



Universitat de Lleida

Aplicació de tecnologies d'alt rendiment per a l'anàlisi d'alteracions moleculars i l'heterogeneïtat intratumoral en el càncer d'endometri

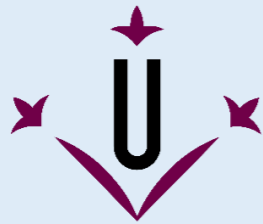
Dolors Cuevas Sánchez

<http://hdl.handle.net/10803/670256>

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



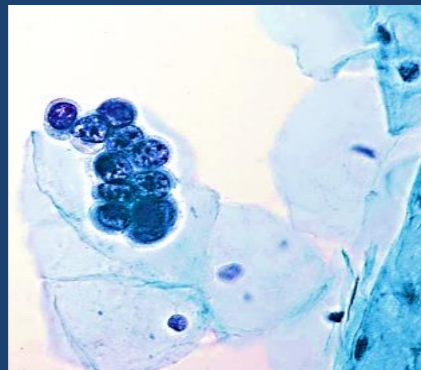
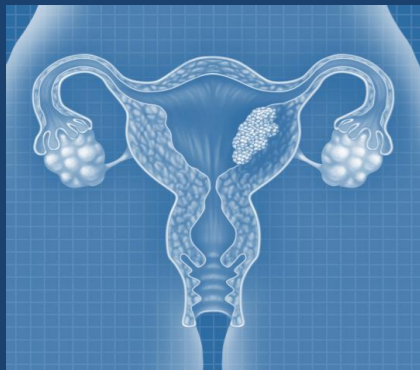
Universitat de Lleida

APLICACIÓ DE TECNOLOGIES D'ALT RENDIMENT PER A L'ANÀLISI D'ALTERACIONS MOLECULARS I L'HETEROGENEÏTAT INTRATUMORAL EN EL CÀNCER D'ENDOMETRI

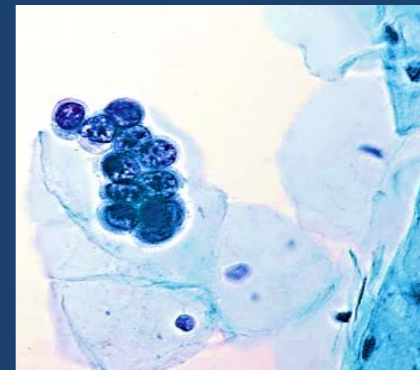
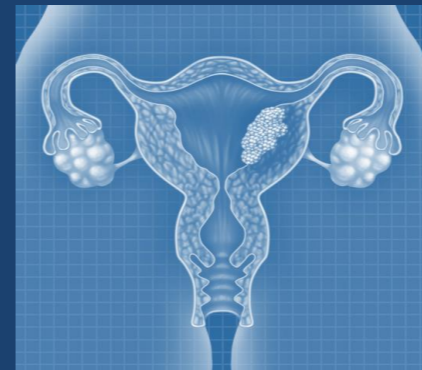
TESI DOCTORAL

APLICACIÓ DE TECNOLOGIES D'ALT RENDIMENT PER A L'ANÀLISI
D'ALTERACIONS MOLECULARS I L'HETEROGENEÏTAT INTRATUMORAL
EN EL CÀNCER D'ENDOMETRI

Els dos subtipus més freqüents de càncer d'endometri (CE) són el carcinoma endometriode (CEE) i el carcinoma serós (CSE). El diagnòstic diferencial entre aquestes dues tipologies no sempre és fàcil, hi ha casos que presenten característiques histològiques molt dubtoses i ambigües que dificulten el correcte diagnòstic final. Aquests dos tipus tumorals presenten diferents perfils moleculars i un pronòstic diferencial, fent que el CSE sigui el subtipus més agressiu i amb el pronòstic més desfavorable. Actualment s'està posant en coneixement la importància que pot jugar l'heterogeneïtat intratumoral (HIT) en els comportaments més agressius dels tumors i en la resistència als tractaments, com és el cas del CSE. En els últims anys, la irrupció de les tecnologies d'alt rendiment com la NGS (*Next Generation Sequencing*) i la MLPA (*Multiple Ligase-dependent Probe Amplification*) han ofert l'opció d'ampliar el coneixement molecular del càncer d'una manera més precisa i resolutiva tal i com mostren els resultats d'aquesta tesi.



DOLORS CUEVAS
SÁNCHEZ



DOLORS CUEVAS SÁNCHEZ

2020

2020



Universitat de Lleida



Institut de Recerca Biomèdica

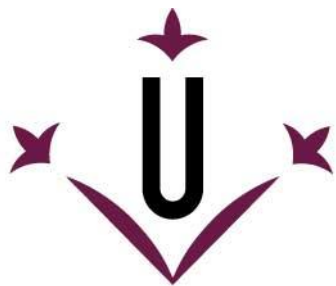
Nota sobre les fotos de la portada

Les fotos de la portada resumeixen d'una manera gràfica els tres principals àmbits d'estudi d'aquesta tesi doctoral:

- *El càncer d'endometri.* Imatge modificada a partir de: New Study Reveals Dozens of Pathways in Endometrial Cancer Development. Clinical OMICs, February 14, 2020. (ESQUERRA).

- *Les cèl·lules tumorals del càncer d'endometri.* Imatge modificada a partir de: Nayar, R. & Wilbur, D.C. 2015. The Bethesda System for Reporting Cervical Cytology. Definitions, Criteria and Explanatory notes. Third edition. Springer, 192 p. ISBN: 978-3-319-11073-8. (CENTRE).

- *L'àcid desoxiribonucleic (ADN).* Imatge modificada a partir de Public Domain Pictures.net created by Bobek Ltd. Registered in England and Wales No. 5951382. (DRETA).



Universitat de Lleida

TESI DOCTORAL

**APLICACIÓ DE TECNOLOGIES D'ALT RENDIMENT PER A L'ANÀLISI
D'ALTERACIONS MOLECULARS I L'HETEROGENEÏTAT
INTRATUMORAL EN EL CÀNCER D'ENDOMETRI**

MARIA DOLORS CUEVAS SÁNCHEZ

Memòria presentada per optar al grau de Doctor per la Universitat de Lleida

Programa de Doctorat en Salut

Director/a

Xavier Matias-Guiu Guia

Ana Velasco Sánchez

Tutor/a

Xavier Matias-Guiu Guia

2020

“L’essencial és invisible als ulls”

Antoine de Saint-Exupéry

Per a tots els que m'han omplert i m'omplen la vida d'alegria, amor i felicitat,

En especial per,

als papes, Benjamin i Victoria,

al mano, Benja,

al millor amic i company de vida, Damià,

als dos grans amors de la meva vida, Ada i Jan,

a la iaia Teodora, per a què t'arribi allà on siguis

a mi

AGRAÏMENTS

Després de tenir la tesi enllestida, i d'haver pogut deixar tota la vessant científica de la tesi dipositada en el paper, ha arribat el moment íntim de descriure la part emocional, i no menys important, que comporta la vivència d'una tesi doctoral, i de donar les gràcies a tots els que han estat al meu costat durant aquesta etapa. És el moment de descriure com vas començar, les persones que t'has trobat, que t'han ensenyat, que t'han fet créixer, tant de manera professional com personal i emocional, les persones que t'han acompanyat, que t'han fet costat, tot el que ha passat durant aquests anys, els moments que has viscut, en com has arribat al que podríem dir el final d'una etapa.

En el meu cas, l'inici va ser molt especial, on va ser la pròpia vida qui em va obrir el camí cap al món de la biologia molecular del càncer. Un món que ja em va fascinar i apassionar al cursar l'assignatura de Biologia molecular del Càncer a la facultat de Biologia de la UB. Després de 4 anys d'haver-la cursat, la vida em va permetre coincidir en un avió en direcció a Bristol, amb la Dra. Ana Balil, una oncòloga especial, humil, humana i molt estimada, qui em va ajudar a atrevir-me a enviar el meu currículum al Dr. Llombart i al Dr. Matias-Guiu. Malauradament, el destí va fer que no pogués arribar a temps de poder-li agrair presencialment la seva ajuda i tampoc a poder gaudir de la seva amistat. Just la setmana que entrava a treballar com a becària del Dr. Matias-Guiu i del Dr. Llombart, de manera inesperada la vida se li va truncar. Així que li agraeixo el fet d'haver-me tret la por de fer un canvi de vida i impulsar-me cap al món de l'estudi del càncer.

La meva primera etapa com a becària predoctoral també va ser especial perquè per una part estava dirigida pel cap d'Oncologia Mèdica, el Dr. Llombart, i per altra pel cap del Servei d'Anatomia Patològica, el Dr. Matias-Guiu, el Xavier. En aquell moment des d'Oncologia es dirigia el projecte, i des d'Anatomia es permetia l'accés al laboratori i l'aprendre les tècniques moleculars, al costat de l'Ana i la Núria. Això em va permetre conèixer i aprendre tots els intrinquis del laboratori de Patologia Molecular, que no són pocs, i estar preparada per a quan la vida novament em va donar una altra oportunitat, oferir-me un lloc dintre al laboratori de Patologia Molecular. En aquesta primera etapa he d'agrair tot el suport i coneixement que em va donar el Dr. Llombart i també tot el que em va aportar la Núria, tant a nivell professional com personal. A partir d'aquest moment, el Xavier vas ser qui va començar a dirigir la meva feina assistencial i el meu projecte d'investigació, amb la codirecció de l'Ana. Gràcies als meus directors de tesi, el Xavier i l'Ana, vaig començar el meu camí en el món de la recerca del càncer.

Gràcies Xavier per haver-me motivat per emprendre aquest camí, per estar insistint en què havíem de fer un projecte bonic i engrescador, en que havíem de fer la tesi. Gràcies a la teva ment tant visionària, a la teva perseverança, i gràcies per fer-nos arribar la teva passió pel món de la

recerca i del coneixement, i per buscar l'excel·lència. Segur que si no hagués estat així aquesta tesi no s'hagués dut a terme.

Gràcies Ana per tot el que m'has ensenyat i m'has donat, que ha sigut moltíssim, tant a nivell científic i professional, però sobretot ha estat indescriptible tot el que m'has ajudat a nivell personal. Gràcies per fer-me créixer i per ensenyar-me a acceptar el curs de la vida, a la gent que no pensa com jo i que té molt a ensenyar-me. És molt bonica aquesta experiència de créixer juntes, infinites gràcies. Ara sabem que ens hem trobat per aprendre a estimar-nos i a estimar als altres, i que ens tenim l'una a l'altra.

A part dels directors de la tesi, també vull agrair a la Marta, la Maria i la Sònia, la seva implicació en aquesta tesi. Gràcies per haver-me donat el vostre suport professional i científic, i per haver revisat la tesi. A tu Marta, gràcies per ser com ets, transparent, sincera, positiva, per escoltar-me, per dir-me que tot passa i tot arriba, per ser com ets. Estic molt contenta de què hagi vingut a format part del nostre equip. Tenim un equip genial. A tu Maria, et dono les gràcies per haver-me ofert la teva amistat, i per totes les converses enriquidores que hem tingut, per tots els moments de confiança i sinceritat, i per saber que quan ho necessitem allí estem les dos, per ajudar-nos en el que calgui. Una de les altres persones que m'han acompanyat és l'Anna Serrate, amb qui vam compartir moltes complicitats al laboratori, moltes inquietuds, alegries i a qui tinc un carinyo molt especial. Gràcies per la teva alegria, per la teva bondat, pel teu recolzament, per la teva humilitat, per tenir un cor molt gran.

Gràcies a les meves súper compis de despatx, la Pilar, la Subi i l'Ana, sobretot per tots els moments d'alegria compartits, per tots els somriures, per ajudar-nos entre nosaltres, per tot el vostre suport i acompanyament durant l'escriptura de la tesi, i pel nostre nou projecte. Per a què podem aterrar algun dia a l'illa paradisiàca dels nostres somnis. Gràcies per fer-me sentir bé en els moments difícils. De les meves compis d'esmorzar la Pilar, la Maria i la Susi, em queda agrair a la Susi el seu súper optimisme i positivisme, per encomanar-lo, i per compartir tots aquelles alegries i preocupacions dels nostres petits.

Gràcies a tota la gent que ha format i forma l'equip del Servei d'Anatomia Patològica, perquè cadascú de vosaltres sou especials i diferents, i perquè amb aquesta especialitat de cadascú heu enriquit la meua vida. Gràcies a l'equip del Biobanc, la Maria, la Ingrid i l'Ana, per estar sempre disposades a ajudar, i a treballar, per fer-te la vida més fàcil. Gràcies a la gent de l'IRBLleida del grup de Patologia Oncològica. I en general gràcies a tota la gent de l'Hospital HUAV, del Santa Maria, i de l'IRBLleida que m'han ajudat.

Gràcies a l'Ester, al Paco i la Núria, perquè m'heu ajudat i m'esteu ajudant a créixer, a obrir-me, a ser lliure, ser més feliç, a ser jo, infinites gràcies.

Gràcies a la Txell, per haver-me ofert la seva amistat de manera incondicional i per haver fet arribar molts rajos de llum i energia positiva a la meva vida.

Gràcies a tots els amics de Roquetes, Lleida i arreu del món, per totes les trobades junts i per totes les alegries compartides. En especial a la família Lara-Pla, per estar sempre disponibles i amb els braços i el cor obert, gràcies per tot.

Gràcies a la meves dos grans famílies, la Cuevas-Sánchez i la Vericat-Querol, amb totes les seves extensions. Tot i la distància, gràcies per fer-me sentir que sempre heu estat a prop, ocupant-vos de mi, parant-nos compte, fent-nos costat. Que mai hem perdut les ganes de retrobar-nos i d'estar junts, perquè això és el que, en part, ens fa ser feliços i estar bé. Gràcies per estimar-me, jo també us estimo. D'aquesta gran família vull tenir un molt especial record, per als que ja no hi són presents, com són els iaïos, el meu padrí i el meu sogre. Perquè sé que estarien súper contents de gaudir d'aquest moment amb mi. Gràcies per tot el que em vau donar i regalar. I per la meva relació especial en ella, moltíssimes gràcies iaia Teodora, perquè sempre seràs un referent de la meva vida, i sé que estàs present en cada moment. Sé que t'hauria encantat poder estar a la primera fila de la sala, escoltant-me i que seria uns dels dies més grans també de la teva vida. Sé que estaràs allí i que em donaràs tota l'energia del món, la teva energia, tota la que tu desprenies.

Infinites gràcies a vatos, papa i mama, per haver-me regalat la vida, per haver-me donat tant d'amor, tanta confiança, per haver-me transmès la vostra bondat, humilitat...perquè una gran part del que sóc és gràcies a vatos, ja ho sabeu, sou uns pares immillorables i únics. Us estimo infinitament i sempre estaré al vostre costat. I espero que us encanti la meva tesi, sé que us agradaria poder entendre molt millor tot el que faig, però només cal entendre que tot això és per a què el càncer sigui una malaltia crònica i que no acabi amb les nostres vides. A tu mano, gràcies per tot, per ser com ets, per ensenyar-me que lo que importa no són les paraules ni els fets, sinó l'amor incondicional que surt de dins. Per estar sempre donant-me el teu suport, confiant en mi, infinites gràcies, T'estimo. I per haver-me donat al meu fillolet Albert, que és un amor i molt especial per a mi.

Gràcies a tu, carinyo, per una banda per tot el teu suport i consell científic, per haver-me ajudat en molts aspectes de la tesi, tant d'aclariments d'idees, de conceptes, de disseny....I per altra banda, gràcies per haver estat tants anys al meu costat, fent que la vida fos fàcil, complicada,

alegre, trista, Perquè tot el que hem viscut i estem vivint ens està fent més grans com a persones. Per continuar així, per continuar buscant nous camins i noves formes. Per haver pogut arribar a veure i viure la vida des d'una altra perspectiva que ens està fent més feliços i lliures. Perquè ara sé més que mai, que sempre estarem junts. T'estimo. I sobretot per haver donat mútuament la vida al Jan i l'Ada, que sens dubte, són les persones més importants de les nostres vides.

Gràcies als meus dos amors i tresors, l'ADA i el JAN, perquè gràcies a ells he après i estic aprenent dia a dia, lo que realment és important en la vida. Ells m'han ajudat a posar un altre ordre de prioritats en la meua vida, un ordre més harmònic i natural, un nou ordre per ser més feliç. Perquè cada dia em donen lliçons de vida, de saber viure-la al 100 %, només pensant en el dia a dia, en la il·lusió i les ganes de viure, sense por, sent lliures. Gràcies a la seva forma de ser, la seva mirada, rialla, alegria, bonança, entusiasme, motivació, amor... Us estimo infinitament, sou el millor de la meua vida.

Gràcies a mi, per haver afrontat com un repte la tesi, la vida en general, per tot el canvi que he fet, tant professional com personal, per tot el que he viscut en aquests anys de la solteria al casament, de l'arribada dels meus tresors, de la pèrdua d'essers molt especials i estimats. Perquè em pensava que això d'una tesi i d'escriure uns articles científics en anglès no estava fet per a mi. Cada dia tinc més clar que els límits te'ls poses tu, o sigui que tu ets capaç de tot i de crear lo inimaginable. Confia en tu i en la vida per sempre.

RESUMS

RESUM

Els dos subtipus més freqüents de càncer d'endometri (CE) són el carcinoma endometriode (CEE) i el carcinoma serós (CSE). El diagnòstic diferencial entre aquestes dues tipologies no sempre és fàcil, hi ha casos que presenten característiques histològiques molt dubtoses i ambigües que dificulten el correcte diagnòstic final. Aquests dos tipus tumorals presenten diferents perfils moleculars i un pronòstic diferencial, fent que el CSE sigui el subtipus més agressiu i amb el pronòstic més desfavorable. Actualment s'està posant en coneixement la importància que pot jugar l'heterogeneïtat intratumoral (HIT) en els comportaments més agressius dels tumors i en la resistència als tractaments, com és el cas del CSE. En els últims anys, la irrupció de les tecnologies d'alt rendiment com la NGS (*Next Generation Sequencing*) i la MLPA (*Multiple Ligase-dependent Probe Amplification*) han ofert l'opció d'ampliar el coneixement molecular del càncer d'una manera més precisa i resolutive.

Per aquests motius, els dos primers objectius d'aquesta tesi s'han centrat en analitzar, en els dos tipus més freqüents de càncer d'endometri, diverses alteracions moleculars que ens permetin obtenir uns perfils genètics específics, i classificar d'una manera més precisa i objectiva aquests dos tipus tumorals. Per realitzar les anàlisis de les alteracions genètiques s'han utilitzat, l'aplicació de *targeted sequencing* (NGS) amb un panell personalitzat de 40 gens per determinar els gens alterats respecte a variants somàtiques, i la tècnica MLPA per determinar les alteracions somàtiques en el número de còpies (ASNC) de 106 gens. Finalment, el tercer objectiu s'ha focalitzat en conèixer el paper que juga l'heterogeneïtat intratumoral en el CSE. L'efecte de l'HIT ha estat estudiat a dos nivells d'alteració molecular, a nivell de l'alteració dels gens determinant variants somàtiques i de l'alteració en el número de còpies dels gens, mitjançant les dues tècniques anteriors.

Els resultats del primer objectiu d'aquesta tesi han demostrat la idoneïtat del nostre estudi personalitzat de NGS com a eina molecular addicional per confirmar la classificació histològica del CE. Aquesta estratègia sembla interessant com una eina per classificar tumors amb troballes microscòpiques inusuals i ambigües. Amb els resultats del segon objectiu s'ha descrit que de manera general el CSE presenta moltes més ASNC que el CEE. Tot i això, també s'ha mostrat que dins el CSE hi ha un percentatge de casos (42%) que presenten unes característiques genètiques que no es corresponen amb el fenotip, i que per tant, en aquests casos pot ser de gran ajuda l'ús del perfil genètic d'ASNC per realitzar una classificació més correcta. A més a més, s'ha suggerit l'ús de la combinació de la determinació de p53 per immunohistoquímica i del número de còpies de la *CCNE1* per classificar els casos dins el grup *serous-like*. Per finalitzar, els resultats del tercer objectiu han mostrat que el CSE, per una banda, presenta unes alteracions moleculars clonals com les variants somàtiques al gen *TP53* i guanys en els gens *CCNE1* i *PIK3CA*, i per altra banda, que la seva heterogeneïtat intratumoral està caracteritzada principalment per les ASNC dels gens. A més a més, s'ha remarcat que un dels gens principalment afectats per aquesta heterogeneïtat és el gen *ERBB2*, que és una diana terapèutica àmpliament utilitzada.

RESUMEN

Los dos subtipos más frecuentes de cáncer de endometrio (CE) son el carcinoma endometriode (CEE) y el carcinoma seroso (CSE). El diagnóstico diferencial entre estas dos tipologías no siempre es fácil, hay casos que presentan características histológicas muy dudosas y ambiguas que dificultan su correcto diagnóstico final. Estos dos tipos tumorales presentan perfiles moleculares y pronósticos diferenciales, haciendo que el CSE sea el subtipo más agresivo y con el pronóstico más desfavorable. Actualmente se está poniendo en conocimiento la importancia que puede jugar la heterogeneidad intratumoral (HIT) en los comportamientos más agresivos de los tumores y en la resistencia a los tratamientos, como es el caso del CSE. En los últimos años, la irrupción de las tecnologías de alto rendimiento como la NGS (Next Generation Sequencing) y la MLPA (Multiple Ligase-dependiente Probe Amplification) han ofrecido la opción de ampliar el conocimiento molecular del cáncer de una manera más precisa y resolutive.

Por estos motivos, los dos primeros objetivos de esta tesis se han centrado en analizar en los dos tipos más frecuentes de cáncer de endometrio varias alteraciones moleculares que nos permitan obtener unos perfiles genéticos específicos, y clasificar de una manera más precisa y objetiva estos dos tipos tumorales. Para realizar los análisis de las alteraciones genéticas se han utilizado, la aplicación de *targeted sequencing* (NGS) con un panel personalizado de 40 genes para determinar los genes alterados respecto a variantes somáticas, y la técnica MLPA para determinar las alteraciones somáticas en el número de copias (ASNC) de 106 genes. Finalmente, el tercer objetivo se ha focalizado en conocer el papel que juega la heterogeneidad intratumoral en el CSE. El efecto de la HIT ha sido estudiado a dos niveles de alteración molecular, a nivel de la alteración de los genes analizando las variantes somáticas y la alteración en el número de copias de los genes, mediante las dos técnicas anteriores.

Los resultados del primer objetivo de esta tesis han demostrado la idoneidad de nuestro estudio personalizado de NGS como herramienta molecular adicional para confirmar la clasificación histológica del CE. Esta estrategia parece interesante como una herramienta para clasificar tumores con hallazgos microscópicos inusuales y ambiguos. Con los resultados del segundo objetivo se ha descrito que de manera general el CSE presenta muchas más ASNC que el CEE. Sin embargo, también se ha mostrado que dentro el CSE hay un porcentaje de casos (42%) que presentan unas características genéticas que no corresponden con el fenotipo, por lo que en estos casos puede ser de gran ayuda el uso del perfil genético de ASNC para realizar una clasificación más correcta. Además, se ha sugerido el uso de la combinación de la determinación de p53 por inmunohistoquímica y del número de copias de la *CCNE1* para clasificar los casos dentro del grupo serous-like. Para finalizar, los resultados del tercer objetivo han mostrado que el CSE, por una parte, presenta unas alteraciones moleculares clonales como las variantes somáticas en el gen *TP53* y ganancias en los genes *CCNE1* y *PIK3CA*, y por otra parte, que su heterogeneidad intratumoral está caracterizada principalmente por las ASNC de los genes. Además, se ha remarcado que uno de los genes principalmente afectados por esta heterogeneidad es el gen *ERBB2*, que es una diana terapéutica ampliamente utilizada.

ABSTRACT

The two most common subtypes of endometrial cancer (CE) are endometrioid carcinoma (CEE) and serous carcinoma (CSE). Differential diagnosis between these two types is not always easy, there are cases with very doubtful and ambiguous histological features that make it difficult their final diagnostic. These two tumor types have different molecular profiles and differential prognosis, making the CSE the most aggressive subtype and with the most unfavorable prognosis. At present, the importance of intratumoral heterogeneity (HIT) can be played out in the most aggressive behaviors of tumors and in resistance to treatment, as in the case of CSE. In recent years, the advent of high-performance technologies such as NGS (*Next Generation Sequencing*) and MLPA (*Multiple Ligase-dependent Probe Amplification*) have offered the option of extending molecular knowledge of cancer in a way more accurate and resolute.

For these reasons, the first two objectives of this dissertation have been to analyze in the two most common types of endometrial cancer, various molecular alterations that allow us to obtain specific genetic profiles, and to classify more precisely and objective these tumor types. To preform genetic alteration analyzes have been used the application of targeted sequencing (NGS) with a personalized panel of 40 genes to determine the altered genes with respect to somatic variants, and the MLPA technique to determine somatic alterations in the number of copies (ASNC) of 106 genes. Finally, the third objective was focused on understanding the role that intratumoral heterogeneity plays in CSE. The effect of HIT has been studied at two levels of molecular alteration, at the level of the genes determining somatic variants and the number of copies of the genes, using the two previous techniques.

The results of the first objective of this thesis have shown the suitability of our personalized NGS study as an additional molecular tool to confirm the histological classification of CE. This strategy seems interesting as a tool for classifying tumors with unusual and ambiguous microscopic findings. The results of the second objective have described that in general the CSE has many more ASNCs than the EEC. However, it has also been shown that in the CSE there is a percentage of cases (42%) that have genetic characteristics that do not correspond to the phenotype, and therefore in these cases the use may be very helpful of the ASNC gene profile for more accurate classification. In addition, the combination of p53 determination by immunohistochemistry and the number of *CCNE1* copies has been suggested to classify cases within the group *serous-like*. Finally, the results of the third objective have shown that CSE, on the one hand, has clonal molecular alterations such as the somatic variants in the *TP53* gene and gains in the *CCNE1* and *PIK3CA* genes, and on the other hand, that its intratumoral heterogeneity is characterized mainly by the ASNC of the genes. It has also been emphasized that one of the genes that is most affected by this heterogeneity is the *ERBB2 gene*, which is a widely used therapeutic target.

ÍNDEX

RESUMS	i
ESTRUCTURA DE LA TESI	xi
CAPÍTOL 1: Introducció	1
1 EL CÀNCER	3
1.1 Cronologia dels estudis moleculars del càncer	5
1.2 Evolució clonal i subclonal del càncer	12
1.3 Heterogeneïtat genètica del càncer	14
2 TÈCNiques D'ALT RENDIMENT PER L'ESTUDI GENÒMIC DEL CÀNCER	16
2.1 Seqüenciació de nova generació (NGS)	16
2.2 <i>Multiple Ligase-dependent Probe Amplification</i> (MLPA)	30
3 EL CÀNCER D'ENDOMETRI	37
3.1 L'endometri	37
3.2 Epidemiologia i factors de risc	41
3.3 Presentació clínica, eines diagnòstiques i característiques macroscòpiques	43
3.4 Classificacions	44
3.5 Estadis i estratificació del risc	52
4 ALTERACIONS MOLECULARS DEL CÀNCER D'ENDOMETRI	54
4.1 Alteracions moleculars identitàries dels grups moleculars TCGA	55
4.2 Altres alteracions moleculars	64
4.3 Teràpia molecular dirigida	68
CAPÍTOL 2: Hipòtesis i Objectius	71
1 HIPÒTESIS	73
2 OBJECTIUS	74
CAPÍTOL 3: Materials i Mètodes	75
1 MOSTRES	78
2 EXTRACCIÓ D'ADN	80
3 <i>MULTIPLEX LIGASE-DEPENDENT PROBE AMPLIFICATION</i> (MLPA)	80
4 <i>TARGETED SEQUENCING</i> (NGS)	83
5 SEQÜENCIACIÓ SANGER	89
6 IMMUNOHISTOQUÍMICA (IHQ)	91

Índex

7	SILVER IN SITU HYBRIDIZATION (SISH)	93
8	ANÀLISI BIOESTADÍSTICA.....	94
CAPÍTOL 4: Resultats.....		97
1	ALTERACIONS GENÈTIQUES A NIVELL DE VARIANTS SOMÀTIQUES (VS).....	100
1.1	Mètriques i estadístiques de l'anàlisi de <i>targeted sequencing</i> (NGS).....	101
1.2	Descripció de les variants somàtiques i gens alterats	105
1.3	Anàlisi d'associació entre gens i tipus histològic	111
1.4	Importància de les variants somàtiques detectades	113
1.5	Anàlisi de doble <i>cluster</i> jeràrquic.....	115
1.6	Validació mitjançant seqüenciació Sanger	117
1.7	Validació mitjançant immunohistoquímica	117
1.8	Classificació molecular amb el sistema alternatiu al TCGA	120
1.9	Integració de resultats.....	122
2	ALTERACIONS SOMÀTIQUES EN EL NÚMERO DE CÒPIES (ASNC) DELS GENS	124
2.1	Descripció de les alteracions somàtiques en el número de còpies dels gens	125
2.2	Gens alterats i localitzacions cromosòmiques implicades.....	129
2.3	Anàlisi d'associació entre gens i tipus histològic	131
2.4	Anàlisi de doble <i>cluster</i> jeràrquic.....	132
2.5	Seqüenciació Sanger del gen <i>POLE</i>	134
2.6	Anàlisi immunohistoquímica de p53	134
2.7	Integració de resultats.....	135
3	HETEROGENEÏTAT INTRATUMORAL (HIT) EN EL CARCINOMA SERÓS D'ENDOMETRI	137
3.1	Heterogeneïtat intratumoral a nivell d'ASNC	138
3.2	Heterogeneïtat intratumoral a nivell de VS.....	150
CAPÍTOL 5: Discussió.....		159
CAPÍTOL 6: Conclusions.....		175
BIBLIOGRAFIA		179
ANNEX		195

ESTRUCTURA DE LA TESI

ESTRUCTURA DE LA TESI

La tesi s'estructura en un total de 6 capítols. En el primer dels capítols es presenta una introducció a la temàtica d'estudi, fent èmfasi en l'estat actual de la recerca del càncer endometri i en les oportunitats que els avanços en la biologia molecular poden aportar en aquest camp. En el segon dels capítols es presenten les hipòtesis i objectius específics de la tesi, mentre que en el tercer capítol es descriuen els materials i mètodes emprats per a la consecució d'aquests objectius. Per últim, el capítol 4 mostra els resultats obtinguts i en el capítol 5 es discuteixen els resultats obtinguts abans de presentar les conclusions de la tesi (Capítol 6). Cal apuntar que, tot i que no es tracta d'una tesi amb format d'articles científics, els resultats corresponents als objectius específics OE1 i OE3 han estat publicats en revistes indexades dintre el JCR-SCI. Aquests articles es presenten a l'annex II. La figura 1 mostra de manera esquemàtica l'estructura de la tesi i la interrelació entre capítols.

