2010). MLPAinter for MLPA interpretation: an integrated approach for the analysis, visualisation and data management of Multiplex Ligation-dependent Probe Amplification. *BMC Bioinformatics*, 11, 67. <https://doi.org/10.1186/1471-2105-11-67>
- van Jaarsveld, R. H., & Kops, G. J. P. L. (2016). Difference Makers: Chromosomal Instability versus Aneuploidy in Cancer. *Trends in Cancer*, 2(10), 561-571. <https://doi.org/https://doi.org/10.1016/j.trecan.2016.09.003>
- Vermeesch, J. R., Fiegler, H., de Leeuw, N., Suzhai, K., Schoumans, J., Ciccone, R., Speleman, F., Rauch, A., Clayton-Smith, J., Van Ravenswaaij, C., Sanlaville, D., Patsalis, P. C., Firth, H., Devriendt, K., & Zuffardi, O. (2007). Guidelines for molecular karyotyping in constitutional genetic diagnosis. *European Journal of Human Genetics*, 15(11), 1105-1114. <https://doi.org/10.1038/sj.ejhg.5201896>
- Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*, 339(6127), 1546-1558. <https://doi.org/10.1126/science.1235122>
- Wagner, G. E., Dabernig-Heinz, J., Lipp, M., Cabal, A., Simantzik, J., Kohl, M., Scheiber, M., Lichtenegger, S., Ehrlich, R., Leitner, E., Ruppitsch, W., & Steinmetz, I. (2023). Real-Time Nanopore Q20+ Sequencing Enables Extremely Fast and Accurate Core Genome MLST Typing and Democratizes Access to High-Resolution Bacterial Pathogen Surveillance. *J Clin Microbiol*, 61(4), e0163122. <https://doi.org/10.1128/jcm.01631-22>
- Wan, J. C. M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J. D., Caldas, C., Pacey, S., Baird, R., & Rosenfeld, N. (2017). Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*, 17(4), 223-238. <https://doi.org/10.1038/nrc.2017.7>
- Wang, H., Nettleton, D., & Ying, K. (2014). Copy number variation detection using next generation sequencing read counts. *BMC Bioinformatics*, 15(1), 109. <https://doi.org/10.1186/1471-2105-15-109>

- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11), 1348-1365. <https://doi.org/10.1038/s41587-021-01108-x>
- Wei, Y.-C., & Huang, G.-H. (2020). CONY: A Bayesian procedure for detecting copy number variations from sequencing read depths. *Scientific Reports*, 10(1), 10493. <https://doi.org/10.1038/s41598-020-64353-1>
- White, M. K., Pagano, J. S., & Khalili, K. (2014). Viruses and human cancers: a long road of discovery of molecular paradigms. *Clin Microbiol Rev*, 27(3), 463-481. <https://doi.org/10.1128/cmr.00124-13>
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, 20(1), 129. <https://doi.org/10.1186/s13059-019-1727-y>
- Wragg, D., Zhang, W., Peterson, S., Yerramilli, M., Mellanby, R., Schoenebeck, J. J., & Clements, D. N. (2024). A cautionary tale of low-pass sequencing and imputation with respect to haplotype accuracy. *Genetics Selection Evolution*, 56(1), 6. <https://doi.org/10.1186/s12711-024-00875-w>
- Wu, G., Xing, M., Mambo, E., Huang, X., Liu, J., Guo, Z., Chatterjee, A., Goldenberg, D., Gollin, S. M., Sukumar, S., Trink, B., & Sidransky, D. (2005). Somatic mutation and gain of copy number of PIK3CA in human breast cancer. *Breast Cancer Res*, 7(5), R609-616. <https://doi.org/10.1186/bcr1262>
- Xu, J., Chen, Y., & Olopade, O. I. (2010). MYC and Breast Cancer. *Genes Cancer*, 1(6), 629-640. <https://doi.org/10.1177/1947601910378691>
- Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., Lawrence, M. S., Zhang, C.-Z., Wala, J., Mermel, C. H., Sougnez, C., Gabriel, S. B., Hernandez, B., Shen, H., Laird, P. W., Getz, G., Meyerson, M., & Beroukhim, R. (2013). Pan-cancer patterns of somatic copy number alteration. *Nature Genetics*, 45(10), 1134-1140. <https://doi.org/10.1038/ng.2760>
- Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S. M., Hellmann, M. D., Barron, D. A., Schram, A. M., Hameed, M., Dogan, S., Ross, D. S., Hechtman, J. F., DeLair, D. F., Yao, J., . . . Berger, M. F. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*, 23(6), 703-713. <https://doi.org/10.1038/nm.4333>
- Zeng, J., Cai, H., Peng, H., Wang, H., Zhang, Y., & Akutsu, T. (2020). Causalcall: Nanopore Basecalling Using a Temporal Convolutional Network [Methods]. *Frontiers in Genetics*, 10. <https://doi.org/10.3389/fgene.2019.01332>
- Zhang, L., Feizi, N., Chi, C., & Hu, P. (2018). Association Analysis of Somatic Copy Number Alteration Burden With Breast Cancer Survival [Original Research]. *Frontiers in Genetics*, 9. <https://doi.org/10.3389/fgene.2018.00421>
- Zhang, Y., Yao, Y., Xu, Y., Li, L., Gong, Y., Zhang, K., Zhang, M., Guan, Y., Chang, L., Xia, X., Li, L., Jia, S., & Zeng, Q. (2021). Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. *Nature Communications*, 12(1), 11. <https://doi.org/10.1038/s41467-020-20162-8>
- Zhao, M., Wang, Q., Wang, Q., Jia, P., & Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics*, 14(11), S1. <https://doi.org/10.1186/1471-2105-14-S11-S1>
- Zheng, Y., Zhu, B., Tan, J., Guan, Y., The Chinese Genomic Structural Variants, C., Morton, C. C., & Lu, G. (2022). Experience of Low-Pass Whole-Genome Sequencing-Based Copy Number Variant Analysis: A Survey of Chinese Tertiary Hospitals. *Diagnostics (Basel)*, 12(5). <https://doi.org/10.3390/diagnostics12051098>
- Zhou, B., Ho, S. S., Zhang, X., Pattni, R., Haraksingh, R. R., & Urban, A. E. (2018). Whole-genome sequencing analysis of CNV using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. *J Med Genet*, 55(11), 735-743. <https://doi.org/10.1136/jmedgenet-2018-105272>

13 APÉNDICES

13.1 Líneas celulares

13.1.1 Reportes de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos de líneas celulares tumorales empleando el software NanoGLADIATOR

Tabla Suplementaria 1. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular HeLa.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana logratio
1	119,030,000	120,450,000	1p12	1p12	1,420	14	Duplicación	3	0.93	0.07	0.5309
3	60,465,000	61,065,000	3p14.2	3p14.2	600	7	Doble deleción	1	1	0	-1.5072
3	97,560,000	99,160,000	3q11.2	3q12.1	1,600	17	Deleción	1	0.78	0.04	-0.7436
4	178,865,000	186,665,000	4q34.3	4q35.1	7,800	79	Deleción	1	0.74	0.02	-0.6729
5	60,000	45,800,000	5p15.33	5p12	45,740	433	Duplicación de múltiples copias	6	1	0	1.6440
6	64,485,000	68,995,000	6q12	6q12	4,510	45	Deleción	1	0.74	0.03	-0.6708
8	128,255,000	146,150,000	8q24.21	8q24.3	17,895	174	Duplicación	3	0.99	0.02	0.5785
9	245,000	38,675,000	9p24.3	9p13.1	38,430	375	Duplicación	3	0.69	0.02	0.4180
9	121,310,000	140,866,997	9q33.1	9q34.3	19,557	190	Duplicación	3	1.37	0.02	0.7294
11	245,000	2,865,000	11p15.5	11p15.4	2,620	26	Deleción	1	0.92	0.03	-0.8572
11	25,595,000	27,495,000	11p14.3	11p14.1	1,900	20	Duplicación de múltiples copias	4	1	0	0.9487
11	45,055,000	50,150,000	11p11.2	11p11.12	5,095	34	Duplicación	3	1.14	0.04	0.6298
11	92,785,000	93,685,000	11q14.3	11q21	900	10	Duplicación	3	1.6	0.03	0.8424
11	102,287,584	111,930,000	11q22.2	11q23.1	9,642	95	Deleción	1	0.65	0.02	-0.6480
11	112,030,000	134,670,000	11q23.1	11q25	22,640	226	Deleción	1	0.76	0.01	-0.7078
13	19,550,000	27,910,000	13q12.11	13q12.2	8,360	74	Deleción	1	0.6	0.02	-0.5125
13	28,010,000	28,110,000	13q12.2	13q12.2	100	2	Duplicación de múltiples copias	4	1	0	1.0342
13	28,210,000	55,855,000	13q12.2	13q21.1	27,645	270	Deleción	1	0.54	0.01	-0.4696
19	305,000	4,205,000	19p13.3	19p13.3	3,900	40	Deleción	1	0.01	0	-0.6787

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana logratio
19	4,305,000	12,875,000	19p13.3	19p13.2	8,570	80	Delección	1	0.64	0.02	-0.5849
20	110,000	25,555,000	20p13	20p11.21	25,445	253	Delección	1	0.73	0.01	-0.6574
20	51,895,000	62,813,369	20q13.2	20q13.33	10,918	107	Duplicación	3	0.96	0.02	0.5650
21	15,530,000	34,920,000	21q11.2	21q22.11	19,390	192	Delección	1	0.48	0.01	-0.4383
21	35,020,000	47,965,000	21q22.11	21q22.3	12,945	128	Delección	1	0.56	0.02	-0.4914

