



VNIVERSITAT
DE VALÈNCIA

Facultad de Medicina y Odontología

Departamento de Medicina

Programa de doctorado 3139: Medicina

**Tecnologías de secuenciación masiva y PCR digital en la detección de ADN
tumoral procedente de tejido y plasma en pacientes de colon localizado**

Roberto Tébar Martínez

Codirector y tutor: Dr. Andrés Cervantes Ruipérez

Codirectora: Dra. Pilar Rentero Garrido

Codirectora: Dra. Noelia Tarazona Llaveró

Octubre 2022

INFORME DIRECTORES/AS Y TUTOR/A PARA DEPÓSITO DE TESIS

Director (es) / Codirector (es):

1.- Apellidos y nombre: Cervantes Ruipérez, Andrés. N.I.F. 22923458W, Departamento/Instituto: Oncología, Centro: INCLIVA

2.- Apellidos y nombre: Rentero Garrido, Pilar. N.I.F. 44876979T, Departamento/Instituto: Unidad de Medicina de Precisión, Centro: INCLIVA

3.- Apellidos y nombre: Tarazona Llavero, Noelia. N.I.F. 53359783J, Departamento/Instituto: Oncología. Centro: INCLIVA

Tutor o tutora (si procede)

Apellidos y nombre: Cervantes Ruipérez, Andrés. N.I.F. 22923458W, Departamento/Instituto: Oncología, Centro: INCLIVA

Directores/as y tutor/a, respectivamente, de la tesis doctoral: "Tecnologías de secuenciación masiva y PCR digital en la detección de ADN tumoral procedente de tejido y plasma en pacientes de colon localizado"

de D. Roberto Tébar Martínez,

estudiante del Programa de Doctorado **3139 Medicina** (RD99/2011) en Medicina de la Universitat de València, emiten informe FAVORABLE (*favorable/desfavorable*) para la realización del depósito y la defensa de la tesis doctoral.

Fecha: 04 de julio de 2022

Fdo.: Andrés Cervantes

ANDRES| Firmado digitalmente por ANDRES|
CERVAN| CERVANTES|
TES| RUIPEREZ
RUIPERE| Fecha:
Z 2022.07.04
17:34:55 +02'00'
Director y tutor

Fdo.: Pilar Rentero

PILAR|RENTERO|
GARRIDO|
Firmado digitalmente por
PILAR|RENTERO|GARRIDO
Fecha: 2022.07.04
15:56:12 +02'00'
Directora

Fdo.: Noelia Tarazona

NOELIA| Firmado digitalmente por
TARAZON| NOELIA|
A| TARAZONA|
LLAVERO| LLAVERO
Director| Fecha: 2022.07.05
10:46:35 +02'00'

ESCUELA DOCTORAL
UNIVERSITAT DE VALÈNCIA

*"I'm not going to conquer anything. The one
who is the most free is the pirate king!"*

- *Monkey D. Luffy (One Piece 507)*

Eiichiro Oda

Resumen

El cáncer de colon se sitúa como uno de los tumores más relevantes a nivel social, con una tasa de entre el 30 y 50% de recaídas tras la resección. Como herramienta de estudio en este tumor, la biopsia líquida se ha abierto paso en diferentes ámbitos, tales como la detección de la heterogeneidad tumoral o la enfermedad mínima residual (EMR). El fundamento de esta estrategia es detectar el ADN tumoral circulante (ctDNA), que se encuentra diluido en el ADN libre circulante (cfDNA), siendo necesario el uso de tecnologías aptas para la detección a frecuencias bajo el 1%.

La hipótesis de esta tesis se apoya en variantes somáticas para detectar el ctDNA y definir la heterogeneidad tumoral, la EMR y las variantes desarrolladas en la recaída, a la vez que se desarrollan los aspectos técnicos para esta detección. Para ello, se obtuvo tejido, sangre y plasma de 150 pacientes de cáncer de colon localizado seguidos a lo largo de 5 años. A partir de la secuenciación del tejido tumoral y de la eliminación de las variantes germinales, se establecieron variantes somáticas para la realización de ensayos personalizados. Para cada paciente se desarrollaron ensayos específicos de PCR digital para detección del tumor y la EMR en las diferentes muestras de plasma: diagnóstico, postoperatorio (6-8 semanas) y seguimientos. También se realizó una descripción de la heterogeneidad tumoral a través de la secuenciación del plasma al diagnóstico y de la enfermedad metastásica mediante la secuenciación de exoma del plasma de los pacientes que recayeron. Por otra parte, con un enfoque más tecnológico, se desarrolló una estrategia basada en reducción de ruido de fondo y otra en el uso de marcaje molecular, ambas por amplicones, para bajar el límite de detección al 0,1%. Todo esto pretende desarrollar la utilidad de la biopsia líquida en cáncer de colon en el contexto clínico y establecer metodologías que permitan un correcto estudio del ADN tumoral circulante.

Palabras clave: cáncer de colon, PCR digital, *Next-Generation Sequencing*, biopsia líquida, ADN tumoral circulante.

Summary

Colon cancer is known as one of the most relevant cancers worldwide. Relapse happens for almost 30-50% of the patients after surgical resection. Helping in its study, liquid biopsy has arisen for different purposes, such as early detection, identification of tumoral heterogeneity or minimal residual disease (MRD). This tool has been focused on detection of fragments of DNA from the tumor among the complete cell-free DNA (cfDNA) that could be isolated from plasma. This tumor-derived fraction is known as circulating tumor DNA (ctDNA), and it usually comprises less than 1% of the total cfDNA.

Our hypothesis is based on tumor somatic variants as main tool for detection of ctDNA in order to determine possible tumor heterogeneity, MRD, and even new mutations which may have arisen during relapse. Moreover, this thesis will also explore the technical aspects of ctDNA detection. For validation of the hypothesis, tumoral tissue, blood, and plasma from 150 patients who were initially diagnosed with localized colon cancer were employed. These patients were also followed for 5 years. From tumor sequencing and germline variant filtering, somatic variants for each patient were selected. Personalized assays for evaluation of MRD employing digital PCR were designed to evaluate plasma samples at diagnosis, postoperative timepoint (6-8 weeks) and follow-ups. Tumor heterogeneity was also explored with targeted sequencing in baseline cfDNA and whole exome sequencing in relapsing plasmas. Simultaneously, from a technological perspective, a strategy of background noise reduction through wild type characterization and other based on molecular tagging were developed. Both techniques were expected to lower the limit of detection to 0.1%. Taken together, all these approaches try to define how useful liquid biopsy could be in the clinical setting of colon cancer and describe suitable methodologies for study of ctDNA.

Key words: Colon cancer, digital PCR, Next-Generation Sequencing, Liquid biopsy, tumor circulating DNA.

Agradecimientos

La realización de una tesis no es un trabajo sencillo, ni que se haga en solitario. Durante estos años, muchas personas han intervenido y me han apoyado durante todo el camino. En primer lugar, quiero agradecer al Dr. Andrés Cervantes la oportunidad de realizar una tesis dentro de su grupo y de la asignación del proyecto, además de su labor de director. Agradecer también a la Dra. Noelia Tarazona toda su orientación y guía en el trabajo con la biopsia líquida y la parte clínica del proyecto. Y a la Dra. Pilar Rentero por su labor en los aspectos técnicos de la tesis y el apoyo diario con las incertidumbres de la bancada y de la escritura. No me quiero olvidar de la Junta Provincial de Valencia de la Asociación Española contra el Cáncer por la financiación recibida y por ser, en última instancia, lo que me ha permitido seguir en Valencia y realizar esta tesis.

Dentro del laboratorio, hay mucha gente a la que debo darles las gracias. Kike, Tani, Pilar (doblete jeje) y Laura OO como miembros de la UMP, que me habéis enseñado todo lo que sé y habéis hecho del laboratorio un lugar de trabajo en el que realmente poder sentirse cómodo y a gusto. No me olvido tampoco de Laura GC, que, aunque hemos coincidido por un periodo corto, sé que vales un montón y tienes mucho que aportar. También he aprendido mucho de vosotros a título personal, y sé que, con subidas y bajadas, nunca olvidaré estos años. También dar las gracias a la gente que ha pasado de prácticas (Laura, Anna y Vero) por los ratos que hemos pasados. Gracias a la gente del laboratorio 23 por los almuerzos, los croissants y todos los ratos que me habéis dejado pasar a molestar y parasitar. A Koke, por comerte todos los análisis de esta tesis y chapazos derivados. Vales un montón y tienes un futuro increíble por delante. A Paco por los congresos compartidos y tardes apoyándome. A María del Labo 4 por hacer que las tardes no fueran tan solitarias. A las chicas del Labo 2 (Redón) por los almuerzos compartidos, a las chicas del Labo 2 de mama por estos años juntos y a la gente del laboratorio 1 de Joan, el auténtico *Rabolatorio*. En especial agradecer a Rai,

quien ha sido como un hermano en el laboratorio para mí. Sé que el Western no se aprende mirando el móvil, pero de ti he aprendido mucho más que ciencia, eres un auténtico reflejo del valor del esfuerzo, la constancia y el no rendirse ante las auténticas adversidades. Gracias también a Quico, sin “K”, por ser mi compañero de memes y de escalada, y a Carlos, Raúl y Paula, de vosotros tampoco me olvido.

Aunque parezca que no, la vida fuera del laboratorio también ha existido. Darles las gracias a mis amigos de Matemáticas, por aguantar discursos y por hacer los findes en Albacete inolvidables; a mis amigos del Máster Dani, Esther, Edu y Tania por hacer del Máster unos años increíbles y por las horas de rutas y de *Biocabañas* que nos han unido; a la puerta 5 de Serpis 26 y 17 de Bernat Fenollar 13 (Peri, Segu, Labrathor, Mimi, Isaías y Juanma), qué habéis sido una familia de verdad aquí en la terreta; a mis recientes compañeros de piso Juanpa, Cris, Lucía y Belen(cebu – Gracias también por esta maravillosa portada) por demostrar lo buenas personas y geniales que son, y en especial a Juanpa, por esta amistad de tantos años y engañarme para introducirme en la escalda; a Anna, por ser una gran amiga y hacerme explicarle las cosas a base de magdalenas y chocolate; a Helen, Lucía y Elena como amigas incondicionales después de todos estos años; a mis compañeros del Kraal del Grupo Scout, y en especial a todos los pioneros y pioneras de la Unidad Kimberley, que os llevo en el corazón; a mis primos Iris y Alejandro por el apoyo incondicional y toda una vida juntos; a Jaime, por una amistad llena de discusiones de verano; y por último, pero no menos importante, a mi padre, a mi hermano y a mi perrito Nemo (descansa tranquilo) por la confianza depositada en mi todos estos años y por haberme apoyado y ayudado a llegar hasta el final.

Sin duda, la mención especial es para ti, Isabel (el auténtico nombre de Paterna sabes que lo reservo, que esto es público), que me has apoyado durante estos años de forma incondicional y continuada. Sé que sigues sin saber lo que es el

cfDNA o la biopsia líquida, pero sí sabes lo que es tener un experimento que no se va de la cabeza o lo que es estar en tensión por unos resultados, ya que lo has tenido que aguantar. Como todos, ha habido ratos mejores y peores, pero siempre has estado ahí desde que empecé en el proyecto, siendo un apoyo de los que no quieres soltar. Sé que la cosa no ha sido la esperada por ambos, pero eres una persona valiosa y sin duda me lo has demostrado año tras año, siguiendo ahí independientemente de las circunstancias, y sé que seguirás. Te quiero un montón.

Por último, la única persona que nunca leerá esto y es motivo del cual estoy realizando esta tesis. No sé qué pensarías de mi después de estos años, de mis errores y aciertos, y, si aquello no hubiera sucedido, sería el tipo de persona que soy hoy. He hecho todo lo que he podido y en fecha de entrega de esta tesis, esto es todo lo que he podido hacer. Nunca olvidaré los momentos que también te pertenecían y no vivirás, como esta tesis (o bueno, te has librado de 4 años de auténtico estrés), pero lo más seguro es que nunca te olvidare a ti, y eso es algo que puedo decir después de lo que ahora es media vida y seguiré sin olvidar después de una vida entera. Este es mi granito de arena frente a la enfermedad que te hizo aquello. Te quiero, mamá.

ÍNDICE

Resumen	7
Summary	8
Agradecimientos.....	9
Siglas	17
Índice de Figuras.....	19
Índice de Tablas.....	22
1. INTRODUCCIÓN.....	25
1.1. Qué es el cáncer	27
1.2. Cáncer colorrectal	28
1.2.1. Origen y causas moleculares del CCR.....	28
1.2.2. Genes de relevancia en CCR.....	32
1.2.3. Diagnóstico del CCR.....	35
1.2.4. Histopatología y localización del CCR.....	36
1.2.5. Clasificaciones moleculares.....	37
1.2.6. Recaída y metástasis en CCR	38
1.2.7. Tratamiento del CCR.....	39
1.2.8. Diagnóstico molecular y biomarcadores en CCR.....	43
1.2.9. Medicina de precisión en CCR	46
1.3. Biopsia líquida	47
1.4. Biopsia líquida en cáncer.....	49
1.4.1. Componentes de la biopsia líquida de plasma.....	49
1.4.2. Tecnologías de evaluación del cfDNA.....	52
1.4.3. Técnicas basadas en PCR para detección del cfDNA.....	53
1.4.4. Secuenciación masiva en el estudio del cfDNA	55
1.4.5. Análisis bioinformáticos y estrategias sin marcaje.....	60
1.4.6. Hematopoyesis clonal como fuente de falsos positivos	61
1.4.7. Translación de la biopsia líquida	62

1.4.8.	Biopsia líquida en CCR.....	68
1.4.9.	Ensayos clínicos de biopsia líquida en CCR	70
2.	HIPÓTESIS Y OBJETIVOS.....	73
3.	METODOLOGÍA	77
3.1.	Reclutamiento de pacientes y obtención de muestras.....	79
3.2.	Procesamiento de las muestras	79
3.2.1.	Extracción de ADN germinal	79
3.2.2.	Extracción de ADN tumoral.....	80
3.2.3.	Extracción de ADN libre circulante en plasma	81
3.3.	Muestras control.....	82
3.4.	Cuantificación y valoración de ácidos nucleicos	83
3.4.1.	Evaluación de integridad y calidad de las muestras	83
3.4.2.	Cuantificación mediante fluorescencia.....	83
3.5.	Diseño de cebadores.....	84
3.6.	Generación y evaluación de librerías.....	85
3.6.1.	<i>QIASEq™ Targeted DNA Panel</i>	87
3.6.2.	<i>HyperExome con HyperCap Workflow v3.0</i>	88
3.6.3.	Librerías de amplicones	90
3.6.4.	Evaluación de librerías	92
3.7.	Secuenciación.....	92
3.7.1.	Procesado bioinformático	93
3.7.2.	Análisis funcional de las variantes de exoma.....	96
3.7.3.	Clasificación de pacientes mediante amplicones	96
3.7.4.	Revisión manual de las variantes	97
3.8.	PCR digital	97
3.8.1.	Elección de las mutaciones para seguimientos.....	97
3.8.2.	Diseño de las sondas ddPCR.....	98
3.8.3.	Protocolo de ddPCR	99

3.8.4.	Evaluación de las muestras de plasma de los pacientes	101
3.9.	Marcaje Molecular Mediante PCR.....	101
3.9.1.	Diseño de los adaptadores	101
3.9.2.	Modificaciones del protocolo de PCR.....	102
3.9.3.	Análisis de las variantes.....	103
3.10.	Estadística.....	104
4.	RESULTADOS	105
4.1.	Pacientes de la cohorte	107
4.2.	Identificación de variantes somáticas en 1ª cohorte	108
4.2.1.	Variantes germinales.....	109
4.2.2.	Variantes somáticas.....	110
4.2.3.	Selección de variantes	113
4.3.	Secuenciación de plasma del diagnóstico	113
4.3.1.	Variantes somáticas y de heterogeneidad tumoral	114
4.3.2.	Asociación del cfDNA y ctDNA al diagnóstico con variables clínicas	117
4.3.3.	Variantes de hematopoyesis clonal.....	118
4.4.	Detección del tumor en plasma mediante ddPCR.....	119
4.4.1.	Diseño y validación de las sondas en tejido	119
4.4.2.	Detección del tumor en plasma al diagnóstico por NGS vs ddPCR ..	120
4.4.3.	Detección del tumor en plasma tras la operación	122
4.4.4.	Enfermedad mínima residual y recaída mediante ddPCR.....	124
4.5.	Secuenciación de los exomas de las recaídas	127
4.5.1.	Variantes germinales y de hematopoyesis clonal	127
4.5.2.	Mutaciones somáticas y concordancia.....	129
4.5.3.	Mutaciones adquiridas accionables y de resistencia	132
4.6.	Detección de EMR en plasma mediante amplicones	133

4.6.1.	Diseño y validación de los paneles.....	133
4.6.2.	Secuenciación en muestras postoperatorias.....	136
4.6.3.	ctDNA y EMR mediante amplicones	137
4.7.	Marcado molecular mediante amplicones	141
4.7.1.	Selección del diseño.....	141
4.7.2.	Optimización del protocolo.....	144
4.7.3.	Sensibilidad del método óptimo	145
5.	DISCUSIÓN	151
6.	CONCLUSIONES.....	171
7.	BIBLIOGRAFÍA	175
8.	ANEXOS.....	199
8.1.	ANEXO I: Listado de cebadores específicos	201
8.2.	ANEXO II: Listado de adaptadores	204
8.3.	ANEXO III: Listado de sondas de ddPCR.....	205
8.4.	ANEXO IV: Listado de pruebas realizadas a los pacientes.....	208
8.5.	ANEXO V: Seguimientos de los pacientes con recaída.....	212
8.6.	ANEXO VI: Concordancia ddPCR y WES	217
8.7.	ANEXO VII: Concordancia en WES en pacientes con tejido de la metástasis	218

Siglas

5-FU: 5-Fluorouracilo

ACS: *American Cancer Society* (Asociación Americana del cáncer)

ADN: Ácido Desoxirribonucleico

ADNg: ADN germinal

ARMS-qPCR: *Amplification Refractory Mutation System-qPCR*

ARN: Ácido Ribonucleico

ARNInc: ARN largo no codificante

ARNm: ARN mensajero

ASC: American cancer Society

ATOM-Seq: Adaptor Template Oligo Mediated Sequencing

BAM: *Binary Alignment Map*

BEAMing: *Beads, Emulsion, Amplification, Magnetics*

BCT: *Blood Collection Tube* (Tubo para la recolección de sangre)

CAP: Capecitabina

CCR: Cáncer Colorrectal

CCRM: Cáncer Colorrectal Metastásico

CDS: *Coding DNA Sequence* (Secuencia de ADN codificante)

CEA: Antígeno Carcinoembriogénico

cfDNA: *Circulating free DNA* (ADN libre circulante)

CGI: *Cancer Genome Interpreter*

CHIP: *Clonal hematopiasis of indeterminate potencial* (Hematopoyesis clonal de significado incierto)

CIMP: *CpG Island Methylation phenotype* (fenotipo metilador de islas CpG)

CIN: *Chromosomal Instability* (Inestabilidad cromosómica)

CMS: *Consensus Molecular Subgroups* (subgrupos de consenso molecular)

CNV: *Copy number variation* (alteración en el número de copias)

COLD-PCR: PCR con co-amplificación a menor temperatura de desnaturalización

CT: *Computerized Tomography* (Tomografía computarizada)

CTCs: *Circulating Tumor Cells* (células tumorales circulantes)

ctDNA: *Circulating tumor DNA* (ADN tumoral circulante)

ddPCR: *droplet digital PCR*

DNMT: *DNA Methyltransferase* (ADN metiltransferasa)

DPD: dihidropirimidina dehidrogenasa

EDTA: ácido etilendiaminatetraacético

EMA: *European Medicines Agency* (Agencia Europea del medicamento)

EMR: Enfermedad Mínima Residual

EMT: *Ephithelial-Mesenchymal Transition* (transición epitelio-mesenchimal)

ERAT: *End-Repair and A-Tailing* (reparación de extremos y adición de cola de A)

ESMO: *European Society of Medical Oncology* (Sociedad Europea de Oncología Médica)

FDA: *Food and Drug Administration* (Agencia de los alimentos y fármacos)

FFPE: *Formalin-Fixed, Paraffin Embedded* (Fijado en formalina y embebido en parafina)

FOL: ácido folínico

GSEA: *Gene Set Enrichment Analysis* (Análisis de enriquecimiento en set de genes)

HCUV: Hospital Clínico Universitario de Valencia

HGVS: *Human Genome Variation Society* (Sociedad de las variaciones genómicas humanas)

HPD: *Hairpin Desing* (Diseño de horquilla)

HR: *Hazard ratio* (cociente de riesgo)

IGV: *Integrative Genomics Viewer*

IR: Iriotecán

LM-PCR: *Ligation-Mediated PCR*
LOH: *Loss of Heterozygosity* (Pérdida de heterocigosidad).
mAb: *monoclonal Antibody* (anticuerpo monoclonal)
MHC-I: *Major Histocompatibility Complex I* (complejo mayor de histocompatibilidad I)
miARN: micro ARN
MMR: *Mismatch Repair* (Reparación de malos apareamientos [de bases])
MSI: *Microsatellite instability* (inestabilidad de microsatélites)
MSS: *Microsatellite stability* (estabilidad de microsatélites)
NGS: *Next Generation Sequencing* (Secuenciación de nueva generación)
NOIR-SS: *Non-overlapping integrated reads Sequencing system* (Sistema de secuenciación con lecturas integradas no solapantes)
NPV: *Negative Predictive Value* (valor predictivo negativo)
OMS: Organización Mundial de la Salud
OXA: Oxaliplatino
pb: pares de bases
PBS: *phosphate buffered saline* (tampon fosfato salino)
PCR: *Polymerase Chain Reaction* (Reacción en cadena de la polimerasa)
PDX: *Patient-derived xenograft* (xenoinjerto derivado de paciente)

PPV: *Positive Predictive Value* (valor predictivo positivo)
qPCR: PCR cuantitativa
RFS: *Relapse Free Survival* (Supervivencia libre de recaída)
RM: Resonancia magnética
RPU: *Read per UMI* (lecturas por UMI)
SNV: *Single Nucleotide Variation*
TAm-Seq: *Tagged-Amplicon Sequencing*
TCGA: *The Cancer Genome Atlas*
TE: Tris-EDTA
TGF: *Tumor Growth Factor* (Factor de crecimiento tumoral)
Tm: Temperatura de disociación (*Temperature of melting*)
TMB: *Tumor Mutation Burden* (carga mutacional del tumor)
TNM: Tumor-Nódulos-Metástasis
TSG: *Tumor Suppressor Gene* (Gen supresor de tumores)
UDI: *Unique Dual Index* (Doble identificador único)
UID: *Unique Identifier* (Identificador Único)
UMI: *Unique Molecular Index* (Identificador de molécula Única)
VAF: *Variant allele frequency* (Frecuencia del alelo alterado)
WES: *Whole-Exome Sequencing* (Secuenciación de exoma completo)
WGS: *Whole-Genome Sequencing* (Secuenciación de exoma completo)

Índice de Figuras

Figura 1: Principales rutas de señalización implicadas en el cáncer colorrectal. .	33
Figura 2: Actuación de los diferentes fármacos para tratamiento del CCR.	40
Figura 3: Fuentes de la biopsia líquida y utilidad en cáncer.....	48
Figura 4: Metodologías de reducción de errores en secuenciación masiva.	56
Figura 5: Principales usos de la biopsia líquida en la clínica del CCR.	63
Figura 6: Esquema representando las muestras disponibles según el momento temporal y de la enfermedad.....	80
Figura 7: Esquema de los protocolos de generación de librerías empleados.	86
Figura 8: Esquema del procesado informático realizado con los datos de secuenciación.	94
Figura 9: Esquema del proceso de filtrado de variantes de las células blancas...	95
Figura 10: Modificaciones de las sondas para ddPCR de acuerdo con los BAMs.	99
Figura 11: Nube de puntos obtenida de la lectura de un tejido mediante ddPCR para la mutación p.G12D de KRAS.	101
Figura 12: Diseño de las modificaciones de los cebadores para introducir UMIs.	102
Figura 13: Distribución de los parámetros de secuenciación de los tejidos tumorales y células blancas de los 149 pacientes.....	108
Figura 14: Variaciones germinales de los pacientes de cáncer de colon localizado.	109
Figura 15: Mutaciones somáticas en regiones relacionadas con la traducción en los 29 genes del panel de CCR.	111
Figura 16: Prevalencia entre los pacientes de genes del panel de CCR con mutaciones entre las 4 variaciones más frecuentes en los pacientes.	113
Figura 17: Estadísticas de la secuenciación de los plasmas al diagnóstico de 120 pacientes de cáncer de colon localizado.....	114
Figura 18: Sensibilidad y especificidad para el panel de CCR con la tecnología de amplicones anclados.	115
Figura 19: Distribución de las mutaciones concordantes en plasma al diagnóstico en 120 pacientes de CCR localizado.	116
Figura 20: Relación del cfDNA y ctDNA al diagnóstico con las características clínicas de los pacientes de cáncer de colon.	118
Figura 21: Variantes de hematopoyesis clonal detectadas con el panel de CCR.	119

Figura 22: Concordancia entre la PCR digital y la secuenciación del panel en la validación con tejidos.....	120
Figura 23: Concordancia entre la PCR digital y la secuenciación del panel en los plasmas al diagnóstico.	120
Figura 24: Detección de ctDNA en muestras pre- y post- operatorias de CCR mediante variantes somáticas evaluadas con ddPCR.	121
Figura 25: Evaluación de los seguimientos de pacientes de CCR mediante ddPCR.	124
Figura 26: Detección de ctDNA en muestras pre- y post- operatorias de CCR mediante las variantes de la selección global evaluadas con ddPCR.	126
Figura 27: Distribución de los parámetros de secuenciación de las muestras empleadas en WES de las recaídas.	127
Figura 28: Principales genes con mutaciones oncogénicas detectadas en el WES de pacientes con recaída.	132
Figura 29: Concordancia de la secuenciación de los amplicones con el panel de personalizado y la ddPCR en muestras de tejido.	135
Figura 30: Concordancia de la secuenciación de los amplicones y el panel de personalizado y la ddPCR en muestras de plasma al diagnóstico.	136
Figura 31: Concordancia entre la secuenciación de los amplicones y la ddPCR en muestras de plasma postoperatorias.	137
Figura 32: Sensibilidad, especificidad y HR de los diferentes análisis de la secuenciación del plasma tras la operación mediante amplicones para detectar la recaída.....	138
Figura 33: Capacidad de detectar la recaída en muestras posteriores a la operación de los diferentes análisis de la secuenciación de amplicones con al menos una variante positiva.....	139
Figura 34: Capacidad de detectar la recaída en muestras posteriores a la operación de los diferentes análisis de la secuenciación de amplicones con al menos dos variantes positivas.	141
Figura 35: Validación del cebador con UMIs con secuencia protectora.....	142
Figura 36: Comparación del rendimiento de los diferentes diseños de cebadores con horquilla.	143
Figura 37: Valoración de la capacidad de los cebadores con horquilla para identificar moléculas.....	144
Figura 38: Optimización de la purificación y de la concentración de cebadores de PCR1.	145

Figura 39: Sensibilidad de los amplicones con horquilla HPD6 en tejidos FFPE.	146
Figura 40: Sensibilidad de los amplicones con horquilla HPD6 en los estándares HD780.....	148
Figura 41: Positividad en los estándares de cfDNA en función de las estrategias de análisis.	149
Figura 42: Seguimientos de los pacientes metastásicos mediante sondas contra mutaciones somáticas y patológicas.	212
Figura 43: Diagramas de Venn para la concordancia entre las variantes somáticas del tejido primario, plasma de la recaída y los diferentes tejidos de la recaída.	218

Índice de Tablas

Tabla 1: Principales características de los estadios del cáncer de colorrectal.....	36
Tabla 2: Principales diferencias entre tumores de colon izquierdo y derecho....	37
Tabla 3: Biomarcadores del cáncer colorrectal.....	44
Tabla 4: Ventajas y limitaciones de la biopsia líquida.....	49
Tabla 5: Características de los controles HD780 y HD816.	82
Tabla 6: Creación de la curva patrón para la cuantificación de ADN de doble cadena mediante el kit <i>Quantifluor dsDNA</i>	84
Tabla 7: Genes incluidos para el panel personalizado de <i>QIAseq™ Targeted DNA Panel</i>	87
Tabla 8: Especificaciones del protocolo del <i>QIAseq™ Targeted DNA Panel</i> en función del material de partida.	88
Tabla 9: Especificaciones del protocolo del <i>HyperExome con HyperCap Workflow v3.0</i> en función del material de partida.....	89
Tabla 10: Características de los equipos de secuenciación empleados.....	93
Tabla 11: Composición del mix de ADN de tejido para evaluación de los amplicones para marcaje molecular.....	103
Tabla 12: Características de los pacientes empleados durante la investigación.	107
Tabla 13: Alteraciones encontradas exclusivamente en plasma en la secuenciación del cfDNA del diagnóstico.	117
Tabla 14: Concordancia a nivel de paciente en la detección de ctDNA en muestras postoperatorias mediante las selecciones somática y patológica.....	122
Tabla 15: Valores diagnósticos de las selecciones de variantes somática y patológica en muestras postoperatorias en función del ctDNA.....	122
Tabla 16: Valores diagnósticos de la selección global en muestras postoperatorias y seguimientos en función del ctDNA.....	126
Tabla 17: Cobertura y variantes obtenidas para cada tipo de material en WES de pacientes con recaída.	127
Tabla 18: Mutaciones de hematopoyesis clonal identificadas en células blancas mediante secuenciación de exoma.....	128
Tabla 19: Promedio de variantes somáticas obtenidas mediante WES.....	129
Tabla 20: Distribución de los diferentes roles de las alteraciones somáticas de las muestras sometidas a WES.	130
Tabla 21: Mutaciones accionables de acuerdo con OncoKB desarrolladas en la recaída e identificadas en el plasma.....	133

Tabla 22: Valores diagnósticos de los métodos de análisis en amplicones muestras postoperatorias con al menos una variante positiva.	138
Tabla 23: Valores diagnósticos de los métodos de análisis en amplicones muestras postoperatorias con al menos dos variantes positivas.	140
Tabla 24: Concordancia a nivel de paciente en la detección de ctDNA en muestras postoperatorias mediante el uso de ddPCR y NGS basada en amplicones.....	140
Tabla 25: Diseños de horquillas con adaptadores Nextera.....	143
Tabla 26: Listado de cebadores específicos empleados en la tesis.....	201
Tabla 27: Listado de adaptadores empleados en la tesis.....	204
Tabla 28: Listado de sondas de PCR digital empleados en la tesis.....	205
Tabla 29: Experimentos realizados en los diferentes pacientes de la cohorte. .	208
Tabla 30: Concordancia entre las variantes seguidas en cada paciente mediante ddPCR y su detección en WES.	217

1. INTRODUCCIÓN

1.1. Qué es el cáncer

De acuerdo con la Organización Mundial de la Salud (OMS), el cáncer es un conjunto de enfermedades que se caracteriza por el crecimiento descontrolado de células anormales de un tejido, derivando en un proceso de superación de los límites del tejido y/o dispersión a otros órganos, denominándose esta última metástasis y siendo la principal causa de muerte en pacientes oncológicos (<https://www.who.int/health-topics/cancer>). Según los últimos datos publicados, más de 19 millones de personas fueron diagnosticadas de cáncer en 2020, mientras que 10 millones murieron de este conjunto de enfermedades (1). En el contexto de España, los tumores han supuesto la tercera causa de muerte en el periodo de enero-mayo de 2020, por detrás de las enfermedades del sistema circulatorio y de las enfermedades infecciosas (2).

Las principales causas ambientales del cáncer son el tabaco y el alcohol, la obesidad, determinados agentes infecciosos, la contaminación ambiental o una dieta inadecuada junto al consumo de productos con sustancias carcinogénicas, entre otros (3,4). Respecto a las causas intrínsecas, existen en la población personas portadoras de mutaciones que puedan incrementar el riesgo de padecer tipos concretos de cáncer, aumentando la predisposición de la célula o disminuyendo los mecanismos naturales de eliminación de células aberrantes (3). En referencia a las características generales de este conjunto de enfermedades con tanta diversidad de histología y funcionamiento se encuentran la capacidad de evitar los supresores del crecimiento, la inducción de la angiogénesis, la señalización proliferativa mantenida, la resistencia a la muerte celular, la inmortalidad replicativa, la inmunovigilancia, reprogramación epigenética, la senescencia, la capacidad invasiva y metastásica, el metabolismo energético alterado, el estado proinflamatorio e inestabilidad genómica y las mutaciones (5).

1.2. Cáncer colorrectal

El cáncer colorrectal (CCR) es un grupo de tumores que abarca aquellos con origen en el colon, recto y ano. Según datos de 2020, el CCR, a nivel global, es el tercer cáncer con mayor incidencia (1,9 millones de casos) y el segundo con mayor mortalidad para ambos sexos (1,2 millones de muertes). En los últimos años, pese a existir una mejora en la detección y prevención de la enfermedad en pacientes mayores de 50 años gracias a las técnicas de cribado, en ciertas regiones se está demostrando un aumento de la incidencia del CCR en edades más tempranas, especialmente en territorios con una tasa de ingresos alta (6,7). Este cambio puede deberse a un aumento de prevalencia de la obesidad, cambios en la alimentación, descenso de la actividad física y aumento del consumo de alcohol y tabaco (6,7).

A nivel nacional, se espera que el CCR sea el tumor más diagnosticado en 2021 con 43581 nuevos casos, mientras que en 2020 se estimó como el segundo cáncer de mayor mortalidad con 16470 muertes, detrás, solamente, de los tumores de pulmón (2).

1.2.1. Origen y causas moleculares del CCR

De forma general, más del 80% de los CCR surgen de pólipos del colon que evolucionan progresivamente a adenomas, y posteriormente pasarán a ser lesiones cancerosas, en un intervalo de más de 10-15 años (8,9). Este tipo de transformación adenoma-carcinoma abarca la mayoría de los tumores de colon espontáneos y presenta una evolución centrada en la activación de la vía Wnt/ β -catenina. La otra principal vía de desarrollo del tumor son las lesiones serradas, que suponen entorno al 10-15% de los CCR espontáneos y se caracterizan por activaciones de B-raf. Dicha vía incluye un amplio espectro de morfologías, entre las que se encuentran pólipos hiperplásicos, lesiones serradas sésiles o lesiones serradas tradicionales (8,10). En cuanto a las situaciones de herencia y

predisposición, entorno a un 25% de los casos de CCR presenta al menos un familiar directo diagnosticado con esta enfermedad. Ciertas patologías con carácter familiar, como síndrome de Lynch, la poliposis adenomatosa familiar, la poliposis asociada a *MUTYH* o el síndrome de poliposis juvenil, suponen menos del 5% de los casos de CCR y pueden llegar a suponer aumentos del riesgo de sufrir la enfermedad al 100% (9).

Pese a distinguirse dos morfologías características de lesiones, existen vías moleculares comunes en el desarrollo de alteraciones: 1) inestabilidad cromosómica (CIN) y mutaciones, 2) defectos del sistema de reparación de mal apareamiento (*Mismatch repair*, MMR) y 3) alteraciones epigenéticas.

Inestabilidad cromosómica (CIN)

El fenotipo CIN es el tipo de alteración más frecuente dentro del CCR, presente en cerca del 80-85% de los tumores colorrectales, y se caracteriza principalmente por la aparición de aberraciones cromosómicas en diversos niveles, como aneuploidías, traslocaciones, amplificaciones o pérdidas de heterocigosidad (*Loss of Heterozygosity*, LOH). Los principales mecanismos responsables de este fenotipo son desregulación en los telómeros, respuesta al daño del ADN y alteración del mecanismo de segregación de los cromosomas. Estos mecanismos acaban afectando a genes de elevada importancia en la célula, como son *APC*, *KRAS*, *PIK3CA*, *SMAD4* y *TP53* (10,11). En cuanto al impacto del CIN en la enfermedad, el CIN se asocia con una peor supervivencia, estableciéndose un peor impacto en un tipo CIN bajo (LOH bajo) frente a un CIN alto (LOH alto) (12).

Defectos del Sistema MMR

El Sistema MMR supone un conjunto de genes que participan en la reparación de pequeñas modificaciones de la secuencia del ADN, como inserciones, deleciones o apareamientos erróneos, durante los procesos de replicación y recombinación del ADN (13). Defectos en este sistema aparecen en cerca del 15% de los pacientes y facilitarán la aparición de mutaciones puntuales que afectarán al

funcionamiento o expresión del gen en el que se produzcan (14). Estas alteraciones se reflejan también en la modificación del número de repeticiones de los microsatélites, hablándose de inestabilidad de microsatélites (MSI) en caso de que se alteren y de estabilidad de microsatélites (MSS) en caso de que no lo estén (15). Genes como *MSH2*, *MSH3*, *MSH6*, *MLH1*, *PMS1* y *PMS2* están implicados en el Sistema MMR, causando alteraciones y el fenotipo MSI debido a mutaciones puntuales o silenciamiento epigenético (16,17).

En los últimos años, el auge de la inmunoterapia y la búsqueda de marcadores predictivos de respuesta ha permitido identificar posibles explicaciones a porqué los tumores MSI presentan mejor prognosis. Algunas de estas respuestas se han asociado a una mayor neoantigenicidad por una elevada carga mutacional del tumor (*Tumor Mutational Burden*, TMB) derivada de las alteraciones acumuladas, exponiéndolo al sistema inmune (18); o a que los tumores con MSI presentan menor frecuencia de alteraciones de peor prognosis como la pérdida de *TP53* o las mutaciones en *KRAS* (17).

Alteraciones epigenéticas (CIMP)

Uno de los principales mecanismos de regulación epigenéticos es la metilación, que afecta tanto a histonas como ADN, concretamente a las islas CpG. Estas últimas estructuras se componen de regiones de entre 200 y 500 bases que son ricas en dinucleótidos CG. Dentro del genoma, estas islas solapan con aproximadamente el 60-70% de los promotores de genes conocidos, asociándose la ausencia de metilación con la permisividad para transcribir dichos genes (19). Este mecanismo juega un rol importante en el desarrollo, causándose metilación *de novo* en genes cuyo rol necesita silenciarse posteriormente y manteniéndose durante las sucesivas replicaciones del genoma gracias a la acción de diferentes ADN metiltransferasas (DNMTs), como DNMT3A y DNMT3B en situaciones *de novo* y DNMT1 en el mantenimiento. Por tanto, errores en estos mecanismos van a conducir a una situación de alteración del patrón de metilación del genoma con

dos vertientes: hipometilación de oncogenes e hipermetilación de genes implicados en los mecanismos de supresión del cáncer (19,24). En el caso de tumores de CCR donde este último mecanismo está activo, se habla del fenotipo metilador de Islas CpG (*CpG Islands Methylator Phenotype*, CIMP), al cual pertenecen la mayoría de los tumores serrados (20,21). En función del grado de metilación se habla de CIMP bajo y CIMP alto. Mientras que el primero se asocia con más mutaciones en *KRAS*, el segundo se relaciona con presencia de mutaciones en *BRAF* y ausencias en *TP53* y con el fenotipo MSI (19,21,22), ya que la hipermetilación de *MLH1* deriva en los defectos del MMR descritos anteriormente (23).

Fenotipo Hipermutado

Aunque diversos autores describen los 3 fenotipos anteriores como los posibles mecanismos de desarrollo del CCR, existe un solapamiento entre los fenotipos MSI y CIMP, debido a la capacidad de la hipermetilación de silenciar genes del Sistema MMR. Es por ello por lo que dentro del CCR también se habla de fenotipo no hipermutado en aquellos pacientes MSS y de hipermutados en aquellos con MSI, ya sea por CIMP o por deficiencias en el Sistema MMR esporádicas o hereditarias en el 75% de los casos, o por los defectos en las polimerasas con capacidad de corrección de errores, como *POLE* Y *POLD1* (24). A nivel molecular, cada uno de estos dos fenotipos va a mostrar las mismas vías alteradas, pero con diferentes genes mutados. Los fenotipos no hipermutados se caracterizan por una mayor frecuencia de mutación en genes como *APC*, *TP53*, *KRAS*, *FBXW7*, *PIK3CA* y *SMAD4* mientras que los hipermutados presentan una mayor tasa de mutación genes como *SYNE1*, *RNF43*, *ACVR2A*, *KMT2D*, *APC*, *FAT1*, *ARID1A* y *BRAF* (25,26).

1.2.2. Genes de relevancia en CCR

En conjunción con los fenotipos descritos en el apartado anterior, en el cáncer de colon existen diferentes rutas moleculares de relevancia que aparecen frecuentemente alteradas en este tipo de tumores (Fig. 1).

Una de las principales vías, alteradas en la mayoría de los tumores de colon, es la ruta canónica de Wnt/ β -catenina, implicada en el control del crecimiento celular y mantenimiento de la homeostasis del tejido (27). Dentro de esta ruta destaca *APC*, un gen supresor de tumores (*Tumor suppressor gene*, TSG) que participa como componente del complejo destructor de la β -catenina (Fig. 1), impidiendo la activación de genes relacionados con la proliferación, supervivencia o apoptosis (28). Las mutaciones en este gen se han descrito como uno de los primeros componentes en ser mutado en los eventos tempranos del desarrollo del tumor (29). *APC* aparece mutado con mayor frecuencia en tumores no hipermutados, mientras que en los hipermutados aparecen mutados otros genes de la vía como *RNF43*, *ARID1A*, *FBXW7*, *SOX9* y *AMER1* (25).

TP53 codifica para p53, uno de los principales controladores del ciclo celular (Fig. 1). Alteraciones de su funcionamiento, mediante diversos mecanismos como mutaciones somáticas puntuales o pérdidas en la región 17q, van a conducir a la célula a crecimientos descontrolados, apareciendo mutado entre el 36-66% de los pacientes de CCR (16,30).

La ruta de las MAPK es otra de relevancia en la dentro del CCR. Dicha ruta comienza con la dimerización de receptores tirosina quinasa como EGFR y HER2, cuya autofosforilación permite la transducción de la señal a través de proteínas como Kras o B-raf (Fig. 1) (31). Alteraciones activantes en EGFR y HER2, como las amplificaciones y sobreexpresiones, se asocian a resistencias a tratamientos anti-EGFR y, en concreto con HER2, a fenómenos metastásicos (32,33). De la misma forma, mutaciones activantes en los codones 12 y 13 del gen *KRAS* hacen que este

oncogén mantenga la vía activa independientemente de la señalización extracelular (34). Apareciendo entre el 25-45% de los pacientes, es otra de las principales resistencias contra anti-EGFRs y se asocia con mal pronóstico (25,35). Por último, *BRAF* se caracteriza por la mutación p.V600E, que aparece en 8-12% de los tumores metastásicos (26,36), suponiendo otro punto de activación constante de la vía. Esta mutación es más frecuente en los fenotipos MSI y CIMP alto (21,37,38), asociándose a una peor supervivencia (36).

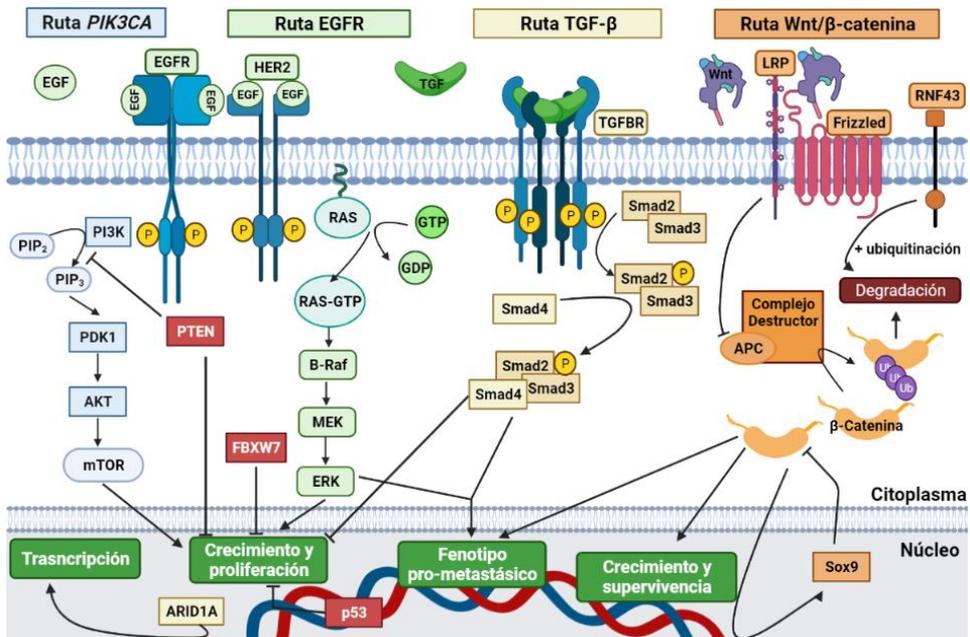


Figura 1: Principales rutas de señalización implicadas en el cáncer colorrectal. La figura refleja las principales cascadas de señalización que intervienen en CCR. Aunque se presentan de forma independiente, algunas de estas vías pueden interactuar, como ocurre con las vías de TGF-β y EGFR. También pueden presentar un origen común, como ocurre con PI3K y Ras, o los mismos receptores pueden iniciar la misma vía, como ocurre con HER2 y EGFR. Flecha: activación, línea: inhibición. (Elaborada en www.biorender.com)

Derivado de la activación de receptores tirosina quinasa y de la insulina, la proteína PI3K, codificada por *PIK3CA*, transduce la señal hacia el núcleo a través de proteínas como AKT y mTOR (Fig. 1). La activación de vías en las que participa

PI3K se ha asociado con funciones como la proliferación, la apoptosis y la supervivencia (39). En CCR, más de la mitad de las alteraciones suceden en los codones 542 y 545 del exón 9 y el codón 1047 del exón 20 (40). Uno de los principales inhibidores de esta vía es *PTEN*, el cual revierte la conversión del fosfatidilinositol bifosfato a trifosfato (41). *PTEN* también actúa junto con p53, colaborando en la detención del ciclo celular (41).

La ruta del TFG- β participa en procesos celulares como la proliferación, la diferenciación, la angiogénesis y el desarrollo temprano, mostrando un rol inhibitorio con funcionamiento normal, mientras que favorece la metástasis en colaboración con otras vías de señalización durante el desarrollo del CCR (42). Mutaciones y LOH en componentes de esta ruta, como *SMAD2*, *SMAD3* y *SMAD4* van a permitir la progresión del tumor de colon (43) (Fig. 1). Además de estos genes, *TGFRB2*, *BMPR2* y *ACVR2A* pueden aparecer alterados, aunque lo hacen con más frecuencia en fenotipos hipermutados (25).

CDX2 ha sido descrito como un TSG en cáncer de colon y está implicado en el desarrollo embrionario del intestino, restringiéndose su expresión del duodeno al recto en la edad adulta (44). Un estudio en líneas celulares asocia su actividad en colon con la ruta Wnt, favoreciendo la proliferación cuando se pierde la expresión de *CDX2* o ejerciendo su actividad como TSG cuando está sobreexpresado (45). *CDX2* ha sido identificado como un marcador de mal pronóstico en la enfermedad metastásica (46).

Como se ha mencionado anteriormente, *POLE* y *POLD1* son responsables del 25% de los tumores hipermutados. Mutaciones en el dominio exonucleasa, el cual participa en la corrección de lectura, son responsables de ese fenotipo (24,47).

Transición epitelio-mesenquimal

La transición epitelio-mesenquimal (*epithelial-mesenchymal transition*, EMT) es un proceso por el que las células cancerosas adquieren un fenotipo invasivo,

abandonando características más epiteliales. Este proceso implica la pérdida de las uniones intercelulares, pérdida de la polaridad celular, adquisición de motilidad y reestructuración del citoesqueleto. En este proceso intervienen: i) factores de transcripción de familias como SNAIL, Twist, Slug, ZEB y bHLH, ii) la activación de vías como TGF- β y Wnt, iii) ácidos ribonucleicos (ARN) con secuencias no codificantes, como micro-ARNs (miARN, como miR200) y ARN largos no codificantes (ARNInc, como HOTAIR y lncRNA H19) (48).

1.2.3. Diagnóstico del CCR

El primer paso en la detección del CCR son los programas de detección temprana en la población de riesgo. Frente al incremento del CCR en población más joven, la Sociedad Americana del Cáncer (ACS) recomienda el comienzo del cribado en pacientes de 45 años, aunque las últimas guías de la Sociedad Europea de Oncología Médica (ESMO) para cáncer de colon localizado no modifica dichas recomendaciones, estableciendo el cribado de 50 a 75 años. Las principales herramientas de cribado son la colonoscopia y la sangre oculta en heces (7,49,50).

La enfermedad causada por el CCR localizado llega a invadir el lumen y la pared de los intestinos y evolucionar hasta extenderse a otros órganos. En estadios más avanzados, o en tumores de mayor tamaño, aparecerán síntomas inespecíficos como dolor abdominal, anemia, debilidad o pérdida de peso, y otros más específicos como pueden ser hemorragias u obstrucción intestinal. Para confirmar el diagnóstico y descartar metástasis, se usan técnicas como la colonoscopia, el escáner mediante tomografía computarizada (*Computerized Tomography, CT*) o la imagen mediante resonancia magnética (RM) (49,51).

Una vez se ha completado el diagnóstico, la resección del tumor será mediante endoscopia o bien mediante cirugía con retirada de segmento del colon si el tumor presenta cierto grado de infiltración. De la muestra retirada se obtendrá

un informe de patología en el que se incluirán marcadores relevantes como el estado del MMR y el estadio mediante tamaño-nódulos-metástasis (TNM). También se determinará el valor del antígeno carcinoembriónico (CEA), asociándose un peor valor pronóstico si el valor postoperatorio es superior a 5 ng/ml (49).

A partir de la cirugía, se recomienda el seguimiento a largo plazo. En un primer periodo de 3 años, en el que se espera encontrar el 80% de las recaídas, se realizará una colonoscopia a partir del primer año de cirugía, una CT cada 6-12 meses y determinación del CEA cada 3-6 meses. En el periodo de 3 a 5 años, donde se espera encontrar el 15% de las recaídas, estas pruebas se realizarán, pero con el doble de tiempo entre unas y otras (49,52).

1.2.4. Histopatología y localización del CCR

La mayoría de los CCR son carcinomas a nivel histológico, y de estos, más del 90% son adenocarcinomas. A su vez, se clasifican según el porcentaje de formación tejido glandular, hablándose un alto y bajo grado de diferenciación (8). En función de la progresión del tumor, se habla de diferentes estadios (53), descritos en la tabla 1. Debido a la ausencia de metástasis, los estadios I-III son referidos como CCR localizado. Además de la histología, se han descrito diferencias entre la localización del tumor en las diferentes zonas del colon, agrupándose clásicamente como colon izquierdo o distal y colon derecho o proximal (54), cuyas principales diferencias se recogen en la tabla 2.

Tabla 1: Principales características de los estadios del cáncer de colorrectal.

Estadio	Invasión	Nodos	Metástasis
I	No	No	No
II	Sí	No	No
III	Sí	Sí	No
IV	Sí	No/Sí	Sí

Tabla 2: Principales diferencias entre tumores de colon izquierdo y derecho. Para cada tumor, se muestran las características más frecuentes.

	Izquierdo (Distal)	Derecho (Proximal)
Incidencia	Mayor	Menor
Sexo más frecuente	Hombre	Mujeres
Fenotipo	CIN	CIMP, MSI
Supervivencia	Mejor	Peor
Inmunogenicidad	Menor	Mayor

1.2.5. Clasificaciones moleculares

Con el fin de agrupar los pacientes y establecer líneas de manejo comunes, se han propuesto diferentes sistemas de clasificación de los pacientes de CCR. La clasificación realizada por el Atlas del Genoma del Cáncer (TCGA) estableció dos tipos principales de tumores: los no hipermutados y los hipermutados (24). Durante los años 2012 y 2013, surgieron varias clasificaciones basadas en expresión. El grupo de Guinney *et al.* decidió realizar un análisis con los trabajos de las diferentes clasificaciones a través de un grupo de evaluación (55). A partir de los perfiles de expresión de todos los grupos de datos, se llegaron a 4 subgrupos de consenso molecular (*Consensus Molecular Subgroups*, CMS). El subgrupo CMS1 o MSI inmune se identificó en el 14% de los pacientes, caracterizándose por MSI y CIMP alto, lo que se asocia con un fenotipo hipermutador, mutaciones en *BRAF* y un perfil más inmunogénico, con mayor tasa de infiltración linfocitaria y actividad inmune. El CMS2 o canónico, propio del 37% de los datos, se caracteriza por un fenotipo CIN y presentar una mayor actividad en la ruta Wnt y la correspondiente sobreexpresión de *MYC*. El CMS3 (13% de los pacientes), denominado Metabólico, se caracteriza por CIMP y CIN bajo, sin tener un estado único de MSI, mutaciones en *KRAS* y presentar un metabolismo alterado. Por último, el CMS4 o mesenquimal, presente en el 23% de los pacientes, se caracteriza por un alto CIN y presentar infiltración estromal, activación de la vía de TGF- β y una mayor angiogénesis. También se asoció un valor pronóstico a los diferentes grupos, observándose una peor supervivencia y

mayor tasa de recaída en el tipo CMS4, una peor supervivencia tras la recaída en el CMS1 y un mejor pronóstico general en el CMS2. Siendo la clasificación actualmente más aceptada, presenta limitaciones, dejando al 13% de los pacientes del análisis inicial sin clasificar (55).

1.2.6. Recaída y metástasis en CCR

Pese a las opciones terapéuticas existentes, un 30-50% de pacientes de CCR presentan recaída, sucediendo la mayoría de éstas en un periodo de 5 años (49,52,56). La confirmación de la metástasis, única o múltiple, se debe realizar mediante la correspondiente obtención de imagen radiológica y análisis histológico de las lesiones. En caso de que la metástasis no sea extirpable, existen diferentes aproximaciones que buscan llevar el tumor a un estado que permita la extirpación (57).

Los principales órganos dónde se desarrolla el cáncer colorrectal metastásico (CCRm) suelen ser hígado, pulmón, hueso, cerebro, y otros órganos cercanos, como el peritoneo y otros órganos abdominales (58–60). A nivel molecular, el CCRm se caracteriza por una frecuencia de casi el 90% de tumores con la ruta Wnt alterada, a través de mutaciones sin sentido de *APC* y mutaciones en genes como *CTNNB1* y *RNF43*. Los genes que se encontraron más frecuentemente mutados en CCRm MSS fueron *APC*, *TP53*, *KRAS*, *PIK3CA* y *SMAD4*. De interés clínico, las principales mutaciones encontradas en CCRm MSS fueron las mutaciones activantes de *KRAS* y *NRAS*, mientras que en CCRm MSI predominaban otras mutaciones accionables, como la p.V600E de *BRAF*, las fusiones de *NTRK* y las alteraciones localizadas en *BRCA1*, *BRCA2*, *PTEN* y *PIK3CA*. En términos de supervivencia, las metástasis de peor pronóstico son aquellas que presentan alteraciones de las rutas RAS o en concomitancia con la ruta de PI3K (61). Las mutaciones que aparecen en el tumor de la recaída no coinciden en su totalidad

con el tumor primario, pudiendo observarse que en ocasiones el clon original de la metástasis no se haya identificado en el tumor primario (62,63), aunque se ha descrito la presencia de mutaciones clonales tempranas que se mantienen en ambos (64). Por el contrario, existe la posibilidad de que una metástasis sea fundada por más de un clon del tumor primario, e incluso que existan recaídas derivadas de clones de origen metastásico (62).

1.2.7. Tratamiento del CCR

La principal opción de tratamiento para las posibles lesiones cancerosas del colon es la retirada del pólipo, adenoma o adenocarcinoma (49,53). De estos pacientes con enfermedad localizada, aquellos que presenten una situación de riesgo de recaída, definida por el TNM, el estado del MMR y por el número de nodos linfáticos afectados, podrán presentar o no necesidad de terapia adyuvante (49,53,65). En el caso de pacientes con metástasis, siempre que sea posible se buscará operar el tumor, pudiendo acompañarse de terapia que busque la reducción del tumor. En caso de que no sea extraíble, existe una serie de terapias que buscan encoger el tumor o crear una situación de control de la enfermedad (57). La mayoría de las quimioterapias no dirigidas son comunes entre la enfermedad localizada y la metastásica, mientras que los fármacos dirigidos, además de usarse en combinación con los quimioterápicos, se emplean solo en la enfermedad metastásica (57). Entre los agentes citotóxicos más usados en CCR están el 5-fluororacilo (5-FU/F), oxaliplatino (OXA), capecitabina (CAP), iriotecán (IRI), ácido folínico o leucovorin (FOL) (Fig. 2), en sus combinaciones FOLFOX, FOLFIRI, CAPOX (dobletes), y FOLFOXIRI (tripleto). Todos estos fármacos buscan dañar el ADN y alterar los mecanismos de funcionamiento celular, con el fin de causar el arresto del ciclo celular, alterar el metabolismo o causar toxicidad, para alcanzar la consiguiente muerte celular (57).