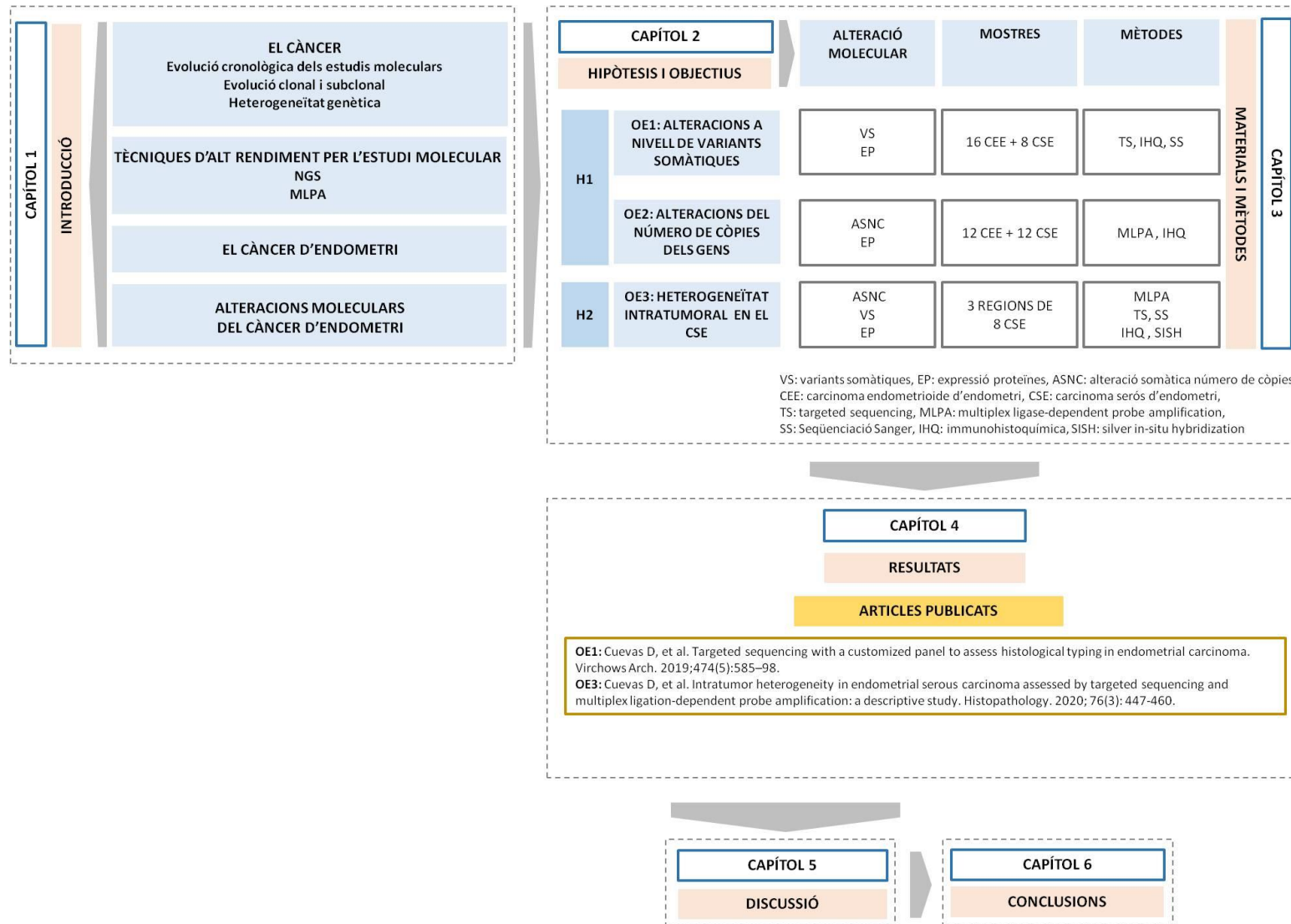


Figura 1. Estructura esquemàtica de la tesi

CAPÍTOL 1

INTRODUCCIÓ

1 EL CÀNCER

El càncer és un terme genèric que designa un grup ampli de malalties, més de 200, que poden afectar a qualsevol part del cos. Aquesta malaltia és un greu problema de Salut Pública a nivell mundial, a Europa per l'any 2018 es va estimar una incidència aproximada de 3,9 milions de casos nous, i de 1,93 milions de morts.¹ A Espanya s'estima que per a l'any 2035 aquesta xifra sigui de 315.413 casos nous. Segons les dades recollides a l'any 2016, a l'estat espanyol el càncer va suposar la segona causa de mort, després de les malalties del sistema circulatori. A la figura 2 es mostra la distribució de la incidència dels tumors més freqüents a Espanya l'any 2017.²

De vegades també es fa referència al càncer com tumors malignes o neoplàsies malignes. Cadascun d'aquest tipus de càncer posseeix unes característiques particulars que fa que es puguin arribar a considerar com a malalties independents, amb les seves causes, evolucions i tractaments específics.

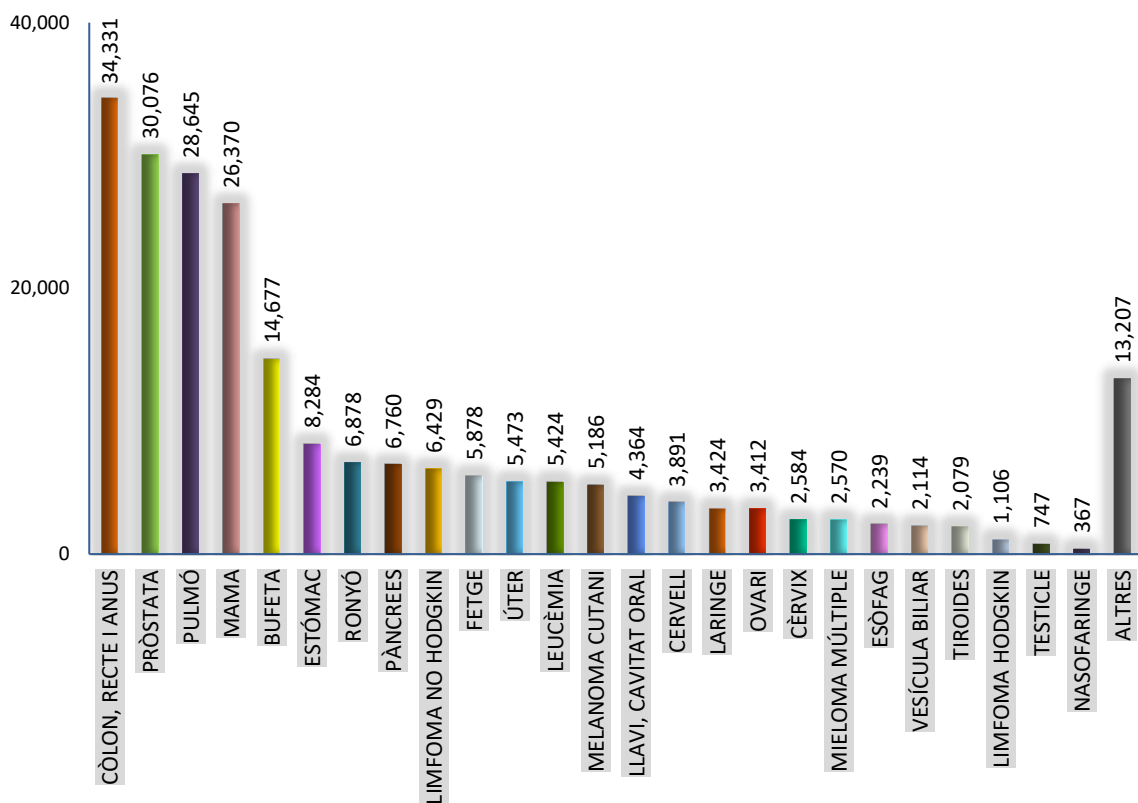


Figura 2. Incidència dels tumors més freqüents a Espanya l'any 2017.²

Capítol 1

El càncer s'esdevé quan les cèl·lules d'un determinat teixit, a causa de la incorporació d'alguna o algunes alteracions moleculars, deixen de respondre a les senyals de control del seu creixement. Així les característiques definitòries del càncer són la capacitat de divisió accelerada i sense control de les seves cèl·lules, i la capacitat d'aquestes d'estendre's més enllà dels seus límits habituals envaint teixits adjacents i disseminant-se a altres òrgans mitjançant el procés anomenat metàstasi. A la vegada, el càncer també es coneix com una malaltia d'origen genètic i epigenètic, ja que es sap que el seu origen i la seva progressió és deguda a l'acumulació d'alteracions moleculars en les seves cèl·lules que li permeten adquirir les característiques definitòries conegudes com els *hallmarks* del càncer (figura 3).³⁻⁵

El gran repte en l'estudi del càncer és esbrinar i entendre com es van adquirint aquestes alteracions, com afecten a la progressió dels tumors i a la capacitat de metàstasi amb l'objectiu de realitzar diagnòstics precoços, pronòstics acurats, i aconseguir una medicina eficient i personalitzada per al tractament del càncer.

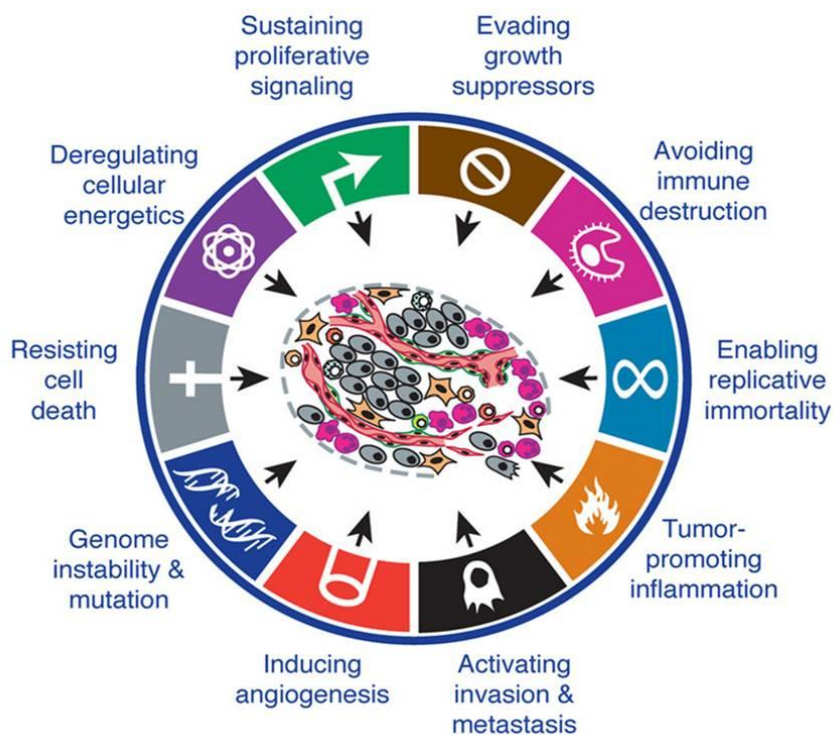


Figura 3. Els 10 *hallmarks* del càncer.⁴

1.1 CRONOLOGIA DELS ESTUDIS MOLECULARS DEL CÀNCER

A continuació es detalla com ha estat l'evolució històrica dels estudis moleculars del càncer, des dels inicis de la humanitat fins a l'actualitat (figura 4).

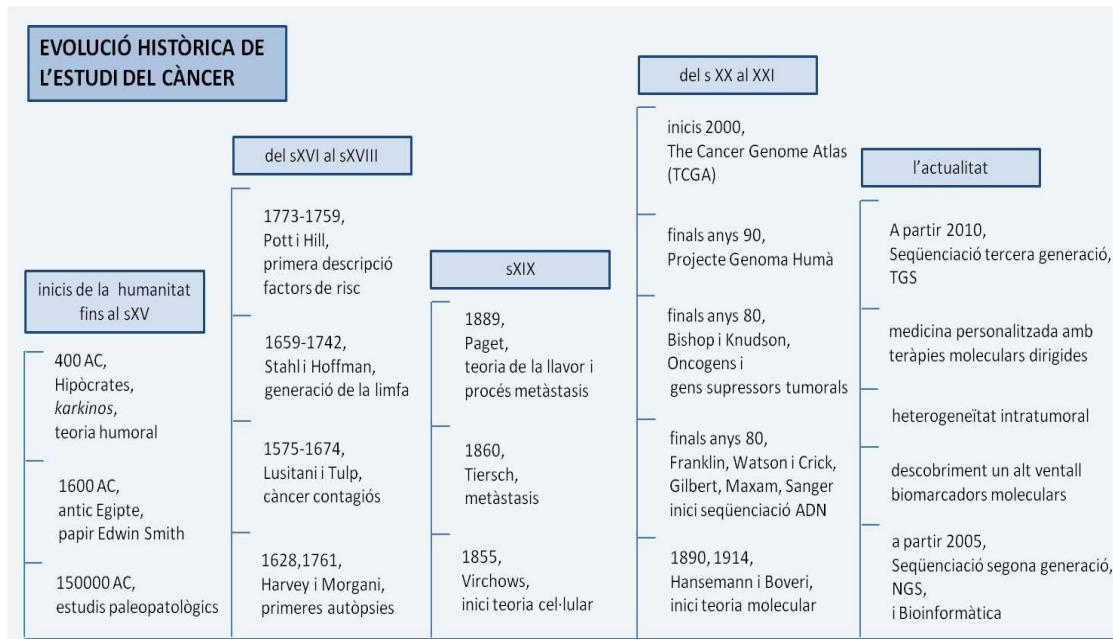


Figura 4. Evolució històrica de l'estudi molecular del càncer.

DELS INICIS DE LA HUMANITAT AL SEGLE XV

En l'actualitat pensem que el càncer és una malaltia moderna conseqüència dels estils de vida actuals i dels canvis mediambientals que s'han anat produint en les últimes dècades; però, es tracta d'una malaltia que ha acompanyat a la humanitat des de la seva aparició.

Estudis paleopatològics han mostrat evidències de lesions comparables amb el càncer en humans des de fa més de 150000 anys. Les primeres referències documentades al càncer les trobem en papirs de l'antic Egipte com el papir d'Edwin Smith (1600 aC) on es descriuen diferents tipus d'úlceres o tumors de la regió de la mama i es diu que no hi havia tractament per a la determinada condició. Però el primer que va batejar aquesta malaltia va ser el grec Hipòcrates (400 aC), el pare de la medicina. Ell la va anomenar *karkinos*, paraula que amb grec vol dir cranc, la va descriure com la presència d'unes lesions ulceroses cròniques, de vegades endurides que es desenvolupen progressivament i sense control, expandint-se pels teixits com si fossin les potes d'un cranc. Més tard, el metge romà Celsus (28-50 aC) va traduir la paraula grega *karkinos* al llatí

Capítol 1

com càncer. Cent anys després el metge romà Galeno (130-200dC) va utilitzar la paraula *onkos* que significa inflamació per descriure els tumors.

Fins aquesta època, les causes de l'origen del càncer havien estat atribuïdes als Déus en l'antic Egipte, o a partir d' Hipòcrates, segons la seva teoria humoral, al desequilibri dels quatre humors i a l'excés de la bilis negra. Cal dir que durant l'època medieval, sobretot per la influència del Cristianisme, es va abandonar l'interès científic iniciat pels grecs, i no va ser fins arribar al Renaixement que va ressorgir l'interès per la ciència.⁶⁻⁸

DELS SEGLES XVI AL XVIII

Entre els segles XVI i XVIII es van realitzar interessants aportacions observacionals i científiques per poder entendre millor el càncer. Així tenim les autòpsies de Harvey (1628) i Juan Morgani de Pàdua (1761) que van portar a una millor comprensió de la circulació de la sang i del càncer. També els estudis de Lusitani (1575-1642) i Tulp (1593-1674) que conclouien que el càncer era contagiós. Els estudis de Stahl (1659-1734) i Hoffman (1660-1742) van suggerir que el càncer era el resultat de la fermentació i la degeneració de la limfa. També, cap al s XVIII comencen a sorgir altres conceptes relacionats amb el càncer com els factors de risc en el seu desenvolupament. Així Pott (1773) va relacionar l'efecte de la pols del carbó i el càncer d'escrot, i Hill (1759) va ser el primer en relacionar el tàbac amb el càncer.⁶⁻⁸

EL SEGLE XIX

A partir del s XIX va ser quan van començar a sorgir les teories relacionades amb la patologia cel·lular gràcies a l'avenç en el camp de la microscòpia. L'alemany Muller va suggerir que el càncer estava format per cèl·lules i no per limfa. En aquest àmbit, un dels grans passos el va donar el seu estudiant Virchow al 1855 qui va demostrar que tota cèl·lula provenia d'una altra cèl·lula, i que per tant les cèl·lules canceroses també devien provenir d'altres cèl·lules, proposant la teoria de la irritació crònica on deia que el càncer s'expandia com un líquid. Al 1860 l'alemany Karl Thiersch va determinar que el càncer es dispersava a través d'un procés anomenat metàstasi on eren les cèl·lules del tumor les que es dispersaven, i no a través d'un fluid no identificat. També en aquest segle es troba l'aportació de Paget que va postular la teoria de la llavor i el terreny per explicar el procés de la metàstasi.⁶⁻⁸

ELS SEGLES XX I XXI

Entre finals del s XIX i inicis del s XX van començar a sorgir les idees de les primeres teories moleculars de la mà de David von Hansemann (1890) i Theodor Boveri (1914), sent els primers en proposar que el càncer era causat per unes alteracions cromosòmiques, o sigui que el càncer era una malaltia causada per alteracions al genoma. Tot i això, no va ser fins a finals dels anys 80, quan es van poder confirmar aquestes teories gràcies als diversos estudis de R. Franklin, F. Sanger, W. Gilbert, A. Maxam, J. Watson i F. Crick, relacionats amb l'estructura de l'ADN (Àcid Desoxiribonucleic) i amb les primeres idees de seqüenciació del genoma.

En aquest període es van descobrir els gens causants del càncer: els oncogens i els gens supressors tumorals, i també es van establir les primeres classificacions dels tipus d'alteracions genètiques que els podien afectar com les substitucions de nucleòtids, les alteracions en el número de còpies dels cromosomes i gens, i els reordenaments cromosòmics (figura 5). El primer oncogen va ser el gen *SRC* (*SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase*) descobert per Bishop gràcies a les seves investigacions amb el Virus del Sarcoma de Rous, i el primer gen supressor tumoral va ser *RB1* (*RB Transcriptional Corepressor 1*) descobert per Knudson amb els seus estudis del retinoblastoma. En aquests anys també es va començar a reconèixer el paper de l'epigenètica com un altre factor mediador del desenvolupament del càncer.

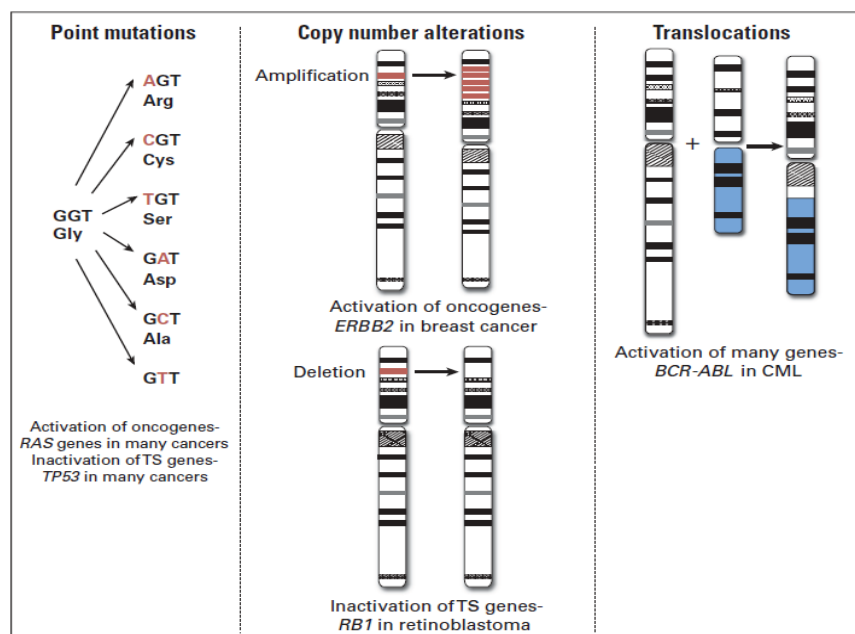


Figura 5. Els tipus d'alteracions moleculars més freqüents en càncer.¹⁰

Capítol 1

Aquests estudis i molts altres que van anar sorgint a posteriori van posar en evidència la gran complexitat molecular del càncer. Cap a finals dels anys 90, gràcies als avenços tecnològics i a l'augment de la capacitat analítica en l'àmbit de la biologia molecular es van començar a definir projectes moleculars molt ambiciosos a nivell mundial. Un d'ells va ser el Projecte del Genoma Humà que va començar el 1990 i va finalitzar l'any 2001, aproximadament. Amb la disponibilitat de la seqüència del genoma es va començar a desenvolupar un nou camp de recerca, l'estudi del genoma del càncer que tenia com a objectiu principal identificar llocs recurrents d'alteració molecular en els diferents tipus de càncer. D'acord amb això, es va engegar el projecte *The Cancer Genome Atlas* (TCGA) que va començar al 2006 i va acabar al 2009.^{5,9,10}

L'ACTUALITAT

El camp de l'estudi de la biologia molecular del càncer, des dels inicis del s XXI fins l'actualitat, ha viscut un procés de constant revolució tecnològica. Aquesta revolució, sobretot, ha estat associada a la millora de la tecnologia de la seqüenciació, la qual permet desxifrar la seqüència de nucleòtids que conformen un fragment d'ADN o ARN (Àcid Ribonucleic). L'inici d'aquesta revolució data cap a l'any 1977 amb l'adveniment de la seqüenciació de primera generació liderada per Frederick Sanger, que va definir el mètode de seqüenciació de terminació de cadena, basat en l'ús de dideoxinucleòtids.^{11,12,13} Amb aquesta laboriosa tecnologia, on s'ha d'anar seqüenciant mostra a mostra i fragment a fragment s'aconsegueix seqüenciar de manera contínua un fragment de fins a 1000pb. Una de les fites més importants de la seqüenciació Sanger, gràcies a l'ús de l'estratègia de seqüenciació *shotgun*, va ser aconseguir la primera seqüenciació del genoma humà complert (3 bilions de bases de longitud) a l'any 2001 amb un temps de 13 anys (1988-2001) i un cost d'uns 2700 milions de dòlars.¹⁴

A partir d'aquest moment va sorgir la necessitat de millorar aquesta tecnologia, per tal de fer-la més ràpida i amb més rendiment. Així va aparèixer a l'any 2005 l'anomenada seqüenciació de segona generació (*Next Generation Sequencing* (NGS)) o també anomenada seqüenciació massiva, la qual veritablement va revolucionar el camp de la seqüenciació de l'ADN. A diferència de la seqüenciació Sanger les longituds dels fragments en la NGS oscil·len entre 30 i 450 pb. L'avantatge principal de la NGS és que té la capacitat de seqüenciar paral·lelament, i de forma massiva, milions de fragments d'ADN en un únic procés, d'una o diferents mostres a la vegada, en un temps rècord i amb un cost cada vegada més reduït. Inicialment aquesta tecnologia era molt cara, però en l'actualitat s'han abaratit els costos d'una manera excepcional.¹⁵

Ara amb la NGS es pot seqüenciar un genoma humà sencer amb un preu aproximat de 1000 dòlars amb un temps d'unes dues setmanes (Illumina 2016). Així en aquest camp ja s'ha deixat de complir la llei de Moore, la qual preveia que es duplicués la capacitat de seqüenciar cada dos anys, i ara és cada 6-9 mesos (figura 6). En l'actualitat la NGS s'està utilitzant en el camp del càncer en aspectes com el diagnòstic, el seguiment de la malaltia i en l'avaluació de l'eficàcia dels tractaments oncològics.¹⁶

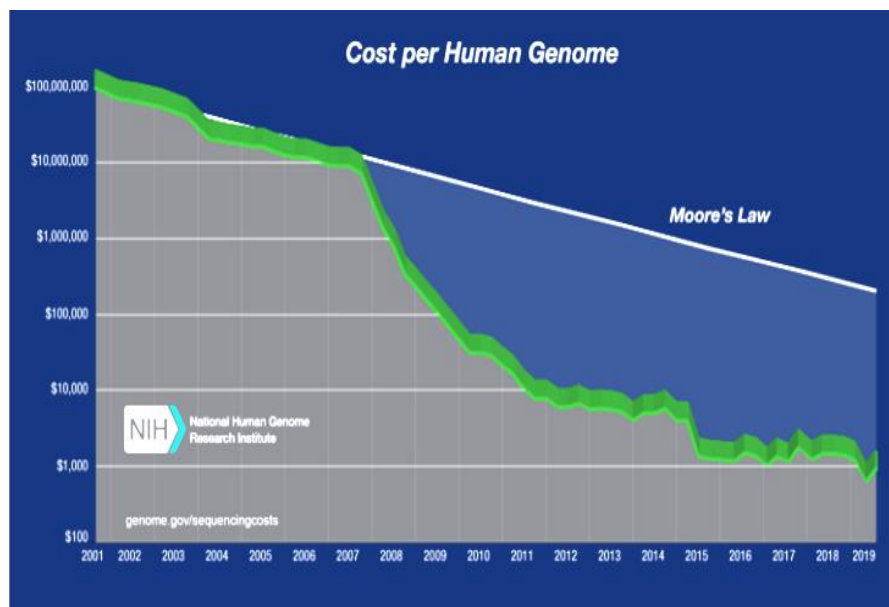


Figura 6. Evolució del cost de la seqüenciació del genoma humà en els últims 18 anys.¹⁶

A remolc de la NGS, a partir de l'any 2010 van començar a sorgir les tecnologies de seqüenciació de tercera generació (*Third Generation Sequencing (TGS)*) amb l'objectiu de poder seqüenciar fragments molts més llargs que el que ofereix la NGS i resoldre els problemes i reptes que aquesta genera. Les tecnologies TGS permeten realitzar la seqüenciació d'una única cèl·lula (*single cell sequencing*) evitant processos d'enriquiment, i optimitzant el cost i el temps. Dues d'aquestes noves tecnologies són la desenvolupada per Oxford Nanopore Technologies (ONT) i la de Pacific Biosciences (PacBio) basada en la metodologia SMRT (*Single Molecule Real Time*). Aquesta novedosa tecnologia té la fita de poder arribar a tenir plataformes tecnològiques on es pugui fer una anàlisi integrada del transcriptoma, epigenoma, genoma, metaboloma, l'anotació clínica i la informació fenotípica.^{14,17}

Capítol 1

Per tal de poder analitzar tota la informació molecular generada a partir d'aquestes tecnologies d'alt rendiment (el que s'anomena *big data*) ha estat necessari la incorporació d'un altre camp interdisciplinar com és la bioinformàtica, encarregada d'aplicar les ciències computacionals i les tecnologies de la informació per al tractament de les dades moleculars.