CN= Número de copias; FA= Fracción alélica; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 2. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular MCF7.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana logratio
1	86,060,000	106,685,000	1p22.3	1p21.1	20,625	201	Delección	1	0.81	0.03	-0.7922
1	106,785,000	108,685,000	1p21.1	1p13.3	1,900	20	Duplicación de múltiples copias	11	1	0	2.4814
1	114,100,000	118,430,000	1p13.2	1p12	4,330	42	Duplicación	3	1.6	0	0.7734
1	145,430,000	152,400,000	1q21.1	1q21.3	6,970	38	Duplicación de múltiples copias	4	1	0	1.1647
1	158,445,000	168,245,000	1q23.1	1q24.2	9,800	96	Delección	1	0.54	0.04	-0.5093
1	168,345,000	181,635,000	1q24.2	1q25.3	13,290	132	Duplicación	3	0.69	0.05	0.4261
1	181,735,000	182,335,000	1q25.3	1q25.3	600	7	Delección	1	0.77	0.14	-0.7637
1	182,435,000	198,370,000	1q25.3	1q31.3	15,935	158	Duplicación	3	0.7	0.05	0.4321
1	198,470,000	207,900,000	1q31.3	1q32.2	9,430	86	Duplicación	3	1.44	0.07	0.8101
1	208,000,000	214,765,000	1q32.2	1q41	6,765	68	Delección	1	0.87	0.04	-0.8504
1	214,865,000	219,665,000	1q41	1q41	4,800	49	Duplicación de múltiples copias	4	1	0	1.0227
3	140,000	24,570,000	3p26.3	3p24.2	24,430	243	Duplicación	3	0.87	0.04	0.5284
3	41,970,000	61,465,000	3p22.1	3p14.2	19,495	195	Delección	1	0.67	0.03	-0.6151
3	61,565,000	64,965,000	3p14.2	3p14.1	3,400	35	Duplicación de múltiples copias	11	1	0	2.4544
3	116,220,000	116,920,000	3q13.31	3q13.31	700	8	Delección	1	0.83	0.13	-0.7747
3	145,815,000	197,750,000	3q24	3q29	51,935	508	Duplicación	3	0.67	0.03	0.4048
4	181,565,000	190,745,000	4q34.3	4q35.2	9,180	89	Delección	1	1.06	0.04	-0.9798

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana log ₂ ratio
5	9,155,000	29,000,000	5p15.31	5p13.3	19,845	184	Duplicación	3	0.66	0.04	0.4019
6	28,520,000	33,405,000	6p22.1	6p21.32	4,885	41	Doble deleción	0	1	0	-2.7004
6	155,395,000	169,992,073	6q25.2	6q27	14,597	141	Deleción	1	0.93	0.03	-0.9434
7	31,695,000	34,895,000	7p14.3	7p14.3	3,200	29	Duplicación	3	1.47	0.12	0.7693
7	100,395,000	114,105,000	7q22.1	7q31.1	13,710	128	Duplicación	3	1.16	0.06	0.6486
8	455,000	43,325,000	8p23.3	8p11.1	42,870	403	Deleción	1	0.88	0.02	-0.8643
8	76,440,000	86,240,000	8q21.11	8q21.2	9,800	99	Duplicación	3	0.72	0.06	0.4371
8	86,340,000	112,355,000	8q21.2	8q23.3	26,015	257	Duplicación de múltiples copias	4	1	0	1.0030
8	112,455,000	121,555,000	8q23.3	8q24.12	9,100	92	Duplicación	3	0.81	0.06	0.4822
8	121,655,000	129,855,000	8q24.12	8q24.21	8,200	83	Duplicación de múltiples copias	4	1	0	1.1357
9	245,000	1,045,000	9p24.3	9p24.3	800	9	Duplicación	3	0.74	0.19	0.4650
9	21,860,000	22,060,000	9p21.3	9p21.3	200	3	Doble deleción	0	1	0	-2.8278
11	245,000	11,895,000	11p15.5	11p15.3	11,650	110	Deleción	1	0.88	0.04	-0.8404
11	43,755,000	45,755,000	11p11.2	11p11.2	2,000	21	Deleción	1	0.95	0.08	-0.9069
11	45,855,000	50,150,000	11p11.2	11p11.12	4,295	26	Duplicación	3	0.67	0.11	0.4427
11	55,130,000	59,325,000	11q11	11q12.1	4,195	42	Deleción	1	0.97	0.05	-0.9003
11	65,475,000	86,515,000	11q13.1	11q14.2	21,040	198	Duplicación	3	0.67	0.04	0.4255
11	86,615,000	117,085,000	11q14.2	11q23.3	30,470	286	Deleción	1	0.86	0.03	-0.8320
11	117,185,000	119,285,000	11q23.3	11q23.3	2,100	22	Duplicación	3	0.84	0.12	0.5194
11	119,385,000	134,670,000	11q23.3	11q25	15,285	153	Deleción	1	0.94	0.03	-0.9018
12	33,100,000	34,280,000	12p11.21	12p11.1	1,180	10	Duplicación	3	0.87	0.18	0.5140
12	60,125,000	85,890,000	12q14.1	12q21.31	25,765	256	Duplicación	3	0.94	0.04	0.4985
13	46,640,000	48,090,000	13q14.13	13q14.2	1,450	14	Duplicación	3	1.45	0.16	0.7930
13	60,515,000	114,989,948	13q21.2	13q34	54,475	534	Deleción	1	0.89	0.02	-0.8655
14	20,240,000	107,165,000	14q11.2	14q32.33	86,925	847	Duplicación	3	0.98	0.02	0.5726
15	49,725,000	54,525,000	15q21.2	15q21.3	4,800	49	Duplicación de múltiples copias	5	1	0	1.2254
15	57,125,000	62,925,000	15q21.3	15q22.2	5,800	59	Deleción	1	0.48	0.06	-0.4493
15	70,725,000	102,185,000	15q23	15q26.3	31,460	293	Duplicación	3	1.06	0.04	0.6094

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana log ₂ ratio
16	46,565,000	67,140,000	16q11.2	16q22.1	20,575	205	Delección	1	0.77	0.03	-0.7235
16	67,240,000	90,090,000	16q22.1	16q24.3	22,850	215	Duplicación	3	0.73	0.04	0.4436
17	28,280,000	38,210,000	17q11.2	17q21.1	9,930	86	Delección	1	0.86	0.04	-0.8322
17	45,040,000	48,925,000	17q21.32	17q21.33	3,885	37	Duplicación	3	1.31	0.1	0.6957
17	49,025,000	67,320,000	17q21.33	17q24.3	18,295	171	Duplicación de múltiples copias	5	1	0	1.3849
17	67,420,000	70,320,000	17q24.3	17q24.3	2,900	30	Duplicación	3	0.79	0.11	0.4977
17	73,320,000	73,820,000	17q25.1	17q25.1	500	6	Duplicación de múltiples copias	5	1	0	1.2695
18	33,890,000	77,921,820	18q12.2	18q23	44,032	436	Delección	1	0.94	0.02	-0.9168
19	305,000	3,905,000	19p13.3	19p13.3	3,600	37	Delección	1	0.86	0.06	-0.7618
19	4,405,000	10,060,000	19p13.3	19p13.2	5,655	53	Delección	1	0.83	0.05	-0.8260
20	6,755,000	25,555,000	20p12.3	20p11.21	18,800	187	Duplicación	3	0.73	0.05	0.4535
20	39,597,085	41,197,085	20q12	20q12	1,600	17	Duplicación de múltiples copias	5	1	0	1.2303
20	45,497,085	62,813,369	20q13.12	20q13.33	17,316	169	Duplicación de múltiples copias	5	1	0	1.3859
21	39,420,000	47,965,000	21q22.13	21q22.3	8,545	84	Duplicación	3	1.23	0.07	0.6720
22	35,340,000	51,064,777	22q12.3	22q13.33	15,725	155	Delección	1	0.93	0.03	-0.9157

CN= Número de copias; FA= Fracción alélica; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 3. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular MDA-MB-361.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana log ₂ ratio
2	141,175,000	142,075,000	2q22.1	2q22.1	900	10	Delección	1	1	0	-1.5498
4	175,365,000	177,365,000	4q34.1	4q34.2	2,000	21	Delección	1	0.54	0.09	-0.4297
5	60,000	45,800,000	5p15.33	5p12	45,740	433	Duplicación	3	0.9	0.03	0.5269
5	49,610,000	54,835,000	5q11.1	5q11.2	5,225	50	Delección	1	0.54	0.06	-0.4591
5	58,595,000	59,495,000	5q11.2	5q12.1	900	10	Delección	1	1	0	-1.3450
6	440,000	1,640,000	6p25.3	6p25.3	1,200	13	Delección	1	0.66	0.11	-0.5762
6	10,840,000	15,140,000	6p24.2	6p23	4,300	44	Duplicación	3	1.15	0.09	0.6648