Terapias dirigidas

En el caso de fármacos dirigidos a dianas moleculares, actualmente hay 3 tipos de terapias aprobadas en el CCRm (53). El primer grupo son los fármacos anti-EGFR, cetuximab y panitumumab, dirigidos a bloquear la activación constitutiva de esta vía mediante bloqueo de la dimerización de los receptores (Fig. 2). Por este mecanismo, solo se usan si el resto de la vía, en concreto, *KRAS* y *BRAF*, no presenta mutaciones activantes. El uso aprobado de los fármacos anti-EGFR incluye su uso en combinación con FOLFOX o FOLFIRI como terapia de primera línea (53,66).

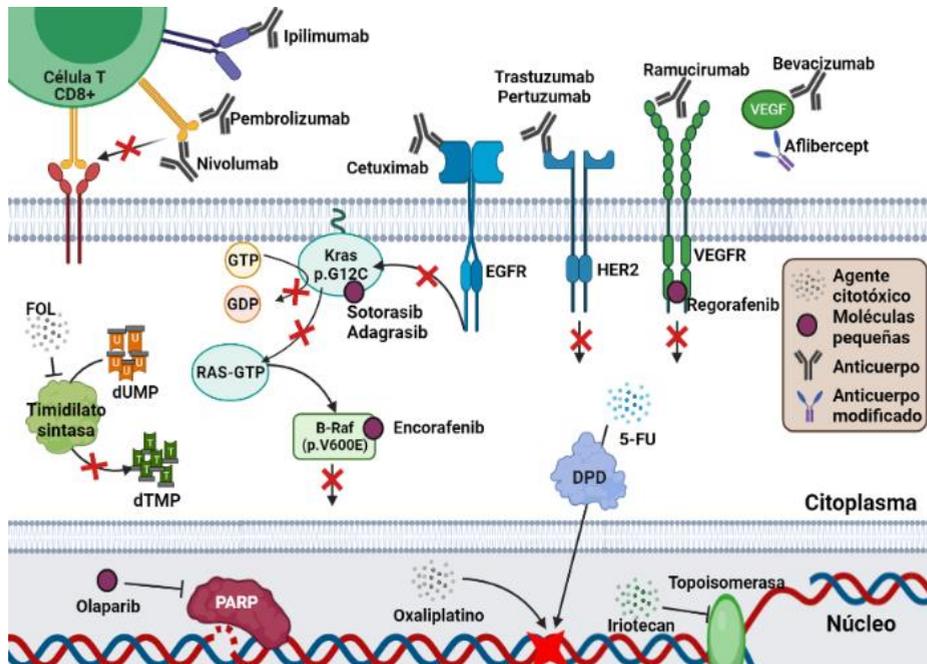


Figura 2: Actuación de los diferentes fármacos para tratamiento del CCR. La imagen contextualiza los diferentes fármacos en la célula, reflejando su actividad y las moléculas a las que afectan. Flecha: activación, línea: inhibición, cruz: interrupción de la actividad. (Elaborada en www.biorender.com)

El segundo grupo son los anti-angiogénicos, diferentes fármacos dirigidos a inhibir este mecanismo de desarrollo tumoral (Fig. 2). El principal fármaco, el

bevacizumab, es un anticuerpo monoclonal (mAb) está dirigido contra el VEGF-A, impidiendo la unión a su receptor VEGFR. Este fármaco está aprobado para administrar en primera línea junto a cualquiera de las quimioterapias combinadas (57). Otros fármacos aprobados tanto por la Agencia Europea del medicamento (*European Medicines Agency*, EMA) como la Agencia de los alimentos y fármacos americana (*Food and drug administration*, FDA) son aflibercept, ramucirumab, y regorafenib (67).

Por último, el tercer grupo abarca diferentes fármacos implicados en inmunoterapia (Fig. 2), los inhibidores de los puntos de control inmune, cuyo fundamento es interrumpir las señales coestimuladoras negativas mediante bloqueo de la unión, impidiendo la inmunosupresión desarrollada por el tumor (68,69). Mientras que la FDA ha aprobado el uso de pembrolizumab (anti-PD1) y nivolumab (anti-PD1) solo o con ipilimumab (anti-CTLA-4) en CCRm con defecto de MMR o MSI alto (68), la EMA solo ha aprobado el pembrolizumab en primera línea con estos tumores (69).

Terapias dirigidas en desarrollo

Aunque las terapias dirigidas aprobadas en CCR suponen el mejor tratamiento disponible para los pacientes, no todos los pacientes pueden optar a ellas. Es el caso de los anti-EGFR, ya que, estando dirigidas al componente inicial de la ruta, activaciones constitutivas a través de otros puntos como RAS o B-raf mutados o amplificaciones de *HER2* o *MET* hacen que el bloqueo de EGFR sea inefectivo (76). Es por ello, que las terapias en desarrollo intentan superar estos problemas, ya sea con nuevas moléculas o con terapias previamente aprobadas en otros tumores.

Las mutaciones en *RAS* (*HRAS*, *KRAS*, *NRAS*) son de las principales causas de mal pronóstico en CCR metastásico, al no presentar beneficio del tratamiento anti-EGFR (70). Dos fármacos están surgiendo como los primeros inhibidores tirosina

quinasa dirigidos directamente contra KRAS, las moléculas sotorasib (71) y adagrasib (72). Ambas moléculas están dirigidas específicamente contra la mutación p.G12C, la cual define un subgrupo de pacientes de peor pronóstico en comparación con pacientes con otras mutaciones de KRAS (73). Ninguna de las dos estructuras ha sido aprobada todavía para CCRm.

BRAF es otro oncogén implicado en la cascada de señalización de EGFR. La mutación más frecuente es la p.V600E. En estos pacientes, como primera línea de tratamiento, se emplean combinaciones con iriotecán y oxalipaltino, con opción de bevacizumab (74). Si bien el fármaco dirigido individualmente contra la mutación p.V600E, encorafenib, no está aprobado para su uso individual, el ensayo BEACON ha explorado su uso en terapia combinada, observándose el mejor incremento en la supervivencia en el triplete encorafenib, cetuximab y binimetinib (inhibidor de la ruta MAPK) frente al control en el análisis primario (75), mientras que en los resultados finales, se asociaba la misma supervivencia tanto al doblete (sin binimetinib) como al triplete, pero menor reactividad adversa de grado 3 al doblete, al carecer de los efectos adversos del inhibidor del MAPK (76). Esto hizo que se haya propuesto el doblete como nueva terapia de segunda o tercera línea en CCRm BRAF p.V600E (74,76).

A parte de EGFR, la sobreexpresión, y no simplemente la amplificación (87), de otros genes que codifican para receptores tirosina quinasa como *MET*, *HER2*, *HER3*... pueden activar la ruta MAPK a pesar del bloqueo de EGFR, actuando como posible mecanismo de resistencia (66). De estas moléculas, una de las que más relevancia presenta es *HER2*, cuya alteración en CCRm varía entre el 2 y 9% de los casos, definiendo un subgrupo de pacientes que no se benefician de la terapia con anti-EGFR. Los principales desarrollos terapéuticos proceden de fármacos aprobados en otros tumores, como los anticuerpos monoclonales (mAbs) trastuzumab y pertuzumab, o el inhibidor tirosina quinasa lapatinib (78). En el caso de *MET*, otro estudio ha propuesto tratar los tumores EGFR resistentes por

sobreexpresión de MET con cetuximab y tivantinib, un inhibidor tirosina quinasa contra MET, alcanzando solo respuesta en un 10% de los pacientes (79).

1.2.8. Diagnóstico molecular y biomarcadores en CCR

El cáncer se presenta como una enfermedad en el que la genética de la célula juega un rol importante (80). Estas mutaciones van a ser responsables de activación de mecanismos que van a generar el fenotipo maligno de las células, y parte de estos mecanismos se convertirán en imprescindibles para el desarrollo tumoral y supervivencia de la célula. La identificación de estas alteraciones y otros mecanismos implicados en la alteración de expresión de genes como la epigenética han sido claves para poder profundizar y entender parte del origen, el desarrollo, la evolución y los mecanismos que hay detrás del cáncer. Estas características que diferencian a la célula tumoral de las células sanas han sido clave para la propuesta y desarrollo de terapias dirigidas contra estos mecanismos (81). De todos estos, se considerarán biomarcadores aquellas “moléculas biológicas encontradas en sangre, cualquier otro fluido o tejidos que sea signo de un proceso normal o alterado o de una condición o enfermedad” (82).

A diferencia del cáncer de mama o de pulmón, el diagnóstico molecular para la decisión del tumor primario de CCR se usa como factor pronóstico (49) y para determinar la toxicidad al tratamiento (83). En el caso del tumor metastásico, el testado molecular es clave para poder definir las terapias recomendadas para el paciente (57). La tabla 3 recoge una selección de marcadores de relevancia dentro del CCR.

Tabla 3: Biomarcadores del cáncer colorrectal. La tabla incluye algunos de los biomarcadores de mayor utilidad en la enfermedad. Dx.: Diagnóstico, Pre.: Predictivo, Pro.: Pronóstico.

Biomarcador	Tipo	Uso	Ref.
APC (germinal)	Dx.	Identificación de casos hereditarios de la Poliposis Adenomatosa Familiar	(16,53)
Genes del MMR (Germinal)	Dx.	Identificación de casos hereditarios del síndrome de Lynch	(16,53)
MUTYH (germinal)	Dx.	Identificación de casos hereditarios de del síndrome de poliposis por <i>MUTYH</i>	(16,53)
Sangre oculta en heces	Dx.	Presencia de enfermedades del colon, como CCR	(84)
ctDNA postoperatorio	Pro.	Asociado a peor desarrollo de la enfermedad y menor supervivencia libre de enfermedad	(85–87)
miR-200	Pro.	Mayor expresión en plasma asociada con progresión, EMT y metástasis	(84)
SEPT9	Dx.	Detección de metilación en plasma en pacientes de CCR en plasma	(88)
DPYD	Pre.	Determinación efectos adversos al tratamiento con 5-FU y capecitabina	(83,89)
UGT1A1	Pre.	Determinación de efectos adversos al iriotecán	(90)
KRAS y NRAS (p.G12 y p.G13)	Pre. Pro.	Resistencia a tratamientos con anticuerpos anti-EGFR Asociado a un peor desarrollo de la enfermedad	(70)
BRAF (p.V600E)	Pre. Pro.	Resistencia a tratamientos con anticuerpos anti-EGFR y respuesta al doblete encorafenib-cetuximab Asociado a un peor desarrollo de la enfermedad	(74,76)
Sobreexpresión de HER2	Pre.	Resistencia a tratamientos con anticuerpos anti-EGFR	(76)
Patrones de expresión	Pro.	Estratificación del riesgo de recaída.	(91,92)
Fusiones NTRK	Pre.	Respuesta inhibidores tirosina quinasa contra las fusiones de NTRK	(93)
MSI o defectos MMR (somático)	Pre.	Respuesta a inmunoterapia (anti-PD-1 y anti-CTLA-4) Menor respuesta a 5-FU (MSI)	(18,94,95)
CEA	Pro.	Asociado a un peor desarrollo de la enfermedad	(49)
Pérdida de CDX2	Pro.	Asociado a menor supervivencia libre de enfermedad	(46)

En tumor primario, uno de los principales usos del diagnóstico molecular es la detección de biomarcadores asociados a casos hereditarios (16,53) y a la identificación de toxicidad al tratamiento. En este segundo caso destacan las determinaciones en *DPYD* como marcadores de aumento de toxicidad con tratamiento con 5-FU y capecitabina (83,89) y *UGTA1* como marcador de efectos

adversos al iriotecán (90). Por otro lado, las mutaciones en los genes del MMR o alteración de expresión detectada mediante inmunohistoquímica, junto al MSI, son biomarcadores que son también importantes para la estratificación de la enfermedad, como se ha comentado anteriormente. A nivel predictivo, la presencia de MSI es marcador de ausencia de beneficio en el tratamiento con 5-FU (94,95) y de respuesta a inmunoterapia (18). Los patrones de expresión del tumor juegan también un rol pronóstico, gracias a paneles como *Oncotype DX* y *GeneFx Colon*, enfocados a establecer un grado de riesgo en la recaída a partir de la expresión génica del tumor, pudiendo colaborar en la decisión de si tratar con terapia adyuvante o no a un paciente de estadio II (91,92). Por último, el ADN tumoral circulante (*circulating tumor DNA*, ctDNA) en sangre posterior a la cirugía ha mostrado jugar un rol en el riesgo de recaída del paciente con enfermedad localizada. Este biomarcador se desarrollará en posteriores apartados.

En enfermedad avanzada y metastásica, además de los marcadores de toxicidad, es importante añadir las mutaciones de los codones 12 y 13 de KRAS y NRAS y la mutación p.V600E de BRAF (Tabla 3), por su valor pronóstico de peor desarrollo de la enfermedad y también como biomarcador predictivo de resistencia a los tratamientos anti-EGFR (53,57,83), como se ha expuesto anteriormente. La expresión de *HER2* es otro biomarcador, asociado también a un peor pronóstico y a mejor respuesta a los fármacos dirigidos contra este receptor (66). Por último, la fusión de cualquiera de los genes *NTRK* se ha asociado con respuesta a uno de los pocos fármacos aprobados independientemente del origen del tumor (93), aunque la frecuencia de dicha alteración es bastante reducida, encontrándose en cerca del 0,2-0,3% de pacientes de CCR (96,97).

Junto a las mutaciones y los perfiles de expresión, las secuencias no codificantes como los miARN, pequeños ARN de entre 20-30 nucleótidos y los ARNinc, superiores a 200 nucleótidos, son mecanismos de regulación epigenéticos que

pueden suponer biomarcadores, pero que no han logrado demostrar utilidad en la clínica (84).

Por último, además de los biomarcadores encontrados en la célula, hay otros que se pueden encontrar fuera del tumor. El CEA sería un biomarcador pronóstico de carácter proteico, encontrado en la sangre (49). Las heces también serían fuente de biomarcadores asociados a CCR, al presentar sangrado oculto o presencia de mutaciones o patrones de metilación concretos en el ADN fecal (84).

1.2.9. Medicina de precisión en CCR

De una forma más tradicional, se ha buscado dar solución concreta a una enfermedad, independientemente del individuo. Los avances en el entendimiento del genoma y el aumento en la comprensión sobre todos los factores que le rodean (mutaciones, epigenética, factores ambientales...) han ido definiendo un avance más centrado en qué características de un paciente de una misma enfermedad hacen que tenga diferente curso que otro paciente. Gracias a la identificación de estas características, poco a poco se ha ido planteando una medicina de precisión en la que el uso de estos biomarcadores permita guiar la terapia más adecuada a una enfermedad en función de las características propias de la enfermedad del individuo (98,99). En CCR, la medicina de precisión va a permitir establecer un riesgo de recaída, identificar posibles registros familiares de cáncer y guiar el tratamiento, tanto a nivel de adyuvancia como terapia dirigida en enfermedad metastásica (100), tal y como se ha desarrollado en el apartado anterior.

1.3. Biopsia líquida

Como se ha comentado previamente, la principal fuente de material para el diagnóstico molecular es el propio tumor, que se conserva fijado en formalina y embebido en parafina (*Formalin-Fixed, Paraffine Embedded, FFPE*). Pese a ser el principal método de diagnóstico en los tumores, este procesado presenta una serie de limitaciones (101). En primer lugar, los cortes que se procesan a partir de la biopsia, e incluso la misma pieza de tejido, pueden suponer una representación parcial del tumor. Cuando se procesa un corte del tumor, se puede estar dejando de incluir diferentes clones y subclones que puedan tener relevancia en la enfermedad, perdiéndose parte de la heterogeneidad tumoral. En segundo lugar, la biopsia puede no ser el único tumor del paciente e incluso puede ser de un momento en el que ya haya sucedido la diseminación de los clones que iniciarían la metástasis (102,103). Por último, la biopsia es una muestra estática. En ocasiones, los análisis se solicitan en fechas posteriores a la extracción y existe la posibilidad de que los resultados no reflejen la situación real del paciente, debido a los cambios que se puedan haber producido tras el paso del tiempo (104,105).

Otro problema que surge es la poca sensibilidad que pueden presentar las técnicas actuales de seguimiento de los pacientes tras la resección del tumor. Como se ha comentado, el seguimiento incluye evaluación del CEA y pruebas por imagen, que, siendo estrategias útiles, presentan falsos positivos entre el 7 y 16% y cerca de un 40% de falsos negativos. Además, las técnicas de imagen carecen de la capacidad de detectar lesiones tempranas y microscópicas (49).

Frente a estas situaciones, surge la biopsia líquida, la cual se define como una técnica no invasiva basada en la obtención de biomarcadores de fluidos corporales. Por esta definición, la sangre, la saliva, la orina o el líquido cefalorraquídeo (Fig. 3A) son las principales fuentes de muestras para biopsia líquida (101,106). Dentro de estas muestras, se podrán encontrar células,

vesículas y material genético procedente de los tejidos con los que interacciona el fluido. A diferencia de las biopsias de tejido, la obtención de una muestra de biopsia líquida es posible de obtener en cualquier momento del transcurso de una situación biológica y nos permitiría evaluar a tiempo real y de forma dinámica dicho evento (101,103). La tabla 4 recoge las ventajas y desventajas que se han discutido e irán discutiendo a lo largo del documento.

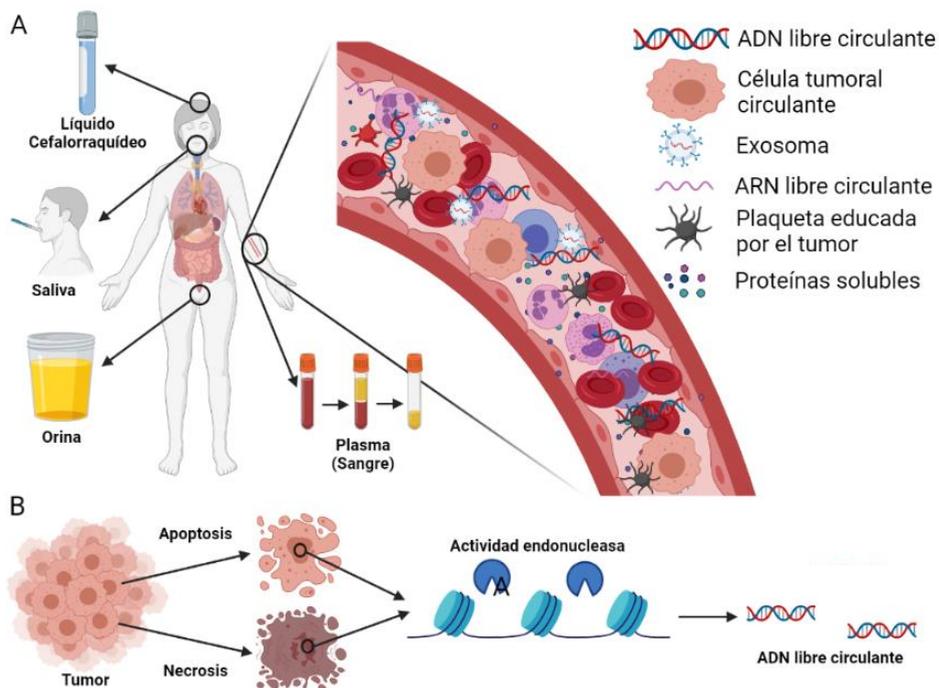


Figura 3: Fuentes de la biopsia líquida y utilidad en cáncer. A) Representación de fluidos que son empleados en la biopsia líquida. Los elementos representados en la circulación sanguínea como las células tumorales, los exosomas o el ADN libre circulante pueden aparecer en el otro tipo de muestras (saliva, líquido cefalorraquídeo, orina). B) Esquema del origen del ADN circulante. Tras los procesos de muerte celular, el ADN es cortado por nucleasas. El ADN protegido por las histonas es el que pasa a circulación. (Elaborada en www.biorender.com)

Tabla 4: Ventajas y limitaciones de la biopsia líquida.

Ventajas	Limitaciones
<ul style="list-style-type: none"> • Técnica poco invasiva para obtener la muestra • Monitorización a tiempo real de enfermedades • Obtención de material cuando el tejido no está disponible • Diferentes tipos de moléculas/células para analizar • Utilidad en diferentes etapas de una enfermedad (diagnóstico temprano, monitorización...) • Primeros biomarcadores para estudiar en plasma aprobados • Captación de la totalidad de la heterogeneidad en tumores 	<ul style="list-style-type: none"> • Falta de estandarización en los métodos de obtención de material • Poco material, según que se pretenda estudiar • Fuentes de falsos positivos • Alta dilución del material de interés • Requerimiento de tecnologías altamente sensibles • Falta de validación clínica

1.4. Biopsia líquida en cáncer

De los posibles fluidos empleados en biopsia líquida, muchos de los estudios en tumores optan por el uso del plasma procedente de sangre periférica (101–103). Aunque en función de la histología del tumor se pueden usar otros fluidos: los tumores del sistema nervioso central y de las metástasis en estos órganos se van a apoyar también en el uso de líquido cefalorraquídeo (107–109), mientras que los del tracto urinario emplearán la orina (110). En el caso del CCR, se emplea el plasma, aunque hay estudios que han intentado emplear la orina en busca de marcadores en enfermedad localizada y metastásica (111–113).

1.4.1. Componentes de la biopsia líquida de plasma

ADN libre circulante

El ADN libre circulante (*circulating free DNA*, cfDNA) está compuesto por el ADN localizado en la fracción plasmática de la sangre (Fig. 3A). En general, este cfDNA procede de las células que circulan en la sangre (células blancas y plaquetas) y, en menor medida, del resto de tejidos del cuerpo. En el caso del cfDNA proveniente

del tumor, éste procede de la apoptosis y necrosis de las células tumorales, ya sea por liberación propia o mediante liberación posterior a la digestión de los restos celulares por parte de macrófagos (Fig. 3B). En la generación de los fragmentos se ha descrito la participación de nucleasas como la DNASE1L3 (114). Además, procesos como la senescencia celular juegan un rol en detener la liberación del cfDNA (115). En individuos sanos la concentración ronda unas 1500 copias/mL (equivalente a 5 ng/mL) y la longitud es próxima a 147pb (nucleosoma) y 166pb (nucleosoma junto a espaciador) (116,117), estimándose su vida media en plasma y 37°C en 2 horas (118). En comparación con controles sanos, los pacientes de cáncer presentan una concentración de cfDNA variable pero superior (119). También se observa que la longitud de los fragmentos de ctDNA difiere ligeramente del procedente de células no malignas, asociándose una mayor proporción de fragmentos más cortos con pacientes de cáncer con mayor cantidad de ctDNA que con individuos sanos, pero no de tal forma que supongan un biomarcador (120).

Células Tumorales Circulantes

Inicialmente, un tumor sólido presenta características más epiteliales. La acumulación de mutaciones lleva a que el tumor adquiera un fenotipo más mesenquimal, caracterizado por la pérdida de la dependencia a la adhesión celular, aumento de la capacidad angiogénica y desarrollo de motilidad (121). Este fenotipo invasivo, adquirido mediante la EMT, permite que la célula alcance el torrente sanguíneo, y aquellas que se intravasen entran en la circulación (Fig. 3A). Estas Células Tumorales Circulantes (*Circulating Tumor Cells*, CTC) serán las que extravasen en determinados nichos y posteriormente desarrollen las posibles metástasis (122). Estudios han visto que la cantidad de células circulantes retenidas en individuos con cáncer es superior a la de donantes sanos, donde el conteo es muy cercano a cero, y se ha asociado su aparición a tumores con características invasivas (metástasis e invasión de nodos) (123). Estas CTCs serán

evaluadas para identificar la presencia de ciertos marcadores proteicos, determinar alteraciones genómicas y establecer modelos *ex vivo* en cultivo o animales para estudiar respuestas a fármacos y otras características. La principal limitación es su escasa presencia en los pacientes, en los que la concentración puede situarse por debajo de 10 células/mL (124).

Exosomas y vesículas extracelulares

Como parte de su comunicación con otras células, el tumor secreta vesículas extracelulares de diversos tipos y tamaños. De ellas, los exosomas son vesículas de entre 30 y 150 nm de diámetro de origen endocítico (Fig. 3A). Este término también se usa para referirse a vesículas derivadas de la membrana (125). Su secreción depende de la activación de la ruta Wnt y de la sobreexpresión de Rab3D (126). Estas estructuras se han identificado en todo tipo de fluidos corporales, incluidos plasma y orina (106). Respecto a su contenido, llevan proteínas, lípidos asociados a la balsa lipídica y ácidos nucleicos, como ARN mensajero (ARNm), ARNlnc, miARN y ADN (127). Por su naturaleza, estas estructuras van a fusionarse con las membranas de otras células y van a modular su actividad. En el contexto del tumor, los exosomas van a colaborar en la angiogénesis, en la migración y en la formación del nicho metastásico por reclutamiento de células derivadas de la médula ósea (126,128). Diferentes estudios relacionan una mayor presencia de exosomas en pacientes de CCR y páncreas que en individuos sanos (129,130), y se espera que la expresión diferencial de miARNs sirva de biomarcador para detección temprana y detección de recaída (126).

Otros componentes de interés

Además del contenido de los exosomas, se han descrito la presencia de ARNm y miARN circulantes en plasma (128). Se cree que ambos tipos de ARNs ajenos a exosomas están asociados a proteínas que los protegen de las RNasas del plasma (131). También se ha descrito la presencia de otros ARNs cortos, como de

transferencia, piwi-interactivo, pequeño nucleolar y pequeño nuclear, en muestras de suero (132).

Otro elemento de interés son las plaquetas educadas por el tumor (Fig. 3A). Mediante la liberación de vesículas con miARNs y otros componentes que son capaces de modular la actividad celular, los tumores son capaces de estimular los mecanismos detenidos de las plaquetas procedentes de los megacariocitos de los que se originaron, causando un cambio en el comportamiento de estas (133). Dichas plaquetas, a su vez, liberarán desde moléculas señalizadoras hasta vesículas con determinados transcritos que van a modular su entorno. Así, estas plaquetas estimuladas por el tumor van a ser cruciales en la metástasis a través de la formación de un agregado entre ellas y células tumorales, el cual preparará un punto para la metástasis y protegerá a las células tumorales de los efectos físicos de la circulación sanguínea (134). A nivel celular, las plaquetas de este agregado colaborarán en la degradación de la matriz celular de los vasos, facilitarán la extravasación de las células metastásicas, promoverán la angiogénesis y ayudarán a la evasión inmune de estas células (134,135).

La biopsia líquida también ha sido empleada para el estudio de marcadores proteicos de relevancia en las patologías (Fig. 3A). En CCR, el CEA plasmático ha mostrado tener mal pronóstico en pacientes positivos y con aumento en muestras posteriores (136), y el marcador CA724 se ha asociado con un peor pronóstico en pacientes con CEA normal (137). El problema del uso de marcadores proteicos como el CEA, CA-125 o CA 19-9 es que se presentan en más patologías y no solo en tumores (138)

1.4.2. Tecnologías de evaluación del cfDNA

El estudio del cfDNA se enfrenta a un gran problema: la proporción de cadenas tumorales frente al total del ADN circulante puede llegar a ser muy reducida (139).

Y teniendo en cuenta que mutaciones de relevancia, como las de resistencia a tratamientos, pueden estar en baja proporción en el tumor, esto hace que la detección de mutaciones en el cfDNA se enfrente a los límites de las tecnologías convencionales a la hora de detectar estas alteraciones. Para determinadas aplicaciones, se espera que una tecnología de estudio de cfDNA sea capaz de distinguir mutaciones hasta un 0,01% de frecuencia en el alelo mutado (*Variant allele frequency, VAF*), el equivalente a 1 cadena mutada entre 10000 (140). Para ello, se han desarrollado nuevas técnicas, las cuales mediante modificaciones de protocolos de la reacción en cadena de la polimerasa (*Polymerase Chain Reaction, PCR*) permiten alcanzar dicha sensibilidad y otras que mediante la modificación de los protocolos de generación de librerías de secuenciación masiva (*Next Generation Sequencing, NGS*) permiten reducir el ruido de fondo (141).

1.4.3. Técnicas basadas en PCR para detección del cfDNA

Mediante el uso de modificaciones en las técnicas clásicas de PCR, se ha conseguido desarrollar metodologías que permitan distinguir entre diferentes alelos con frecuencias inferiores al 1%. Teniendo en cuenta la naturaleza del cfDNA y las características de la PCR, la principal limitación, en el diseño de todas estas técnicas, se va a encontrar en la distancia entre el principio y el final del amplicón. Al componerse de fragmentos de aproximadamente 140-166pb, el amplicón debe ser corto, recomendablemente menor a 100pb y próximo a 80pb, para un mejor rendimiento (142–144). A su vez, según la tecnología, entre el extremo 3' de ambos cebadores, debe existir espacio suficiente como para que hibriden sondas.

La *Amplification Refractory Mutation System-qPCR (ARMS-qPCR)* se basa en el uso del extremo 3' de los cebadores, amplificando solo si se encuentra con el alelo específico, ya que sin hibridación no se puede realizar la extensión (162). Algunas

de las estrategias se apoyan también en el uso de cebadores con 3' bloqueado, impidiendo la extensión de uno de los dos alelos (146,147). Esto permite discernir mutaciones puntuales y determinadas delinsersiones, habiéndose reportado frecuencias tan bajas como el 0,015% (148).

La PCR con co-amplificación a menor temperatura de desnaturalización (COLD-PCR) se basa en el uso de pequeñas diferencias en la temperatura de disociación (T_m) de dos cadenas de menos de 200pb para conseguir una amplificación de ambas (149,150). Esta técnica es útil como paso previo a otras tecnologías de detección, como la secuenciación, además de permitir un enriquecimiento de la secuencia mutada sin depender completamente del tipo de mutación (141,149). Esta tecnología permite detectar mutaciones al 0,01% cuando hay un input superior a 1 ng (151,152).

La PCR digital es una modificación de las PCRs cuantitativas (*quantitative* PCR, qPCR), en las que en lugar de buscar una señal minoritaria en una mezcla de alelos que emitirán dos señales, se cuentan las señales de partículas generadas mediante la creación de pequeños reactores con la solución necesaria para la reacción y una única cadena de la secuencia a analizar (141). Así, cada reactor se espera que emita la señal de un alelo u otro, según cual contenga. Para generar esta distinción, entre los reactivos se incluyen sondas *TaqMan* que puedan distinguir entre mutado y no mutado, cubriendo normalmente mutaciones puntuales y pequeñas delinsersiones, aunque también se ha descrito su uso para cuantificación del número de copias y para expresión (141). El principal tipo de PCR digital es la *droplet digital* PCR (ddPCR), en la que los pequeños reactores o droplets se crean mediante emulsiones. De este tipo, destacan el *Beads, Emulsión, Amplification and Magnetics* o BEAMing, en el que se cuentan esferas magnéticas envueltas en la cadena amplificada e hibridadas con una sonda complementaria a la mutación (153), y la metodología desarrollada por Bio-Rad, donde se detecta la fluorescencia de la emulsión, generada por la ruptura de sondas *TaqMan*

durante la amplificación del alelo (154). Esta modificación de la PCR ha permitido alcanzar límites de detección de hasta el 0,003-0,005% entre 10000 copias totales (155,156).

1.4.4. Secuenciación masiva en el estudio del cfDNA

La secuenciación masiva está establecida hoy en día como una de las principales herramientas de diagnóstico en tumor sólido, ya que permite la detección simultánea de una gran cantidad de pequeñas alteraciones que permiten orientar el tratamiento del paciente. Dado el coste-beneficio que supone el uso de la NGS, se ha trasladado su uso también al cfDNA procedente de muestras de biopsia líquida.

Encontramos dos tipos de procesos de generación de librerías: basados en captura y basados en amplicones. Los primeros están basados en hibridar sondas marcadas con regiones específicas del genoma que se recuperan mediante interacciones específicas, como la biotina y la estreptavidina (Fig. 4A). Esta química usa paneles muchos más amplios y el uso de sondas elimina algunas de las limitaciones de la PCR, al ser una única hibridación de la región de interés. Además, en caso de querer realizarse estrategias de secuenciación del genoma completo (*Whole Genome Sequencing*, WGS), los pasos de hibridación no son necesarios. Como principal contra, estos paneles suelen ser más caros y requerir de mucha más capacidad de secuenciación para poder recurrir a ellos, junto a una mayor exigencia de material de partida para obtener resultados adecuados (141,157). En un punto intermedio se encuentran los protocolos de PCR anclada (Fig. 4B), en los que la incorporación de un extremo con una secuencia universal permite amplificar las regiones de interés en un posterior paso con un cebador universal y otro específico. Por último, los métodos basados en amplicones (Fig. 4C) utilizan cebadores para seleccionar las regiones de interés. Al igual que los

métodos de PCR, el tamaño del amplicón afectará al porcentaje de regiones de interés que se podrán amplificar del cfDNA. Pese a estos inconvenientes, dentro de sus principales beneficios suele entrar la versatilidad de los paneles y un coste reducido (141,157).

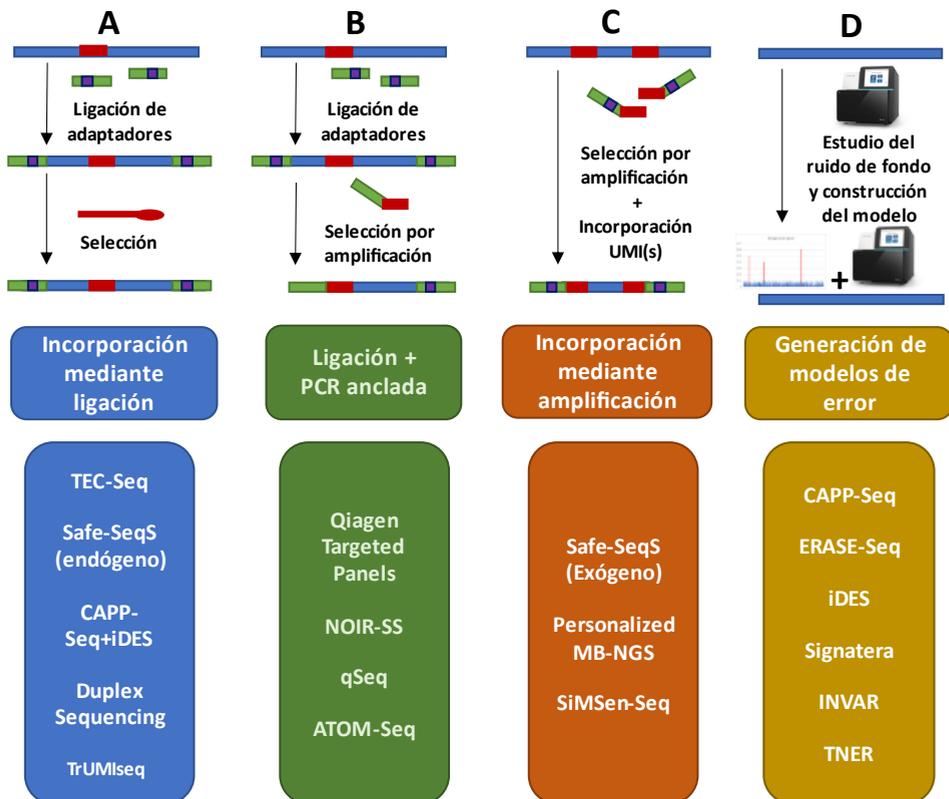


Figura 4: Metodologías de reducción de errores en secuenciación masiva. Se representa brevemente el funcionamiento de cada tecnología y las metodologías asociadas a cada una, para A) incorporación de UMIs con ligación, B) uso de la adición de un adaptador y PCR anclada, C) mediante uso de amplicones y D) mediante modelos de reducción de error apoyados en desarrollos bioinformáticos. Azul: ADN libre circulante. Verde: Adaptadores Illumina. Rojo: regiones para la selección. Violeta: Índices de secuenciación.

Pese a que la NGS está bien establecida, ello no quita que la biopsia líquida deba hacer frente a otros desafíos. La baja frecuencia en la que se pueden encontrar

las variantes en el cfDNA se enfrenta a los límites de detección establecidos en la NGS, situados en el 0,5% en algunos casos. También es importante un procesamiento adecuado de las muestras, debido a la baja vida media del cfDNA sin separar de la sangre. Además, entran en juego implicaciones biológicas, como otras mutaciones somáticas no debidas a cáncer, como las descritas mutaciones de hematopoyesis clonal, y pacientes cuyos tumores liberen poco ctDNA, junto a la baja cantidad de cfDNA de la mayoría de las muestras, siendo casi imposible detectarlo (158–161). La NGS, además, presenta errores con frecuencias de hasta el 0,1% en cambios como A>G y T>C (162). Contra estas situaciones, la secuenciación masiva se ha apoyado en técnicas de etiquetado molecular y análisis informático para reducir el ruido de fondo y los falsos positivos (Fig. 4D) y en la secuenciación pareada de células blancas para evitar falsos positivos derivados de mutaciones somáticas no procedentes del tumor.

Etiquetado molecular

Con el fin de aumentar la sensibilidad de la NGS, se han desarrollado las denominadas etiquetas moleculares, conocidas como UMI (*Unique molecular identifier*) o UID (*Unique Identifier*). El fundamento es la asignación de una secuencia aleatoria de entre 8 y 15 nucleótidos antes de cualquier paso de amplificación, de forma que todas las secuencias que hibriden en la misma región y contengan el mismo UMI, serán duplicados de PCR de la misma molécula original. La finalidad, por tanto, es eliminar falsos positivos a través de las secuencias consenso procedentes de la misma familia de UMIs. También posibilita un conteo más próximo al real de las cadenas analizadas y de las frecuencias de las variantes (141,160). A nivel de secuenciación, esto requerirá de mayores coberturas para que, con mayor número de repeticiones del correspondiente UMI, se puedan alcanzar buenas coberturas consenso. La forma de añadir estos UMIs es mediante ligación, aunque existen también estrategias basadas en hibridación y extensión, empleando un bajo número de ciclos de PCR.

La principal estrategia, el uso de ligación para incorporar los UMIs (Fig. 4A), aprovecha que el tamaño de cfDNA es el adecuado para las lecturas del secuenciador. Según el diseño, el marcaje irá junto a uno de los índices de la muestra o junto a la misma secuencia, afectando a en qué momento de la secuenciación se leerá el identificador único. El diseño también podrá etiquetar y reconocer cuál de las dos cadenas se está leyendo, reduciendo errores de PCR tempranos. Entre los primeros diseños propuestos está el *Safe-SeqS* con “marcaje” endógeno, en el que se aprovechaban los extremos de la cadena para identificar duplicados de PCR y establecer consensos, defendiendo la reducción de errores en al menos 70 veces (163). Estrategias como *TEC-Seq* (138), *CAPP-Seq+iDES* (164), *Duplex Sequencing* (165) o *Singleton Correction* (166) presentan la incorporación de un UMI en cada extremo de la molécula mediante ligación, lo que permite también orientar la lectura y saber cuál de las dos cadenas está siendo leída. Esto permitirá hacer un consenso de cada uno de los sentidos, y luego un posterior consenso de ambos, alcanzándose frecuencias del 0,1% y del 0,05% (138), e incluso del 0,034% en posiciones de interés (165,166).

Una variación a las estrategias con captura son las estrategias con PCR anclada (Fig. 4B). Dicha tecnología es representada por las metodologías de QIAGEN, las cuales han demostrado encontrar variantes próximas al 0,1% (167–169). Bajo el mismo fundamento, la *Non-Overlapping Integrated Reads Sequencing System* o NOIR-SS usa una amplificación lineal con cebador específico, amplificando solo una de las hebras, o bien usando la estrategia de QIAGEN con cebador universal y oligonucleótido específico para incluir ambas. NOIR-SS ha llegado a reducir el error en las bases por debajo del 0,01% y confirmar variantes somáticas al 0,02% (170,171). Otra tecnología, *Adaptor Template Oligo Mediated Sequencing* o ATOM-Seq usa un cebador complejo con una región aleatoria que actuará como UMI y otra que permitirá unirse de forma aleatoria a los extremos del cfDNA.

ATOM-Seq ha demostrado capacidad de detectar con seguridad variantes al 0,1% con al menos 20 ng de cfDNA (172).

En cuanto a la incorporación mediante PCR (Fig. 4C), se basa en añadir el UMI a uno o dos de los cebadores, como parte de la cola del adaptador. De las primeras estrategias en emplearlo fue *Safe-SeqS*, que introducía el marcaje mediante 2 ciclos de PCR y considerando positiva una mutación si aparecía en el 95% de las lecturas (163). *AmpliSeq*, la estrategia empleada por la compañía ThermoFisher, introduce UMIs en el extremo 5' de ambos cebadores, marcando la molécula con 3 ciclos de PCR. Esta estrategia ha permitido detectar variantes por encima del 0,25% (173,174). Aunque debido a la inespecificidad introducida por la secuencia aleatoria del marcaje, otras metodologías han recurrido a estrategias o modificaciones complejas que permitan reducir la formación de productos inespecíficos o autocomplementarios. Por ejemplo, en lugar de usar una secuencia de 15 nucleótidos aleatorios, el grupo de Masunaga *et al.* ha descrito el uso de un cebador de la primera PCR de 15 ciclos con la combinación de nucleótidos BDHVBDHVBDHVBDH (de acuerdo con las nomenclaturas internacionales), excluyéndose un nucleótido en cada una de las posiciones. Esta estrategia, a la que han denominado *Personalized MB-NGS*, les ha permitido detectar con alta seguridad mutaciones somáticas identificadas en el tumor previamente a frecuencias tan bajas como el 0,1% en estadios tempranos y metastásicos de cáncer de mama (175). Otro protocolo introducido por el grupo de Ståhlberg *et al.* consiste en el uso de una horquilla dentro de uno de los cebadores donde se esconda la secuencia aleatoria en una PCR de 3 ciclos. Así, a temperaturas de alineamiento, la horquilla aparece como una estructura secundaria cerrada e impidiendo la inespecificidad, mientras que a temperaturas de extensión y de hibridación de los cebadores con índices en una segunda PCR, la horquilla se rompe y permite una extensión normal. Esta técnica, llamada *SiMSen-Seq* (contracción de *Simple, multiplexed, PCR-based barcoding of DNA for*

sensitive mutation detection using sequencing) reduce el error a frecuencias inferiores a 0,05%, permitiendo identificar variantes por encima del 0,1% (144,176).

1.4.5. Análisis bioinformáticos y estrategias sin marcaje

El proceso de llamado de variantes es también importante una vez obtenidas las familias de UMIs. Diferentes softwares van a permitir la obtención de las secuencias consenso a partir de los diferentes duplicados. Entre estos softwares se encuentran *Umi-Tools* (177); *Caleb*, basado en *clustering* (178); *UMI-VarCal*, cuya reducción de ruido de fondo es centrado en la calidad de las bases, haciéndolo independiente de muestras negativas (179) o *gencore*, cuyas correcciones están enfocadas en el genoma de referencia (180).

La bioinformática también puede jugar otro rol definiendo modelos de ruido de fondo que permitan facilitar distinguir los verdaderos positivos (Fig. 4D). Estos modelos pueden ser por sí solos, como propone el modelo inicial de *CAPP-Seq* (181), o pueden complementar al uso de los UMIs, como ocurre con el *CAPP-Seq+iDES*, en el que existe un estudio de controles negativos para detectar los errores inherentes al procesado y secuenciación de las muestras (164). En el caso de la compañía comercial Natera y su estrategia *Signatera*, la reducción de error también se basa en generar modelos de ruido de fondo de sus amplicones en individuos sanos (142). El software *ERASE-Seq* trabaja de forma similar, comparando al menos triplicados técnicos tanto de muestras control como de las de interés en amplicones, mostrando sensibilidades y especificidades del 100% en frecuencias del 0,3% y buenos rendimientos al 0,1% (182). El software *INVAR* archivos de lecturas colapsadas, procedentes de UMIs o no, del tumor y del plasma para encontrar variantes concordantes y determinar aquellas que no lo son para posteriormente analizarlas, teniendo en cuenta la ratio de falsos

positivos obtenidos a partir de muestras control (156). Por último, *TNER* estudia el contexto en base a trinucleótidos para determinar un modelo de error (183). En cualquier caso, todas estas estrategias se basan en la construcción de un modelo de error para un panel específico. Modificaciones del panel requerirían crear nuevos modelos.

Otras estrategias sin marcaje juegan con propiedades físicas de los procesos de secuenciación, como el método de *Tagged-Amplicon Sequencing* (TAm-Seq) y su versión modificada *enhanced TAm-Seq* (eTAm-Seq), apoyada por los ensayos de la compañía Inivata, que crean un paso de preamplificación combinado con la separación de las cadenas y amplificación en un array, que junto a un mejorado software de análisis en el eTAM-Seq, permiten identificar variantes por encima del 0,1% (184,185).

1.4.6. Hematopoyesis clonal como fuente de falsos positivos

Como se ha comentado previamente, el cfDNA procedente de muestras de biopsia líquida y, de plasma en concreto, acumula ADN que no solo procede del tumor, sino que también es aportado por otros tipos celulares. Esto da lugar a que no solo los errores asociados al procesamiento sean los únicos en detectarse, ya que otras mutaciones, no procedentes del tumor, pueden aparecer en baja frecuencia en las muestras de biopsia líquida, haciendo que se identifiquen auténticas alteraciones pero que no proceden del tejido tumoral (186,187). Parte de estas variantes proceden de la hematopoyesis clonal, un proceso por el cual las células precursoras de los linajes hematopoyéticos desarrollan y acumulan mutaciones somáticas que acaban apareciendo en las versiones maduras (186,188). Estas mutaciones se asocian a pacientes de mayor edad (186,187) y se detectan en genes como *DNMT3A*, *TET2*, *ASXL1*, *JAK2*, *PPM1D* y *TP53* (189). Aunque aparecen en genes que se pueden asociar a tumores, muchas de estas

variantes presentan un contexto clínico variable, asociándose en algunos contextos con riesgos de enfermedad hematológica (189,190). Debido a la incertidumbre en su significado, el proceso y estas variantes son llamadas de hematopoyesis clonal con significado incierto o CHIP (*Clonal Hematopoiesis of Indeterminate Potential*).

En relación con este tipo de variaciones somáticas, se realizó un estudio en cfDNA de pacientes metastásicos e individuos sanos, empleando una tecnología con reducción de ruido de fondo y fiabilidad por encima del 0,2% de frecuencia (186). Solo el 24,4% de las alteraciones encontradas en plasma se pudieron asociar como concordantes a las descritas en el tumor. Al cuestionar el origen del resto de alteraciones, se concluyó que la mitad y cuatro de cada cinco provenían de células blancas en pacientes y controles, respectivamente. En estos últimos, dos de cada tres individuos mostraban al menos una variante de origen desconocido o no detectada en células blancas. En pacientes, aquellos con fenotipos hipermutados concentraban un alto número de cambios no concordantes, los cuales se podrían asociar con heterogeneidad tumoral (186). Otro estudio asoció el desarrollo o la selección de mutaciones en genes de respuesta al daño al ADN, como *TP53*, *PPM1D* y *CHEK2*, en pacientes tratados con agentes citotóxicos o radioterapia (191). Es importante, por tanto, caracterizar también estas variantes CHIP para distinguirlas de aquellas de origen tumoral somático, aunque algunos paneles comerciales y estrategias cuentan con softwares que clasifican variantes del plasma como CHIP sin necesidad de pareado germinal (192).

1.4.7. Translación de la biopsia líquida

Con el uso de las tecnologías de secuenciación masiva, recientemente se ha planteado una clasificación de las estrategias de detección del tumor en la biopsia líquida, en función del tipo de señales detectadas. La primera generación se basa

en detectar señales digitales de unas pocas mutaciones, siempre con enfoques dirigidos a mutaciones concretas, como los que se puede encontrar a través de la PCR convencional o la ddPCR. La segunda generación ampliaría el rango de mutaciones estudiadas mediante el uso de paneles de genes. La tercera generación estaría centrada en análisis de señales analógicas como patrones de metilación, tamaño de fragmentos o las regiones cubiertas por los nucleosomas (193). Tanto la segunda como la primera, las más estudiadas en la última década, usan las mutaciones somáticas para describir la presencia o no y las características moleculares del tumor. Por tanto, estas mutaciones detectadas a través del plasma permitirán detectar enfermedad mínima residual (EMR) o establecer un diagnóstico temprano, identificar la aparición de resistencias a los tratamientos, estudiar la posible heterogeneidad tumoral o establecer la evolución clonal del tumor (Fig. 5) (101,193).

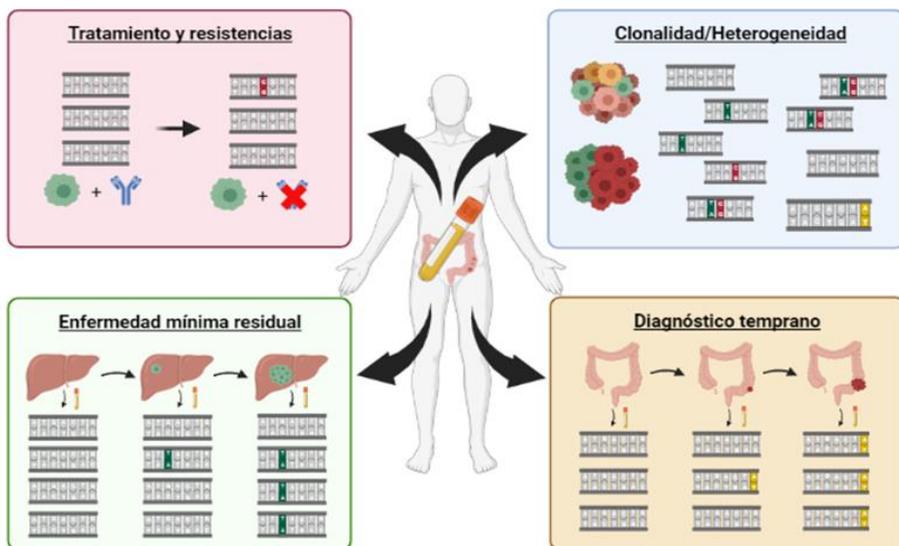


Figura 5: Principales usos de la biopsia líquida en la clínica del CCR. La imagen representa como, con muestras tomadas en tiempos sucesivos, pueden presentar mutaciones que permitan identificar mutaciones de interés en plasma. Sin un tratamiento, se espera que en muestras sucesivas se incremente la proporción de cadenas mutadas de ctDNA en el cfDNA. Se espera también que la biopsia líquida permita identificar situaciones de heterogeneidad tumoral. (Elaborada en www.biorender.com)

Diagnóstico temprano

Uno de los desafíos de la oncología es la detección temprana y, por ello, el uso del ctDNA está siendo probado en diversos tipos de tumores para tal fin (Fig. 5) (194). El estudio en el que se describe TEC-seq identificó mutaciones en cerca del 75% de pacientes de CCR, el 66% de cáncer de ovario y en la mayoría de pulmón y mama en estadios I y II (138). El principal problema que surge para la detección temprana, junto a los propios de la biopsia líquida, es superar las sensibilidades requeridas y enfrentarse a un tumor que no libere ctDNA. Por lo general, un tumor en fase metastásica va a presentar más probabilidad de tener ctDNA detectable que uno en estadios más tempranos (195–197).

Enfermedad Mínima Residual

La enfermedad mínima residual o EMR es constituida por el resto de los tumores que permanece tras los tratamientos y que posteriormente supondrán el riesgo de recaída (198). Esta EMR es detectable a través de técnicas como PCRs cuantitativas dirigidas a mutaciones específicas o PCR digital hasta grandes paneles de NGS en busca de fusiones o mutaciones somáticas que puedan surgir de clones no identificados previamente (Fig. 5).

Los principales tumores en los que se ha descrito la EMR han sido de origen hematológico, asociándose la no detección de EMR con una mejor supervivencia y una menor probabilidad de recaída (85–87). También es usado para valorar la respuesta al tratamiento (85). En cuanto a los tumores sólidos, un estudio determinó que la presencia de EMR se asocia con peor pronóstico del tumor, habiéndose detectado una posible correlación negativa entre el número de mutaciones identificadas y su frecuencia, mientras que se definió una mejor supervivencia libre de enfermedad en aquellos pacientes que apenas tenían ctDNA detectable (195). Si bien en cáncer de colon se ha descrito un rol para la detección de la EMR, como se explicará más adelante, en otros tumores sólidos también. En el caso del cáncer de mama se ha descrito el uso de la detección de

la EMR, estableciéndose como criterio la positividad o no del ctDNA durante el seguimiento. Así, se ha descrito dicho potencial en plasma a las 2-4 semanas con una sensibilidad del 50% y del 80% en muestras de seguimientos (199) y anticipándose la recaída con una mediana de 10 meses, en aquellos sujetos con ctDNA positivo en muestras posteriores a la operación (200). En cáncer gástrico, la detección de mutaciones somáticas en ctDNA de plasma, incluyendo filtrado de variantes germinales mediante secuenciación de células blancas, en plasma postoperatorio de una pequeña cohorte de 20 individuos consiguió distinguir completamente los pacientes que recaían de los que no con una mediana de 8,9 meses de antelación (201). Todos estos estudios describen cómo el cfDNA postoperatorio, en periodos de 2-6 semanas o en seguimientos posteriores, permite anticipar la recaída, aunque difieren en las sensibilidades alcanzadas.

Seguimiento de respuesta y resistencia

Parte de las terapias dirigidas que existen hoy en día muestran el desarrollo de resistencias, a través de mecanismos que permitan mantener la activación de la misma vía en un punto más avanzado de la señalización o mediante el uso de vías paralelas (Fig. 5). Ejemplos de esto son las mutaciones de *KRAS* en los tratamientos anti-EGFR en CCR o la mutación p.T790M en pacientes *EGFR* positivos en cáncer de pulmón no microcítico (202). Con el fin de monitorizar la respuesta al fármaco y de pautar posibles cambios en el tratamiento, es posible seguir estas mutaciones conocidas en plasma. En CCR se han descrito las discrepancias entre muestras pareadas de plasma y tumor, tanto pre-tratamiento como tras la progresión, detectándose la presencia de mutaciones de RAS en ctDNA no detectadas en tejido. Esto refleja la posible presencia de subclones resistentes que condicionen el tratamiento (203).

El plasma puede ser usado también para descubrir nuevas alteraciones durante el tratamiento, ya sea mediante el uso de paneles dirigidos a los genes implicados en la terapia, como puede ser el estudio de *PTEN* y *ESR1* al tratar con inhibidores

de PI3K y aromatasa en cáncer de mama (204), o mediante el uso de estrategias con paneles más amplios, como el uso de la secuenciación de exoma completo (*Whole Exome Sequencing*, WES). Esta aproximación, WES, se ha utilizado para caracterizar mutaciones de resistencia en pacientes con tumores gastrointestinales y diferentes tratamientos, donde se han detectado posibles mutaciones de resistencia en un 78% de los pacientes (205).

Heterogeneidad tumoral y clonalidad

Los tumores no se componen de una masa homogénea de células, sino que, debido a su naturaleza, las células van acumulando errores en el ADN, diferentes modificaciones epigenéticas y otro tipo de alteraciones que las hace divergir entre ellas (Fig. 5). Este hecho, conocido como heterogeneidad tumoral (206), es una característica propia de la enfermedad y es responsable de fenómenos como las resistencias a fármacos, reflejadas como células que proliferarán con la presión selectiva de un tratamiento determinado (207,208). En concreto, se habla de heterogeneidad intratumoral al hablar de los diferentes clones y subclones que se pueden encontrar en la misma masa de células. Se distingue entre clones, si aparecen en etapas tempranas y con mayor peso en el tumor, y subclones, si son eventos más tardíos. Su detección es posible gracias a la identificación de genes denominados *drivers*, los cuales juegan un rol importante en las etapas tempranas de la enfermedad (209), y de complejos algoritmos que, en base a la proximidad con otras variaciones, las frecuencias de la mutación en la muestra y las alteraciones en el número de copias (*Copy Number Alteration*, CNV), establecen una predicción del rol en el desarrollo del tumor (210,211). También ha sido descrito cómo las metástasis presentan mutaciones de algunos de los clones y subclones y surgen nuevas mutaciones distintas a las que lleva el tumor primario (62), siendo posible detectar éstas en plasma (199,207). También se ha visto, en concreto en CCRm, como, pese a la disponibilidad de tejido de la progresión, un porcentaje elevado de las alteraciones de resistencia son solo observables a

través de las muestras de plasma, estando ausentes en las de tejido (212). Esto recalca la importancia de la biopsia líquida en la detección de resistencias, pero a su vez, el papel que juega en la búsqueda de esta heterogeneidad, no siempre identificada en las muestras de tejidos.