Amb tots aquests estudis s'ha aconseguit identificar una gran quantitat i diversitat de biomarcadors tumorals com: una gran diversitat de mutacions, translocacions, amplificacions i pèrdues gèniques, sobreexpressió o pèrdua d'expressió de proteïnes, detecció d'exosomes, de microARNs, de cèl·lules circulants tumorals i d'ADN circulant tumoral. A més a més, també han posat de manifest l'existència de l'heterogeneïtat genètica del càncer, tant a nivell intratumoral com intertumoral. A la taula 1 es mostra un llistat de diferents gens relacionats amb el càncer, els quals han estat descoberts gràcies a la revolució tecnològica. L'ús d'aquests biomarcadors està permetent tenir diagnòstics més específics i més precoços, avaluar millor el pronòstic, disposar d'una medicina més personalitzada amb teràpies dirigides, i predir d'una manera més eficient les respostes a les teràpies. D'altra banda, totes aquestes dades han permès conèixer un ampli ventall de processos moleculars que dirigeixen els processos de carcinogènesi, i han posat de manifest que és necessària una comprensió molt més profunda dels factors que governen la integritat del genoma i l'evolució del genoma del càncer.^{5,9,10}

Taula 1. Biomarcadors moleculars descoberts en la caracterització genòmica del càncer, i la seva relació amb els processos cel·lulars en què estan implicats.⁹

Table 2. Discoveries from Cancer Genome Characterization	
Cellular Process Altered by Genomic Alterations	Examples of Cancer Genes Discovered (or Extended to New Cancers ^a) by Genomics
RTK signaling	<i>EGFR</i> ^a , <i>ERBB2</i> ^{a,b} , <i>MET</i> ^{a,b} , <i>ALK</i> ^{a,b} , <i>JAK2</i> ^a , <i>RET</i> ^{a,b} , <i>ROS</i> ^{a,b} , <i>FGFR1</i> ^{a,b} , <i>FGFR2</i> ^a , <i>PDGFRA</i> ^{a,b} and <i>CRKL</i> ^a
MAPK signaling (oncogenes)	<i>KRAS</i> ^{a,b} , <i>NRAS</i> ^{a,b} , <i>BRAF</i> ^a and <i>MAP2K1</i> ^a
MAPK signaling (TSG)	<i>NF1</i> ^{a,b}
PI3K signaling (oncogenes)	<i>PIK3CA</i> ^a , <i>AKT1</i> ^a and <i>AKT3</i> ^a
PI3K signaling (TSG)	<i>PTEN</i> ^{a,b} and <i>PIK3R1</i> ^b
Notch signaling (oncogene or TSG)	<i>NOTCH1</i> ^c , <i>NOTCH2</i> ^c and <i>NOTCH3</i> ^b
TOR signaling (TSG)	<i>STK11</i> ^{a,b} , <i>TSC1</i> ^{a,b} and <i>TSC2</i> ^{a,b}
Wnt/ β -catenin signaling (TSG)	<i>APC</i> ^{a,b} and <i>CTNNB1</i> ^{a,b}
TGF- β signaling (TSG)	<i>SMAD2</i> ^{a,b} , <i>SMAD4</i> ^{a,b} and <i>TGFBR2</i> ^b
NF- κ B signaling (oncogene)	<i>MYD88</i> ^a
Other signaling	<i>RAC1</i> ^a , <i>RAC2</i> ^a , <i>CDC42</i> ^a , <i>KEAP1</i> ^b , <i>MAP3K1</i> ^b , <i>MAP2K4</i> ^b , <i>ROBO1</i> ^b , <i>ROBO2</i> ^b , <i>SLIT2</i> ^b , <i>SEMA3A</i> ^b , <i>SEMA3E</i> ^b , <i>ELMO1</i> ^d and <i>DOCK2</i> ^d
Epigenetics DNA methylation	<i>DNMT3A</i> ^b
Epigenetics DNA hydroxymethylation	<i>TET2</i> ^b
Chromatin histone methyltransferases	<i>MLL</i> ^{a,b} , <i>MLL2</i> ^b , <i>MLL3</i> ^b , <i>EZH2</i> ^c , <i>NSD1</i> ^b and <i>NSD3</i> ^b
Chromatin histone demethylases	<i>JARID1A</i> ^b , <i>UTX</i> ^b , <i>KDM5A</i> ^b and <i>KDM5C</i> ^b
Chromatin histone acetyltransferases	<i>CREBP</i> ^b and <i>EP300</i> ^b
Chromatin SWI/SNF complex	<i>SMARCA1</i> ^{a,b} , <i>SMARCA4</i> ^b , <i>ARID1A</i> ^b , <i>ARID2</i> ^b , <i>ARID1B</i> ^b and <i>PBRM1</i> ^b
Chromatin other	<i>CHD1</i> ^b , <i>CHD2</i> ^b and <i>CHD4</i> ^b
Transcription factor lineage dependency or oncogene	<i>MITF</i> ^a , <i>NKX2-1</i> ^a , <i>SOX-2</i> ^a , <i>ERG</i> ^a , <i>ETV1</i> ^a and <i>CDX2</i> ^a
Transcription factor other	<i>MYC</i> ^{a,b} , <i>RUNX1</i> ^b , <i>GATA3</i> ^b , <i>FOXA1</i> ^b , <i>NKX3.1</i> ^b , <i>SOX9</i> ^a , <i>NFE2L2</i> ^a and <i>MED12</i> ^d
Splicing	<i>SF3B1</i> ^d , <i>U2AF1</i> ^d , <i>SFRS1</i> ^d , <i>SFRS7</i> ^d , <i>SF3A1</i> ^d , <i>ZRSR2</i> ^b , <i>SRSF2</i> ^d , <i>U2AF2</i> ^d and <i>PRPF40B</i> ^d
RNA abundance	<i>DIS3</i> ^d
Translation/protein homeostasis/ubiquitination	<i>SPOP</i> ^d , <i>FBXW7</i> ^{a,b} , <i>WWP1</i> ^{a,b} , <i>FAM46C</i> ^d and <i>XBP1</i> ^d
Metabolism	<i>IDH1</i> ^a and <i>IDH2</i> ^a
Genome integrity	<i>TP53</i> ^{a,b} , <i>MDM2</i> ^a , <i>MSH</i> ^{a,b} , <i>MLH</i> ^{a,b} and <i>ATM</i> ^{a,b}
Telomere stability	<i>TERT</i> promoter mutations ^a
Cell cycle (oncogene)	<i>CCND1</i> ^{a,b} and <i>CCNE1</i> ^{a,b}
Cell cycle (TSG)	<i>CDKN2A</i> ^{a,b} , <i>CDKN2B</i> ^{a,b} and <i>CDKN1B</i> ^b
Apoptosis regulation	<i>MCL1</i> ^a , <i>BCL2A1</i> ^a and <i>BCL2L1</i> ^a

^aActivating mutation or amplification.
^bInactivating mutation or deletion.
^cBoth activating and inactivating genomic events observed.
^dEffect of mutations on protein function unknown.

1.2 EVOLUCIÓ CLONAL I SUBCLONAL DEL CÀNCER

Fa uns quaranta anys Peter Nowell (1976) va proposar el model d'evolució clonal del càncer.¹⁸ Aquest model s'ha pogut confirmar i matisar gràcies a la gran resolució de les anàlisis moleculars dels últims anys. Segons aquest model, la progressió tumoral segueix un microprocés evolutiu darwinià que pot durar dècades i que una gran part del temps es manté clínicament ocult. La iniciació d'aquest procés ve determinada per l'adquisició d'una o unes quantes alteracions moleculars genètiques o epigenètiques per part d'alguna cèl·lula del teixit. Aquestes alteracions li conferiran un avantatge selectiu positiu (per exemple un augment de la proliferació i menys mort programada (apoptosi) en relació a les altres cèl·lules del teixit, i desembocarà en una expansió clonal a partir d'aquesta cèl·lula fundadora o clon fundador. Cal esmentar que hi ha estudis que també parlen de la importància de la selecció negativa en aquest procés.¹⁹ Les alteracions moleculars que permeten aquest avantatge selectiu positiu i que són adquirides per la cèl·lula fundadora del tumor es coneixen com *driver* o conductores, en canvi, altres alteracions que no produeixen un avantatge se les anomena *passenger* o passatgeres.⁵

Un cop iniciat el procés tumoral, aquest es va repetint d'una manera dinàmica donant-se durant tota la seva vida, així a partir del clon inicial en cada etapa de la progressió es van generant noves mini poblacions de cèl·lules (subclons) que experimentaran noves expansions clonals (figura 7).²⁰ D'aquesta manera, en un tumor existiran unes alteracions *driver* i *passenger* clonals que estaran presents en totes les cèl·lules del tumor perquè provindran de la cèl·lula clonal fundadora, i unes alteracions *drivers* i *passengers* subclonals que provindran de poblacions subclonals generades a posteriori i que només estaran presents en unes determinades poblacions cel·lulars.

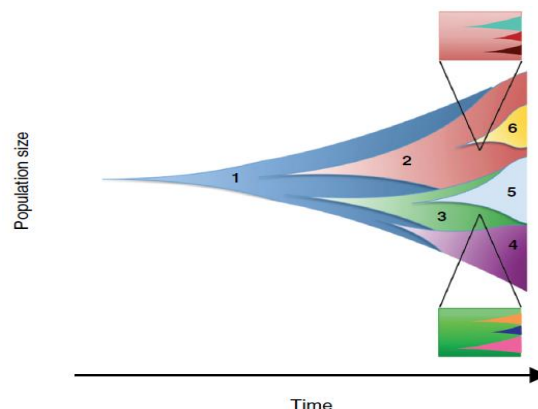


Figura 7. Il·lustració esquemàtica de l'evolució clonal del càncer al llarg del temps. Múltiples subclons evolucionen a partir del clon fundador (blau).²⁰

A mesura que el tumor progressa, expandint-se en un microambient tumoral que té limitació de recursos, les cèl·lules amb més capacitat de dividir-se, d'utilitzar més eficientment l'energia i de migrar aconseguiran un avantatge selectiu. Aquestes aptituds funcionals avantatjoses estan incloses a la revisió de Hanahan i Weinberg del 2011⁴, on es van redefinir els trets identificatius del càncer (*hallmarks*) (figura 3). Aquests trets són el manteniment de senyals proliferatives de manera sostinguda, l'evasió de senyals supressores de divisions cel·lulars, la resistència a la mort cel·lular, la immortalitat, la inducció de l'angiogènesi i la capacitat d'invasió i de metastatitzar. A més a més, també es van incloure altres punts clau en l'evolució del càncer com són la reprogramació del metabolisme energètic, l'evasió dels atacs del sistema immune i la capacitat de generar un microambient tumoral favorable. En la figura 8 es mostren els components que formen part d'aquest microambient tumoral.

Per resumir, podem dir que en un determinant moment del procés tumoral el tumor estarà format per una diversitat de poblacions cel·lulars subclonals genòmicament diferents entre elles. Tota aquesta diversitat genòmica intrínseca dels tumors és el que entenem per heterogeneïtat genètica intratumoral.²⁰

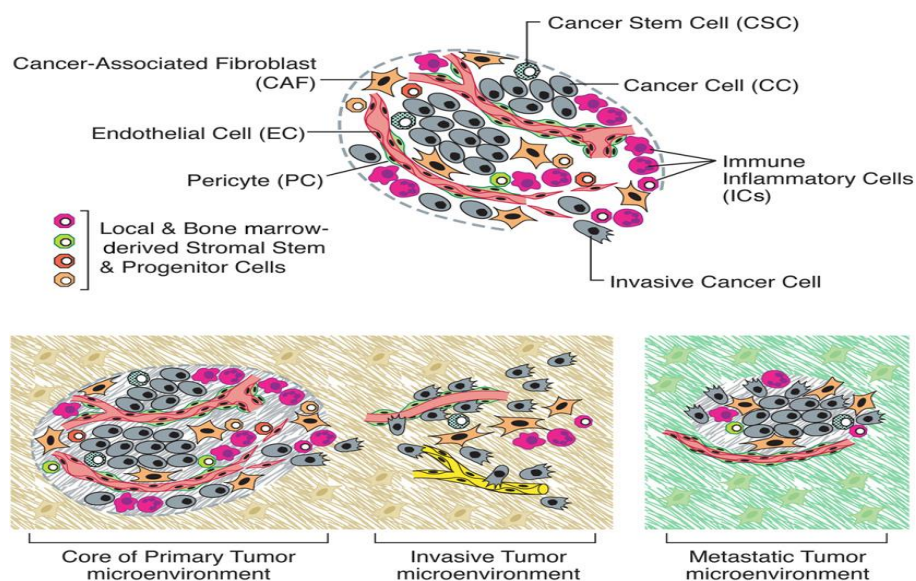


Figura 8. Components del microambient tumoral.⁴

1.3 HETEROGENEÏTAT GENÈTICA DEL CÀNCER

D'acord amb la teoria de l'evolució clonal, l'heterogeneïtat intratumoral (HIT) és atribuïda a l'adquisició contínua d'alteracions genètiques i epigenètiques heretables i a la selecció natural produïda pel canviant microambient tumoral dins un únic tumor. Cal assenyalar que també existeix l'heterogeneïtat intertumoral que es refereix al fet que tumors del mateix tipus tinguin característiques genètiques i fenotípiques diferents, per exemple tots els subtipus histològics i moleculars del càncer d'endometri.

La magnitud de l'HIT en cada tipus de càncer depèn de factors com la taxa de mutacions, la deriva genètica, l'estructura de les poblacions cel·lulars en els teixits, el microambient i la selecció natural. L'HIT pot ser estudiada a nivell espacial analitzant diferents regions d'un mateix tumor, i a nivell temporal analitzant el tumor primari i les seves metàstasis.^{21,22,23,20} Fins al moment actual hi ha pocs estudis d'HIT a nivell temporal sobretot degut a la dificultat d'obtenir mostres de les metàstasis ja que de vegades estan en llocs inaccessibles, o passa massa temps des del tumor primari fins a la metàstasis. L'estudi espacial es pot fer a dos nivells de resolució, a nivell de la macroheterogeneïtat, és a dir analitzant múltiples regions d'un mateix tumor amb les tècniques estàndards de seqüenciació massiva, i a nivell de la microheterogeneïtat que consisteix en estudiar subclons cel·lulars incipients que potser encara no han començat l'expansió clonal. Per fer l'anàlisi de la microheterogeneïtat s'han d'utilitzar tecnologies d'altíssima resolució més recents com la seqüenciació de ADN ultra profunda (*ultra-deep DNA sequencing*) i la seqüenciació cèl·lula per cèl·lula (*single cell sequencing*).²⁰

Principalment hi ha dos models que intenten reconstruir la dinàmica de la generació de les diferents poblacions subclonals durant el procés evolutiu, aquests són el model de l'evolució lineal seqüencial i el de l'evolució ramificada (*branched evolution*). El primer model es basa en el fet que una població cel·lular amb una alteració molecular avantatjosa té temps per escombrar la resta de població abans de l'adquisició d'altres alteracions avantatjoses, és a dir que la producció d'un clon no es dona fins que l'anterior hagi acabat. El segon model, el de *branched evolution*, es correspon amb l'escenari en què múltiples subclons diferents coexisteixen i competeixen durant el creixement tumoral.²⁴ Gràficament aquests models es representen a través d'arbres filogenètics que són estructures canòniques que permeten la descripció de les relacions genealògiques entre les diferents poblacions que formen part del tumor. A la figura 9 es mostra l'arbre filogenètic que exemplifica la *branched evolution*.^{21, 22,23, 20}

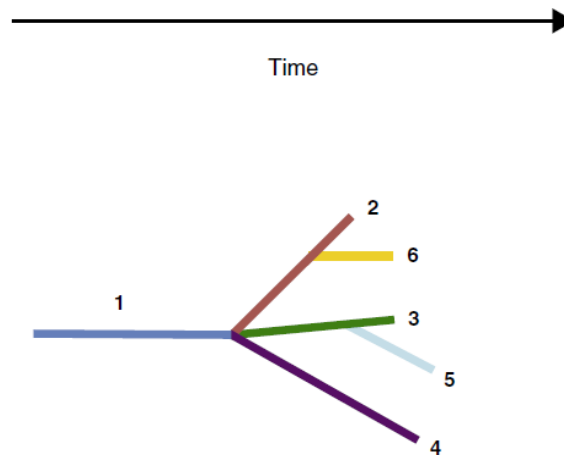


Figura 9. Arbre filogenètic que il·lustra la trajectòria de l'evolució ramificada (*branched evolution*) en el temps. El clon fundador (línia blava) representa el tronc de l'arbre.²⁰

Amb tots els estudis realitzats s'ha posat de manifest que l'HIT té una gran implicació en la progressió del càncer, i que representa un gran desafiament per al seu maneig i per a la medicina de precisió. L'HIT té tota una sèrie d'implicacions clíniques ja que és responsable: - dels biaixos en la presa de mostres degut a la disposició espacial aleatòria dels diferents subclons que fan que es donin diagnòstics moleculars incorrectes, - dels fracassos terapèutics degut a l'aparició de subclons resistents a les teràpies, - de la presència d'alteracions *driver* tractades que estan restringides només en una població subclonal, i per tant aquest tumor serà tractat de manera subòptima. En definitiva l'HIT és també responsable de la progressió del càncer i d'un pitjor pronòstic, de tal manera que s'ha parlat que l'HIT es pot considerar un potencial biomarcador pronòstic.^{20,22,25-28}

2 TÈCNiques D'ALT RENDIMENT PER L'ESTUDI GENÒMIC DEL CÀNCER

A continuació es descriuen les dues tecnologies moleculars d'alt rendiment que han estat utilitzades pel desenvolupament d'aquesta tesi. Aquestes són la tècnica de la seqüenciació de nova generació (NGS) i la tècnica MLPA (*Multiplex Ligase –dependent Probe Amplification*).

2.1 SEQÜENCIACIÓ DE NOVA GENERACIÓ (NGS)

La NGS (*Next Generation Sequencing*) és una tecnologia que ens permet determinar la seqüència d'ADN o ARN d'una mostra, amb la finalitat d'estudiar la variació genètica associada a una malaltia, com el càncer, o a un altre fenomen biològic. En el nostre cas ens centrarem en la NGS aplicada a la seqüenciació de l'ADN. Aquesta tecnologia permet l'anàlisi, en una mateixa carrera, de centenars de milers de gens en una o diverses mostres. La NGS permet l'anàlisi i detecció de diferents tipus de característiques genòmiques com variants de nucleòtids o mutacions, alteracions en el número de còpies, translocacions, entre d'altres. Enfront la seqüenciació automàtica Sanger, la NGS ofereix importants avantatges com: la millora del rendiment, per la disminució dels temps d'anàlisi i la major quantitat de mostres que es poden analitzar a la vegada; la necessitat de menys input d'ADN; i la capacitat detectar variants al·lèliques en freqüències més baixes. Tot i això, com que encara és una tecnologia bastant cara i dona certs errors de seqüència, en l'actualitat la seqüenciació Sanger encara es manté com la tècnica *gold-estandard* ja que és la utilitzada per validar els resultats obtinguts amb la NGS, i per continuar seqüenciant en la pràctica diària de molts laboratoris de biologia molecular.¹⁵

2.1.1 APLICACIONS

La NGS ofereix un ampli ventall d'aplicacions en funció de dues categories principals: les lectures (*reading*) i el comptatge de les lectures (*counting*). Les aplicacions de lectura es focalitzen en la pròpia seqüència d'ADN o ARN p.e. per identificar variants genòmiques o en l'assemblatge d'un genoma desconegut; en canvi les aplicacions de comptatge de lectures es basen en la capacitat de comptar quantitats de lectures i comparar aquests comptatges amb mostres tumorals i no tumorals p.e. per determinar nivells d'expressió i alteracions en el número de còpies dels gens.²⁹

Així, en el cas de l'aplicació per lectures tenim:

-la seqüenciació de tot el genoma (*Re-sequencing*) on es busquen variants en una determinada mostra al comparar-les amb un genoma de referència, en aquest cas tindriem la seqüenciació del genoma complert (*Whole Genome Sequencing* (WGS)).

-la seqüenciació amb enriquiment de regions d'interès que és una forma específica de reseqüenciació que es focalitza només en seqüenciar unes determinades regions d'interès del genoma (*Targeted Sequencing* (TS)) o tots els exons del genoma (*Whole Exome Sequencing* (WES)).

-l'assemblatge de-novo (*De-novo assembly*) que es basa en identificar seqüències genòmiques noves sense disposar d'un genoma de referència.

En el cas de les aplicacions basades amb el comptatge tenim:

-el ChiP-Seq/RIP-Seq que es basa en trobar dominis d'unió a proteïnes en ADN i ARN.

-el ARN-Seq que es basa en detectar noves isoformes i comparar nivells d'expressió gènica.

-el microARN-Seq que es basa en detectar i comptar microARNs.

Per al desenvolupament d'aquesta tesi s'ha utilitzat l'aplicació de *targeted sequencing* (NGS).

2.1.2 FLUX DE TREBALL

Tot i que en funció de l'aplicació i la plataforma de NGS utilitzada hi haurà uns requeriments específics, el protocol general de la NGS, per l'aplicació de *targeted sequencing* consisteix principalment en: A. la preparació de la llibreria, B. l'amplificació clonal i la seqüenciació massiva, i C. l'anàlisi de les dades (*pipeline* bioinformàtic) (figura 10).³⁰

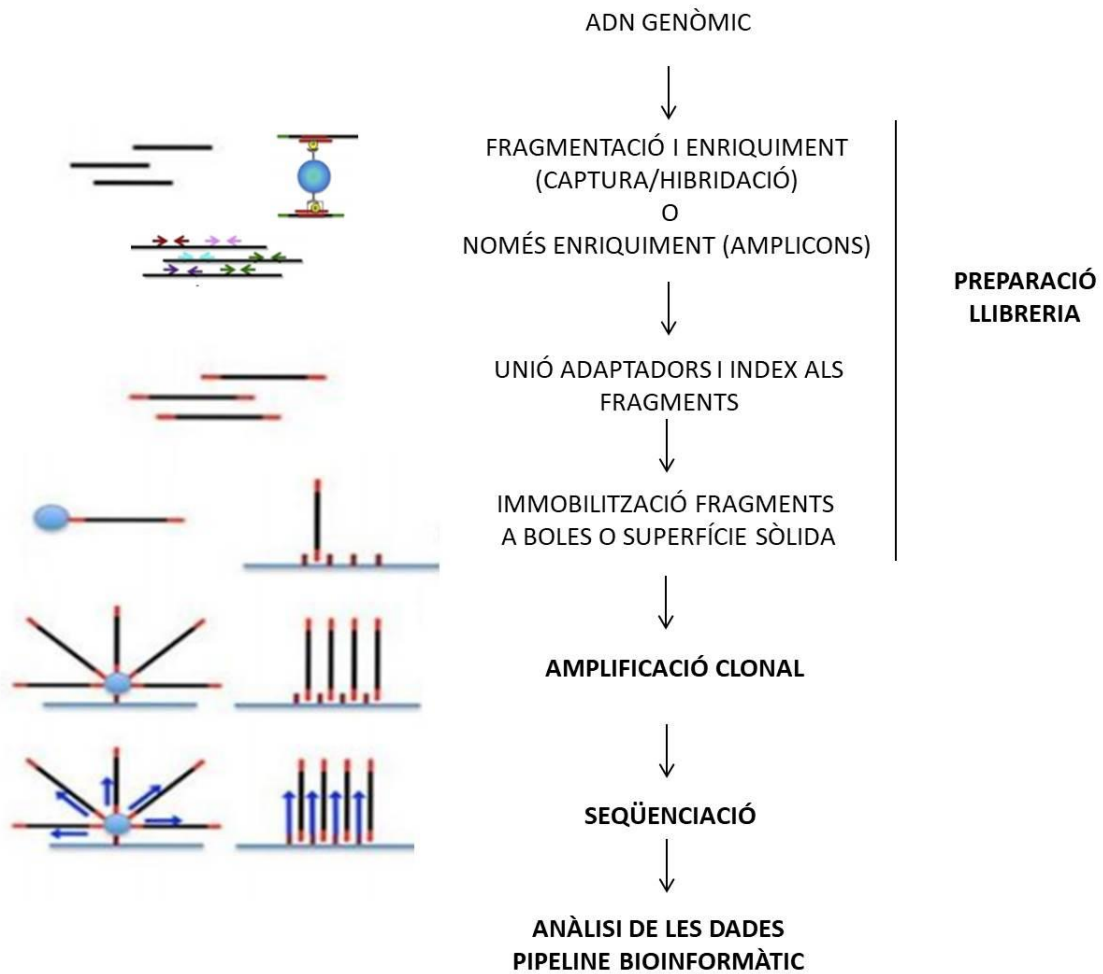


Figura 10. Esquema del flux de treball de la NGS en l'aplicació de *targeted sequencing*.³⁰

A. PREPARACIÓ DE LA LLIBRERIA

Per molts estudis de recerca en càncer, focalitzats en la seqüenciació de l'ADN, l'aplicació de WGS encara és molt costosa i cara, sobretot per les necessitats del processament computacional i informàtic. Aquest sobrecost i la seva complexitat són poc beneficioses quan es té l'objectiu d'estudiar regions genòmiques específiques, i quan es vol donar als resultats obtinguts una aplicació clínica. Per tal d'abordar aquest problema, en molts estudis del càncer s'ha adoptat l'aproximació de la seqüenciació dirigida (*targeted sequencing (TS)*) d'un llistat de gens (panell) associats a un determinat càncer o a un grup de càncers. Amb aquest enfocament s'aconsegueix millorar la cobertura de les regions específiques, analitzar diverses mostres a la vegada, simplificar les anàlisis i la interpretació del resultat, i reduir els costos totals de la seqüenciació.³¹

La preparació de la llibreria consisteix en preparar una col·lecció de fragments d'ADN d'una longitud específica per seqüenciar. En el cas de l'aplicació de *targeted sequencing* on l'objectiu és capturar del genoma complet només aquelles regions que siguin d'interès i es vulguin seqüenciar, la preparació de la llibreria inclou el pas de l'enriquiment. Depenent de la metodologia utilitzada es requerirà un pas previ de fragmentació de l'ADN genòmic. El tipus de fragmentació serà diferent en funció del tipus d'aplicació que utilitzem, pot ser una fragmentació amb un mètode físic o amb una digestió enzimàtica. A posteriori, els fragments d'ADN resultants de l'enriquiment s'han de lligar a unes seqüències específiques anomenades "adaptadors" que són les que permetran seqüenciar els fragments en una determinada plataforma de NGS. A més a més, en el cas que s'utilitzin diferents mostres a la vegada, els fragments d'ADN també s'han de lligar a unes seqüències "índex" que són necessàries per marcar l'ADN de cada mostra. L'aproximació d'utilitzar moltes mostres a la vegada, anomenada *pooling* o *multiplexing*, permet estalviar temps i costos en els experiments de seqüenciació.³¹

Existeixen dos mètodes d'enriquiment, el mètode basat amb una múltiple PCR (*Polymerase Chain Reaction*) amb generació d'amplicons, o el mètode basat amb la captura/hibridació (figura 11). La decisió de quins d'aquests mètodes escollir depèn de diferents factors com la plataforma de NGS, la quantitat d'ADN de partida, la quantitat de regions que es volen capturar, i el cost i el temps que es vol invertir. En el nostre cas vam escollir l'enriquiment amb múltiple PCR amb generació d'amplicons.^{32,33}

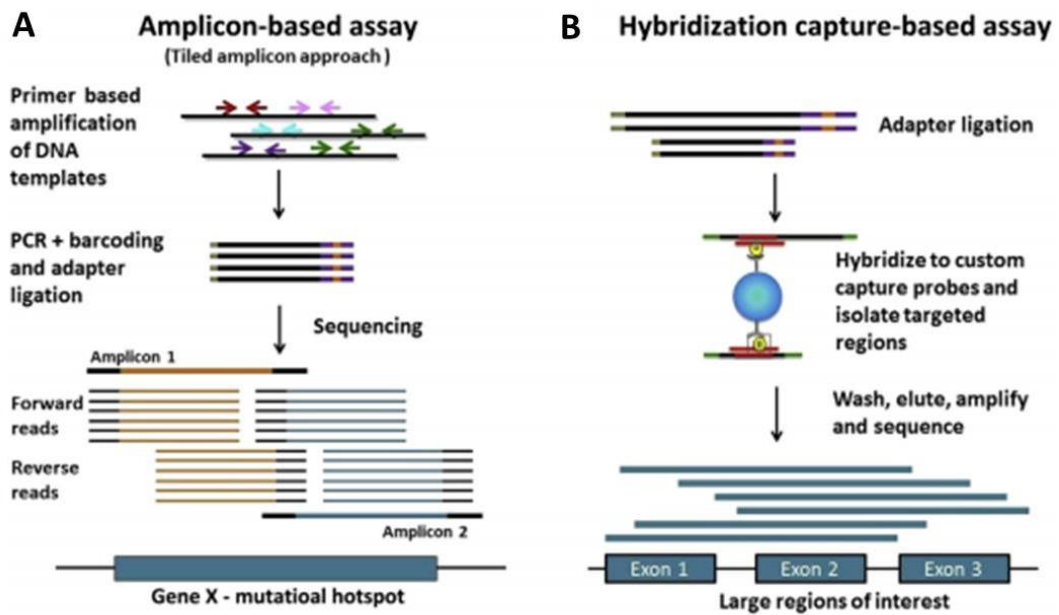


Figura 11. Il·lustració dels dos mètodes d'enriquiment, el basat en la generació d'amplicons (A) i el basat en la captura/hibridació (B).³⁰

El mètode de múltiple PCR amb generació d'amplicons es basa en amplificar les diferents regions d'interès utilitzant sondes específiques que flanquejaran aquestes regions. Abans de realitzar la PCR s'uneixen les sondes i es produeix un procés d'extensió-lligació de les regions d'interès. En l'amplificació els productes resultants de l'extensió-lligació s'amplifiquen amb l'ús de *primers* que, a més a més, afegiran als fragments anteriors les seqüències índex i els adaptadors. En el nostre cas els índex (i5, i7) i adaptadors (P5, P7) van ser els de la plataforma Illumina.