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana log ₂ ratio
6	28,520,000	33,405,000	6p22.1	6p21.32	4,885	41	Doble deleción	0	1	0	-2.4517
6	33,505,000	39,705,000	6p21.31	6p21.2	6,200	63	Duplicación	3	1.03	0.07	0.6408
6	52,730,000	57,130,000	6p12.2	6p11.2	4,400	45	Duplicación	3	0.93	0.09	0.5406
6	70,195,000	110,420,000	6q13	6q21	40,225	396	Duplicación	3	0.84	0.03	0.4817
7	71,760,000	75,945,000	7q11.22	7q11.23	4,185	23	Duplicación	3	1.25	0.12	0.7995
8	455,000	2,905,000	8p23.3	8p23.2	2,450	23	Deleción	1	0.51	0.08	-0.4297
8	3,005,000	34,050,000	8p23.2	8p12	31,045	289	Deleción	1	1	0	-1.3132
8	75,840,000	146,150,000	8q21.11	8q24.3	70,310	694	Duplicación	3	0.96	0.02	0.5605
9	245,000	30,335,000	9p24.3	9p21.1	30,090	299	Deleción	1	1	0	-1.3440
9	31,335,000	37,245,000	9p21.1	9p13.2	5,910	55	Duplicación	3	1.47	0.09	0.6648
10	170,000	5,120,000	10p15.3	10p15.1	4,950	49	Deleción	1	0.49	0.06	-0.4306
10	11,845,000	19,720,000	10p14	10p12.31	7,875	73	Deleción	1	0.45	0.05	-0.4005
10	19,820,000	20,820,000	10p12.31	10p12.31	1,000	11	Duplicación	3	0.91	0.17	0.5288
10	129,016,069	135,377,527	10q26.2	10q26.3	6,361	62	Deleción	1	0.55	0.05	-0.4972
11	245,000	6,830,000	11p15.5	11p15.4	6,585	60	Deleción	1	1	0	-1.3268
11	6,930,000	50,150,000	11p15.4	11p11.12	43,220	412	Deleción	1	0.49	0.02	-0.4268
11	68,565,000	70,245,000	11q13.3	11q13.3	1,680	15	Duplicación de múltiples copias	5	1	0	1.2874
11	70,345,000	134,670,000	11q13.3	11q25	64,325	617	Deleción	1	0.48	0.02	-0.4085
12	230,000	34,280,000	12p13.33	12p11.1	34,050	326	Duplicación	3	1	0.03	0.5718
12	79,990,000	105,720,000	12q21.2	12q23.3	25,730	257	Duplicación de múltiples copias	4	1	0	1.0324
12	105,820,000	127,660,000	12q23.3	12q24.32	21,840	215	Duplicación	3	1.15	0.04	0.6507
14	48,190,000	51,790,000	14q21.3	14q22.1	3,600	37	Duplicación de múltiples copias	4	1	0	0.9279
15	22,370,000	31,175,000	15q11.2	15q13.2	8,805	57	Deleción	1	1	0	-1.3576
16	140,000	35,075,000	16p13.3	16p11.1	34,935	266	Duplicación	3	0.98	0.04	0.5833
16	66,740,000	90,090,000	16q22.1	16q24.3	23,350	220	Duplicación	3	1.35	0.04	0.7370
17	25,580,000	29,420,000	17q11.1	17q11.2	3,840	36	Duplicación	3	1.24	0.1	0.7001
17	31,485,000	33,485,000	17q11.2	17q12	2,000	21	Duplicación	3	0.68	0.12	0.4943
17	33,585,000	34,385,000	17q12	17q12	800	9	Duplicación de múltiples copias	8	1	0	1.9203

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana log ₂ ratio
17	34,865,000	36,065,000	17q12	17q12	1,200	13	Delección	1	1.15	0.09	-1.1481
17	36,515,000	38,110,000	17q12	17q21.1	1,595	16	Duplicación de múltiples copias	4	1	0	1.1435
17	38,210,000	41,035,000	17q21.1	17q21.31	2,825	28	Delección	1	1	0	-1.2725
17	45,925,000	76,720,000	17q21.32	17q25.3	30,795	296	Duplicación de múltiples copias	5	1	0	1.4485
18	30,790,000	77,921,820	18q12.1	18q23	47,132	467	Delección	1	0.55	0.02	-0.4898
19	54,150,000	57,495,000	19q13.42	19q13.43	3,345	31	Duplicación de múltiples copias	4	1	0	0.9421
20	14,555,000	15,455,000	20p12.1	20p12.1	900	10	Delección	1	1	0	-1.5122
20	29,885,000	37,797,085	20q11.21	20q12	7,912	79	Duplicación	3	0.86	0.06	0.5251
20	49,495,000	62,813,369	20q13.13	20q13.33	13,318	131	Duplicación	3	1.12	0.06	0.5555
21	15,530,000	47,965,000	21q11.2	21q22.3	32,435	320	Delección	1	0.49	0.03	-0.4034

CN= Número de copias; FA= Fracción alélica; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 4. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular SK-BR-3.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana log ₂ ratio
1	145,430,000	229,330,000	1q21.1	1q42.13	83,900	780	Duplicación	3	0.99	0.03	0.5642
3	140,000	4,310,000	3p26.3	3p26.1	4,170	41	Duplicación	3	0.64	0.1	0.4297
3	4,410,000	8,010,000	3p26.1	3p26.1	3,600	37	Duplicación	3	1.6	0	0.8610
3	13,010,000	14,210,000	3p25.2	3p25.1	1,200	13	Duplicación	3	1.26	0.2	0.6781
3	18,910,000	25,070,000	3p24.3	3p24.2	6,160	62	Duplicación	3	1.2	0.09	0.6572
3	37,970,000	39,570,000	3p22.2	3p22.1	1,600	17	Duplicación de múltiples copias	4	1	0	0.9564
3	138,960,000	142,115,000	3q23	3q23	3,155	32	Duplicación de múltiples copias	4	1	0	1.0495
3	151,085,000	151,585,000	3q25.1	3q25.1	500	6	Duplicación de múltiples copias	6	1	0	1.5239
3	152,185,000	156,785,000	3q25.2	3q25.31	4,600	47	Duplicación	3	1.15	0.11	0.6521
3	158,270,000	165,480,000	3q25.32	3q26.1	7,210	71	Duplicación	3	1.2	0.08	0.6707
3	169,080,000	172,665,000	3q26.2	3q26.31	3,585	36	Duplicación	3	1.6	0	0.8969

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana logratio
3	178,065,000	190,915,000	3q26.32	3q28	12,850	129	Duplicación de múltiples copias	4	1	0	0.9244
3	191,015,000	197,750,000	3q28	3q29	6,735	61	Duplicación	3	0.83	0.09	0.4805
4	125,000	48,947,096	4p16.3	4p11	48,822	468	Delección	1	0.88	0.02	-0.8656
4	88,500,000	120,740,000	4q22.1	4q26	32,240	316	Duplicación	3	0.65	0.04	0.4183
5	14,470,000	15,550,000	5p15.2	5p15.1	1,080	11	Delección	1	1.1	0.11	-1.1494
5	129,045,000	137,340,000	5q23.3	5q31.2	8,295	81	Delección	1	0.86	0.05	-0.8851
5	148,187,073	154,387,073	5q32	5q33.2	6,200	63	Duplicación	3	1.6	0	0.8995
6	440,000	7,440,000	6p25.3	6p24.3	7,000	71	Delección	1	0.76	0.05	-0.8150
6	8,840,000	25,040,000	6p24.3	6p22.3	16,200	163	Duplicación	3	1.22	0.06	0.6845
6	28,020,000	33,405,000	6p22.1	6p21.32	5,385	46	Doble delección	0	1	0	-2.6584
6	33,505,000	34,205,000	6p21.31	6p21.31	700	8	Duplicación	3	0.69	0.23	0.5059
7	95,000	57,365,000	7p22.3	7p11.2	57,270	526	Duplicación	3	1.6	0	0.8942
7	63,425,000	128,560,000	7q11.21	7q32.1	65,135	583	Duplicación de múltiples copias	4	1	0	0.9709
7	128,660,000	129,060,000	7q32.1	7q32.1	400	5	Delección	1	0.82	0.2	-0.8656
7	129,160,000	131,604,523	7q32.1	7q32.3	2,445	24	Duplicación de múltiples copias	6	1	0	1.6933
7	131,704,523	132,804,523	7q32.3	7q33	1,100	12	Delección	1	0.98	0.12	-1.0012
8	71,240,000	81,840,000	8q13.3	8q21.13	10,600	107	Duplicación de múltiples copias	13	1	0	2.6646
8	83,540,000	86,985,000	8q21.13	8q21.3	3,445	32	Duplicación	3	0.7	0.11	0.4256
8	87,085,000	92,520,000	8q21.3	8q21.3	5,435	55	Duplicación de múltiples copias	8	1	0	1.9222
8	92,620,000	98,620,000	8q21.3	8q22.1	6,000	61	Delección	1	1	0	-1.6787
8	104,755,000	108,055,000	8q22.3	8q23.1	3,300	34	Delección	1	1	0	-1.8074
8	108,955,000	109,755,000	8q23.1	8q23.1	800	9	Delección	1	1	0	-1.6147
8	109,855,000	126,655,000	8q23.1	8q24.13	16,800	169	Duplicación de múltiples copias	14	1	0	2.7583
8	127,855,000	129,055,000	8q24.21	8q24.21	1,200	13	Duplicación de múltiples copias	17	1	0	3.1177
8	131,655,000	132,255,000	8q24.22	8q24.22	600	7	Delección	1	0.85	0.16	-0.6314

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana logratio
8	132,355,000	132,655,000	8q24.22	8q24.22	300	4	Duplicación de múltiples copias	4	1	0	0.9753
8	132,755,000	133,255,000	8q24.22	8q24.22	500	6	Delección	1	1	0	-1.9045
8	137,330,000	142,130,000	8q24.23	8q24.3	4,800	49	Delección	1	0.92	0.06	-0.9143
10	29,855,000	33,255,000	10p11.23	10p11.22	3,400	34	Delección	1	0.9	0.08	-0.8656
10	57,025,000	73,490,000	10q21.1	10q22.1	16,465	165	Duplicación	3	1.43	0.06	0.7182
10	73,590,000	82,170,000	10q22.1	10q23.1	8,580	78	Duplicación de múltiples copias	4	1	0	1.0857
10	82,270,000	113,250,000	10q23.1	10q25.2	30,980	299	Delección	1	0.85	0.03	-0.8425
12	108,120,000	109,320,000	12q23.3	12q24.11	1,200	13	Delección	1	0.84	0.12	-0.7661
12	116,173,470	122,730,623	12q24.21	12q24.31	6,557	65	Delección	1	0.82	0.06	-0.7567
13	95,060,324	114,989,948	13q32.1	13q34	19,930	194	Duplicación	3	1.24	0.05	0.6891
14	52,295,000	53,095,000	14q22.1	14q22.1	800	9	Duplicación de múltiples copias	5	1	0	1.2971
14	57,995,000	58,495,000	14q22.3	14q23.1	500	6	Duplicación de múltiples copias	4	1	0	1.1255
14	85,505,000	92,485,000	14q31.3	14q32.12	6,980	70	Duplicación de múltiples copias	5	1	0	1.2410
16	61,640,000	90,090,000	16q21	16q24.3	28,450	271	Delección	1	0.9	0.03	-0.8656
17	19,390,000	21,070,000	17p11.2	17p11.2	1,680	11	Delección	1	0.92	0.13	-0.9164
17	25,580,000	29,165,000	17q11.1	17q11.2	3,585	34	Duplicación de múltiples copias	4	1	0	1.0137
17	29,265,000	37,210,000	17q11.2	17q12	7,945	69	Delección	1	0.9	0.05	-0.8656
17	37,310,000	40,535,000	17q12	17q21.2	3,225	32	Duplicación de múltiples copias	7	1	0	1.7294
17	40,635,000	48,725,000	17q21.2	17q21.33	8,090	71	Delección	1	0.89	0.05	-0.7859
17	49,625,000	60,425,000	17q21.33	17q23.2	10,800	104	Delección	1	0.88	0.05	-0.8656
17	60,695,000	69,320,000	17q23.2	17q24.3	8,625	81	Duplicación	3	1.51	0.09	0.7947
17	69,420,000	76,620,000	17q24.3	17q25.3	7,200	73	Duplicación	3	0.75	0.08	0.4436
17	76,720,000	78,175,000	17q25.3	17q25.3	1,455	13	Duplicación de múltiples copias	6	1	0	1.5460
17	78,275,000	81,009,049	17q25.3	17q25.3	2,734	27	Duplicación	3	0.76	0.13	0.4436
18	33,290,000	54,359,136	18q12.2	18q21.31	21,069	210	Delección	1	0.96	0.03	-0.9181