WES y WGS en biopsia líquida

Las estrategias de secuenciación dirigida suelen ir enfocadas a genes de interés o cuya relevancia ha sido descrita previamente en cada uno de los tipos de tumores de interés. Otras estrategias como el WES y el WGS buscan estas variantes a través de todas las regiones codificantes y del genoma completo, respectivamente. Debido a la capacidad de secuenciación que requiere cada estrategia, la detección de las variantes no puede ser realizada con la misma profundidad de lectura, haciendo que las mutaciones identificadas no se encuentren a tan baja frecuencia como con las estrategias con reducción de error (213). Esto hace que algunas estrategias opten por realizar WES en aquellos tumores en los que previamente han concretado la presencia de ctDNA con VAF alto (214,215), ya que, por lo general, los resultados descritos en plasma empleando WES se mueven en coberturas medias de 50-200X (216), mientras que los de WGS oscilan entre 0,3-35,0X (217–219). Estos estudios se han usado tanto en la caracterización del plasma al diagnóstico como para evaluar seguimientos de pacientes al tratamiento y la concordancia con tejidos como el metastásico (216). A nivel molecular, el WES combina el análisis de variaciones de nucleótido único (*Single Nucleotide Variation*, SNV) y CNVs, mientras que el WGS se enfoca en detectar CNVs y cambios estructurales.

En CCR se han realizado pocos estudios específicos o en combinación con otros tumores gastrointestinales. En una investigación de Parikah *et al.* (214) se buscaron mutaciones en plasmas a la progresión de individuos con resistencia al tratamiento en CCR y otros tumores gastrointestinales, observando un mayor número de pacientes con resistencias en las muestras de plasma que en las de

tejido primario. También encontraron que estas mutaciones se observaron en posteriores biopsias de tejido metastásico (214). Mediante WGS, el grupo de Zviran *et al.* ha trabajado en la detección de ctDNA en muestras de bajo cfDNA, combinando SNV y CNV, observando en muestras de CCR una sensibilidad del 100%, pero con un valor predictivo positivo del 42% (219). Por último, el grupo de García-Olmo y Olmedillas-López ha publicado dos estudios sobre un análisis denominado presencia diferencial de exones mediante el uso de datos de WES de plasmas de CCRm en pacientes y en ratones inyectados con líneas celulares en comparación con pacientes sin recaída, encontrando una firma de 379 y de 1028 exones, respectivamente, con representación diferencial en el plasma (220,221).

1.4.8. Biopsia líquida en CCR

De la misma forma que los tumores presentan características distintas en función de su histología, las perspectivas de la biopsia líquida en cáncer de colon son específicas en comparación con otros tumores. Las áreas donde la biopsia líquida ha tenido más desarrollo en CCR son la detección temprana con marcadores de metilación, el seguimiento de los genes *RAS* y la EMR, y enfermedad metastásica (222).

Uno de los principales usos de la biopsia líquida es el seguimiento de las mutaciones en genes *RAS*. Un estudio identificó estas mutaciones en pacientes metastásicos en plasma con un 93% de concordancia, observándose mayores frecuencias en pacientes que presentaban recaída hepática. También se observa como las frecuencias de las variantes reflejan la respuesta a los fármacos y cuando estos comienzan a fallar, volviendo a bajar los niveles de ctDNA con un nuevo tratamiento (223). En la misma cascada de señalización, la mutación p.V600E de *BRAF*, biomarcador de tratamiento, ha estado también en el punto de mira por la aparición de resistencias. Mediante WES y paneles de genes en pareados de tejido

y plasma, se ha identificado la aparición de resistencias en *KRAS*, *NRAS* y *MEK*, potenciando la activación de la vía (207,224).

Otro de los usos del plasma en los que más se ha profundizado en CCR ha sido la detección de EMR y predicción de la recaída, para la que se han empleado diferentes aproximaciones. La detección de ctDNA mediante mutaciones somáticas en muestras postoperatorias se asocia con una mayor frecuencia de recaída, mientras que los seguimientos longitudinales pueden llegar a una sensibilidad del 88% (142). Otro estudio realizado con un panel comercial mostró un valor predictivo positivo del 100% y sensibilidad del 55% para detectar recaída tras 1 mes de la cirugía o de la quimioterapia con un seguimiento superior a 1 año, incrementando la sensibilidad con muestras posteriores (192). Resultados similares se han alcanzado en otro estudio, empleando mutaciones patológicas identificadas en el tumor, observándose una tasa de recurrencia en la mitad de los pacientes ctDNA positivos en muestras a las 6 semanas de la operación y con aumento de la sensibilidad con muestras posteriores, identificándose el 82% de las recaídas con una antelación de 11,5 meses respecto a la recaída radiológica (225). Mediante el uso de otra tecnología, *Safe-Seq*, con pacientes en estadio II de cáncer de colon, se consiguió detectar ctDNA en el 8% de los pacientes, de los cuales el 80% recayó frente al 10% de los ctDNA negativos (226). Otro trabajo reciente consideró la presencia de ctDNA si la concentración de cfDNA superior a 7 ng/mL o si existían mutaciones detectables cuando el valor era inferior, reportando una sensibilidad del 72,4% y especificidad del 80,6% en muestras a 3-6 meses de la operación (173). Además de su capacidad de detectar la recaída, el ctDNA se ha descrito en CCR como marcador de respuesta a la quimioterapia, observándose disminución como respuesta al tratamiento (226), y como marcador de mal pronóstico y de recaída tras la quimioterapia (225,226). En concreto, un estudio en CCR en estadio III asoció la positividad tras la quimioterapia con una mayor probabilidad de recurrencia, observándose ese

aumento en el riesgo independientemente de que fueran negativos o positivos para ctDNA tras la resección del tumor (227). Aunque muchos de los resultados descritos en la literatura muestran sensibilidades variables entre el 50-80%, la mayoría presenta buenos valores predictivos positivos, es decir, que un alto porcentaje de los pacientes ctDNA positivo, tras el postoperatorio o en el seguimiento, acaba recayendo (192,228).

1.4.9. Ensayos clínicos de biopsia líquida en CCR

La posibilidad de acceso a ADN del tumor sin recurrir a la biopsia del tumor posibilita que más pacientes puedan acceder a tratamientos dirigidos y a ensayos clínicos a través de la biopsia líquida. Este beneficio ha sido descrito por ensayos como TARGET, donde 41 de 100 pacientes presentaron mutaciones accionables al seleccionar un VAF superior a 2,5% en plasma, y 11 de ellos recibieron tratamiento (229). En una cohorte pan-cáncer de cerca de 40 tipos tumorales, la secuenciación del plasma identificó 63 marcadores distintos asociados a mutaciones potencialmente accionables, de acuerdo con la base de datos OncoKB, en el 41,2% de un total de 8032 pacientes (195). Otro estudio, centrado solo en pacientes de colon, identificó mutaciones accionables en un 81,8% de las recaídas con ctDNA positivo (142).

Respecto a la traslación a la clínica, diferentes ensayos clínicos pretenden determinar qué utilidad presenta la detección del ctDNA para la práctica habitual. En CCR, existen más de 100 ensayos clínicos en la base de datos www.clinicaltrials.gov que responden a la búsqueda de los términos cfDNA/ctDNA en CCR y cáncer de colon. De éstos, la plataforma CIRCULATE-Japan explora el cfDNA mediante el uso de *Signatera* en 3 aspectos de la clínica: monitorización de CCR operable (GALAXY), comparativa de la operación con o sin adyuvancia en pacientes estadio II de alto riesgo o III de bajo riesgo en ctDNA

negativos tras la operación (VEGA) y beneficio del tratamiento en pacientes que positivicen en un intervalo de 2 años tras la operación (ALTAIR) (230). De forma similar, en el ensayo clínico Pegasus se pretende estratificar a los pacientes según su ctDNA tras la operación y tras el tratamiento y comprobar el beneficio del uso de la biopsia líquida para guiar los tratamientos (231). A menor escala, los ensayos DYNAMIC II y III quieren ver el efecto de intensificar el tratamiento de pacientes estadio II y III, respectivamente, que testan positivo para ctDNA tras la cirugía (232), mientras que el ensayo de Tie *et al.* refleja la utilidad del ctDNA tras la quimioterapia para monitorizar la respuesta al tratamiento (227). A nivel de enfermedad metastásica, el ensayo COLOMATE usa biopsia líquida para decidir la terapia dirigida que puedan recibir pacientes con CCRm tratados previamente con quimioterapia (233). A nivel español, se ha concedido, en nuestro grupo de investigación, la financiación para el ensayo clínico independiente CIRCULATE-Spain-01 (EudraCT: 2021-000507-20) por parte del Instituto de Salud Pública Carlos III. El objetivo de este ensayo es establecer si los pacientes con CCR localizado y ctDNA positivos, presentan un mayor beneficio al ser tratados como pacientes metastásicos. En dicho ensayo, se usará el panel diseñado en una tesis del grupo y de esta misma tesis para identificar las variantes somáticas de cada paciente y se seguirán mediante la técnica de amplicones desarrollada en el presente trabajo.

2. HIPÓTESIS Y OBJETIVOS

A partir de lo expuesto anteriormente, nuestra **hipótesis** es que la biopsia líquida es una herramienta con utilidad clínica en cáncer de colon debido a que: 1) la valoración de variantes somáticas en plasma va a ser una herramienta con valor predictivo y capacidad de evaluar los seguimientos en cáncer colorrectal y, 2) la evaluación del ADN libre circulante procedente del plasma nos va a permitir identificar la heterogeneidad tumoral en el momento del diagnóstico y la detección de la dinámica de las variantes tumorales, identificando posibles variantes de resistencia al tratamiento.

A partir de la hipótesis expuesta, se establecen los siguientes **objetivos generales** con objetivos concretos que permitirán evaluar su consecución:

- Identificar mutaciones somáticas en los tumores primarios que fueran susceptibles de seguimiento.
 - Desarrollar una estrategia que permita la eliminación de falsos positivos.
 - Determinar si existen genes que puedan ser introducidos en el panel disponible mediante la evaluación de los plasmas de las recaídas.
- Desarrollar una herramienta de secuenciación masiva que permitiera evaluar los seguimientos y detectar mutaciones de resistencia en estos.
 - Desarrollar una herramienta de secuenciación masiva que sea capaz de alcanzar una sensibilidad del 0,1% en muestras de ADN circulante procedente de plasma.
 - Desarrollar un panel apto para la monitorización de variantes en seguimientos para la detección de enfermedad mínima residual.

- Establecer el potencial de la biopsia líquida como herramienta en la recurrencia e identificación de la heterogeneidad tumoral frente a las técnicas establecidas:
 - Identificar a través de la secuenciación del plasma al diagnóstico posibles variantes no identificadas en el tejido primario.
 - Establecer la capacidad de la biopsia líquida para identificar la enfermedad mínima residual y anticipar la recaída.

3. METODOLOGÍA

3.1. Reclutamiento de pacientes y obtención de muestras

Los pacientes incluidos en el proyecto fueron reclutados en el servicio de Oncología del Hospital Clínico Universitario de Valencia (HCUV, España) entre octubre de 2015 y octubre de 2017. Los criterios de inclusión fueron: ser mayor de 18 años, diagnóstico de la enfermedad como CCR sin evidencia radiológica de recaída, y ausencia de cualquier tipo de malignidad en los 5 años previos. Todos los procedimientos descritos en esta tesis han sido aprobados por el Comité Ético del HCUV y todos los pacientes concedieron su consentimiento informado.

En este periodo se reclutaron un total de 150 pacientes de los que se obtuvieron las siguientes muestras: sangre del momento previo a la cirugía para la obtención de plasma y células blancas; tumor de la cirugía inicial, el cual fue parafinado; sangre considerada como postoperatoria a las 6-8 semanas de la cirugía y cada 4 meses a partir de la cirugía para obtención de plasma y tumor de la recaída en al menos 4 pacientes que presentaron metástasis. El número de muestras obtenidas en cada paciente se indica en la figura 6.

De estos pacientes se dispuso de datos como: características demográficas, parámetros clínico-patológicos, fecha de recaída y fecha de los seguimientos, así como de los tratamientos posteriores a la cirugía.

3.2. Procesamiento de las muestras

3.2.1. Extracción de ADN germinal

El ADN germinal (ADNg) fue obtenido a partir de extracciones de las muestras de sangre de los pacientes en el momento de la operación. De esta extracción, 1 ml de sangre o de células mononucleares de sangre periférica resuspendidas en PBS 1X fueron extraídas mediante el equipo automatizado Chemagen (Perkin-Elmer), usando el kit Chemagic DNA Blood 1 ml (Perkin-Elmer) siguiendo las indicaciones

del fabricante. La muestra de ADN, eluída en agua o Tris-HCl 10 mM pH 8.0, fue almacenada y gestionada por Biobanco de INCLIVA a -80°C.

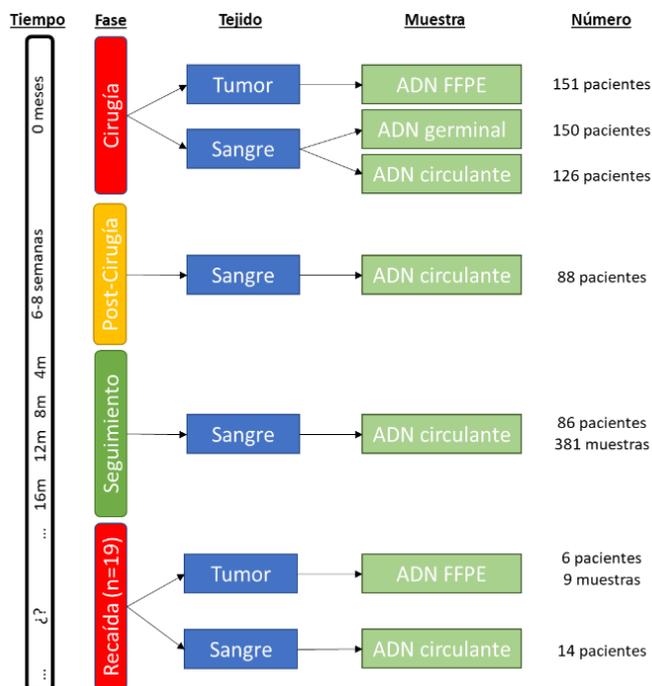


Figura 6: Esquema representando las muestras disponibles según el momento temporal y de la enfermedad.

3.2.2. Extracción de ADN tumoral

Para la obtención de ADN procedente del tumor, se empleó el tejido tumoral previamente parafinado en el servicio de Anatomía Patológica del HCUV. Se utilizaron de 5 cortes de 10 µm del bloque de parafina, los cuales se procesaron mediante el Kit *AllPrep® DNA/RNA FFPE* (QIAGEN) para la extracción conjunta de ADN y ARN. Tras un primer paso de desparafinado, las muestras se trataron con proteinasa K obteniendo una fase líquida, sobrenadante, de la que se extrajo ARN y una fase sólida, precipitado, con el que siguió el protocolo para la extracción de ADN. Tras un segundo tratamiento del pellet con proteinasa K, se purificó el ADN

en las columnas *QIAamp MinElute*. Tras lavar la membrana y dejar evaporar el etanol, el ADN fue eluído en 50 µl de agua libre de nucleasas.

Las muestras de ADN procedente de tejido FFPE se conservaron a -20°C hasta su posterior uso. Por su parte, el ARN obtenido del protocolo se almacenó a -80°C para otros trabajos del grupo.

3.2.3. Extracción de ADN libre circulante en plasma

Las muestras de sangre obtenidas se recogieron en tubos Vacutainer EDTA *Blood Collection Tubes* (BCT). Se realizaron dos centrifugaciones a 1600 g durante 20 min para separar el plasma y las células blancas, este proceso se realizó en una ventana de tiempo inferior a dos horas desde la extracción de la sangre para evitar la degradación del cfDNA. El plasma resultante fue almacenado a -80°C hasta el momento de su uso.

Las muestras de ADN circulante se extrajeron mediante el kit *QIAamp® Circulating Nucleic Acid* (QIAGEN). Para la extracción se utilizó 4 ml de plasma en tandas de 12. En caso de no alcanzar el volumen, se añadió PBS para alcanzar 1, 2, 3 o 4 ml. El plasma se trató con proteinasa K, un *carrier* de ARN y un tampón de lisis. Finalizado el tratamiento, la muestra se hizo pasar a través de la membrana de purificación de las columnas *QIAamp Mini* mediante una bomba de vacío. A continuación, se realizaron una serie de lavados con tampones basados en etanol y finalmente las muestras se eluyeron con agua libre de nucleasas en un volumen de 70 µl o de 50 µl en caso de plasmas para exoma.

Las muestras de cfDNA se almacenaron a -20°C hasta su posterior uso.

3.3. Muestras control

Para asegurar el correcto funcionamiento de las técnicas descritas, se emplearon controles comerciales representando todos los tipos de muestras que fueron empleados en cada panel. Para tejido y sangre se usaron las líneas celulares NA12877 y NA12878 (Coriell Institute), permitiendo evaluar múltiples variantes a diferentes frecuencias.

Como controles del ADN circulante, se emplearon los estándares HD780 (Tris-HCl 10 mM, EDTA 1 mM, pH 8.0) y HD816 (conservado en plasma sintético y extraído con el kit *QIAamp® Circulating Nucleic Acid*), ambos de la casa comercial Horizon Discovery. Ambas muestras contienen variaciones en los genes *KRAS*, *NRAS*, *PIK3CA* y *EGFR*, incluyendo inserciones y deleciones, al 0,0%, 0,1%, 1,0% y 5,0% (tabla 5). Si bien estas muestras no presentan el perfil característico de una muestra de plasma, proceden de líneas celulares fragmentadas para obtener un pico similar al de una muestra de ADN circulante.

Tabla 5: Características de los controles HD780 y HD816. A) Características de las muestras incluidas en cada control y su codificación. El código en las columnas HD780 y HD816 codifica la muestra específica del kit. B) Características de las mutaciones compartidas entre los estándares. La clasificación de alta y baja corresponde con las frecuencias 6,30%, 1,00% y 0,13% por un lado y 5,00%, 1,00% y 1,00% por otro, respectivamente. CDS: *Coding DNA Sequence*.

A)				B)				
Frec.	HD780	HD816	Frec.	Gen	Frec.	Variante	CDS	
5,00%	HD777	HD812	5,00%	<i>EGFR</i> (NM_005228)	Baja	p.E746_A750	c.2235_2249del	
			6,30%			p.V769-D770ins ASV	c.2300_2308dup	
			1,00%			p.T790M	c.2369C>T	
0,10%	HD779	HD814	0,10%		Alta	p.L858R	c.2573T>G	
			0,13%				p.A59T	c.175G>A
0,00%	HD776	HD815	0,00%		<i>NRAS</i> (NM_002524)	Alta	p.Q61K	c.181C>A
			0,00%	<i>KRAS</i> (NM_033360)			p.G12D	c.35G>A
			0,00%				<i>PIK3CA</i> (NM_006218)	p.E545K

3.4. Cuantificación y valoración de ácidos nucleicos

3.4.1. Evaluación de integridad y calidad de las muestras

Con el fin de determinar la integridad y el tamaño de las muestras de ADN, se utilizó el equipo TapeStation 4200 (Agilent) para evaluarlas. Los tipos de cartuchos empleados fueron: *Genomic DNA* para ADN nuclear, *Cell-free DNA* para cfDNA, y *High Sensitivity D1000* y *High Sensitivity D5000* para diferentes tipos de librerías. Los resultados obtenidos fueron analizados mediante el programa *TapeStation Analysis Software* (Agilent) y revisados manualmente. Para las muestras de ADN circulante, la concentración de ctDNA se calculó como el producto de la concentración de cfDNA por el promedio de frecuencia de las variantes somáticas.

3.4.2. Cuantificación mediante fluorescencia

Para conocer la concentración de las muestras, se utilizó el kit *Quantifluor dsDNA* (Promega). Basado en la fluorescencia emitida por muestras marcadas con un intercalante fluorescente del ADN, se partió de un estándar de ADN de concentración conocida (*Lambda DNA Standard*, 100 ng/ μ l) incluido en el kit. De dicho estándar se realizaron diluciones seriadas a 1:2 para obtener una curva patrón (Tabla 6). Usando una dilución 1:200 en TE 1X de QuantiFluor concentrado, se añadió un volumen en proporción 1:1 al pocillo de la muestra. La lectura de la emisión de fluorescencia se realizó en el equipo Glomax Discovery (Promega), mediante el uso de un filtro de excitación a 490 nm y emisión a 510-570 nm (azul). La concentración de las muestras fue finalmente determinada mediante la ecuación de regresión lineal obtenida a partir de los valores de fluorescencia de la curva patrón y aplicando el factor de corrección de la dilución de la muestra.

Tabla 6: Creación de la curva patrón para la cuantificación de ADN de doble cadena mediante el kit *Quantifluor dsDNA*. Para el punto “Std” , el volumen procede del estándar propio del kit (100 ng/ μ l).

Punto	Concentración (ng/ μ l)	Volumen del punto anterior (μ l)	Volumen TE 1X (μ l)	Dilución
Std	4.000	10	240	1:25
A	2.000	110	110	1:2
B	1.000	110	110	1:2
C	0.500	110	110	1:2
D	0.250	110	100	1:2
E	0.125	110	110	1:2
F	0.062	110	110	1:2
G	0.031	110	110	1:2
H	0.000	0	110	-

3.5. Diseño de cebadores

Para diseñar los cebadores necesarios para los amplicones seleccionados se usaron secuencias de ADN_g procedentes de la base de datos pública Ensembl (<https://www.ensembl.org/index.html>). A partir de esta información, se delimitaron las regiones de interés y se identificaron los polimorfismos que podían afectar al diseño de los cebadores. Las secuencias seleccionadas se evaluaron con *Primer3* (<https://primer3.ut.ee/>, v.4.1.0) empleando las opciones por defecto junto a las siguientes especificaciones: longitud entre 25 y 100 pb; diferencia de temperatura de *melting* (T_m) de 2°C entre las parejas de cebadores, rango de T_m de 58°C a 64°C, poli(X) inferior a 5 y un máximo de 3 GC en los 5 nucleótidos del extremo 3'.

Los cebadores se evaluaron mediante la herramienta web *BLAST* (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), aceptando como válidos aquellos con al menos dos órdenes de magnitud de diferencia en el parámetro *Expected Value* entre la región esperada y otra región inespecífica al comparar con el genoma humano de referencia y las secuencias de RefSeq, y mediante la herramienta web

UCSC *In-Silico* PCR (<https://genome.ucsc.edu/cgi-bin/hgPcr>) para comprobar la ausencia de interacciones inespecíficas.

Estos cebadores específicos requieren de un segundo paso de amplificación para hacerlos compatibles con los equipos de secuenciación Illumina e incorporar las secuencias identificativas específicas para cada muestra, por lo que es necesario que se les añada unas secuencias adaptadoras en sentido 5'→3'. Estas secuencias adaptadoras se corresponden con los adaptadores Nextera P5 para el cebador en directo (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) y P7 para el cebador reverso (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG).

Finalmente, para comprobar la compatibilidad entre los cebadores en las mezclas individualizadas, se usó el programa de Java *FastPCR* (<http://primerdigital.com/fastpcr/>). Se consideraron incompatibles aquellos cebadores que interaccionaran con una T_m de más de 28°C, con 12 o más nucleótidos consecutivos y con una ΔG inferior a -9 kcal/mol. Todas las secuencias de cebadores específicos y adaptadores están recogidas en el anexo I y anexo II, respectivamente.

3.6. Generación y evaluación de librerías

Con el fin de preparar el material genético para la secuenciación, se realizaron diversos protocolos de generación de librerías, diferenciándose cada uno en cómo son introducidos los adaptadores y cómo se seleccionan las regiones de interés (Fig. 7). En esta tesis, se emplearon 3 métodos: amplicones (Fig. 7A), PCR anclada (Fig. 7B) y captura con sondas (Fig. 7C).

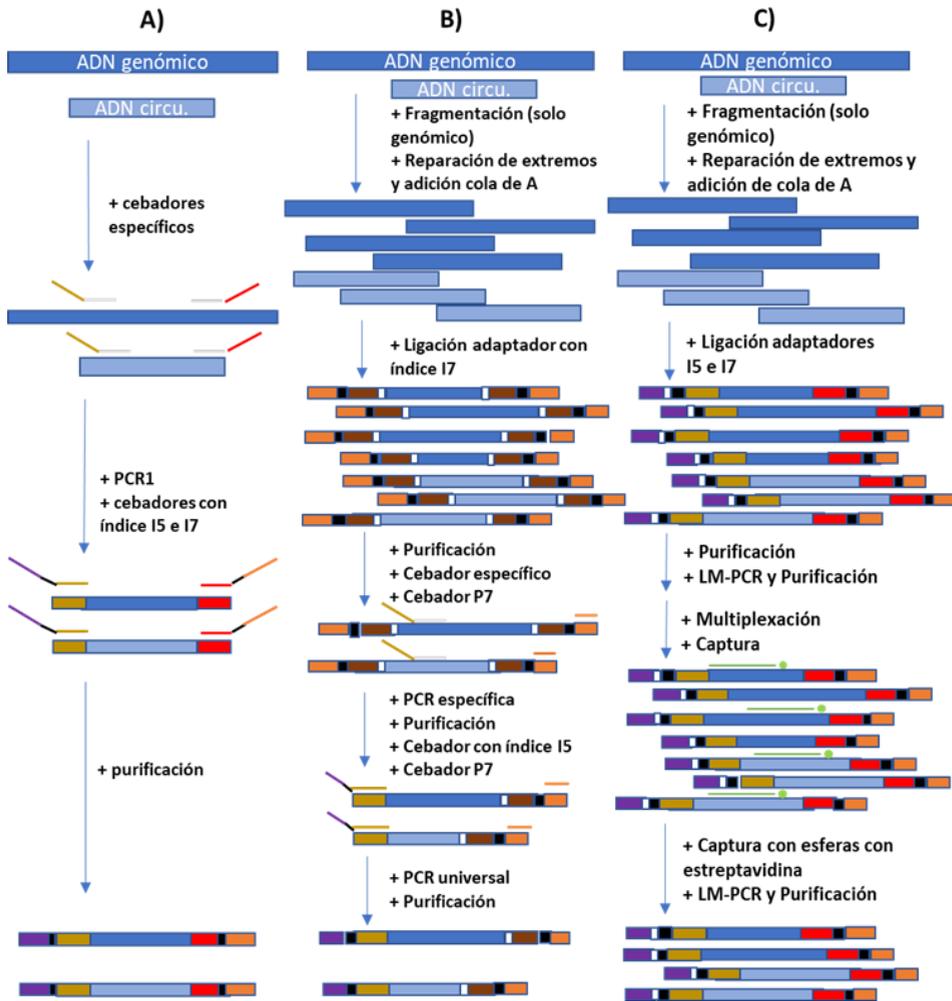


Figura 7: Esquema de los protocolos de generación de librerías empleados. A) Protocolo con amplicones (Nextera); B) Protocolo de PCR anclada de QiaSeq (Nextera); C) Protocolos de captura de KAPA (TruSeq). Por colores: Violeta: adaptador P5; Negro: Índice 5 o 7; Amarillo: Adaptador 5 (TruSeq o Nextera); Rojo: Adaptador 7 (TruSeq o Nextera); Marrón: Adaptador 7 específico de QiaSeq (Nextera); Naranja: Adaptador P7; Blanco: UMI; Verde: Sondas de captura; Gris: Cebador específico.

3.6.1. *QIAseq™ Targeted DNA Panel*

Todas las muestras del proyecto se procesaron mediante el kit *QIAseq™ Targeted DNA Panel* (QIAGEN), empleando un panel customizado diseñado para otra tesis de nuestro grupo de investigación (234). El diseño incluye un total de 29 genes (Tabla 7) que fueron seleccionados por su implicación en el cáncer de colon, búsqueda bibliográfica y la de información en bases de datos oncológicos como MDAnderson o COSMIC. En total, el panel se compone de 1385 amplicones cubriendo 105kb.

Tabla 7: Genes incluidos para el panel personalizado de *QIAseq™ Targeted DNA Panel*

Genes del <i>QIAseq™ Targeted DNA Panel</i>					
<i>ACVR2A</i>	<i>BRAF</i>	<i>FAT4</i>	<i>KRAS</i>	<i>PPP2R1A</i>	<i>SOX9</i>
<i>AKT1</i>	<i>CTNNB1</i>	<i>FLNA</i>	<i>MEK1</i>	<i>PTEN</i>	<i>TCF7L2</i>
<i>AMER1</i>	<i>EGFR</i>	<i>FBXW7</i>	<i>NRAS</i>	<i>RNF43</i>	<i>TGFBR2</i>
<i>APC</i>	<i>ERBB3</i>	<i>HRAS</i>	<i>PIK3CA</i>	<i>SMAD2</i>	<i>TP53</i>
<i>ARID1A</i>	<i>ERBB4</i>	<i>KMT2C</i>	<i>POLE</i>	<i>SMAD4</i>	

El protocolo (Fig. 7b) se inició con la adición mediante ligación de un adaptador que incluye un UMI en el índice 7 y un extremo universal. Esta última secuencia supone la base para una amplificación selectiva mediante un cebador específico contra la región de interés, y un cebador general para el extremo del adaptador añadido. Tras lavar esta primera PCR, el adaptador con el índice 5 se introdujo mediante una PCR gracias a su región complementaria contra la cola del cebador específico y un segundo cebador contra el extremo universal del adaptador 7. Tras purificar, las muestras se recogieron en 20 µl de agua libre de nucleasas. La tabla 8 recoge todas las especificaciones en las que difiere el protocolo en función del tipo de material de partida. Las librerías eluidas en el último paso del protocolo se cuantificaron mediante fluorescencia y evaluación de perfiles con el alineamiento 15-3000 de Qiaxcel y se almacenaron a -20°C hasta su posterior uso.

Tabla 8: Especificaciones del protocolo del *QIAseq™ Targeted DNA Panel* en función del material de partida. La cantidad de ADN circulante es dependiente del volumen máximo de muestra que se ha podido alcanzar, debido a las bajas concentraciones de este tipo de muestras.

	Germinal	Parafina	ADN circulante
ADN de partida	40 ng	100 ng	Variable
Volumen de partida	16,75 µl	16,75 µl	15,50 µl
Incubación a 32°C	24 min	14 min	14 min
H₂O en la ligación	0,0 µl	0,0 µl	2,0 µl
Adaptador IL-N7	2,8 µl	2,8 µl	0,8 µl
H₂O tras ligación	50 µl	50 µl	30 µl
Esferas tras ligación	100 µl	100 µl	112 µl
H₂O tras 1er lavado	50 µl	50 µl	50 µl
Esferas tras 1er lavado	50 µl	50 µl	70 µl
Ciclos de PCR específica	8 ciclos	8 ciclos	8 ciclos
H₂O tras PCR esp.	80 µl	80 µl	70 µl
Esferas tras PCR esp.	100 µl	100 µl	108 µl
Ciclos de PCR universal	18 ciclos	20 ciclos	17 ciclos
H₂O tras PCR U.	80 µl	80 µl	70 µl
Esferas tras PCR U.	80 µl	80 µl	81 µl

3.6.2. *HyperExome con HyperCap Workflow v3.0*

Para la generación de las librerías de exoma se utilizó una química de captura basada en hibridación con sondas biotiniladas. Las librerías se han generado mediante el uso del kit *KAPA HyperPlus* (Roche Diagnostics S.L) con *Illumina Primer Mix* como cebador de la *Ligation-mediated PCR* (LM-PCR), usando *Unique Dual Index* con UMI (UMI-UDI, IDT) e *HyperPure Beads* (Roche Diagnostics S.L) como bolas magnéticas de purificación. Las capturas son realizadas con el kit *HyperExome* (Roche Diagnostics S.L) y las bolas magnéticas con estreptavidina *HyperPure Capture Beads* (Roche Diagnostics S.L). Todo el protocolo se ha realizado siguiendo las recomendaciones del *KAPA HyperCap Workflow v3.0*, añadiendo las modificaciones necesarias para el uso de *Illumina Primer Mix* en lugar de los índices recomendados y cada material de partida (tabla 9). Para este protocolo, se dispuso de 14 grupos de tejido primario, plasma de la recaída y

células blancas de 14 pacientes que recayeron de los 19 de la cohorte. Además, se pudo emplear tejidos de recaída de 4 de ellos.

Tabla 9: Especificaciones del protocolo del *HyperExome* con *HyperCap Workflow v3.0* en función del material de partida. La cantidad de ADN circulante es dependiente del volumen máximo de muestra que se ha podido alcanzar, debido a las bajas concentraciones de este tipo de muestras.

	Geminal	Parafina	ADN circulante
ADN de partida	100 ng	100 ng	Variable
Volumen de partida	35 µl	35 µl	50 µl
Tiempo Fragmentación	25 min	20 min	0 min
Tiempo a 20°C en ERAT	0 min	0 min	30 min
Adaptador IDT	5,0 µl	5,0 µl	0,1-5,0 µl
Temperatura de ligación	20°C	20°C	16°C
Tiempo de ligación	0,33 h	0,33 h	16,00 h
Incubación con <i>HyperPure</i>	5 min	5 min	10 min
Volumen <i>HyperPure</i> tras ligación	88 µl	88 µl	88 µl
Volumen de elución tras ligación	22 µl	22 µl	22 µl
Ciclos de LM-PCR pre-captura	6 ciclos	6 ciclos	11 ciclos
Volumen <i>HyperPure</i> tras 1ª LM-PCR	70 µl	70 µl	50 µl+50 µl
Volumen de elución tras LM-PCR	30 µl	30 µl	50 µl+30 µl
Multiplexado capturas	4 muestras	4 muestras	1 muestra
Volumen <i>HyperPure</i> tras 2ª LM-PCR	70 µl	70 µl	50 µl
Volumen de elución tras 2ª LM-PCR	22 µl	22 µl	22 µl

El protocolo de las muestras de parafina y sangre se inició con una fragmentación enzimática, paso del que carecen las muestras de ADN libre circulante al poseer ya un tamaño reducido. Seguidamente se realizó un paso de *end-repair and A-tailing* (ERAT) con un único momento de incubación a 65°C durante el tiempo indicado en la tabla 9 en función del tipo de muestra. A partir de este punto, todas las muestras se sometieron a la ligación de los adaptadores UDI-UMI, añadiéndose los adaptadores a la concentración adecuada, incrementándose 3 µM por cada 10 ng de ADN circulante añadido, y durante los tiempos indicados en la tabla 9. Tras la ligación, se realizó la purificación con las esferas *HyperPure*, y eluyendo en Tris-HCl 10 mM pH 8.0. Tras recuperar el ADN, se llevó a cabo la LM-PCR con el número de ciclos determinado en la tabla 9 y se purificó con el

volumen de esferas correspondiente. En el caso de librerías de ADN circulante, antes de lavar las esferas con etanol, se añadieron 50 µl de Tris-HCl 10 mM pH 8.0, se incorporó otro volumen 1X de esferas para repetir la captura del ADN. Las librerías pre-captura fueron eluidas en Tris-HCl 10 mM pH 8.0 (Tabla 10) y evaluadas con D1000 HS ScreenTape para determinar su tamaño y concentración antes de proceder con los pasos de captura. En el caso de las de ADN libre circulante se empleó una dilución 1:10 de la muestra.

Para capturar las librerías, se multiplexaron las muestras hasta 1500 ng como se indica en la tabla 9, realizándose una purificación para capturar todos los fragmentos. Estas librerías se purificaron para eluir una solución de captura conteniendo las librerías, reactivos para optimizar la hibridación, y las sondas *HyperExome* contra las regiones de interés y marcadas con biotina continuación, se desnaturalizó durante 5 min y las librerías se dejaron a 55°C durante al menos 16 horas. Antes de proseguir, se lavaron alícuotas de 100 µl las esferas con estreptavidina *HyperPure Capture Beads* con las soluciones correspondientes. A continuación, se añadió cada una de las librerías a una de las alícuotas de esferas y se realizó la LM-PCR post-captura tras los correspondientes lavados. Seguidamente, las librerías se purificaron con el volumen de esferas *HyperPure* correspondientes (Tabla 9) y las librerías capturadas se eluyeron en Tris-HCl 10 mM pH 8.0 (Tabla 9). Las librerías capturadas fueron evaluadas con D5000 HS ScreenTape, empleando una dilución 1:10 de la muestra, con el fin de detectar la posible aparición de heterodúplex en la muestra.

Las librerías fueron almacenadas a -20°C hasta su posterior uso.

3.6.3. Librerías de amplicones

La generación de librerías de amplicones, empleada en muestras de tejido, plasma al diagnóstico, plasma postoperatorio y plasma de individuos sanos constó de dos

pasos de PCR. En un primer paso, se realizó una PCR específica donde se usó un mix compuesto de 7,5 μl de *Master Mix 2X de Multiplex Q* (QIAGEN), 3 μl de pool de cebadores a 0,2 μM , hasta 4 μl de ADN con un total inferior a 400 ng y 0,5 μl de agua. Las muestras se incubaron en un termociclador con el siguiente programa: 15 min a 98°C para activar el enzima, 25 ciclos de 30 seg a 98°C de desnaturalización, 4 min a 60°C de alineamiento y 60 seg a 70°C de extensión, seguidos de una extensión final a 70°C durante 20 min y una incubación a 15°C hasta el momento de la retirada. A continuación, se siguió con la PCR de etiquetado, usando entre 3 y 5 μl de producto de PCR específica, 7,5 μl de *Master Mix 2X de Multiplex Q*, 1,1 μl de un par de índices I5-I7 Nextera específicos para cada muestra y agua libre de nucleasas hasta completar los 15 μl . La solución se incubó en el termociclador con el siguiente programa: 15 min a 98°C para activar el enzima, 38 ciclos de 30 seg a 98°C de desnaturalización, 30 seg a 60°C de alineamiento y 60 seg a 70°C de extensión, seguidos de una extensión final a 70°C durante 10 min y una incubación final a 15°C. Los productos de PCR se evaluaron mediante electroforesis capilar en Qiaxcel para comprobar que existía amplificación y se almacenaron a -20°C hasta su posterior uso.

Para generar el pool de secuenciación, se mezclaron con volúmenes iguales las muestras con diferentes índices. De dicha librería, se purificaron 15 μl con las esferas *MagSi-NGS PREP* (Magnativio) a una proporción de 0.7X. Dicha solución se incubó 5 min a temperatura ambiente y posteriormente se introdujo en un imán para retirar el sobrenadante y realizar 2 lavados con etanol al 70%, añadiendo 20 μl , incubando 5 min y retirando todo el alcohol. Tras los lavados, el etanol remanente se dejó evaporar durante 10 min para posteriormente resuspender en 15 μl de agua libre de nucleasas.

Las librerías se almacenaron a -20°C hasta su posterior carga en el secuenciador.

3.6.4. Evaluación de librerías

Con el fin de determinar el tamaño de las librerías de ADN, tanto en pasos intermedios como al final del protocolo, se empleó el equipo de electroforesis capilar QIAxcel (QIAGEN) con el cartucho *DNA Screening* (QIAGEN). Para ello, se combinaron 8 μ l de agua libre de nucleasas con 2 μ l de la librería en un soporte adecuado. Junto a las muestras, y siguiendo el protocolo de la casa comercial, se cargó el alineamiento pertinente: 15-1000 en librerías de amplicones y 15-3000 en librerías del panel; y se seleccionó el marcador de pesos moleculares previamente validado en el equipo y adecuado para el alineamiento. Los resultados se analizaron mediante el software *QIAxcel ScreenGel* (QIAGEN) y revisaron manualmente. Estas librerías se cuantificaron con el ya descrito *Quantifluor dsDNA*.

En el caso específico de las librerías de exoma, se emplearon los cartuchos de *High Sensitivity D5000* y *High Sensitivity D1000* de TapeStation para evaluación de las librerías intermedias y finales, respectivamente.

3.7. Secuenciación

Para la secuenciación de las librerías, los pools purificados se diluyeron hasta alcanzar una concentración final de 10 nM y se procedió a la carga de los equipos siguiendo las indicaciones de la casa comercial Illumina. En función del número de lecturas necesario, se emplearon hasta cuatro equipos de secuenciación distintos, cuya principal diferencia era la capacidad de lecturas (Tabla 10).

Tabla 10: Características de los equipos de secuenciación empleados. Se incluyen los equipos y los aspectos más importantes sobre su utilización. Además, se añadió el uso de PhiX, secuencias empleadas en el control del porcentaje de errores en los equipos Illumina. MO: *Mid-Output*, HO: *High-Output*.

Equipos	Ciclos	Lecturas	Conc. Carga	Conc. PhiX	Material
Miseq	300+25	22 millones	9-10 pM	5-20%	Amplicones
NextSeq 550	300+18	MO: 130 millones HO: 400 millones	MO: 1.2 pM HO: 1.4 pM	5-30%	Librerías QiaSeq Amplicones
Hiseq X	300+25	1 <i>lane</i> : 312 millones	-	10%	Lib. QiaSeq cfDNA Exomas
NovaSeq 6000	300+25	1 <i>lane</i> : 2000 a 2500 millones	-	10%	Exomas cfDNA

3.7.1. Procesado bioinformático

Los datos generados por los equipos de secuenciación fueron analizados en la unidad de Bioinformática de INCLIVA. El proceso viene recogido en la figura 8. En términos generales, el dato obtenido fue filtrado en base a la calidad mediante el paquete *prinseq*, revisándose ambos grupos de datos con la herramienta web *FastQC*. Mediante el algoritmo *bwa mem* se alinearon las lecturas contra el genoma de referencia hg38 y se evaluaron las métricas y la especificidad mediante *PicardTools* y *BEDTools*. Los archivos BAM generados se procesaron mediante *FGBIOs*, herramienta que también se usó para generar las secuencias consensos de las familias de UMIs. A partir de los BAM, se generaron las estadísticas correspondientes del consenso. Las bases fueron recalibradas mediante el paquete *GATK* y a continuación se realizó el llamado de variantes mediante *LoFreq* y *Mutect2*. Las variantes generadas por ambos se compararon y condensaron mediante *VCftools* y *Haplotype*. Por último, se emplearon los softwares *VEP* y *Snpsift* junto a scripts propios del servicio para la anotación y clasificación de las variantes.



Figura 8: Esquema del procesamiento informático realizado con los datos de secuenciación.

En el caso de las secuenciaciones de las librerías producidas con el panel de QiaSeq, el consenso de UMIs (secuencia final derivada de una familia de UMIs o conjunto de secuencias con el mismo marcador molecular) se realizó también con el software propio de QiaSeq, y tras la recalibración de bases con GATK, el llamado de variantes se consiguió con smCounter, integrándose con las variantes obtenidas por LoFreq y Mutect2. Además, para cada familia de UMIs se estableció un valor de Read per UMI o RPU, reflejando el número de veces que un mismo UMI es leído. La eficacia del algoritmo se determinó mediante el uso del estándar HD780 de Horizon Discovery, con el estándar de cada frecuencia en diferentes cantidades de partida (10 ng, 20 ng, 40 ng y 80 ng).

Las variantes que se obtuvieron de las células blancas se utilizaron para eliminar el componente germinal de las variantes obtenidas del plasma y los tejidos tumorales, dejando las potenciales variantes somáticas (Fig. 9).

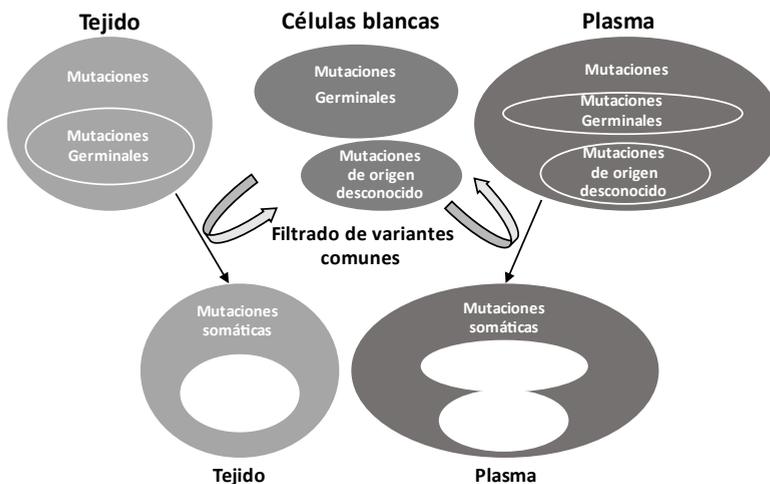


Figura 9: Esquema del proceso de filtrado de variantes de las células blancas. En los tres tipos de muestras, se pueden localizar las variantes germinales del paciente. Además, en la sangre se localizan otras de origen desconocido, como las variantes CHIP. La comparación de las variantes de las células blancas con las totales obtenidas de la secuenciación del tejido y del plasma nos va a permitir filtrar las comunes y obtener las exclusivas y potencialmente somáticas de ambos.

En el análisis de la secuenciación de exomas, para cada paciente se determinó la carga mutacional o *Tumor Mutation Burden* (TMB) considerando solo las variantes no sinónimas con efecto en la proteína (mutaciones sin sentido, pequeñas delinserciones y alteraciones de la pauta de lectura). El cálculo se realizó de acuerdo con lo descrito previamente en la literatura (235) y clasificando como TMB alto aquellos pacientes con un valor superior a 10.

3.7.2. Análisis funcional de las variantes de exoma

Las variantes clasificadas como germinales en exoma fueron empleadas para un análisis de sobrerrepresentación mediante la herramienta *Panther*, mientras que con las variantes somáticas se usó la base de datos *Hallmarks* de la herramienta *Gene Set Enrichment Analysis* (GSEA). El rol de las variantes en la tumorigénesis, expresado como *Driver* o *Passenger* fue establecido mediante el uso de *Cancer Genome Interpreter* (CGI) (236). Por último, la oncogenicidad del tumor fue anotada mediante la base OncoKb, incluyendo anotación sobre aquellas mutaciones potencialmente accionables con los niveles 4 referido a mutaciones con evidencia biológica de ser accionables, 3B para mutaciones que en otros tumores son marcadores de respuesta y R1 para mutaciones de resistencia aprobadas por la FDA (237).

3.7.3. Clasificación de pacientes mediante amplicones

Para determinar si un paciente es ctDNA positivo o no a través de muestras postoperatorias evaluadas con amplicones dirigidos a mutaciones somáticas, se revisaron 4 estrategias para el llamado de variantes: i) datos filtrados del secuenciador, ii) uso de los programas *LoFreq* y *Mutect2*, iii) análisis incorporando *VarScan* para el llamado de variantes y iv) análisis incorporando *VarScan* junto a la clasificación de los amplicones como de “Buena calidad” o “Mala calidad. Para

las variantes llamadas en cada uno de los protocolos, se ha generado una medida del ruido de fondo a partir de muestras de cfDNA de individuos sanos y se ha considerado la positividad en la detección de la variante: i) sin tener en cuenta un control negativo, ii) un punto de corte respecto de la frecuencia de dicho control negativo con un valor de 1,0 vez dicha medida, iii) 1,5 veces y iv) 2,0 veces. Por último, se revisaron estos resultados considerando la positividad del paciente a partir de 1 o 2 variantes positivas. Los criterios de eficacia de cada análisis se basaron en la sensibilidad, especificidad, valor predictivo positivo (*Positive Predictive Value*, PPV), valor predictivo negativo (*Negative Predictive Value*, NPV) y *Hazard Ratio* (HR).

3.7.4. Revisión manual de las variantes

En las muestras de plasma al diagnóstico, se realizó una revisión manual de todas las variantes obtenidas a partir del procesado bioinformático a partir de los archivos BAM utilizando el programa *Integrative Genomics Viewer* (IGV, Broad Institute). En este proceso de revisión, se generó una lista de variantes artefactuales, compuesta por posibles falsos positivos. Los criterios para incluir variantes en la lista fueron: i) Situación en región homopolimérica, con deleciones irregulares, ii) Aparición de la variante en muestras en las que el software no las ha detectado, iii) Localización en solo uno de los sentidos de la lectura (*strand-bias*).

3.8. PCR digital

3.8.1. Elección de las mutaciones para seguimientos

En el análisis inicial presentado en la ya mencionada tesis del grupo (234), se seleccionaron mutaciones clasificadas como patológicas por las bases de datos disponibles en ese momento. Para este trabajo, al disponer de la información de tejido sano pareado, se utilizaron las dos variantes somáticas de mayor frecuencia

alélica para seguimiento mediante PCR digital y las cuatro de mayor frecuencia para seguimiento mediante amplicones, teniendo en cuenta los siguientes criterios de exclusión: filtrado de las variantes clasificadas con impacto *MODIFIER*, filtrado de variantes en regiones UTR o intrónicas (sin incluir regiones de ajustamiento) y mutaciones clasificadas como posibles artefactos por la revisión manual. En ambos casos, el fin es la detección de ctDNA para considerar la presencia o no de EMR. Para esta parte del trabajo, se seleccionaron un total de 68 pacientes.

3.8.2. Diseño de las sondas ddPCR

Para el diseño de las sondas se introdujo la información de la variación a estudiar en la plataforma de búsqueda y diseño de sondas de Bio-Rad (<https://www.bio-rad.com/digital-assays>), escribiendo el gen y la variación proteica en código de una letra de acuerdo con las recomendaciones de la *Human Genome Variation Society* (HGSV) o indicando la variación en el CDS en caso de que la mutación pudiera ser representada por varios cambios. En los casos en los que no hubo ninguna sonda prediseñada para la variante, se procedió al diseño manual mediante la introducción de la secuencia genómica, obtenida de la base de datos ENSEMBL, e indicación del cambio a evaluar en la secuencia.

Se establecieron los siguientes criterios para intercambiar una de las cuatro de las variantes más frecuentes por la siguiente o modificar la mutación de la sonda: i) Imposibilidad de diseño por parte de la plataforma de Bio-Rad, ii) Proximidad de la mutación a una segunda variación de carácter germinal y en la misma cadena, rediseñándose la sonda para abarcar las dos variaciones (Fig. 10A), y iii) No concordancia entre la mutación del BAM y la sonda (Fig. 10B y 10C). En estos casos, se comprobó mediante IGV que coincidieran en la misma cadena las múltiples alteraciones y se anotó manualmente la variante definitiva, en caso de

que se encontraran en el mismo alelo. La figura 10D muestra los casos concretos de estas modificaciones mostradas en las figuras 10A, 10B y 10C. El identificador de las sondas de ddPCR empleadas se detalla en el anexo III.

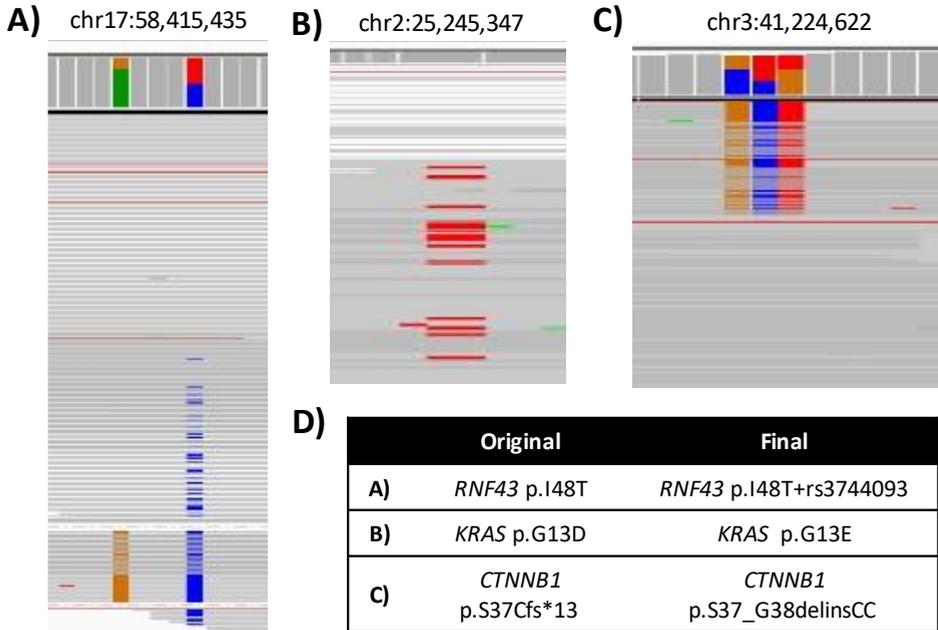


Figura 10: Modificaciones de las sondas para ddPCR de acuerdo con los BAMs. Cada uno de los casos A-C representa una modificación que se realizó en el diseño de las sondas, indicándose en la tabla D la variación representada en cada caso. A) La variante somática *RNF43* p.I48T ocurre en un alelo específico, con proximidad a una variante germinal que puede afectar a la hibridación de la sonda. B) Una doble mutación puntual (*KRAS* p.G13E) identificada inicialmente como una variación de un único nucleótido (*KRAS* p.G13D). C) La variante identificada con mayor frecuencia por los softwares, una mutación puntal, es distinta a la observada en el BAM, en la que ocurre un cambio de un triplete completo en el mismo alelo.

3.8.3. Protocolo de ddPCR

Los ensayos de ddPCR, entendidos como la preparación de placas para su posterior lectura, se diseñaron incluyendo para cada sonda empleada un tejido con la mutación como control positivo y validación de esta, y un pocillo con agua como control negativo.

La composición de cada reacción fue: 10 μ l de Supermix sin dUTP X2 de ddPCR para sondas (Bio-Rad), 2 μ l de sondas FAM/HEX 20X (Bio-Rad) contra la mutación a evaluar, 1 μ l de agua libre de nucleasas y hasta 8 μ l de ADN con muestra, limitada a 20 ng en el caso de tejido FFPE. Una vez finalizada la preparación, la placa termosellada se introdujo en el sistema *Automated Droplet Generator* (Bio-Rad) para la generación de las gotas en emulsión. La emulsión se realizó la PCR usando el siguiente programa: 95°C a 10 min para activar el enzima, 40 ciclos de 30 segundos de desnaturalización a 94°C y 60 seg de alineamiento junto a extensión a 55°C con una rampa de 2°C/seg, 10 min a 98°C para inactivar el enzima e incubación final a 4°C.

Tras la PCR, la placa se leyó en el equipo *QX200 Droplet Digital PCR System* (Bio-Rad), configurado mediante el uso del software *QuantaSoft v1.7*. Finalizada la lectura, los datos se analizaron con el software *QX Manager 1.1 Standard Edition* (Bio-Rad). El criterio de aceptación fue la presencia de pocillos con más de 10000 gotas evaluadas. Mediante la edición manual, se editaron las nubes de puntos para definir las regiones mutadas en el eje del FAM y las de referencia en el HEX, usando el tejido de referencia para definir dichas regiones (Fig. 11A). Las gotas positivas tanto para FAM como para HEX se descartaron (Fig. 11B). Las muestras se consideraron fallidas si el control negativo tenía más de 2 puntos positivos para cualquiera de los fluoróforos o si no se distinguieron las regiones mutadas (HEX) y de referencia (FAM). Para las muestras de plasma se clasificaron como positivas si se detectaban 2 puntos o más con HEX, considerándose la fracción alélica (gotas positivas para HEX entre la suma de gotas positivas en HEX o FAM) como la frecuencia de la mutación en la muestra de plasma.

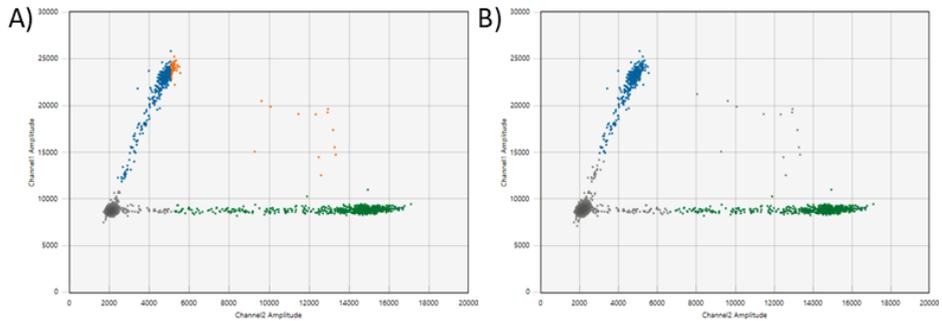


Figura 11: Nube de puntos obtenida de la lectura de un tejido mediante ddPCR para la mutación p.G12D de KRAS. A) Análisis automático del software. B) Edición manual. El eje Y muestra la intensidad de FAM de la gota y el X para el HEX. Cada color representa una evaluación para cada gota: Gris, rechazada; Verde, Alelo de Referencia; Azul, Alelo mutado; Naranja, Alelo mutado y de referencia.