Respecte el mètode basat en la captura/hibridació cal dir que existeixen diferents tipus de sistemes d'enriquiment, però tots ells funcionen de manera similar. Aquests són els sistemes de captura per circularització, el d'hibridació en solució i el d'hibridació en fase sòlida. El primer pas és una fragmentació prèvia de l'ADN, després es produeix la captura o hibridació utilitzant sondes específiques biotinilades complementàries a les regions específiques d'interès, i a posteriori es realitza un aïllament final de les regions d'interès amb l'ús de boles d'estreptavidina.

Per seleccionar les regions específiques d'interès o panells de gens disposem de dues opcions. Una és utilitzar un kit d'una casa comercial que ja té un contingut de molècules preseleccionat, i l'altra és fer un disseny customitzat del panell específic de gens. Actualment, les cases comercials, disposen de *softwares* amb algorismes específics per fer aquests dissenys. La cobertura del disseny (*coverage*) dependrà de l'algorisme utilitzat i de diferents factors que poden afectar

l'eficiència de l'enriquiment. Un d'aquests factors, inherents a les seqüències d'ADN, són aquelles regions que tinguin valors extrems en els percentatges de guanines i citosines (<25% o >65%) ja que tenen un considerable impacte negatiu en l'eficiència de l'enriquiment, aquestes zones solen ser les 5'UTR/promotores i els primers exons de cada gen.

B. AMPLIFICACIÓ CLONAL I SEQÜENCIACIÓ

Abans de la seqüenciació, la llibreria s'ha d'immobilitzar en una superfície sòlida i s'ha d'amplificar clonalment per tal d'incrementar les senyals i que puguin ser detectades durant la seqüenciació. Durant aquest procés, cada fragment únic de la llibreria és lligat, mitjançant els adaptadors, o a una bola (figura 12B) o a una cel·la de flux (*flow cell*) (figura 12A) i és amplificat per poder crear un conjunt de còpies idèntiques o també anomenat clon o *cluster*. Tots els fragments amplificats de manera clonal seran seqüenciats a la vegada en una plataforma de NGS.

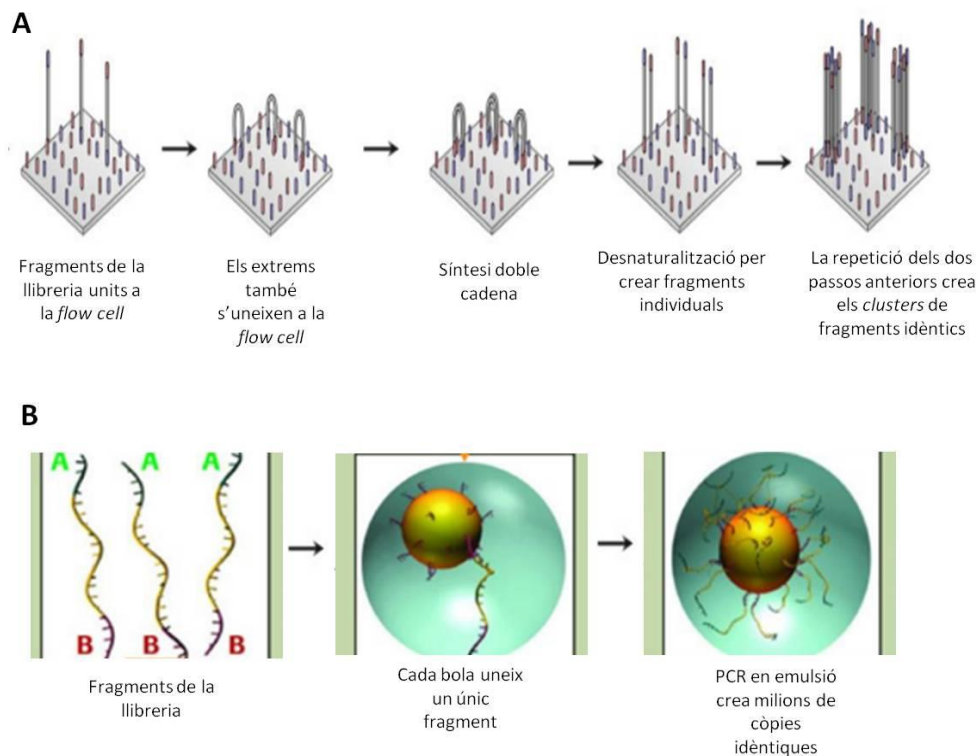


Figura 12. Il·lustració dels dos mètodes d'amplificació clonal, A) el de *bridge amplification*, i B) el de PCR d'emulsió. ³⁰

Capítol 1

Durant aquests anys han sorgit diferents plataformes comercials per realitzar la NGS com Roche-454, Termo Fisher Scientific i Illumina. Aquestes plataformes es diferencien en diversos aspectes com el tipus de mètode d'amplificació clonal, el tipus de química de seqüenciació, la longitud de les lectures obtingudes, la capacitat d'emmagatzematge de dades, els temps dels protocols, i el cost, entre d'altres (taula 2).³⁴

Taula 2. Descripció de les característiques de les principals plataformes de NGS.

	ROCHE/454	TERMO FISHER SCIENTIFIC	SOLiD (4, 4hq, PI)	ILLUMINA
Equips	GS FLX, GS Junior	ION TORRENT PGM (xip 316, xip 318, Proton)	SOLiD (4, 4hq, PI)	HiSeq, MiSeq
Capacitat dades	700Mb, 70Mb	100 Mb, 1Gb, <1Gb	100 Gb, 300 Gb, 77 Gb	1000 Gb, 15Gb
Longitud lectures (pb)	1000, 700	200	75	2x125, 2x300
Mètode seqüenciació	Piroseqüenciació amb detecció Ppi alliberats	Seqüenciació per síntesi amb detecció H ⁺ alliberats	Seqüenciació per lligació amb detecció sondes fluorescents	Seqüenciació per síntesi amb terminació reversible i detecció nucleòtids fluorescents
Càrrega a l'equip	Els adaptadors dels fragments s'uneixen als oligonucleòtids de les boles, i les boles es distribueixen en pous d'una placa picotiter	Els adaptadors dels fragments s'uneixen als oligonucleòtids de les boles, i les boles es distribueixen en pous d'una placa semiconductora	Els adaptadors dels fragments s'uneixen als oligonucleòtids complementaris de les boles, i les boles s'uneixen a una <i>flow cell</i>	Els adaptadors dels fragments s'uneixen als oligonucleòtids complementaris de la <i>flow cell</i>
Amplificació clonal	PCR en emulsió	PCR en emulsió	PCR en emulsió	<i>Bridge amplification</i>
Enzim seqüenciació	ADN polimerasa	ADN polimerasa	ADN lligasa	ADN polimerasa

Els seqüenciadors **454 de Roche** van ser els primers en sortir al mercat i en l'actualitat estan en desús. En aquesta plataforma l'amplificació clonal es realitza mitjançant una PCR en emulsió (figura 12B), i el mecanisme de seqüenciació està basat en la piroseqüenciació, que és un tipus de seqüenciació per síntesi. La PCR en emulsió es basa en l'ús d'unes boles que estan recobertes per oligonucleòtids específics, que són complementaris als adaptadors que els fragments de la llibreria ja tenen incorporats. Cadascuna de les boles s'associa a un únic fragment, i cada complex bola/fragment és aïllat en una micel·la d'aigua-oli que conté els reactius necessaris per la PCR. Així a l'acabar la PCR en cada micel·la es produeixen aproximadament un milió de còpies de cada fragment d'ADN anclat a la superfície de les boles. Per a la piroseqüenciació aquestes boles es disposen individualment en els pous d'una placa picotiter on iterativament flueixen els quatre nucleòtids. Quan un nucleòtid s'incorpora a la cadena creixent, el pirofosfat alliberat desencadena una cascada de reaccions que alliberen una quantitat de llum proporcional al nombre de nucleòtids incorporats. La placa picotiter està acoblada a una càmera que va captant imatges després de cada ronda de nucleòtids. Aquest sistema permet la seqüenciació de seqüències llargues que faciliten el seu alineament amb el genoma de referència, però la seva limitació és que tenen una alta taxa d'error en regions amb homopolímers de més de 6pb, fet que posa en entredit la fiabilitat de les seqüències obtingudes. Els dos equips que funcionen amb aquesta tecnologia són el GS FLX que produeix 1.000.000 de lectures amb un rendiment de 700 Mb, i el GS Junior que produeix unes 100.000 lectures amb un rendiment de 70 Mb.

Termo Fisher Scientific disposa de la plataforma SOLiD (*Sequence by Oligo Ligation Detection*) i la Ion Torrent. En el cas de SOLiD s'utilitza la PCR en emulsió per a l'amplificació clonal i el mecanisme de seqüenciació és per lligació. La PCR en emulsió utilitza unes boles magnètiques que porten unides l'adaptador P1 a la superfície. En aquest cas, abans de la seqüenciació, les emulsions es trenquen per alliberar les boles magnètiques dels microreactors i es purifiquen mitjançant diferents rentats. Posteriorment, les boles s'uneixen covalentment a una superfície sòlida (*flow cell*) per fer la seqüenciació. En aquesta reacció és necessari el *primer* universal, complementari a l'adaptador, una lligasa i una barreja de quatre sondes octamèriques fluorescents basades en la codificació de dos bases per interrogar la seqüència d'ADN motlle. Aquest tipus de seqüenciació té la particularitat que enlloc d'una polimerasa utilitza una lligasa, i que utilitza un sistema basat en la codificació de dos bases. Com a resultat s'obté una seqüència de nucleòtids interpretada a partir d'una seqüència de colors mitjançant l'ús de les bases de codificació. Així, durant la lligació cada nucleòtid de la cadena és interrogat dos cops i el color després en cada cicle de lligació és emmagatzemat digitalment. Una de les avantatges d'aquest

Capítol 1

mètode és que presenta la menor taxa d'errors de seqüència. Aquesta tecnologia té un rendiment per carrera d'entre 77 i 300 Gb dependent de l'equip que s'utilitzi.

El seqüenciador **Ion Torrent** va ser el primer seqüenciador de taula (*benchtop*) que va sortir al mercat. Aquests seqüenciadors estan guanyant molta popularitat degut al seu baix cost i a la facilitat dels seus requeriments tècnics i de manteniment. En aquest cas també es parteix d'una PCR en emulsió, i per la seqüenciació utilitza la tecnologia de semiconductors basada en la detecció d'ions d'hidrogen que s'alliberen durant la seqüenciació per síntesi de la cadena d'ADN. En aquest cas les boles resultants de l'amplificació clonal són dipositades en uns *chips* semiconductors amb milions de pous, on com en els altres casos, cada pou conté una única molècula d'ADN motlle. Els pous són seqüencialment inundats amb nucleòtids sense modificar. Quan un nucleòtid introduït és complementari al següent nucleòtid desaparellat, aquest és incorporat a la cadena per la polimerasa i s'allibera un ió hidroegn. Aquest ió alliberat canvia el pH de la solució i aquest canvi de voltatge és detectat per un ISFET (*Ion-Sensitive Field-Effect Transistor*), que tradueix els impulsos elèctrics a una seqüència d'ADN, sense la necessitat de conversió de les senyals intermitges. En aquest cas s'abarateix el cost gràcies a que no s'utilitzen nucleòtids modificats ni processos òptics. Aquesta tecnologia té un rendiment per carrera d'entre 100 Mb i més d'1Gb en funció del *chip* que s'utilitzi.

Per últim, els sequenciadors **illumina** són els que en l'actualitat lideren la NGS de lectures curtes. La seva tecnologia combina l'amplificació clonal amb pont (*bridge amplification*) i la seqüenciació per síntesi (*Sequencing By Synthesis* (SBS)) amb terminació reversible (figura 13). Dels aparells que disposa tenim l'equip MiSeq, que és el de mitjana capacitat, amb un rendiment per carrera d'unes 15 Gb i és el que s'ha utilitzat per aquesta tesi doctoral. Primerament, els milions de fragments que conformen la llibreria s'han d'immobilitzar sobre una superfície sòlida i transparent anomenada *flow cell*. La immobilització es dona gràcies a que aquesta superfície presenta uns oligonucleòtids que són complementaris als adaptadors P5 i P7 que havien estat afegits als fragments en el punt de l'amplificació de la llibreria. Després de la immobilització es produeix una amplificació clonal en pont (figura 12A), on per cada fragment immobilitzat inicialment es generen múltiples còpies idèntiques, que és el que s'anomena generació de *clusters*. L'ideal és que cada *cluster* tingui aproximadament 1000 molècules idèntiques d'ADN per tal d'obtenir suficient intensitat de senyal per a una detecció fiable. La *bridge amplification* genera centenars de milions de *clusters* i permet una densitat de seqüenciació molt elevada. Un cop la generació de *clusters* s'ha completat, els fragments ja estan llestos per a ser seqüenciats.

La seqüenciació per síntesi (SBS) utilitza el mètode basat en la terminació reversible. Per a aquesta reacció són necessaris els *primers* de seqüència, la polimerasa i els quatre nucleòtids modificats com a terminadors reversibles, cadascun marcat amb una molècula fluorescent diferent. En cada cicle, d'acord amb la complementarietat de bases de cada fragment, s'incorporen els nucleòtids terminadors reversibles marcats amb les seves respectives fluorescències. En aquest moment la llum fluorescent emesa es grava utilitzant una càmera, i posteriorment es desxifra amb un *software*. Així en el primer cicle s'uneixen els nucleòtids complementaris a la primera posició de cadascun dels fragments, i es capturen imatges amb moltes senyals fluorescentes diferents. Per poder començar el següent cicle, una reacció química modifica el nucleòtid terminador ja incorporat desprenent-ne la fluorescència i tornant-lo apte per a què s'uneixi el següent nucleòtid marcat. D'aquesta manera, cicle a cicle, es va obtenint la seqüència del fragment. La base de la SBS és molt similar a la seqüenciació Sanger.^{35,36}

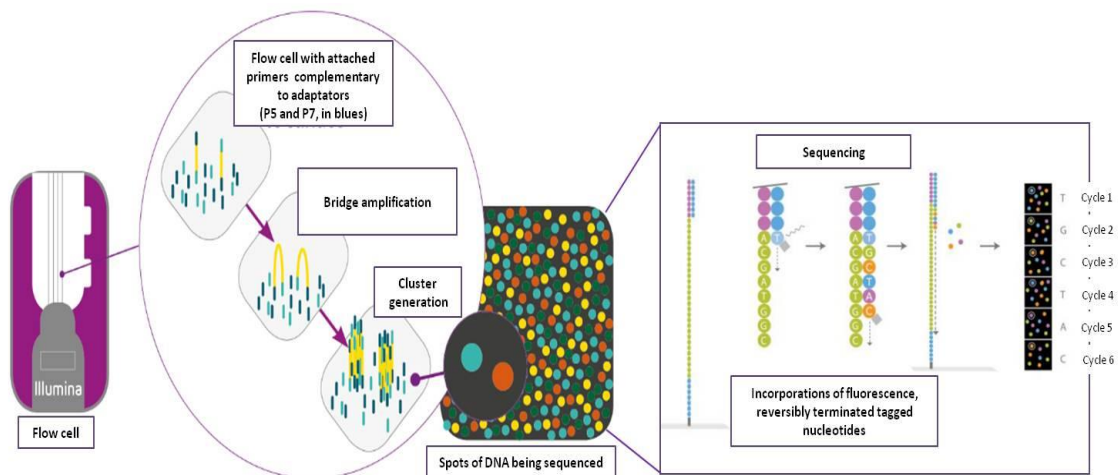


Figura 13. Il·lustració de l'amplificació clonal i la seqüenciació (SBS) que tenen lloc dintre la *flow cell* a la plataforma MiSeq d'Illumina.

C. ANÀLISI DE LES DADES

Els experiments de NGS generen grans quantitats de dades complexes. Tot i que, cada plataforma té els seus algoritmes d'anàlisi, a la vegada també n' existeixen d'altres lliures als comercials. Tots ells comparteixen un fluxe de treball o *pipeline* similar, i utilitzen mètriques comuns per avaluar la qualitat de les dades obtingudes. En relació a aquesta anàlisi de dades, la bioinformàtica és la disciplina encarregada d'aplicar les ciències computacionals i les tecnologies de la informació per a l'anàlisi de les dades biològiques.

L'anàlisi de les dades, o anàlisi bioinformàtica, permet convertir les senyals amb dades, les dades en informació interpretable, i la informació en coneixements científics aplicables. De manera generalitzada el *pipeline* (figura 14) inclou una anàlisi primària, una anàlisi secundària i una anàlisi terciària. Breument, la primària consisteix en convertir les senyals digitals crues generades en els equips de NGS en nucleòtids (bases). La secundària inclou l'alineament de les lectures amb el genoma de referència o l'assemblatge de novo de les lectures, i la subsegüent detecció de les variants. I l'anàlisi terciària és qui posa en context tota la informació generada associant a les variants detectades les diferents anotacions descriptives. A més a més, un altre component important del *pipeline* és l'obtenció de les estadístiques i mètriques de la seqüenciació que permeten fer un control de qualitat al llarg de tot el *pipeline*.^{37,38,29}

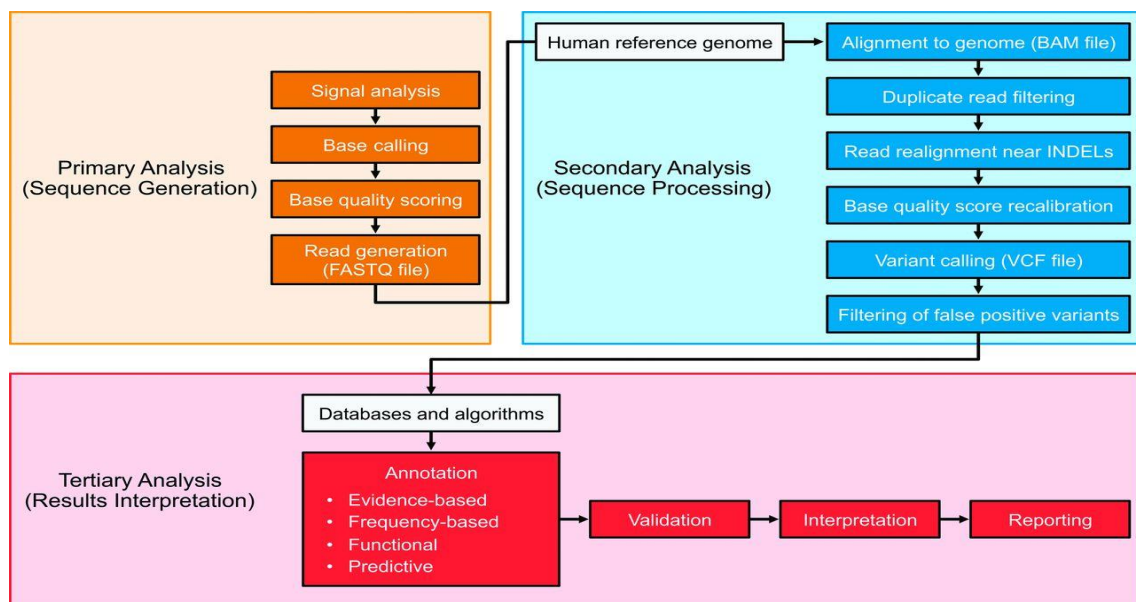


Figura 14. Diagrama de flux que il·lustra els components principals del *pipeline* bioinformàtic.³⁷

A continuació es desglossen cadascun dels passos del *pipeline*:

a. Generació de seqüències

Es tracta d'una anàlisi automàtica que, normalment, té lloc en els propis equips de NGS. L'objectiu d'aquesta anàlisi és convertir les senyals crues (dades primàries o *raw data*) que s'obtenen als equips de NGS en seqüències de nucleòtids (lectures), i assignar a cada nucleòtid (base) un paràmetre de qualitat. Els arxius amb que s'extreuen aquestes dades tenen el format estàndard "fastq". Aquests arxius contenen les dades de seqüència de totes les lectures generades per cada mostra i per tant són arxius molt pesants i difícilment manejables, els quals en els passos posteriors s'hauran de transformar per fer-los més manejables. El paràmetre de qualitat que s'assigna a cada base és el que coneixem com a *Phred* o *Q score*, que ja es va introduir per poder avaluar la qualitat de les bases de la seqüenciació Sanger.³⁹ Aquest paràmetre és un valor enter relacionat logarítmicament amb la probabilitat que una determinada base s'hagi assignat per error (P) que s'obté amb la següent fórmula: $Q = -10 \log_{10}P$. Així, p.e. un valor de $Q=30$ significa que la base té una probabilitat d'1 entre 1000 de ser incorrecta, que és el mateix que una probabilitat de que sigui correcta del 99,9%.

b. Alineament

El procés d'alineament té com a finalitat localitzar i organitzar cadascuna de les lectures en la posició correcta de la seqüència del genoma de referència. És molt important saber en quina versió del genoma de referència es vol treballar ja que es disposen de diferents versions. En el cas de *Homo sapiens* tenim la del *Genome Reference Consortium* (GRC), on la versió actual és la 38, que s'abreua GRCh38 (hg38). La versió anterior a aquesta era la GRCh37 (hg19). En el cas que no hi hagi seqüència de referència perquè no es coneix, el que s'intenta és col·locar les lectures una a continuació de l'altra sobreposant les bases coincidents i construint una seqüència consens. En l'actualitat es disposa de molts *softwares* que permeten transformar els arxius "fastq" en arxius d'alineament com el format SAM (*Sequence Alignment Map*) o MAP (*Binary Alignment Map*). L'arxiu BAM és el format binari equivalent al SAM. Alguns exemples són el BWA (*Burrows-Wheeler Aligner*) i el Bowtie2. En l'alineament també és molt important determinar les qualitats dels mapatges per tal de saber la probabilitat que una lectura hagi estat alineada incorrectament, i així poder seleccionar els alineaments que siguin més fiables i ens assegurin una correcta detecció de variants.

Hi ha *softwares* que ens permeten fer la visualització dels alineaments utilitzant els arxius d'alineament SAM/BAM. Un dels més utilitzats és l'IGV (*Integrative Genomics Viewer*).

c. Detecció de variants

Aquest pas permet detectar les diferències que hi ha entre la seqüència d'una mostra i la del genoma de referència. En NGS, quan trobem una diferència en una posició de la seqüència l'anomenem variant. Aquestes variants poden ser insercions, delecions, indels (inserció/deleció), SNV (*Single Nucleotide Variant*), ASNC (Alteració Somàtica del Número de Còpies) i reordenaments. Per assolir una correcta detecció de variants és molt important haver realitzat un filtrat òptim dels alineaments per tal de minimitzar els falsos positius i negatius. Existeixen *softwares* diferents anomenats *variant callers* per realitzar la detecció de variants, alguns d'ells són el SAMtools, el GATK, el VarScan2 i el MuTect. La majoria d'aquests *softwares* generen els resultats en arxius en format VCF (*Variant Caller Format*). Cal dir que per a la detecció d'ASNC i reordenaments es requereixen *softwares* amb algorismes complexos i necessiten assaigs amb elevades profunditats de lectura i lectures *paired-end*. En aquest apartat és també molt útil l'ús del *software* IGV, ja que ens permet avaluar les profunditats de lectura de cada posició, i observar si les variants estan en zones de poca o molta profunditat de lectura, pròximes a regions repetitives o a regions de seqüències complicades, entre d'altres.

d. Anotació

Aquest pas té l'objectiu de posar en context totes les variants detectades i així caracteritzar-les. Això significa incorporar informació provinent de diferents bases de dades per tenir de cadascuna de les variants una gran quantitat d'informació diversa com: les profunditats de lectura, les freqüències al·lèliques, el gen, la posició en el gen, el codó, l'efecte sobre el transcrit, l'efecte sobre la proteïna, saber si estan descrites en bases de dades de polimorfismes, si estan descrites com a variants somàtiques, saber quina és la predicció de patogenicitat *in-silico*, entre molta d'altra informació. Per a tot això, a la vegada, es fa una consulta d'una enorme quantitat d'informació de diferents bases de dades públiques. Per a això és necessari utilitzar tota la informació disponible i actualitzada que contenen les bases de dades públiques i que fan referència a diferents categories. Un dels programes utilitzats per l'anotació és l'AnnoVar.

e. Filtrat de variants

Aquest és l'últim pas de l'anàlisi i fa referència a la selecció de només aquelles variants que ens interessin en funció del disseny i de l'objectiu final del propi estudi. Per fer aquesta selecció s'han d'anar fent filtratges de les diferents categories de l'anotació i anar establint valors líndars de profunditat de lectura, de freqüència al·lèlica, de quins efectes sobre els transcrits i proteïna volem etc. I Així tenir un llistat de totes les variants identificades que són les definitives.

f. Control de qualitat: estadístiques i mètriques

Les estadístiques i les mètriques mostren els valors característics i representatius del tot el procés de NGS. En cadascun dels passos del *pipeline* és important realitzar un control de qualitat d'aquests paràmetres per tal de poder detectar irregularitats en el procés de seqüenciació.

Per al control de qualitat de les dades primàries ("fastq") tenim p.e. la distribució de les qualitats mitjanes de les bases en funció de la posició dins de les lectures, el percentatge mitjà de cadascun dels nucleòtids per cada posició dins de les lectures, la distribució de les qualitats mitjanes sobre totes les seqüències, la distribució del percentatge mitjà de GC (guanina/citosina) de les lectures, distribució de la llargada mitjana de les lectures, el percentatge mitjà de duplicats de la seqüència. Un dels *softwares* més utilitzats per fer aquestes anàlisis de qualitat és el FastQC.

Per al control de l'alineament tenim la profunditat de lectura (*read depth*) mitjana de les regions d'interès, el percentatge de nucleòtids amb un mínim de *read depth*, l'avaluació del rendiment de l'alineament, i l'avaluació de les regions no cobertes. En aquest apartat cal donar importància a la profunditat de lectura, que fa referència al número de lectures que tenim alineades en una determinada posició del genoma. L'ideal seria que tinguessin una quantitat de lectures uniformes alineades al llarg del genoma de referència, però la realitat no és així, i la profunditat de lectura no és uniforme, ja que la seqüenciació no és perfecta, que hi ha regions del genoma que són més difícils de seqüenciar i regions que són difícils d'alinejar. Les regions amb dificultat per seqüenciar són p.e. les regions amb alt contingut de nucleòtids GC (>65%), i les difícils d'alinejar són p.e. els elements repetitius, com les repeticions en tàndem i els pseudogens. És necessària una profunditat de lectura suficient (en funció de l'aplicació) per poder donar validesa a les variants detectades. Així quan es tenen valors de profunditat de lectura elevats es poden detectar variants amb baixes freqüències al·lèliques que estiguin distribuïdes heterogèniament en una mostra tumoral. En canvi amb valors insuficients hi haurà una incapacitat de poder detectar variants d'interès.

Per a la detecció de variants també hi ha paràmetres que ens donen la qualitat del genotipat i del nomenament de la variant. Algunes de les mètriques són el comptatge de variants, i la densitat de variants per una determinada regió.

2.2 MULTIPLE LIGASE-DEPENDENT PROBE AMPLIFICATION (MLPA)

La tècnica MLPA (*Multiple Ligase-dependent Probe Amplification*) disposa de diferents aplicacions, una d'elles és l'aplicació per detectar nivells de metilació de l'ADN, i l'altra permet avaluar la variació en el número de còpies de diferents gens de manera simultània. Molts dels estudis que s'han dut a terme per avaluar l'estat del número de còpies dels gens han utilitzat les tecnologies de *microarrays* de SNPs (*Single Nucleotide Polymorphism*), *arrays* de CGH (*Comparative Genòmic Hybridization*), i l'eina GISTIC (*Genomic Identification of Significant Targets in Cancer*). Aquestes permeten identificar regions cromosòmiques que estan significativament amplificades o delecionades. El seu inconvenient és que no permeten detectar de manera específica quins són els gens implicats en aquestes regions cromosòmiques. Per això la MLPA és una tècnica d'alt rendiment que permet avaluar les alteracions somàtiques en el número de còpies dels gens (ASNC) de manera directa sobre un conjunt específic de gens. A més a més, és menys costosa i més ràpida que les tècniques anteriors.⁴⁰⁻⁴³ Per al desenvolupament d'aquesta tesi aquesta tècnica ha estat la seleccionada per analitzar les alteracions somàtiques en el número de còpies dels gens.