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana logratio
19	20,345,000	23,810,000	19p12	19p12	3,465	5	Doble deleción	0	1	0	-4.1446
20	110,000	25,555,000	20p13	20p11.21	25,445	253	Duplicación	3	1.6	0.03	0.8506
20	32,985,000	37,597,085	20q11.22	20q11.23	4,612	46	Duplicación de múltiples copias	5	1	0	1.3438
20	41,297,085	62,813,369	20q12	20q13.33	21,516	211	Duplicación de múltiples copias	5	1	0	1.3976
22	46,965,000	47,565,000	22q13.31	22q13.31	600	7	Deleción	1	1	0	-1.6724

CN= Número de copias; FA= Fracción alélica; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 5. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular MDA-MB-231.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana logratio
2	206,520,000	242,900,000	2q33.3	2q37.3	36,380	358	Deleción	1	0.56	0.02	-0.4594
3	108,160,000	162,070,000	3q13.13	3q26.1	53,910	530	Deleción	1	0.52	0.02	-0.4366
6	440,000	28,520,000	6p25.3	6p22.1	28,080	278	Duplicación	3	1.01	0.04	0.6101
6	28,620,000	33,405,000	6p22.1	6p21.32	4,785	40	Doble deleción	0	1	0	-2.0228
6	33,505,000	34,805,000	6p21.31	6p21.31	1,300	14	Duplicación	3	0.99	0.17	0.5682
6	34,905,000	43,105,000	6p21.31	6p21.1	8,200	83	Duplicación de múltiples copias	5	1	0	1.2224
6	43,205,000	57,130,000	6p21.1	6p11.2	13,925	137	Duplicación	3	1.04	0.05	0.6176
6	62,285,000	75,195,000	6q11.1	6q13	12,910	128	Duplicación	3	1	0.06	0.6110
6	78295000	166380000	6q14.1	6q27	88,085	872	Deleción	1	0.5	0.02	-0.4001
8	455,000	43,325,000	8p23.3	8p11.1	42,870	403	Deleción	1	0.5	0.02	-0.4064
8	106,755,000	113,355,000	8q23.1	8q23.3	6,600	67	Duplicación	3	0.71	0.07	0.4475
8	125,455,000	126,355,000	8q24.13	8q24.13	900	10	Deleción	1	1	0	-1.3294
9	3,665,000	18,065,000	9p24.2	9p22.2	14,400	145	Deleción	1	1	0	-1.4948
9	18,165,000	20,765,000	9p22.2	9p21.3	2,600	27	Deleción	1	0.57	0.09	-0.4867
9	20,865,000	26,035,000	9p21.3	9p21.2	5,170	51	Doble deleción	0	1	0	-4.8086
11	89,985,000	134,670,000	11q14.3	11q25	44,685	441	Duplicación	3	0.68	0.03	0.4339
13	19,550,000	114,989,948	13q12.11	13q34	95,440	925	Deleción	1	0.54	0.02	-0.4406
14	47,290,000	57,795,000	14q21.3	14q22.3	10,505	105	Duplicación	3	0.8	0.06	0.5040

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana log ₂ ratio
15	22,370,000	102,185,000	15q11.2	15q26.3	79,815	729	Delección	1	0.54	0.02	-0.4306
16	140,000	23,860,000	16p13.3	16p12.2	23,720	193	Delección	1	0.53	0.03	-0.4623
16	46,565,000	90,090,000	16q11.2	16q24.3	43,525	420	Delección	1	0.49	0.03	-0.4108
18	43,990,000	77,921,820	18q21.1	18q23	33,932	335	Delección	1	0.74	0.03	-0.6387
20	29,885,000	62,813,369	20q11.21	20q13.33	32,928	324	Duplicación	3	0.84	0.04	0.5215
21	15,530,000	21,330,000	21q11.2	21q21.1	5,800	59	Delección	1	0.5	0.06	-0.4163
22	17,570,000	51,064,777	22q11.1	22q13.33	33,495	301	Delección	1	0.57	0.03	-0.4790

CN= Número de copias; FA= Fracción alélica; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 6. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la línea celular Hs578T.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana log ₂ ratio
2	3,629,312	3,729,312	2p25.3	2p25.3	100	2	Duplicación de múltiples copias	18	1	0	3.1812
3	93,695,000	178,565,000	3q11.1	3q26.32	84,870	836	Duplicación	3	0.83	0.02	0.4552
3	178,665,000	197,750,000	3q26.32	3q29	19,085	184	Duplicación	3	1.6	0.05	0.8088
5	60,000	45,800,000	5p15.33	5p12	45,740	433	Duplicación	3	0.92	0.03	0.4883
6	440,000	57,130,000	6p25.3	6p11.2	56,690	552	Duplicación	3	1.25	0.03	0.6605
7	63,425,000	66,310,000	7q11.21	7q11.21	2,885	14	Duplicación	3	1.13	0.18	0.5768
8	455,000	43,325,000	8p23.3	8p11.1	42,870	403	Duplicación	3	1.33	0.04	0.6862
8	47,855,000	64,840,000	8q11.1	8q12.3	16,985	168	Duplicación	3	1.52	0.06	0.7727
8	64,940,000	132,655,000	8q12.3	8q24.22	67,715	674	Duplicación de múltiples copias	4	1	0	1.0453
8	132,755,000	146,150,000	8q24.22	8q24.3	13,395	129	Duplicación	3	1.6	0.02	0.8291
11	127,985,000	134,670,000	11q24.3	11q25	6,685	67	Delección	1	1.06	0.05	-1.1468
13	53,655,000	53,855,000	13q14.3	13q14.3	200	3	Doble delección	0	1	0	-5.9542
16	80,245,000	90,090,000	16q23.2	16q24.3	9,845	95	Duplicación	3	0.74	0.07	0.4312
17	31,385,000	33,285,000	17q11.2	17q12	1,900	20	Delección	1	0.73	0.09	-0.7021
18	185,000	13,990,000	18p11.32	18p11.21	13,805	135	Duplicación	3	0.72	0.05	0.4081

CN= Número de copias; FA= Fracción alélica; pb= pares de bases; kpb= kilopares de bases

13.1.2 Reportes de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina de líneas celulares tumorales empleando el software XCAVATOR

Tabla Suplementaria 7. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina en la línea celular MCF7.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
1	86,028,083	106,713,906	1p22.3	1p21.1	20,686	Delección	1	-0.9042
1	106,713,907	109,013,906	1p21.1	1p13.3	2,300	Duplicación	3	0.7052
1	114,113,907	115,713,906	1p13.2	1p13.2	1,600	Duplicación de múltiples copias	8	1.9758
1	142,535,435	152,509,645	1q12	1q21.3	9,974	Duplicación de múltiples copias	4	1.0386
1	158,409,646	168,309,645	1q23.1	1q24.2	9,900	Delección	1	-0.8666
1	198,409,646	207,882,221	1q31.3	1q32.2	9,473	Duplicación	4	0.8349
1	207,882,222	214,782,221	1q32.2	1q41	6,900	Delección	1	-0.9363
1	214,782,222	219,782,221	1q41	1q41	5,000	Duplicación de múltiples copias	4	1.0591
2	90,371,526	92,295,103	2p11.2	2p11.1	1,924	Duplicación de múltiples copias	7	1.8960
3	41,360,001	61,560,000	3p22.1	3p14.2	20,200	Delección	1	-0.8969
3	61,560,001	63,960,000	3p14.2	3p14.1	2,400	Duplicación de múltiples copias	9	2.1557
3	116,104,855	116,904,854	3q13.31	3q13.31	800	Delección	1	-1.7356
4	10,001	48,997,096	4p16.3	4p11	48,987	Delección	1	-0.4037
4	48,997,097	49,588,941	4p11	4p11	592	Duplicación de múltiples copias	9	2.1181
4	181,452,280	190,452,279	4q34.3	4q35.2	9,000	Delección	1	-1.0127
4	190,452,280	190,952,279	4q35.2	4q35.2	500	Duplicación de múltiples copias	4	1.0730
5	10,001	1,610,000	5p15.33	5p15.33	1,600	Duplicación de múltiples copias	5	1.2138
5	1,610,001	28,880,657	5p15.33	5p14.1	27,271	Duplicación	3	0.4270
5	168,888,728	180,888,727	5q35.1	5q35.3	12,000	Duplicación	3	0.5161
6	155,330,544	170,042,073	6q25.2	6q27	14,712	Delección	1	-0.9486
7	31,682,485	35,282,484	7p14.3	7p14.2	3,600	Duplicación	3	0.6278
7	57,510,632	58,010,631	7p11.2	7p11.1	500	Duplicación de múltiples copias	4	1.1343