3.8.4. Evaluación de las muestras de plasma de los pacientes
 Un paciente se consideró positivo para una muestra concreta (diagnóstico, posterior a la cirugía, seguimiento) si cualquiera de las mutaciones evaluadas mediante ddPCR fue positiva con los criterios establecidos en el punto anterior. En los seguimientos, se consideró que se detectaba EMR cuando el paciente era positivo para al menos una alteración en uno de los seguimientos.

3.9. Marcaje Molecular Mediante PCR

3.9.1. Diseño de los adaptadores

Con el fin de desarrollar una tecnología de PCR capaz de detectar variantes por debajo del 1% de frecuencia con seguridad, se diseñaron dos adaptadores: un adaptador basado en UMIs separados de la región específica mediante una secuencia protectora de 2-3 nucleótidos (Fig. 12A) y un adaptador basado en un cebador con una horquilla basada en SiMSen-Seq (estructura secundaria generada por auto-apareamiento) que protege el UMI (Fig. 12B), modificando el diseño original de Ståhlberg *et al.* (144,176) con adaptadores TruSeq a

adaptadores Nextera. La formación de la estructura secundaria u horquilla fue evaluada mediante el programa *UNAFold* de IDT (<https://eu.idtdna.com/unafold>) con solo el adaptador y junto a la secuencia de cebador específica del genoma. Ambas propuestas fueron estudiadas en viabilidad mediante comparación de PCR1 de 3 ciclos con las de 25 ciclos. Tras la selección del diseño de horquilla adecuado entre las 5 iniciales (*Hairpin Desing*, HPD1-5), se introdujo un sexto diseño con dos nucleótidos aleatorios extras (HPD6). Todos los diseños están recogidos en el anexo II.

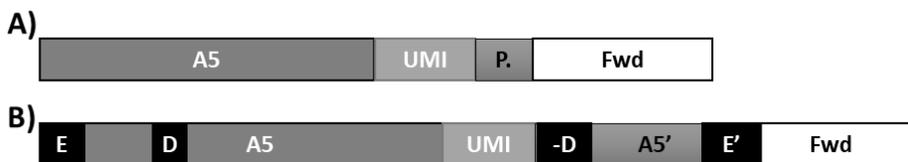


Figura 12: Diseño de las modificaciones de los cebadores para introducir UMIs. Se presenta el diseño para el cebador tanto sin horquilla (A) como con horquilla (B). A5: adaptador en 5'; UMI: marcaje molecular; P.: secuencia espaciadora; Fwd: secuencia específica. E: estabilizador. D: desestabilizador. “-” señala mismas secuencias que no aparean. “'” indica secuencias complementarias.

3.9.2. Modificaciones del protocolo de PCR

Para optimizar el uso de los cebadores con horquilla y UMI, se modificó el protocolo de PCR para amplicones anteriormente descrito:

- El volumen de cebadores a 0,2 μ M se modificó en un rango entre 0,5 y 3,0 μ l y el de agua se aumentó hasta completar los 15 μ l de reacción total.
- El número de ciclos de la primera PCR se redujo a 3.
- Previo a la segunda PCR y para eliminar cebadores, se realizó una purificación 1,10X con AMPure o con *HyperPure Beads* (Roche) con 10 min de incubación, 2 lavados con etanol al 80% y elución en 7 μ l para pasar 5,5 μ l a la siguiente PCR.

- La temperatura de alineamiento se subió a 68°C en la segunda PCR y se realizaron 10 ciclos con 150 seg de alineamiento y 25 con 30 seg. También se utilizaron cebadores con índices con mayor secuencia hibridando contra el adaptador de los productos de PCR1 debido al aumento de temperatura (Anexo II).

Los diferentes pasos mencionados fueron sujetos a prueba y optimización. Los protocolos finales fueron probados en mezclas de los estándares de tejido HD112 (*PIK3CA* p.E545K), HD200 (*KRAS* p.G12D, p.G13D y *PIK3CA* p.E545K) y HD255 (*KRAS* p.G12A) alcanzando diferentes concentraciones, probando varias diluciones seriadas del mix (tabla 11), con una cantidad total de 15 ng y 30 ng de ADN. También fue probado en los mencionados estándares de cfDNA.

Tabla 11: Composición del mix de ADN de tejido para evaluación de los amplicones para marcaje molecular. La tabla incluye la frecuencia final esperada en cada dilución empleada para cada variante y la proporción de material añadida en la preparación de los mixes, con material del mix de la dilución previa y un mix de ADN no mutado. Dil: dilución.

Dil.	Mutaciones				Preparación	
	p.G12A <i>KRAS</i>	p.G12D <i>KRAS</i>	p.G13D <i>KRAS</i>	p.E545K <i>PIK3CA</i>	Mix dil. previa	Mix wt
1:1	3,18%	4,95%	12,38%	11,00%	-	0,00%
1:2	1,59%	2,48%	6,19%	5,50%	50,00%	50,00%
1:4	0,79%	1,24%	3,09%	2,75%	50,00%	50,00%
1:10	0,31%	0,49%	1,22%	1,09%	40,00%	60,00%
1:20	0,16%	0,24%	0,61%	0,54%	50,00%	50,00%

3.9.3. Análisis de las variantes

En las pruebas de sensibilidad del marcaje molecular con amplicones en muestras de cfDNA, se establecieron 2 fases. En la primera, se evaluaron las posiciones específicas en función de la frecuencia obtenida tras el consenso con los UMIs. En la segunda, para evaluar la reducción del ruido de fondo, se probaron tres

estrategias: i) Llamado de variantes utilizando *VarScan* sin UMIs, ii) Llamado de variantes tras el consenso obtenido con los UMIs, iii) Llamado de variantes utilizando *VarScan* en las lecturas consenso generadas con los UMIs. En todos los casos, se estableció un margen mínimo del 0,05% de frecuencia respecto a dicha variante en comparación con el no mutado para considerarla positiva, además de tener que ser detectada en al menos de dos de los triplicados realizados. Cualquier variante por debajo de 0,05% también se consideró descartada.

3.10. Estadística

Los datos generados se analizaron mediante el uso del software Prism 8.0.2 (GraphPad). La descripción general de la cohorte fue realizada mediante parámetros estadísticos descriptivos. Para comparar grupos definidos por variables clínicas cualitativas y/o la positividad para ctDNA se usó el t-test en caso de asumir normalidad y el t-test de Welch en caso de que la varianza fuera distinta entre los dos grupos. La supervivencia libre de recaída o la supervivencia global de los pacientes en base a dos condiciones se representó mediante el uso de la Kaplan-Meier y se analizó mediante el Logrank test. Para comparar la similitud de una misma medida mediante técnicas distintas, se usaron la regresión lineal para tendencia y las representaciones de Bland-Altman para comparar los dos grupos de datos. En esta segunda visualización se ha presentado, para cada medida, el sesgo, reflejado como el porcentaje en la diferencia de las medidas en comparación con el promedio de estas. Por último, la eficacia de los métodos para determinar una variable clínica se evaluó mediante Test de Fisher y la herramienta de sensibilidad, especificidad y valores predictivos de Prism.

4. RESULTADOS

4.1. Pacientes de la cohorte

Entre octubre de 2013 y octubre de 2015 se reclutaron 151 pacientes, de los cuales dos fueron descartados por caso previo de cáncer y por no tener disponibilidad de células blancas, respectivamente (Tabla 12). La media de edad fue de 71,15 años (Rango 41-93), con un 56,38% de los tumores localizados en el colon izquierdo. A nivel histopatológico, 16 pacientes presentaban defectos en el sistema MMR (10,74%) y 55 invasión linfática (36,91%). Un total de 19 pacientes (12,75%) habían recaído en junio de 2021 y 54 pacientes (36,24%) habían recibido algún tipo de terapia adyuvante, siendo 13 de ellos metastásicos. La mediana de seguimiento fue de 36,47 meses (Rango 0,00-53,19). El anexo IV recoge los experimentos realizados en cada paciente.

Tabla 12: Características de los pacientes empleados durante la investigación.

Características		Pacientes (n=149)
Edad		71,15 (Rango 41-93)
Genero	Hombre	94 (63,08%)
	Mujer	55 (36,91%)
Localización	Derecha	65 (43,62%)
	Izquierda	84 (56,37%)
Estadio	I	29 (19,59%)
	II	66 (44,59%)
	III	53 (35,81%)
Estadio II Riesgo	Alto	37 (56,06%)
	Bajo	29 (43,93%)
Invasión linfática	No	94 (63,08%)
	Sí	55 (36,91%)
Invasión perivascolar	No	107 (71,81%)
	Sí	39 (26,17%)
MMR	Proficiente	133 (89,26%)
	Defectivo	16 (10,73%)
Terapia	Capecitabine	32 (21,47%)
	Cap+Ox	21 (14,09%)
	FOLFOX	1 (0,67%)
	Nada	95 (63,75%)
Recaída	I	0 (0,00%)
	II	4 (21,05%)
	III	15 (78,94%)

4.2. Identificación de variantes somáticas en 1ª cohorte
 A partir de la secuenciación pareada de 149 tejidos tumorales y sus correspondientes células blancas se realizó la identificación de las variantes somáticas de cada paciente. La mediana de la cobertura obtenida ha sido de 773X (rango 102-4089) en tejido tumoral y de 848X (rango 293-3770) en células blancas (Fig. 13A). La mediana del porcentaje de bases cubiertas a 100X ha sido 97,71% en tumor y 99,16% en células blancas, y 79,55% y 86,43% para 500X, respectivamente (Fig. 13B y 13C).

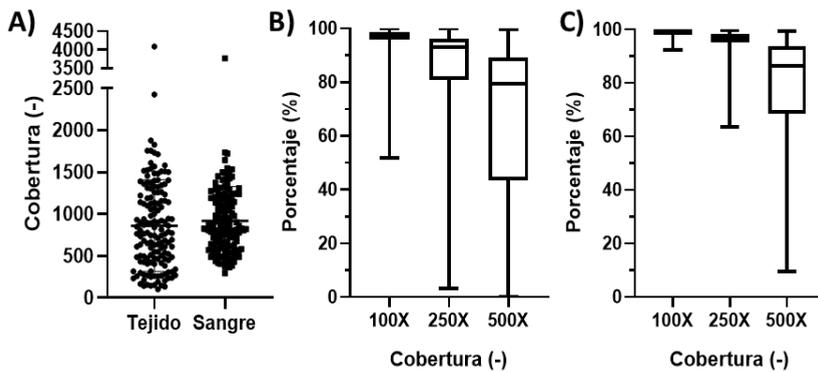


Figura 13: Distribución de los parámetros de secuenciación de los tejidos tumorales y células blancas de los 149 pacientes. A) Distribución de las medianas de la secuenciación de ambos tipos de material genético. B) y C) Distribución de los porcentajes de bases secuenciadas que superan 100X, 250X y 500X para tejido tumoral (B) y células blancas (C).

Junto a los pacientes se incluyeron los diferentes controles indicados en el apartado 3.3. de “Metodología”. La sensibilidad y especificidad de las muestras NA12877 y NA12878, al compararlas con la información genética contenida en las bases de datos, fue del 100,00%. Mientras que para las muestras HD734 y HD827 como controles de tejido tumoral, se obtuvieron como Valor Predictivo Verdadero (PPV) porcentajes del 82,00% y 98,75%, respectivamente.

4.2.1. Variantes germinales

Mediante la secuenciación de las células blancas, se detectaron 8972 variantes, 441 de ellas únicas en los 29 genes del panel que no fueran artefactuales, detectándose una mediana de 61 alteraciones por paciente (rango 30-81). Estas variantes correspondieron en su mayoría a los genes *KMT2C* (16,01%), *FAT4* (11,13%) y *APC* (11,13%). Otros genes con gran número de variantes fueron *POLE* (9,96%), *EGFR* (8,52%), *AKT1* (7,39%) y *TP53* (5,00%) (Fig. 14A).

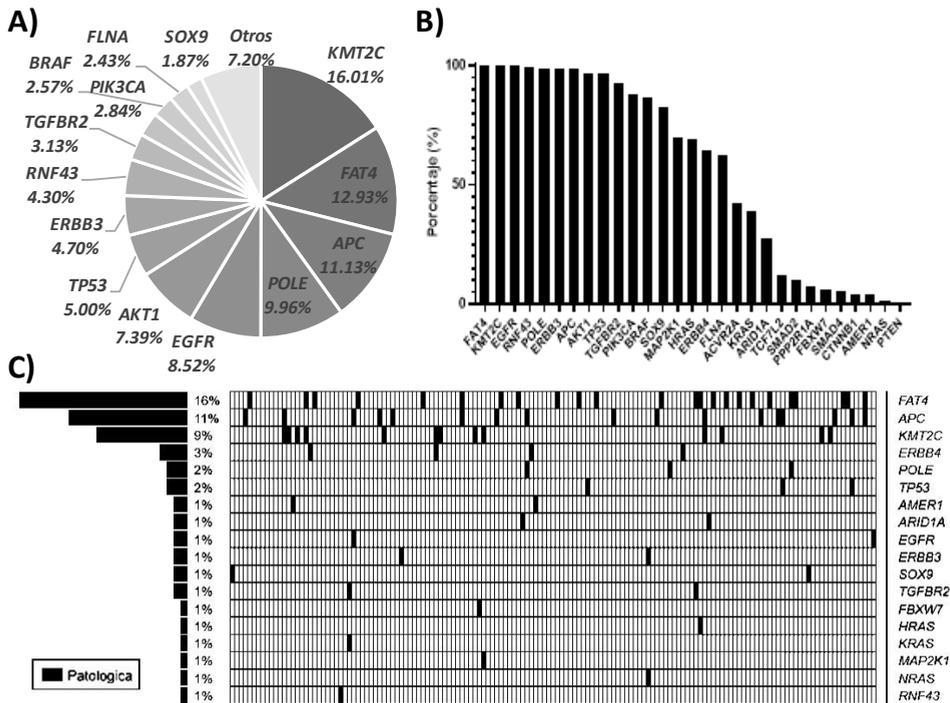


Figura 14: Variaciones germinales de los pacientes de cáncer de colon localizado. A) Porcentaje de las mutaciones asociadas a cada gen. B) Porcentaje de pacientes con una mutación germline en cada uno de los genes del panel. C) Distribución de las mutaciones germinales con posible efecto negativo encontradas.

Se identificaron alteraciones germinales en más del 95,00% de los pacientes en los genes *FAT4*, *KMT2C*, *EGFR* (100,00%), *RNF43* (99,33%), *ERBB3*, *POLE*, *APC* (98,66%), *TP53* y *AKT1* (96,64%) (Fig. 14B). De estas variaciones, el 38,92% se encontraron en regiones no codificantes (intrones, regiones no transcritas...) y el

33,82% se correspondieron con variantes sinónimas no localizadas en puntos de ajustamiento. Únicamente 88 variantes en 66 pacientes (44,30%) fueron anotadas con posibilidad de tener efecto negativo en la proteína, presentando el 16,11% de los pacientes estas mutaciones en *FAT4*, el 11,41% en *APC* y el 8,72% en *KMT2C* (Fig. 14C). En 16 pacientes (10,74%) se encontraron 2 mutaciones concomitantes, estando las alteraciones en al menos uno de los 3 genes más frecuentes para 14 de ellos.

4.2.2. Variantes somáticas

La secuenciación completa del tumor permitió detectar 113 alteraciones (rango 81-291) independientemente de su posible origen. Tras la substracción de las variantes germinales, se identificaron un total de 2313 mutaciones somáticas con un VAF superior al 5% y una profundidad de mínima de 50X a lo largo de los 149 tejidos tumorales (Fig. 15). Cada paciente presentó una mediana de 7 variantes somáticas (Rango 2-210). En total, 289 mutaciones (12,21%) se localizaron en *APC*, 246 (10,63%) en *FLNA* y 233 (10,07%) en *FAT4*. En menor medida, estas mutaciones pertenecieron a otros genes como *TP53* (6,74%), *KMT2C* (6,57%) y *ARID1A* (4,71%). Entre todas las alteraciones identificadas se encontraron 1871 variantes únicas, siendo las más repetidas la alteración p.G12D de *KRAS* en 25 pacientes, p.G12V de *KRAS* en 14 pacientes y p.E545K de *PIK3CA* en 12 pacientes.

1331 alteraciones fueron seleccionadas al estar identificadas como mutaciones con efecto directo sobre la proteína, mediante afectación en algún punto de ajustamiento, por ser variante no sinónima o tratarse de algún tipo de delinscripción. Se obtuvo una mediana de 5 mutaciones por paciente (rango 1-23). El 93,29% de los pacientes presentaban al menos una mutación en *APC* o *TP53*. Por otro lado, el 84,56% de éstos presentaba alguna mutación en *APC*, el 63,76% en *TP53* y el 44,97% en *KRAS*.

Otros genes altamente mutados en la cohorte fueron *FAT4* (31,54%), *ARID1A* (30,20%), *PIK3CA* (28,19%), *SOX9* (22,82%), *FLNA* (22,14%), *SMAD4* (20,81%), *KMT2C* (20,81%), *FBXW7* (20,13%), *AMER1* (16,10%) y *BRAF* (15,76%).

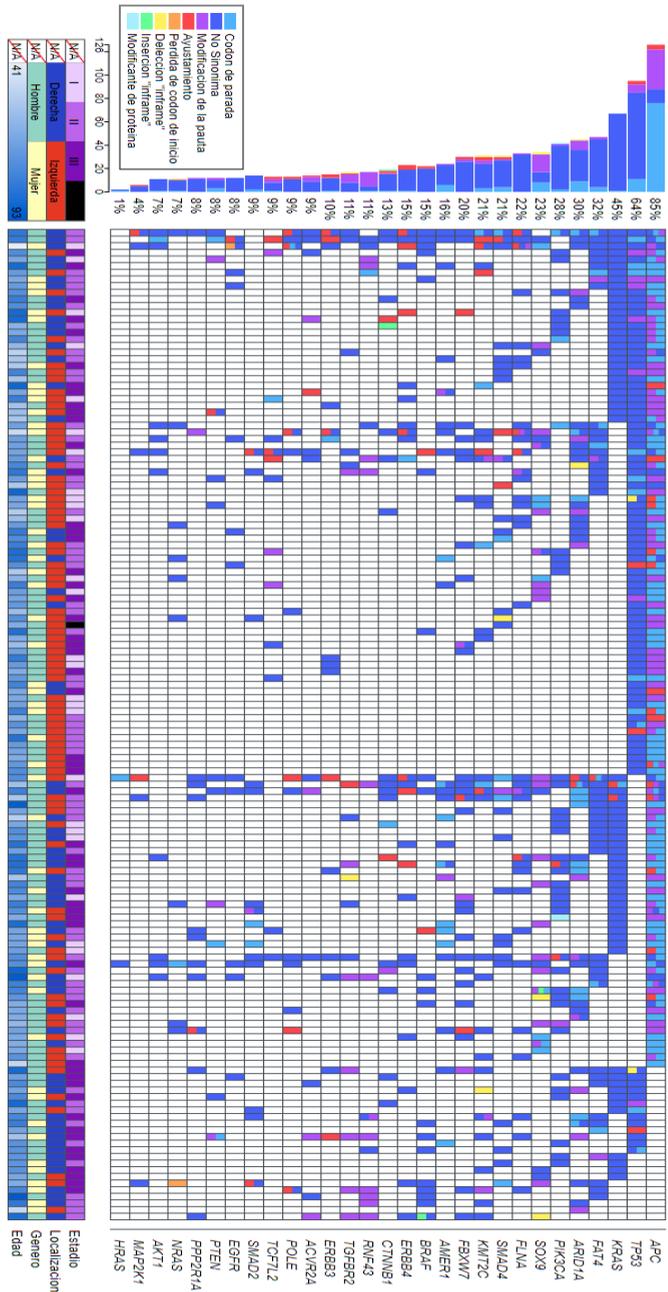


Figura 15: Mutaciones somáticas en regiones relacionadas con la traducción en los 29 genes del panel de CCR. En la figura se representan los diferentes tipos de mutaciones que se identificaron en cada paciente (columnas) y en cada gen (filas), reflejándose hasta tres tipos de alteraciones que sucedan en el mismo gen y paciente. La columna de la izquierda resume el porcentaje de pacientes con alteraciones y el peso de cada tipo de alteración en este porcentaje. Se incluye para cada paciente el estadio, la localización, el género y la edad (escala de color). Elaborado con la herramienta *Landscape* de <https://bio.oviz.org>.

Únicamente el 8,72% de los pacientes presentaron alteraciones en *POLE*, de las cuales solo p.P286R, en el paciente CCR023, corresponde con una de las mutaciones frecuentes del dominio exonucleasa. El grupo de pacientes con *POLE* mutado, independientemente del dominio donde se localizase la alteración, presentó una mayor cantidad de cambios somáticos frente a los no mutados (73,77 [rango 4-210] vs 10,07 [rango 2-76], Welch's t-test p=0,0020). Al definir un subgrupo de pacientes *POLE* mutado o MSI frente a un grupo MSS y *POLE* wt, estas diferencias se redujeron, pero siguieron siendo significativas (41,65 [rango 3 -210] vs 7,94 [rango 2-41], Welch's t-test p=0,0002).

Mientras que en esta selección de mutaciones somáticas predominaron las variaciones no sinónimas, en genes como *APC*, *SOX9* y *RNF43* predominaron las variaciones relacionadas con truncamiento de la proteína, destacando en *APC* la aparición de codones de parada mientras que en los genes *SOX9* y *RNF43* destacaron los truncamientos derivados de la modificación de la pauta de lectura. En los genes *TGFRB2* y *ARID1A* también existió un aumento de las mutaciones sin sentido. A nivel de vías celulares, el 95,97% de los pacientes mostró alguna mutación en la ruta Wnt, mientras que los porcentajes se reducían al 64,00% para p53, al 61,75% en la vía de EGFR, al 38,93% para la de PI3K y al 36,24% para la de TGF- β .

Respecto a las variantes de carácter patológico, se encontraron 339 mutaciones en 142 pacientes, siendo la mediana de 2 variantes patológicas por paciente (rango 0-6). Además de las alteraciones más frecuentes mencionadas anteriormente, se encontraron otras mutaciones patológicas relevantes como la p.V600E, p.D594Y y p.T599delinsKK en *BRAF* y la p.H1057R en *PIK3CA*. Respecto a las mutaciones de *BRAF*, considerando aquellas con efecto significativo en la actividad de la proteína (p.V600E, p.D594Y, p.T599delinsKK), se encontró un incremento significativo de la frecuencia de estas mutaciones en pacientes MSI comparado con los MSS (31,25% vs 3,76%; Test de Fisher, p=0,0014).

4.2.3. Selección de variantes

6 pacientes (4,03%) carecieron de mutaciones clasificadas como patogénicas y 42 (28,19%) solo tenían 1. Al realizar una selección de variantes somáticas en lugar de patogénicas, encontramos que el 100,00% de los pacientes poseía 2 o más cambios somáticos para su seguimiento y 4 o más en el 90,60%. Al elegir las 4 variantes de mayor frecuencia en el tejido, se encontró que el 75,17% de los pacientes tenían entre su selección mutaciones de *APC*, el 62,42% de *TP53* y el 36,24% de *KRAS* (Fig. 16). En menor medida, se encontraron también en *FAT4* (16,78%), *SOX9* (14,77%), *PIK3CA* (14,09%), *ARID1A* (12,08%), *AKT1* (10,07%) y *SMAD4* (10,07%).

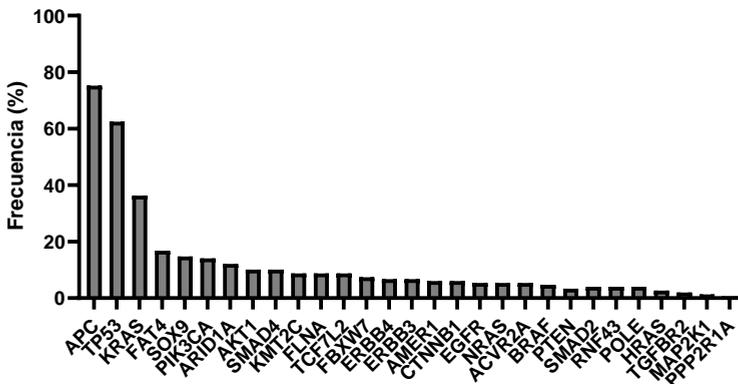


Figura 16: Prevalencia entre los pacientes de genes del panel de CCR con mutaciones entre las 4 variaciones más frecuentes en los pacientes.

4.3. Secuenciación de plasma del diagnóstico

De los 125 plasmas disponibles, 5 fueron descartados por criterios de la calidad del plasma (bajo contenido en cfDNA) o mala cobertura consenso tras la secuenciación. El promedio de lecturas de cada UMI (o *read per UMI* en inglés, RPU) fue de 13,72 (rango 3,22-10,97), con el 93,6% por encima de 6. La mediana de la cobertura consenso para la secuenciación fue 1263X (rango 263-5803, Fig. 17A). De media se cubrió el 99,85% (rango 96,10-100,00%) de las bases a 100X y el 95,38% (rango 3,26-99,79%) a 500X (Fig. 17B). Se observó una tendencia lineal

de aumento de las coberturas consenso en relación con el ADN libre circulante empleado en el protocolo ($R^2=0,3720$, $p<0,0001$, Fig. 17C).

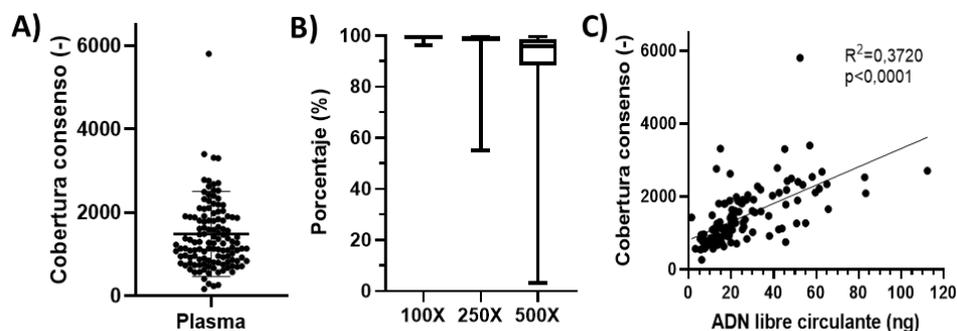


Figura 17: Estadísticas de la secuenciación de los plasmas al diagnóstico de 120 pacientes de cáncer de colon localizado. A) Distribución de las coberturas consenso obtenidas. B) Distribución del porcentaje de bases cubierto a 100X, 250X y 500X. C) Relación entre la cobertura consenso y el ADN circulante empleado.

4.3.1. Variantes somáticas y de heterogeneidad tumoral

En la evaluación del algoritmo para el llamado de las variantes de baja frecuencia se observó una relación lineal entre la mediana de la cobertura y el ADN empleado tras la obtención de las profundidades de lectura derivada del consenso ($R^2=0,9763$, $p<0,0001$, Fig. 18A). La especificidad obtenida para para cada una de las diferentes cantidades empleadas fue del 100%, mientras que la sensibilidad fue variable (Fig. 18B). Con 40 ng de partida todas las variantes a las distintas frecuencias alcanzaron el 100,00% de sensibilidad. Sin embargo, en el caso de las variantes al 0,10%, con 10 ng no se identificó ninguna, mientras que con 20 ng solo se detectó la p.G12D de KRAS y con 40 ng se añadió la p.L858R de EGFR. Finalmente, partiendo de 80 ng se detectaron 7 de las 8 variantes incluidas en el control. En todos los casos en las que no se detectó al menos una variante, como mínimo esta correspondió con una inserción en el exón 20 de EGFR.

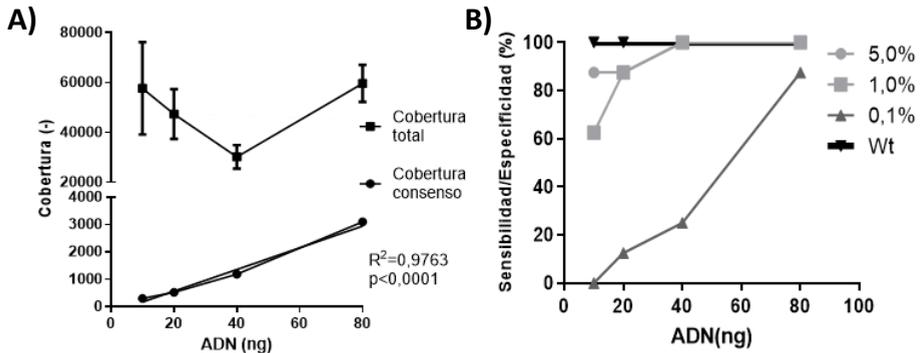


Figura 18: Sensibilidad y especificidad para el panel de CCR con la tecnología de amplicones anclados. A) Cobertura total y cobertura consenso para el estándar HD780 a diferentes cantidades de ADN. B) Sensibilidad y especificidad para las diferentes frecuencias abarcadas por el estándar HD780. Wt: No mutado.

El análisis de las muestras de pacientes dio como resultado 383 mutaciones, siendo 232 mutaciones identificadas por los algoritmos de análisis y 151 detectadas mediante revisión manual en IGV de las variantes somáticas más frecuentes en el tejido, de las cuales 370 eran concordantes entre el tejido y el plasma (Fig. 19). Sobre estas alteraciones, 78 se han detectado en *APC* (20,37%), 55 en *TP53* (14,36%), 27 en *KRAS* (7,05%) y 26 en *FAT4* (6,79%). En el caso de otros genes se identificaron 19 en *ARID1A* (4,96%), 18 en *SOX9* (4,70%), 16 en *FLNA* (4,18%), 12 en *PIK3CA* (3,13%), 11 en *SMAD4*, *TCF7L2* y *FBXW7* (2,87% cada uno) y 10 en *PTEN*, *ACVR2A* y *KMT2C* (2,61%). En total, se detectaron alteraciones en 79 pacientes mediante los algoritmos (65,83%), los cuales ascenderían a 108 teniendo en cuenta la revisión manual. El 51,67% de los pacientes ha presentado mutaciones en *APC*, el 40,83% en *TP53*, el 21,67% en *KRAS*, el 14,17% en *FAT4*, el 14,17% en *ARID1A*, el 13,33% en *SOX9* y el 10,83% en *FLNA*.

Si nos centramos en las mutaciones de interés en CCR, 28 de las 72 variantes de los genes de la familia RAS que se identificaron en los tejidos fueron detectadas en el plasma (38,89%), acumulándose 23 de ellas en los codones G12 y G13 de *KRAS*, 4 en *NRAS* y 1 en el codón Q61 de *HRAS*. En *BRAF*, solo 4 de las 8 mutaciones

p.V600E identificadas en tumor se detectaron en plasma, mientras que de otras 4 alteraciones cercanas a este *hotspot* y con efecto patológico solo se ha identificado como concordante la p.D594G en un paciente. Respecto a la concordancia en *PIK3CA*, se anotaron como coincidentes 3 de las 13 alteraciones p.E545K y 2 de las 6 mutaciones p.H1047R. En total, 8 de las 27 alteraciones de relevancia identificadas en *PIK3CA* en tejido se han encontrado en el plasma. Por último, 46 de las 191 alteraciones somáticas con efecto en la proteína codificada por *TP53* han sido identificadas en ambos tipos de muestras.

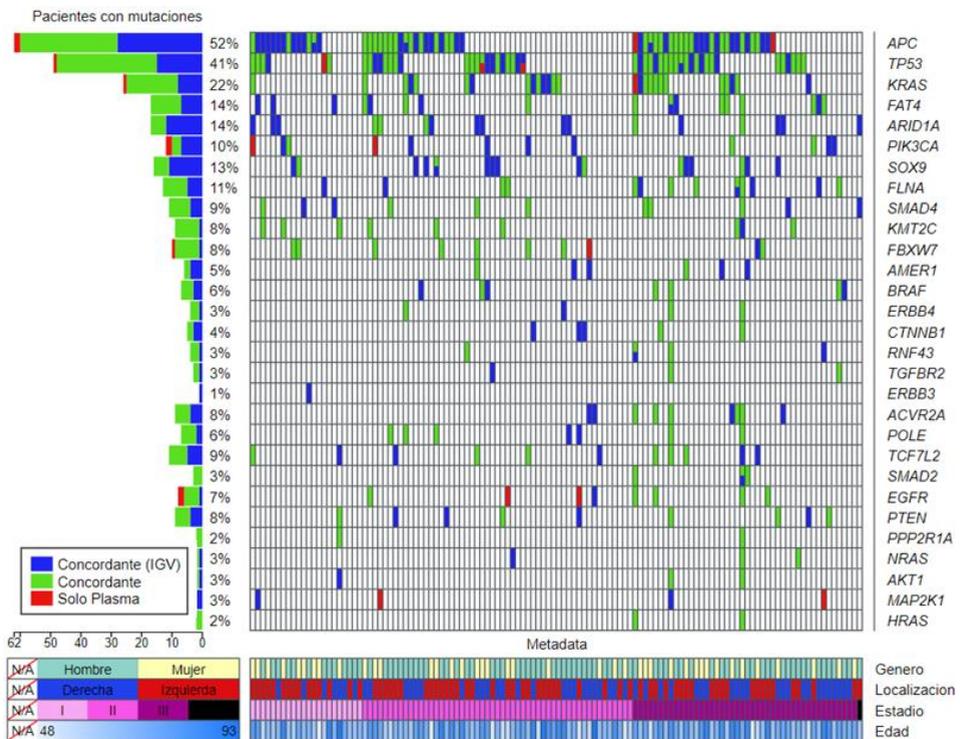


Figura 19: Distribución de las mutaciones concordantes en plasma al diagnóstico en 120 pacientes de CCR localizado. Los porcentajes de la izquierda representan el número de pacientes con mutación concordante en dicho gen. El orden de la derecha mantiene el de la figura. Elaborado con la herramienta *Landscape* de <https://bio.oviz.org>.

En cuanto a las mutaciones exclusivas de plasma, estas se han indicado en la tabla 13. Estas mutaciones 13 mutaciones, que se detectaron en 12 pacientes, parecieron en genes como *TP53*, *EGFR*, *PIK3CA*, *APC*, *MAP2K1* o *FBWX7*.

Tabla 13: Alteraciones encontradas exclusivamente en plasma en la secuenciación del cfDNA del diagnóstico.

Paciente	Gen	Variación	Posición (hg38)	Alteración
CCR003	<i>MAP2K1</i>	p.P124L	chr15:66436825	C>T
CCR015	<i>TP53</i>	p.R175H	chr17:7675088	C>T
CCR026	<i>PIK3CA</i>	p.R93W	chr3:179199102	C>T
CCR062	<i>APC</i>	p.T1556fs*3	chr5:112840254	G>GA
CCR062	<i>KRAS</i>	p.T20M	chr12:25245326	G>A
CCR067	<i>EGFR</i>	p.T790M	chr7:55181378	C>T
CCR076	<i>MAP2K1</i>	p.P124Q	chr15:66436825	C>A
CCR082	<i>FBWX7</i>	p.S582L	chr4:152324294	G>A
CCR084	<i>APC</i>	p.E1286*	chr5:112839450	G>T
CCR110	<i>TP53</i>	p.V157F	chr17:7675143	C>A
CCR111	<i>TP53</i>	p.S241C	chr17:7674241	G>C
CCR113	<i>PIK3CA</i>	p.A1046V	chr3:179234294	C>T
CCR146	<i>EGFR</i>	p.T790M	chr7:55181378	C>T

4.3.2. Asociación del cfDNA y ctDNA al diagnóstico con variables clínicas

Al relacionar el cfDNA y ctDNA con variables clínicas, se obtuvo una relación estadísticamente significativa entre la concentración de ctDNA y el tamaño del tumor (Welch's ANOVA test, $p=0,0041$, Fig. 20A), así como con el estadio del tumor (Welch's ANOVA test, $p=0,0298$, Fig. 20B) y con el CMS (Welch's ANOVA test, $p=0,0043$), observándose una tendencia ascendente entre el ctDNA con el tamaño del tumor y el estadio y definiéndose el grupo CMS3 como el que menos ctDNA presentaba (Fig. 20C). También se encontraron diferencias significativas en la concentración de cfDNA entre los pacientes con invasión linfocelular y los que no la presentaban (t-test, $p=0,0160$, Fig. 20D). No se ha encontrado ninguna relación con otras de las características clínicas de las que se disponen datos.

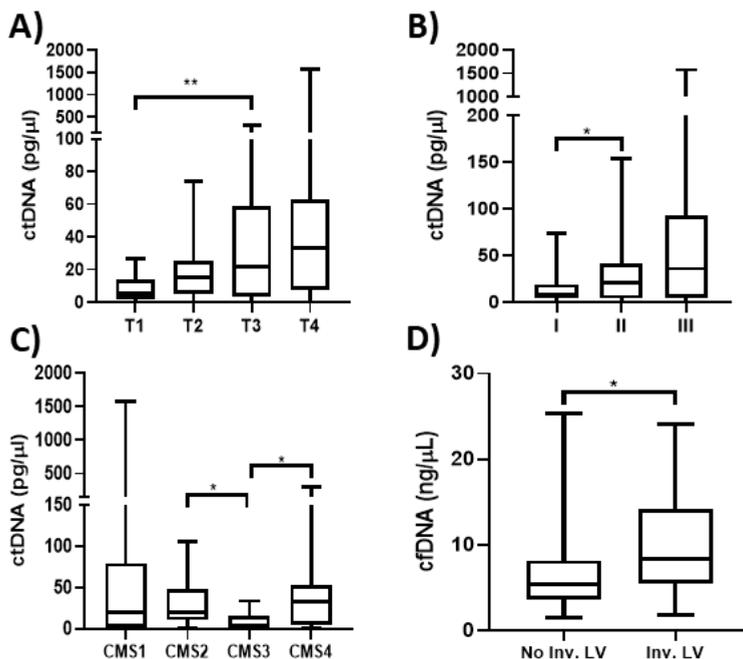


Figura 20: Relación del cfDNA y ctDNA al diagnóstico con las características clínicas de los pacientes de cáncer de colon. En esta figura se separa por grupos la concentración de ctDNA en función del tamaño (A), estadio (B) y CMS (C) y la de cfDNA en función de la invasión linfovascular (D). CMS: *Consensus Molecular Subtypes*. LV: Linfovascular.

4.3.3. Variantes de hematopoyesis clonal

Mediante el algoritmo empleado, se identificaron un total de 33 mutaciones clasificadas como CHIP en 30 de los 120 pacientes analizados (25,00%). 13 de ellas se han localizado en *TP53*, 5 en *KRAS*, 3 en *ERBB3* y 3 en *KMT2C* (Fig. 21A). 17 mutaciones (51,52%) presentan algún efecto sobre la proteína, de las cuales, 7 se corresponden con *TP53* y 5 con *KRAS*. Respecto a la edad, los pacientes con mutaciones CHIP presentaron una mayor edad media que los pacientes que carecían de ellas, aunque dichas diferencias no fueron estadísticamente significativas (71,73 vs 74,41 años, $p=0,1333$, Fig. 21B).

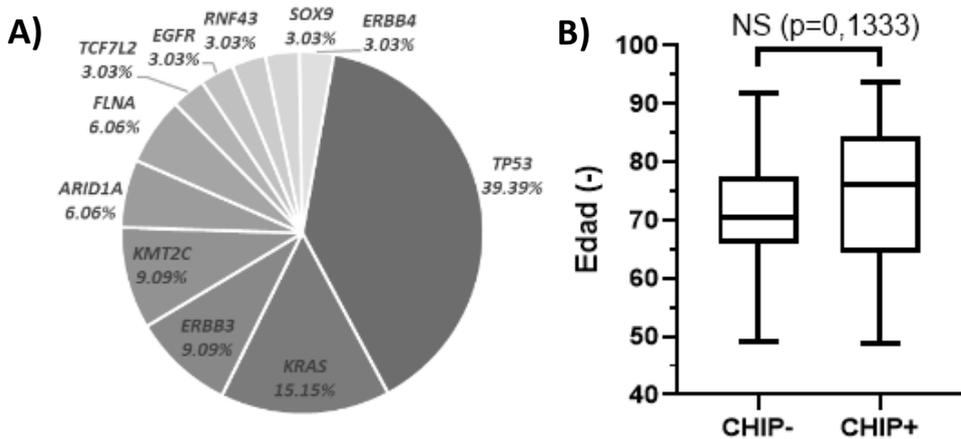


Figura 21: Variantes de hematopoyesis clonal detectadas con el panel de CCR. A) Porcentaje de las mutaciones identificadas en sangre y plasma. B) Distribución de las edades de los pacientes en función de si presentan variantes CHIP o no.

4.4. Detección del tumor en plasma mediante ddPCR

4.4.1. Diseño y validación de las sondas en tejido

Para cubrir todas las 101 variantes seleccionadas entre los 68 pacientes, se requirió de un total de 95 sondas de ddPCR, de las cuales 93 fueron dirigidas a mutaciones únicas y 2 con variantes multiplexadas (codones 12 y 13 de *KRAS* y *NRAS*).

En la validación de las sondas de ddPCR en los propios tejidos, el 96,32% de las mutaciones fueron detectadas adecuadamente, mientras que aquellas que fallaron se debieron a diseños poco eficientes. Esto llevó a descartar dos de los pacientes, reduciendo a 66 el número de pacientes evaluados. Entre las variantes evaluadas, el grado de concordancia fue de $R^2=0,7335$ ($p<0,0001$, Fig. 22A) y el sesgo entre las medidas de ambos métodos fue de 2,51% y una dispersión del 29,45% (Fig. 22B).

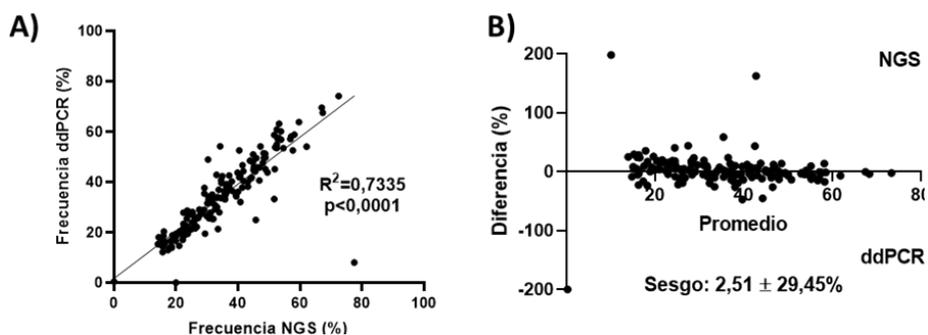


Figura 22: Concordancia entre la PCR digital y la secuenciación del panel en la validación con tejidos. El análisis incluye variantes de la selección somática y patológica. A) Correlación entre las medidas de los VAFs de ambas técnicas. B) Representaciones Bland-Altman, mostrando el porcentaje de variación entre la media de ambos VAF.

4.4.2. Detección del tumor en plasma al diagnóstico por NGS vs ddPCR
 Se evaluó la concordancia con los datos de la secuenciación del plasma al diagnóstico con los datos de ddPCR, revisándose un total de 160 posiciones. El grado de concordancia determinado mediante regresión lineal presentó un coeficiente de correlación $R^2=0,8654$ ($p<0,0001$, Fig. 23A) y un sesgo del $-2,34\%$, con una dispersión del $\pm 156,10\%$ (Fig. 23B). 36 mutaciones (22,50%) no fueron detectadas por ninguna de las técnicas, 31 (19,38%) únicamente detectadas mediante ddPCR y 34 (21,25%) solo mediante NGS.

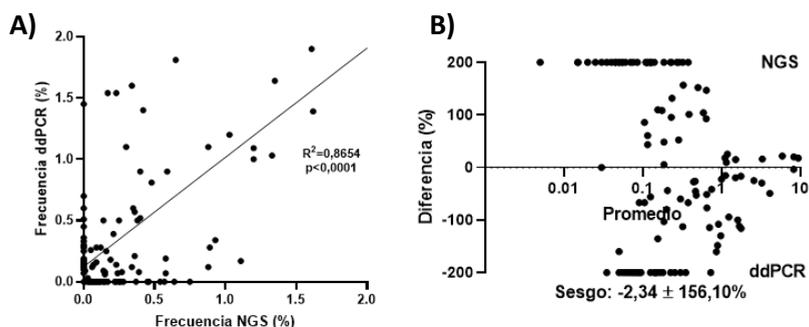


Figura 23: Concordancia entre la PCR digital y la secuenciación del panel en los plasmas al diagnóstico. El análisis incluye variantes de la selección somática y patológica. A) Correlación entre las medidas de los VAFs de ambas técnicas. B) Representaciones Bland-Altman, mostrando el porcentaje de variación entre la media de ambos VAF.

Comparando ambas selecciones, se identificaron un total de 45 mutaciones de la selección somática en 36 pacientes, y 59 variantes en 42 pacientes en el caso de la patológica. La detección de ctDNA al diagnóstico en pacientes de CCR mediante la selección somática no se asoció con la recaída (HR=0,5526, $p=0,2211$, Fig. 24A).

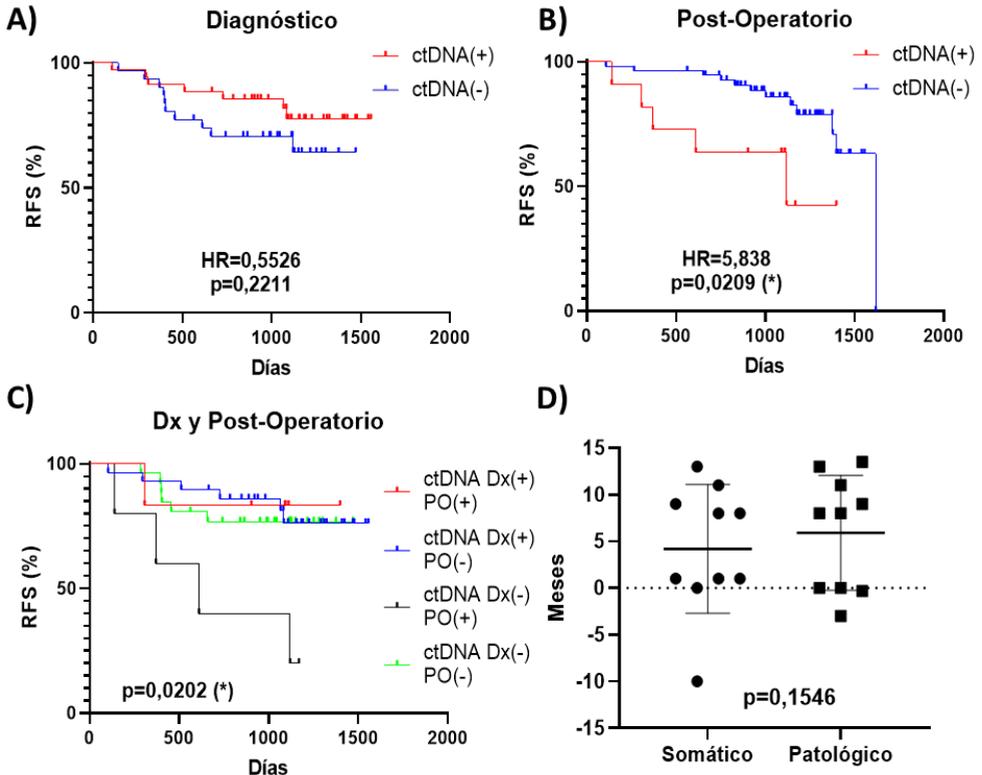


Figura 24: Detección de ctDNA en muestras pre- y post- operatorias de CCR mediante variantes somáticas evaluadas con ddPCR. Curvas Kaplan-Meier para la recaída en función del ctDNA en muestras al diagnóstico (A) y a las 6-8 semanas de la operación (B). C representa la recaída en función de la dinámica del ctDNA antes y después de la operación. D representa la capacidad de la ddPCR para anticipar la recaída en los pacientes ctDNA positivos, comparando la selección somática y patológica. PO: Postoperatorio. Som: Somático. Pat: patológico. RFS: supervivencia libre de recaída.

4.4.3. Detección del tumor en plasma tras la operación

La capacidad de las selecciones somáticas y patológicas para detectar ctDNA con solo dos alteraciones evaluadas y anticipar la recaída en muestras postoperatorias a las 6-8 semanas viene reflejadas en las tablas 14 y 15. La tabla 14 recoge los resultados de ambas selecciones considerando la detección de ctDNA por paciente, observándose un mayor número de positivos mediante la selección patológica. Por otro lado, la tabla 15 muestra los valores diagnósticos de ambas metodologías. En este caso, la selección patológica alcanzó una mejor sensibilidad y PPV, además de que mostró capacidad para discernir a los pacientes ($p=0,005$).

Tabla 14: Concordancia a nivel de paciente en la detección de ctDNA en muestras postoperatorias mediante las selecciones somática y patológica.

	Somático (+)	Somático (-)	Total
Patológico (+)	8 (11,76%)	6 (8,82%)	14 (20,58%)
Patológico (-)	3 (4,41%)	51 (75,00%)	54 (79,42%)
Total	11 (16,17%)	57 (83,83%)	68 (100,0%)

Tabla 15: Valores diagnósticos de las selecciones de variantes somática y patológica en muestras postoperatorias en función del ctDNA. PPV: Valor predictivo verdadero. NPV: Valor predictivo negativo. NS: No significativo.

	Sensibilidad	Especificidad	PPV	NPV	Test de Fisher
Somático	29,41%	87,76%	45,45%	78,18%	NS
Patológico	47,06%	87,76%	57,15%	82,69%	**

En cuanto a la selección somática, se observaron diferencias entre las dos curvas que reflejan la recaída al clasificar a los pacientes según la presencia o no de ctDNA ($HR=5,838$, $p=0,0209$, Fig. 24B). Al observar la dinámica del ctDNA entre el momento pre- y postoperatorio, se distinguió un subgrupo de pacientes, inicialmente negativos, que positivizaban para ctDNA tras la cirugía, el cual presentó una mayor probabilidad de recurrencia ($p=0,0202$). No se apreciaron diferencias entre el resto de los grupos (Fig. 24C).

Al considerar muestras posteriores a las 6-8 semanas en la selección somática solo en los pacientes que recayeron, se detectó ctDNA en 10 de los 16 disponibles (62,50%). Se anticipó la detección de la recaída con un promedio de 4,2 (rango -10,0-13,0) y de 5,9 meses (rango -3,0-13,5) con la evaluación somática y patológica, respectivamente, no existiendo diferencias estadísticamente significativas entre ambas selecciones (Fig. 24D).

La figura 25 recoge ejemplos de dinámicas de las variantes evaluadas, tanto somáticas como patológicas, a lo largo de diferentes muestras. Las figuras 25A y 25B muestran ejemplos de conteos de gotas de un ensayo de ddPCR para una variante en un paciente que permanece negativo tras la operación y otro que permaneció positivo tras la cirugía y cuya frecuencia aumenta en el seguimiento, respectivamente. En el apartado C de esta misma figura se observa como un paciente concreto presentó ctDNA tras la cirugía en las variantes seleccionadas de *PIK3CA* y *APC*, para disminuir tras la terapia adyuvante y reaparecer de cara a la posterior recaída. El caso presentado en el apartado D mostró una tendencia similar, con disminución de todas las variantes seleccionadas tras la quimioterapia y cambio a una tendencia ascendente al finalizarla. Únicamente en dos pacientes se produce el aumento de los niveles de ctDNA durante el tratamiento para luego no volver a detectarse (Fig. 25E). Por último, el paciente mostrado la figura 25F es un ejemplo de aquellos que mostraron variantes con una frecuencia elevada (>1,00%) debido a la alta liberación de ctDNA, llegando a detectarse sus 3 alteraciones por encima del 10% en uno de los seguimientos. El anexo V recoge los gráficos para todos los pacientes no incluidos en la figura 25.

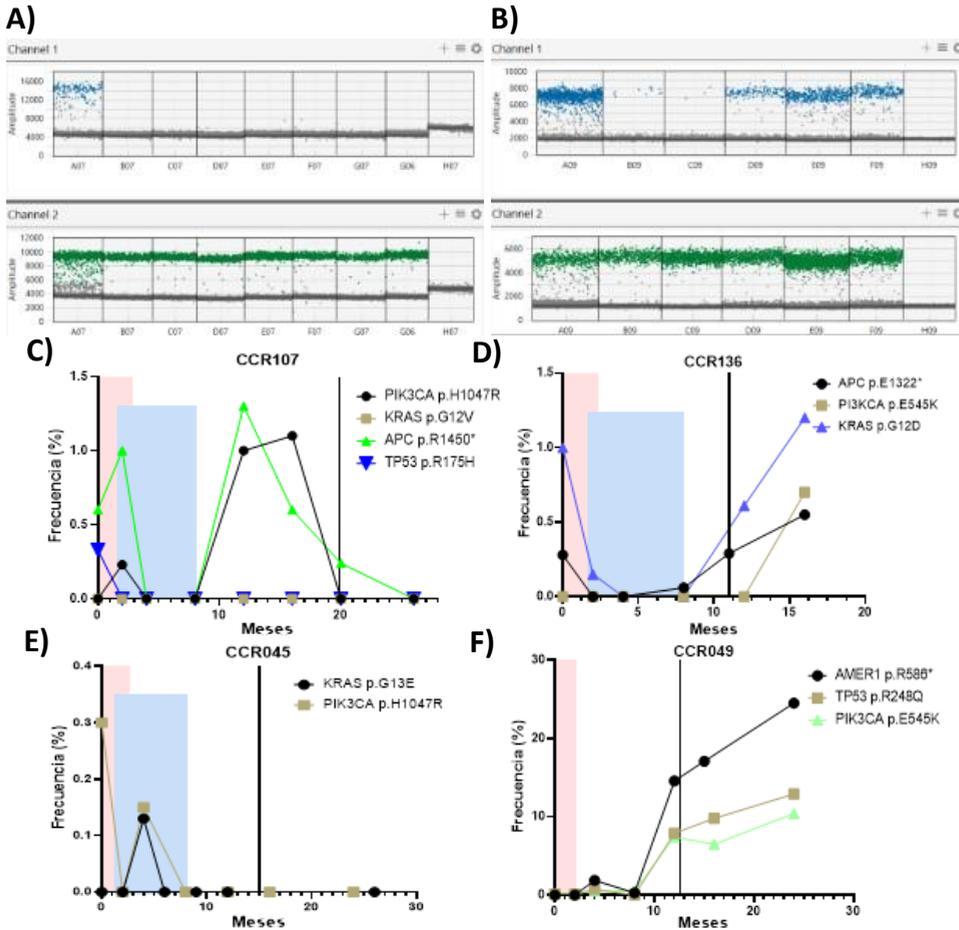


Figura 25: Evaluación de los seguimientos de pacientes de CCR mediante ddPCR. A) y B) visualización de la ddPCR para un paciente con resultado negativo y positivo, respectivamente. Ambos incluyen un pocillo sin muestra al final. C) y D) Representación de pacientes que positivizan tras responder a la quimioterapia. E) Paciente que recae tras la quimioterapia sin positivizar en el ctDNA para las variantes monitorizadas. F) Ejemplo de paciente con variantes que positivizan durante los seguimientos. Área rosa: periodo de 6-8 semanas tras la cirugía. Área azul: periodo de quimioterapia adyuvante.

4.4.4. Enfermedad mínima residual y recaída mediante ddPCR

Al considerar todas las mutaciones evaluadas en función de su VAF, se observó que, en los 48 plasmas al diagnóstico (72,72%) de aquellos pacientes que fueron ctDNA(+), el 68,75% lo fue para su variante más frecuente, mientras que el 22,92%

lo fue con la segunda, detectándose todos los positivos utilizando las 3 variantes más frecuentes. En el caso de las muestras postoperatorias, de los 17 positivos, el 47,06% se identificó con la variante más frecuente, mientras que al emplear una segunda solo se consiguió un incremento del 11,76%. El aumento más notable se notó al incorporar una tercera variante, ya que el 94,12% de positivos tenía alguna mutación detectada.