De manera resumida, la MLPA es basa en la hibridació específica i simultània de múltiples sondes sobre diferents seqüències d'interès, la posterior amplificació amb una única reacció de PCR i la resolució final dels productes amplificats mitjançant una anàlisi d'electroforesi capil·lar. En cada anàlisi de MLPA per cada mostra tumoral també s'han d'analitzar un *pool* de mostres no tumorals (controls), per tal de poder determinar de manera relativa les variacions en el número de còpies dels gens entre les dues mostres.

Principalment es divideix en 5 passos :

A. DESNATURALITZACIÓ DE L'ADN GENÒMIC I HIBRIDACIÓ DE LES SONDES

En aquest primer pas l'ADN genòmic es desnatura i s'incuba durant tota la nit amb la mescla de sondes incloses als kits MLPA que hibridaran sobre les seqüències diana d'interès. Cadascuna de les sondes està constituïda per dos oligonucleòtids (LPO i RPO), que inclouen una seqüència d'hibridació específica a la seqüència diana (*hybridization sequence*) i una seqüència de reconeixement de primers (*forward (X)* en LPO i *reverse (Y)* en RPO) que serà igual en totes les sondes MLPA. A més a més, l'oligonucleòtid RPO conté una seqüència de longitud variable en funció de la sonda (*stuffer*) que és la que permet diferenciar les diferents sondes en funció de la seva longitud mitjançant l'anàlisi de fragments. (figura 15)

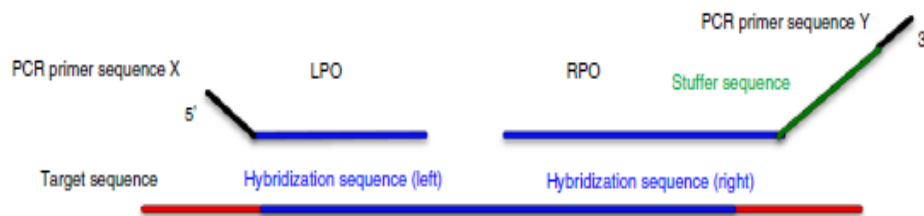


Figura 15. Primer pas de la tècnica MLPA: desnaturalització de l'ADN genòmic i hibridació de les sondes.⁴³

B. REACCIÓ DE LLIGACIÓ

Un cop hagin hibridat els dos oligonucleòtids de cada sonda, una lligasa uneix els extrems 5' i 3', i a posteriori els primers es poden unir a les seves seqüències complementàries per a que s'iniciï l'amplificació. En el cas que no s'hagi trobat una seqüència complementària no es produirà ni la hibridació ni la posterior amplificació. (figura 16)

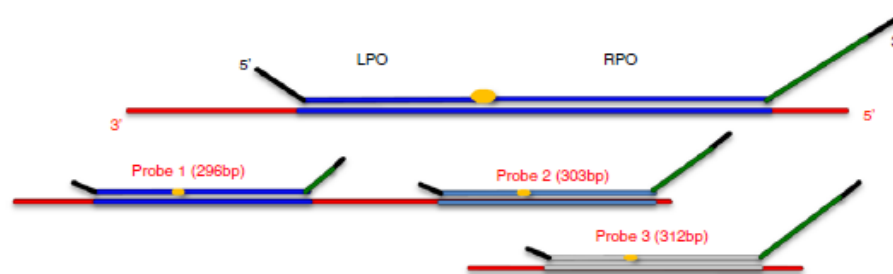


Figura 16. Segon pas de la tècnica MLPA: reacció de lligació. El punt groc indica el punt de lligació.⁴³

C. REACCIÓ D'AMPLIFICACIÓ PER PCR

Amb l'ús de només un parell de *primers* universals s'amplificaran tots els fragments. El *primer forward* està marcat amb un fluorocrom i és el que permet la posterior detecció dels fragments. (figura 17)

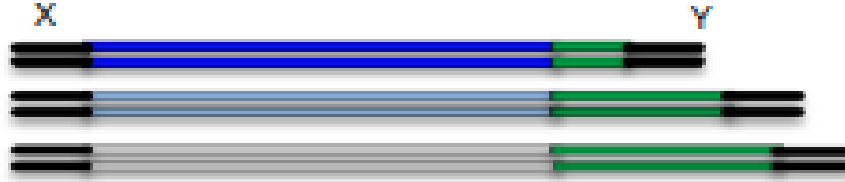


Figura 17. Tercer pas de la tècnica MLPA: reacció de PCR. La X i Y indiquen els *primers forward* i *reverse*, respectivament. El fragment marcat amb verd és la seqüència *stuffer*.⁴³

D. SEPARACIÓ DELS PRODUCTES RESULTANTS DE L'AMPLIFICACIÓ MITJANÇANT ELECTROFORESI CAPIL·LAR

A continuació els diferents fragments corresponents a l'amplificació de cada sonda es resolen a través d'una electroforesi capil·lar, gràcies a les diferents longituds que tenen els fragments. A posteriori s'analitzen els fragments resultants amb un *software* específic. El primer pas de l'anàlisi de fragments és avaluar si la tècnica s'ha produït correctament. Per a això, cada kit MLPA conté 9 fragments que permeten fer l'anàlisi de qualitat. Aquests són: un fragment de control que té una longitud de 92 nt (nucleòtids) que és la referència amb qui comparar els altres fragments de qualitat, els fragments Q amb unes longituds de 64, 70, 76, 82 nt que permeten avaluar si hi havia suficient ADN i si la lligació s'ha produït correctament (la seva alçada ha de ser inferior al 33% de l'alçada del fragment de 92 nt), els fragments D amb unes longituds de 88 i 96 nt que permeten avaluar si la desnaturalització ha estat correcta (la seva alçada ha de ser superior al 40% del fragment de 92 nt) i els fragments sexuals X-Y de 100 i 105 nt respectivament. (figura 18)

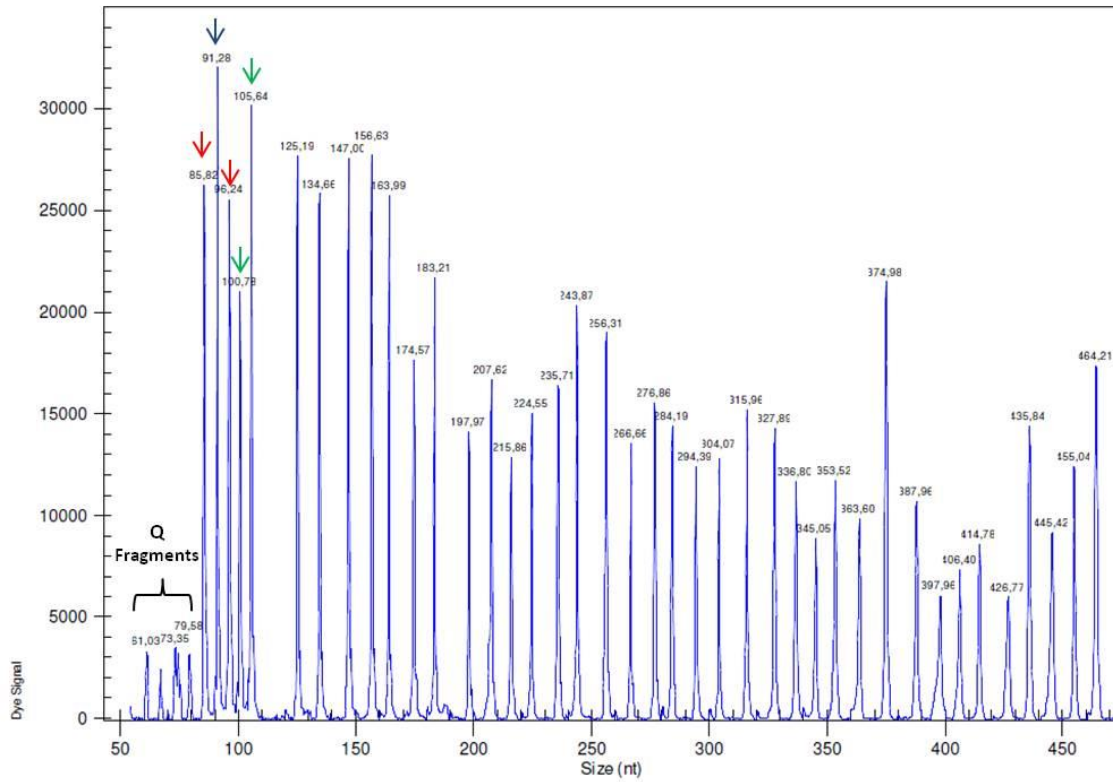


Figura 18. Quart pas de la tècnica MLPA: resolució dels fragments. Les fletxes vermelles indiquen els fragments D, la blava el fragment de control i les verdes els fragments X-Y. La resta de pics indiquen totes les sondes del kit que s'han amplificat.⁴³

E. ANÀLISI DE LES DADES I INTERPRETACIÓ DELS RESULTATS OBTINGUTS

Un cop s’ha avaluat que la tècnica ha funcionat correctament ja es pot procedir al posterior anàlisi de resultats. De cada mostra s’obtenen tants pics com sondes s’han utilitzat (figura 18), així es quantifica l’àrea de cada pic (sonda), i finalment totes les àrees són importades en un full Excel (figura 19).

mostra	sonda	tamany	àrea	control 1	sonda	tamany	àrea	control 2	sonda	tamany	àrea	control 3	sonda	tamany	àrea	mitjana de les àrees
tumoral	1	112	23127	no tumoral 1	1	112	55589	no tumoral 2	1	112	50967	no tumoral 3	1	112	48471	51675,66667
tumoral	2	136	17541	no tumoral 1	2	136	26756	no tumoral 2	2	136	23360	no tumoral 3	2	136	23085	24397
tumoral	3	142	63652	no tumoral 1	3	142	33691	no tumoral 2	3	142	30577	no tumoral 3	3	142	29057	31108,33333
tumoral	4	148	29029	no tumoral 1	4	148	25813	no tumoral 2	4	148	24609	no tumoral 3	4	148	24400	24907,33333
tumoral	5	153	18971	no tumoral 1	5	153	21364	no tumoral 2	5	153	20543	no tumoral 3	5	153	19769	20569,66667
tumoral	6	160	18783	no tumoral 1	6	160	21694	no tumoral 2	6	160	18784	no tumoral 3	6	160	16690	19122,66667
tumoral	7	166	26235	no tumoral 1	7	166	27373	no tumoral 2	7	166	26048	no tumoral 3	7	166	25017	26146
tumoral	8	172	26132	no tumoral 1	8	172	25854	no tumoral 2	8	172	23308	no tumoral 3	8	172	22746	23969,33333
tumoral	9	179	30718	no tumoral 1	9	179	26360	no tumoral 2	9	179	26147	no tumoral 3	9	179	24273	26593,33333
tumoral	10	185	12716	no tumoral 1	10	185	19971	no tumoral 2	10	185	18202	no tumoral 3	10	185	17946	18706,33333
tumoral	11	193	17775	no tumoral 1	11	193	18588	no tumoral 2	11	193	17636	no tumoral 3	11	193	18246	18166,66667
tumoral	12	202	17238	no tumoral 1	12	202	24931	no tumoral 2	12	202	21714	no tumoral 3	12	202	21017	22554
tumoral	13	211	32784	no tumoral 1	13	211	26405	no tumoral 2	13	211	24922	no tumoral 3	13	211	23499	24932
tumoral	14	220	22732	no tumoral 1	14	220	23336	no tumoral 2	14	220	20951	no tumoral 3	14	220	18583	20966,66667
tumoral	15	229	23001	no tumoral 1	15	229	26018	no tumoral 2	15	229	24168	no tumoral 3	15	229	24640	24938,66667
tumoral	16	238	28033	no tumoral 1	16	238	27255	no tumoral 2	16	238	28671	no tumoral 3	16	238	26648	27524,66667
tumoral	17	247	20186	no tumoral 1	17	247	21171	no tumoral 2	17	247	20041	no tumoral 3	17	247	19434	19882
tumoral	18	255	15778	no tumoral 1	18	255	18204	no tumoral 2	18	255	14804	no tumoral 3	18	255	15508	15538
tumoral	19	265	22719	no tumoral 1	19	265	21462	no tumoral 2	19	265	21409	no tumoral 3	19	265	21141	21334
tumoral	20	273	13983	no tumoral 1	20	273	19429	no tumoral 2	20	273	19159	no tumoral 3	20	273	18714	19100,66667
tumoral	21	283	37521	no tumoral 1	21	283	29301	no tumoral 2	21	283	32743	no tumoral 3	21	283	27576	29873,33333
tumoral	22	292	21793	no tumoral 1	22	292	18206	no tumoral 2	22	292	18210	no tumoral 3	22	292	18095	18170,33333
tumoral	23	301	23926	no tumoral 1	23	301	25642	no tumoral 2	23	301	28951	no tumoral 3	23	301	22855	25816
tumoral	24	310	38909	no tumoral 1	24	310	20427	no tumoral 2	24	310	21009	no tumoral 3	24	310	19812	20416
tumoral	25	319	17632	no tumoral 1	25	319	18887	no tumoral 2	25	319	17552	no tumoral 3	25	319	16461	16960
tumoral	26	328	26191	no tumoral 1	26	328	25234	no tumoral 2	26	328	26160	no tumoral 3	26	328	23954	24916
tumoral	27	337	38521	no tumoral 1	27	337	19637	no tumoral 2	27	337	21541	no tumoral 3	27	337	20419	20532,33333
tumoral	28	346	12639	no tumoral 1	28	346	11798	no tumoral 2	28	346	12371	no tumoral 3	28	346	12680	12283
tumoral	29	355	18102	no tumoral 1	29	355	23859	no tumoral 2	29	355	25085	no tumoral 3	29	355	22255	23733
tumoral	30	364	16639	no tumoral 1	30	364	13748	no tumoral 2	30	364	14647	no tumoral 3	30	364	14410	14267,66667
tumoral	31	373	12506	no tumoral 1	31	373	12183	no tumoral 2	31	373	15500	no tumoral 3	31	373	12108	13263,66667
tumoral	32	383	13113	no tumoral 1	32	383	18282	no tumoral 2	32	383	20750	no tumoral 3	32	383	17769	18933,66667
tumoral	33	391	16128	no tumoral 1	33	391	19474	no tumoral 2	33	391	20266	no tumoral 3	33	391	18888	19539,33333
tumoral	34	400	19046	no tumoral 1	34	400	20440	no tumoral 2	34	400	22108	no tumoral 3	34	400	20216	20921,33333
tumoral	35	409	23619	no tumoral 1	35	409	15908	no tumoral 2	35	409	20691	no tumoral 3	35	409	17467	18022
tumoral	36	418	28020	no tumoral 1	36	418	24547	no tumoral 2	36	418	25375	no tumoral 3	36	418	24327	24749,66667
tumoral	37	427	21951	no tumoral 1	37	427	20224	no tumoral 2	37	427	20953	no tumoral 3	37	427	18769	19982
tumoral	38	436	23419	no tumoral 1	38	436	18940	no tumoral 2	38	436	22228	no tumoral 3	38	436	19620	20262,66667
tumoral	39	445	25588	no tumoral 1	39	445	17125	no tumoral 2	39	445	20685	no tumoral 3	39	445	18174	18661,33333
tumoral	40	454	19600	no tumoral 1	40	454	13336	no tumoral 2	40	454	16005	no tumoral 3	40	454	15501	14947,33333
tumoral	41	463	21102	no tumoral 1	41	463	16772	no tumoral 2	41	463	19074	no tumoral 3	41	463	16576	17474
tumoral	42	472	18773	no tumoral 1	42	472	15519	no tumoral 2	42	472	17615	no tumoral 3	42	472	14843	15992,33333

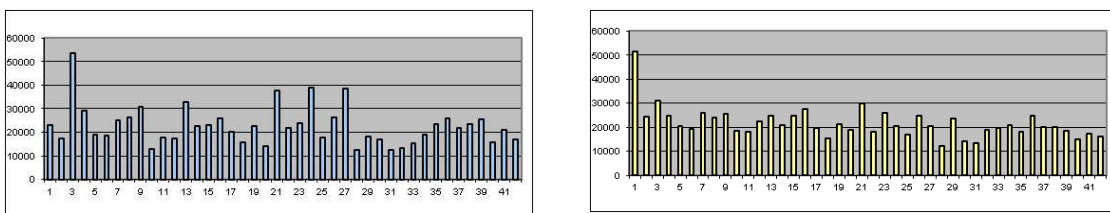


Figura 19. Il·lustració del format full Excel on s’importen les quantificacions de les àrees (blau, mostra tumoral) i groc (mostres no tumorals de control), i representació gràfica de les àrees de cada sonda.

A continuació es pot procedir amb la normalització intra-mostra, que es basa en obtenir una constant de normalització (*Normalisation Constant; NC*) per cada una de les sondes d'una mostra (figura 20). Aquesta constant resulta de dividir, en cada mostra (tumoral i control), l'àrea del pic de cada sonda respecte cadascuna de les àrees de la resta de sondes i finalment obtenir un valor mitjà (NC).

mostra	sonda	tamany	àrea	NC
tumoral	1	112	23127	1.13
tumoral	2	136	17541	0.86
tumoral	3	142	53652	2.62
tumoral	4	148	25029	1.42
tumoral	5	153	18971	0.93
tumoral	6	160	18783	0.92
tumoral	7	166	25235	1.24
tumoral	8	172	26132	1.28
tumoral	9	179	30719	1.60
tumoral	10	185	12716	0.62
tumoral	11	193	17775	0.87
tumoral	12	202	17238	0.84
tumoral	13	211	32794	1.60
tumoral	14	220	22732	1.11
tumoral	15	229	23001	1.13
tumoral	16	238	26633	1.27
tumoral	17	247	20196	0.99
tumoral	18	255	15778	0.77
tumoral	19	265	22719	1.11
tumoral	20	273	13983	0.68
tumoral	21	283	37321	1.94
tumoral	22	292	21793	1.07
tumoral	23	301	23926	1.17
tumoral	24	310	36969	1.91
tumoral	25	319	17632	0.86
tumoral	26	328	26191	1.26
tumoral	27	337	38521	1.89
tumoral	28	346	12639	0.62
tumoral	29	355	18192	0.89
tumoral	30	364	18339	0.82
tumoral	31	373	12506	0.61
tumoral	32	383	13113	0.74
tumoral	33	391	15126	0.84
tumoral	34	400	19046	0.93
tumoral	35	409	23619	1.16
tumoral	36	418	26020	1.27
tumoral	37	427	21951	1.07
tumoral	38	436	23419	1.15
tumoral	39	445	25588	1.25
tumoral	40	454	15600	0.76
tumoral	41	463	21102	1.03
tumoral	42	472	16773	0.82

Figura 20. Il·lustració del format full Excel amb la normalització intra-mostra de la mostra tumoral i l'obtenció del valor NC de cada sonda.

Capítol 1

Després es procedeix a calcular la normalització inter-mostra, que es basa en obtenir una ratio anomenada *dosage quotient* (DQ) per cada sonda que permetrà saber la variació en el número de còpies relativa entre la mostra tumoral i la no tumoral (control) (figura 21). Aquest DQ es calcula dividint la constant de normalització d'una determinada sonda en la mostra tumoral respecte la constant de la mateixa sonda en la mostra no tumoral (control). Finalment amb l'obtenció dels DQ s'estableix l'estat de número de còpies d'un gen definint les següents categories: a) si el DQ per una sonda és inferior a 0,7 considerarem una pèrdua o deleció, b) si el DQ per una sonda està entre 0,7 i 1,4 el considerarem normal o diploide, i c) si el DQ per una sonda és superior a 1,4 ho considerarem com guany, duplicació o amplifacació.

Sonda	NC mostra tumoral	NC mostres control	DQ	Número còpies
1	1,13	2,54	0,45	PÈRDUA
2	0,86	1,20	0,72	NORMAL
3	2,62	1,53	1,71	GUANY
4	1,42	1,22	1,16	NORMAL
5	0,93	1,01	0,92	NORMAL
6	0,92	0,94	0,98	NORMAL
7	1,24	1,28	0,96	NORMAL
8	1,28	1,18	1,09	NORMAL
9	1,50	1,26	1,20	NORMAL
10	0,62	0,92	0,68	PÈRDUA
11	0,87	0,89	0,98	NORMAL
12	0,84	1,11	0,76	NORMAL
13	1,60	1,23	1,31	NORMAL
14	1,11	1,03	1,08	NORMAL
15	1,13	1,23	0,92	NORMAL
16	1,27	1,35	0,94	NORMAL
17	0,99	0,98	1,01	NORMAL
18	0,77	0,76	1,01	NORMAL
19	1,11	1,05	1,06	NORMAL
20	0,68	0,94	0,73	NORMAL
21	1,84	1,47	1,25	NORMAL
22	1,07	0,89	1,19	NORMAL
23	1,17	1,27	0,92	NORMAL
24	1,91	1,00	1,90	GUANY
25	0,86	0,83	1,04	NORMAL
26	1,28	1,22	1,05	NORMAL
27	1,89	1,01	1,87	GUANY
28	0,62	0,60	1,02	NORMAL
29	0,89	1,17	0,76	NORMAL
30	0,82	0,70	1,18	NORMAL
31	0,61	0,65	0,94	NORMAL
32	0,64	0,93	0,69	PÈRDUA
33	0,74	0,96	0,77	NORMAL
34	0,93	1,03	0,91	NORMAL
35	1,16	0,89	1,31	NORMAL
36	1,27	1,22	1,05	NORMAL
37	1,07	0,98	1,09	NORMAL
38	1,15	1,00	1,15	NORMAL
39	1,25	0,92	1,37	NORMAL
40	0,76	0,73	1,04	NORMAL
41	1,03	0,86	1,20	NORMAL
42	0,82	0,79	1,04	NORMAL

Figura 21. Il·lustració del format full Excel amb la normalització inter-mostra i obtenció del valor DQ i l'equivalència en el número de còpies.

3 EL CÀNCER D'ENDOMETRI

3.1 L'ENDOMETRI

La primera referència al càncer uterí, de la qual es té constància, prové d'un metge bizantí d'Alexandria anomenat Paul d'Aegia (625-690). Aquest metge va escriure un compendi de llibres de medicina on va fer descripcions del càncer en general, i en particular del càncer d'úter i de mama.

El càncer d'endometri (CE) és un tipus de càncer que es desenvolupa a l'úter. L'úter és l'òrgan de l'aparell reproductor femení encarregat de la gestació, i com a tal les seves funcions principals són acollir el fetus, mantenir-lo i protegir-lo durant tot l'embaràs, i expulsar-lo en el moment del part. És un òrgan de localització intra-pèlvica, amb una cavitat buida a l'interior i una paret que l'envolta que està organitzada en tres capes, el perimetri, el miometri, i l'endometri (figura 22). En aquesta última capa, l'endometri, és on s'inicia el CE. Cal esmentar, que a l'úter també es desenvolupen altres càncers com els sarcomes uterins els qual s'inicien en l'estroma i el teixit muscular del miometri, però són molt menys comuns que el CE.^{44,45}

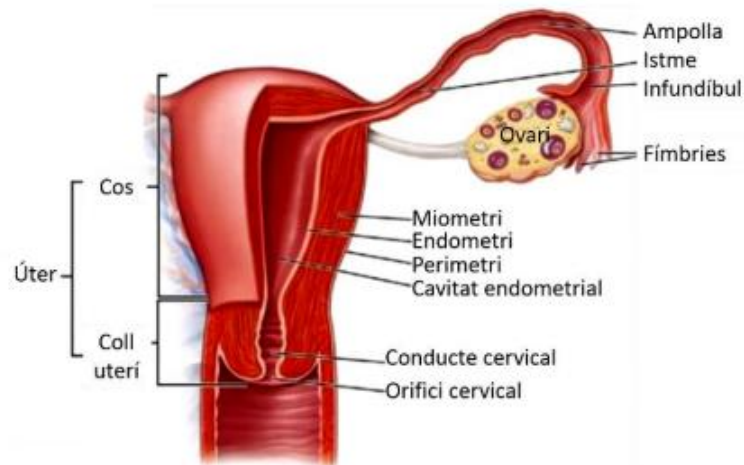


Figura 22. Esquema de l'úter.⁴⁴

Capítol 1

L'endometri és un teixit molt dinàmic que sofreix tota una sèrie de canvis en resposta a l'alliberament d'hormones, estrògens i progesterona, durant el cicle ovàric. La seva principal funció és servir de receptacle per a la implantació de l'òvul fecundat i posterior desenvolupament de l'embrió. Histològicament, l'endometri està constituït per glàndules de forma tubular revestides per un epiteli columnar de cèl·lules estretes i allargades, envoltades per un estroma vascularitzat (figura 23). En funció del moment del cicle, les glàndules presenten morfologies diferents passant de glàndules simples i lleugerament ramificades, a glàndules amb una complexitat remarcable. L'estroma està format per vasos sanguinis, fibres nervioses i cèl·lules immunitàries, i també presenta morfologies diferents en funció del moment del cicle ovàric. Funcionalment, l'endometri es subdivideix en una capa basal prima, adjacent al miometri, que no es descama durant el cicle menstrual; i per una funcional molt dinàmica a sobre que respon als canvis hormonals produïts durant el cicle ovàric. La capa funcional, a la vegada, es subdivideix en dues capes més, una amb poques glàndules i estroma abundant, anomenada compacta, i per una altra més profunda que presenta moltes glàndules i menys estroma, anomenada esponjosa.⁴⁶

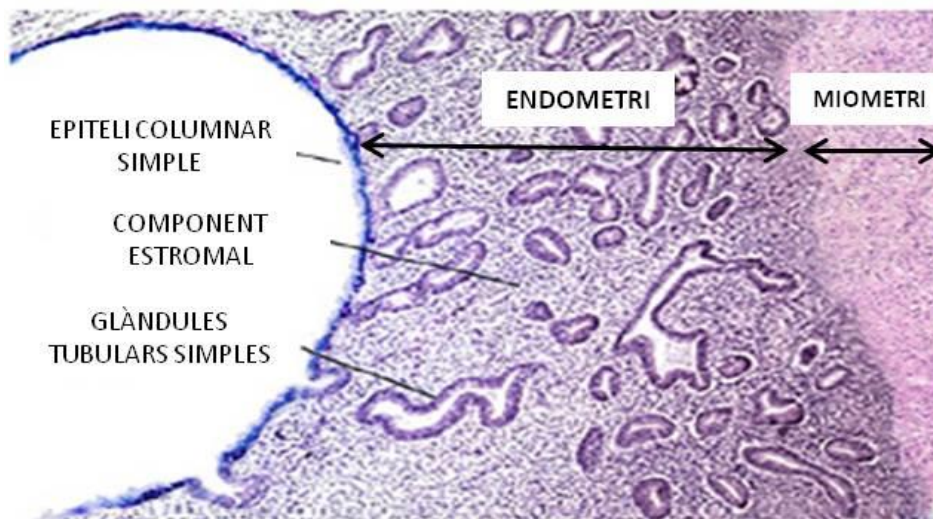


Figura 23. Imatge de la histologia de l'endometri.⁴⁶

D'aquesta manera podem descriure diferents tipus d'endometri com a conseqüència de les diferents morfologies que adopta aquest teixit en funció de les variacions hormonals produïdes durant el cicle ovàric (figura 24).^{44,45} Així tenim els següents tipus:

- l'endometri en fase descamativa, que es dona als inicis del cicle menstrual (dia 1 a 4 del cicle) on hi ha una reducció dels nivells de progesterona. Es caracteritza per una dissociació estructural, una marcada infiltració leucocitària, hemorràgia i necrosi en cas de no haver-se realitzat la fecundació.

- l'endometri proliferatiu, que es dona en la fase fol·licular (dia 5 a 14 del cicle) on hi ha un augment progressiu dels estrògens arribant al seu pic el dia 14 quan es produeix l'ovulació. Es caracteritza per glàndules endometrials rectilínies delimitades per cèl·lules que presenten una marcada pseudoestratificació nuclear i figures de mitosi, i l'estroma dens amb cèl·lules amb citoplasma escàs.

- l'endometri secretor, que es dona en la fase lútea (dia 15 al 28 del cicle) on hi ha un augment de progesterona. Es caracteritza per glàndules endometrials molt serrades amb secreció luminal i estroma amb transformació predecidual, aparició de vacuoles al pol basal de les cèl·lules que contenen glucògen i mucopolisacàrids que seran secretats, i aparició a l'estroma de les arterioles espirals i augment del citoplasma de les seves cèl·lules adoptant una forma cúbica.

Capítol 1

A més a més també es poden considerar dos tipus addicionals fora de la influència del cicle menstrual:

-l'endometri gestacional, que es caracteritza per la hipertrofia i el quadre secretor més activat tant del teixit glandular com de l'estroma, per unes glàndules constituïdes per un epiteli baix i ric en glucògen, i unes cèl·lules de l'estroma més grans i poligonals, amb un ampli citoplasma típic de les cèl·lules decidues disposades en forma de mosaic.

- l'endometri atròfic, que es dona en l'etapa del climateri (pas de l'etapa reproductiva a no reproductiva) i està caracteritzat per la pèrdua de la capacitat proliferativa i per tant disminució de l'activitat mitòtica de les cèl·lules. Les glàndules perden la morfologia ramificada i s'organitzen de forma esfèrica, nuclis cel·lulars reduïts i inactius, i l'estroma presenta una estructura densa i fibrosa amb cèl·lules fines i allargades i una coloració blanca característica.