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
7	100,365,725	114,206,043	7q22.1	7q31.1	13,840	Duplicación	3	0.6192
8	10,001	43,741,854	8p23.3	8p11.1	43,732	Delección	1	-0.9597
8	46,838,888	86,235,599	8q11.1	8q21.2	39,397	Duplicación	3	0.4805
8	86,235,600	129,926,451	8q21.2	8q24.21	43,691	Duplicación de múltiples copias	5	1.2164
8	129,926,452	146,232,588	8q24.21	8q24.3	16,306	Duplicación	3	0.6835
9	65,467,680	66,168,360	9q12	9q13	701	Delección	1	-0.7979
11	60,001	11,912,759	11p15.5	11p15.3	11,853	Delección	1	-0.9112
11	51,090,854	51,590,853	11p11.12	11p11.12	500	Delección	1	-0.5634
11	54,694,206	59,494,205	11q11	11q12.1	4,800	Delección	1	-0.9131
11	86,574,696	117,137,584	11q14.2	11q23.3	30,563	Delección	1	-0.9405
11	117,137,585	119,337,584	11q23.3	11q23.3	2,200	Duplicación	3	0.6029
11	119,337,585	134,937,584	11q23.3	11q25	15,600	Delección	1	-0.9664
12	37,856,695	38,256,694	12q11	12q12	400	Duplicación de múltiples copias	5	1.3090
13	48,120,001	60,320,000	13q14.2	13q21.2	12,200	Delección	1	-0.4012
13	60,320,001	115,039,948	13q21.2	13q34	54,720	Delección	1	-0.9618
14	19,000,001	107,200,000	14q11.1	14q32.33	88,200	Duplicación	3	0.4111
15	20,000,001	22,562,114	15q11.1	15q11.2	2,562	Duplicación	3	0.4915
15	49,709,444	54,609,443	15q21.2	15q21.3	4,900	Duplicación de múltiples copias	5	1.2047
15	63,009,444	102,434,473	15q22.2	15q26.3	39,425	Duplicación	3	0.7315
16	46,385,802	67,185,801	16q11.2	16q22.1	20,800	Delección	1	-0.9221
17	28,263,007	38,525,848	17q11.2	17q21.2	10,263	Delección	1	-0.9479
17	45,125,849	67,360,760	17q21.32	17q24.3	22,235	Duplicación de múltiples copias	4	1.1342
18	33,810,899	77,971,820	18q12.2	18q23	44,161	Delección	1	-0.9583
19	60,001	10,037,198	19p13.3	19p13.2	9,977	Delección	1	-0.9915
20	60,001	6,660,000	20p13	20p12.3	6,600	Delección	1	-0.4105
20	6,660,001	26,260,000	20p12.3	20p11.1	19,600	Duplicación	3	0.5761
20	29,419,570	30,803,908	20q11.21	20q11.21	1,384	Duplicación	3	0.5874
20	39,447,086	41,347,085	20q12	20q12	1,900	Duplicación de múltiples copias	5	1.2367

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
20	46,847,086	52,147,085	20q13.13	20q13.2	5,300	Duplicación	4	0.9061
20	53,547,086	62,963,369	20q13.2	20q13.33	9,416	Duplicación de múltiples copias	5	1.2855
21	9,411,194	11,097,896	21p11.2	21p11.1	1,687	Duplicación de múltiples copias	7	1.8475
21	14,338,130	39,338,129	21q11.2	21q22.13	25,000	Delección	1	-0.4365
21	39,338,130	48,082,664	21q22.13	21q22.3	8,745	Duplicación	3	0.7825
22	16,050,001	17,047,850	22q11.1	22q11.1	998	Duplicación	3	0.6585
22	17,047,851	35,109,431	22q11.1	22q12.3	18,062	Delección	1	-0.4065
22	35,109,432	51,214,777	22q12.3	22q13.33	16,105	Delección	1	-0.9411

CN= Número de copias; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 8. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina en la línea celular MDA-MB-361.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
1	142,535,435	145,410,724	1q12	1q21.1	2,875	Duplicación de múltiples copias	4	0.8245
2	91,595,104	92,295,103	2p11.1	2p11.1	700	Duplicación de múltiples copias	5	1.4495
2	108,226,172	156,590,582	2q12.3	2q24.1	48,364	Delección	1	-0.4681
4	48,997,097	49,588,941	4p11	4p11	592	Duplicación de múltiples copias	8	1.9779
5	10,001	46,380,657	5p15.33	5p11	46,371	Duplicación	3	0.4342
5	49,405,642	49,605,641	5q11.1	5q11.1	200	Duplicación de múltiples copias	6	1.5836
6	31,760,001	33,560,000	6p21.33	6p21.31	1,800	Delección	1	-0.4349
7	61,054,332	61,827,020	7q11.1	7q11.21	773	Duplicación	3	0.5101
7	61,967,158	71,667,157	7q11.21	7q11.22	9,700	Delección	1	-0.4839
7	71,667,158	76,665,724	7q11.22	7q11.23	4,999	Duplicación	3	0.4420
8	10,001	3,010,000	8p23.3	8p23.2	3,000	Delección	1	-0.5503
8	3,010,001	34,041,854	8p23.2	8p12	31,032	Delección	1	-1.3027
8	75,735,600	146,232,588	8q21.11	8q24.3	70,497	Duplicación	3	0.4695
9	10,001	30,410,000	9p24.3	9p21.1	30,400	Delección	1	-1.0717

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log₂ratio
9	30,410,001	37,310,000	9p21.1	9p13.2	6,900	Duplicación	3	0.5675
9	37,310,001	41,815,793	9p13.2	9p12	4,506	Delección	1	-0.5413
9	65,467,680	66,168,360	9q12	9q13	701	Delección	1	-0.6962
10	60,001	5,160,000	10p15.3	10p15.1	5,100	Delección	1	-0.4516
10	11,660,001	39,068,835	10p14	10p11.1	27,409	Delección	1	-0.4645
10	70,948,846	74,948,845	10q22.1	10q22.2	4,000	Delección	1	-0.4260
10	129,066,070	135,427,527	10q26.2	10q26.3	6,361	Delección	1	-0.4998
11	60,001	6,912,759	11p15.5	11p15.4	6,853	Delección	1	-1.2565
11	6,912,760	51,590,853	11p15.4	11p11.12	44,678	Delección	1	-0.4345
11	68,694,206	70,274,695	11q13.3	11q13.3	1,580	Duplicación de múltiples copias	4	1.1232
12	145,740	34,839,876	12p13.33	12p11.1	34,694	Duplicación	3	0.4538
12	37,856,695	38,156,694	12q11	12q11	300	Duplicación de múltiples copias	6	1.5674
12	80,056,695	105,856,694	12q21.2	12q23.3	25,800	Duplicación de múltiples copias	4	0.9065
12	105,856,695	127,680,623	12q23.3	12q24.32	21,824	Duplicación	3	0.4649
14	19,000,001	20,400,000	14q11.1	14q11.2	1,400	Duplicación de múltiples copias	4	1.0553
14	48,100,001	51,800,000	14q21.3	14q22.1	3,700	Duplicación	3	0.7002
15	22,746,194	30,409,443	15q11.2	15q13.2	7,663	Delección	1	-1.2047
16	60,001	32,086,921	16p13.3	16p11.2	32,027	Duplicación	3	0.4872
16	32,086,922	35,273,150	16p11.2	16p11.1	3,186	Duplicación de múltiples copias	4	1.0823
16	46,385,802	51,985,801	16q11.2	16q12.1	5,600	Delección	2	-0.4111
16	66,785,802	90,239,383	16q22.1	16q24.3	23,454	Duplicación	3	0.5428
17	25,263,007	29,463,006	17q11.1	17q11.2	4,200	Duplicación	3	0.5112
17	31,463,007	34,563,006	17q11.2	17q12	3,100	Duplicación	3	0.4809
17	34,563,007	36,225,848	17q12	17q12	1,663	Delección	1	-1.0825
17	36,225,849	38,125,848	17q12	17q21.1	1,900	Duplicación	3	0.7519
17	38,125,849	41,025,848	17q21.1	17q21.31	2,900	Delección	1	-1.0488
17	45,825,849	51,125,848	17q21.32	17q22	5,300	Duplicación de múltiples copias	5	1.2446

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
17	51,125,849	56,425,848	17q22	17q22	5,300	Duplicación	3	0.5170
17	56,425,849	76,860,760	17q22	17q25.3	20,435	Duplicación de múltiples copias	5	1.3461
18	30,810,899	77,971,820	18q12.1	18q23	47,161	Delección	1	-0.5156
19	54,131,783	57,631,782	19q13.42	19q13.43	3,500	Duplicación	3	0.6798
20	25,660,001	26,260,000	20p11.1	20p11.1	600	Duplicación	3	0.6140
20	29,419,570	37,847,085	20q11.21	20q12	8,428	Duplicación	3	0.4705
20	49,447,086	62,963,369	20q13.13	20q13.33	13,516	Duplicación	3	0.5292
21	9,411,194	11,097,896	21p11.2	21p11.1	1,687	Duplicación de múltiples copias	5	1.3985
21	14,338,130	14,538,129	21q11.2	21q11.2	200	Duplicación de múltiples copias	4	1.0981
21	14,538,130	48,082,664	21q11.2	21q22.3	33,545	Delección	1	-0.5178
22	16,050,001	16,947,850	22q11.1	22q11.1	898	Duplicación de múltiples copias	4	1.0477
22	39,909,432	51,214,777	22q13.1	22q13.33	11,305	Delección	1	-0.4501

CN= Número de copias; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 9. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina en la línea celular SK-BR-3.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
1	10,001	1,721,368	1p36.33	1p36.33	1,711	Duplicación	3	0.5342
1	16,107,163	17,375,658	1p36.21	1p36.13	1,268	Duplicación	4	0.8594
1	120,413,907	121,386,695	1p12	1p11.2	973	Duplicación de múltiples copias	8	1.9330
1	142,535,435	162,809,645	1q12	1q23.3	20,274	Duplicación	3	0.7463
1	162,809,646	229,797,846	1q23.3	1q42.13	66,988	Duplicación	3	0.6282
2	32,578,114	32,878,113	2p22.3	2p22.3	300	Duplicación de múltiples copias	27	3.7450
2	89,830,437	92,295,103	2p11.2	2p11.1	2,465	Duplicación de múltiples copias	8	1.9487
2	111,351,338	114,051,337	2q13	2q13	2,700	Duplicación	3	0.7429
2	140,951,338	142,051,337	2q22.1	2q22.1	1,100	Duplicación de múltiples copias	8	2.0468
2	169,890,583	170,190,582	2q31.1	2q31.1	300	Duplicación de múltiples copias	22	3.4558