En cuanto a las muestras previas a la cirugía, la detección de ctDNA no tuvo impacto en la detección de la recaída (HR=0,6548, p=0,3988, Fig. 26A). En cambio, la capacidad de detección observada en las muestras postoperatorias para la EMR se mantuvo, con una mejoría en la sensibilidad (52,94% vs 29,41%) y especificidad similar (84,31% vs 87,76%) en comparación con la selección somática (Fig. 26B). También se observó una mejora en los valores de PPV y NPV, aunque el HR fue menor (HR=4,754, p=0,0004). Al añadir la información del seguimiento a las muestras postoperatorias, se consiguió un aumento en la sensibilidad del 23,53%, alcanzando el 76,46% (Fig. 26C y Tabla 16). Este aumento fue a costa de la especificidad, que se vio reducida al 69,57%. Respecto al resto de valores, el PPV disminuyó al 48,15%, el NPV aumentó al 88,89% y el HR fue ligeramente inferior, de 4,291 (p=0,0032, Fig. 26C).

Por último, al considerar la dinámica de la detección del ctDNA antes de la operación y en cualquier muestra posterior de la operación, se volvió a observar que positivizar tras la operación suponía uno de los peores pronósticos para recaer, mientras que el grupo que se mantenía constantemente negativo fue de pronóstico favorable (p<0,0001, Fig. 26D). En caso de detectar ctDNA en ambos periodos o solo positivos al diagnóstico, las curvas fueron similares, aunque ligeramente inferiores en pacientes con doble positivo (Fig. 26D).

Tabla 16: Valores diagnósticos de la selección global en muestras postoperatorias y seguimientos en función del ctDNA. La figura incluye la diferencia entre solo a las 6-8 semanas y al incluir los seguimientos. PPV: Valor predictivo verdadero. NPV: Valor predictivo negativo. NS: No significativo.

	Sensibilidad	Especificidad	PPV	NPV	Test de Fisher
PO	52,94%	84,31%	52,94%	84,31%	**
PO + Seguimiento	76,47%	69,57%	48,15%	88,89%	**
Diferencia	+23,53%	-14,74%	-4,79%	+4,58%	

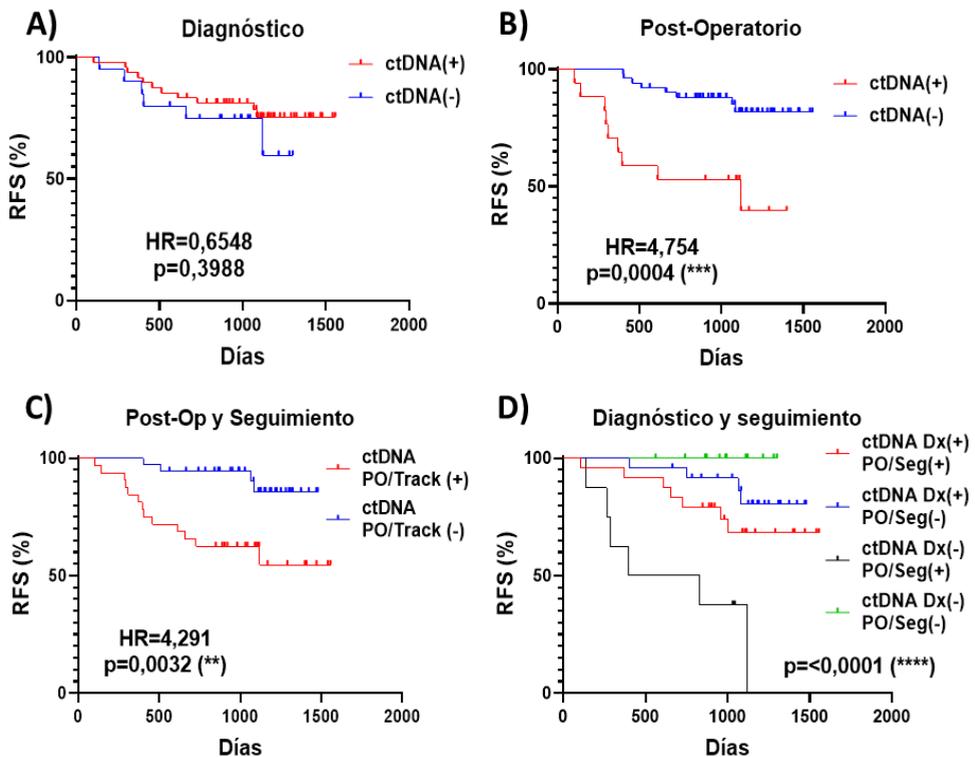


Figura 26: Detección de ctDNA en muestras pre- y post- operatorias de CCR mediante las variantes de la selección global evaluadas con ddPCR. Curvas Kaplan-Meier para la recaída en función del ctDNA en muestras al diagnóstico (A), a las 6-8 semanas de la operación (B) y para cualquier muestra posterior a la operación (C). D) dinámica del ctDNA antes y después de la operación. PO: Postoperatorio. PPV: Valor predictivo verdadero. NPV: Valor predictivo negativo. RFS: supervivencia libre de recaída.

4.5. Secuenciación de los exomas de las recaídas

Los datos de secuenciación están recogidos en la tabla 17 y la figura 28. El RPU en torno a 1 indica que casi todas las lecturas obtenidas corresponden a cadenas únicas.

Tabla 17: Cobertura y variantes obtenidas para cada tipo de material en WES de pacientes con recaída. Mediana [Mínimo-máximo]. RPU: Reads per UMI.

	Tej. Primario	Tej. Recaída	Germinal	Plasma
Cobertura	274X [112-414]	266X [143-311]	158X [53-217]	307X [212-479]
Variantes	110275 [65402-220517]	79125 [49859-291355]	69130 [36192-109262]	36869 [34883-40489]
RPU	1,17 [1,15-1,19]	1,18 [1,17-1,20]	1,14 [1,13-1,17]	1,07 [1,06-1,15]

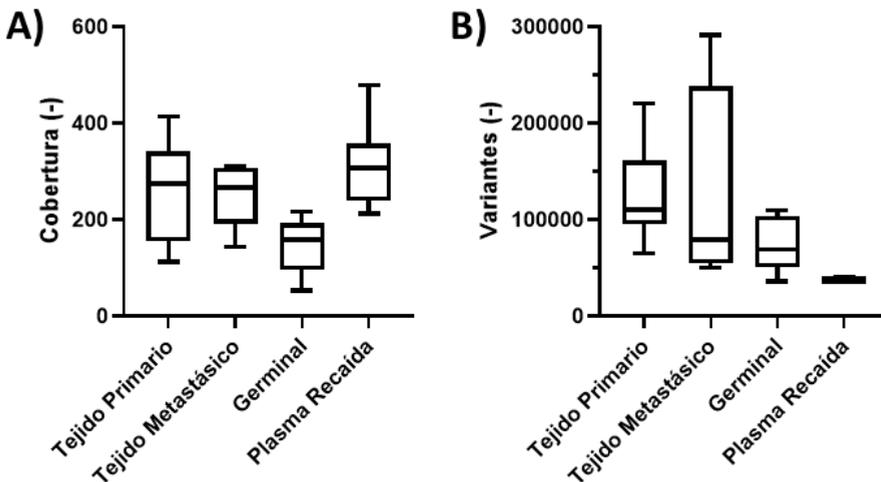


Figura 27: Distribución de los parámetros de secuenciación de las muestras empleadas en WES de las recaídas. A) Distribución de las medianas de la secuenciación según el tipo de material empleado. B) Distribución de las variantes obtenidas según el tipo de material.

4.5.1. Variantes germinales y de hematopoyesis clonal

En los 14 pacientes estudiados, se detectó un promedio de 4,29 mutaciones (rango 2-7) clasificadas mediante CGI como *drivers* conocidos. Entre éstas, se identificaron algunas alteraciones que juegan un papel importante en otros

tumores sólidos como fueron p.R621C en *RB1* en el paciente CCR062, p.S1088F de *FANCA* en CCR063 y CCR150, p.K666N de *SF3B1* en CXR049, y p.Q472H de *KDR* en CCR013, CCR048, CCR062, CCR104, CCR107 y CCR136. Ninguna de estas mutaciones se ha relacionado directamente con CCR. Al centrarnos en las variaciones en genes implicados en CCR, se observó un promedio de 3,92 variaciones (rango 1-6) clasificadas como *drivers* de acuerdo con las predicciones del CGI. Gran parte de estas mutaciones suceden en genes como *ERBB2* y *ARID1A*. Por último, se identificaron mutaciones germinales asociadas a respuesta a fármacos: p.S1088F de *FANCA* (en tumores no CCR) y p.R202C de *RAD54L* se han relacionado con respuesta a Olaparib, y p.P489L en *ERBB2* con respuesta a Trastuzumab, siendo esta última la única mutación potencialmente accionable identificada en la línea germinal.

En el análisis de las células blancas se detectó un promedio de 26,57 alteraciones clasificadas como CHIP (rango 5-67). No se observó una relación entre la edad y la cantidad de alteraciones identificadas ni por correlación lineal ($R^2=0,0004$, $p=0,9462$) ni agrupando por intervalos de 5 años (ANOVA test, $p=0,3161$). Ningún paciente presentó mutaciones en genes clásicos de hematopoyesis clonal (*DNMT3A*, *TET2*, *ASXL1*, *JAK2*, *PPM1D* y *TP53*), mientras que los identificados se asociaron a funciones como a la interacción con el MHC tipo I, la regulación de la hematopoyesis, la regulación de la respuesta defensiva y la señalización mediante citoquinas. Los genes más mutados vienen recogidos en la tabla 18.

Tabla 18: Mutaciones de hematopoyesis clonal identificadas en células blancas mediante secuenciación de exoma.

Gen	Mutaciones	Pacientes
<i>MUC6</i>	55	11 (78,57%)
<i>MUC19</i>	25	7 (50,00%)
<i>LILRB2</i>	24	4 (28,57%)
<i>MUC3A</i>	20	6 (42,86%)
<i>NBPF1</i>	14	9 (64,29%)
<i>MUC12</i>	12	8 (57,14%)

4.5.2. Mutaciones somáticas y concordancia

Los promedios de variantes somáticas obtenidos en cada uno de los tipos de muestra vienen recogidos en la tabla 19. 3 pacientes (21,43%) han mostrado un valor alto de TMB tanto en el tejido primario como en la recaída, mientras que 2 (14,29%) de ellos pasan de un TMB bajo a alto. Ningún paciente ha pasado de un TMB alto a bajo en la metástasis. La tabla 20 muestra la distribución de las mutaciones obtenidas tras el filtrado de las variantes germinales en función de su papel en el tumor y del tipo de material. De acuerdo con el análisis de sobrerepresentación de genes del GSEA, se han identificado mutaciones *Driver* en genes relacionados con la respuesta a estrógenos solo en plasma de pacientes. Estas han sido p.R348W de *RET* en CCR008, p.Y1668N de *NCOR2* en CCR045, p.R406H en *FKBP5* en CCR048 y p.R813W en *RET* y p.R406H en *NCOR2* de CCR062. Estos 4 pacientes han sido tratados con Capecitabine o CAPOX y, salvo el paciente CCR062, que es MSI, no han presentado variantes positivas por ddPCR en el momento del seguimiento. Por otro lado, 3 de nuestros pacientes (21,43%) mostraron un valor alto de TMB tanto en el tejido primario como en la recaída, mientras que 2 (14,29%) de ellos incrementaron su TMB a la recaída. Ningún paciente disminuyó se TMB en la metástasis.

Tabla 19: Promedio de variantes somáticas obtenidas mediante WES. Media [rango].

	Tumor primario	Plasma Recaída	Tejido Recaída
Variantes somáticas	4270,36 [927-8731]	3073,21 [1903-5413]	1303,17 [535-2002]

Tabla 20: Distribución de los diferentes roles de las alteraciones somáticas de las muestras sometidas a WES. Media [rango].

Alteración	Tumor primario	Plasma	Recaída	Tejido	Recaída
Driver	1,93% [1,34-2,64%]	1,37% [0,82-2,01%]	1,79% [1,31-2,15%]		
Passenger	62,32% [58,92-65,83%]	54,20% [46,93-61,21%]	62,04% [56,08-67,09%]		
Neutra	32,50% [29,21-35,97%]	31,26% [27,17-34,76%]	31,62% [26,84-38,69%]		
No clasificada	3,26% [2,24-4,56%]	13,18% [6,00-18,19%]	4,55% [3,19-6,19%]		

En relación con la concordancia entre el tumor primario y las variantes detectadas en la recaída, una media de 425,16 variantes (rango 106-1689) fueron comunes entre ambos materiales, mientras que 3845,21 (rango 794-8475) se encontraron exclusivamente en tejidos primarios y 2648,07 (rango 1756-3724) únicamente en el plasma. El paciente 62, MSI con variantes detectables en la ddPCR por encima del 2,00%, fue el que presentó el mayor número de mutaciones concordantes. El anexo VI recoge los valores de concordancia entre las sondas empleadas por ddPCR y los datos de exoma. No se obtuvo relación estadísticamente significativa entre la concordancia y MSI ($p=0,5517$), detección por ddPCR ($p=0,8373$) o TMB, a pesar de que el promedio de variantes somáticas concordantes fue mayor en pacientes con TMB alto, independientemente de que se excluyeran los individuos MSI (608,70 TMB alto vs 245,80 TMB bajo, $p=0,1691$) o no (748,00 vs 245,80, $p=0,1311$).

El anexo VII recoge la información de la concordancia en las muestras de las que se dispone de al menos un tejido metastásico. Se observó un mayor número de variantes exclusivas en los tejidos primarios en comparación con los de las recaídas. En el caso del paciente CCR107, el cual disponía de dos piezas obtenidas en la misma cirugía, dos tercios de las variantes de cada tumor no fueron coincidentes con las del otro. En el caso de CCR063, la pieza de procedente de la

metástasis de 2019 presentó menos variantes que la de 2018 y una menor concordancia con el plasma, obtenido también en 2018.

Los genes con mayor número de mutaciones exclusivas en plasma fueron *MUC6*, *MUC19*, *MUC16*, *MUC4*, *MUC3A* (100,00%), *GOLGA6L2*, *CTBP2* (92,86%), *IGFN1*, *GXYLT1* (85,71%) y *RRBP1* (78,57%). Teniendo en cuenta aquellos mutados en al menos de la mitad de los pacientes, las principales funciones de estos genes se relacionan con la actividad de la respuesta inmune innata y con la glicosilación de proteínas.

Tras seleccionar solo las variantes catalogadas como *Oncogenic* o *Likely Oncogenic*, se observó que no todos los pacientes presentan variantes concordantes, suponiendo estos un 42,86%. Todos los individuos presentaron nuevas mutaciones patogénicas en el plasma. Estas mutaciones no detectadas en los tejidos primarios se relacionan con la ruta WNT, la actividad de p53, con la actividad transcripcional de E2F y la modulación de la cromatina, aunque dichas funciones se encontraron ya alteradas en los tumores iniciales.

Considerando los 20 genes con más mutaciones oncogénicas (Fig. 28), genes como *ERBB2* o *KMT2C* son los que presentaron un mayor número de mutaciones exclusivas en plasma sin estar mutados en tejido. Otros genes como *CUX1*, *RAD51B* o *ANKRD11* incluyeron también mutaciones oncogénicas exclusivas en 2 pacientes cada uno. Por otro lado, se observó la no detección de mutaciones del tejido primario en el plasma en genes como *APC*, *TP53* o *KRAS*. Entre los genes que presentaron alteraciones oncogénicas y no se encontraban en el panel inicial se incluyen *HLA-B*, *ATXN2*, *ATR* y *DNMT3A*.

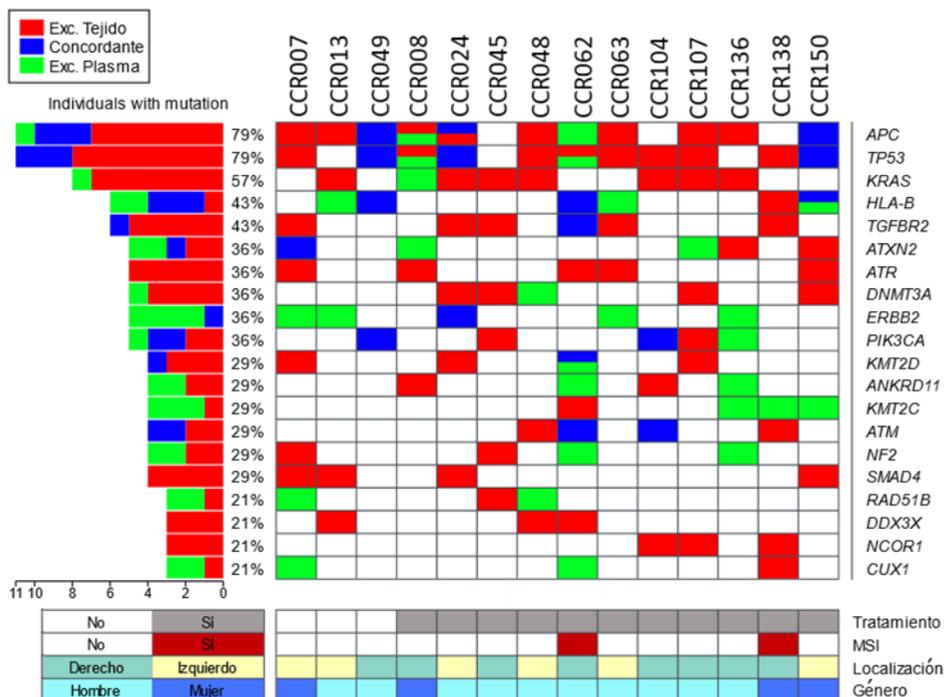


Figura 28: Principales genes con mutaciones oncogénicas detectadas en el WES de pacientes con recaída. El gráfico muestra si esta mutación es exclusiva del tumor (rojo), si se ha originado en la recaída (verde) o se mantiene en el desarrollo de la metástasis (azul). Elaborado con la herramienta *Landscape* de <https://bio.oviz.org>.

4.5.3. Mutaciones adquiridas accionables y de resistencia

En relación con las mutaciones accionables y de resistencia que han aparecido en las recaídas, un total de 10 pacientes (71,43%) presentaron nuevas mutaciones accionables, recogidas en la tabla 21. Solamente uno de los pacientes (CCR008) desarrolló resistencias reconocidas a tratamientos aprobados para colon metastásico, concretamente la mutación fue p.G12D en *KRAS*. De aquellos pacientes de los que se dispuso de tejido de la recaída, solo una mutación en el paciente CCR063 (en sus dos muestras) y otra en el CCR136 se identificaron como concordantes. Ambas mutaciones aparecen subrayadas en la tabla 21.

Tabla 21: Mutaciones accionables de acuerdo con OncoKB desarrolladas en la recaída e identificadas en el plasma. Se incluyen los niveles de evidencia 3B, 4B y de resistencia. Las mutaciones concordantes con tejido metastásico aparecen subrayadas.

Paciente	Gen	Variación	OncoKb	Fármaco
CCR007	<i>RAD51B</i>	p.C118F	3B	Olaparib
	<i>RAD51B</i>	p.M120I	3B	Olaparib
	<i>ERBB2</i>	p.S907C	3B	Trastuzumab
CCR008	<i>RET</i>	p.R348W	3B	Selpercatinib, Pralsetinib
	<i>KRAS</i>	p.G12D	4	Cobimetinib, Binimetinib
	<i>PTCH1</i>	p.Q389*	R1 3B	Cetuximab Sonidegib
CCR013	<i>PALB2</i>	p.R663C	3B	Olaparib
	<i>ERBB2</i>	p.D904V	3B	Trastuzumab
CCR045	<i>RET</i>	p.E505V	3B	Selpercatinib, Pralsetinib
	<i>NF1</i>	c.3198-3C>A	4	Cobimetinib, Trametinib
CCR048	<i>RAD51B</i>	p.C118F	3B	Olaparib
	<i>RAD51B</i>	p.M120I	3B	Olaparib
CCR049	<i>BRCA1</i>	p.Q1811*	3B	Olaparib
CCR062	<i>ARID1A</i>	p.A339Gfs*61	4	Tazemetostat, PLX2853
	<i>MTOR</i>	p.T2446M	3B	Everolimus, Temozolomida
	<i>MTOR</i>	p.W1916*	3B	Everolimus, Temozolomida
	<i>RET</i>	p.R813W	3B	Olaparib
	<i>CHEK1</i>	p.T242Hfs*14	3B	Olaparib
	<i>NF1</i>	p.A188E	4	Trametinib, Cobimetinib
CCR063	<i>FGFR3</i>	p.R223C	3B	Erdafitinib, AZD4547
	<i>TSC2</i>	p.A272T	3B	Crizotinib
	<i>ERBB2</i>	p.Y908F	3B	Olaparib
CCR136	<i>MET</i>	p.P664A	3B	Everolimus
	<i>TEP1</i>	p.W192G	4	GSK2636771, AZD8186
	<i>ERBB2</i>	p.D904V	3B	Trastuzumab
	<i>ARAF</i>	p.M98I	3B	Sorafenib
CCR138	<i>PIK3CA</i>	p.E545K	3B	Fulvestrant, Alpelisib
	<i>TSC2</i>	p.Y587*	3B	Everolimus

4.6. Detección de EMR en plasma mediante amplicones

4.6.1. Diseño y validación de los paneles

Se diseñaron 155 amplicones (anexo II), distribuidos en 67 paneles, cubriendo un total de 267 mutaciones. Cada panel incluía 4 diseños para mutaciones independientes específicos de cada paciente.

Cada panel fue testado en el tejido del paciente correspondiente a modo de control positivo. 3 de los amplicones diseñados no pudieron ser analizados por limitaciones de la técnica (SOX9_p.E388*, ARID1A_p.P293S y PTEN_p.I101L). La mediana de cobertura de los amplicones en tejido fue de 12218X (rango 148-36180). 30 de las variantes analizadas (11,24%) resultaron ser artefactuales y 8 alteraciones aparecieron en porcentajes inferiores a los esperados. Teniendo en cuenta estos resultados, la distribución de variantes seguidas en cada paciente quedó modificada de la siguiente manera: 8 pacientes se siguieron con 2 variantes, 20 con 3 y 39 mantuvieron las 4 alteraciones.

Respecto a los valores de concordancia, las regresiones comparando las determinaciones mediante NGS (n=221) como con el análisis con ddPCR (n=143) dieron un valor de $R^2=0,6513$ y $R^2=0,7073$ respectivamente (Fig. 29A y 29B). En base a los gráficos Bland-Altman, las frecuencias obtenidas son más precisas con los valores de la ddPCR en comparación a los del panel personalizado para NGS (3,46% vs 7,75% de variación respecto al promedio de las medidas) con una dispersión similar (20,19% vs 24,12%, Fig. 29C y 29D).

Al disponer tanto de los datos de ddPCR como de NGS de los plasmas al diagnóstico, se decidió hacer la misma comparación con estas muestras. 48 de los 67 pacientes con plasmas al diagnóstico fueron secuenciados. La mediana de cobertura de los amplicones fue de 131919X (rango 2693-606528), con el 91,30% de los amplicones por encima de 50000X y el 67,70% por encima de 100000X. Los valores de correlación de la ddPCR (n=74) y los datos del panel NGS (n=113) con los amplicones de buena calidad definidos mediante el software *VarScan* se reflejan en las figuras 30A y 30B. El porcentaje de alteraciones con frecuencia mayor que 0,00% en los amplicones y que no fueron identificados ni en la secuenciación ni en la ddPCR es del 48,67% en el caso de la comparación amplicones-panel (Fig. 30C) y 39,28% para la de *VarScan*-ddPCR (Fig. 30D). Por otro lado, se obtuvieron dispersiones superiores al 100,00% en los sesgos de las

representaciones Bland-Altman de ambas comparaciones (Fig. 30C y 30D). El sesgo promedio de los amplicones es del 17,30% respecto al panel dirigido y de 30,14% respecto a la ddPCR, siendo más del doble y más de 10 veces mayor que sus promedios en las medidas en tejido, respectivamente.

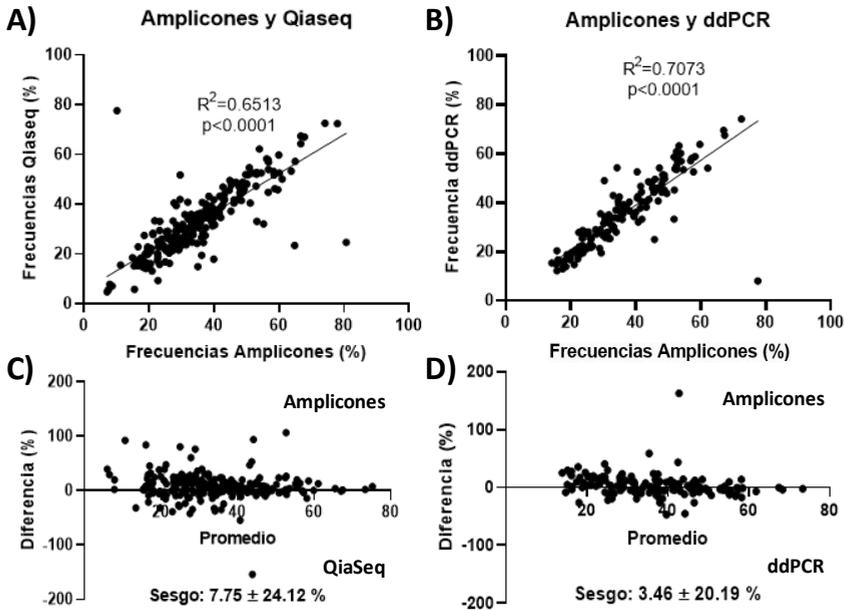


Figura 29: Concordancia de la secuenciación de los amplicones con el panel de personalizado y la ddPCR en muestras de tejido. Las figuras A y C corresponden a la comparativa con el panel y las B y D con la ddPCR. Las figuras A y B reflejan la correlación entre las medidas de los VAFs de ambas técnicas, mientras que la C y D son las representaciones Bland-Altman, mostrando el porcentaje de variación entre la media de ambos VAF.

Respecto a las muestras evaluadas como controles negativos, empleadas para evaluación de ruido de fondo, la mediana de la cobertura obtenida fue de 133018X (rango 179-1442691). El 84,00% de los amplicones mostró una cobertura superior a 50000X y el 66,22% a 100000X.

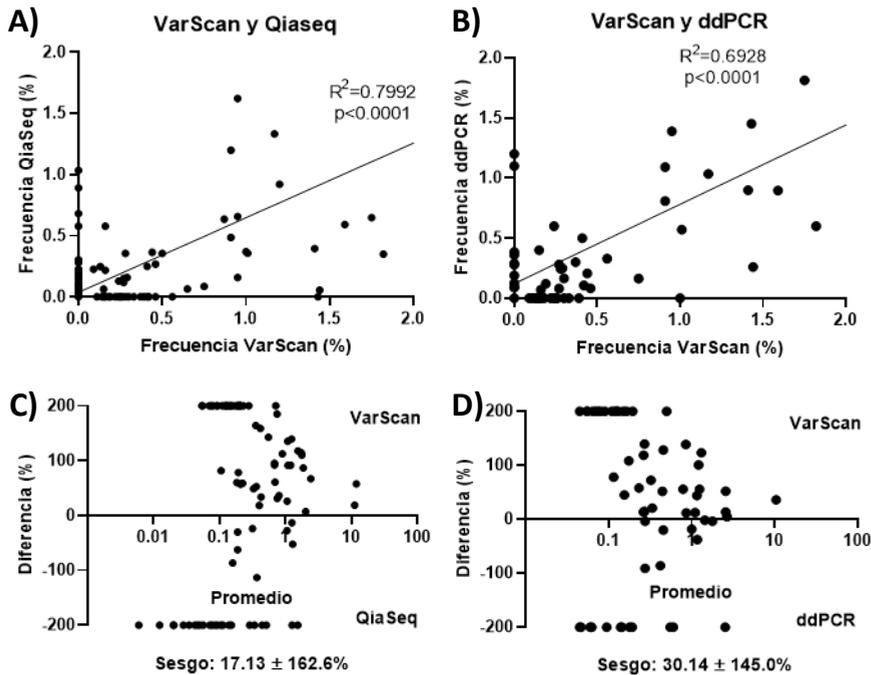


Figura 30: Concordancia de la secuenciación de los amplicones y el panel de personalizado y la ddPCR en muestras de plasma al diagnóstico. Las figuras A y C corresponden a la comparativa con el panel y las B y D con la ddPCR. Las figuras A y B reflejan la correlación entre las medidas de los VAFs de ambas técnicas, mientras que la C y D son las representaciones Bland-Altman, mostrando el porcentaje de variación entre la media de ambos VAF.

4.6.2. Secuenciación en muestras postoperatorias.

La secuenciación de los plasmas obtenidos tras la operación se realizó por duplicado, considerándose el promedio de las frecuencias para los análisis realizados. La mediana de la cobertura obtenida fue 135906X (rango 827-2321765), estando el 87,95% de los amplicones por encima de 50000X y el 65,00% por encima de 100000X. Debido a la falta de cobertura, 2 de los pacientes fueron excluidos de los análisis con datos postoperatorios. En comparación con las medidas obtenidas por ddPCR (n=98), la correlación entre ambas mediciones es

$R^2=0,0208$ con un $p=0,1564$ (Fig. 31A). La representación de Bland-Altman muestra 5 puntos (5,10%) encontrados únicamente mediante ddPCR y 40 variantes (40,82%) mediante el análisis del *VarScan*. Estas discordancias llevan a situar el sesgo de las medidas en un 138,2% (Fig. 31B).

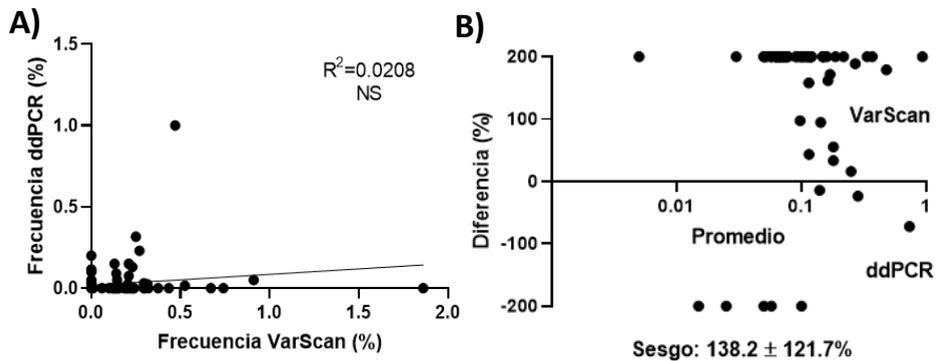


Figura 31: Concordancia entre la secuenciación de los amplicones y la ddPCR en muestras de plasma postoperatorias. A) Correlación entre las medidas de los VAFs de ambas técnicas. B) Representaciones Bland-Altman, mostrando el porcentaje de variación entre la media de ambos VAF.

4.6.3. ctDNA y EMR mediante amplicones

La figura 32 recoge las sensibilidades, especificidades (Fig. 32A) y los HR (Fig. 32B) de los diferentes análisis valorados. Para poder realizar una comparación adecuada, se recalculó la capacidad de la ddPCR de detectar recaídas en muestras a las 6-8 semanas con las dos muestras excluidas. En este nuevo análisis, el HR fue de 5,008 ($p=0,0002$), mientras que la sensibilidad y especificidad lo fueron del 53,94% y 85,42%, respectivamente, y el PPV y NPV de 56,25% y 83,67% ($p=0,0032$).

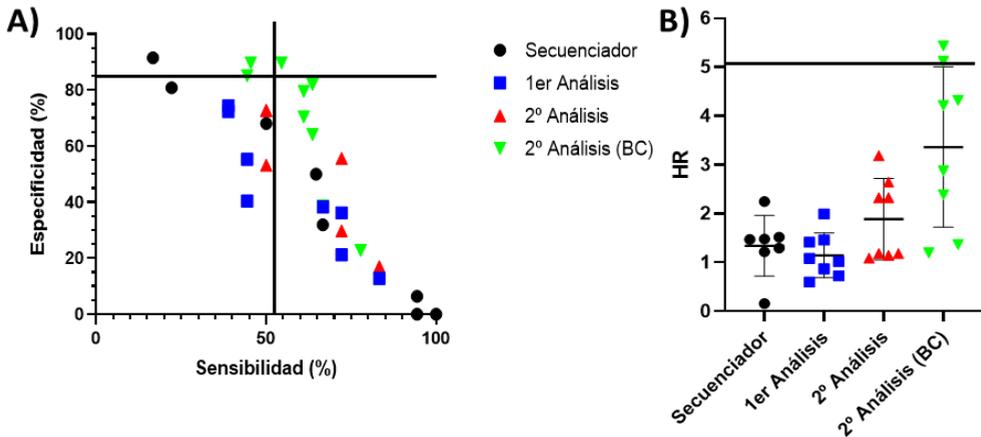


Figura 32: Sensibilidad, especificidad y HR de los diferentes análisis de la secuenciación del plasma tras la operación mediante amplicones para detectar la recaída. A) representa los valores de sensibilidad y especificidad mientras que **B)** refleja los HR, por cada tipo de análisis de los datos del secuenciador.

Considerando los análisis que calificarían la variante como válida si su detección era superior al ruido de fondo y que consideraban al paciente ctDNA positivo con una variante somática detectada (Fig. 33), ninguno de ellos mostró resultados significativos ni en las curvas de supervivencia (Fig. 33A-D) ni a nivel de evaluación de la detección (Tabla 22).

Tabla 22: Valores diagnósticos de los métodos de análisis en amplicones muestras postoperatorias con al menos una variante positiva. PPV: Valor predictivo verdadero. NPV: Valor predictivo negativo. NS: No significativo. BC: Buena calidad

	Sensibilidad	Especificidad	PPV	NPV	Test de Fisher
Secuenciador	94,44%	6,38%	27,87%	75,00%	NS
LoFreq+Mutect2	72,22%	21,28%	26,00%	66,67%	NS
VarScan	72,22%	29,79%	28,26%	73,68%	NS
VarScan (BC)	66,67%	38,64%	30,77%	73,91%	NS
ddPCR	52,94%	85,42%	56,25%	83,67%	**

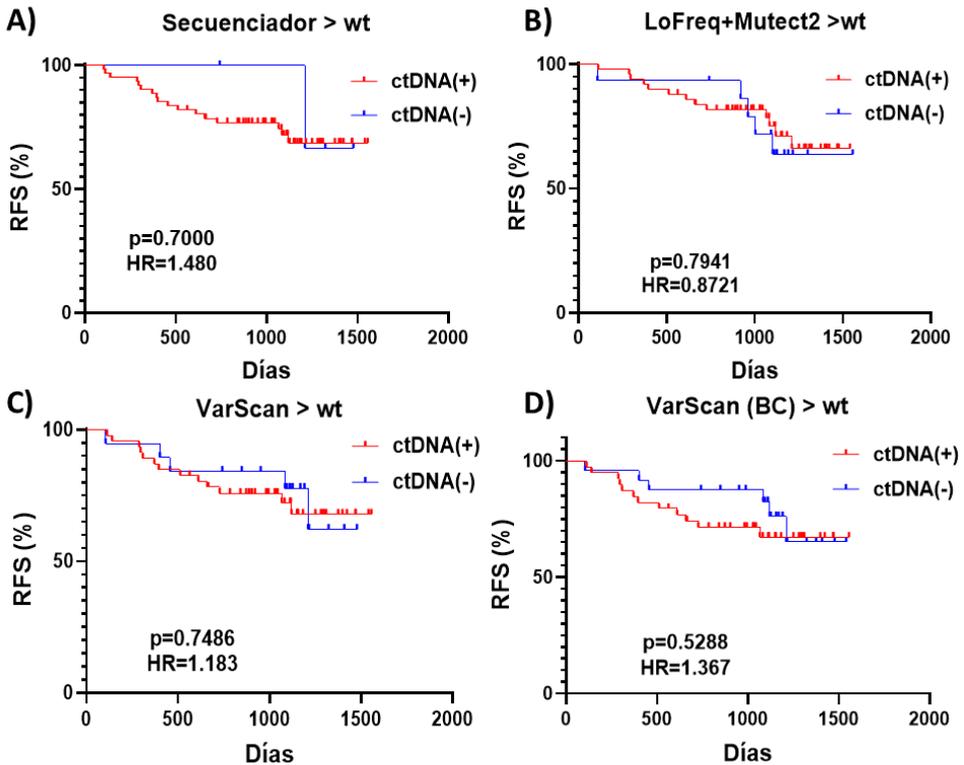


Figura 33: Capacidad de detectar la recaída en muestras posteriores a la operación de los diferentes análisis de la secuenciación de amplicones con al menos una variante positiva. Las gráficas A-D representan las curvas Kaplan-Meier de cada análisis, siendo estos A) los resultados crudos del secuenciador, B) análisis con LoFreq y Mutect2, C) análisis con *VarScan* y D) amplicones de buena calidad según el *VarScan*. PPV: Valor predictivo positivo. NPV: Valor predictivo negativo. RFS: supervivencia libre de recaída. HR: *Hazard Ratio*. BC: Buena calidad.

Cuando en lugar de una variante, se requirió que el paciente fuera positivo con dos, en el mismo contexto (Fig. 34), los datos del secuenciador y para el primer software no produjeron resultados significativos (Fig. 34A y 34B). En cambio, en caso del análisis incluyendo *VarScan*, se produjo un incremento del HR (Fig. 34C y 34D). En el caso de mantener todos los amplicones, los resultados no fueron significativos, aunque se observó una tendencia a generar 2 curvas diferenciadas (p=0,0633, HR=2,330, Fig. 34C). Seleccionando únicamente los amplicones que

pasaron los criterios de calidad, se perdieron 15 pacientes, de los cuales 7 presentaron metástasis, quedando un total de 50 pacientes para el análisis. De estos, recayeron 10 (20,00%). Esta situación fue la única que mostró una curva Kaplan-Meier estadísticamente significativa ($p=0,0037$) capaz de separar ambos grupos (Fig. 34D), con un HR muy similar al de la ddPCR (5,116 vs 5,008). Además, este método demostró una sensibilidad superior a la de la ddPCR, aunque con mayor PPV en la ddPCR y especificidades y NPV similares (tabla 23). A nivel de concordancia entre este análisis y la ddPCR, el 64,00% de los pacientes se clasificaron de forma concordante entre ambas técnicas (tabla 24). En base a esta última clasificación con los resultados de NGS, se relacionó una mayor proporción de positivos en el grupo de pacientes CMS1 y CMS4 en comparación con los CMS2 y CMS3 (OR=8,566, $p=0,0352$).

Tabla 23: Valores diagnósticos de los métodos de análisis en amplicones muestras postoperatorias con al menos dos variantes positivas. PPV: Valor predictivo verdadero. NPV: Valor predictivo negativo. NS: No significativo. BC: Buena calidad.

	Sensibilidad	Especificidad	PPV	NPV	Test de Fisher
Secuenciador	66,67%	31,91%	27,27%	71,43%	NS
LoFreq+Mutect2	44,44%	55,32%	27,59%	72,22%	NS
VarScan	50,00%	72,74%	40,91%	79,07%	NS
VarScan (BC)	63,64%	82,05%	50,00%	88,89%	**
ddPCR	52,94%	85,42%	56,25%	83,67%	**

Tabla 24: Concordancia a nivel de paciente en la detección de ctDNA en muestras postoperatorias mediante el uso de ddPCR y NGS basada en amplicones.

	ddPCR (+)	ddPCR (-)	Total
Amplicones (+)	5 (10,00%)	9 (18,00%)	14 (28,00%)
Amplicones (-)	9 (18,00%)	27 (54,00%)	36 (72,00%)
Total	14 (28,00%)	36 (72,00%)	50 (100,0%)

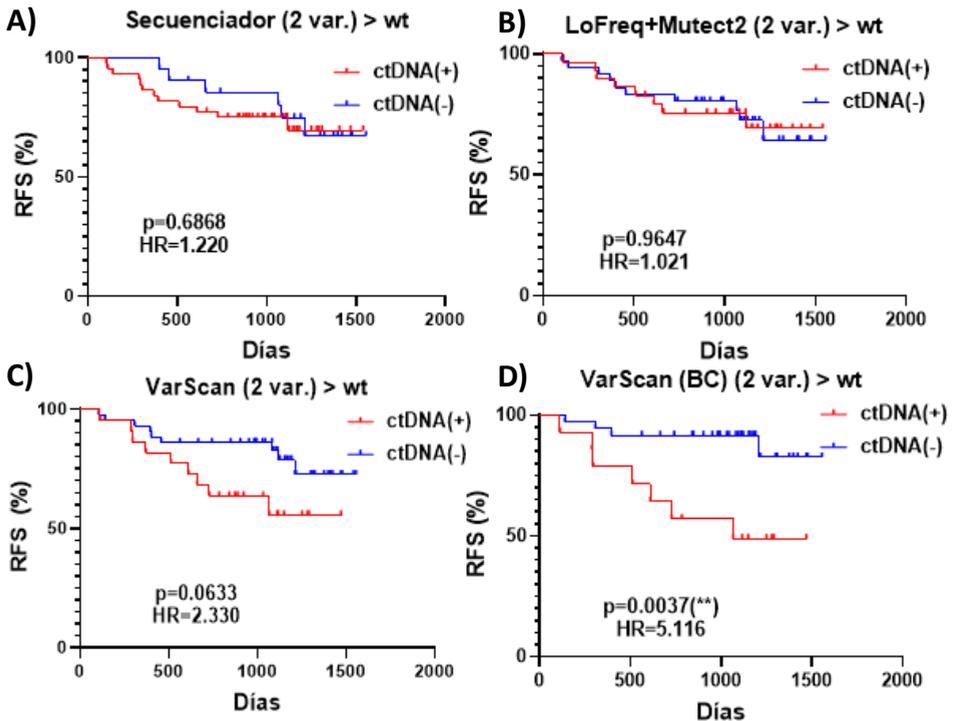


Figura 34: Capacidad de detectar la recaída en muestras posteriores a la operación de los diferentes análisis de la secuenciación de amplicones con al menos dos variantes positivas. Las gráficas A-D representan las curvas Kaplan-Meier de cada análisis, siendo estos A) los resultados crudos del secuenciador, B) análisis con LoFreq y Mutect2, C) análisis con *VarScan* y D) amplicones de buena calidad según el *VarScan*. PPV: Valor predictivo positivo. NPV: Valor predictivo negativo. RFS: supervivencia libre de recaída. HR: *Hazard Ratio*.

4.7. Marcado molecular mediante amplicones

4.7.1. Selección del diseño

El primer diseño, sin horquilla, se probó en comparación con cebadores sin UMIs (Fig. 35). Para el protocolo de 3 ciclos, se usó como referencia el empleo o no de purificación con AMPure XP en proporción 1:1 para decidir si este paso era indispensable o no. Los resultados mostraron que el único contexto en el que sucede amplificación es a partir de los 26 ciclos de PCR2 en muestras que han sido purificadas con un volumen 1:1 de AMPure XP tras la PCR1 (Fig. 35). En base a esto, se decidió descartar esta estructura.

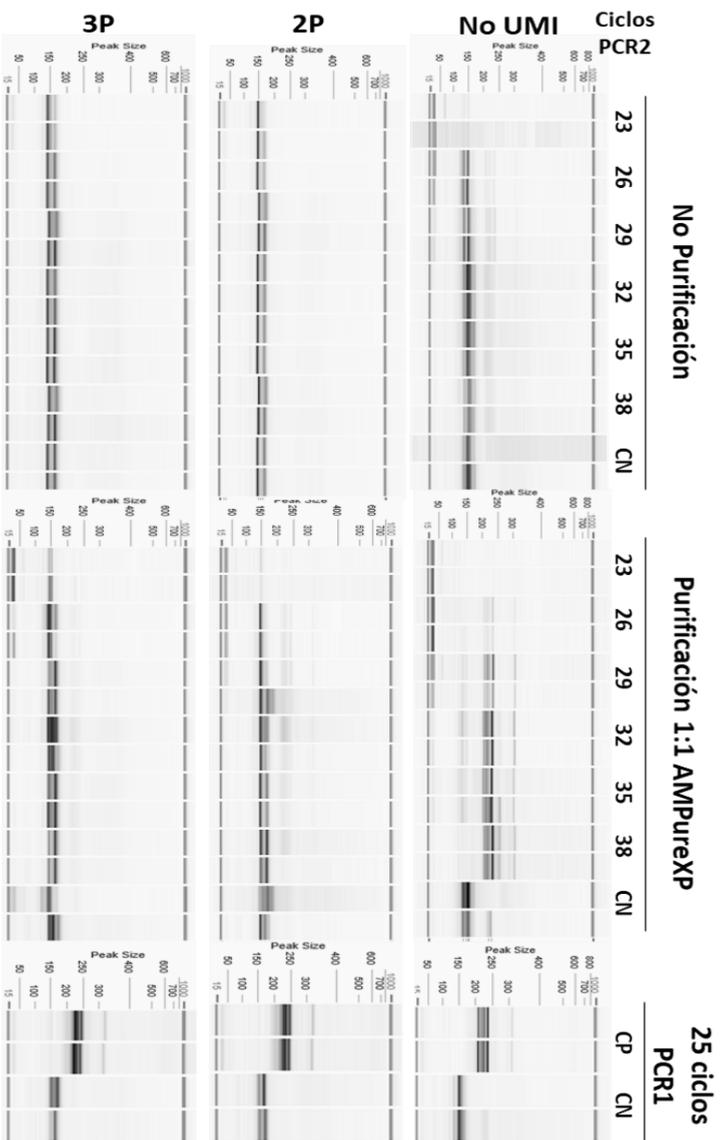


Figura 35: Validación del cebador con UMIs con secuencia protectora. La imagen muestra el electroferograma de 3 tipos de cebadores: sin UMI (control de la técnica), con una secuencia protectora de 2 nucleótidos (2P) y con tres nucleótidos (3P). Todas las muestras comenzaron con 3 ciclos de PCR1 y el número de ciclos de PCR2 indicado en la parte superior. Se comparó el efecto de la purificación y sin su uso. Como control positivo, se incluye una muestra con 25 ciclos de PCR1, mostrando la banda a la que deberían aparecer los oligos si funcionarían. CN: control negativo. CP: Control positivo.

El segundo diseño se basó en la horquilla propuesta por SiMSen-Seq. Las características principales de los 5 diseños inicialmente propuestos vienen recogidas en la tabla 25. Todos funcionaron en las condiciones planteadas para demostrar su funcionamiento. Respecto al rendimiento en secuenciación, la figura 36 recoge los aspectos relevantes de los 5 diseños, probados contra la

mutación p.E545K de *PIK3CA* y los residuos G12-G13 de *KRAS*. El análisis de la varianza o ANOVA a nivel de cada parámetro mostró que el diseño de la horquilla era un factor significativo en las lecturas mapeadas ($p < 0,0001$) y en las lecturas consenso ($p = 0,0016$) pero no en la especificidad del kit ($p = 0,2602$). En las lecturas mapeadas, se observaron diferencias significativas en HPD1 y HPD2 con los HPD3, HPD4 y HPD5 (Fig. 36A), mientras que en las lecturas consenso, se observó un mayor nivel de lecturas en los diseños HPD2 a HPD5 en comparación con el HPD1 (Fig. 36B). Por tanto, en base a estos datos, se eligió el diseño HPD2 como el óptimo para seguir con el desarrollo.

Tabla 25: Diseños de horquillas con adaptadores Nextera. Los nucleótidos autocomplementarios se muestran como suma nucleótidos procedentes del propio adaptador y de los procedentes de las bases añadidas. Tm calculados con el software UMAFold de IDT. Tm: Temperatura de disociación.

Diseño	Nucleótidos añadidos 5'	Nucleótidos autocomplementarios	Nucleótidos Aleatorios	Tm (°C)
HPD1	2 (GG)	15 (13+2)	8	80,0 °C
HPD2	2 (GG)	14 (12+2)	8	75,4 °C
HPD3	2 (GG)	13 (11+2)	8	65,1 °C
HPD4	3 (GGA)	14 (11+3)	8	65,1 °C
HPD5	3 (GGG)	14 (11+3)	8	68,2 °C
HPD6	2 (GG)	14 (12+2)	10	75,4 °C

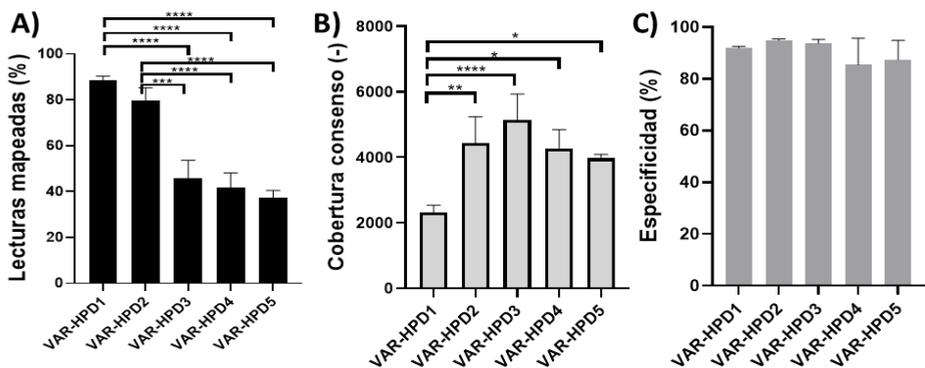


Figura 36: Comparación del rendimiento de los diferentes diseños de cebadores con horquilla. Los criterios seleccionados para valorar el rendimiento fueron el porcentaje de lecturas mapeadas (A), el número de lecturas consenso (B) y la especificidad de los amplicones (C).

Se probó la capacidad y el potencial del diseño HPD2 para marcar diferentes cantidades de ADN de partida en la primera PCR. Se observó un incremento del porcentaje de lecturas mapeadas con el aumento de material añadido (Fig. 37A). También se estableció una relación lineal entre el material de partida y el número de UMIs obtenidos (Fig. 37B), tanto para *KRAS* ($R^2=0,9386$, $p=0,0066$) como para *PIK3CA* ($R^2=0,9124$, $p=0,0113$).

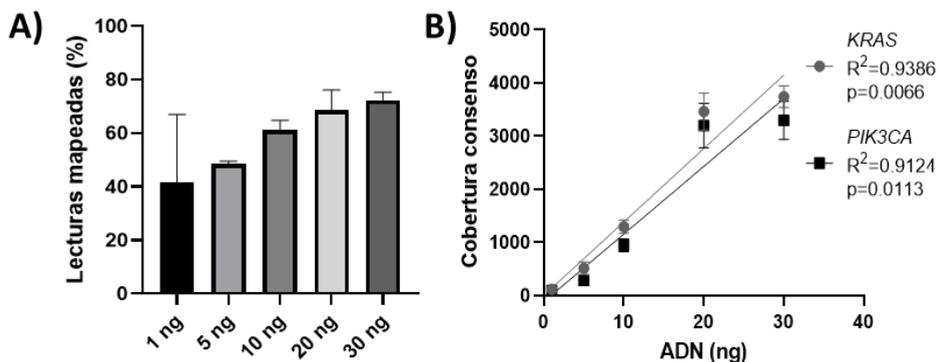


Figura 37: Valoración de la capacidad de los cebadores con horquilla para identificar moléculas. A) Lecturas mapeadas para cada punto de ADN añadido. B) Cobertura consenso obtenida en función del ADN empleado. Debido al concepto de UMI (marcaje de molécula única) se espera que cuanto mayor más ADN se haya introducido en el protocolo, mayor sea el número de UMIs diferentes que se encuentre, además de guardar una proporcionalidad directa con dicha cantidad.

4.7.2. Optimización del protocolo

Como parte de la optimización del protocolo, se empleó el protocolo descrito en la metodología (punto 3.9.2.), junto con el uso del diseño HPD6, incluido en la tabla 25. Las modificaciones incluyeron uso de esferas *HyperPure* (Roche Diagnostics) y 10 ciclos de PCR2 con mayores tiempos de hibridación.

De todas las pruebas realizadas, únicamente la concentración del cebador mostró cambios significativos al ser modificado. En función de si los 3 μ l empleados se encontraban a 0,03 μ M, 0,07 μ M, 0,13 μ M o 0,20 μ M, se ha observado una relación lineal entre la cantidad de *primer* añadida y la cantidad de lecturas

consenso obtenida (Fig. 38A), detectándose un estancamiento entre $0,13 \mu\text{M}$ y $0,20 \mu\text{M}$. Además, se observó que el aumento del cebador añadido favorecía una disminución de las lecturas mapeadas (Fig. 38B), siendo un factor significativo de la variación en este parámetro ($p=0,0002$). En dicho experimento se compararon dos métodos de purificación de la PCR1: (1) con $15 \mu\text{l}$ de volumen final y (2) añadiendo $15 \mu\text{l}$ de agua para doblar el número de esferas añadidas. Aunque no se observaron diferencias significativas, la purificación tenía un efecto leve en la variación del porcentaje de lecturas mapeadas ($p=0,0147$) y ningún efecto en las lecturas consenso ($p=0,3520$). Por ello, se decidió reducir la concentración de cebador empleado a $0,13 \mu\text{M}$, pero no modificar el paso de purificación.

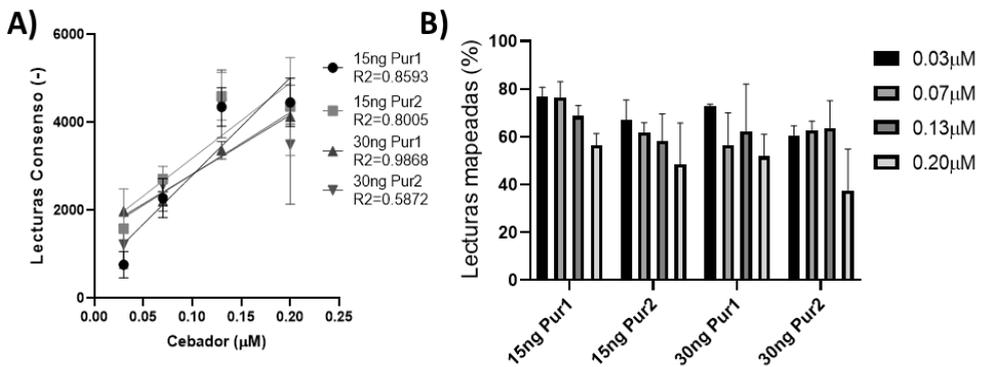


Figura 38: Optimización de la purificación y de la concentración de cebadores de PCR1. Lecturas consenso (A) y porcentaje de lecturas mapeadas (B) teniendo en cuenta el ADN de partida, la concentración de cebador empleada y el método de purificación. Pur1: $15 \mu\text{l}$ de PCR1 con $15 \mu\text{l}$ de esferas. Pur2: $15 \mu\text{l}$ de PCR1 más $15 \mu\text{l}$ de agua con $30 \mu\text{l}$ de esferas.

4.7.3. Sensibilidad del método óptimo

Una vez establecidas las condiciones adecuadas para la PCR con cebadores con horquilla, se decidió probar la sensibilidad del protocolo tanto en tejido parafinado como en cfDNA. Respecto a los tejidos, los resultados mostraron una correlación de $R^2=0,8717$, con un $p<0,0001$ para las frecuencias obtenidas en comparación con las esperadas al realizar la dilución de los estándares. Las

muestras de 15 ng mostraron unas frecuencias ligeramente superiores a las esperadas, mientras que las de 30 ng fueron más ajustadas, representado por la proximidad a la línea de puntos (Fig. 39A). Respecto al error relativo, se clasificaron las variantes en función de la concentración esperada y en función de si habían superado o no una profundidad consenso de 2000 lecturas (Fig. 39B). Se determinó que para variantes por debajo del 4% existían diferencias significativas entre el error relativo observado, siendo superior en las variantes con menos de 2000 UMIs identificados. En el caso de las variantes por debajo del 1 por ciento, el incremento era de 25,95% de error relativo promedio en muestras de más de 2000 UMIs a 161,60% en las de menos de 2000 ($p=0,0101$, Fig. 39B).

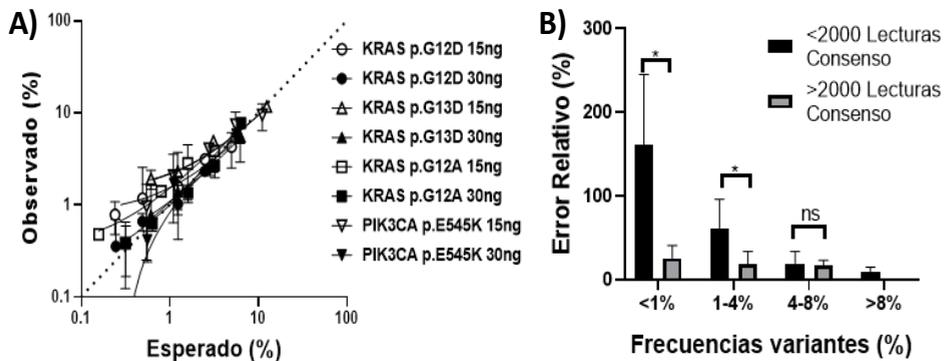


Figura 39: Sensibilidad de los amplicones con horquilla HPD6 en tejidos FFPE. A) Representación de las frecuencias calculadas frente a las determinadas por secuenciación para 15 ng y 30 ng totales de las diferentes mezclas de ADN para diluir las variantes. B) Promedio de los errores relativos para las variantes de frecuencia de los intervalos establecidos.