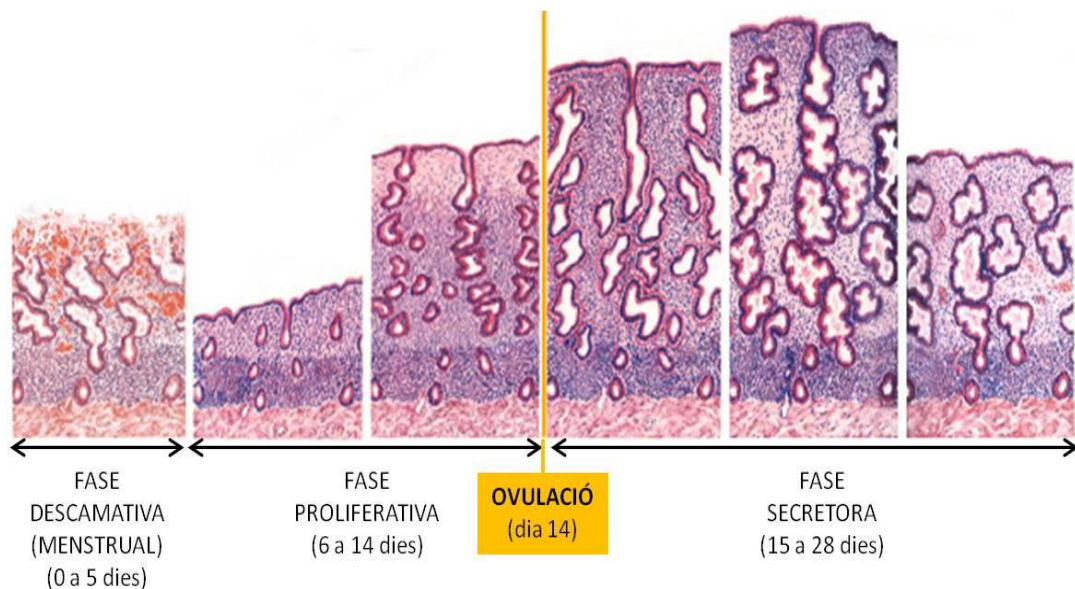


Figura 24. Figura de les diferents fases i morfologies que adopta l'endometri durant el cicle ovàric.⁴⁵

3.2 EPIDEMIOLOGIA I FACTORS DE RISC

A nivell mundial, el 2012, el càncer d'endometri va ocupar la sisena posició pel que fa a la seva incidència (320000 casos nous) i la catorzena pel que fa a la seva mortalitat (76000 casos).⁴⁷ A nivell espanyol, el 2017, va tenir una incidència de 5473 casos nous, ocupant així la tercera posició, darrera del càncer de mama i de colon (figura 25). I pel que fa a la mortalitat va suposar per a l'any 2016 unes 1615 defuncions.² De manera general les morts per càncer d'endometri representen al voltant del 2-3% del total de morts produïdes pel càncer, i dins les pacients diagnosticades amb CE té una taxa de mortalitat al voltant del 20%.

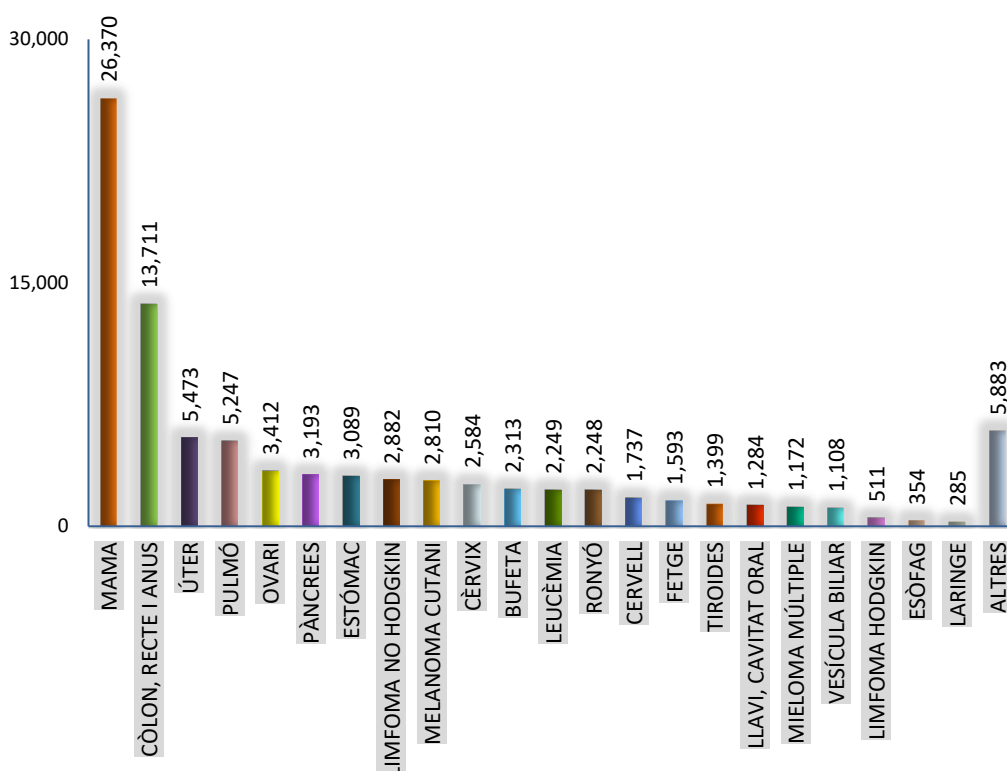


Figura 25. Incidència dels diferents tipus de tumor que afecten a les dones a Espanya per l'any 2017.²

Les hormones juguen un paper molt important en l'etiologia d'aquest càncer. Així, els estrògens són contribuïdors claus de cadascun dels factors de risc establerts per aquest càncer, com són la menarquia precoç, la menopausa tardana, la nuliparitat, l'ús d'hormones per la menopausa, l'obesitat, la diabetis, la hipertensió i tenir una història familiar de càncer d'endometri. Contràriament, la multiparitat, l'ús d'anticonceptius orals combinats d'estrògen-progesterona i l'edat avançada del primer part s'ha vist que són factors que redueixen el seu risc.⁴⁷

Capítol 1

En un estudi que va recollir dades de 43 països de 5 continents diferents en el període comprès entre l'any 1978 i el 2013, es va observar una tendència a l'augment de la incidència del càncer d'endometri en la meitat de les poblacions estudiades, especialment en països que havien sofert una ràpida transició socioeconòmica (figura 26). Algunes de les explicacions a aquesta tendència generalitzada serien: l'augment de l'ús d'estrògens exògens en dones peri i post menopàusiques; l'augment de l'exposició a estrògens endògens (nulipariat); canvis en els factors que influeixen la reproducció com la disminució de les taxes de fertilitat i increment de la nuliparitat, que fan que en molts països la mitjana de fills estigui per sota de 2; un major nombre de casos d'obesitat; l'augment de la diabetis i increment de la conscienciació de les revisions ginecològiques i les millores de la sanitat en general. No obstant això, en alguns països d'Europa no s'ha observat aquesta tendència i s'hipotitza que podria ser degut a l'augment de l'ús d'anticonceptius orals en aquests països.⁴⁷

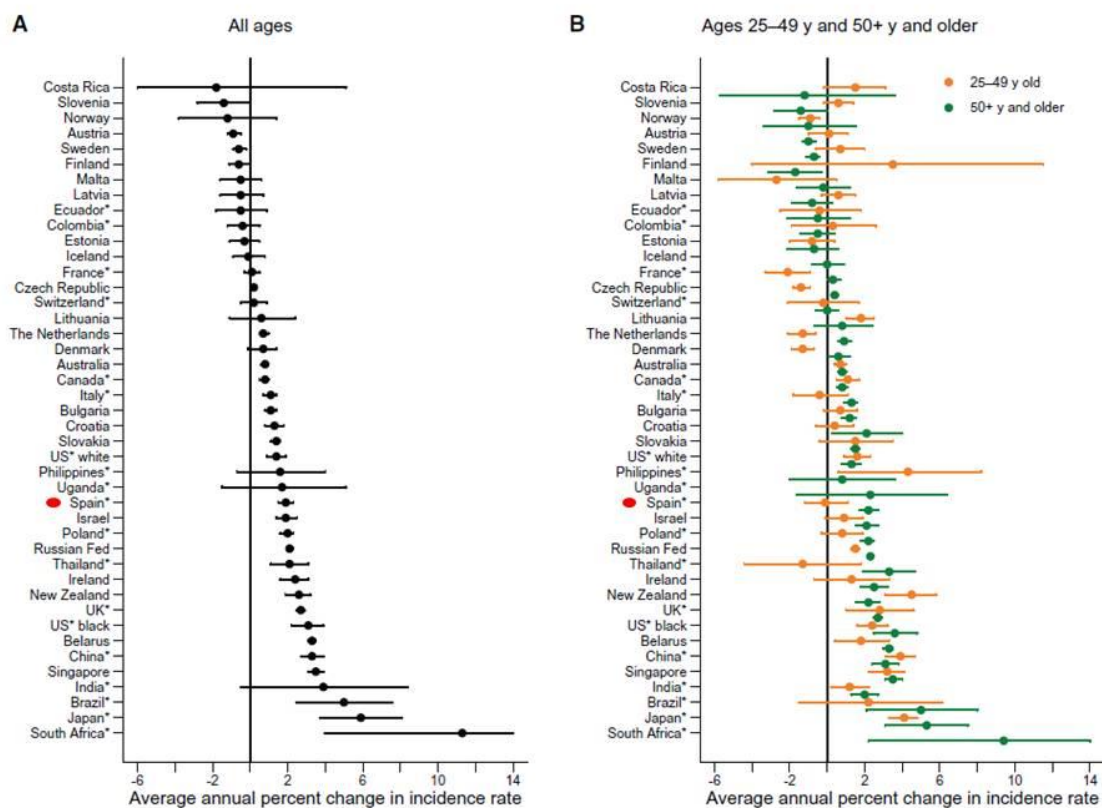


Figura 26. Canvi percentual mitjà anual en les taxes d'incidència del càncer d'endometri estandarditzat per edat, (A) totes les edats i (B) de 25 a 49 anys, i a partir de 50 anys; en els darrers 10 anys disponibles segons estudi de Lortet-Tieulent et al.⁴⁷

3.3 PRESENTACIÓ CLÍNICA, EINES DIAGNÒSTIQUES I CARACTERÍSTIQUES MACROSCÒPIQUES

La forma de presentació més freqüent (90%) del càncer d'endometri és la presència d'un sagnat uterí anormal, sobretot post menopàusic ja que la majoria d'aquest càncer es diagnostica en pacients post menopàusiques. Tot i això, cal tenir en compte altres factors associats a aquests sagnats que no estiguin relacionats amb les neoplàsies malignes. El dolor pèlvic i la distensió abdominal també poden ser altres formes de presentació, sobretot en casos de càncer d'endometri en estadi més avançat. També cal esmentar que un alt percentatge de les pacients amb càncer d'endometri presenten obesitat i cicles anovulatoris.⁴⁸

Les tècniques majoritàriament utilitzades per al diagnòstic del CE són l'anàlisi histopatològica de la biòpsia endometrial i les tècniques de diagnòstic per la imatge, com l'ús d'ultrasons transvaginals, la tomografia computeritzada i la ressonància magnètica nuclear. Per fer la presa de la biòpsia endometrial s'utilitza la tècnica de la histeroscòpia, que permet la valoració de la cavitat endometrial i del canal endocervical a temps real. Les tècniques d'imatge són àmpliament utilitzades per poder excloure el cas de malaltia metastàtica, i per avaluar l'estat d'afectació ganglionar.

A nivell macroscòpic, l'aspecte del teixit tumoral sol ser pàl·lid i carnós, i ocasionalment pot presentar una estructura sorrosa i sòlida. També és freqüent trobar focus hemorràgics i necrosi. La presentació d'aquest tumor pot ser local o difusa, afectant àrees diferents de l'endometri o la seva totalitat. Aquest carcinoma a mesura que es desenvolupa pot anar envaint els diferents teixits adjacents com el miometri, la serosa uterina, el cervix, les trompes, i com més avançat està pot arribar a estendre's a vagina, ovaris i cavitat peritoneal, i també disseminar-se a través dels vasos limfàtics i vasculars. Conseqüentment, pot afectar ganglis limfàtics regionals de la pelvis, i també els més distants com els paraòrtics, a més en els casos més agressius pot afectar altres òrgans com els pulmons.

3.4 CLASSIFICACIONS

3.4.1 HISTOPATOLÒGICA

Histològicament el càncer d'endometri és molt heterogeni i per això es classifica en diferents tipus com carcinoma endometriode, serós, cèl·lula clara, mucinós, tumor neuroendocrí, carcinoma indiferenciat, carcinoma mixte i carcinoma desdiferenciat entre d'altres.⁴⁹ De tots aquests, els tipus més freqüents són el carcinoma endometriode d'endometri (CEE) i el carcinoma serós d'endometri (CSE) representant un 70-80% i 10-20%, respectivament (figura 27).

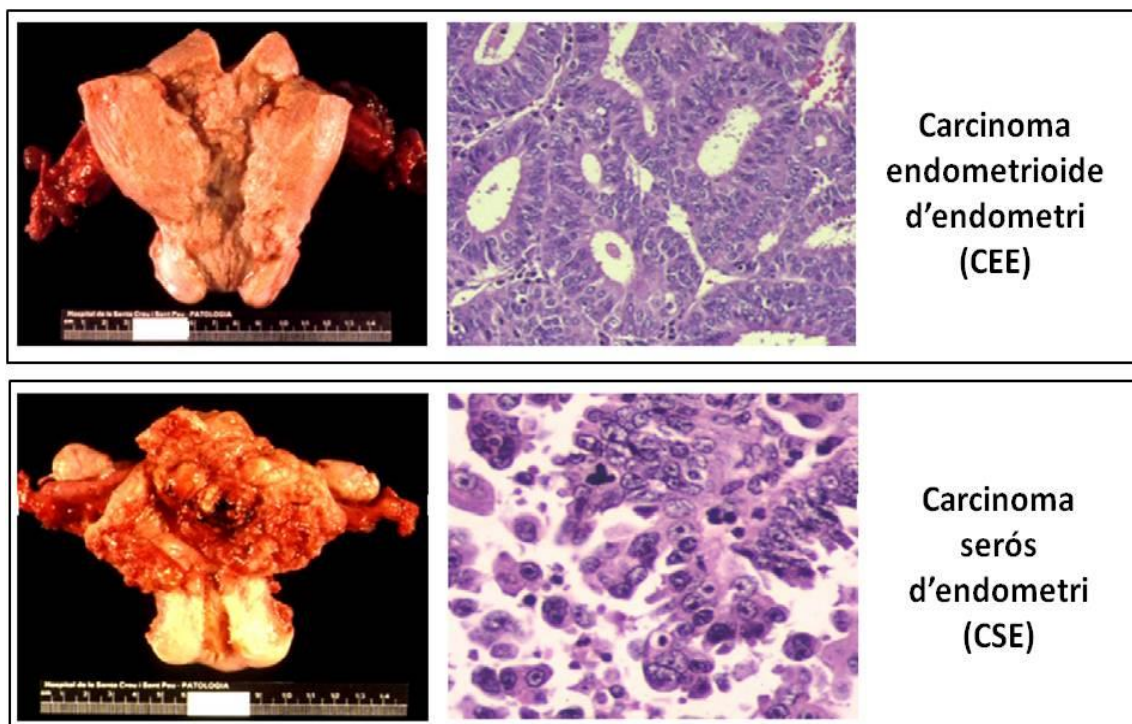


Figura 27. Representació a nivell macroscòpic i histològic d'un carcinoma endometriode d'endometri i d'un carcinoma serós d'endometri.

El CEE presenta estructures glandulars o villoglandulars, i se li reconeixen diferents variants com diferenciació escamosa, villoglandular, secretora i de cèl·lules ciliades. Una característica molt important del CEE és que la seva histologia pot presentar diferents nivells de diferenciació glandular, que poden anar des d'un carcinoma ben diferenciat on les estructures glandulars siguin molt semblants a les de l'endometri normal, fins a un carcinoma mínimament diferenciat que es pot confondre amb un carcinoma indiferenciat i també amb algun sarcoma.

D'aquesta manera el CEE es classifica en funció del grau histològic (FIGO (*International Federation of Gynecology and Obstetrics*) grades) en tres tipus: el grau 1 (ben diferenciat), grau 2 (moderadament diferenciat) i grau 3 (pobrament diferenciat) caracteritzats per un grau de creixement sòlid no glandular, inferior al 5%, comprés entre el 6 i 50% o superior al 50%, respectivament.⁵⁰ Així, el grau 3 es considerat el de més alt risc. Cal esmentar que l'observació d'una atípia notable dels nuclis, com el pleomorfisme o nuclèols prominents, inapropiada per al grau histològic podrà augmentar en una unitat els grau histològic pertinent.⁵¹ En els últims anys, alguns grups han proposat el canvi del sistema de tres graus a l'establiment d'un sistema de grau binari (*binary* FIGO) amb les categories de baix risc (grau 1, 2) i l'alt risc (grau 3).^{52,53} Contràriament, el CSE es caracteritza per un patró de papil·les complex, amb cèl·lules i nuclis generalment ovalats i sense presentar una orientació perpendicular a la membrana basal. Els nuclis estan poc diferenciats, en situació apical i amb macro nuclèols eosinofílics, i les mitosis, les cèl·lules multinucleades i els cossos de psammoma hi són bastant presents. Aquest carcinoma per definició ja és considerat de grau 3, i per tant, d'alt risc.⁵⁴

3.4.2 TRADICIONAL

La classificació tradicional del CE, també anomenada dualística, prové de la proposta que va fer Bokhman a l'any 1983⁵⁵, basada en classificar aquest càncer en dues categories clínicopatològiques, la tipus I i la tipus II, tenint en compte les característiques histològiques, l'expressió dels receptors hormonals i el grau (taula 3).

Segons aquesta classificació el CE tipus I es caracteritza per ser de baix grau (grau 1 i 2), de tipus histològic endometriode (CEE), diploide a nivell de dotació gènica, presentar expressió d'estrògens, ser de bon pronòstic, i per provenir d'una hiperplàsia complexa atípica. Contràriament, el CE tipus II es caracteritza per ser d'histologia endometriode grau 3 o no endometriode, principalment pel carcinoma serós (CSE), aneuploide pel que fa a la dotació gènica, per presentar mutacions del gen *TP53*, absència d'expressió d'estrògens, per desenvolupar-se a partir de l'endometri atrofíic i per estar associat amb un risc més alt de metastasi i pitjor pronòstic.⁵¹ Els estudis moleculars també van recolzar aquesta classificació dicotòmica. El tipus I està associat amb mutacions en gens implicats en la via PI3K-AKT com *PTEN*, *KRAS*, *PIK3CA*, mutacions en gens implicats en la via WNT com *CTNNB1* i mutacions al gen *ARID1A*. A més a més, aquest tipus freqüentment també presenta inestabilitat de microsatèl·lits majoritàriament com a conseqüència de la hipermetilació del promotor del gen *MLH1*. Contràriament, el tipus II presenta mutacions recurrents al gen *TP53* i amplificacions d'*ERBB2*.

Taula 3. Característiques diferencials entre els carcinomes tipus I i tipus II. ⁵⁶

	Tipus I	Tipus II
Característiques clíniques, hormonals i morfològiques		
Freqüència	60-70%	30-40%
Funció reproductiva	Reduïda	No afectada
Inici menopausa	Després dels 50 anys	Abans dels 50 anys
Lesió precursora	Hiperplàsia	Atròfia
Associat a estrògens	Sí	No
Associació amb obesitat, diabetes mellitus	Sí	No
Grau	Baix (grau 1-2)	Alt (grau 3)
Invasió miometri	Superficial	Profunda
Pronòstic	Favorable	Desfavorable
Supervivència lliure de malaltia (5 anys)	86%	59%
Característiques clínicopatològiques i moleculars		
Tipus histològic	Endometrioides	Serós
Expressió de receptors d'estrògens o progesterona	Alt	Baix
Estadi al diagnòstic	Inicials (FIGO I-II)	Avançats (FIGO III-IV)
Alteracions moleculars		
<i>PTEN</i>	52-78%	1-11%
<i>PIK3CA</i>	36-52%	24-42%
<i>PIK3R1</i>	21-43%	0-12%
<i>KRAS</i>	15-43%	2-8%
<i>ARID1A</i>	25-48%	6-11%
<i>CTNNB1</i>	23-24%	0-3%
<i>TP53</i>	9-12%	60-91%
<i>PPP2R1A</i>	5-7%	15-43%
<i>ERBB2</i> (amplificació)	0	27-44%
Inestabilitat de microsatèl·lits	28-40%	0-2%

Actualment les evidències clíniques, epidemiològiques i moleculars suggereixen que aquesta classificació dualística basada en la composició de dos entitats homogènies no reflecteix la veritable naturalesa del CE. L'entitat dels carcinomes d'endometri d'alt grau (CEAG), que inclouen tant carcinomes amb grau FIGO 3 com carcinomes endometrioides, serosos, cèl·lula clara, indiferenciats/desdiferenciats, i carcinosarcomes, són el reflex de la naturalesa complexa del CE. Aquests carcinomes són un subgrup amb característiques biològiques, genètiques, morfològiques i clíniques molt heterogènies. Alguns d'aquests casos (30%) no presenten característiques histològiques típiques i els resultats immunohistoquímics tampoc són conclouents, d'aquesta manera és difícil classificar-los adientment. Inclús s'ha observat que la reproductibilitat interobservador entre els patòlegs especialistes és també subòptima.⁵⁷ Per tant el gran repte recau en poder discernir si un determinat CE d'alt grau es comportarà com un dels subtipus més agressius com són el carcinoma serós o de cèl·lula clara.⁵⁸

3.4.3 MOLECULAR-TCGA

L'any 2013 es va fer un gran avenç en la classificació del CE com a resultat d'un gran estudi molecular integrat portat a terme pel TCGA.⁵⁹ Segons les dades d'aquest estudi el carcinoma d'endometri es pot classificar en quatre grups moleculars (taula 4):

1) *POLE-ultramutat* (~8%), on trobem CE amb unes càrregues mutacionals molt elevades (*ultramutated*) (232×10^{-6} mutacions/Mb), amb mutacions que afecten al domini exonucleasa del gen *POLE* (100%), amb microsatèl·lits estables (MSS), i on s'inclouen carcinomes predominantment d'histologia endometrioides (CEE) d'alt grau però amb molt bon pronòstic, i que solen presentar mutacions en altres gens com: *PTEN* (94%), *PIK3R1* (65%), *PIK3CA* (71%), *FBXW7* (82%), *KRAS* (53%).

2) *MSI-H* (~30%), amb inestabilitat de microsatèl·lits (MSI), majoritàriament CEE, i també amb una elevada taxa de mutacions (*hypermuted*) (18×10^{-6} mutacions/Mb), i que presenten mutacions en altres gens com: *PTEN* (88%), *PIK3CA* (55%), *KRAS* (38%), *CTNNB1* (19%) i *FGFR2* (16%).

3) *CN low o MSS* (~65%), amb CE majoritàriament CEE, amb un baix número de gens amb alteració en el número de còpies (*copy-number low*), amb microsatèl·lits estables (MSS), i amb pronòstic intermig i que presenten mutacions en gens com: *PTEN* (77%), *PIK3CA* (53%), *CTNNB1* (53%), *PIK3R1* i *KRAS* (15%), *FGFR2* (11%) i *SOX17* (8%).

4) CN high (copy-number high) (~26%), amb CE amb alt número de gens amb alteració en el número de còpies, sobretot en gens com *ERBB2* i *CCNE1*, mutacions al gen *TP53* i caracteritzat per tumors d'alt grau que majoritàriament són carcinomes serosos i sobre un 35% de carcinomes d'alt grau amb histologia endometriode, tots ells amb molt mal pronòstic. Cal remarcar que, tot i que hi ha gens alterats exclusius de grups moleculars com *POLE* per al grup 1 i *TP53* per al grup 4, les alteracions en gens de la via PI3K/AKT/mTOR i RTK/KRAS/CTNNB1 són comuns a múltiples grups com el 1, 2 i 3.

Taula 4. Descripció de les alteracions moleculars en cada grup del TCGA. ⁵⁶

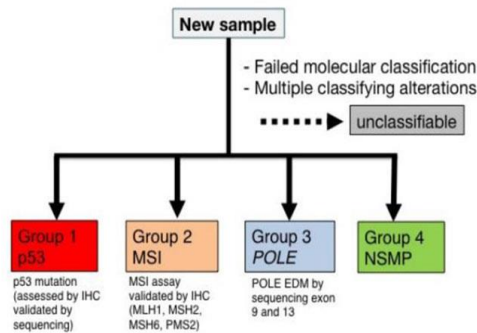
	GRUPS MOLECULARS TCGA			
	<i>POLE</i> (ultramutated)	MSI-H (hypermutated)	CN-Low (copy-number low)	CN-High (copy-number high)
ASNC	baix	baix	baix	alt (<i>ERBB2</i> , <i>CCNE1</i>)
Estat dels microsatèl·lits	estables i inestables	inestables	estables	estables
Taxa de mutacions	molt elevada ($232 \cdot 10^{-6}$ mutacions/Mb)	elevada ($18 \cdot 10^{-6}$ mutacions/Mb)	baixa ($2,9 \cdot 10^{-6}$ mutacions/Mb)	baixa ($2,3 \cdot 10^{-6}$ mutacions/Mb)
Gens freqüentment alterats	<i>POLE</i> (100%)	<i>PTEN</i> (88%)	<i>PTEN</i> (77%)	<i>TP53</i> (92%)
	<i>PTEN</i> (94%)	<i>RPL22</i> (37%)	<i>CTNNB1</i> (52%)	<i>PP2R1A</i> (22%)
	<i>PIK3CA</i> (71%)	<i>KRAS</i> (35%)	<i>PIK3CA</i> (53%)	<i>PIK3CA</i> (47%)
	<i>PIK3R1</i> (65%)	<i>PIK3CA</i> (54%)	<i>PIK3R1</i> (33%)	
	<i>FBXW7</i> (82%)	<i>PIK3R1</i> (40%)	<i>ARID1A</i> (42%)	
	<i>ARID1A</i> (76%)	<i>ARID1A</i> (37%)		
	<i>KRAS</i> (53%)			
Tipus histologic	Endometriode	Endometriode	Endometriode	Endometriode i Serós
Grau	graus 1 a 3	graus 1 a 3	graus 1 i 2	grau 3
Supervivència lliure de malaltia	alta	intermitja	intermitja	baixa

3.4.4 MOLECULAR POST-TCGA

En els últims 4 anys hi ha hagut un gran interès en poder portar a la pràctica clínica diària l'aplicació de la classificació molecular del TCGA sobretot per al diagnòstic dels CEAG que són histològicament ambigus i difícils de classificar.^{60,53} En aquest sentit, diferents grups han proposat l'ús d'un sistema de classificació molecular alternatiu (*surrogate*) al del TCGA amb una aproximació metodològica més senzilla per tal de poder aplicar-la en els laboratoris de patologia molecular en la pràctica clínica diària.^{61,62} De manera generalitzada aquest sistema es basen en la combinació de diferents tècniques com la IHQ (immunohistoquímica) per l'anàlisi de l'expressió de les proteïnes p53, MSH2, MSH6, MLH1 i PMS2; l'anàlisi de microsatèl·lits mitjançant l'anàlisi de fragments per determinar la inestabilitat de microsatèl·lits; i la seqüenciació Sanger per l'anàlisi de mutacions del gen *POLE*. Les dades d'aquests estudis han revelat que aquest sistema proveeix informació pronòstica independent més enllà de la de les variables clínico-patològiques tradicionals, que té una capacitat discriminatòria del risc de les pacients comparable al sistema d'estratificació del risc ESMO (*European Society for Medical Oncology*), i que pot ser utilitzat en la pràctica clínica.⁶³ A més a més, també s'ha posat de manifest que els CEAG són una col·lecció d'almenys els 4 subtipus moleculars representats pel TCGA.⁶⁴

En base a aquesta metodologia, el grup d'Stello (Leiden/TransPORTEC)⁶¹ va classificar el CE en els següents grups: 1) p53, amb casos *TP53* mutat, 2) MSI, amb casos amb inestabilitat de microsatèl·lits, 3) *POLE* mutat, i 4) NSMP (*No Specific Molecular Profile*) amb casos amb un perfil molecular no específic que no corresponien a cap dels altres 3 grups. I el grup de Talhouk (ProMisE/Vancouver)⁶² els va classificar com: 1) MMR-D, amb casos amb inestabilitat de microsatèl·lits, 2) *POLE* EDM (*POLE Exonuclease Domain Mutations*), 3) p53 wt amb casos on la proteïna p53 tenia funció normal, i 4) p53 abn amb casos on hi havia una expressió aberrant de *TP53*. A la figura 28 es mostra l'algoritme a seguir en funció d'aquestes dos classificacions.

a Leiden/TransPORTEC molecular classification



b ProMisE /Vancouver group molecular classification

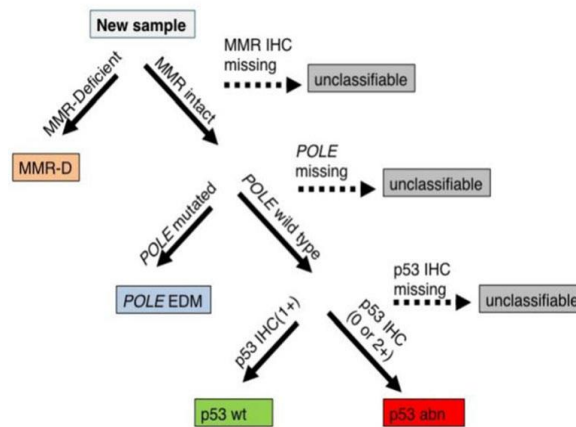


Figura 28. Algoritmes utilitzats per als sistemes de classificació molecular del grup Leiden/TransPORTEC (a) i ProMisE/Vancouver (b).