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
2	179,390,583	179,690,582	2q31.2	2q31.2	300	Duplicación de múltiples copias	126	5.9818
2	201,690,583	202,790,582	2q33.1	2q33.1	1,100	Duplicación de múltiples copias	7	1.8851
2	209,090,583	209,290,582	2q34	2q34	200	Duplicación de múltiples copias	17	3.0569
2	216,090,583	216,390,582	2q35	2q35	300	Duplicación de múltiples copias	7	1.7349
2	230,990,583	231,490,582	2q36.3	2q37.1	500	Duplicación de múltiples copias	6	1.5332
3	4,360,001	8,060,000	3p26.1	3p26.1	3,700	Duplicación	4	0.8795
3	36,660,001	39,560,000	3p22.2	3p22.1	2,900	Duplicación	4	0.9150
3	50,260,001	51,460,000	3p21.31	3p21.2	1,200	Duplicación	3	0.7388
3	138,904,855	142,304,854	3q23	3q23	3,400	Duplicación de múltiples copias	4	0.9753
3	150,904,855	165,504,854	3q25.1	3q26.1	14,600	Duplicación	3	0.5038
3	168,804,855	172,704,854	3q26.2	3q26.31	3,900	Duplicación de múltiples copias	4	1.0160
3	178,004,855	193,104,854	3q26.32	3q29	15,100	Duplicación	4	0.8492
4	10,001	48,997,096	4p16.3	4p11	48,987	Delección	1	-0.8892
4	48,997,097	49,588,941	4p11	4p11	592	Duplicación de múltiples copias	5	1.4350
4	190,352,280	190,952,279	4q35.2	4q35.2	600	Duplicación de múltiples copias	4	1.1041
5	14,210,001	14,510,000	5p15.2	5p15.2	300	Duplicación de múltiples copias	6	1.6233
5	14,510,001	15,610,000	5p15.2	5p15.1	1,100	Delección	1	-1.1429
5	79,905,642	80,605,641	5q14.1	5q14.1	700	Duplicación de múltiples copias	6	1.5252
5	127,586,129	127,886,128	5q23.3	5q23.3	300	Duplicación de múltiples copias	10	2.3793
5	127,886,129	134,286,128	5q23.3	5q31.1	6,400	Delección	1	-0.7260
5	134,286,129	137,586,128	5q31.1	5q31.2	3,300	Delección	1	-0.8823
5	137,586,129	142,837,073	5q31.2	5q31.3	5,251	Duplicación	3	0.5402
5	147,937,074	154,437,073	5q32	5q33.2	6,500	Duplicación	3	0.7751
6	60,001	7,360,000	6p25.3	6p24.3	7,300	Delección	1	-0.7810
6	8,260,001	8,760,000	6p24.3	6p24.3	500	Delección	1	-0.8369
6	8,760,001	25,960,000	6p24.3	6p22.2	17,200	Duplicación	3	0.6364
6	38,660,001	39,060,000	6p21.2	6p21.2	400	Duplicación de múltiples copias	15	2.8788

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log₂ratio
6	51,460,001	51,960,000	6p12.3	6p12.2	500	Duplicación de múltiples copias	9	2.2252
6	56,260,001	56,660,000	6p12.1	6p12.1	400	Duplicación de múltiples copias	29	3.8422
6	79,578,590	79,778,589	6q14.1	6q14.1	200	Duplicación de múltiples copias	15	2.8693
6	117,530,544	117,930,543	6q22.1	6q22.1	400	Duplicación de múltiples copias	6	1.5353
7	10,001	58,010,631	7p22.3	7p11.1	58,001	Duplicación	4	0.8222
7	61,054,332	131,654,523	7q11.1	7q32.3	70,600	Duplicación	4	0.8811
7	131,654,524	132,854,523	7q32.3	7q33	1,200	Delección	1	-0.9898
7	151,797,898	152,097,897	7q36.1	7q36.1	300	Duplicación de múltiples copias	31	3.9551
8	37,941,855	38,341,854	8p11.23	8p11.22	400	Duplicación de múltiples copias	9	2.1418
8	49,435,600	50,835,599	8q11.21	8q11.21	1,400	Delección	1	-0.6943
8	67,335,600	81,935,599	8q13.1	8q21.13	14,600	Duplicación de múltiples copias	10	2.2672
8	86,726,452	92,626,451	8q21.2	8q21.3	5,900	Duplicación de múltiples copias	8	1.9248
8	92,626,452	98,726,451	8q21.3	8q22.1	6,100	Delección	1	-1.7810
8	99,526,452	101,226,451	8q22.2	8q22.2	1,700	Duplicación de múltiples copias	6	1.6014
8	105,026,452	109,726,451	8q22.3	8q23.1	4,700	Delección	1	-1.7755
8	109,726,452	126,726,451	8q23.1	8q24.13	17,000	Duplicación de múltiples copias	16	2.9794
8	127,726,452	129,126,451	8q24.21	8q24.21	1,400	Duplicación de múltiples copias	18	3.1344
8	132,726,452	133,326,451	8q24.22	8q24.22	600	Delección	1	-1.6496
8	137,326,452	141,626,451	8q24.23	8q24.3	4,300	Delección	1	-0.9494
9	8,310,001	8,610,000	9p24.1	9p24.1	300	Duplicación de múltiples copias	9	2.1173
9	35,510,001	35,810,000	9p13.3	9p13.3	300	Duplicación de múltiples copias	9	2.1202
10	16,860,001	17,260,000	10p13	10p13	400	Duplicación de múltiples copias	16	2.9548
10	29,824,676	33,124,675	10p11.23	10p11.22	3,300	Delección	1	-0.8307
10	42,354,936	42,696,687	10q11.21	10q11.21	342	Duplicación de múltiples copias	168	6.3892
10	55,548,846	69,548,845	10q21.1	10q21.3	14,000	Duplicación	3	0.6085

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log₂ratio
10	69,548,846	82,248,845	10q21.3	10q23.1	12,700	Duplicación de múltiples copias	4	0.9313
10	82,248,846	113,048,845	10q23.1	10q25.2	30,800	Delección	1	-0.8593
11	107,937,585	108,837,584	11q22.3	11q22.3	900	Duplicación de múltiples copias	5	1.4220
12	145,740	1,445,739	12p13.33	12p13.33	1,300	Duplicación de múltiples copias	5	1.4326
12	18,339,877	19,539,876	12p12.3	12p12.3	1,200	Duplicación	3	0.8027
12	26,539,877	26,939,876	12p11.23	12p11.23	400	Duplicación de múltiples copias	9	2.2338
12	40,556,695	42,056,694	12q12	12q12	1,500	Duplicación de múltiples copias	4	1.1354
12	78,156,695	78,656,694	12q21.2	12q21.2	500	Duplicación de múltiples copias	15	2.9024
12	108,056,695	109,356,694	12q23.3	12q24.11	1,300	Delección	1	-0.9011
12	109,423,471	110,023,470	12q24.11	12q24.11	600	Duplicación de múltiples copias	5	1.2244
13	23,820,001	24,020,000	13q12.12	13q12.12	200	Duplicación de múltiples copias	27	3.7713
13	28,120,001	29,620,000	13q12.2	13q12.3	1,500	Duplicación	4	0.9200
13	48,820,001	49,320,000	13q14.2	13q14.2	500	Duplicación de múltiples copias	8	2.0076
13	95,110,325	115,039,948	13q32.1	13q34	19,930	Duplicación	3	0.5007
14	19,000,001	23,100,000	14q11.1	14q11.2	4,100	Duplicación	3	0.6204
14	50,100,001	53,100,000	14q21.3	14q22.1	3,000	Duplicación	3	0.7007
14	85,500,001	92,500,000	14q31.3	14q32.12	7,000	Duplicación de múltiples copias	4	1.1097
14	93,900,001	94,200,000	14q32.12	14q32.12	300	Duplicación de múltiples copias	11	2.4750
15	20,000,001	23,864,853	15q11.1	15q11.2	3,865	Duplicación de múltiples copias	5	1.1784
15	28,264,854	29,064,853	15q13.1	15q13.1	800	Duplicación de múltiples copias	6	1.6163
15	40,209,444	43,809,443	15q15.1	15q15.3	3,600	Duplicación	3	0.6224
16	29,386,922	35,273,150	16p11.2	16p11.1	5,886	Duplicación	3	0.5539
16	46,385,802	46,785,801	16q11.2	16q11.2	400	Duplicación de múltiples copias	5	1.2790
16	53,185,802	53,585,801	16q12.2	16q12.2	400	Duplicación de múltiples copias	16	2.9776
16	61,485,802	90,239,383	16q21	16q24.3	28,754	Delección	1	-0.8606
17	396,627	10,196,626	17p13.3	17p13.1	9,800	Delección	1	-0.5526