Posteriormente, se probó la capacidad de los amplicones con horquilla HPD6 en estándares de cfDNA (Fig. 40), utilizando el wt de HD780 como muestras sin variaciones para la reducción de ruido de fondo. Los estándares se probaron a 5, 10, 15 y 20 ng siguiendo el protocolo establecido y a 1 ng mediante purificación con PippinPrep (SAGE) debido a la alta cantidad de artefactos generados entre los cebadores. Al igual que en tejido, se observó una relación lineal entre el cfDNA de partida y la cobertura consenso ($R^2=0,7928$, $p<0,0001$, Fig. 40A). En relación con

la frecuencia alcanzada (Fig. 40B), se identificaron las variantes al 1% con tan solo 1 ng de partida, aunque en el caso de la variación p.G12D, las diferencias con las muestras no mutadas fueron mínimas. Respecto a las variantes al 0,1%, todas fueron detectadas por encima del ruido de fondo desde los 5 ng, mientras que, para 1 ng, todas fallaron (Fig. 40B). Se observó una tendencia de menor error relativo para mayores cantidades de cfDNA empleado (Fig. 40C).

Debido a las bajas frecuencias con las que se está trabajando, se decidió evaluar también los falsos positivos que pasaban los análisis (Fig. 40D). Se observó una reducción de los falsos positivos detectados en los análisis que incluían consenso molecular, aunque sin diferencias significativas (Fig. 40D). Las únicas en la que no hubo diferencias fue con las muestras a 15 ng, observándose un número similar de falsos positivos entre los 3 análisis. Se observó que un mayor *input* causó un aumento de los falsos positivos, no detectándose más de 10 si incluían consensos. En la figura 40E y 40F se tiene un ejemplo de cómo funcionó esta reducción de ruido de fondo, desapareciendo casi todas las variantes detectadas sin UMIs y sin criterios de filtrado. En el caso de la figura 40E, de 23 positivos no esperados en el amplicón de *KRAS*, se reducen a 4, de los cuales 1 no supera el mínimo de 0,05% exigido. En el caso de *PIK3CA* (Fig. 40F), desaparecen 12 de los 14 falsos positivos. 1 de ellos se ha detectado a una frecuencia superior al 5% en todas las muestras, indicando un posible artefacto del amplicón. Con los criterios establecidos, también se han visto afectados algunas variantes positivas (Fig. 41). El llamado de variantes de *VarScan* ha presentado problemas para detectar la delección de 15 nucleótidos de *EGFR*, con o sin UMIs. Se perdió la variante al 0,1% de *KRAS* a 5 ng en tres de los análisis y 10 ng en dos (Fig. 41A). La variante p.T790M de *EGFR* mostró problemas a 1 ng al analizarse sin UMIs al 1,0% (Fig. 41B) y 5,0% (Fig. 41C), mientras que fue negativa a 10 ng al 0,1%. Por otro lado, ningún análisis con *VarScan* pudo detectar la delección de *EGFR*.

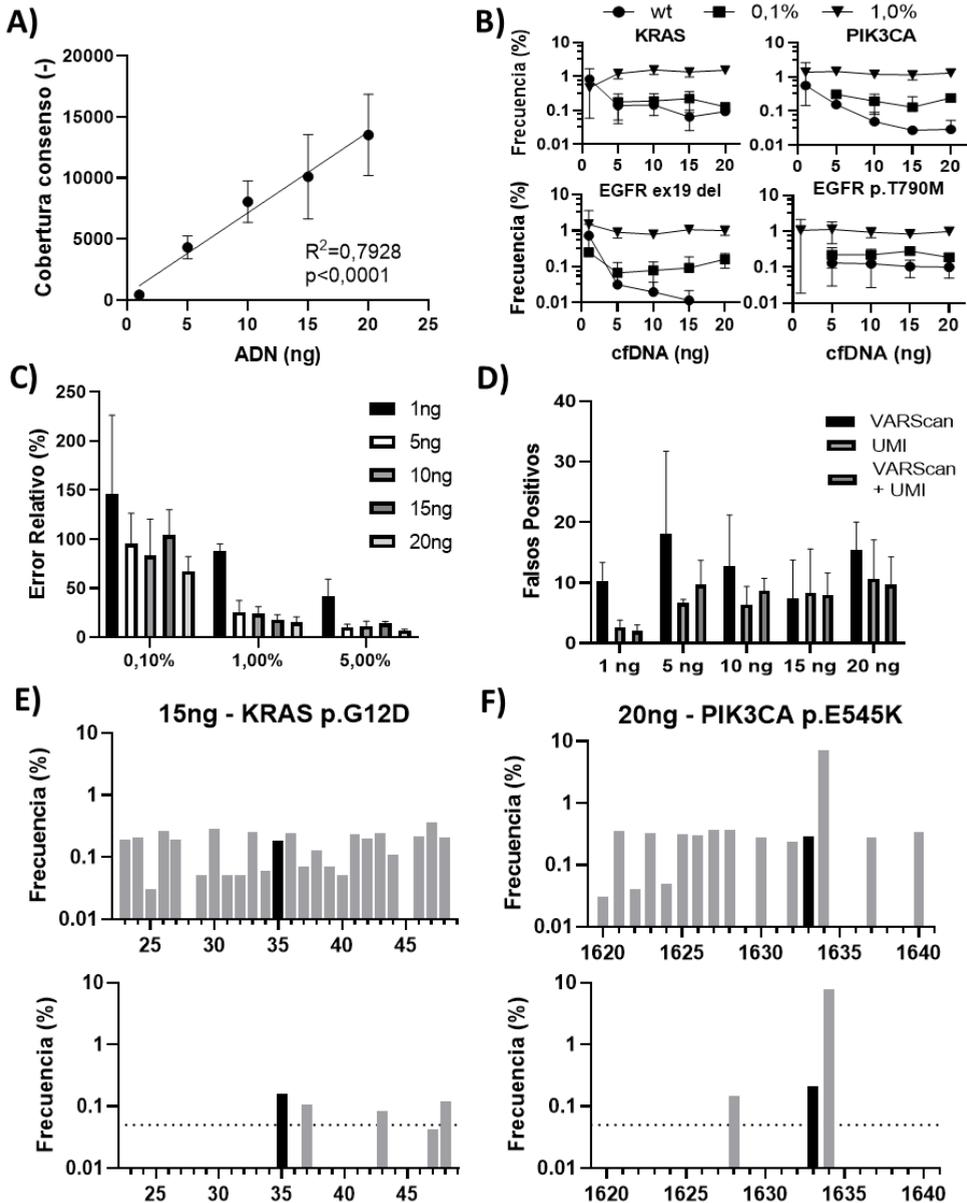


Figura 40: Sensibilidad de los amplicones con horquilla HPD6 en los estándares HD780. A) Relación entre el cfDNA cargado y la cobertura consenso obtenida. B) Frecuencia obtenida en la cobertura consenso para los estándares normal, al 0,1% y al 1,0% en las 4 variantes seguidas. C) Error relativo para los estándares al 0,1%, 1,0% y 5,0%. D) Falsos positivos obtenidos según el procesamiento de los datos en función del cfDNA empelado. E) y F) Representación de la frecuencia obtenida sin consenso (superior) y tras la eliminación del ruido de fondo (inferior) para los amplicones de las variantes al 0,1% de KRAS con 15 ng y PIK3CA con 20 ng de material. En negro aparece la alteración de interés.

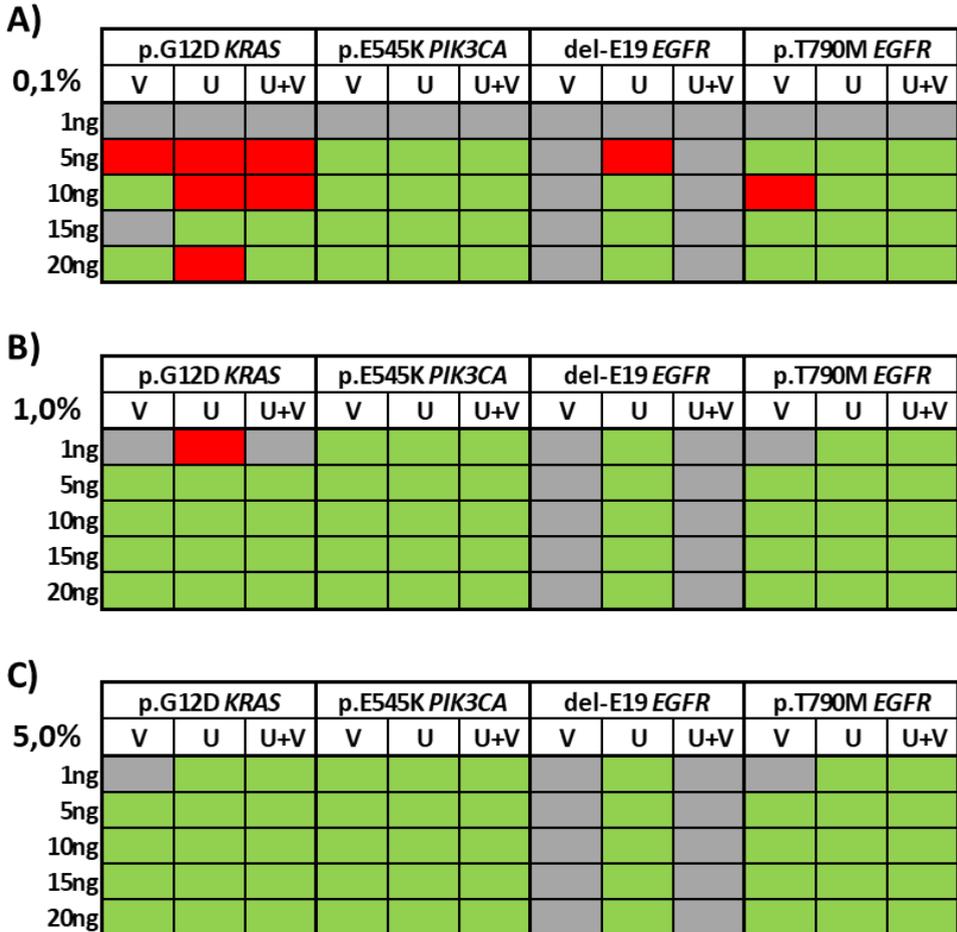


Figura 41: Positividad en los estándares de cfDNA en función de las estrategias de análisis. En el gráfico se muestra para cada variante y cantidad de partida el resultado de la evaluación de cada una de las variantes de los estándares a 0,1% (A), 1,0% (B) y 5,0% (C) con las estrategias empleadas y, al menos, una diferencia de frecuencia del 0,05%. El verde representa una evaluación positiva, el rojo una evaluación negativa y el gris la no identificación de la variante. V: *VarScan*, U: solo UMIs, U+V: UMIs y *VarScan*.

5. DISCUSIÓN

La biopsia líquida es una herramienta que ha ido ganando peso en diferentes contextos biológicos durante los últimos años, como el testado prenatal o los trasplantes (106). En el caso del cáncer, esta técnica no invasiva se ha mostrado como una estrategia con diferentes y prometedores potenciales, como la detección de enfermedad mínima residual, la identificación de variantes adicionales a las detectadas en el tejido (heterogeneidad tumoral), el seguimiento de pacientes, la detección de marcadores de resistencias o la respuesta del tratamiento en pacientes (101). En este trabajo, la biopsia líquida se ha evaluado en pacientes de cáncer de colon localizado para determinar heterogeneidad en el tratamiento, anticipar la recaída tras la cirugía mediante diferentes seguimientos, así como caracterizar la recaída mediante una visión más amplia. Por otro lado, se ha explorado también el apartado tecnológico para cada uno de estos contextos, empleándose la PCR digital y la secuenciación dirigida, ya sea con paneles pequeños dirigidos a 4 mutaciones o paneles más amplios estudiando el exoma completo.

Esta tesis se ha sustentado en la identificación de mutaciones somáticas exclusivas del tumor que establezcan una firma que sea reconocible en el cfDNA del paciente. Si bien el trabajo inicial establecido por una tesis anterior del grupo se apoyó en variantes patológicas, varios trabajos posteriores han establecido este seguimiento a través de variantes somáticas propias del tumor (199,201,238). Para poder diferenciar las variantes somáticas de mayor frecuencia de aquellas de origen germinal, se realizó un filtrado mediante secuenciación del ADN germinal procedente de las células sanguíneas de la línea blanca. Esta secuenciación tuvo dos vertientes. La primera, fue la eliminación de variantes catalogadas como de origen germinal por su localización en las células de la línea blanca a altas frecuencias (199,201,238). La segunda, consistió en encontrar variantes de hematopoyesis clonal, que, debido a su frecuencia (<5%), fueron una fuente potencial de falsos positivos en la secuenciación del cfDNA

(174,186,188). Mediante la secuenciación del material germinal se detectaron mutaciones con efecto negativo sobre la proteína en el 44,30% de los pacientes, un porcentaje elevado en comparación con el de otros estudios, aunque hay que tener en cuenta que la composición de los paneles de genes es distinta (239). Que dichas variantes estén identificadas como tal, no quiere decir que ni sean directamente patogénicas ni que se haya demostrado una implicación directa en CCR en todas, pero para algunas sí. En concreto, las mutaciones de *APC* detectadas han sido asociadas con predisposición a CCR hereditario según ClinVar, y las identificadas en *TP53* a síndromes asociados con una mayor prevalencia de cáncer.

Gracias a esta secuenciación de las células blancas, se ha producido una reducción importante en el número de variantes presentes en cada tumor, quedando únicamente aquellas desarrolladas como parte del proceso tumoral. Con nuestro panel, se ha encontrado una mediana de 7 variantes somáticas, existiendo un mínimo de 2 alteraciones en cada paciente. Esto supone una mejora frente a la aproximación anterior del grupo, donde solo se contaba con variantes patológicas, y en la que 48 pacientes carecían de al menos 2 variantes de carácter patológico. Así, se define una ventaja clave en el uso de variantes somáticas frente a las patológicas: la capacidad de poder seguir a la gran mayoría de pacientes. En estudios anteriormente comentados se usan genes de interés en el tumor o en cáncer para limitar los genes de las variantes somáticas a seguir (199,201,238), aunque también es posible seguir variantes somáticas independientemente del rol del gen en el tumor (142). La principal limitación en este punto es que la cobertura obtenida en el material procedente de células blancas es bastante inferior al obtenido en el cfDNA del plasma al diagnóstico. Esto supone que los primeros no van a representar de forma fiel las variantes de baja frecuencia que luego se pudieran encontrar en la secuenciación del plasma, por falta de cobertura (186,201).

Mediante la identificación de las mutaciones somáticas, en nuestra cohorte los 3 genes más mutados se corresponden con *APC*, *TP53* y *KRAS*. Estos 3 genes han coincidido también con los que más variantes han presentado entre los pacientes al realizar la selección somática para los seguimientos, concordando estos resultados con lo descrito bibliografía (240,241), teniendo en cuenta que la mayoría de los tumores son del tipo no hipermutados. Respecto a genes incluidos en el panel y que la literatura indica como frecuentes en hipermutados, se encuentran *RNF43*, *ACVR2A* y *TGFBR2*, estando los tres entre el 10-15% de los pacientes de la cohorte, próximo al 16% esperado de individuos con dichos tumores (24). Otros genes que han presentado una alta tasa de mutación han sido *FAT4*, *ARID1A*, *PIK3CA*, *SOX9*, *FLNA* y *SMAD4*, estando la mayoría genes relacionados con la patología del CCR. Tanto *ARID1A* como *SOX9* fueron descritos por Zaidi *et al.* como genes frecuentemente mutados, habiéndose identificado también en nuestra cohorte con una mayor frecuencia (25). En el caso de *ARID1A*, se trata de un gen implicado en CCR a través de la concurrencia con mutaciones relacionadas con la ruta de EGFR, la ruta WNT, la reparación del ADN y modulación de la cromatina. Para este gen, la frecuencia obtenida en nuestra cohorte es superior a la detectada en otros estudios (30% vs 10%) (241). En parte, esto se puede deber a la tasa de falsos positivos detectada en dicho gen o *KMT2C* en el panel. Otras tecnologías tal vez permitirían eliminar estas variantes que en general cumplen los criterios exigidos (5%, más de 50X) pero presentan frecuencias y coberturas próximas a los puntos de corte establecidos y los programas de análisis informáticos las dan por buenas. Respecto a *BRAF*, uno de los principales marcadores pronósticos y de resistencia a anti-EGFR del CCRm, se han detectado 10 mutaciones próximas al dominio *A-loop* de activación, 8 de las cuales han correspondido con el *hotspot* p.V600E, la principal mutación en este gen en CCR (36). Los porcentajes de mutaciones observados en nuestra cohorte son similares a los de la literatura y, aunque de las 10 mutaciones del total, la

mitad se han detectado en pacientes MSI, lo que hace que el porcentaje de los pacientes con inestabilidad mutados en *BRAF* sea superior al de MSS, también en concordancia con lo descrito (21,24,36).

Por último, *POLE* ha aparecido mutado en cerca del 9% de los pacientes, aunque al carecer de la clasificación en hipermutados o no hipermutados no podemos saber a qué porcentaje de dichos pacientes afecta. Sí que se ha observado, que considerando pacientes MSS con *POLE wt* frente a pacientes MSI y/o *POLE* con mutación somática (aquellos con mayor tendencia a acumular mutaciones), existe una diferencia significativa entre ambos grupos, observándose más mutaciones en el segundo, en concordancia con los resultados de Zaidi *et al.* (25), donde los pacientes hipermutados (por MSI, *POLE* con mutaciones en el dominio exonucleasa o ambos) muestran más mutaciones somáticas que los MSS. Por otro lado, en esta cohorte solo se ha encontrado una de las mutaciones más frecuentes en el dominio corrector de *POLE* (p.P286R), las cuales ocurren en menos del 0,5% de los pacientes. Otras mutaciones frecuentes para este gen no detectadas en la cohorte son la p.V411L y la p.S459G (242).

Otro de los aspectos trabajados en esta tesis ha sido la secuenciación del ADN circulante obtenido del plasma al diagnóstico mediante el uso de una tecnología de PCR anclada. La muestra empleada procede de la cirugía, de forma que, si bien no se pueden inferir resultados sobre diagnóstico precoz, sí que se puede describir la concordancia y la heterogeneidad detectada en comparación con la biopsia sólida extirpada en la operación. En referencia al diagnóstico temprano, en cáncer de colon la mayoría de las herramientas basadas en plasma están en desarrollo, enfocándose en firmas de miARN y en marcas de metilación más que en la detección de mutaciones somáticas (243). Respecto a la detección de variantes concordantes, la cantidad de mutaciones identificadas en plasma ha sido inferior en comparación con el total de mutaciones que se han detectado en el tejido, un 10% en total considerando el software y casi un 16% incluyendo las

identificadas mediante revisión de los archivos de mapeo. Casi la mitad de las mutaciones detectadas mediante software y revisión manual coinciden con los genes que más han sido representados para la selección de variantes somáticas, es decir, los de mayor frecuencia en tejido, aunque no se ha identificado una correlación entre el VAF en el tejido y en el plasma (244). Por otra parte, solo se han visto 13 variantes exclusivas de plasma en un total de 12 pacientes. La heterogeneidad tumoral ha sido descrita en CCR en tejido mediante revisión en parafinas (245) y en cfDNA (138), aunque en el segundo estudio citado solo comprueban la heterogeneidad en un paciente de CCR, sugiriendo que es posible detectar esa heterogeneidad en plasma (138). Respecto a estos valores de concordancia con el total de variantes influyen varios factores. Por una parte, está el rendimiento de la secuenciación y la sensibilidad del panel, y por otra se ha visto que no todos los tumores liberan ctDNA de la misma forma, por mecanismos aún desconocidos, habiéndose observado que hay pacientes que no positivizan en el seguimiento pese a recaer (192,226,246). Con respecto al rendimiento y la sensibilidad, como se ha visto con los estándares de plasma y los pacientes, existe una correlación entre la mediana de la cobertura consenso y la cantidad de ADN usada en el protocolo. Con 40 ng iniciales, solo se han visto 3 de las 8 variantes al 0,1% en los estándares, lo que junto a una cobertura irregular y con la baja cantidad de ADN de algunas muestras justifica la posibilidad de no detectar más variantes sin arriesgarse a detectar ruido de fondo. En cuanto a las diferencias entre pacientes en relación con la liberación de ctDNA, esto se refleja en los extremos vistos donde un paciente ha podido tener variantes al 10% en el plasma del momento de la cirugía mientras que en otros no se detecta ninguna variante. También se ha visto que el tamaño y el estadio influyen en la probabilidad de liberación (247) y en la cantidad de ADN circulante en plasma (195,196), en concordancia con lo visto en la cohorte, en las que se vio una mayor cantidad de ctDNA en función del tamaño del tumor y el estadio.

Pese a que a nivel global de mutaciones la concordancia con el plasma al diagnóstico ha sido más baja, a nivel de pacientes se ha detectado al menos una variante en dos de cada tres pacientes teniendo en cuenta el análisis y en el 90% si se considera la revisión manual. En el caso de la ddPCR, la sensibilidad ha sido más similar a la del análisis, detectándose al menos una variante en el 72,72% de los pacientes con la selección global. Los valores obtenidos para el análisis y la PCR digital son similares para los descritos por Phallen *et al.* en un análisis con varios tipos de tumores (138), mientras que los de la revisión son similares a los obtenidos con la estrategia de *Signatera* (16 variantes somáticas elegidas mediante exoma) en el estudio de Reinert *et al.* (142). Respecto a la utilidad de la detección de ctDNA en plasma al diagnóstico, se ha visto que la detección de las variantes tumorales no presenta valor predictivo para la detección de la recaída, a diferencia del plasma de las seis-ocho semanas, en concordancia con lo descrito en otros artículos e incluso de otros tumores (199). El estudio de Phallen *et al.* sí asocia la presencia de mutaciones con VAF superior al 2% en muestras de CCR al diagnóstico con la recaída de pacientes diagnosticados con enfermedad localizada, aunque dicho grupo está compuesto solo de 3 pacientes, cuestionando el resultado obtenido (138). Otro estudio del grupo de Allegretti *et al.* solo encontró una relación débil pero no significativa entre la cantidad de ctDNA en el diagnóstico y la pT y el tamaño del tumor (244).

Con relación a las variantes de hematopoyesis clonal, nuestro panel incluye muy pocos genes que han sido descritos en este proceso. Uno de estos genes es *TP53*, uno de los principales relacionados con estas variantes CHIP (189). De hecho, la mayoría de las mutaciones identificadas han sido en este gen, junto a *KRAS*. Entre ambos, han supuesto más de la mitad de las mutaciones de hematopoyesis clonal con efecto sobre la proteína. En comparación con el trabajo de Park *et al.*, el porcentaje de pacientes con estas variantes ha sido similar (188). El estudio de Zink *et al.* definió la prevalencia general en el 12,5% de la población, aunque para

el rango de edad de la cohorte, la prevalencia era de entre el 5% para individuos de 45-55 y el 25% entre 65 y 75 años (248). Por otro lado, otros estudios detectan un porcentaje superior al 50% de pacientes de cáncer con variantes de hematopoyesis clonal (186,249). El principal motivo asociable a este hecho ha sido la composición del panel, enfocada a genes de relevancia en CCR y no a detectar mutaciones CHIP. Un panel conteniendo otros genes como *JAK2*, *DNMT3A*, *TET2* o *ASXL1* permitiría una mejor visión de este fenómeno, aunque se aleja de la idea de un panel enfocado a identificar mutaciones somáticas relevantes en CCR. Esto también se puede asociar al motivo por el que, aunque al ser mayor la media de edad de los pacientes con variantes CHIP, las diferencias no son significativas, pese a que estudios asocian la aparición de estas variantes con una mayor edad (174,186,248). Adicionalmente, se ha observado una ausencia de mutaciones CHIPS en genes donde tienen mayor frecuencia en la secuenciación del exoma, posiblemente debido a la menor cobertura obtenida en la secuenciación del material germinal, siendo, por esto, difícil a nivel técnico poder detectarlas.

La biopsia líquida también ha sido empleada en este trabajo para la caracterización de la enfermedad metastásica en 14 de los pacientes que han sufrido recaída. Aunque el exoma del cfDNA ha sido estudiado en diferentes tumores y otros tipos de muestras, como líquido cefalorraquídeo (250), pocos artículos se centran en su uso en el plasma de la recaída en CCR (214,220,221, 224,251). Esta parte de la tesis supone una novedad frente a lo presentado en la literatura en CCR. Una de las principales aportaciones es la identificación de mutaciones somáticas potencialmente accionables identificadas en la recaída, permitiendo el acceso de pacientes a potenciales ensayos clínicos. En la cohorte presentada, el 71,43% de los pacientes tenía al menos una de estas mutaciones. Sí que ha sido descrito cómo el uso de exoma, a nivel terapéutico, permite caracterizar nuevas mutaciones, incluso de resistencia (224,251), monitorizar la

respuesta a inmunoterapia (252) e identificar estas mutaciones accionables, ya sea en tejido o en plasma (253). También ha sido posible identificar mutaciones germinales accionables, aunque éstas no se han correspondido con mutaciones implicadas en los principales casos de CCR hereditario (9). Adicionalmente, en las células blancas se identificaron potenciales mutaciones CHIP en los catorce pacientes, a diferencia de un estudio reciente en el que identificaron en el 52,69% de pacientes de pulmón (254). Aunque si bien a diferencia de este estudio, donde se usa un número significativo de individuos con tumores pulmonares (más de 1000), la cobertura que tienen es aproximadamente un tercio de la empleada en esta tesis.

En relación con el estudio del papel que juegan los genes identificados mediante WES, se ha encontrado enriquecimiento de genes *driver* relacionados con la respuesta a estrógenos exclusivamente mutados en plasma. Ha sido descrita la relación de los genes implicados en dicha vía con la ruta de MMR, mediante potenciación de ésta (255). De hecho, se ha demostrado la capacidad de esta vía de reducir la formación de tumores de colon (255,256) y ciertos estudios epidemiológicos asocian una menor incidencia del CCR en mujeres menores de 50 años debido al efecto protector de los estrógenos mediante el receptor ER β (257). Uno de los genes que se ha identificado con mutaciones exclusivas en plasma, *NCOR1*, codifica para un correpresor que actúa mediante el reclutamiento de una deacetilasa de histonas a los promotores, observándose que en CCR predomina una forma aberrante que lo mantiene excluido del núcleo (258). A parte de los genes calificados como *driver*, dentro de los que se han encontrado más mutados, tanto en plasma como en tejido, han sido los genes de la familia *MUC* y *HLA*. La presencia de mutaciones en aquellos que codifican para diferentes mucinas, como *MUC2*, *MUC6*, *MUC16* o *MUC19*, se ha asociado con una mayor carga mutacional y de neoantígenos y una mejor respuesta a inmunoterapia en tumores sólidos (259). Por otro lado, en CCR se asocia una alta expresión de *MUC1* con

mayor actividad tumoral y metástasis y un posible rol protector para MUC2, sin referenciar el estado mutacional (260). En el caso de *HLA-B*, que codifica para una de las cadenas del complejo mayor de histocompatibilidad I (*Major Histocompatibility Complex I*, MHC-I), el gen ha aparecido mutado en 42,86% de los pacientes, ya sea en recaída o tejido primario. Este complejo, relacionado con la presentación de antígenos intracelulares, es conocido por presentar alteraciones de su expresión en CCR (261), y las pérdidas de heterocigosidad somáticas se relacionan con peor supervivencia en pacientes tratados con bloqueo de *immune-checkpoints* (262). En cuanto a mutaciones somáticas, se ha descrito que las desarrolladas en genes del HLA se asocian con una mayor carga de neoantígenos en tumores MSS, observándose una mayor citotoxicidad en tumores de colon con HLA mutado en comparación con los no alterados (263).

La concordancia observada entre el tejido inicial y las variantes identificadas en las recaídas se ha situado entre el 3,65% y el 31,20% con referencia en el tejido, correspondiendo con valores inferiores a los descritos, situados entre el 15% y el 49% (216). Hay que destacar ausencia de concordancia en variantes patogénicas, como ocurre en *KRAS* (Anexo VI). Ha sido descrito que el desarrollo de la metástasis no tiene porqué suceder siempre a partir del clon mayoritario, pudiendo proceder de otros sub-clones (62), lo que justificaría la pérdida de determinadas mutaciones *driver*, aunque no se haya descrito el significado biológico de esta pérdida. Otro aspecto para tener en cuenta relacionado con la no detección de variantes es la proporción de ctDNA entre el ADN circulante. Teniendo en cuenta que el rango de detección ha podido variar desde pacientes con menos del 1% frente a otros con variantes por encima del 10%, es posible que variantes patogénicas que dejan de detectarse en el WES de la recaída lo hagan por una cuestión tecnológica (216). Las coberturas medias obtenidas, aunque superiores en términos generales a las de la literatura (216), tampoco son las óptimas a nivel técnico para detectar variantes a tan baja frecuencia (140,172), lo

que explica la detección de algunas variantes en ddPCR pero no exoma. En este aspecto, existe un balance entre invertir recursos en una mayor cobertura para llegar a variantes en muy baja proporción y alcanzar un mayor número de genes y regiones, en busca de alteraciones menos comunes en cáncer de colon. Por último, uno de los aspectos mejorables es la falta de WES en las muestras de la cirugía. La heterogeneidad tumoral es un fenómeno descrito también al diagnóstico (138,208), por lo que, sin esta muestra, no podemos confirmar con total seguridad que las variantes obtenidas no se encontraban en el momento inicial, siendo más acertado expresar que las variantes detectadas no se encontraban en la biopsia de tejido primario empleada.

La biopsia líquida también ha permitido trabajar en la detección de la enfermedad mínima residual. En este trabajo se han empleado la ddPCR y la secuenciación masiva con enriquecimiento en regiones de interés mediante amplicones usando muestras control para reducir el ruido de fondo. Este trabajo ha comparado la estrategia anterior del grupo, basada en usar solo variantes patogénicas en ddPCR, frente al uso de variantes somáticas de genes de relevancia en CCR. Los resultados han mostrado que ambas estrategias por separado presentan una sensibilidad inferior a una estrategia combinada, analizando hasta 4 variantes somáticas, independientemente de su carácter patogénico, de alta frecuencia en el tejido. Esto ha sido notable en las muestras postoperatoria, de menor frecuencia, donde la mejor sensibilidad se ha obtenido a partir de la tercera variante de mayor frecuencia incluida. En definitiva, una clara ventaja aportada por esta estrategia es independiente del criterio de selección “somático” o “patológico”, y subyace en el uso de un mayor número de alteraciones, observándose un claro beneficio al usar más de tres. Por otro lado, NGS y ddPCR han sido capaces de distinguir la recaída en muestras de plasma en muestras a las 6-8 semanas de la cirugía con una sensibilidad del 63,64% para los amplicones y el 52,94% para la ddPCR, dando también valores similares de especificidad, PPV y

NPV. Este hecho solo confirma los hallazgos definidos por otros grupos, en el que la detección de ctDNA por diferentes medios (mutaciones somáticas, variaciones estructurales...) en plasma posterior a la cirugía se define como un predictor de la recaída, variando el porcentaje en función de la tecnología empleada (142,226,227,264). El grupo de Suzuki *et al.* presentó una estrategia similar, empleando ddPCR obtuvo un valor predictivo verdadero del 50%, aunque manejaron muestras a una semana de la cirugía (238), momento que algunos artículos defienden como viable para detectar ctDNA con carácter pronóstico para la recaída (265–267). Aunque hay que tener en cuenta que en periodos tan cortos no todos los pacientes pueden haber eliminado el cfDNA procedente de procesos inflamatorios derivados de la operación, siendo este un campo que aún necesita más investigación (266,268,269). El grupo de Parikh *et al.*, por otro lado, ha reportado una sensibilidad del 55,6% pero con un PPV del 100% al emplear la secuenciación dirigida para detectar mutaciones somáticas junto con análisis de metilación en muestras al mes de la cirugía o la adyuvancia (192). La estrategia del grupo de Reinert *et al.* es similar a la aquí empleada, empleando 16 mutaciones procedentes de exoma y usando al menos 2 variantes para considerar al paciente positivo, consiguiendo una sensibilidad de 41,17% con muestras a 30 días de la operación (142). La estrategia presentada usa un panel de genes relevantes en CCR, obteniendo una sensibilidad ligeramente superior e invirtiendo menor coste en la secuenciación inicial. Algo que tienen en común estas estrategias es que incorporan secuenciación con reducción de error para trabajar en plasma posterior a la cirugía. Poco a poco, la ddPCR está siendo desplazada del uso en biopsia líquida para la búsqueda de mutaciones en la medida en la que la NGS se convierte en una tecnología más sensible y permite analizar muchas más posiciones que las mutaciones únicas a las que están limitadas las sondas de ddPCR. De hecho, recientemente se está empezando a recomendar la NGS antes que la ddPCR en cáncer de colon para la identificación

de la EMR (270). Por otra parte, ESMO ya considera la EMR como posible marcador bajo investigación para adelantar la recaída (49).

Además del postoperatorio, en este trabajo se han evaluado seguimientos mediante el uso de la ddPCR. Junto a la detección de la EMR en postoperatorio, la evaluación de muestras cada 3-4 meses es donde la biopsia líquida ha mostrado un rol importante: en este trabajo se ha comprobado como en estas muestras posteriores es posible detectar con antelación la recaída en comparación con las tecnologías ya establecidas de imagen y diagnóstico. Este hecho ha sido realizado también por diferentes grupos, obteniendo diferentes meses de antelación en la detección de la recaída (142,238,271). En los análisis anteriores del grupo presentados por Tarazona *et al.*, la sensibilidad en el seguimiento fue del 82,40% en comparación con el 76,47% obtenido en este nuevo análisis (225). Tras este cambio se encuentra la recaída de pacientes de la cohorte que para el trabajo inicial no habían recaído, por lo que, si recaen, pero no son detectados, suponen una caída en la sensibilidad. Adicionalmente, en los pacientes con recaída, esta evaluación ha permitido ver dinámicas interesantes de la evolución de la presencia de ctDNA en plasma. Algunos de los pacientes con recaída no han mostrado ctDNA detectable con las variantes seleccionadas, independientemente del momento. Esto puede deberse a que el tumor no libera ctDNA (246), como se ha comentado anteriormente, o a la fundación de la metástasis mediante un clon cuyas mutaciones no hayan sido las elegidas (62,63). Otra tendencia observada es la de pacientes en los que la recaída ha sido detectada con VAFs elevados, habiendo pacientes donde las variantes se podían detectar al 10% e incluso al 40%. Por último, la tendencia más interesante ha sido la de aquellos pacientes que han recibido quimioterapia. En 6 de los casos en los que el ctDNA es detectable, 2 muestran un pequeño aumento para no volver a positivar, mientras que en los otros 4, se produce una disminución de la presencia de ctDNA para luego crecer tras la adyuvancia. En todos los casos, el tratamiento supone un

descenso del ADN tumoral circulante detectable, suponiendo un reflejo de la eficacia de la quimioterapia en esos pacientes. Este fenómeno de depuración del ctDNA ha sido descrito por otros artículos, y la positividad después de la quimioterapia se ha asociado con un peor pronóstico (142,225,226). De todas formas, al no haberse analizado todos los seguimientos de los pacientes sin recaída con la selección de las variantes somáticas, no es posible extraer conclusiones adecuadas de todos los pacientes tratados.

Frente a estrategias basadas en uso de grandes paneles de genes o de sondas específicas para un cambio concreto, en este trabajo se ha planteado el uso de una estrategia más dirigida a mutaciones de interés, requiriendo de menores costes y de reducida capacidad de secuenciación para detectar el ctDNA, siguiendo la línea de otros estudios (142,174,175). El primer tipo de diseño planteado, basado en una cola aleatoria después de los oligonucleótidos específicos del amplicón no ha progresado, principalmente a causa de la posible inespecificidad introducida por la secuencia aleatoria. Frente a esta situación, el grupo de Masunaga *et al.* (2018) planteó la reducción de las combinaciones posibles mediante restricción de un tipo de nucleótido en cada una de las posiciones, como propone en su estrategia *Personalized MB-NGS* (175). En cambio, la estrategia empleada por *AmpliSeq HD* introduce el UMI sin reportar ningún tipo de modificación (174). El segundo diseño empleado ha sido el basado en SiMSen-Seq (176), que incluye una modificación en el cebador con el oligonucleótido, de forma que en la fase de hibridación la secuencia aleatoria está bloqueada dentro de una horquilla. ATOM-Seq es otra estrategia que también mantiene el UMI escondido mediante una secuencia aleatoria complementaria unida mediante una molécula de unión (172). En general, la mayoría de las estrategias que incorporan UMI mediante PCR difieren en cómo lidian con la inespecificidad que introduce en el cebador. Este problema no surge en los protocolos de captura o PCR anclada que introducen el UMI en la ligación, ya que

la adición del adaptador es mediada por colas de Adenina y el resto del protocolo emplea oligonucleótidos dirigidos a regiones en las que no interviene el marcaje molecular.

Este trabajo ha adaptado el diseño de los adaptadores de SiMSen-Seq, basados en TruSeq, al diseño de los adaptadores Nextera. Mediante modificaciones del protocolo establecido en el laboratorio, centradas en el número de ciclos, las temperaturas de alineamiento de la PCR2 y la longitud de los cebadores de PCR2, se ha conseguido trasladar el diseño para lograr una librería funcional. Nuestros resultados muestran que mediante el uso de esta tecnología se consigue detectar variantes al 0,1% por encima del ruido de fondo a partir de los 5 ng de ADN de partida y con alta fiabilidad al 1,0% solo empleando 1 ng de material. Para este tipo de muestras, son necesarias unas condiciones en las que se ha conseguido reducir la presencia de dímeros de adaptador, para no reducir el rendimiento de la secuenciación. Otras tecnologías basadas en amplicones, sin determinar el material de partida, alcanzan valores similares, como es el caso de *Ampliseq HD*, que llega a reportar variantes al 0,16% (272), o SiMSen-Seq, donde se llega a detectar una variante al 0,063% (176). En el caso de la *Personalized MB-NGS*, con un máximo de 3 ng se llegó a identificar una frecuencia del 0,1% con seguridad, requiriendo un mínimo de 30 lecturas por cada marcador molecular (175). La estrategia ATOM-Seq permite llegar a variantes al 0,1% con al menos 20 ng y menor cobertura que el anteriormente mencionado, con la ventaja de que el *primer* específico puede ser diseñado junto a la mutación de interés, siendo más versátil y económico que un panel de PCR anclada que requiere ligación (172). Aunque el llamado de variantes realizado por *VarScan* ha mostrado problemas con las grandes delinserciones, los archivos BAM mostraban la presencia de las 4 variantes seguidas por encima del ruido de fondo, alcanzando el objetivo de tener una tecnología que permita detectar dichas variantes a frecuencias próximas a 0,1%. Es importante definir la capacidad de la tecnología en base a la cantidad de

ADN circulante empleado, ya que es la principal limitante que se va a encontrar en muchas de las muestras de cfDNA, siendo necesario establecer un protocolo previo a la PCR para alcanzar la concentración deseada. Entre los puntos de actuación encontraríamos el volumen de elución adecuado, un volumen de plasma alto y un posible paso de concentración de la muestra para adaptarlo al volumen de partida de este protocolo. Como se ha visto en los pasos de optimización, tener una cantidad de cebadores adecuada para la cantidad de material de partida ayuda a la reducción de la formación de estructuras secundarias, de la misma forma que un mayor repertorio de moléculas analizadas mejora la capacidad de detectar variantes a menor frecuencia. Por otro lado, establecer un valor mínimo de diferencia entre el ruido de fondo y la frecuencia detectada en la muestra ayuda a reducir los falsos positivos, aunque en ciertos amplicones pueda suponer la pérdida de alguna variante. Una estrategia posible es no usar un valor arbitrario y emplear una prueba basada en la media y desviaciones de las medidas de las variantes y el ruido de fondo, lo cual conseguiría una gran reducción de los falsos positivos, aunque debido a la precisión y el propio ruido de los secuenciadores, se perderían variantes de baja frecuencia por la dispersión de las medidas, en base a revisión en los propios datos. Surge otro punto de optimización, siendo este las polimerasas empleadas, buscando reducir el error mediante el uso de enzimas con menores tasas de error (273). Hay que recordar que la idea de esta estrategia es simplemente valorar las variantes específicas del paciente y gran parte de las requeridas se detectan con los criterios establecidos. Otro aspecto a tener en cuenta es que el estándar empleado en la tesis como representación del cfDNA es generado en líneas celulares. Esto implica que variantes procedentes de las líneas parentales puedan aparecer a baja frecuencia y no hayan sido validadas, lo cual causa que una alteración pueda estar y ser detectada, pero no haya sido descrita previamente a la comercialización, considerándose erróneamente como falso positivo. Además,

otra limitación es la no evaluación en otro tipo de estándar negativo de ADN circulante, puesto que la muestra no mutada del estándar con el que determinar dicho valor ha sido empleada para establecer la medida del ruido de fondo. En última instancia, el auténtico valor de esta herramienta sería evaluado mediante la detección de la EMR de muestras postoperatorias de pacientes. Actualmente, existen ensayos clínicos que usan tecnologías basadas en amplicones para determinar la EMR de pacientes de CCR. Algunos de ellos usan tecnologías con UMIs como *Safe-Seq* en los estudios de Tie *et al.* (226,227,274) o *Guardant360™* como en los ensayos del estudio GOZILA (233) y otros basados en reducción de ruido de fondo como *Signatera* en los ensayos del estudio CIRCULATE-Japan (230).

A lo largo del trabajo se han ido encontrando diferentes limitaciones, que o bien se han ido comentado en cada punto de la discusión o se comentan a continuación. A nivel de caracterización del tumor, se carece del fenotipo principal del tumor (CIMP, CIN), aunque si se tiene aquellos que tienen defectos en el MMR y, gracias a la secuenciación, aquellos mutados en *POLE*. Tampoco se han realizado análisis teniendo en cuenta la localización (izquierdo-derecho), si bien los principales objetivos de la tesis no han ido a caracterizar tan al detalle y han buscado un enfoque más global del CCR. Por otro lado, tanto la tecnología de amplicones anclados como los amplicones con marcaje molecular han presentado la introducción de falsos positivos difíciles de diferenciar informáticamente, que mediante tecnologías de captura o con menor uso de ciclos de PCR permita no producirlos. Esto hace necesario seguir mejorando en la reducción de ruido de fondo, especialmente en los amplicones para evaluar los plasmas posteriores a la cirugía. A nivel de la secuenciación de exoma, se podría haber empleado el plasma al diagnóstico también para caracterizar mejor la heterogeneidad tumoral inicial y discernir que las variantes obtenidas son por nuevos clones y no porque no se han detectado inicialmente, dado el potencial de dicha heterogeneidad mencionada recurrentemente en el texto. Por último, los estándares comerciales

como el HD780 es una mezcla de ADN genómico fragmentado centrada en 166pb, no siendo una réplica exacta del cfDNA, aunque sí próxima.

6. CONCLUSIONES

Las conclusiones obtenidas en la presente tesis son las siguientes:

1. La secuenciación de ADN germinal permite detectar en los pacientes las alteraciones no somáticas, lo cual permite eliminar falsos positivos, entendidos como mutaciones no desarrolladas en el proceso tumoral.
2. El seguimiento de hasta 4 mutaciones somáticas de alta frecuencia en tejido y/o relevancia en el tumor mediante PCR digital permite detectar la recaída con una sensibilidad del 52,94% y especificidad del 84,31% en plasma tras la cirugía, incrementando la sensibilidad al 76,47% en las muestras de los seguimientos.
3. La secuenciación del ADN circulante en el momento de la cirugía presenta un uso limitado en cáncer de colon, aunque sí permite detectar alteraciones no presentes en el tejido en un bajo porcentaje de pacientes.
4. La biopsia líquida puede jugar un papel importante en la descripción de las alteraciones metastásicas cuando no existe tejido para su análisis, descubriendo mutaciones accionables y de resistencia en cáncer de colon.
5. A su vez, la secuenciación del exoma en la recaída permite generar una perspectiva de la dinámica de las alteraciones en comparativa con el estadio localizado.
6. El uso de pequeños paneles específicos para los pacientes, cubriendo las 4 mutaciones somáticas más frecuentes, permite detectar la recaída con una sensibilidad del 63,64% en plasma posterior a la cirugía, superior a la sensibilidad ofrecida por la PCR digital.
7. El uso de UMIs con horquilla junto a la secuenciación de muestras de ADN circulante no mutado permite reducir el ruido de fondo para detectar variantes al 0,1%, aunque un pequeño porcentaje de variantes permanecería como falsos positivos.

7. BIBLIOGRAFÍA

1. Sung, H.; Ferlay, J.; Siegel, R. L.; *et al.* (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 71(3), 209–249. DOI: 10.3322/caac.21660
2. SEOM. (2021). Las cifras del cáncer en España 2021. *SEOM* 1–40
3. Golemis, E. A.; Scheet, P.; Beck, T. N.; *et al.* (2018) Molecular mechanisms of the preventable causes of cancer in the United States. *Genes Dev* 3232(13-14), 868–902. DOI: 10.1101/gad.314849.118
4. Theodoratou, E.; Timofeeva, M.; Li, X.; *et al.* (2017). Nature, Nurture, and Cancer Risks: Genetic and Nutritional Contributions to Cancer. *Annu Rev Nutr* 37, 293–320. DOI: 10.1146/annurev-nutr-071715-051004
5. Hanahan, D. (2022) Hallmarks of Cancer: New Dimensions. *Cancer Discov* 12(1), 31–46. DOI: 10.1158/2159-8290.CD-21-1059
6. Siegel, R. L.; Torre, L. A.; Soerjomataram, I.; *et al.* (2019). Global patterns and trends in colorectal cancer incidence in young adults. *Gut* 68(12), 2179–2185. DOI: 10.1136/gutjnl-2019-319511
7. Vuik, F. E. R.; Nieuwenburg, S. A. V.; Bardou, M.; *et al.* (2019). Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. *Gut*, 68(10), 1820-1826. DOI: 10.1136/gutjnl-2018-317592
8. Recio-Boiles, A. & Cagir, B. (StatPearls Publishing, 2021) in *StatPearls [Internet]* at <<https://www.ncbi.nlm.nih.gov/books/NBK430685>>
9. Conteduca, V.; Sansonno, D.; Russi, S.; *et al.* (2013) Precancerous colorectal lesions (Review). *Int J Oncol* 43(4), 973–984. DOI: 10.3892/ijo.2013.2041
10. Bateman, A. C. (2021) The spectrum of serrated colorectal lesions—new entities and unanswered questions. *Histopathology* 78(6), 780–790. DOI: 10.1111/his.14305
11. Bateman, A. C. (2014) Pathology of serrated colorectal lesions. *J Clin Pathol* 67(10), 865–874. DOI: 10.1136/jclinpath-2014-202175
12. Watanabe, T.; Kobunai, T.; Yamamoto, Y.; *et al.* (2012). Chromosomal instability (CIN) phenotype, CIN high or CIN low, predicts survival for colorectal cancer. *J Clin Oncol* 30(18), 2256–2264. DOI: 10.1200/JCO.2011.38.6490
13. Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol* 7(5), 335–346. DOI: 10.1038/nrm1907
14. Grady, W. M. & Carethers, J. M. (2008) Genomic and Epigenetic Instability in Colorectal Cancer Pathogenesis. *Gastroenterology* 135(4), 1079–1099. DOI: 10.1053/j.gastro.2008.07.076
15. Nojadeh, J. N.; Sharif, S. B. & Sakhinia, E (2018) Microsatellite instability in colorectal cancer. *EXCLI J* 17, 159–168. DOI: 10.17179/excli2017-948
16. Mármol, I.; Sánchez-de-Diego, C.; Dieste, A. P.; *et al.* (2017) Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. *Int J Mol Sci* 18(1), 197. DOI: 10.3390/ijms18010197

17. Gupta, R.; Sinha, S. & Paul, R. N (2018) The impact of microsatellite stability status in colorectal cancer. *Curr Probl Cancer* 42(6), 548–559. DOI: 10.1016/j.currprobcancer.2018.06.010
18. Pramil, E.; Dillard, C.; Escargueil, A. E.; *et al.* (2021). Colorectal Cancer and Immunity: From the Wet Lab to Individuals. *Cancers (Basel)* 13(7), 1713. DOI: 10.3390/cancers13071713
19. Lao, V. V. & Grady, W. M. (2011) Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* 8, 686–700. DOI: 10.1038/nrgastro.2011.173
20. de Palma, F. D. E.; D'argenio, V.; Pol, J.; *et al* (2019) The molecular hallmarks of the serrated pathway in colorectal cancer. *Cancers (Basel)* 11(7), 1017. DOI: 10.3390/cancers11071017
21. Testa, U.; Castelli, G. & Pelosi, E. (2020). Genetic alterations of metastatic colorectal cancer. *Biomedicines* 8(10), 414. DOI: 10.3390/biomedicines8100414
22. Kozłowska, J.; Karpinski, P.; Szmida, E.; *et al.* (2012). Assessment of chromosomal imbalances in CIMP-high and CIMP-low/CIMP-0 colorectal cancers. *Tumour Biol* 33(4), 1015–1019. DOI: 10.1007/s13277-012-0334-2
23. Ward, R.; Meagher, A.; Tomlinson, I.; *et al.* (2001). Microsatellite instability and the clinicopathological features of sporadic colorectal cancer. *Gut* 48(6), 821–829. DOI: 10.1136/gut.48.6.821
24. Muzny, D. M.; Bainbridge, M. N.; Chang, K.; *et al.* (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487(7407), 330–337. DOI: 10.1038/nature11252
25. Zaidi, S. H.; Harrison, T. A.; Phipps, A. I.; *et al.* (2020). Landscape of somatic single nucleotide variants and indels in colorectal cancer and impact on survival. *Nat Commun* 11(1), 3644. DOI: 10.1038/s41467-020-17386-z
26. Testa, U.; Pelosi, E. & Castelli, G. (2018). Colorectal Cancer: Genetic Abnormalities, Tumor Progression, Tumor Heterogeneity, Clonal Evolution and Tumor-Initiating Cells. *Medical Sciences* 6(2), 31. DOI: 10.3390/medsci6020031
27. Aghabozorgi, A. S.; Bahreyni, A.; Soleimani, A.; *et al.* (2019). Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives. *Biochimie* 157, 64–71. DOI: 10.1016/j.biochi.2018.11.003
28. Herbst, A.; Jurinovic, V.; Krebs, S.; *et al.* (2014). Comprehensive analysis of β -catenin target genes in colorectal carcinoma cell lines with deregulated Wnt/ β -catenin signaling. *BMC Genomics* 15, 74. DOI: 10.1186/1471-2164-15-74
29. Wolff, R. K.; Hoffman, M. D.; Wolff, E. C.; *et al.* (2018). Mutation analysis of adenomas and carcinomas of the colon: Early and late drivers. *Genes Chromosomes Cancer* 57(7), 366–376. DOI: 10.1002/gcc.22539
30. Tsilimigras, D. I.; Ntanasis-Stathopoulos, I.; Bagante, F.; *et al* (2018) Clinical significance and prognostic relevance of KRAS, BRAF, PI3K and TP53 genetic mutation analysis for resectable and unresectable colorectal liver metastases: A

- systematic review of the current evidence. *Surg Oncol* 27(2), 280–288. DOI: 10.1016/j.suronc.2018.05.012
31. Brambilla, E. & Gazdar, A. (2009) Pathogenesis of lung cancer signalling pathways: Roadmap for therapies. *Eur Respir J* 33(6), 1485–1497. DOI: 10.1183/09031936.00014009
 32. Sartore-Bianchi, A.; Amatu, A.; Porcu, L.; *et al.* (2019). HER2 Positivity Predicts Unresponsiveness to EGFR-Targeted Treatment in Metastatic Colorectal Cancer. *Oncologist* 24(10), 1395–1402. DOI: 10.1634/theoncologist.2018-0785
 33. Ross, J. S.; Fakih, M.; Ali, S. M.; *et al.* (2018). Targeting HER2 in colorectal cancer: The landscape of amplification and short variant mutations in ERBB2 and ERBB3. *Cancer* 124(7), 1358–1373. DOI: 10.1002/cncr.31125
 34. Stolze, B.; Reinhart, S.; Bullinger, L.; *et al.* (2014). Comparative analysis of KRAS codon 12, 13, 18, 61, and 117 mutations using human MCF10A isogenic cell lines. *Sci Rep* 5, 8535. DOI: 10.1038/srep08535
 35. Yiu, A. & Yiu, C. (2016). Biomarkers in Colorectal Cancer. *Anticancer Res* 36, 1093–1102. PMID: 26977004.
 36. Nakayama, I.; Hirota, T. & Shinozaki, E (2020) BRAF mutation in colorectal cancers: From prognostic marker to targetable mutation. *Cancers (Basel)* 12(11), 3236. DOI: 10.3390/cancers12113236
 37. Leggett, B. & Whitehall, V. (2010). Role of the Serrated Pathway in Colorectal Cancer Pathogenesis. *Gastroenterology* 138(6), 2088–2100. DOI: 10.1053/j.gastro.2009.12.066
 38. Rajagopalan, H.; Bardelli, A.; Lengauer, C.; *et al.* (2002). RAF/RAS oncogenes and mismatch-repair status. *Nature* 418(6901), 934–934. DOI: 10.1038/418934a
 39. Samuels, Y.; Diaz, L. A.; Schmidt-Kittler, O.; *et al.* (2005). Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell* 7(6), 561–573. DOI: 10.1016/j.ccr.2005.05.014
 40. Haley, L.; Tseng, L. H.; Zheng, G.; *et al.* (2015). Performance characteristics of next-generation sequencing in clinical mutation detection of colorectal cancers. *Modern Pathology* 28(10), 1390–1399. DOI: 10.1038/modpathol.2015.86
 41. Salvatore, L.; Calegari, M. A.; Loupakis, F.; *et al.* (2019). PTEN in colorectal cancer: Shedding light on its role as predictor and target. *Cancers (Basel)* 11(11), 1765. DOI: 10.3390/cancers11111765
 42. Cheruku, H. R.; Mohamedali, A.; Cantor, D. I.; *et al.* (2015). Transforming growth factor- β , MAPK and Wnt signaling interactions in colorectal cancer. *EuPA Open Proteom* 8, 104–115. DOI: 10.1016/j.euprot.2015.06.004
 43. Itatani, Y.; Kawada, K. & Sakai, Y. (2019) Transforming growth factor- β signaling pathway in colorectal cancer and its tumor microenvironment. *Int J Mol Sci* 20(23), 5822. DOI: 10.3390/ijms20235822