En vistes de la gran heterogeneïtat molecular i morfològica del CE, és evident que és necessari establir sistemes de classificació del CE més refinats, i que encara es necessiten més estudis per esclarir les bases moleculars i biològiques de certs tipus de CE com els casos ambigus que presenten solapament d'histologies, els casos amb perfils moleculars no específics, i els casos que pertanyen a més d'un grup molecular (*multiple classifiers*). Aquests últims casos, un 2-3% segons Stello et al i Talhouk et al ^{61,62}, poden presentar a la vegada mutacions a *POLE* i MMR-D, mutacions a *TP53* i MMR-D, mutacions a *POLE* i *TP53*, i mutacions a *TP53* i MMR-D. Respecte a això, s'ha descrit que les mutacions del gen *TP53* poden ser un efecte secundari del fenotip mutat dels casos *POLE* i MMR-D. D'acord amb això seria recomanable, a l'hora de fer la classificació molecular del CE, primer determinar l'estat mutacional del gen *POLE*, ja que a més a més, s'ha vist que aquests pacients presenten pronòstics molt favorables. ⁶³

En la següent taula (taula 5) es mostra la comparativa dels quatre tipus de classificacions del CE.

Taula 5. Comparació dels quatre sistemes de classificació del càncer d'endometri.⁵⁶

	TRADICIONAL (BOKHMAN)	HISTOLÒGICA (WHO)	MOLECULAR TCGA	MOLECULAR POST-TCGA
BASE	característiques clínicas i histopatològiques	característiques histològiques	caracterització genòmica	IHQ (p53, MSH2, MSH6, MLH1, PMS2) i seq. Sanger (<i>POLE</i>)
CATEGORIES	Tipus I	Endometriode	<i>POLE</i> , MSI, CN-L, CN-H	<i>POLE</i> EDM, MMR-D, p53 wt/NSMP, p53 abn
	Tipus II	Serós	CN-H	p53 abn

3.5 ESTADIS I ESTRATIFICACIÓ DEL RISC

La determinació de l'estadi tumoral (estadificació) es realitza tenint en compte el grau d'extensió d'un tumor que s'ha iniciat en una localització concreta (tumor primari) i el grau de metàstasi. Un dels principals objectius de l'estadificació és poder definir grups de pacients amb pronòstics similars per tal de poder estandaritzar el seu maneig i donar les opcions terapèutiques més adequades per cadascun d'ells. En l'actualitat, la classificació dels estadis més utilitzada en el càncer d'endometri és la que combina la classificació FIGO (*International Federation of Gynecology and Obstetrics*) del 2009 i el sistema TNM (*Tumor Node Metastasis*).⁵⁰ Aquestes classificacions es basen en l'anàlisi de la peça quirúrgica amb el càlcul de la mida del tumor, del grau d'invasió del miometri, del grau d'afectació ganglionar i del grau de metàstasi local i distant. D'aquesta manera el CE es classifica en un total d'onze estadis (taula 6).

Taula 6. Estadiatge FIGO/TNM 2009 i supervivència als 5 anys.

ESTADIS FIGO	T	N	M	% SUPERVIVÈNCIA ALS 5 ANYS	DESCRIPCIÓ
I					
IA	T1a	N0	M0	96	Tumor limitat a l'úter, invasió miometrial absent o inferior al 50%
IB	T1b	N0	M0	87	Tumor limitat a l'úter, invasió miometrial igual o superior al 50%
II					
	T2	N0	M0	80	Invasió de l'estroma cervical, però no més enllà de l'úter
III					
IIIA	T3a	N0	M0	48	El tumor invaeix serosa o annexes
IIIB	T3b	N0	M0	53	Implicació de la vagina o parametris
IIIC1	T1-T3	N1	M0	60	Ganglis pèlvics implicats
IIIC2	T1-T3	N1	M0	53	Implicació para-aòrtica
IV					
IVA	T4	Qualsevol N	M0	57	El tumor envaeix la bufeta i/o la mucosa intestinal
IVB	Qualsevol T	Qualsevol N	M1	16	Metàstasi a distància que inclou metàstasi abdominal i/o ganglis limfàtics inguinals

De manera global, la gran majoria dels CE (75%) són diagnosticats en estadis primerencs com estadi I o II, amb una supervivència lliure de malaltia als 5 anys que oscil·la entre el 80% i 96%; per a l'estadi III oscil·la entre el 48-60%; i per a l'estadi IV entre el 16-57%. La supervivència lliure de malaltia als 5 anys és del 90% en pacients sense metàstasis ganglionars, del 60-70% en pacients amb metàstasis en ganglis pèlvics, i 30-40% en pacients amb metàstasis als ganglis para-aòrtics.⁴⁸ Cal dir que en aquest sistema de classificació no es tenen en compte altres factors pronòstics que s'ha vist que també són importants per definir la supervivència global i el risc de recurrència de les pacients.⁶⁵ Així, en els últims anys, s'han desenvolupat sistemes d'avaluació del risc de recurrència en el CE que han incorporat aquests factors pronòstics per tal de definir els diferents grups de risc. Així el grup ESMO, al 2014, va identificar que el tipus i grau histològic, una invasió del miometri superior al 50%, la invasió de l'espai linfovacular, les metàstasis ganglionars i un diàmetre de tumor superior a 2 cm eren factors que determinaven un alt risc de recurrència en pacients en estadis aparentment inicials (estadi I). Per exemple, el grup d'alt risc era definit per pacients amb estadis IB, grau 3 amb histologia endometriode i per pacients amb histologia no endometriode de qualsevol grau histològic.⁶⁶

A més a més, altres grups, també han suggerit la incorporació de les característiques moleculars per poder definir millor el pronòstic i les teràpies per a les pacients. D'aquesta manera, els quatre grups del TCGA es classifiquen segons el risc en: el grup de baix risc on s'inclou el grup *POLE-ultramutated*, el grup de risc intermig amb el grup MSI-H i CN-low, i el grup d'alt risc amb el grup *CN-high/serous-like*.⁵⁶ Està clar que hi ha d'haver una integració de les classificacions histopatològiques i moleculars, ja que moltes pacients tenen la probabilitat de poder ser tractades només amb cirurgia, i conseqüentment el sobretractament d'aquestes pacients pot ser evitat. Així, la identificació de marcadors pronòstics i predictius fiables, és per tant, de gran importància clínica.

En resum, podem dir que cada cop hi ha més evidència de la gran complexitat i heterogeneïtat del càncer d'endometri, i de la importància d'aplicar a la pràctica clínica un sistema de classificació integrat global, on es combinin diferents factors com els anatòmics, els histopatològics, i els moleculars, amb l'objectiu de poder fer una òptima estratificació de les pacients per poder definir uns òptims grups de risc, i per poder donar així el millor maneig quirúrgic i terapèutic.

4 ALTERACIONS MOLECULARS DEL CÀNCER D'ENDOMETRI

En els últims anys, la recerca en l'àmbit de la biologia molecular del CE ha sofert un gran avenç gràcies a l'aplicació de la NGS. Aquesta recerca ha posat en evidència l'alta heterogeneïtat que presenta el CE, establint diferents grups moleculars, i ha descrit la múltiple diversitat d'alteracions moleculars que estan implicades en la seva carcinogènesi (taula 4). A continuació es detallen les alteracions moleculars més freqüentment associades al CE, i de les que es té més evidència de la seva implicació clínica, tant a nivell diagnòstic, pronòstic i de predicció de resposta als tractaments.

4.1 ALTERACIONS MOLECULARS IDENTITÀRIES DELS GRUPS MOLECULARS TCGA

4.1.1 POLE

El gen *POLE* (*Polymerase Epsilon*) codifica per una ADN polimerasa que presenta una activitat exonucleasa de correcció d'errors de la replicació de l'ADN, i que conseqüentment, permet que es produeixi una replicació d'alta fidelitat.⁶⁷ Aproximadament, el 8% dels càncers d'endometri, en general, i entre el 15-20% dels CEE FIGO grau 3 presenten mutacions al domini exonucleasa d'aquest gen.⁵⁹ Les mutacions en aquest domini fan que aquesta polimerasa no pugui corregir els errors d'introducció de nucleòtids erronis a mesura que es replica l'ADN (figura 29). S'ha descrit que les conseqüències d'aquesta alteració estan associades a un increment excepcional del número de mutacions en altres posicions del genoma donant un genotip anomenat ultramutat (232×10^{-6} mutacions/Mb), i a un increment de la proporció de transversions C→A.⁶⁸ Aquests tumors presenten mutacions en altres gens com *PTEN*, *PIK3R1*, *PIK3CA*, *FBXW7* i *KRAS*, molt poques alteracions en el número de còpies de gens, i es caracteritzen per presentar estabilitat dels microsatèl·lits (MSS, *microsatellite stable*). En l'estudi de Temko⁶⁹ s'ha documentat que les mutacions somàtiques al gen *POLE* solen ser precoces i que són events iniciadors de la tumorigenèsis endometrial. A la vegada també han documentat que l'adquisició d'aquestes mutacions provoca un patró diferent de mutacions *driver*, una càrrega de mutacions molt elevada, i una millor resposta immune.

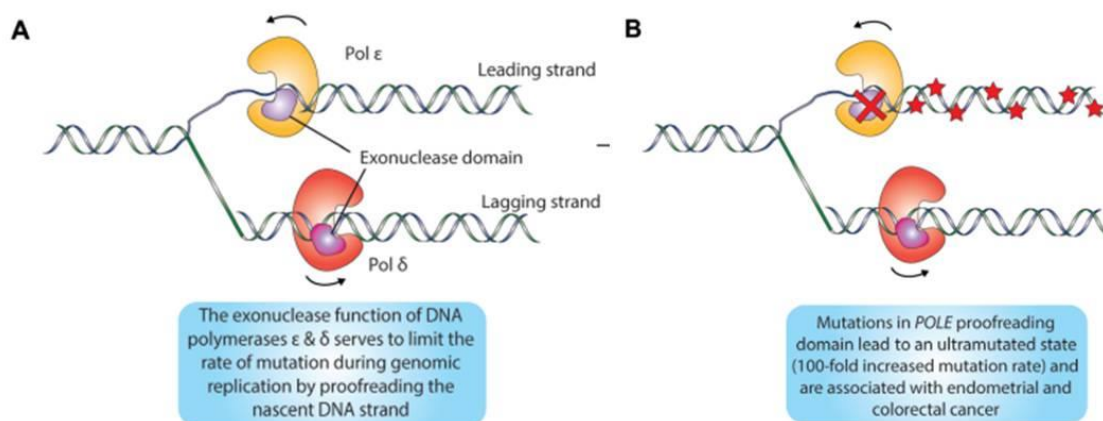


Figura 29. Figura de la funció normal de la polimerasa *POLE* (A), i del resultat de la disfunció de *POLE* al incorporar mutacions al domini exonucleasa (B).⁷⁰

Capítol 1

La gran majoria de les mutacions somàtiques identificades es troben localitzades entre els exons 9, 13 i 14. En el cas de la detecció i valoració de les mutacions, a l'hora de fer el diagnòstic molecular, és molt important tenir en compte només les variants patogèniques. Per això s'ha proposat classificar com a patogèniques aquelles variants que compleixen els requisits següents: que es demostrï que la variant dona com a resultat un genotip ultramutat (més de 100 mutacions per Mb), que suposin un increment de la proporció de transversions C→A (superant el 20% de totes les substitucions) i que la mutació escaigui en un residu que es trobi recurrentment alterat en càncer.⁶⁷

Una gran proporció dels tumors *POLE* mutats també presenten les següents característiques: són tumors amb histologia endometrioide, FIGO estadi I, grau 3 (60%), són rics en limfòcits infiltrants tumorals CD8+ (*TILs*) i/o limfòcits peritumorals (*Crohn's like*), i presenten heterogeneïtat intratumoral amb morfologia ambigua incloent característiques similars a les del CSE.⁷¹ Múltiples estudis han demostrat que les pacients amb tumors *POLE* mutats tenen pronòstics excel·lents, independentment d'altres característiques clinicopatològiques com el grau histològic i els estadis avançats.⁷²⁻⁷⁴ Per aquest motiu és molt important detectar les pacients que són *POLE* mutades, per tal d'evitar-les tractaments innecessaris. Les raons associades a aquest bon pronòstic encara no estan del tot clares, però s'han especulat diferents explicacions com que el genotip ultramutat i el número elevat de substitucions d'aquests tumors és incompatible amb la proliferació i supervivència de les cèl·lules⁷⁴, o que aquests tumors produeixen una gran multitud de neoantígens immunogènics que promouen una resposta immune molt potent.⁷⁵ Estudis recents també han proposat que aquests tumors poden ser altament sensibles a les innovadores immunoteràpies que inhibeixen els punts de controls dels sistema immune (*immune checkpoint inhibition*), com les teràpies anti-PD-L1, ja que s'ha vist que aquests tenen sobreexpressió de PD-L1.⁷⁵ Tot i que hi ha pocs casos, existeixen tumors que presenten mutacions al gen *POLE* i també mutacions al gen *TP53*, també anomenats *double positive* o *multiple classifiers* perquè molecularment presenten característiques que els classificarien a la vegada en dos grups diferents (el grup 1 i 4 del TCGA). No obstant, en aquests casos s'ha observat que, tot i presentar mutacions a *TP53*, aquestes pacients segueixen retenint les característiques de bon pronòstic i pertanyent al grup de baix risc.^{71, 76}

4.1.2 GENS REPARADORS D'ERRORS EN L'APARELLAMENT DE BASES DE L'ADN (MMR) I HIPERMETILACIÓ PROMOTOR MLH-1

Els microsatèl·lits (MS) són petites seqüències repetitives, de un a cinc nucleòtids, que es constitueixen amb blocs de repeticions en tàndem i que es troben dispersats al llarg de tot el genoma. Per la seva naturalesa repetitiva, les regions del genoma amb MS són difícils de replicar, i els enzims encarregats de la síntesi de l'ADN poden originar petits bucles, a causa d'insercions o delecions d'alguns nucleòtids, en la nova cadena sintetitzada. En condicions normals, les cèl·lules disposen de tota una maquinària per reconèixer i reparar aquests errors i així mantenir la integritat del genoma. Un d'aquests mecanismes és l'anomenat sistema de reparació d'aparellaments erronis de bases de l'ADN (MMR, *mismatch repair*), que actua sobre els nucleòtids que han estat incorporats de manera incorrecta (*mismatched nucleotide*) durant el procés de replicació de l'ADN (figura 30). Una de les formes d'instabilitat genòmica més estudiada és la instabilitat que afecta els MS, anomenada instabilitat de microsatèl·lits (MSI, *microsatellite instability*), que es va descriure per primer cop en càncer de colon esporàdic i en càncer colorectal hereditari no polipòsic (CCHNP).⁷⁷ La MSI ve determinada per l'alteració en la longitud dels MS al comparar la longitud dels MS de l'ADN tumoral i de l'ADN no tumoral corresponent (figura 31). Aquesta alteració dels MS és la conseqüència de l'actuació d'un sistema de reparació d'errors defectuós, també anomenat, *mismatch repair defective* (MMR-D). La pèrdua de funcionalitat de les proteïnes implicades en el sistema MMR com MSH2, MSH6, MLH1, PMS2 entre d'altres, pot ser causada per mutacions somàtiques o germinals, o per silenciament transcripcional per metilació.⁷⁸⁻⁸⁰

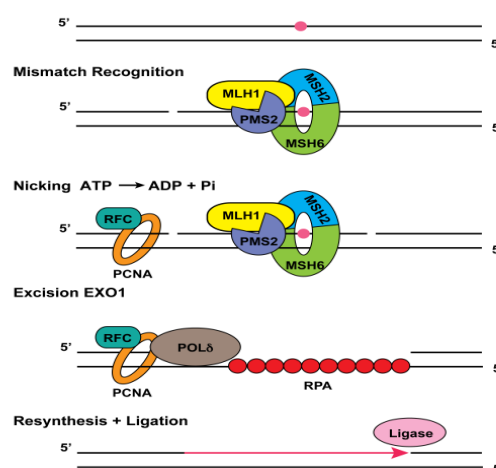


Figura 30. Representació del mecanisme de reparació de l'ADN mitjançant el sistema MMR i les proteïnes implicades. Font: Peggy Hsieh (NIDDK)

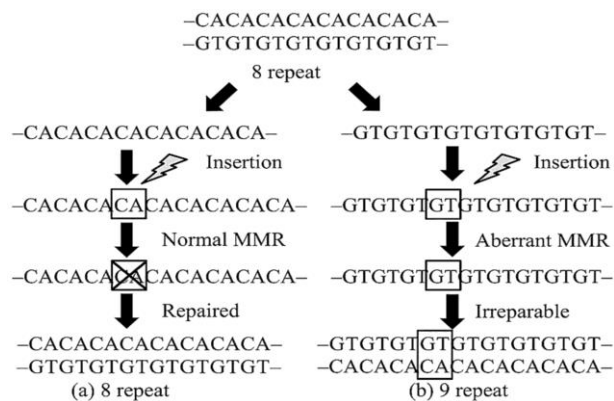


Figura 31. Exemple de la inestabilitat de microsatèl·lits (MSI) causada per defectes en les proteïnes del sistema de reparació MMR. En a) hi ha un correcte funcionament del sistema MMR, i en el b) el sistema MMR està defectuós. ⁷⁰

Aproximadament entre el 20-30% de CEE esporàdics presenten MSI causada principalment per la metil·lació de promotor del gen *MLH1* (*MutL Homolog 1*), que és més o menys el doble del descrit en càncer de colon.⁸¹ Aquests tumors es caracteritzen per presentar unes taxes de mutacions elevades de 18×10^{-6} mutacions/Mb, per una histologia majoritàriament endometrioides, tot i que també s'han descrit tumors amb histologies no endometrioides, i per presentar limfòcits infiltrants tumorals i limfòcits peritumorals. A més a més, també se'ls ha associat amb un patró d'invasió tipus MELF (*Microcystic Elongated and Fragmented*) i amb LVSI (*Lymphovascular Space Invasion*). Per les seves característiques aquest tipus de tumor, igual que els tumors *POLE* mutats, també són candidats potencials a l'ús d'immunoteràpies que inhibeixen els punts de control del sistema immunològic (*immune checkpoint inhibition*), com les teràpies anti-PD-L1.

Tot i aquestes característiques, hi ha controvèrsia quant a l'estratificació del risc dels pacients amb aquests tumors, ja que hi ha estudis que els associen amb indicadors pronòstics desfavorables incloent alt grau histològic i presència de LVSI. Mentre que, altres autors no han mostrat associació amb el pronòstic o alguns els associen amb una millor supervivència. Aquestes dades són un possible reflex del maneig diferencial entre centres pel que fa a l'estadiatge i a l'ús de teràpies adjuvants.⁸²

4.1.3 TP53

El gen *TP53* (*Tumor Protein P53*) codifica per un factor de transcripció nuclear que s'expressa en resposta a diverses senyals d'estrés com el dany a l'ADN, xocs de calor i la hipòxia. La proteïna p53, un cop activada, té la funció de mantenir la integritat i estabilitat del genoma desencadenant l'aturada del cicle cel·lular, la reparació de l'ADN, i l'apoptosis; per això a aquesta proteïna se la coneix com el guardià del genoma.⁸³ P53 és considerada una proteïna supressora tumoral amb una longitud de 393 amino àcids constituïda per un domini de transactivació N-terminal, un domini ric en prolines, un domini d'unió al ADN, un domini de tetramerització i un domini regulador C-terminal.⁸⁴ En condicions normals, la síntesis i degradació d'aquesta proteïna està estrictament regulada, fent que el seu nivell d'expressió es mantingui molt baix. Una proteïna p53 anòmala té un recanvi més prolongat resultant en una sobreexpressió de la proteïna que pot ser detectada per IHQ.⁸⁵ La gran majoria de gens supressors tumorals presenten mutacions *frameshift* o *nonsense* i resulten en una pèrdua d'expressió de la proteïna. En canvi pel gen *TP53* la majoria de mutacions detectades són *missense* i es localitzen principalment al domini d'unió a l'ADN (exons 5 al 8), concretament afectant els residus R175, G245, R248, R249, R273 i R282 (figura 32).⁸⁶

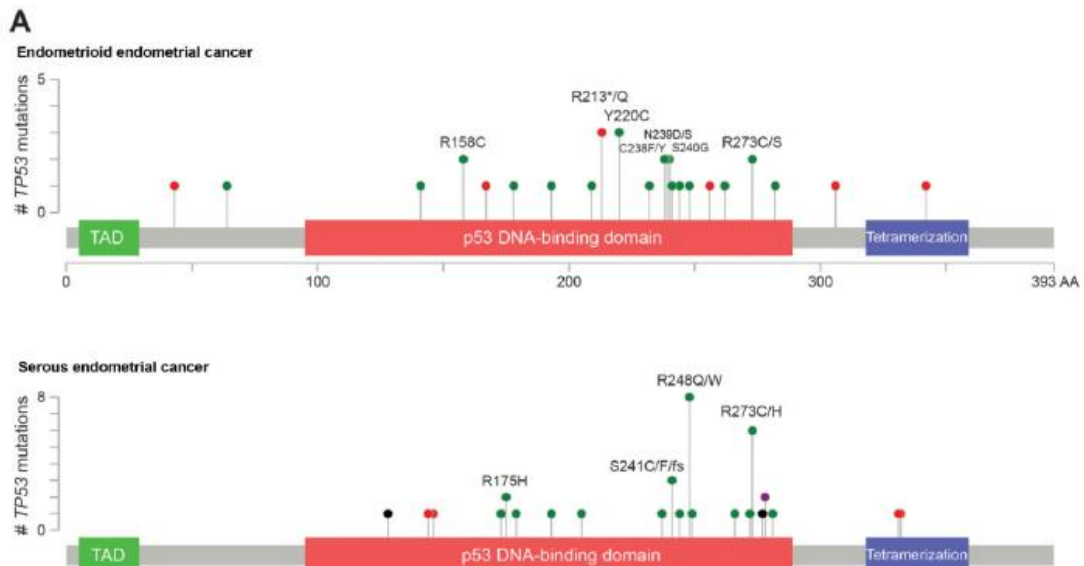


Figura 32. Distribució i espectre de les mutacions de *TP53* en el càncer d'endometri.⁸⁶ Els cercles verds indiquen mutacions *missense*, els vermells mutacions truncadores, i els negres insercions *in-frame* i delecions.

Capítol 1

S'ha vist que aquestes mutacions repercuteixen amb una pèrdua de funcionalitat, com a conseqüència poden exercir activitats dominants negatives o poden obtenir noves propietats oncogèniques. Els diferents tipus de mutacions que es detecten en aquest gen tenen efectes diferents a nivell proteic. L'estudi de Murnyák ⁸⁷, on analitza les dades de la base de dades IARC (*International Agency for Research on Cancer*) *TP53*, ha demostrat que la gran majoria de mutacions al gen *TP53* són *missense* i produeixen una sobreexpressió detectable mitjançant IHQ, contràriament les mutacions *nonsense* que són menys freqüents i són responsables de produir proteïnes truncades donen resultats d'IHQ negatius (*null pattern*), és a dir, que no es detecta la proteïna.

Està plenament reconegut el paper que juga el gen *TP53* en el desenvolupament tumoral, ja que s'han trobat alteracions en aquest gen en més del 50% de tumors humans ⁸⁸, i a més a més les mutacions en aquest gen són les que han estat més associades a la pitjor supervivència dels pacients, més que qualsevol altre gen.⁸⁹ En el CE, de manera global, s'han detectat freqüències mutacionals del gen *TP53* del 25% principalment associades a CEAG.⁵⁹ Específicament en el subtipus molecular *CN high/serous-like*, s'ha descrit una altíssima proporció de mutacions en aquest gen (91 %), i en el grup *POLE-ultramutated* s'han trobat freqüències del 35%. D'aquesta manera les alteracions en aquest gen han estat plenament relacionades amb càncer d'alt risc, com és el *CN-high*, el carcinoma serós d'ovari i el càncer de mama *basal-like*. En relació amb això, diferents estudis han evidenciat que les mutacions a *TP53* estan associades amb tumors que presenten duplicacions en tàndem al llarg de tot el genoma.^{40,90,91} Tot i això, cal esmentar que sobre un 10% de casos del grup *CN high* no presenten mutacions en aquest gen, i que per tant la detecció d'alteracions en aquest gen no pot ser utilitzada com una característica definitiva de carcinomes d'alt grau i tampoc pot servir per discriminar entre CEE i CSE.⁸⁶ De la mateixa manera, s'ha demostrat que existeixen tumors (2-3%) amb més d'una característica molecular, com els tumors amb mutacions a *TP53* i *POLE*, i els tumors amb mutacions a *TP53* i MSI. Conseqüentment, és molt important establir un bon ordre dels tests moleculars, com per exemple primer realitzar la determinació de *POLE* i a posteriori la de p53.^{61,62} Les alteracions moleculars de *TP53* s'ha vist que són events iniciadors de la tumorigènesis ja que s'han descrit en lesions premalignes, en displasies endometrials glandulars i en carcinoma intraepitelial endometrial.⁹²

4.1.4 ALTERACIÓ SOMÀTICA EN EL NÚMERO DE CÒPIES (ASNC) DELS GENS

En el nucli, els gens, de manera normal es troben en una dotació gènica diploide (2N), és a dir dos còpies, una per cromosoma homòleg. Quan existeixen alteracions somàtiques en el número de còpies dels gens (ASNC), significa que hi ha una alteració en les dotacions gèniques dels gens produïda per un guany o amplificació del número de còpies (3N, 4N... o amplificació massiva), o per una pèrdua o deleció dels gens (1N, 0N).⁹³ (figura 33). En els últims anys molts estudis han subratllat que els tumors amb un alt nombre d'ASNC tendeixen a tenir pitjor pronòstic que els tumors diploides.⁸⁹

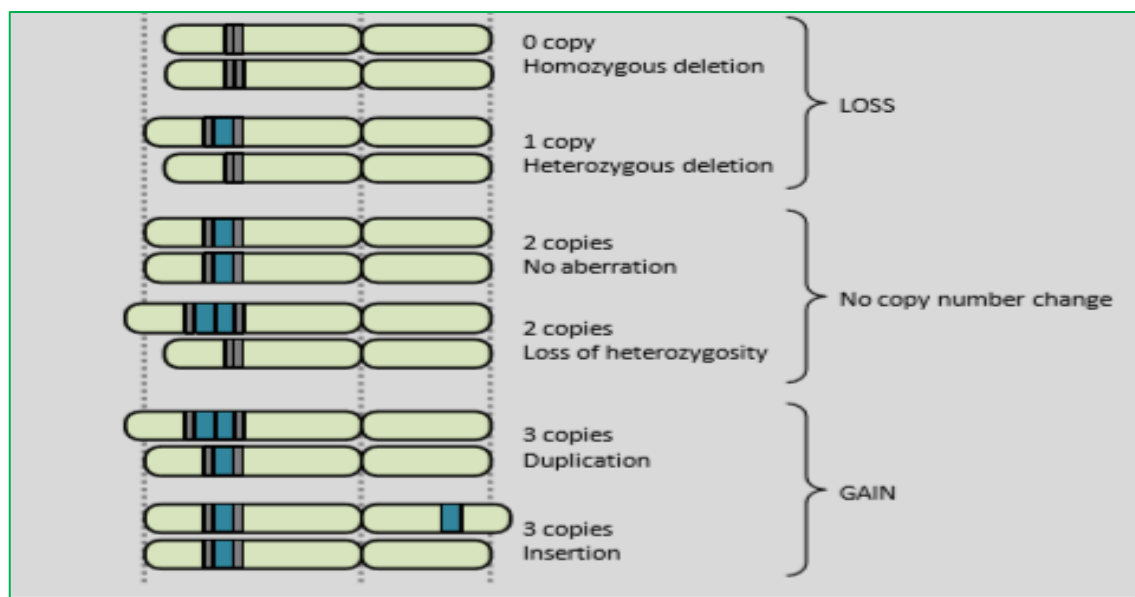


Figura 33. Exemples de les alteracions en el número de còpies dels gens (ASNC), com els guanys (*gain*) i les pèrdues (*loss*). (Font: Gaëlle Marenne, 2012).