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
17	10,196,627	10,596,626	17p13.1	17p13.1	400	Duplicación de múltiples copias	12	2.5523
17	10,596,627	22,166,608	17p13.1	17p11.2	11,570	Delección	1	-0.5765
17	25,263,007	29,763,006	17q11.1	17q11.2	4,500	Duplicación de múltiples copias	4	1.1568
17	29,763,007	34,063,006	17q11.2	17q12	4,300	Delección	1	-0.8784
17	34,063,007	37,225,848	17q12	17q12	3,163	Delección	1	-0.6278
17	37,225,849	40,525,848	17q12	17q21.2	3,300	Duplicación de múltiples copias	9	2.1889
17	41,825,849	56,525,848	17q21.31	17q22	14,700	Delección	1	-0.8740
17	58,325,849	59,625,848	17q23.2	17q23.2	1,300	Delección	1	-0.9132
17	59,625,849	68,860,760	17q23.2	17q24.3	9,235	Duplicación	4	0.8773
17	76,760,761	79,696,461	17q25.3	17q25.3	2,936	Duplicación de múltiples copias	5	1.2729
18	10,001	210,000	18p11.32	18p11.32	200	Duplicación de múltiples copias	13	2.7471
18	18,510,899	18,710,898	18q11.1	18q11.1	200	Duplicación de múltiples copias	21	3.3692
18	33,210,899	54,509,136	18q12.2	18q21.31	21,298	Delección	1	-0.9120
18	56,109,137	56,509,136	18q21.31	18q21.32	400	Duplicación de múltiples copias	7	1.8868
19	20,337,199	24,573,415	19p12	19p11	4,236	Doble delección	0	-3.8916
19	27,731,783	27,931,782	19q11	19q11	200	Duplicación de múltiples copias	21	3.4086
20	60,001	26,260,000	20p13	20p11.1	26,200	Duplicación	4	0.8317
20	32,903,909	37,647,085	20q11.22	20q12	4,743	Duplicación de múltiples copias	5	1.3193
20	41,247,086	62,963,369	20q12	20q13.33	21,716	Duplicación de múltiples copias	5	1.4078
21	9,411,194	11,097,896	21p11.2	21p11.1	1,687	Duplicación de múltiples copias	8	1.9189
22	16,050,001	16,450,000	22q11.1	22q11.1	400	Duplicación de múltiples copias	5	1.2917
22	46,909,432	47,609,431	22q13.31	22q13.31	700	Delección	1	-1.5071

CN= Número de copias; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 10. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación Illumina en la línea celular MDA-MB-231.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
1	16,807,163	17,275,658	1p36.13	1p36.13	468	Duplicación de múltiples copias	4	1.1300
1	142,535,435	145,410,724	1q12	1q21.1	2,875	Duplicación de múltiples copias	4	1.1639
1	148,176,039	149,404,460	1q21.2	1q21.2	1,228	Duplicación	4	0.8143
2	91,595,104	92,095,103	2p11.1	2p11.1	500	Duplicación de múltiples copias	7	1.7120
2	206,290,583	243,009,132	2q33.3	2q37.3	36,719	Delección	1	-0.4882
3	75,670,271	75,870,270	3p12.3	3p12.3	200	Duplicación de múltiples copias	6	1.6984
3	90,270,271	90,470,270	3p11.1	3p11.1	200	Duplicación de múltiples copias	5	1.3997
3	108,404,855	162,204,854	3q13.13	3q26.1	53,800	Delección	1	-0.4607
4	49,197,097	49,588,941	4p11	4p11	392	Duplicación de múltiples copias	9	2.1448
4	190,452,280	190,952,279	4q35.2	4q35.2	500	Duplicación de múltiples copias	4	1.1363
6	60,001	34,860,000	6p25.3	6p21.31	34,800	Duplicación	3	0.6738
6	34,860,001	43,260,000	6p21.31	6p21.1	8,400	Duplicación de múltiples copias	5	1.3111
6	43,260,001	58,737,659	6p21.1	6p11.1	15,478	Duplicación	3	0.6985
6	61,880,167	74,978,589	6q11.1	6q13	13,098	Duplicación	3	0.6740
6	74,978,590	78,178,589	6q13	6q14.1	3,200	Duplicación	3	0.4069
6	78,178,590	166,291,300	6q14.1	6q27	88,113	Delección	1	-0.4237
7	57,510,632	58,010,631	7p11.2	7p11.1	500	Duplicación de múltiples copias	4	0.9814
7	61,054,332	61,827,020	7q11.1	7q11.21	773	Duplicación	3	0.6682
8	10,001	43,741,854	8p23.3	8p11.1	43,732	Delección	1	-0.4326
8	46,838,888	81,135,599	8q11.1	8q21.13	34,297	Delección	1	-0.4242
8	113,426,452	116,126,451	8q23.3	8q23.3	2,700	Delección	1	-0.4257
8	125,426,452	126,426,451	8q24.13	8q24.13	1,000	Delección	1	-1.4646
9	3,610,001	18,110,000	9p24.2	9p22.2	14,500	Delección	1	-1.4678
9	18,110,001	20,810,000	9p22.2	9p21.3	2,700	Delección	1	-0.4286

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log₂ratio
9	20,810,001	26,110,000	9p21.3	9p21.2	5,300	Doble deleción	0	-5.3460
9	39,210,001	41,815,793	9p13.1	9p12	2,606	Deleción	1	-0.5891
9	43,313,699	43,813,698	9p12	9p11.2	500	Deleción	1	-1.9902
9	43,813,699	44,596,569	9p11.2	9p11.2	783	Deleción	1	-0.4310
9	68,337,999	69,088,946	9q13	9q21.11	751	Duplicación de múltiples copias	7	1.7916
10	42,354,936	42,896,687	10q11.21	10q11.21	542	Duplicación	3	0.5229
10	46,296,688	46,876,964	10q11.22	10q11.22	580	Deleción	1	-0.7471
10	46,876,965	47,176,964	10q11.22	10q11.22	300	Duplicación	3	0.7151
11	48,712,760	49,012,759	11p11.2	11p11.12	300	Duplicación de múltiples copias	5	1.3638
11	54,694,206	55,094,205	11q11	11q11	400	Duplicación	3	0.5989
11	88,438,379	134,937,584	11q14.3	11q25	46,499	Duplicación	3	0.4427
12	37,856,695	38,156,694	12q11	12q11	300	Duplicación de múltiples copias	6	1.5644
12	124,980,624	133,806,992	12q24.31	12q24.33	8,826	Deleción	1	-0.4463
13	19,020,001	115,039,948	13q11	13q34	96,020	Deleción	1	-0.4934
14	19,000,001	20,200,000	14q11.1	14q11.2	1,200	Duplicación	3	0.7530
15	20,000,001	20,200,000	15q11.1	15q11.1	200	Duplicación de múltiples copias	4	1.1520
15	22,262,115	102,434,473	15q11.2	15q26.3	80,172	Deleción	1	-0.4837
16	60,001	32,086,921	16p13.3	16p11.2	32,027	Deleción	1	-0.4580
16	34,273,151	35,273,150	16p11.2	16p11.1	1,000	Deleción	1	-0.5355
16	46,385,802	70,885,801	16q11.2	16q22.2	24,500	Deleción	1	-0.5206
16	70,885,802	71,185,801	16q22.2	16q22.2	300	Duplicación	4	0.8119
16	71,185,802	90,239,383	16q22.2	16q24.3	19,054	Deleción	1	-0.5069
18	10,001	210,000	18p11.32	18p11.32	200	Duplicación de múltiples copias	4	1.0520
18	43,910,899	77,971,820	18q21.1	18q23	34,061	Deleción	1	-0.5574
19	27,731,783	27,931,782	19q11	19q11	200	Duplicación de múltiples copias	4	1.1352
20	25,660,001	26,260,000	20p11.1	20p11.1	600	Duplicación	3	0.6253
20	29,419,570	29,903,908	20q11.21	20q11.21	484	Duplicación de múltiples copias	7	1.7115

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Mediana log ₂ ratio
20	29,903,909	62,963,369	20q11.21	20q13.33	33,059	Duplicación	3	0.4670
21	9,411,194	11,097,896	21p11.2	21p11.1	1,687	Duplicación de múltiples copias	6	1.5410
21	14,338,130	21,138,129	21q11.2	21q21.1	6,800	Delección	1	-0.5372
22	16,050,001	17,047,850	22q11.1	22q11.1	998	Duplicación	3	0.6215
22	17,047,851	51,214,777	22q11.1	22q13.33	34,167	Delección	1	-0.5124

CN= Número de copias; pb= pares de bases; kpb= kilopares de bases

13.2 Muestras de pacientes con cáncer de mama

13.2.1 Reportes de alteraciones en el número de copias de DNA detectadas mediante microarreglos de alta resolución en muestras de biopsia de tejido tumoral empleando el software Chromosome Analysis Suite

Tabla Suplementaria 11. Reporte de alteraciones en el número de copias de DNA detectadas mediante microarreglos de alta resolución en la muestra de biopsia de tejido tumoral BL-147.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Importancia clínica
1	145,038,815	249,212,878	1q21.1	1q44	104,174	Duplicación	3	Patogénica
1	35,578,909	37,003,273	1p34.3	1p34.3	1,424	Delección	1	Desconocida
1	45,798,555	46,495,942	1p34.1	1p34.1	697	Delección	1	Desconocida
11	62,083,908	66,309,040	11q12.3	11q13.2	4,225	Delección	1	Patogénica
11	84,543,946	134,938,847	11q14.1	11q25	50,395	Delección	1	Patogénica
11	51,196,206	62,068,685	11p11.12	11q12.3	10,872	Delección	1	Patogénica
11	68,044,804	71,344,441	11q13.2	11q13.4	3,300	Duplicación	3	Patogénica
11	71,407,173	72,181,733	11q13.4	11q13.4	775	Delección	1	Desconocida
11	80,790,741	81,599,288	11q14.1	11q14.1	809	Duplicación	3	Desconocida
16	32,608,733	90,158,005	16p11.2	16q24.3	57,549	Delección	1	Patogénica
20	19,836,217	20,838,118	20p11.23	20p11.23	1,002	Delección	1	Desconocida