44. Wu, C. C.; Hsu, T. W.; Yeh, C. C.; *et al.* (2020) The role of transcription factor caudal-related homeobox transcription factor 2 in colorectal cancer. *Tzu Chi Med J* 32(4), 305–311. DOI: 10.4103/tcmj.tcmj_49_20
45. Yu, J.; Liu, D.; Sun, X.; *et al.* (2019). CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/ β -catenin signaling via transactivation of GSK-3 β and Axin2 expression. *Cell Death Dis* 10(1), 26. DOI: 10.1038/s41419-018-1263-9
46. Bruun, J.; Sveen, A.; Barros, R.; *et al.* (2018). Prognostic, predictive, and pharmacogenomic assessments of CDX2 refine stratification of colorectal cancer. *Mol Oncol* 12(9), 1639–1655. DOI: 10.1002/1878-0261.12347
47. Briggs, S. & Tomlinson, I. (2013). Germline and somatic polymerase ϵ and δ mutations define a new class of hypermutated colorectal and endometrial cancers. *J Pathol* 230(2), 148–153. DOI: 10.1002/path.4185
48. Vu, T. & Datta, P. K. (2017). Regulation of EMT in colorectal cancer: A culprit in metastasis. *Cancers (Basel)* 9(12), 171. DOI: 10.3390/cancers9120171
49. Argilés, G.; Tabernero, J.; Labianca, R.; *et al.* (2020). Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol* 31(10), 1291–1305. DOI: 10.1016/j.annonc.2020.06.022
50. Wolf, A. M. D.; Fontham, E. T. H.; Church, T. R.; *et al.* (2018). Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin* 68(4), 250–281. DOI: 10.3322/caac.21457
51. van der Geest, L. G. M.; Lam-Boer, J.; Koopman, M.; *et al.* (2015). Nationwide trends in incidence, treatment and survival of colorectal cancer patients with synchronous metastases. *Clin Exp Metastasis* 32(5), 457–465. DOI: 10.1007/s10585-015-9719-0
52. Sargent, D.; Sobrero, A.; Grothey, A.; *et al.* (2009). Evidence for cure by adjuvant therapy in colon cancer: Observations based on individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol* 27(6), 872–877. DOI: 10.1200/JCO.2008.19.5362
53. Lawler, M.; Johnston, B.; van Schaeybroeck, S.; *et al.* (Elsevier, 2020). in *Abeloff's Clinical Oncology* (eds. Niederhuber, J. E.; Armitage, J. O.; Doroshow, J. H.; *et al.*) 1219–1280
54. Lee, G. H.; Malietzis, G.; Askari, A.; *et al.* (2015) Is right-sided colon cancer different to left-sided colorectal cancer? - A systematic review. *Eur J Surg Oncol* 41(3), 300–308. DOI: 10.1016/j.ejso.2014.11.001
55. Guinney, J.; Dienstmann, R.; Wang, X.; *et al.* (2015). The consensus molecular subtypes of colorectal cancer. *Nat Med* 21(11), 1350–1356. DOI: 10.1038/nm.3967
56. Seo, S. I.; Lim, S. B.; Yoon, Y. S.; *et al.* (2013). Comparison of recurrence patterns between ≤ 5 years and > 5 years after curative operations in colorectal cancer patients. *J Surg Oncol* 108(1), 9–13. DOI: 10.1002/jso.23349

57. van Cutsem, E.; Cervantes, A.; Adam, R.; *et al.* (2016). ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol* 27(8), 1386–1422. DOI: 10.1093/annonc/mdw235
58. Qiu, M.; Hu, J.; Yang, D.; *et al.* (2015). Pattern of distant metastases in colorectal cancer: a SEER based study. *Oncotarget* 6(36), 38658–38666. DOI: 10.18632/oncotarget.6130
59. Lee-Ying, R.; Bernard, B.; Gresham, G.; *et al.* (2017). A Comparison of Survival by Site of Metastatic Resection in Metastatic Colorectal Cancer. *Clin Colorectal Cancer* 16(2), e23–e28. DOI: 10.1016/j.clcc.2016.07.018
60. Margalit, O.; Shacham-Shmueli, E.; Lawrence, Y. R.; *et al.* (2019). Lung Metastasis Predicts Better Prognosis in Metastatic Colorectal Cancer With Mutated KRAS. *Clin Colorectal Cancer* 18(3), e300–e307. DOI: 10.1016/j.clcc.2019.06.001
61. Yaeger, R.; Chatila, W. K.; Lipsyc, M. D.; *et al.* (2018). Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. *Cancer Cell* 33(1), 125–136.e3. DOI: 10.1016/j.ccell.2017.12.004
62. Dang, H. X.; Krasnick, B. A.; White, B. S.; *et al.* (2020). The clonal evolution of metastatic colorectal cancer. *Sci. Adv* 6(24), eaay9691. DOI: 10.1126/sciadv.aay9691
63. del Carmen, S.; Sayagués, J. M.; Bengoechea, O.; *et al.* (2018). Spatio-temporal tumor heterogeneity in metastatic CRC tumors: a mutational-based approach. *Oncotarget* 9(76), 34279–34288. DOI: 10.18632/oncotarget.26081
64. Hu, Z.; Ding, J.; Ma, Z.; *et al.* (2019). Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat Genet* 51(7), 1113–1122. DOI: 10.1038/s41588-019-0423-x
65. Xu, W.; He, Y.; Wang, Y.; *et al.* (2020). Risk factors and risk prediction models for colorectal cancer metastasis and recurrence: an umbrella review of systematic reviews and meta-analyses of observational studies. *BMC Med* 18(1), 172. DOI: 10.1186/s12916-020-01618-6
66. Li, Q. H.; Wang, Y. Z.; Tu, J.; *et al.* (2020). Anti-EGFR therapy in metastatic colorectal cancer: Mechanisms and potential regimens of drug resistance. *Gastroenterol Rep (Oxf)* 8(3), 179–191. DOI: 10.1093/gastro/goaa026
67. Kanat, O. & Ertas, H. (2019). Existing anti-angiogenic therapeutic strategies for patients with metastatic colorectal cancer progressing following first-line bevacizumab-based therapy. *World J Clin Oncol* 10(2), 52–61. DOI: 10.5306/wjco.v10.i2.52
68. Damilakis, E.; Mavroudis, D.; Sfakianaki, M.; *et al.* (2020). Immunotherapy in metastatic colorectal cancer: Could the latest developments hold the key to improving patient survival? *Cancers (Basel)* 12(4), 889. DOI: 10.3390/cancers12040889
69. Trullas, A.; Delgado, J.; Genazzani, A.; *et al.* (2021). The EMA assessment of pembrolizumab as monotherapy for the first-line treatment of adult patients with

- metastatic microsatellite instability-high or mismatch repair deficient colorectal cancer. *ESMO Open* 6(3), 100145. DOI: 10.1016/j.esmoop.2021.100145
70. Königsberg, R.; Hulla, W.; Klimpfinger, M.; *et al.* (2012). Clinical and economic aspects of KRAS mutational status as predictor for epidermal growth factor receptor inhibitor therapy in metastatic colorectal cancer patients. *Oncology* 81(5-6), 359–364. DOI: 10.1159/000334919
 71. Canon, J.; Rex, K.; Saiki, A. Y.; *et al.* (2019). The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* 575(7781), 217–223. DOI: 10.1038/s41586-019-1694-1
 72. Fell, J. B.; Fischer, J. P.; Baer, B. R.; *et al.* (2020). Identification of the Clinical Development Candidate MRTX849, a Covalent KRAS^{G12C} Inhibitor for the Treatment of Cancer. *J Med Chem* 63(13), 6679–6693. DOI: 10.1021/acs.jmedchem.9b02052
 73. Schirripa, M.; Nappo, F.; Cremolini, C.; *et al.* (2020). KRAS G12C Metastatic Colorectal Cancer: Specific Features of a New Emerging Target Population. *Clin Colorectal Cancer* 19(3), 219–225. DOI: 10.1016/j.clcc.2020.04.009
 74. Grothey, A.; Fakih, M. & Tabernero, J. (2021). Management of BRAF-mutant metastatic colorectal cancer: a review of treatment options and evidence-based guidelines. *Ann Oncol* 32(8), 959-967. DOI: 10.1016/j.annonc.2021.03.206
 75. Kopetz, S.; Grothey, A.; Yaeger, R.; *et al.* (2019). Encorafenib, Binimetinib, and Cetuximab in BRAF V600E–Mutated Colorectal Cancer. *N Engl J Med* 381(17), 1632–1643. DOI: 10.1056/nejmoa1908075
 76. Tabernero, J.; Grothey, A.; van Cutsem, E.; *et al.* (2021). Encorafenib plus cetuximab as a new standard of care for previously treated BRAF V600E-mutant metastatic colorectal cancer: Updated survival results and subgroup analyses from the BEACON study. *J Clin Oncol* 39(4), 273–284. DOI: 10.1200/JCO.20.02088
 77. Takegawa, N.; Tsurutani, J.; Kawakami, H.; *et al.* (2019). [fam-] trastuzumab deruxtecan, antitumor activity is dependent on HER2 expression level rather than on HER2 amplification. *Int J Cancer* 145(12), 3414–3424. DOI: 10.1002/ijc.32408
 78. Wang, G.; He, Y.; Sun, Y.; *et al.* (2020). Prevalence, prognosis and predictive status of HER2 amplification in anti-EGFR-resistant metastatic colorectal cancer. *Clin Transl Oncol* 22(6), 813–822. DOI: 10.1007/s12094-019-02213-9
 79. Rimassa, L.; Bozzarelli, S.; Pietrantonio, F.; *et al.* (2019). Phase II Study of Tivantinib and Cetuximab in Patients With KRAS Wild-type Metastatic Colorectal Cancer With Acquired Resistance to EGFR Inhibitors and Emergence of MET Overexpression: Lesson Learned for Future Trials With EGFR/MET Dual Inhibition. *Clin Colorectal Cancer* 18(2), 125-132.e2. DOI: 10.1016/j.clcc.2019.02.004
 80. Hanahan, D. & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell* 144(5), 646–674. DOI: 10.1016/j.cell.2011.02.013
 81. Seebacher, N. A.; Stacy, A. E.; Porter, G. M.; *et al.* (2019). Clinical development of targeted and immune based anti-cancer therapies. *J Exp Clin Cancer Res* 38(1), 156. DOI: 10.1186/s13046-019-1094-2

82. Henry, N. L. & Hayes, D. F. (2012). Cancer biomarkers. *Mol Oncol* 6(2), 140–146. DOI: 10.1016/j.molonc.2012.01.010
83. Chung, C. (2021). Predictive and prognostic biomarkers with therapeutic targets in colorectal cancer: A 2021 update on current development, evidence, and recommendation. *J Onco Pharm Pract* 28(4), 860-869. DOI: 10.1177/10781552211005525
84. Oh, H.-H. & Joo, Y.-E. (2020). Novel biomarkers for the diagnosis and prognosis of colorectal cancer. *Intest Res* 18(2), 168–183. DOI: 10.5217/ir.2019.00080
85. Sung, P. J. & Luger, S. M. (2017). Minimal Residual Disease in Acute Myeloid Leukemia. *Curr Treat Options Oncol* 18(1), 1. DOI: 10.1007/s11864-017-0447-3
86. Patil, P. P.; Jafa, E. & Aggarwal, M. (2021). Minimal Residual Disease in Acute Lymphoblastic Leukemia. *Ind J Med Paediatr Oncol* 42, 71–76. DOI: 10.1055/s-0041-1729730
87. Kothari, S.; Hillengass, J.; McCarthy, P. L.; *et al.* (2019). Determination of Minimal Residual Disease in Multiple Myeloma: Does It Matter? *Curr Hematol Malig Rep* 14(1), 39–46. DOI: 10.1007/s11899-019-0497-7
88. Hu, J.; Hu, B.; Gui, Y. C.; *et al.* (2019). Diagnostic value and clinical significance of methylated SEPT9 for colorectal cancer: A meta-analysis. *Med Sci Monit* 25, 5813–5822. DOI: 10.12659/MSM.915472
89. Hodroj, K.; Barthelemy, D.; Lega, J.-C.; *et al.* (2021). Issues and limitations of available biomarkers for fluoropyrimidine-based chemotherapy toxicity, a narrative review of the literature. *ESMO Open* 6(3), 100125. DOI: 10.1016/j.esmoop.2021.100125
90. Campbell, J. M.; Stephenson, M. D.; Bateman, E.; *et al.* (2017). Irinotecan-induced toxicity pharmacogenetics: An umbrella review of systematic reviews and meta-analyses. *Pharmacogenomics J* 17(1), 21–28. DOI: 10.1038/tpj.2016.58
91. Niedzwiecki, D.; Frankel, W. L.; Venook, A. P.; *et al.* (2016). Association Between Results of a Gene Expression Signature Assay and Recurrence-Free Interval in Patients With Stage II Colon Cancer in Cancer and Leukemia Group B 9581 (Alliance). *J Clin Oncol* 34(25), 3047–3053. DOI: 10.1200/JCO.2015.65.4699
92. Gray, R. G.; Quirke, P.; Handley, K.; *et al.* (2011). Validation study of a quantitative multigene reverse transcriptase- polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer. *J Clin Oncol* 29(35), 4611–4619. DOI: 10.1200/JCO.2010.32.8732
93. Cocco, E.; Scaltriti, M. & Drilon, A. (2018). NTRK fusion-positive cancers and TRK inhibitor therapy. *Nat Rev Clin Oncol* 15(12), 731–747. DOI: 10.1038/s41571-018-0113-0
94. Sargent, D. J.; Marsoni, S.; Monges, G.; *et al.* (2010). Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *J Clin Oncol* 28(20), 3219–3226. DOI: 10.1200/JCO.2009.27.1825

95. Ribic, C. M.; Sargent, D. J.; Moore, M. J.; *et al.* (2003). Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. *N Engl J Med* 349(3), 247–257. DOI: 10.1056/NEJMoa022289
96. Chou, A.; Fraser, T.; Ahadi, M.; *et al.* (2020). NTRK gene rearrangements are highly enriched in MLH1/PMS2 deficient, BRAF wild-type colorectal carcinomas—a study of 4569 cases. *Mod Pathol* 33(5), 924–932. DOI: 10.1038/s41379-019-0417-3
97. Yamashiro, Y.; Kurihara, T.; Hayashi, T.; *et al.* (2021). NTRK fusion in Japanese colorectal adenocarcinomas. *Sci Rep* 11(1), 5635. DOI: 10.1038/s41598-021-85075-y
98. Goetz, L. H. & Schork, N. J. (2018). Personalized medicine: motivation, challenges, and progress. *Fertil Steril* 109(6), 952–963. DOI: 10.1016/j.fertnstert.2018.05.006
99. di Sanzo, M.; Cipolloni, L.; Borro, M.; *et al.* (2017). Clinical Applications of Personalized Medicine: A New Paradigm and Challenge. *Curr Pharm Biotechnol* 18(3), 194–203. DOI: 10.2174/1389201018666170224105600
100. Binefa, G.; Rodríguez-Moranta, F.; Teule, À.; *et al.* (2014). Colorectal cancer: From prevention to personalized medicine. *World J Gastroenterol* 20(22), 6786–6808. DOI: 10.3748/wjg.v20.i22.6786
101. Poulet, G.; Massias, J. & Taly, V. (2019). Liquid Biopsy: General Concepts. *Acta Cytol* 63(6), 449–455. DOI: 10.1159/000499337
102. Strickler, J. H.; Loree, J. M.; Ahronian, L. G.; *et al.* (2018). Genomic landscape of cell-free DNA in patients with colorectal cancer. *Cancer Discov* 8(2), 164–173. DOI: 10.1158/2159-8290.CD-17-1009
103. Pantel, K. & Alix-Panabières, C. (2019). Liquid biopsy and minimal residual disease — latest advances and implications for cure. *Nat Rev Clin Oncol* 16(7), 409–424. DOI: 10.1038/s41571-019-0187-3
104. Millán-Esteban, D.; Reyes-García, D.; García-Casado, Z.; *et al.* (2018). Suitability of melanoma FFPE samples for NGS libraries: Time and quality thresholds for downstream molecular tests. *Biotechniques* 65(2), 79–85. DOI: 10.2144/BTN-2018-0016
105. Nagahashi, M.; Shimada, Y.; Ichikawa, H.; *et al.* (2017). Formalin-fixed paraffin-embedded sample conditions for deep next generation sequencing. *J Surg Res* 220, 125–132. DOI: 10.1016/j.jss.2017.06.077
106. Zhou, B.; Xu, K.; Zheng, X.; *et al.* (2020). Application of exosomes as liquid biopsy in clinical diagnosis. *Signal Transduct Target Ther* 5(1), 144. DOI: 10.1038/s41392-020-00258-9
107. Mouliere, F.; Mair, R.; Chandrananda, D.; *et al.* (2018). Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients. *EMBO Mol Med* 10(12), e9323. DOI: 10.15252/emmm.201809323

108. Seoane, J.; de Mattos-Arruda, L.; le Rhun, E.; *et al.* (2019). Cerebrospinal fluid cell-free tumour DNA as a liquid biopsy for primary brain tumours and central nervous system metastases. *Ann Oncol* 30(2), 211–218. DOI: 10.1093/annonc/mdy544
109. Li, N.; Liu, Y.; Duan, J.; *et al.* (2019). Prognostic significance of molecular characteristics of cerebrospinal fluid for non-small cell lung cancer patients with leptomeningeal metastasis. *Thorac Cancer* 10(8), 1673–1682. DOI: 10.1111/1759-7714.13123
110. Oshi, M.; Murthy, V.; Takahashi, H.; *et al.* (2021). Urine as a source of liquid biopsy for cancer. *Cancers (Basel)* 13(11), 2652. DOI: 10.3390/cancers13112652
111. Arasaradnam, R. P.; Mcfarlane, M. J.; Ryan-Fisher, C.; *et al.* (2014). Detection of colorectal cancer (CRC) by urinary volatile organic compound analysis. *PLoS One* 9(9), e108750. DOI: 10.1371/journal.pone.0108750
112. Crisafulli, G.; Mussolin, B.; Cassingena, A.; *et al.* (2019). Whole exome sequencing analysis of urine trans-renal tumour DNA in metastatic colorectal cancer patients. *ESMO Open* 4(6), 3000572. DOI: 10.1136/esmoopen-2019-000572
113. Kim, E. R.; Kwon, H. N.; Nam, H.; *et al.* (2019). Urine-NMR metabolomics for screening of advanced colorectal adenoma and early stage colorectal cancer. *Sci Rep* 9(1), 1–10. DOI: 10.1038/s41598-019-41216-y
114. Jiang, P.; Sun, K.; Peng, W.; *et al.* (2020). Plasma DNA end-motif profiling as a fragmentomic marker in cancer, pregnancy, and transplantation. *Cancer Discov* 10(5), 664–673. DOI: 10.1158/2159-8290.CD-19-0622
115. Rostami, A.; Lambie, M.; Yu, C. W.; *et al.* (2020). Senescence, Necrosis, and Apoptosis Govern Circulating Cell-free DNA Release Kinetics. *Cell Rep* 31(13), 107830. DOI: 10.1016/j.celrep.2020.107830
116. Snyder, M. W.; Kircher, M.; Hill, A. J.; *et al.* (2016). Cell-free DNA Comprises an *In Vivo* Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* 164(1-2), 57–68. DOI: 10.1016/j.cell.2015.11.050
117. Alborelli, I.; Generali, D.; Jermann, P.; *et al.* (2019). Cell-free DNA analysis in healthy individuals by next-generation sequencing: a proof of concept and technical validation study. *Cell Death Dis* 10(7), 534. DOI: 10.1038/s41419-019-1770-3
118. Yao, W.; Mei, C.; Nan, X.; *et al.* (2016). Evaluation and comparison of *in vitro* degradation kinetics of DNA in serum, urine and saliva: A qualitative study. *Gene* 590(1), 142–148. DOI: 10.1016/j.gene.2016.06.033
119. Meddeb, R.; Dache, Z. A. A.; Thezenas, S.; *et al.* (2019). Quantifying circulating cell-free DNA in humans. *Sci Rep* 9(1), 5220. DOI: 10.1038/s41598-019-41593-4
120. Mouliere, F.; Chandrananda, D.; Piskorz, A. M.; *et al.* (2018). Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* 10(466), eaat4921. DOI: 10.1126/scitranslmed.aat4921
121. Bonnomet, A.; Brysse, A.; Tachsidis, A.; *et al.* (2010). Epithelial-to-mesenchymal transitions and circulating tumor cells. *J Mammary Gland Biol Neoplasia* 15(2), 261–273. DOI: 10.1007/s10911-010-9174-0

122. Lowes, L. E. & Allan, A. L. (Elsevier Inc., 2018). in *Advances in Clinical Chemistry* (eds Makowski, G. S.) 83, 121-181. DOI: 10.1016/bs.acc.2017.10.004
123. Xie, X.; Wang, L.; Wang, X.; *et al.* (2021). Evaluation of Cell Surface Vimentin Positive Circulating Tumor Cells as a Diagnostic Biomarker for Lung Cancer. *Front Oncol* 11, 672687. DOI: 10.3389/fonc.2021.672687
124. Chen, L.; Bode, A. M. & Dong, Z. (2017). Circulating tumor cells: Moving biological insights into detection. *Theranostics* 7(10), 2606–2619. DOI: 10.7150/thno.18588
125. Becker, A.; Thakur, B. K.; Weiss, J. M.; *et al.* (2016). Extracellular Vesicles in Cancer: Cell-to-Cell Mediators of Metastasis. *Cancer Cell* 30(6), 836–848. DOI: 10.1016/j.ccell.2016.10.009
126. Steinbichler, T. B.; Dudás, J.; Riechelmann, H.; *et al.* (2017). The role of exosomes in cancer metastasis. *Semin Cancer Biol* 44, 170–181. DOI: 10.1016/j.semcancer.2017.02.006
127. Wubbolts, R.; Leckie, R. S.; Veenhuizen, P. T. M.; *et al.* (2003). Proteomic and biochemical analyses of human B cell-derived exosomes: Potential implications for their function and multivesicular body formation. *J Biol Chem* 278(13), 10963–10972. DOI: 10.1074/jbc.M207550200
128. Junqueira-Neto, S.; Batista, I. A.; Costa, J. L.; *et al.* (2019). Liquid Biopsy beyond Circulating Tumor Cells and Cell-Free DNA. *Acta Cytol* 63(6), 479–488. DOI: 10.1159/000493969
129. Silva, J.; Garcia, V.; Rodriguez, M.; *et al.* (2012). Analysis of exosome release and its prognostic value in human colorectal cancer. *Genes Chromosomes Cancer* 51(4), 409–418. DOI: 10.1002/gcc.21926
130. Melo, S. A.; Luecke, L. B.; Kahlert, C.; *et al.* (2015). Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* 523(7559), 177–182. DOI: 10.1038/nature14581
131. Mitchell, P. S.; Parkin, R. K.; Kroh, E. M.; *et al.* (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* 105(30), 10513–10518. DOI: 10.1073/pnas.0804549105
132. Umu, S. U.; Langseth, H.; Bucher-Johannessen, C.; *et al.* (2018). A comprehensive profile of circulating RNAs in human serum. *RNA Biol* 15(2), 242–250. DOI: 10.1080/15476286.2017.1403003
133. in 't Veld, S. G. J. G. & Wurdinger, T. (2019). Tumor-educated platelets. *Blood* 133(22), 2359–2364. DOI: 10.1182/blood-2018-12-852830
134. Asghar, S.; Parvaiz, F. & Manzoor, S. (2019). Multifaceted role of cancer educated platelets in survival of cancer cells. *Thromb Res* 177, 42–50. DOI: 10.1016/j.thromres.2019.02.026
135. Naderi-Meshkin, H. & Ahmadiankia, N. (2018). Cancer metastasis versus stem cell homing: Role of platelets. *J Cell Physiol* 233(12), 9167–9178. DOI: 10.1002/jcp.26937

136. You, W.; Yan, L.; Cai, Z.; *et al.* (2020). Clinical Significances of Positive Postoperative Serum CEA and Post-preoperative CEA Increment in Stage II and III Colorectal Cancer: A Multicenter Retrospective Study. *Front Oncol* 10, 671. DOI: 10.3389/fonc.2020.00671
137. Kuang, J.; Gong, Y.; Xie, H.; *et al.* (2020). The prognostic value of preoperative serum CA724 for CEA-normal colorectal cancer patients. *PeerJ* 8, e8936. DOI: 10.7717/peerj.8936
138. Phallen, J.; Sausen, M.; Adleff, V.; *et al.* (2017). Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med* 9(403), ean2415. DOI: 10.1126/scitranslmed.aan2415
139. Wan, J. C. M.; Heider, K.; Gale, D.; *et al.* (2020). ctDNA monitoring using patient-specific sequencing and integration of variant reads. *Sci Transl Med* 12(548), eaaz8084. DOI: 10.1126/scitranslmed.aaz8084
140. Papadopoulos, N. (Springer, 2020). in *Tumor Liquid Biopsy. Recent results in cancer research* (eds. Schaffner, F.; Merlin, J. L.; von Bubnoff, N.). 215, 163–180. DOI: 10.1007/978-3-030-26439-0_9
141. Gilson, P. (Springer, 2020). in *Tumor Liquid Biopsy. Recent results in cancer research* (eds. Schaffner, F.; Merlin, J. L.; von Bubnoff, N.). 215, 181–211. DOI: 10.1007/978-3-030-26439-0_10
142. Reinert, T.; Henriksen, T. V.; Christensen, E.; *et al.* (2019). Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients with Stages I to III Colorectal Cancer. *JAMA Oncol* 5(8), 1124–1131. DOI: 10.1001/jamaoncol.2019.0528
143. Egyud, M.; Tejani, M.; Pennathur, A.; *et al.* (2019). Detection of Circulating Tumor DNA in Plasma: A Potential Biomarker for Esophageal Adenocarcinoma. *Ann Thorac Surg* 108(2), 343–349. DOI: 10.1016/j.athoracsur.2019.04.004
144. Ståhlberg, A.; Krzyzanowski, P. M.; Egyud, M.; *et al.* (2017). Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing. *Nat Protoc* 12(4), 664–682. DOI: 10.1038/nprot.2017.006
145. Lo, Y. M. D. (Humana Press, 2003) in *Clinical Applications of PCR. Methods in Molecular Medicine* (eds. Lo, Y. M. D.). 16, 61–70. DOI: 10.1385/0-89603-499-2:61
146. Liu, Q.; Nguyen, V. Q.; Li, X.; *et al.* (2006). Multiplex dosage pyrophosphorolysis-activated polymerization: Application to the detection of heterozygous deletions. *Biotechniques* 40(5), 661–668. DOI: 10.2144/000112164
147. Moulriere, F.; el Messaoudi, S.; Pang, D.; *et al.* (2014). Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol Oncol* 8(5), 927–941. DOI: 10.1016/j.molonc.2014.02.005
148. Zhang, X.; Chang, N.; Yang, G.; *et al.* (2017). A comparison of ARMS-Plus and droplet digital PCR for detecting EGFR activating mutations in plasma. *Oncotarget* 8(67), 112014–112023. DOI: 10.18632/oncotarget.22997

149. Li, J.; Wang, L.; Mamon, H.; *et al.* (2008). Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. *Nat Med* 14(5), 579–584. DOI: 10.1038/nm1708
150. Milbury, C. A.; Li, J. & Makrigiorgos, G. M. (2009). PCR-based methods for the enrichment of minority alleles and mutations. *Clin Chem* 55(4), 632–640. DOI: 10.1373/clinchem.2008.113035
151. How-Kit, A. & Tost, J. (Springer New York, 2015). in *Pyrosequencing: Methods and Protocols* (eds. Lehmann, U. & Tost, J.) 83–101. DOI: 10.1007/978-1-4939-2715-9_7
152. How-Kit, A.; Lebbé, C.; Bousard, A.; *et al.* (2014). Ultrasensitive detection and identification of BRAF V600 mutations in fresh frozen, FFPE, and plasma samples of melanoma patients by E-ice-COLD-PCR. *Anal Bioanal Chem* 406(22), 5513–5520. DOI: 10.1007/s00216-014-7975-5
153. Diehl, F.; Li, M.; He, Y.; *et al.* (2006). BEAMing: Single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods* 3(7), 551–559. DOI: 10.1038/nmeth898
154. Hindson, C. M.; Chevillet, J. R.; Briggs, H. A.; *et al.* (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods* 10(10), 1003–1005. DOI: 10.1038/nmeth.2633
155. Lee, J. Y.; Qing, X.; Xiumin, W.; *et al.* (2016). Longitudinal monitoring of EGFR mutations in plasma predicts outcomes of NSCLC patients treated with EGFR TKIs: Korean Lung Cancer Consortium (KLCC-12-02). *Oncotarget* 7(6), 6984–6993. DOI: 10.18632/oncotarget.6874
156. Oxnard, G. R.; Paweletz, C. P.; Kuang, Y.; *et al.* (2014). Noninvasive Detection of Response and Resistance in EGFR -Mutant Lung Cancer Using Quantitative Next-Generation Genotyping of Cell-Free Plasma DNA. *Clin Cancer Res* 20(6), 1698–1705. DOI: 10.1158/1078-0432.CCR-13-2482
157. Tebaldi, M. & Salvi, S. (Springer New York, 2019). in *Cell-free DNA as Diagnostic Markers: Methods and Protocols* (eds. Casadio, V. & Salvi, S.) 119–125. DOI: 10.1007/978-1-4939-8973-7_9
158. Rolfo, C.; Cardona, A. F.; Cristofanilli, M.; *et al.* (2020). Challenges and opportunities of cfDNA analysis implementation in clinical practice: Perspective of the International Society of Liquid Biopsy (ISLB). *Crit Rev Oncol Hematol* 151, 102978. DOI: 10.1016/j.critrevonc.2020.102978
159. el Achi, H.; Khoury, J. D. & Loghavi, S. (2019). Liquid Biopsy by Next-Generation Sequencing: a Multimodality Test for Management of Cancer. *Curr Hematol Malig Rep* 14(5), 358–367. DOI: 10.1007/s11899-019-00532-w
160. Luca, G. de & Dono, M. (2021). The Opportunities and Challenges of Molecular Tagging Next - Generation Sequencing in Liquid Biopsy. *Mol Diagn Ther.* 25(5), 537-547. DOI: 10.1007/s40291-021-00542-6

161. Deveson, I. W.; Gong, B.; Lai, K.; *et al.* (2021). Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. *Nat Biotechnol* 39(9), 1115–1128. DOI: 10.1038/s41587-021-00857-z
162. Ma, X.; Shao, Y.; Tian, L.; *et al.* (2019). Analysis of error profiles in deep next-generation sequencing data. *Genome Biol* 20(1), 1–15. DOI: 10.1186/s13059-019-1659-6
163. Kinde, I.; Wu, J.; Papadopoulos, N.; *et al.* (2011). Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* 108(23), 9530–9535. DOI: 10.1073/pnas.1105422108
164. Newman, A. M.; Lovejoy, A. F.; Klass, D. M.; *et al.* (2016). Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol* 34(5), 547–555. DOI: 10.1038/nbt.3520
165. Schmitt, M. W.; Kennedy, S. R.; Salk, J. J.; *et al.* (2012). Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A* 109(36), 14508–14513. DOI: 10.1073/pnas.1208715109
166. Wang, T. T.; Abelson, S.; Zou, J.; *et al.* (2019). High efficiency error suppression for accurate detection of low-frequency variants. *Nucleic Acids Res* 47(15), e87. DOI: 10.1093/nar/gkz474
167. Calapre, L.; Giardina, T.; Robinson, C.; *et al.* (2019). Locus-specific concordance of genomic alterations between tissue and plasma circulating tumor DNA in metastatic melanoma. *Mol Oncol* 13(2), 171–184. DOI: 10.1002/1878-0261.12391
168. Keup, C.; Benyaa, K.; Hauch, S.; *et al.* (2020). Targeted deep sequencing revealed variants in cell-free DNA of hormone receptor-positive metastatic breast cancer patients. *Cell Mol Life Sci* 77(3), 497–509. DOI: 10.1007/s00018-019-03189-z
169. Crowgey, E. L.; Mahajan, N.; Wong, W. H.; *et al.* (2020). Error-corrected sequencing strategies enable comprehensive detection of leukemic mutations relevant for diagnosis and minimal residual disease monitoring. *BMC Med Genomics* 13(1), 1–11. DOI: 10.1186/s12920-020-0671-8
170. Kukita, Y.; Matoba, R.; Uchida, J.; *et al.* (2015). High-fidelity target sequencing of individual molecules identified using barcode sequences: De novo detection and absolute quantitation of mutations in plasma cell-free DNA from cancer patients. *DNA Res* 22(4), 269–277. DOI: 10.1093/dnares/dsv010
171. Kukita, Y.; Ohkawa, K.; Takada, R.; *et al.* (2018). Selective identification of somatic mutations in pancreatic cancer cells through a combination of next-generation sequencing of plasma DNA using molecular barcodes and a bioinformatic variant filter. *PLoS One* 13(2), e0192611. DOI: 10.1371/journal.pone.0192611
172. Dunwell, T. L.; Dailey, S. C.; Ottestad, A. L.; *et al.* (2021). Adaptor Template Oligo-Mediated Sequencing (ATOM-Seq) is a new ultra-sensitive UMI-based NGS library preparation technology for use with cfDNA and cfrRNA. *Sci Rep* 11(1), 3138. DOI: 10.1038/s41598-021-82737-9

173. Yeh, Y. M.; Lin, P. C.; Lee, C. T.; *et al.* (2020). Treatment monitoring of colorectal cancer by integrated analysis of plasma concentration and sequencing of circulating tumor DNA. *Mol Cancer* 19, 150. DOI: 10.1186/s12943-020-01273-8
174. Chin, Y. M.; Takahashi, Y.; Chan, H. T.; *et al.* (2021). Ultradeep targeted sequencing of circulating tumor DNA in plasma of early and advanced breast cancer. *Cancer Sci* 112(1), 454–464. DOI: 10.1111/cas.14697
175. Masunaga, N.; Kagara, N.; Motooka, D.; *et al.* (2018). Highly sensitive detection of ESR1 mutations in cell-free DNA from patients with metastatic breast cancer using molecular barcode sequencing. *Breast Cancer Res Treat* 167(1), 49–58. DOI: 10.1007/s10549-017-4487-y
176. Ståhlberg, A.; Krzyzanowski, P. M.; Jackson, J. B.; *et al.* (2016). Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. *Nucleic Acids Res* 44(11), e105. DOI: 10.1093/nar/gkw224
177. Smith, T.; Heger, A. & Sudbery, I. (2017). UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* 27(3), 491–499. DOI: 10.1101/gr.209601.116
178. Orabi, B.; Erhan, E.; McConeghy, B.; *et al.* (2019). Alignment-free clustering of UMI tagged DNA molecules. *Bioinformatics* 35(11), 1829–1836. DOI: 10.1093/bioinformatics/bty888
179. Sater, V.; Vially, P. J.; Vially, P. J.; *et al.* (2020). UMI-VarCal: A new UMI-based variant caller that efficiently improves low-frequency variant detection in paired-end sequencing NGS libraries. *Bioinformatics* 36(9), 2718–2724. DOI: 10.1093/bioinformatics/btaa053
180. Chen, S.; Zhou, Y.; Chen, Y.; *et al.* (2019). Gencore: An efficient tool to generate consensus reads for error suppressing and duplicate removing of NGS data. *BMC Bioinformatics* 20(Suppl 23), 606. DOI: 10.1186/s12859-019-3280-9
181. Newman, A. M.; Bratman, S. v, To, J.; *et al.* (2014). An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* 20(5), 548–554. DOI: 10.1038/nm.3519
182. Kamps-Hughes, N.; McUsic, A.; Kurihara, L.; *et al.* (2018). ERASE-Seq: Leveraging replicate measurements to enhance ultralow frequency variant detection in NGS data. *PLoS One* 13(4), e0195272. DOI: 10.1371/journal.pone.0195272
183. Deng, S.; Lira, M.; Huang, D.; *et al.* (2018). TNER: A novel background error suppression method for mutation detection in circulating tumor DNA. *BMC Bioinformatics* 19(1), 387. DOI: 10.1186/s12859-018-2428-3
184. Forshew, T.; Murtaza, M.; Parkinson, C.; *et al.* (2012). Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* 4(136), 136ra68. DOI: 10.1126/scitranslmed.3003726
185. Plagnol, V.; Woodhouse, S.; Howarth, K.; *et al.* (2018). Analytical validation of a next generation sequencing liquid biopsy assay for high sensitivity broad molecular profiling. *PLoS One* 13(3), e0193802. DOI: 10.1371/journal.pone.0193802

186. Razavi, P.; Li, B. T.; Brown, D. N.; *et al.* (2019). High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat Med* 25(12), 1928–1937. DOI: 10.1038/s41591-019-0652-7
187. Chin, Y. M.; Takahashi, Y.; Chan, H. T.; *et al.* (2021). Ultradeep targeted sequencing of circulating tumor DNA in plasma of early and advanced breast cancer. *Cancer Sci* 112(1), 454–464. DOI: 10.1111/cas.14697
188. Park, S. J. & Bejar, R. (2020). Clonal hematopoiesis in cancer. *Exp Hematol* 83, 105–112. DOI: 10.1016/j.exphem.2020.02.001
189. Gibson, C. J. & Steensma, D. P. (2018). New insights from studies of clonal hematopoiesis. *Clin Cancer Res* 24(19), 4633–4642. DOI: 10.1158/1078-0432.CCR-17-3044
190. Jaiswal, S.; Fontanillas, P.; Flannick, J.; *et al.* (2014). Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N Engl J Med* 371(26), 2488–2498. DOI: 10.1056/nejmoa1408617
191. Bolton, K. L.; Ptashkin, R. N.; Gao, T.; *et al.* (2020). Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat Genet* 52(11), 1219–1226. DOI: 10.1038/s41588-020-00710-0
192. Parikh, A. R.; van Seventer, E. E.; Siravegna, G.; *et al.* (2021). Minimal Residual Disease Detection using a Plasma-only Circulating Tumor DNA Assay in Patients with Colorectal Cancer. *Clin Cancer Res* 27(20), 5586–5594. DOI: 10.1158/1078-0432.ccr-21-0410
193. Im, Y. R.; Tsui, D. W. Y.; Diaz, L. A.; *et al.* (2021). Next-Generation Liquid Biopsies: Embracing Data Science in Oncology. *Trends Cancer* 7(4), 283–292. DOI: 10.1016/j.trecan.2020.11.001
194. Schiffman, J. D.; Fisher, P. G. & Gibbs, P. (2015). Early Detection of Cancer: Past, Present, and Future. *Am Soc Clin Oncol Educ Book*, 57–65. DOI: 10.14694/EdBook_AM.2015.35.57
195. Zhang, Y.; Yao, Y.; Xu, Y.; *et al.* (2021). Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. *Nat Commun* 12(1), 11. DOI: 10.1038/s41467-020-20162-8
196. Ørntoft, M. B. W.; Jensen, S. Ø.; Øgaard, N.; *et al.* (2021). Age-stratified reference intervals unlock the clinical potential of circulating cell-free DNA as a biomarker of poor outcome for healthy individuals and patients with colorectal cancer. *Int J Cancer* 148(7), 1665–1675. DOI: 10.1002/ijc.33434
197. Chen, M. & Zhao, H. (2019). Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Hum Genomics* 13(1), 34. DOI: 10.1186/s40246-019-0220-8
198. Lion, T. (1999). Minimal residual disease. *Curr Opin Hematol* 6(6), 406–411. DOI: 10.1097/00062752-199911000-00009

199. Garcia-Murillas, I.; Schiavon, G.; Weigelt, B.; *et al.* (2015). Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* 7(302), 302ra133. DOI: 10.1126/scitranslmed.aab0021
200. Garcia-Murillas, I.; Chopra, N.; Comino-Méndez, I.; *et al.* (2019). Assessment of Molecular Relapse Detection in Early-Stage Breast Cancer. *JAMA Oncol* 5(10), 1473–1478. DOI: 10.1001/jamaoncol.2019.1838
201. Leal, A.; van Grieken, N. C. T.; Palsgrove, D. N.; *et al.* (2020). White blood cell and cell-free DNA analyses for detection of residual disease in gastric cancer. *Nat Commun* 11(1), 525. DOI: 10.1038/s41467-020-14310-3
202. Vollbrecht, C.; Lehmann, A.; Lenze, D.; *et al.* (2018). Validation and comparison of two NGS assays for the detection of EGFR T790M resistance mutation in liquid biopsies of NSCLC patients. *Oncotarget* 9 (26), 18529–18539. DOI: 10.18632/oncotarget.24908
203. Parikh, A. R. & Corcoran, R. B. (2018). Monitoring resistance through liquid biopsy. *Ann Oncol* 29(1), 8–11. DOI: 10.1093/annonc/mdx650
204. Razavi, P.; Dickler, M. N.; Shah, P. D.; *et al.* (2020). Alterations in PTEN and ESR1 promote clinical resistance to alpelisib plus aromatase inhibitors. *Nat Cancer* 1(4), 382–393. DOI: 10.1038/s43018-020-0047-1
205. Parikh, A. R.; Leshchiner, I.; Elagina, L.; *et al.* (2019). Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med* 25, 1415–1421. DOI: 10.1038/s41591-019-0561-9
206. Marusyk, A.; Janiszewska, M. & Polyak, K. (2020). Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* 37(4), 471–484. DOI: 10.1016/j.ccell.2020.03.007
207. Siravegna, G.; Mussolin, B.; Buscarino, M.; *et al.* (2015). Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* 21(7), 795–801. DOI: 10.1038/nm.3870
208. Dentre, S. C.; Leshchiner, I.; Haase, K.; *et al.* (2021). Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* 184(8), 2239–2254.e39. DOI: 10.1016/j.cell.2021.03.009
209. Martínez-Jiménez, F.; Muiños, F.; Sentís, I.; *et al.* (2020). A compendium of mutational cancer driver genes. *Nat Rev Cancer* 20(10), 555–572. DOI: 10.1038/s41568-020-0290-x
210. Dentre, S. C.; Wedge, D. C. & van Loo, P. (2017). Principles of Reconstructing the Subclonal Architecture of Cancers. *Cold Spring Harb Perspect Med* 7(8), a026625. DOI: 10.1101/cshperspect.a026625
211. Zaccaria, S. & Raphael, B. J. (2020). Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. *Nat Commun* 11(1), 4301. DOI: 10.1038/s41467-020-17967-y
212. Woolston, A.; Khan, K.; Spain, G.; *et al.* (2019). Genomic and Transcriptomic Determinants of Therapy Resistance and Immune Landscape Evolution during Anti-

- EGFR Treatment in Colorectal Cancer. *Cancer Cell* 36(1), 35-50.e9. DOI: 10.1016/j.ccell.2019.05.013
213. Gremel, G.; Lee, R. J.; Girotti, M. R.; *et al.* (2016). Distinct subclonal tumour responses to therapy revealed by circulating cell-free DNA. *Ann Oncol* 27(10), 1959–1965. DOI: 10.1093/annonc/mdw278
214. Parikh, A. R.; Leshchiner, I.; Elagina, L.; *et al.* (2019). Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med* 25(9), 1415–1421. DOI: 10.1038/s41591-019-0561-9
215. Ahlborn, L. B.; Tuxen, I. V.; Mouliere, F.; *et al.* (2018). Circulating tumor DNA as a marker of treatment response in BRAF V600E mutated non-melanoma solid tumors. *Oncotarget* 9(66), 32570–32579. DOI: 10.18632/oncotarget.25948
216. Bos, M. K.; Angus, L.; Nasserinejad, K.; *et al.* (2020). Whole exome sequencing of cell-free DNA – A systematic review and Bayesian individual patient data meta-analysis. *Cancer Treat Rev* 83, 101951. DOI: 10.1016/j.ctrv.2019.101951
217. Aguilar-Mahecha, A.; Lafleur, J.; Brousse, S.; *et al.* (2021). Early, on-treatment levels and dynamic changes of genomic instability in circulating tumor dna predict response to treatment and outcome in metastatic breast cancer patients. *Cancers (Basel)* 13(6), 1–12. DOI: 10.3390/cancers13061331
218. Chen, X.; Chang, C. W.; Spoerke, J. M.; *et al.* (2019). Low-pass whole-genome sequencing of circulating cell-free DNA demonstrates dynamic changes in genomic copy number in a squamous lung cancer clinical cohort. *Clin Cancer Res* 25(7), 2254–2263. DOI: 10.1158/1078-0432.CCR-18-1593
219. Zviran, A.; Schulman, R. C.; Shah, M.; *et al.* (2020). Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. *Nat Med* 26(7), 1114–1124. DOI: 10.1038/s41591-020-0915-3
220. García-Olmo, D. C.; Peiró-Pastor, R.; Picazo, M. G.; *et al.* (2020). Liquid biopsy by NGS: Differential presence of exons (DPE) is related to metastatic potential in a colon-cancer model in the rat. *Transl Oncol* 13(11), 100837. DOI: 10.1016/j.tranon.2020.100837
221. Olmedillas-López, S.; García-Olmo, D. C.; García-Arranz, M.; *et al.* (2018). Liquid biopsy by NGS: differential presence of exons (DPE) in cell-free DNA reveals different patterns in metastatic and nonmetastatic colorectal cancer. *Cancer Med* 7(5), 1706–1716. DOI: 10.1002/cam4.1399
222. Marcuello, M.; Vymetalkova, V.; Neves, R. P. L.; *et al.* (2019). Circulating biomarkers for early detection and clinical management of colorectal cancer. *Mol Aspects Med* 69, 107–122. DOI: 10.1016/j.mam.2019.06.002
223. Vidal, J.; Muinelos, L.; Dalmases, A.; *et al.* (2017). Plasma ctDNA RAS mutation analysis for the diagnosis and treatment monitoring of metastatic colorectal cancer patients. *Ann Oncol* 28(6), 1325–1332. DOI: 10.1093/annonc/mdx125
224. Hazar-Rethinam, M.; Kleyman, M.; Han, G. C.; *et al.* (2018). Convergent Therapeutic Strategies to Overcome the Heterogeneity of Acquired Resistance in BRAF V600E

- Colorectal Cancer. *Cancer Discov* 8(4), 417–427. DOI: 10.1158/2159-8290.CD-17-1227
225. Tarazona, N.; Gimeno-Valiente, F.; Gambardella, V.; *et al.* (2019). Targeted next-generation sequencing of circulating-tumor DNA for tracking minimal residual disease in localized colon cancer. *Ann Oncol* 30(11), 1804–1812. DOI: 10.1093/annonc/mdz390
226. Tie, J.; Wang, Y.; Tomasetti, C.; *et al.* (2016). Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med* 8(346), 346ra92. DOI: 10.1126/scitranslmed.aaf6219
227. Tie, J.; Cohen, J. D.; Wang, Y.; *et al.* (2019). Circulating tumor DNA analyses as markers of recurrence risk and benefit of adjuvant therapy for stage III colon cancer. *JAMA Oncol* 5(12), 1710–1717. DOI: 10.1001/jamaoncol.2019.3616
228. Wong, R.; Tie, J.; Lee, M.; *et al.* (2019). The potential role of circulating tumor DNA (ctDNA) in the further investigation of colorectal cancer patients with nonspecific findings on standard investigations. *Int J Cancer* 145(2), 540–547. DOI: 10.1002/ijc.32117
229. Rothwell, D. G.; Ayub, M.; Cook, N.; *et al.* (2019). Utility of ctDNA to support patient selection for early phase clinical trials: the TARGET study. *Nat Med* 25(5), 738–743. DOI: 10.1038/s41591-019-0380-z
230. Taniguchi, H.; Nakamura, Y.; Kotani, D.; *et al.* (2021). CIRCULATE-Japan: Circulating tumor DNA–guided adaptive platform trials to refine adjuvant therapy for colorectal cancer. *Cancer Sci* 112(7), 2915–2920. DOI: 10.1111/cas.14926
231. Lonardi, S.; Montagut, C.; Pietrantonio, F.; *et al.* (2020). The PEGASUS trial: Post-surgical liquid biopsy-guided treatment of stage III and high-risk stage II colon cancer patients. *J Clin Oncol* 38(15). DOI: 10.1200/JCO.2020.38.15_suppl.TPS4124
232. Siravegna, G.; Mussolin, B.; Venesio, T.; *et al.* (2019). How liquid biopsies can change clinical practice in oncology. *Ann Oncol* 30(10), 1580–1590. DOI: 10.1093/annonc/mdz227
233. Bando, H.; Nakamura, Y.; Kotani, D.; *et al.* (2021). Comparing GOZILA and COLOMATE: Ongoing Umbrella/Basket Trials Examining Genetic Testing in Gastrointestinal Malignancies. *Oncology (Williston Park)* 35(7), 382–389. DOI: 10.46883/ONC.2021.3507.0382
234. Tarazona, N. (2019). A multi-omic liquid biopsy-based signature as a valuable tool to assess minimal residual disease in localised colorectal cancer. at <<https://roderic.uv.es/handle/10550/70814>>
235. Fenizia, F.; Pasquale, R.; Roma, C.; *et al.* (2018) Measuring tumor mutation burden in non-small cell lung cancer: Tissue versus liquid biopsy. *Transl Lung Cancer Res* 7(6), 668–677. DOI: 10.21037/tlcr.2018.09.23
236. Tamborero, D.; Rubio-Perez, C.; Deu-Pons, J.; *et al.* (2018). Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med* 10(1), 25. DOI: 10.1186/s13073-018-0531-8

237. Chakravarty, D.; Gao, J.; Phillips, S.; *et al.* (2017). OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* 2017, PO.17.00011. DOI: 10.1200/PO.17.00011
238. Suzuki, T.; Suzuki, T.; Yoshimura, Y.; *et al.* (2020). Detection of circulating tumor DNA in patients of operative colorectal and gastric cancers. *Oncotarget* 11(34), 3198–3207. DOI: 10.18632/oncotarget.27682
239. Yurgelun, M. B.; Kulke, M. H.; Fuchs, C. S.; *et al.* (2017). Cancer susceptibility gene mutations in individuals with colorectal cancer. *J Clinl Oncol* 35(10), 1086–1095. DOI: 10.1200/JCO.2016.71.0012
240. Bailey, M. H.; Tokheim, C.; Porta-Pardo, E.; *et al.* (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* 173(2), 371–385.e18. DOI: 10.1016/j.cell.2018.02.060
241. Tokunaga, R.; Xiu, J.; Goldberg, R. M.; *et al.* (2020). The impact of ARID1A mutation on molecular characteristics in colorectal cancer. *Eur J Cancer* 140, 119–129. DOI: 10.1016/j.ejca.2020.09.006
242. Kawai, T.; Nyuya, A.; Mori, Y.; *et al.* (2021). Clinical and epigenetic features of colorectal cancer patients with somatic POLE proofreading mutations. *Clin Epigenetics* 13(1), 117. DOI: 10.1186/s13148-021-01104-7
243. Tepus, M. & Yau, T. O. (2020). Non-Invasive Colorectal Cancer Screening: An Overview. *Gastrointest Tumors* 7(3), 62–73. DOI: 10.1159/000507701
244. Allegretti, M.; Cottone, G.; Carboni, F.; *et al.* (2020). Cross-sectional analysis of circulating tumor DNA in primary colorectal cancer at surgery and during post-surgery follow-up by liquid biopsy. *J Exp Clin Cancer Res* 39(1), 69. DOI: 10.1186/s13046-020-01569-z
245. Myint, N. N. M.; Verma, A. M.; Fernandez-Garcia, D.; *et al.* (2018). Circulating tumor DNA in patients with colorectal adenomas: assessment of detectability and genetic heterogeneity. *Cell Death Dis* 9(9), 894. DOI: 10.1038/s41419-018-0934-x
246. Kim, Y. W.; Kim, Y. H.; Song, Y.; *et al.* (2019). Monitoring circulating tumor DNA by analyzing personalized cancer-specific rearrangements to detect recurrence in gastric cancer. *Exp Mol Med* 51(8), 1–10. DOI: 10.1038/s12276-019-0292-5
247. Abbosh, C.; Birkbak, N. J. & Swanton, C. (2018). Early stage NSCLC — challenges to implementing ctDNA-based screening and MRD detection. *Nat Rev Clin Oncol* 15(9), 577–586. DOI: 10.1038/s41571-018-0058-3
248. Zink, F.; Stacey, S. N.; Norddahl, G. L.; *et al.* (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* 130(6), 742–752. DOI: 10.1182/blood-2017-02-769869
249. Liu, J.; Chen, X.; Wang, J.; *et al.* (2019). Biological background of the genomic variations of cf-DNA in healthy individuals. *Ann Oncol* 30(3), 464–470. DOI: 10.1093/annonc/mdy513
250. Duan, H.; Hu, J. L.; Chen, Z. H.; *et al.* (2020). Assessment of circulating tumor DNA in cerebrospinal fluid by whole exome sequencing to detect genomic alterations of

- glioblastoma. *Chin Med J (Engl)* 133(12), 1415–1421. DOI: 10.1097/CM9.0000000000000843
251. Toledo, R. A.; Garralda, E.; Mitsi, M.; *et al.* (2018). Exome sequencing of plasma DNA portrays the mutation landscape of colorectal cancer and discovers mutated VEGFR2 receptors as modulators of antiangiogenic therapies. *Clin Cancer Res* 24(15), 3550–3559. DOI: 10.1158/1078-0432.CCR-18-0103
 252. Hastings, R. K.; Openshaw, M. R.; Vazquez, M.; *et al.* (2021). Longitudinal whole-exome sequencing of cell-free DNA for tracking the co-evolutionary tumor and immune evasion dynamics: longitudinal data from a single patient. *Ann Oncol* 32(5), 681–684. DOI: 10.1016/j.annonc.2021.02.007
 253. Butler, T. M.; Johnson-Camacho, K.; Peto, M.; *et al.* (2015). Exome sequencing of cell-free DNA from metastatic cancer patients identifies clinically actionable mutations distinct from primary disease. *PLoS One* 10(8), e0136407. DOI: 10.1371/journal.pone.0136407
 254. Hong, W.; Li, A.; Liu, Y.; *et al.* (2022). Clonal Hematopoiesis Mutations in Patients with Lung Cancer Are Associated with Lung Cancer Risk Factors. *Cancer Res* 82(2), 199–209. DOI: 10.1158/0008-5472.CAN-21-1903
 255. Lu, J.-Y.; Jin, P.; Gao, W.; *et al.* (2017). Estrogen enhances mismatch repair by induction of MLH1 expression via estrogen receptor- β . *Oncotarget* 8(24): 38767-38779. DOI: 10.18632/oncotarget.16351
 256. Liu, D. (2016). Gene signatures of estrogen and progesterone receptor pathways predict the prognosis of colorectal cancer. *FEBS J* 283(16), 3115–3133. DOI: 10.1111/febs.13798
 257. Abancens, M.; Bustos, V.; Harvey, H.; *et al.* (2020) Sexual Dimorphism in Colon Cancer. *Front Oncol* 10, 607909. DOI: 10.3389/fonc.2020.607909
 258. Triki, M.; Lapiere, M.; Cavailles, V.; *et al.* (2017). Expression and role of nuclear receptor coregulators in colorectal cancer. *World J Gastroenterol* 23(25), 4480-4490. DOI: 10.3748/wjg.v23.i25.4480
 259. Li, Y.; Duan, Q. & Tan, Y. (2021). A pan-cancer analysis of MUC family genes as potential biomarkers for immune checkpoint therapy. *J Clin Oncol* 39, 2598–2598. DOI: 10.1200/JCO.2021.39.15_suppl.2598
 260. Li, C.; Zuo, D.; Liu, T.; *et al.* (2019) Prognostic and Clinicopathological Significance of MUC Family Members in Colorectal Cancer: A Systematic Review and Meta-Analysis. *Gastroenterol Res Pract* 2019, 2391670. DOI: 10.1155/2019/2391670
 261. Bodmer, W. F.; Cottrell, S.; Frischauf, A. M.; *et al.* (1989). Genetic analysis of colorectal cancer. *Princess Takamatsu Symp* 20, 49–59
 262. Chowell, D.; Morris, L. G. T.; Grigg, C. M.; *et al.* (2018). Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* 359(6375), 582–587. DOI: 10.1126/science.aao4572

263. Castro, A.; Ozturk, K.; Pyke, R. M.; *et al.* (2019). Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Med Genomics* 12(Suppl 6), 107. DOI: 10.1186/s12920-019-0544-1
264. Reinert, T.; Schøler, L. v.; Thomsen, R.; *et al.* (2016). Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut* 65(4), 625–634. DOI: 10.1136/gutjnl-2014-308859
265. Zhao, X.; Dai, F.; Mei, L.; *et al.* (2021). The Potential Use of Dynamics Changes of ctDNA and cfDNA in the Perioperative Period to Predict the Recurrence Risk in Early NSCLC. *Front Oncol* 11, 671963. DOI: 10.3389/fonc.2021.671963
266. Chen, K.; Zhao, H.; Shi, Y.; *et al.* (2019). Perioperative dynamic changes in circulating tumor DNA in patients with lung cancer (Dynamic). *Clin Cancer Res* 25(23), 7058–7067. DOI: 10.1158/1078-0432.CCR-19-1213
267. Fleming, C. A.; O’Leary, D. P.; Wang, J.; *et al.* (2020). Association of Observed Perioperative Cell-Free DNA Dynamics With Early Recurrence in Patients With Colon Cancer. *JAMA Surg* 155(2), 168-170. DOI: 10.1001/jamasurg.2019.4706
268. Merkle, J.; Daka, A.; Deppe, A. C.; *et al.* (2019). High levels of cell-free DNA accurately predict late acute kidney injury in patients after cardiac surgery. *PLoS One* 14(6), e0218548. DOI: 10.1371/journal.pone.0218548
269. Jordan, P. R.; O’Leary, D. P.; Fleming, C. A.; *et al.* (2021). The relevance of liquid biopsy in surgical oncology: The application of perioperative circulating nucleic acid dynamics in improving patient outcomes. *The Surgeon* 20(4), e163-e173 DOI: 10.1016/j.surge.2021.06.006
270. Dasari, A.; Morris, V. K.; Allegra, C. J.; *et al.* (2020). ctDNA applications and integration in colorectal cancer: an NCI Colon and Rectal–Anal Task Forces whitepaper. *Nat Rev Clin Oncol* 17(12), 757–770. DOI: 10.1038/s41571-020-0392-0
271. Masfarré, L.; Vidal, J.; Fernández-Rodríguez, C.; *et al.* (2021). ctDNA to Guide Adjuvant Therapy in Localized Colorectal Cancer (CRC). *Cancers (Basel)* 13(12), 2869. DOI: 10.3390/cancers13122869
272. Hirotsu, Y.; Otake, S.; Ohyama, H.; *et al.* (2020). Dual-molecular barcode sequencing detects rare variants in tumor and cell free DNA in plasma. *Sci Rep* 10(1), 3391. DOI: 10.1038/s41598-020-60361-3
273. Filges, S.; Yamada, E.; Ståhlberg, A.; *et al.* (2019). Impact of Polymerase Fidelity on Background Error Rates in Next-Generation Sequencing with Unique Molecular Identifiers/Barcodes. *Sci Rep* 9(1), 3503. DOI: 10.1038/s41598-019-39762-6
274. Tie, J.; Wang, Y.; Cohen, J.; *et al.* (2021). Circulating tumor DNA dynamics and recurrence risk in patients undergoing curative intent resection of colorectal cancer liver metastases: A prospective cohort study. *PLoS Med* 18(5), e1003620. DOI: 10.1371/journal.pmed.1003620

8. ANEXOS

8.1. ANEXO I: Listado de cebadores específicos

Tabla 26: Listado de cebadores específicos empleados en la tesis.

Amplicón	Cebador directo	Cebador reverso
APC_p.V1320/E1322	AAGAAAAGATTGGAAGTGGTACAGC	TGGTTCTAGGGTGTGTGAC
APC_p.Y1031*	GATAATGATGGAGAAGTACACCA	CTTTGCCTCCAGAGTTCAACT
ARID1A_p.344+350	ACATGGCCTCGCAGTGTG	CTCCTTTGTTGGGCCCTC
ARID1A_p.757+764	CCCACAAATAGGTTATATGCAGAGG	GACGCGGAGATAAGGCTGAG
ARID1A_p.F2141Sfs*5 9	ATGTGGACCTGATTCTGGCC	ACTGAGGAAGCGCACCATAG
ARID1A_p.P293S	CTCAGCAGCGCTTCGGG	CGACGTGAGCAGTTGGTTG
ARID1A_p.P550L	GACGACACAGCAGCACCC	CTGGCTGTGAGGTTGCTG
ARID1A_p.Q379P	GGGGCCCAACAAAGGAGC	CCCTCCCCTCAGCTGT
ARID1A_p.Q758Rfs*7 5	TTTGTTCGCATTGTATAAAGCTAATAAC	GGGAAGTGTACTGGGCATC
ARID1A_p.R1721Pfs* 6	GCTCCCAGGGTTGCTAGAG	CCACCTCATACTCCTTAAAATGCC
ARID1A_p.S11Afs*90	GGATCATGGCCGCGCAG	TGCTGCTCGGCTTCTTCAG
BRAF_p.G30/G32	TGTTCAACGGGGACATGGAG	CACTCACCTCCTCCGAATG
BRAF_p.G6=	GGCTCTCGTTATAAGATGGCG	ATGTCCCGTTGAACAGAGC
BRAF_p.V600E	GAAGACCTCACAGTAAAAATAGG	CAAAATGGATCCAGACAAC
CTNNB1_p.A149P	CCCCTAATGTCCAGCGTTTG	CAGTTCAGGGATTGCACGTG
CTNNB1_p.S37_G38d elinsCC	AAGCGGCTGTTAGTCACTGG	CCTTACCCTCAGAGAAGGAGC
CTNNB1_p.S45F	CTGGCAGCAACAGTCTTACCT	TCCACATCCTTCTCTCAGG
EGFR_delex18_HPD	AAGGTGAGAAAGTAAAAATCCCGTC	TCCTTGTGGCTTTCGGAGATG
EGFR_delex18_UMI1	GTTAAAATCCCGTCGCTATCAAG	AGCAAAGCAGAAACTCACATCG
EGFR_delex18_UMI2	TGAGAAAGTAAAAATCCCGTCGC	AGCAAAGCAGAAACTCACATCG
EGFR_insex20	CTGACGTGCCTCTCCCTCC	GCACACGTGGGGTTGTGTC
EGFR_p.P753=	TCCCGTCGCTATCAAGGAATTAAG	GGACCCCAACAGCAAAAG
EGFR_pL858R_UMI1	GCATGTCAAGATCACAGATTTGGG	TTCTTCTCTCCGACCCAG
EGFR_pL858R_UMI2	CAGCATGTCAAGATCACAGATTTTG	TTCTTCTCTCCGACCCAG
EGFR_p.T790M	CATCTGCCTCACCTCCACCG	GTCTTTGTGTTCCCGACATAGTC
ERBB3_c.2938-1G>T	AGGAGGAGTTGGCTGGAACC	GCTATTCCAGGCCACTCTC
ERBB3_p.D297Y	TGTCCTCTCTCTCCCTAG	GCCCATTTTATCTACTTCCATCTG
ERBB3_p.S869G	TGTCGCCTCTGATGATAAGC	ATCCCCATGCTTCACTCCAC
ERBB4_p.H809N	GTTACTCAACTTATGCCCCATGG	CACCAGTTAAGCAGCAGTTGTG
ERBB4_p.M831/L834	AAACCTGGCATTAACTGAATATCA	GCTGCCAAATCCCGATGAAC
ERBB4_p.N626=	TGCATGTCACTGTGGATGGAG	CCGTCCATGGGTAGTAAATGCAG
FAT4_p.A4555=	AGAAAAAGGGAAGTGAACGTT	GGTTTTGGGTTCCCTCAGG
FAT4_p.D276A	TGAGGACGCGTTGTGGG	GGTCCCCTCGCTGCAG
FAT4_p.E1007*	TGACCAACTCTTATGAAGTCACC	ACCATTTGCTCTGAATCCTTATC
FAT4_p.E3293*	AAGGTGTAGCTTGCCTACTG	TTAGGAGCTGCTTCTGAGATTC
FAT4_p.E3467Dfs*2	CGCAGACAGGACAGATCACC	ACTGAGAGATTATAGATGGGAAGGG
FAT4_p.F2536Gfs*9	GCCGCCGACCACTAAAC	TGGAATGAGCCCACTCTTTTAC
FAT4_p.G4257=	TTTTGGTGTGAACAGTCTGGAAG	AGTAGTGAATTGCTGCTTCTTGG
FAT4_p.I3057Yfs*8	GGGTGGATTTTCACTAGCATCCTC	TCCCTTATCCTTTGCAGTGACTG
FAT4_p.K205E	CATCACCTGAACCCGAGC	TTGTCTCCACTCAACCAG
FAT4_p.L2134=	TCTAACAGTGGTGGCTACAGAC	TGTCAAGTACCATAACTACAACCTC
FAT4_p.L4591I	CCCAAAACAGATATCATTGAAAGGG	ATGGTTTCTGAGTTGTGAGGAATATC
FAT4_p.P774=	CAGCTTATCAGTTGCAAAATAGTAGC	GTGGGTTGTCTTGTGAGTGCC
FAT4_p.R175H	ATCGGCTCAAACGGTGTGG	CTCGGGTTCAGGGTGTGTC

Amplicón	Cebador directo	Cebador reverso
FAT4_p.R336=	GCAGGGCATGGACAGAGG	CCGGGTCATTGTCCATTACG
FAT4_p.R4106=	ATTTATTTAGGACTCTTGATGTTTCAGC	TCCACGTGTCCTCTTCTCTG
FAT4_p.S4282=	TGTTTTACAGATTAAGAATGGCAAAGT	TATTTCTCTCCACTTTCCACAGCA
FBXW7_p.A677Efs*18	GGGAGTTGTGTGGCGGATC	CCCATTCCGACTCCCAACTG
FBXW7_p.D520Y	TGATGGCAGGAGGGTTGTTAG	CAGTCTCTGGATCCACACC
FBXW7_p.R465H	ACACACCTTATATGGGCATACTTCC	CCCAACCATGACAAGATTTTCCC
FBXW7_p.R479Q	TTTCTGTTTCTCCCTCTGCAGA	CACTGGCCTGTCTCAATATCCC
FLNA_p.L140F	TCACAGACAGCAAGGCCATC	CCTCCTCGTCCCACATGG
FLNA_p.T1508=	TGGACGTGGTAGACAACGCT	ACTGAGATGCTGTAGGGCCC
KMT2C_p.A30=	CCCCGAGGAGCCTGGAG	GAAAGGGGAAGCGCCATCT
KMT2C_p.F4496Lfs*2 1	ACTTCACCTGCGCCATTAAGC	TTCCCTTTGGTTTGTGCATGG
KMT2C_p.I575=	ATGTCAACGGTCAGGAGTCC	TCTTCAAAAAACAGAGTATATCCATTAACA
KRAS_p.8-14	TTATAAGGCCTGCTGAAAATGACTG	CTGAATTAGCTGTATCGTCAAGGC
KRAS_p.A146T	AGACACAAAACAGGCTCAGGAC	TGTATTTATTTAGTGTACTTACCTGTC
KRAS_p.Q61L	ACCTGTCTCTGGATATTCTCGACAC	GAAAGCCCTCCCCAGTCTCT
KRAS_pG12D_HPD	TGCTGAAAATGACTGAATATAAACTTGTG G	GAATTAGCTGTATCGTCAAGGCAC
KRAS_pG12D_UMI1	GCTGAAAATGACTGAATATAAACTTGTG GT	GAATTAGCTGTATCGTCAAGGCAC
NRAS_p.G12	AGGTTCTTGCTGGTGTGAAATGAC	GTGGTTCTGGATTAGCTGGATTGTC
NRAS_p.Q61	TGAAACCTGTTTGTGGACATACTG	TACACAGAGGAAGCCTTCGC
NRAS_pA59T_UMI	ACAAGTGGTTATAGATGGTGAAC	TCTCATGGCACTGTACTCTTC
PIK3CA_p.C420R	GAAGAAAAGTGTTTGAAATGTGTTTT	AGTTTATATTTCCCATGCCAATG
PIK3CA_p.D64Y	AGCAAGAAAATACCCCTCCATC	CCCTTCTGCTTCTTGAGTAACAC
PIK3CA_p.E542/E545	GACAAAGAACAGCTCAAAGCAA	TTAGCACTTACTGTGACTCCA
PIK3CA_p.H1047R	CAAGAGGCTTTGGAGTATTTTCATG	AAGATCCAATCCATTTTGTGTGCC
PIK3CA_p.K111/L113	TCAACCTTTTTAAAAGTAATTGAACCAG	AATAGGATATTGTATCATACCAATTTCTCG
PIK3CA_p.N1068Kfs*5	CATCATGGTGGCTGGACAAC	TCCAGAGTGAGCTTTCATTTTCTC
PIK3CA_p.Y1021C	GATGCTTGGCTCTGGAATGC	CCTCTTGGCTCAGTTTTATCTAAGGC
PIK3CA_pE545K_HPD	AAGCAATTTCTACACGAGATCCTC	TTACCTGTGACTCCATAGAAAATCTTTC
PIK3CA_pE545K_UMI 1	CTCAAAGCAATTTCTACACGAGA	TGTGACTCCATAGAAAATCTTCTCC
POLE_c.1360-8C>T	CACTTGACCCTGGGCTCTTG	ACAGAATACGTGGCCAGAGTC
POLE_p.G1923=	CTCTCGATGCTGGGAATTTCTTC	ACCAGTCCACAGTGAATACGAG
POLE_p.G963S	TCAGGTATGCTGTGTTCAATGAAG	CGAGGATTGGAAGATCTTAATCAGC
POLE_p.L1235=	CCTCGTAAAGCTGCCTCACC	CAGTCGGCGTGAGGTCCTG
PPP2R1A_p.R183W	ACTTCCGGAACTGTGCTC	TTGACGTTGTCCAGTCCAG
PTEN_p.I101L	AAGACCATAACCCACCACAGC	CATCTTCACTTAGCCATTGGTCAAG
PTEN_p.R15*	TCAAAGAGATCGTTAGCAGAAACA	AGAAAGGTAAAGAGGAGCAGCC
RNF43_c.-12_35del	CAGTATGTATGGTTGAAGTGCATTG	CCTGCAGGGTAGCCATCAG
RNF43_p.I48T+rs	TGAAAGATCAGCAGAACAGAAAGC	CCAGCAAACACACCTTCAAAG
RNF43_p.R117	GCATCGTCAAGCTGGAGAGTC	GGGCGAAGTGTGAGTCTACC
RNF43_p.R392Efs*27	ATCCAGGAGCCAGGCAT	CATAGGGGTGCTGGGCTC
SMAD2_p.G379S	ACCTGCAACAGTGTGTAATAATTC	AAGCAAGTTGACATGATAGGTTTATG
SMAD2_p.L446I	CTTGCTGGATTGAATTCATCTGA	CAACGCACTGAAGGGGATCC
SMAD2_p.S464*	TCTACAGTGGTTGGACAAAGTATTAAC	GGGACTTGATTGGTGAAGCTTT
SMAD4_p.351-363	GTTCTTCAAGCTGCCCTATTG	TGGACATTGGAGAGTTGACCC
SMAD4_p.D537del	AGCATCAAAGAAACACCTTGCTG	GTTGTGGGTCTGCAATCGG
SMAD4_p.F339V	AGCTCTGAGTATTGGTGTCC	GTAAACAATAGGGCAGCTGAAGG
SMAD4_p.K428*	GAGAAGCTGGGCGTGAC	TGTAACCTTATATATGCACTTGGG
SMAD4_p.R189C	ACCATCCAGCATCCACCAAG	GACTCAGATGGGGCTAACAGAG
SMAD4_p.S32*	ATAGTTTGTATGTCCATAGACAAGG	TTCTCTTCTCAGCTTCTTACCAAAC
SMAD4_p.Y301F	CCAAAACGGCCATCTTCAGC	GCCCTTACAACAAAAACAGAGC

Amplificón	Cebador directo	Cebador reverso
SOX9_p.E388*	CGCACACGCTGACCACG	CTGCTCGTGTAGTGGCTG
SOX9_p.F270/R271	CCTGAAGCGAGAGGGCC	GAGATGACGTGCTGCTCAG
SOX9_p.N45H+p.P48Efs*201	CCCCACCATGTCCGAGGA	TCCTTCTCAGATCGGGCTC
SOX9_p.P426Sfs*45	CCCCAACAGATCGCTACAG	GGTGGTCGGTGTAGTCGTAC
SOX9_p.Q505Afs*72	CTTCCATCCCAGACCC	CCTTCGTGGGAGGCCTC
SOX9_p.S391fs*213	AGCCCCACCATGTCCGAG	CGCGTGTCTCGGTGTCC
TCF7L2_c.616+3A>C	AAACCCACCTCCACACTTACC	CAGGACCAGCACTCTACCTTG
TCF7L2_p.L336V	AGAAAAGAAGAAGCCCCACATAAAG	GCTACGACCTTTGCTCTCATTTT
TCF7L2_p.Q390fs*37	CCAGAGAAGAGCAAGCGAAATAC	CAGCCGGGGTACAGTTGC
TCF7L2_p.R402Q	ATATGCAACTGTACCCCGGC	ACAGCTCTACAGACTCCAATCC
TGFBR2_p.R528C	TGAGACGTTGACTGAGTGCTG	AGCCTGTCCAGATGCTCCAG
TP53_c.75-1G>C	TCCATGGGACTGACTTTCTGC	GCCCAACCTTGTCTTACC
TP53_c.782+1G>C	CCCATCCTCACCATCATCACAC	GGTCAGAGGCAAGCAGAGG
TP53_p.173-176	CTACAAGCAGTCACAGCATG	GCTGCTCACCATCGCTATCTG
TP53_p.234-245	GCTCTGACTGTACCACCATCC	GTGATGATGGTGAGGATGGGC
TP53_p.266-275	TGCTTCTCTTTTCTATCCTGAGTAG	TCTGTGCCCGGTCTCTC
TP53_p.E298*	GGCGCACAGAGGAAGAGAATC	TGCTTGCTTACCTCGCTTAGTG
TP53_p.G105V	GCCCTTGCATCTTCTGTCC	AAGAAGCCAGACGAAACC
TP53_p.G199/E204	CCTCAGCATCTTATCCGAGTGG	AGGCGGCTCATAGGGCAC
TP53_p.I195Kfs*53	GGCCTCTGATTCCTCACTGATTG	TCCACACGCAAATTTCTTCC
TP53_p.R248/P250	CCACCTCCACTACAACACTATG	CCAGTGTGATGATGGTGAGG
TP53_p.R282/E285	CTTTGAGGTGCGTGTGTGTGC	GGCTCCCCTTTCTTGCGG
TP53_p.R342*	TTCTCCCCCTCTCTGTTGC	CCTGGGCATCCTTGAGTTCC
TP53_p.S127/L130	CAACTCTGTCTCTTCTCTTCC	GAATCAACCCACAGCTGCAC
TP53_p.S15Rfs30*	CAGTCAGATCCTAGCGTCGAG	GCCCTTCCAATGGATCCACTC
TP53_p.T211/Y220	GGAAGGAAATTTGCGTGTGGAG	TTGCAAACCAGACCTCAGGC
TP53_p.V157/A159	TGTGCAGCTGTGGTTGATTC	TGCTGTGACTGCTTGTAGATGG
TP53_p.W53R	GCAATGGATGATTTGATGCTGTCC	TCTGGGAGCTTCATCTGGACC

8.2. ANEXO II: Listado de adaptadores

Tabla 27: Listado de adaptadores empleados en la tesis.

Tipo	Secuencia
Nextera Directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-Directo
Nextera Reverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-Reverso
Nextera BC5'	AATGATACGGCGACCACCGAGATCTACACXXXXXXXXTCGTCGGCAGCGT C
Nextera BC3'	CAAGCAGAAGACGGCATAACGAGATXXXXXXXXGTCTCGTGGGCTCGG
Nextera BC5'L	AATGATACGGCGACCACCGAGATCTACACXXXXXXXXTCGTCGGCAGCGT CAGATGTG
Nextera BC3'L	CAAGCAGAAGACGGCATAACGAGATXXXXXXXXGTCTCGTGGGCTCGGAG ATGTG
Truseq 5'	AATGATACGGCGACCACCGAGATCTACACXXXXXXXXACACTTTCCCTA CACGACGCTCTCCGATCT
Truseq 3'	CAAGCAGAAGACGGCATAACGAGATNNNNNNNNNNXXXXXXXXGTGACTG GAGTTCAGACGTGTGCTCTCCGATC T
Nextera UMI 0P	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNN-Directo
Nextera UMI 2P	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNAG- Directo
Nextera UMI 3P	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNCAG- Directo
HPD1	GGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNCTGGC TGCCGACGACC-Directo
HPD2	GGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNTGCCT GCCGACGACC-Directo
HPD3	GGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNGCGTG CCGACGACC-Directo
HPD4	GGATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNGCGTG GCCGACGATCC-Directo
HPD5	GGGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNGCGTG GCCGACGACCC-Directo
HPD6	GGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNTGC CTGCCGACGACC-Directo

8.3. ANEXO III: Listado de sondas de ddPCR

Tabla 28: Listado de sondas de PCR digital empleados en la tesis. CDS: Coding DNA Sequence.

Gen	Variante	Posición hg38	CDS	ID Sonda
ACVR2A	p.H304R	chr2:147918541	A>G	dHsaMDS254185447
ACVR2A	p.R400C	chr2:147923093	C>T	dHsaMDS490220689
ACVR2A	p.V408Gfs*2	chr2:147926036	GTA>G	dHsaMDS355099155
AKT1	p.Q79R	chr14:104776710	T>C	dHsaMDS985787291
AMER1	p.R586*	chrX:64191531	G>A	dHsaMDS738051030
AMER1	p.R497*	chrX:64191798	G>A	dHsaMDS698651556
AMER1	p.R353*	chrX:64192230	G>A	dHsaMDS695093712
AMER1	p.F173Lfs*36	chrX:64192767	TA>T	dHsaMDS439470130
APC	p.R216*	chr5:112792446	C>T	dHsaMDS2511334
APC	p.E241Sfs*10	chr5:112792519	CAG>C	dHsaMDS677144709
APC	p.R283*	chr5:112815507	C>T	dHsaMDS2511652
APC	p.L472Tfs*13	chr5:112827106	A>AG	dHsaMDS981882556
APC	p.R564*	chr5:112828919	C>T	dHsaMDS913746157
APC	c.1958+3A>G	chr5:112835168	A>G	dHsaMDS561532702
APC	p.S678Vfs*40	chr5:112837624	TC>T	dHsaMDS244920112
APC	p.N813Ffs*2	chr5:112838030	CAA>C	dHsaMDS421650228
APC	p.R876*	chr5:112838220	C>T	dHsaMDS444452080
APC	p.Q1062*	chr5:112838773	TAAAC>T	dHsaMDS522639417
APC	p.R1114*	chr5:112838934	C>T	dHsaMDS657757205
APC	p.Q1244*	chr5:112839324	C>T	dHsaMDS385873727
APC	p.E1286*	chr5:112839450	G>T	dHsaMDS2510676
APC	p.Q1294*	chr5:112839474	C>T	dHsaMDS2515338
APC	p.E1295*	chr5:112839477	G>T	dHsaMDS562376414
APC	p.E1309*	chr5:112839519	G>T	dHsaMDS2515252
APC	p.E1322*	chr5:112839558	G>T	dHsaMDS2514046
APC	p.H1349Ifs*21	chr5:112839636	AGGCACAA AGCTGTT>A	dHsaMDS309440556
APC	p.Q1367*	chr5:112839693	C>T	dHsaMDS2511856
APC	p.D1394Gfs*2	chr5:112839773	T>TG	dHsaMDS885201106
APC	p.N1455Ifs*18	chr5:112839953	TA>T	dHsaMDS883051030
APC	p.D1486Vfs*21	chr5:112840050	GA>G	dHsaMDS946792127
ARID1A	p.P550L	chr1:26731450	C>T	dHsaMDS418949883
ARID1A	p.Q758Rfs*75	chr1:26762167	AC>A	dHsaMDS652643153
BRAF	p.V600E	chr7:140753336	A>T	dHsaMDV2010027
CTNNB1	p.S37_G38delin sCC	chr3:41224622	CTG>GCT	dHsaMDS779488754
CTNNB1	p.S45F	chr3:41224646	C>T	dHsaMDV2010117
ERBB3	c.2938-1G>T	chr12:56099837	G>T	dHsaMDS736075161
ERBB4	p.N626=	chr2:211657822	G>A	dHsaMDS153154737
FAT4	p.P774=	chr4:125318733	C>A	dHsaMDS442542138
FAT4	p.G4257=	chr4:125481693	C>T	dHsaMDS196686310
FAT4	p.S4282=	chr4:125487374	C>T	dHsaMDS864011250

Gen	Variante	Posición hg38	CDS	ID Sonda
FLNA	p.T1508=	chrX:154358519	G>A	dHsaMDS691619610
KMT2C	p.I575=	chr7:152250863	A>G	dHsaMDS363378119
KRAS	p.A146T	chr12:25225628	C>T	dHsaMDV2010079
KRAS	p.Q61L	chr12:25227342	T>A	dHsaMDV2010101
KRAS	p.V14I	chr12:25245345	C>T	dHsaMDS860027581
KRAS	p.G13E	chr12:25245346	GC>TT	dHsaMDS163593566
KRAS	p.G12A	chr12:25245350	C>G	ID:1863506
KRAS	p.G12V	chr12:25245350	C>A	ID:1863506
KRAS	p.G12D	chr12:25245350	C>T	ID:1863506
KRAS	p.G12R	chr12:25245351	C>G	ID:1863506
KRAS	p.G12C	chr12:25245351	C>A	ID:1863506
KRAS	p.G12S	chr12:25245351	C>T	ID:1863506
NRAS	p.Q61R	chr1:114713908	T>C	dHsaMDS882187944
NRAS	p.G12V	chr1:114716126	C>A	ID:12001627
NRAS	p.G12C	chr1:114716127	C>A	ID:12001627
PIK3CA	p.K111E	chr3:179199156	A>G	dHsaMDS635800075
PIK3CA	p.C420R	chr3:179210192	T>C	dHsaMDS823193175
PIK3CA	p.Y1021C	chr3:179234219	A>G	dHsaMDS814336169
PIK3CA	p.H1047R	chr3:179234297	A>G	dHsaMDV2010077
PTEN	p.R15*	chr10:87864512	A>T	dHsaMDS358644896
RNF43	p.R392Efs*27	chr17:58358602	TG>T	dHsaMDS509966829
RNF43	p.I48T+rs37440 93	chr17:58415435	ATAAT>GTA AC	dHsaMDS909393116
SMAD2	p.S464*	chr18:47841840	G>C	dHsaMDS2514276
SMAD4	p.S32*	chr18:51047134	GGA>G	dHsaMDS367053456
SMAD4	p.D351Y	chr18:51065518	G>T	dHsaMDS156813217
SMAD4	p.C363G	chr18:51065554	T>G	dHsaMDS547413285
SMAD4	p.K428*	chr18:51067161	A>T	dHsaMDS476921044
SOX9	p.S39Lfs*213	chr17:72121501	C>CG	dHsaMDS182158765
SOX9	p.N45H+p.P48Efs* 201	chr17:72121524	AACACGCG GCCCCAG> CACACGC	dHsaMDS612155605
SOX9	p.V71Cfs*39	chr17:72121597	TC>T	dHsaMDS362795240
SOX9	p.F270Sfs*25	chr17:72123665	TTC>T	dHsaMDS418679127
SOX9	p.R271Pfs*25	chr17:72123666	T>TC	dHsaMDS181648474
SOX9	p.Q505Afs*72	chr17:72124364	TAC>T	dHsaMDS630098247
TCF7L2	p.R402Q	chr10:113152430	G>A	dHsaMDS877174646
TP53	p.R342*	chr17:7670685	G>A	dHsaMDV2516914
TP53	p.E298*	chr17:7673728	C>A	dHsaMDS2512032
TP53	p.E285V	chr17:7673766	T>A	dHsaMDS904130891
TP53	p.R282W	chr17:7673776	G>A	dHsaMDV2516902
TP53	p.C275F	chr17:7673796	C>A	dHsaMDS2513452
TP53	p.R273H	chr17:7673802	C>T	dHsaMDV2010109
TP53	p.R273C	chr17:7673803	G>A	dHsaMDV2510538
TP53	p.R267W	chr17:7673821	G>A	dHsaMDS2511682
TP53	p.G266E	chr17:7673823	C>T	dHsaMDV2516854
TP53	p.P250L	chr17:7674214	G>A	dHsaMDS2512020
TP53	p.R248Q	chr17:7674220	C>T	dHsaMDV2010127

Gen	Variante	Posición hg38	CDS	ID Sonda
TP53	p.R248W	chr17:7674221	G>A	dHsaMDV2010107
TP53	p.G245S	chr17:7674230	C>T	dHsaMDV2516746
TP53	p.C238Y	chr17:7674250	C>T	dHsaMDV2516850
TP53	p.Y234C	chr17:7674262	T>C	dHsaMDV2516900
TP53	p.Y220C	chr17:7674872	T>C	dHsaMDV2510536
TP53	p.T211I	chr17:7674899	G>A	dHsaMDS422580936
TP53	p.E204*	chr17:7674921	C>A	dHsaMDS2511762
TP53	p.G199V	chr17:7674935	C>A	dHsaMDS2513584
TP53	p.I195Kfs*53	chr17:7674947	A>AAA	dHsaMDS984144988
TP53	p.C176F	chr17:7675085	C>A	dHsaMDS318375984
TP53	p.R175H	chr17:7675088	C>T	dHsaMDV2010105
TP53	p.V173L	chr17:7675095	C>A	dHsaMDS2511406
TP53	p.V157G	chr17:7675142	A>C	dHsaMDS2511960
TP53	p.L130F	chr17:7675224	G>A	dHsaMDS2511872
TP53	p.S127F	chr17:7675232	G>A	dHsaMDS2513586
TP53	p.S15Rfs30*	chr17:7676550	A>ACT	dHsaMDS870747583

8.4. ANEXO IV: Listado de pruebas realizadas a los pacientes

Tabla 29: Experimentos realizados en los diferentes pacientes de la cohorte. T: secuenciación del tumor. N: Secuenciación de material germinal. P: Secuenciación de Plasma al diagnóstico. ddPCR Som.: PCR digital de variantes somáticas. ddPCR Pat.: PCR digital de variantes patológicas. ET: WES de tejido primario. EN: WES de material germinal. EP: WES del plasma a la recaída. EM: WES de tejido metastásico.

CCR	Recaída	T	N	P	ddPCR Som.	ddPCR Pat.	ET	EN	EP	EM
001	No	Sí	Sí	Sí	Sí	Sí				
002	Sí	Sí	Sí	Sí	Sí	Sí				
003	No	Sí	Sí	Sí	Sí	Sí				
004	No	Sí	Sí	Sí						
005	No	Sí	Sí							
006	No	Sí	Sí							
007	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
008	Sí	Sí	Sí	Sí		Sí	Sí	Sí	Sí	
009	No	Sí	Sí	Sí	Sí	Sí				
010	No	Sí	Sí	Sí						
011	No	Sí	Sí	Sí	Sí	Sí				
012	No	Sí	Sí	Sí						
013	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
014	No	Sí	Sí	Sí						
015	No	Sí	Sí	Sí	Sí	Sí				
016	No	Sí	Sí	Sí						
017	No	Sí	Sí							
018	No	Sí	Sí							
019	No	Sí	Sí							
020	No	Sí	Sí	Sí						
021	No	Sí	Sí	Sí	Sí	Sí				
022	No	Sí	Sí	Sí						
023	No	Sí	Sí	Sí						
024	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
025	No	Sí	Sí	Sí						
026	No	Sí	Sí	Sí						
027	No	Sí	Sí							
028	No	Sí	Sí	Sí						
029	Sí	Sí	Sí	Sí	Sí	Sí				
030	No	Sí	Sí	Sí						
031	No	Sí	Sí	Sí						
032	No	Sí	Sí	Sí						

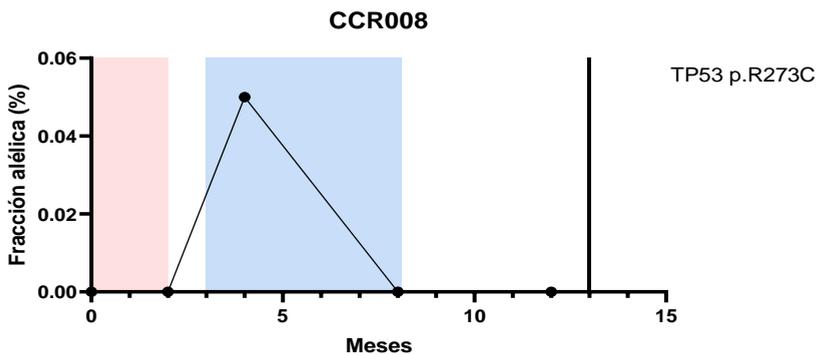
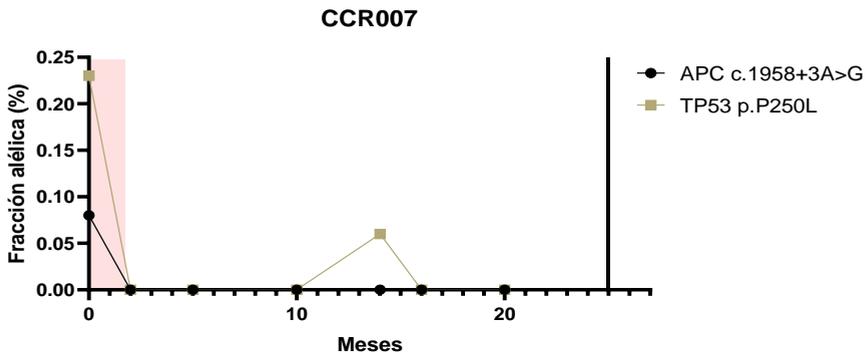
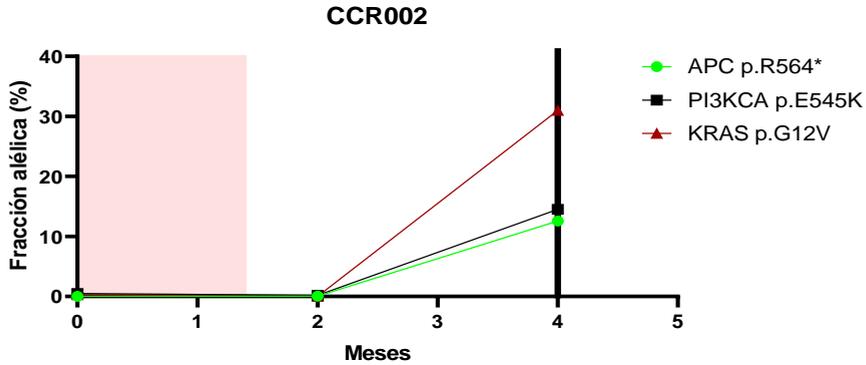
CCR	Recaída	T	N	P	ddPCR Som.	ddPCR Pat.	ET	EN	EP	EM
033	No	Sí	Sí							
034	No	Sí	Sí							
035	No	Sí	Sí	Sí	Sí	Sí				
036	No	Sí	Sí	Sí						
037	No	Sí	Sí							
038	No	Sí	Sí							
039	No	Sí	Sí	Sí		Sí				
040	No	Sí	Sí	Sí	Sí	Sí				
041	No	Sí	Sí	Sí	Sí	Sí				
042	No	Sí	Sí							
043	No	Sí	Sí	Sí	Sí	Sí				
044	No	Sí	Sí	Sí	Sí	Sí				
045	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
046	No	Sí	Sí	Sí	Sí	Sí				
047	No	Sí	Sí	Sí	Sí	Sí				
048	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
049	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
050	Sí	Sí	Sí	Sí	Sí	Sí				
051	No	Sí	Sí							
052	No	Sí	Sí	Sí						
053	No	Sí	Sí	Sí						
054	No	Sí	Sí	Sí	Sí	Sí				
055	No	Sí	Sí	Sí						
056	No	Sí	Sí	Sí						
057	No	Sí	Sí	Sí						
058	No	Sí	Sí	Sí	Sí	Sí				
059	No	Sí	Sí	Sí						
060	No	Sí	Sí	Sí						
061	No	Sí	Sí	Sí	Sí	Sí				
062	Sí	Sí	Sí	Sí						
063	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
064	No	Sí	Sí							
065	No	Sí	Sí	Sí						
066	No	Sí	Sí	Sí	Sí	Sí				
067	No	Sí	Sí	Sí						
068	No	Sí	Sí	Sí	Sí	Sí				
069	No	Sí	Sí	Sí						
070	No	Sí	Sí	Sí						
071	No	Sí	Sí	Sí						
072	No	Sí	Sí	Sí						

CCR	Recaída	T	N	P	ddPCR Som.	ddPCR Pat.	ET	EN	EP	EM
073	No	Sí	Sí	Sí	Sí	Sí				
074	No	Sí	Sí	Sí		Sí				
075	No	Sí	Sí	Sí	Sí	Sí				
076	No	Sí	Sí	Sí	Sí	Sí				
077	No	Sí	Sí	Sí	Sí	Sí				
078	No	Sí	Sí	Sí						
079	No	Sí	Sí	Sí						
080	No	Sí	Sí	Sí	Sí	Sí				
081	No	Sí	Sí	Sí						
082	No	Sí	Sí	Sí	Sí	Sí				
083	No	Sí	Sí							
084	No	Sí	Sí	Sí						
085	No	Sí	Sí	Sí						
086	No	Sí	Sí	Sí	Sí	Sí				
087	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
088	No	Sí	Sí	Sí	Sí	Sí				
089	No	Sí	Sí	Sí	Sí	Sí				
090	No	Sí	Sí	Sí	Sí	Sí				
091	No	Sí	Sí	Sí						
092	No	Sí	Sí	Sí	Sí	Sí				
093	No	Sí	Sí	Sí	Sí	Sí				
094	No	Sí	Sí							
095	No	Sí	Sí	Sí	Sí	Sí				
096	No	Sí	Sí	Sí						
097	No	Sí	Sí	Sí	Sí	Sí				
098	No	Sí	Sí							
099	No	Sí	Sí	Sí						
100	No	Sí	Sí	Sí	Sí	Sí				
101	No	Sí	Sí	Sí						
102	No	Sí	Sí	Sí	Sí	Sí				
103	No	Sí	Sí	Sí	Sí	Sí				
104	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	
105	No	Sí	Sí	Sí		Sí				
106	No	Sí	Sí	Sí						
107	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
108	No	Sí	Sí	Sí	Sí	Sí				
109	No	Sí	Sí	Sí						
110	No	Sí	Sí	Sí	Sí	Sí				
111	No	Sí	Sí	Sí						
112	No	Sí								

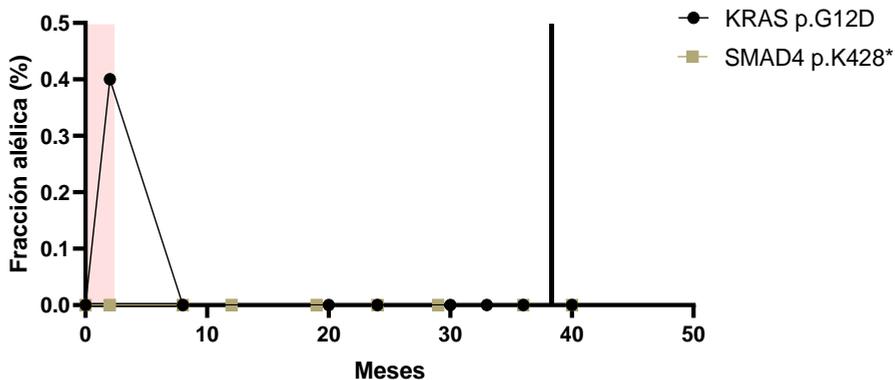
CCR	Recaída	T	N	P	ddPCR Som.	ddPCR Pat.	ET	EN	EP	EM
113	No	Sí	Sí	Sí						
114	No	Sí	Sí							
115	No	Sí	Sí	Sí						
116	No	Sí	Sí							
117	No	Sí	Sí							
118	No	Sí	Sí	Sí	Sí	Sí				
119	No	Sí	Sí	Sí	Sí	Sí				
120	No	Sí	Sí	Sí						
121	No	Sí	Sí	Sí						
122	No	Sí	Sí	Sí						
123	No	Sí	Sí	Sí	Sí	Sí				
124	No	Sí	Sí	Sí	Sí	Sí				
125	No	Sí	Sí	Sí	Sí	Sí				
126	No	Sí	Sí	Sí						
127	No	Sí	Sí	Sí						
128	No	Sí	Sí	Sí						
129	No	Sí	Sí	Sí						
130	No	Sí	Sí							
131	No	Sí	Sí	Sí						
132	No	Sí	Sí	Sí	Sí	Sí				
133	No	Sí	Sí	Sí						
134	No	Sí	Sí	Sí						
135	No	Sí	Sí	Sí	Sí	Sí				
136	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
137	No	Sí	Sí	Sí						
138	Sí	Sí	Sí		Sí	Sí	Sí	Sí	Sí	
139	No	Sí	Sí	Sí						
140	No	Sí	Sí	Sí	Sí	Sí				
141	No	Sí	Sí							
142	Sí	Sí	Sí	Sí	Sí	Sí				
143	No	Sí	Sí							
144	No	Sí	Sí			Sí				
145	No	Sí	Sí	Sí	Sí	Sí				
146	No	Sí	Sí	Sí	Sí	Sí				
147	No	Sí	Sí	Sí	Sí	Sí				
148	No	Sí	Sí	Sí						
149	No	Sí	Sí							
150	Sí	Sí	Sí	Sí			Sí	Sí	Sí	Sí
151	No	Sí	Sí							

8.5. ANEXO V: Seguimientos de los pacientes con recaída

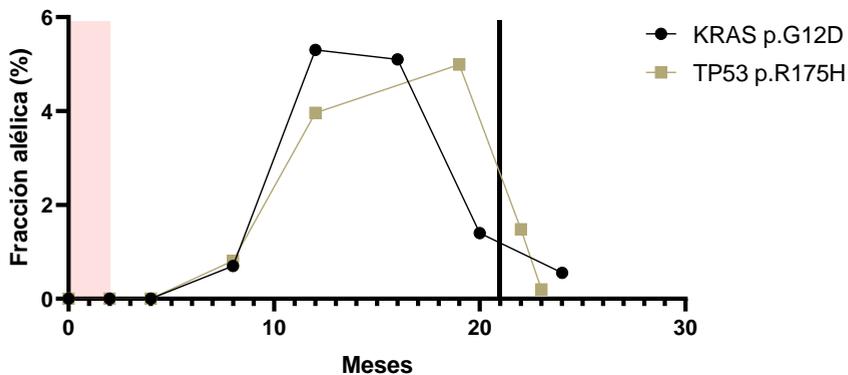
Figura 42: Seguimientos de los pacientes metastásicos mediante sondas contra mutaciones somáticas y patológicas. Línea negra: recaída. Cuadrado rosa: periodo de 6-8 semanas tras la cirugía. Cuadrado azul: quimioterapia adyuvante.



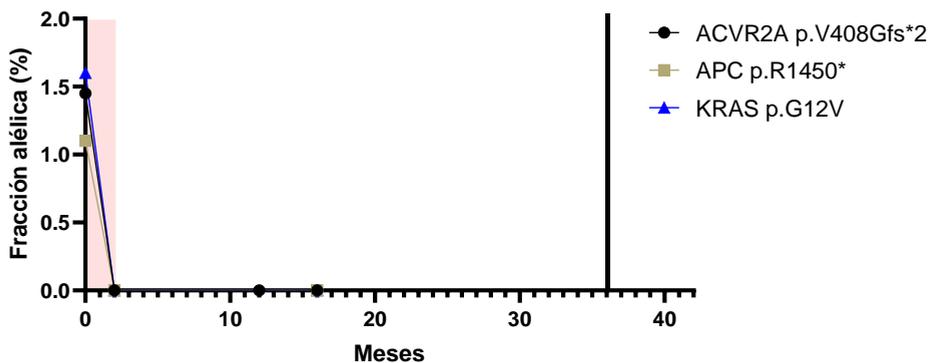
CCR013



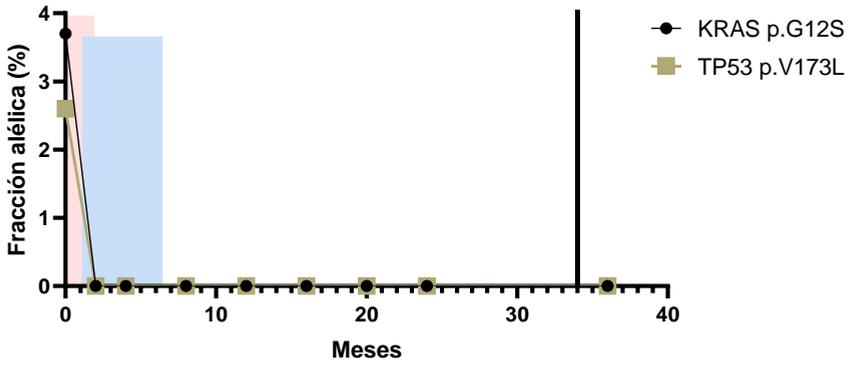
CCR024



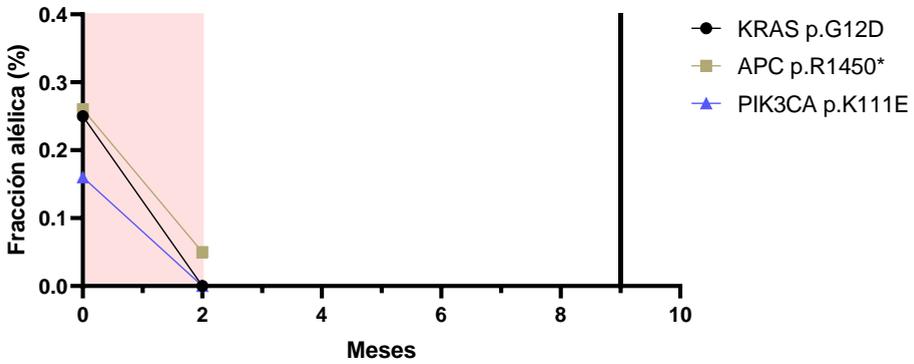
CCR029



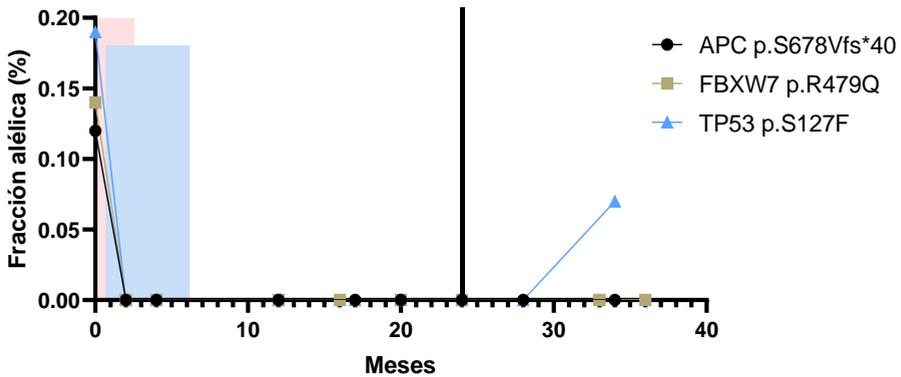
CCR048

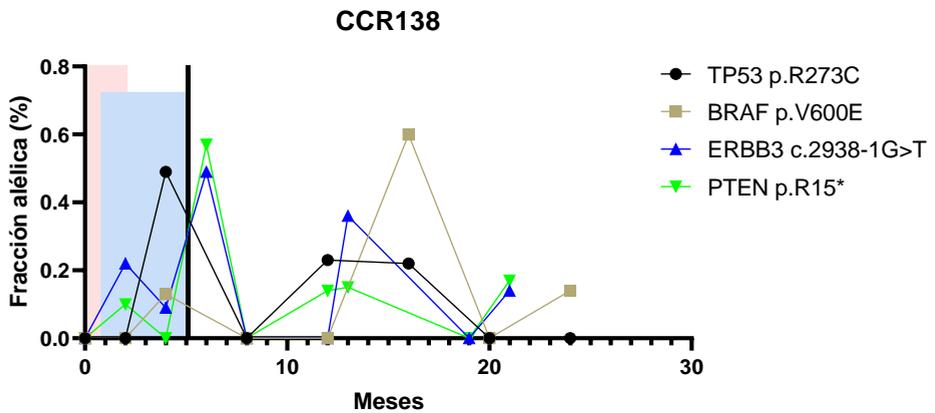
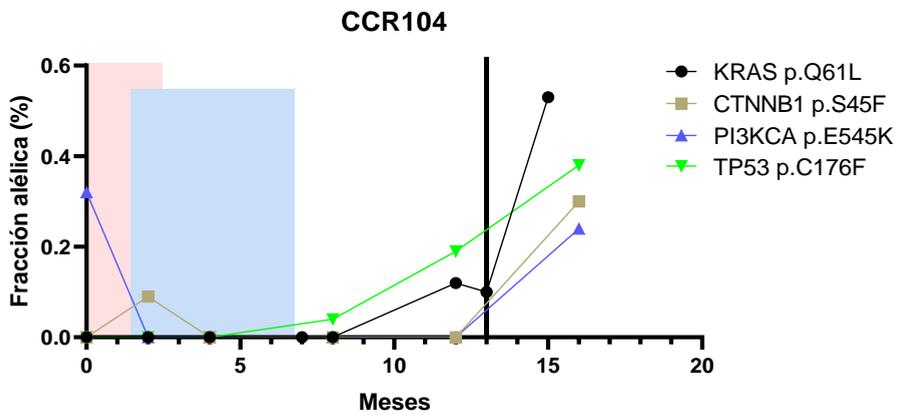
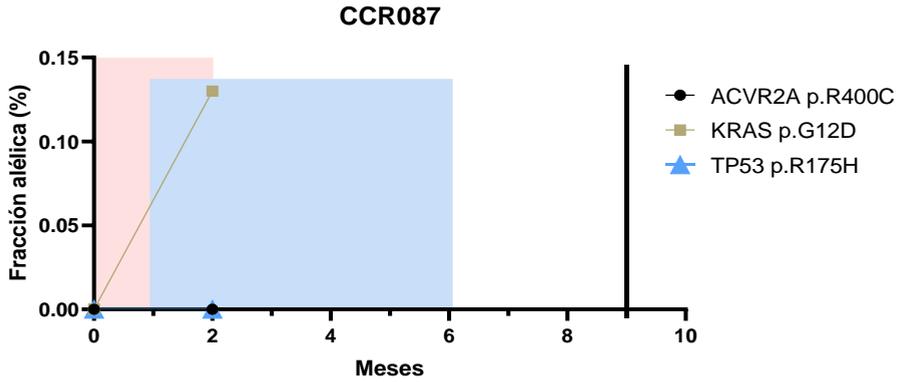


CCR050

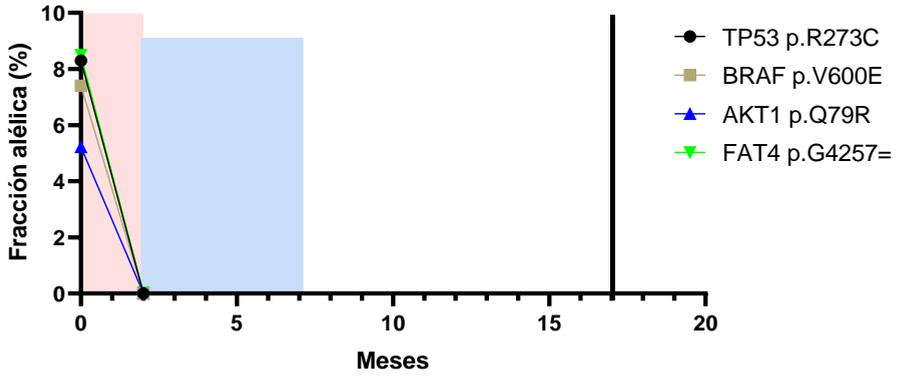


CCR063

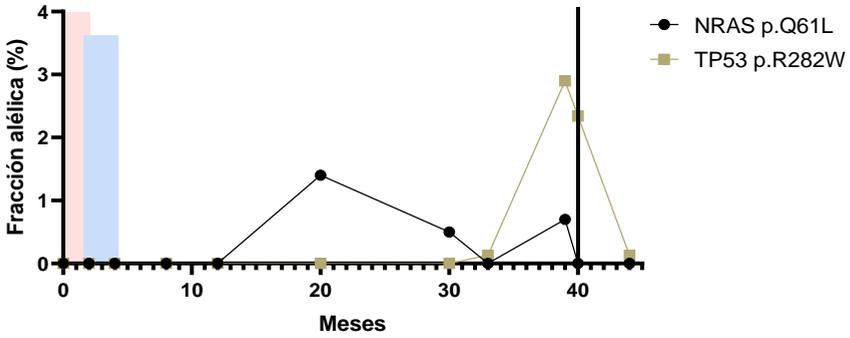




CCR142



CCR150



8.6. ANEXO VI: Concordancia ddPCR y WES

Tabla 30: Concordancia entre las variantes seguidas en cada paciente mediante ddPCR y su detección en WES.

Paciente	Alteración	ddPCR		Exoma	
		ctDNA	Frec.	ctDNA	Frec.
CCR007	c.1958+3A>G <i>APC</i>	-	0.00%	-	0.00%
	p.P250L <i>TP53</i>	-	0.00%	-	0.00%
CCR008	p.R273C <i>TP53</i>	-	0.00%	-	0.00%
	p.F339V <i>SMAD4</i>	-	-	-	0.00%
CCR013	p.G12D <i>KRAS</i>	-	0.00%	-	0.00%
	p.K428* <i>SMAD4</i>	-	0.00%	-	0.00%
CCR024	p.G12D <i>KRAS</i>	+	1.40%	-	0.00%
	p.R175H <i>TP53</i>	+	1.48%	+	2.03%
CCR045	p.G13E <i>KRAS</i>	-	0.00%	-	0.00%
	p.H1047R <i>PIK3CA</i>	-	0.00%	-	0.00%
CCR048	p.G12S <i>KRAS</i>	-	0.00%	-	0.00%
	p.V173L <i>TP53</i>	-	0.00%	-	0.00%
CCR049	p.R586* <i>AMER1</i>	+	14.57%	+	26.74%
	p.R248Q <i>TP53</i>	+	7.90%	+	13.46%
	p.E545K <i>PIK3CA</i>	+	7.40%	+	12.53%
CCR063	p.S678Vfs*40 <i>APC</i>	-	0.00%	-	0.00%
	p.R479Q <i>FBXW7</i>	-	0.00%	-	0.00%
	p.S127F <i>TP53</i>	-	0.00%	-	0.00%
CCR104	p.Q61L <i>KRAS</i>	+	0.10%	-	0.00%
	p.S45F <i>CTNNB1</i>	-	0.00%	-	0.00%
	p.E545K <i>PIK3CA</i>	-	0.00%	-	0.00%
	p.C176F <i>TP53</i>	+	0.19%	-	0.00%
CCR107	p.H1047R <i>PIK3CA</i>	-	0.00%	-	0.00%
	p.G12V <i>KRAS</i>	-	0.00%	-	0.00%
	p.R1450* <i>APC</i>	+	0.24%	-	0.00%
	p.R175H <i>TP53</i>	-	0.00%	-	0.00%
CCR136	p.E1322* <i>APC</i>	+	0.29%	-	0.00%
	p.E545K <i>PIK3CA</i>	-	0.00%	-	0.00%
	p.G12D <i>KRAS</i>	+	0.61%	-	0.00%
CCR138	p.R273C <i>TP53</i>	-	0.00%	-	0.00%
	p.V600E <i>BRAF</i>	-	0.00%	-	0.00%
	c.2938-1G>T <i>ERBB3</i>	+	0.49%	-	0.00%
	p.R15* <i>PTEN</i>	+	0.57%	-	0.00%
CCR150	p.R282W <i>TP53</i>	+	2.34%	+	2.23%
	p.S1415Rfs*4 <i>APC</i>	-	-	+	5,18%

8.7. ANEXO VII: Concordancia en WES en pacientes con tejido de la metástasis

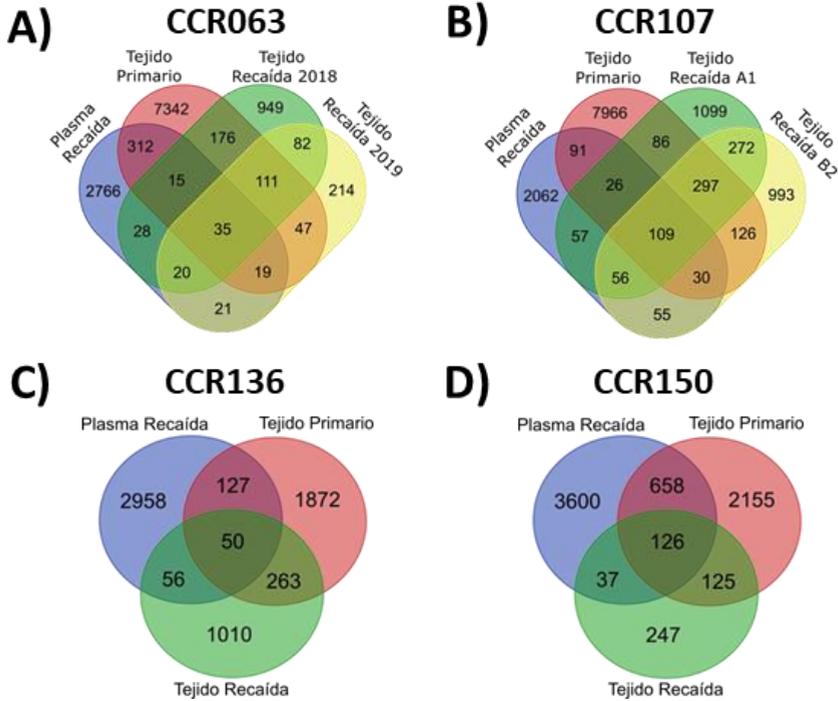


Figura 43: Diagramas de Venn para la concordancia entre las variantes somáticas del tejido primario, plasma de la recaída y los diferentes tejidos de la recaída. Se incluyen los pacientes a) CCR063, b) CCR107, c) CCR136 y d) CCR150.