CN= Número de copias; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 12. Reporte de alteraciones en el número de copias de DNA detectadas mediante microarreglos de alta resolución en la muestra de biopsia de tejido tumoral BL-158.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Importancia clínica
1	754,192	28,906,940	1p36.33	1p35.3	28,153	Delección	1	Patogénica
1	210,644,714	249,212,878	1q32.2	1q44	38,568	Delección	1	Patogénica
1	44,171,266	46,567,425	1p34.1	1p34.1	2,396	Delección	1	Patogénica
1	187,064,038	202,470,658	1q31.1	1q32.1	15,407	Delección	1	Patogénica
1	153,718,746	155,812,355	1q21.3	1q22	2,094	Duplicación de múltiples copias	6	Probablemente patogénica
1	31,731,208	33,540,018	1p35.2	1p35.1	1,809	Duplicación de múltiples copias	6	Desconocida
1	150,276,022	150,990,292	1q21.2	1q21.3	714	Delección	1	Desconocida
1	30,509,439	31,426,815	1p35.2	1p35.2	917	Duplicación de múltiples copias	5	Desconocida
1	42,978,550	44,161,980	1p34.2	1p34.1	1,183	Duplicación de múltiples copias	5	Desconocida
1	47,055,992	48,588,004	1p33	1p33	1,532	Delección	1	Desconocida
1	161,153,836	162,440,134	1q23.3	1q23.3	1,286	Duplicación de múltiples copias	5	Desconocida
1	164,421,385	165,695,726	1q23.3	1q24.1	1,274	Duplicación de múltiples copias	5	Desconocida
1	165,787,477	168,849,822	1q24.1	1q24.2	3,062	Duplicación de múltiples copias	4	Desconocida
1	168,852,043	171,151,991	1q24.2	1q24.3	2,300	Duplicación	3	Desconocida
1	171,168,545	173,541,768	1q24.3	1q25.1	2,373	Duplicación de múltiples copias	4	Desconocida
1	173,558,234	174,928,375	1q25.1	1q25.1	1,370	Duplicación de múltiples copias	5	Desconocida
1	174,951,068	176,001,826	1q25.1	1q25.2	1,051	Duplicación de múltiples copias	4	Desconocida
1	179,847,877	180,930,050	1q25.2	1q25.3	1,082	Duplicación de múltiples copias	6	Desconocida
1	180,931,508	182,777,751	1q25.3	1q25.3	1,846	Duplicación de múltiples copias	4	Desconocida
1	182,809,835	184,110,720	1q25.3	1q25.3	1,301	Duplicación	3	Desconocida
1	184,123,374	185,542,871	1q25.3	1q25.3	1,419	Duplicación de múltiples copias	7	Desconocida
1	185,563,418	186,390,303	1q25.3	1q31.1	827	Duplicación de múltiples copias	4	Desconocida

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Estado de CN	CN	Importancia clínica
1	186,401,206	187,054,897	1q31.1	1q31.1	654	Duplicación de múltiples copias	6	Desconocida
1	202,503,815	204,717,211	1q32.1	1q32.1	2,213	Duplicación de múltiples copias	5	Desconocida
1	206,875,697	208,212,397	1q32.1	1q32.2	1,337	Duplicación de múltiples copias	10	Desconocida
1	209,618,109	210,633,845	1q32.2	1q32.2	1,016	Duplicación de múltiples copias	5	Desconocida
5	101,349,874	102,526,175	5q21.1	5q21.1	1,176	Delección	1	Desconocida
6	57,797,218	170,913,051	6p11.2	6q27	113,116	Delección	1	Patogénica
8	3,093,409	39,114,949	8p23.2	8p11.22	36,022	Duplicación de múltiples copias	4	Patogénica
8	39,412,457	43,767,534	8p11.22	8p11.1	4,355	Duplicación de múltiples copias	4	Desconocida
8	128,744,202	140,467,205	8q24.21	8q24.3	11,723	Duplicación de múltiples copias	4	Desconocida
11	54,795,357	87,220,603	11q11	11q14.2	32,425	Delección	1	Patogénica
11	104,769,543	134,938,847	11q22.3	11q25	30,169	Delección	1	Patogénica
11	92,573,911	100,978,052	11q14.3	11q22.1	8,404	Delección	1	Probablemente patogénica
11	101,880,373	102,669,409	11q22.1	11q22.2	789	Delección	1	Desconocida
11	90,364,163	91,547,465	11q14.3	11q14.3	1,183	Delección	1	Desconocida
14	65,980,170	107,282,024	14q23.2	14q32.33	41,302	Delección	1	Patogénica
15	93,263,066	94,798,406	15q26.1	15q26.2	1,535	Delección	1	Patogénica
15	53,025,692	57,565,148	15q21.3	15q21.3	4,539	Delección	1	Probablemente patogénica
15	58,595,933	62,171,479	15q21.3	15q22.2	3,576	Delección	1	Desconocida
15	100,246,452	102,397,317	15q26.3	15q26.3	2,151	Delección	1	Desconocida
16	46,461,309	90,158,005	16q11.2	16q24.3	43,697	Delección	1	Patogénica
16	34,326,140	35,271,725	16p11.2	16p11.1	946	Duplicación	3	Benigna
22	27,062,038	33,786,170	22q12.1	22q12.3	6,724	Delección	1	Patogénica
22	36,418,475	41,961,831	22q12.3	22q13.2	5,543	Delección	1	Patogénica
22	42,243,618	45,254,083	22q13.2	22q13.31	3,010	Delección	1	Patogénica
22	24,394,088	25,944,478	22q11.23	22q12.1	1,550	Delección	1	Desconocida
22	16,054,713	17,989,378	22q11.1	22q11.21	1,935	Delección	1	Desconocida

CN= Número de copias; pb= pares de bases; kpb= kilopares de bases

13.2.2 Reportes de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en muestras de biopsia líquida empleando el software NanoGLADIATOR

Tabla Suplementaria 13. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la muestra de biopsia líquida BL-147.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana log ₂ ratio
1	57,790,000	57,890,000	1p32.2	1p32.2	100	2	Duplicación	3	0.77	0.17	0.4542
1	90,660,000	90,760,000	1p22.2	1p22.2	100	2	Duplicación de múltiples copias	6	1	0	1.5412
2	153,010,000	153,110,000	2q23.3	2q23.3	100	2	Duplicación de múltiples copias	9	1	0	2.1539
2	162,185,000	162,285,000	2q24.2	2q24.2	100	2	Duplicación de múltiples copias	6	1	0	1.5623
3	25,070,000	25,170,000	3p24.2	3p24.2	100	2	Duplicación	3	0.76	0.17	0.4550
4	1,728,646	1,928,646	4p16.3	4p16.3	200	3	Duplicación de múltiples copias	7	1	0	1.8604
5	130,745,000	130,845,000	5q31.1	5q31.1	100	2	Duplicación	3	0.82	0.17	0.4938
6	119,520,000	119,620,000	6q22.31	6q22.31	100	2	Duplicación de múltiples copias	4	1	0	1.0973
6	124,920,000	125,120,000	6q22.31	6q22.31	200	3	Duplicación	3	0.84	0.14	0.6774
7	151,985,000	152,085,000	7q36.1	7q36.1	100	2	Duplicación	3	1	0.17	0.6045
8	4,605,000	4,905,000	8p23.2	8p23.2	300	4	Duplicación	3	1	0.13	0.5079
8	22,110,000	22,210,000	8p21.3	8p21.3	100	2	Duplicación	3	1	0.07	0.8442
8	32,110,000	32,210,000	8p12	8p12	100	2	Duplicación	3	0.96	0.17	0.5465
8	39,250,000	39,350,000	8p11.22	8p11.22	100	2	Delección	1	0.76	0.11	-0.6852
8	89,385,000	89,585,000	8q21.3	8q21.3	200	3	Duplicación	3	1	0.14	0.8611
9	14,965,000	15,165,000	9p22.3	9p22.3	200	3	Duplicación	3	1	0.15	0.7533
11	12,595,000	12,695,000	11p15.3	11p15.3	100	2	Duplicación	3	1	0.18	0.5924
11	119,485,000	119,585,000	11q23.3	11q23.3	100	2	Duplicación	3	0.82	0.17	0.4724
13	28,010,000	28,110,000	13q12.2	13q12.2	100	2	Duplicación de múltiples copias	8	1	0	1.9907
14	70,595,000	71,100,000	14q24.2	14q24.2	505	4	Duplicación	3	0.89	0.12	0.4798
16	25,560,000	25,660,000	16p12.1	16p12.1	100	2	Duplicación	3	0.79	0.17	0.4496

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana logratio
17	41,335,000	41,435,000	17q21.31	17q21.31	100	2	Duplicación	3	1.0	0.18	0.6020
17	45,185,000	45,285,000	17q21.32	17q21.32	100	2	Duplicación de múltiples copias	4	1	0	0.9658
17	59,955,000	60,155,000	17q23.2	17q23.2	200	3	Duplicación	3	1	0.15	0.7013
20	54,545,000	54,645,000	20q13.2	20q13.2	100	2	Duplicación	3	0.81	0.17	0.4752

CN= Número de copias; pb= pares de bases; kpb= kilopares de bases

Tabla Suplementaria 14. Reporte de alteraciones en el número de copias de DNA detectadas mediante secuenciación genómica con nanoporos en la muestra de biopsia líquida BL-158.

Cromosoma	Posición de inicio (pb)	Posición de fin (pb)	Citobanda de inicio	Citobanda de fin	Tamaño (kpb)	Num. de ventanas genómicas	Estado de CN	CN	FA	Error de FA	Mediana logratio
2	162,185,000	162,285,000	2q24.2	2q24.2	100	2	Duplicación de múltiples copias	4	1	0	0.9835
4	1,728,646	1,928,646	4p16.3	4p16.3	200	3	Duplicación de múltiples copias	4	1	0	1.1030
4	69,230,000	69,740,000	4q13.2	4q13.2	510	3	Delección	1	1	0.08	-1.0324
6	28,520,000	33,405,000	6p22.1	6p21.32	4,885	41	Doble delección	0	1	0	-2.7923
6	89,815,000	89,915,000	6q15	6q15	100	2	Duplicación	3	0.93	0.18	0.5269
6	124,920,000	125,120,000	6q22.31	6q22.31	200	3	Duplicación	3	1	0.15	0.5333
7	151,985,000	152,085,000	7q36.1	7q36.1	100	2	Duplicación	3	1	0.19	0.7292
10	2,970,000	3,570,000	10p15.3	10p15.2	600	7	Duplicación	3	0.95	0.1	0.5812
12	120,965,000	121,065,000	12q24.31	12q24.31	100	2	Duplicación	3	0.71	0.17	0.4088
16	83,445,000	83,645,000	16q23.3	16q23.3	200	3	Duplicación	3	0.72	0.14	0.4264
17	43,435,000	44,160,000	17q21.31	17q21.31	725	6	Delección	1	0.93	0.06	-0.8921

CN= Número de copias; pb= pares de bases; kpb= kilopares de bases