Segons les dades del TCGA⁵⁹ la gran majoria de CSE (94%) s'inclouen dins el grup 4 (*CN high /serous-like*). Aquest grup es caracteritza per un alt grau d'ASNC com amplificacions focals recurrents (guanys) que afecten localitzacions on trobem gens com *ERBB2* (17q12) (27%) i *CCNE1* (19q12) (42%). Altres laboratoris^{94,95} també han descrit en els CSE amplificacions focals en altres gens com *PIK3CA* (3q26.32) (60%), i *EVI1* (*MECOM*) (3q26.2) (21.7%). El grup 4 del TCGA també es caracteritza per tenir una alta freqüència de mutacions al gen *TP53* (90%), poca MSI (6%) i poques mutacions al gen *PTEN* (11%) en comparació a la freqüència que es troba en els CEE (84%). Aquests tumors tenen molt pitjor pronòstic que els tumors de tipus endometrioides. D'aquesta manera es pot afirmar que les ASNC juguen un paper molt important en la patogènesi del CSE. Recentment, s'ha descrit que les ASNC proveeixen més informació pronòstica que les mutacions

Capítol 1

puntuals, i que les amplificacions focals tendeixen a conferir un pitjor pronòstic comparat amb les amplificacions que engloben més amplitud. Això és consistent amb el fet que en models animals s'ha vist que els desequilibris a gran escala de les dosis gèniques fan que es produeixi un estrès proteotòxic que perjudica la supervivència de les cèl·lules tumorals. Addicionalment, també s'ha observat que els tumors amb mutacions al gen *TP53* presenten més grau d'aneuploidia i ASNC que els tumors *TP53 wild type*.⁸⁹

ERBB2 (*Erb-B2 receptor Tyrosine Kinase 2*) és una glicoproteïna transmembrana, membre de la família dels receptors dels factors de creixement epidèrmics, concretament és el receptor del factor de creixement epidèrmic humà 2. És un receptor tirosina cinasa que quan dimeritza activa la via de transducció de senyal a través de la fosforilació dels seus residus tirosina cinasa intracel·lulars, i d'aquesta manera activa vies, com la via Ras/Raf/MAPK i la PI3K/AKT/mTOR. Conseqüentment, l'amplificació genòmica i sobreexpressió d'*ERBB2* està associada amb un increment de la proliferació, pitjor supervivència i resistència a tractaments en diferents tipus de tumors incloent el CSE.⁹⁶ Diferents estudis han correlacionat l'estat del gen *ERBB2* i les característiques clinicopatològiques del CSE, associant la seva amplificació amb temps de supervivència menors que en els casos amb estats normals d'aquest gen.⁹⁷ Amb tot el descrit, el gen *ERBB2* és una potencial diana terapèutica per al CSE i recentment, s'han publicat diferents estudis més prometedors en aquesta línia.^{98,99}

El gen *CCNE1* (*Cyclin E1*) codifica per la ciclina E1 que està implicada en la regulació de la transició de la fase G1 a la fase S del cicle cel·lular a través d'interaccions amb les proteïnes cdk2, la fosforilació del Rb (retinoblastoma) i el factor E2F-1. Els alts nivells d'aquesta ciclina fan que s'acceleri la transició a través de la fase G1, i s'ha vist que la seva acumulació és freqüent en un alt nombre de tumors.¹⁰⁰⁻¹⁰² La degradació de la *CCNE1* està mediada per la unió a *FBXW7* seguit de la ubiquïtinització mitjançant la via del complex SCF (Skp, Cullin, F-box). D'aquesta manera tant les mutacions al gen *FBXW7*, que també han estat descrites al CSE (20%), com les amplificacions de *CCNE1* produeixen l'acumulació de ciclina E1 en el CSE. Així veiem que quasi la meitat del CSE està afectat per aquestes alteracions.⁹⁴ A més a més, també s'ha descrit que les amplificacions de *CCNE1*, juntament amb les mutacions de *TP53*, estan associades a tumors que presenten duplicacions en tàndem al llarg de tot el genoma.^{40,90,91} Conseqüentment, l'activació de *CCNE1* pot jugar un paper primordial en conduir la tumorigènesis en el CSE, i també es pot considerar com una potencial diana terapèutica.

El gen *PIK3CA* (*Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha*) codifica per la proteïna p110 α que és la subunitat catalítica de PI3K. Aquesta subunitat juga un paper principal en la via de senyalització PI3K/AKT/mTOR, eix central de la senyalització cel·lular. Tant les mutacions, com les amplificacions de *PIK3CA* produeixen una desregulació d'aquesta via, que és la via més freqüentment alterada en les neoplàsies humanes i es considera fonamental per al desenvolupament tumoral desencadenant senyals de proliferació cel·lular.¹⁰³ Tot i això, recentment s'ha observat que les amplificacions de *PIK3CA* s'associen a algun tipus d'activació de la via PI3K, però no es correlaciona amb marcadors d'activitat d'AKT i mTOR. Això suggereix que en el cas de les amplificacions de *PIK3CA*, les associacions entre l'amplificació i l'activació de la via PI3K pot ser indirecta, resultat de la propensió que tenen els tumors *serous-like* a tenir alt ASNC i per tant, a tenir activades altres vies intercomunicants. Aquestes dades posarien en dubte l'ús d'inhibidors de la via PI3K/AKT/mTOR per al tractament del CSE. El que s'ha vist és que l'amplificació focal de *PIK3CA* està associada amb l'expressió de gens implicats en el punt de control de la fase G2 del cicle cel·lular i la reparació de l'ADN. A més a més, també s'ha vist que les amplificacions de *PIK3CA* estan associades a paràmetres clínico-patològics de fenotips més agressius com edats més avançades, estadis FIGO més elevats, alts graus, ganglis positius, infiltració miomètrica i els subtipus CSE.^{104,105}

EVI1 (*Ecotropic Viral Integration Site-1 Protein*) o també anomenat *MECOM* (*MDS1 and EVI1 Complex Locus*) està localitzat cromosòmicament al locus 3q26.2. És un factor de transcripció *zinc finger* que juga un paper important en el desenvolupament normal i en la oncogènesi. En diferents neoplàsies hematològiques s'ha vist que els seus reordenaments sovint activen la seva expressió, i en tumors sòlids d'endometri, ovari, pulmó, fetge també s'ha descrit la seva sobreexpressió i amplificació.¹⁰⁶⁻¹⁰⁹ A més a més, aquestes alteracions s'han associat significativament amb tumors més agressius i amb pitjor pronòstic indicant que *EVI1* és un oncogen dominant en molts tipus de tumor.¹¹⁰ *EVI1* s'ha relacionat amb diferents vies de senyalització com PI3K/AKT i RAS/ERK, i s'ha descrit que pot promoure la transcripció de PBK (AKT) en el carcinoma serós d'ovari d'alt grau, i inhibir a TGF β que es sap que actua com a supressor tumoral. A través d'aquestes vies s'ha suggerit que la sobreexpressió de *EVI1* pot estar promovent l'adquisició de quimioresistència a través de la inducció de l'autofàgia.¹¹¹ En el treball de Fields et al ¹¹² s'ha posat de manifest la importància i la funcionalitat dels freqüents ASNC que afecten la regió 3q26, apuntant que els gens alterats en aquesta regió, com *PIK3CA* i *EVI1* entre altres, actuen de manera altament coordinada, que és el que anomenen "*oncassette*", per tal de conduir el procés tumorigènic.

4.2 ALTRES ALTERACIONS MOLECULARS

4.2.1 PIK3CA, PIK3R1, PTEN (VIA PI3K/AKT/mTOR)

La via PI3K/AKT/mTOR regula aspectes claus de la biologia del càncer incloent el metabolisme, el creixement cel·lular i la supervivència (figura 34).¹⁰³ Després de l'estimulació, per diferents factors de creixement, dels receptors tirosina cinasa (RTK) com poden ser EGFR, HER2, s'activa la PI3K i fosforila el lípid PIP2 donant com a producte el lípid PIP3. Aleshores PIP3 recluta a la membrana diferents quinases com l'AKT, i aquesta cinasa llavors es fosforilada i activada a través del complex mTOR2 i de PDK1. Entre les seves dianes, AKT (PKB) fosforila i inhibeix TSC2, la qual indirectament inhibeix el complex mTORC1.^{113,114} PTEN s'encarrega d'hidrolitzar PIP3 a PIP2, i així desactiva la via PI3K/AKT/mTOR. Respecte altres tipus de càncer, el CE presenta les taxes més elevades (93%) d'alteració d'aquesta via ⁵⁹, amb mutacions als gens *PIK3CA*, *PIK3R1*, *PTEN* i amplificacions de *PIK3CA* entre d'altres, que es troben en diferents freqüències en funció del grup molecular.

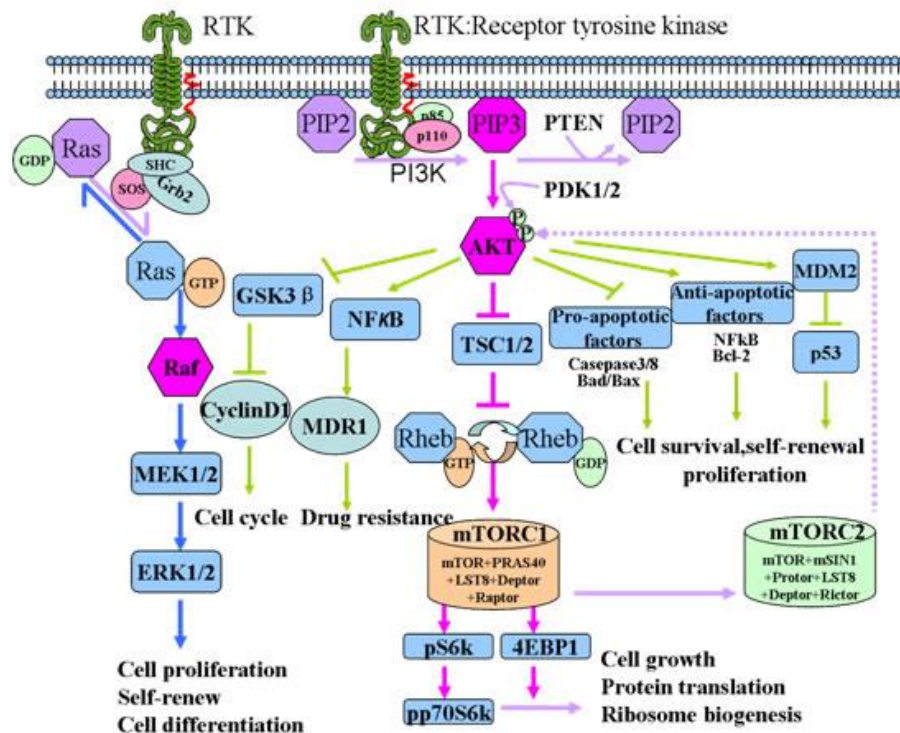


Figura 34. Exemple de la via de senyalització PI3K/AKT/mTOR. Les fletxes signifiquen activació de la via i les barres inhibició.¹¹³

Existeixen quatre classes de PI3K, de les quals la classe IA és la que s'ha relacionat més amb el càncer. Estructuralment està constituïda per una subunitat catalítica, la PIK3CA o també anomenada p110 α , i una subunitat reguladora, la PIK3R1 o també anomenada p85. Les mutacions en els gens *PIK3CA* i *PIK3R1* generalment són excloents entre elles, però contràriament aquestes poden coexistir amb les mutacions de *PTEN*.¹¹⁵ S'ha vist que els carcinomes d'endometri presenten un patró específic de mutacions de *PIK3CA* que es caracteritza per presentar les mutacions en els dominis ABD i C2. Les mutacions d'aquests gens provoquen l'activació constitutiva de la via PI3K/AKT/mTOR.^{116,117}

PTEN (Phosphatase and Tensin Homolog) és un gen supressor tumoral localitzat al cromosoma 10q23.2 que codifica per una fosfatasa lípida i actua com un regulador negatiu de la via PI3K/AKT/mTOR, ja que la seva funció és defosforilar el lípid PIP3. La pèrdua de funció d'aquesta proteïna ha estat proposada com una de les alteracions moleculars iniciadores del desenvolupament de la carcinogènesi del CEE, ja que la disminució de l'activitat de PTEN augmenta la proliferació i la supervivència de les cèl·lules.¹¹⁸ Les causes de la seva pèrdua de funció poden ser mutacions, la pèrdua d'heterozigositat (LOH) i la metilació del seu promotor.¹¹⁹ Cal dir que s'ha descrit que les mutacions d'aquest gen són una de les primeres alteracions moleculars que es donen en el desenvolupament del CEE, ja que ocorren en un 20-27% en hiperplàsies endometrials i en un 55 % en neoplàsies intraepiteliales. Contràriament, les mutacions de *PIK3CA* en hiperplàsies complexes atípiques són rares, i apareixen com events moleculars més tardans en el desenvolupament del CEE.¹²⁰

4.2.2 KRAS

El gen *KRAS (Kirsten Rat Sarcoma Viral Oncogene Homolog)* codifica per la proteïna K-Ras que forma part de la via de senyalització RAS/RAF/MEK/ERK. K-ras és una GTPasa, la qual s'encarrega de transformar les molècules GTP a GDP. Amb aquesta activitat funciona com un interruptor intracel·lular que dona senyals que indiquen a la cèl·lula que creixi i prolifereixi, o que maduri i assumeixi funcions especialitzades de diferenciació. Les mutacions somàtiques en aquest gen produeixen una inhibició de l'activitat GTPasa fent que es mantinguin de manera constitutiva les senyals de proliferació i per tant permetent la proliferació del tumor.¹²¹ Les mutacions de *KRAS* són molt més freqüents en el CEE (18%) que en el CSE (3%). S'ha observat que les mutacions de *KRAS* apareixen en etapes primerenques del desenvolupament tumoral del CEE, detectant-se en hiperplàsies endometrials atípiques. Aquestes mutacions s'ha vist que poden coexistir amb mutacions en altres gens com *PTEN*, *PIK3CA* i *PIK3R1*, fet que suggereix que les mutacions de

KRAS no són funcionalment redundants amb les mutacions dels gens de la via PI3K/AKT/mTOR.¹²²⁻

124

4.2.3 *FGFR2*

El gen *FGFR2* (*Fibroblast Growth Factor Receptor 2*) codifica pel receptor amb activitat tirosina cinasa anomenat receptor 2 del factor de creixement de fibroblast (*FGFR2*). Aquest receptors, a l'activar-se, participen en processos importants com ara el creixement i divisió cel·lular, la maduració cel·lular, la formació de vasos sanguinis, la cicatrització de ferides i el desenvolupament embrionari.¹²⁵ En el CEE s'han detectat mutacions en aquest gen en un 12% de CEE. Aquestes mutacions són mútuament excloents amb les mutacions de *KRAS*, en canvi, la majoria de casos CEE *FGFR2* mutats (77%) també són *PTEN* mutats. S'ha vist que aquest gen presenta un *hotspot*, el residu S252W, que està localitzat en el domini extracel·lular que intervé en la unió al lligand, i és on es troben al voltant del 41% de mutacions reportades. En línies cel·lulars s'ha observat que els casos amb mutacions activadores de *FGFR2* són més propenses a respondre a inhibidors d'aquest receptor, fet que senyala les mutacions de *FGFR2* com a potencials dianes terapèutiques.¹²⁶⁻¹²⁸

4.2.4 *ARID1A*

Recentment, *ARID1A* (*AT-Rich Interaction Domain 1A*) ha estat descrit com un gen supressor tumoral que codifica per la proteïna BAF250a, que és un component del complex SWI/SNF remodelador de la cromatina (*SWI/SNF Chromatin-Remodeling Complex*). Les mutacions d'*ARID1A* poden resultar en una activació de la via PI3K a través de la desregulació de PIK3IP1 (*PI3K Interacting Protein 1*). La pèrdua d'expressió d'*ARID1A* mitjançant immunohistoquímica ha estat observada en CEE de baix grau (26-29%) i d'alt grau (39%). Les mutacions s'han detectat al voltant del 40% en CEE de baix grau, i el 50% d'aquests tumors mutats presentaven pèrdua d'expressió de la proteïna.¹²⁹⁻¹³¹

4.2.5 *PPP2R1A*

PPP2R1A (*Protein Phosphatase 2 Scaffold Subunit Aalpha*) codifica per la subunitat A α de la fosfatasa específica Ser/Thr PP2A, que és una proteïna supressora tumoral. La fosfatasa PP2A regula diferents vies com ERK, PI3K/AKT/mTOR, i WNT. Cal remarcar que les mutacions són bastant freqüents en el CSE (fins un 40%), mentre que es troben en molt baixa freqüència en el CEE (inferior al 7%). A més a més, les mutacions trobades als pocs CEE són en el subtipus d'alt grau, suggerint que aquestes mutacions estan associades amb l'agressivitat dels tumors i la pitjor

supervivència dels pacients. Les mutacions estan confinades als exons 5 i 6, en el CE la majoria d'elles afecten els hotspots 179, 256 i 257.¹³²⁻¹³⁴

4.2.6 *FBXW7*

FBXW7 (*F-Box And WD Repeat Domain Containing 7*) codifica per la proteïna supressora tumoral FBOX, que és un component del complex ubiquitin lligasa SCF (Skp, Cullin, F-box). Aquest complex té com a funció d'ubiquitar substrats fosfoproteïcs per poder ser degradats pel proteosoma. Les mutacions en aquest gen afecten les repeticions WD40 d'unió a substrat de la proteïna FBOX resultant en una pèrdua de la funció del complex SCF i produint una acumulació d'oncoproteïnes.¹³⁵ La ciclina E1 és una de les proteïnes que es veu afectada per les mutacions a *FBXW7*, fent que s'observi un increment de l'expressió de ciclina E1 en els CSE amb mutacions a *FBXW7*. En els CSE també s'ha vist que les mutacions a *FBXW7* i les amplificacions de *CCNE1* són events moleculars excloents. Així s'ha descrit que sobre un total d'un 57 % de casos CSE tenen afectat el sistema CCNE1-*FBXW7*, tant per amplificacions al gen *CCNE1* com per mutacions (20%) o delecions a *FBXW7*. També s'ha observat que les mutacions a *FBXW7* es donen en etapes inicials del desenvolupament tumoral ja que s'han detectat en carcinomes serosos intraepitelials.⁹⁴

4.2.7 *L1CAM / CTNNB1*

L1CAM (*L1 Neuronal Cell-Adhesion Molecule*) és una glicoproteïna transmembrana de la superfamília de les immunoglobulines. En models experimentals s'ha observat que l'expressió de *L1CAM* indueix la migració de les cèl·lules tumorals, la invasió, la transició epiteli mesènquima (EMT, *Epithelial Mesenchymal Transition*) i la quimioresistència.¹³⁶⁻¹³⁸ Recentment, *L1CAM* ha guanyat molta força com a marcador pronòstic específic, i potencial diana terapèutica en el CE i altres tumors.^{139,140} En l'estudi de Kommos ¹⁴¹, s'ha demostrat que l'expressió de *L1CAM* es correlaciona amb el casos de CEE d'alt grau, d'estadis avançats i amb pronòstic desfavorable que estan inclosos en el grup 3 del TCGA, és a dir, els p53 wt/NSMP. També s'ha definit una associació entre l'expressió de *L1CAM* i el CSE, els quals són coneguts per presentar una alta taxa de mutacions al gen *TP53*. Alguns autors han suggerit una associació robusta entre l'expressió de *L1CAM* amb tumors amb mutacions a *TP53*. A més a més, s'ha vist que després de la histerectomia, les pacients del grup p53 wt/NSMP *L1CAM* positives tenen un risc similar a aquelles pacients del grup p53 mutat. D'aquesta manera aquestes pacients es podrien beneficiar de teràpies adjuvants, tot i estar incloses en el grup de baix risc. Un efecte pronòstic similar de l'expressió de *L1CAM* també s'ha observat en en el grup MMR-D. Amb tot això, es podria considerar els casos *L1CAM* positius com un subgrup pronòstic dins els grups p53 wt/NSMP i

Capítol 1

MMR-D, i així redefinir el grup de risc intermig que ocupen aquests dos grups que constitueixen un gran nombre de casos de CE (64-79%).¹⁴¹

Tot i que el grup p53wt/NSMP no té cap gen amb efecte pronòstic, s'ha observat que fins a un 50% de pacients d'aquest grup presenten mutacions activadores al gen *CTNNB1* (*Catenin Beta 1*), que codifica per la proteïna β -catenina.⁵⁹ Aquesta proteïna és un component integral de la via canònica WNT. Juntament amb E-cadherina conformen una unitat que és important per la diferenciació cel·lular, pel manteniment de l'arquitectura dels teixits normals i per la transducció de senyals. β -catenina està negativament regulada per GSK-3 β (*Glycogen Synthase Kinase 3 β*) ja que la fosforila i activa la seva degradació. En els CEE s'ha vist que la majoria de mutacions de *CTNNB1* es localitzen a l'exó 3 i que resulten en l'evasió de la fosforilació mediada per GSK-3 β produint la seva acumulació nuclear i conseqüentment un augment de la proliferació cel·lular.^{142,143} L'estudi de Kommos et al.¹⁴¹ suggereix que les mutacions de *CTNNB1* poden activar l'expressió de L1CAM i així establir un comportament més agressiu d'aquests tumors. D'aquesta manera la determinació d'aquestes mutacions també podria donar informació pronòstica addicional al classificador ProMisE ja que permetria identificar casos amb més risc de recurrència dins el grup p53wt/NSMP.

4.2 TERÀPIA MOLECULAR DIRIGIDA

La histerectomia total i l'exèresi del cèrvix, els ovaris i les trompes de fal·lopi són els tractaments quirúrgics estàndards que es realitzen a les pacients amb càncer d'endometri amb aparent estadi I i que resulten efectius en la gran majoria de casos. No obstant això, moltes pacients en estadis avançats i algunes en estadis inicials presentaran recidiva en el temps. Molts estudis han avaluat aquest risc de recurrència post-cirurgia i han intentat definir les teràpies adjuvants més adequades per tal de reduir aquest risc. Així s'ha vist que els factors d'alt risc inclouen el grau del tumor, la mida del tumor, el tipus histològic, la profunditat de la invasió del miometri, la presència d'invasió de l'espai linfovacular (LVSI) i l'estadi.¹⁴⁴ De manera resumida, tenint en compte les dades dels assaigs clínics portats a terme en els últims anys, podem dir que la quimioteràpia adjuvant és la que aporta més benefici en pacients en estadis III i IV, i que la radioteràpia pot ser considerada per al control local. En els casos en què existeix malaltia recurrent i metàstasis, es considera que el càncer és incurable i està associat a molt mal pronòstic.

En aquests casos es pot pensar en altres teràpies en assaigs clínics com la hormonoteràpia, immunoteràpia i les teràpies dirigides.^{144,145} En l'actualitat, gràcies al coneixement de la classificació molecular del CE, s'està intentant incorporar aquesta informació molecular als algoritmes d'estratificació del risc, per tal de poder millorar els grups i conseqüentment poder recomanar millors teràpies.

Sobretot s'està posant molt esforç en aconseguir teràpies efectives per als carcinomes d'endometri d'estadis avançats i amb malaltia recurrent, ja que fins ara els tractaments que hi ha són inadequats i els pronòstics no han millorat.¹⁴⁶ A continuació es mostra un resum actualitzat de les teràpies moleculars dirigides (amb el codi de l'assaig clínic) en el CE en funció dels grups moleculars i dels marcadors moleculars afectats.¹⁴⁶ (figura 35).

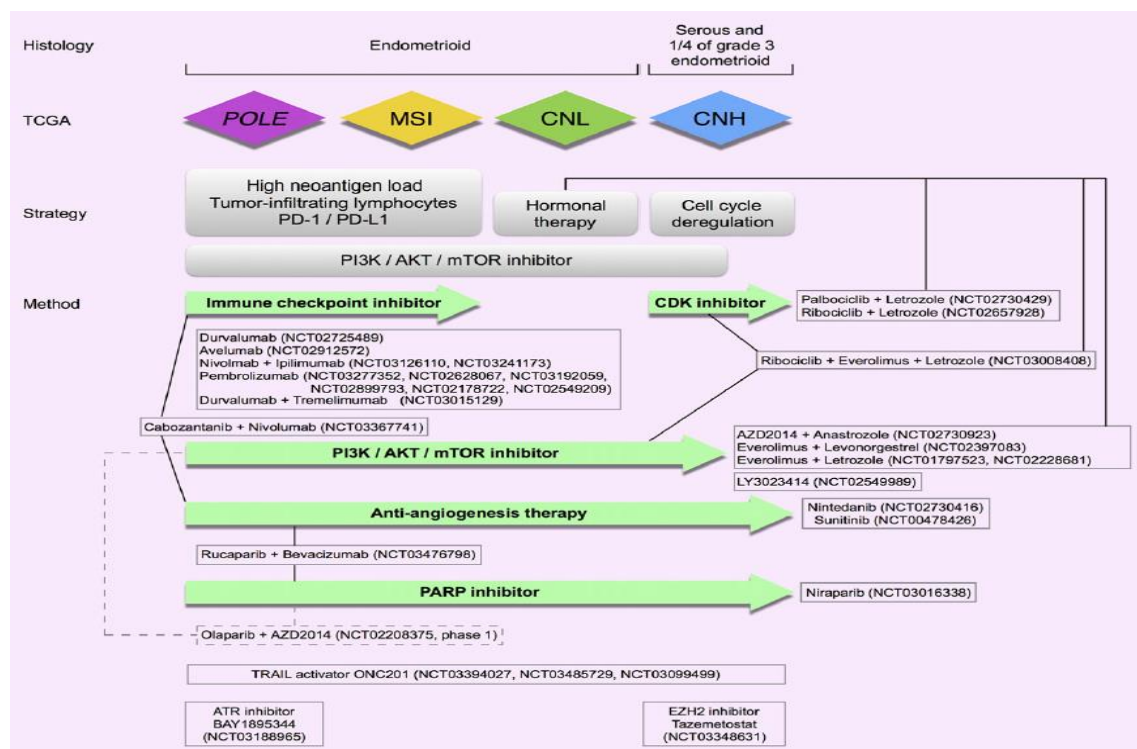


Figura 35. Teràpies dirigides i assaigs clínics actuals, en funció dels grups moleculars TCGA i de les alteracions moleculars associades en cada grup.¹⁴⁶

CAPÍTOL 2

HIPÒTESIS I OBJECTIUS

1 HIPÒTESIS

Els dos subtipus més freqüents de càncer d'endometri (CE) són el carcinoma endometriode d'endometri (CEE) i el carcinoma serós d'endometri (CSE). El diagnòstic diferencial entre aquestes dues tipologies no sempre és fàcil, hi ha casos que presenten característiques histològiques dubtoses i ambigües que dificulten el correcte diagnòstic final. És sabut que aquests dos tipus presenten diferents perfils moleculars, i també un pronòstic diferencial, fent que el carcinoma serós d'endometri sigui el subtipus més agressiu i amb el pronòstic més desfavorable. Actualment també s'està posant en coneixement la importància que pot jugar l'heterogeneïtat intratumoral (HIT) en els comportaments més agressius dels tumors i en la resistència als tractaments, com és el cas del CSE. En els últims anys, la irrupció de les tecnologies d'alt rendiment com la NGS (*Next Generation Sequencing*) i la MLPA (*Multiple Ligase-dependent Probe Amplification*) han ofert l'opció d'ampliar el coneixement molecular del càncer d'una manera més precisa i resolutive.

Per tot això s'han plantejat les següents hipòtesis (H):

H1. Les tecnologies d'alt rendiment com la NGS i MLPA permeten fer una anàlisi de diferents alteracions moleculars en gens responsables dels processos tumorigènics. La identificació d'aquests gens en mostres pertanyents als tipus més freqüents de càncer d'endometri (CEE i CSE) permet establir perfils genètics per classificar les mostres d'una manera més precisa i objectiva, i realitzar un diagnòstic final més acurat.

H2. La presència d'heterogeneïtat intratumoral en el tipus de càncer CSE condiciona el comportament agressiu d'aquest tumor.

2 OBJECTIUS

Per avaluar les hipòtesis s'han definit un total de 3 objectius específics (OE):

OE1. Analitzar les alteracions de diferents gens a nivell de variants somàtiques (VS) mitjançant una anàlisi de seqüenciació dirigida (*targeted sequencing*) amb la tecnologia NGS per tal d'establir un perfil de gens que permeti classificar aquests tipus tumorals. Aquestes alteracions s'han determinat en mostres pertanyents als dos tipus més freqüents de càncer d'endometri (CEE i CSE).

OE2. Analitzar les alteracions somàtiques en el número de còpies de gens (ASNC) mitjançant la tècnica MLPA per tal d'establir un perfil de gens que permeti classificar aquests tipus tumorals. Aquestes alteracions s'han determinat en mostres pertanyents als dos tipus més freqüents de càncer d'endometri (CEE i CSE).

OE3. Estudiar l'heterogeneïtat intratumoral (HIT) en mostres de carcinoma serós d'endometri a diferents nivells d'alteració molecular: (a) a nivell de les alteracions somàtiques en el número de còpies dels gens (ASNC), i (b) a nivell de variants somàtiques (VS).

CAPÍTOL 3

MATERIALS I MÈTODES

En aquest capítol es descriuen els materials i mètodes emprats per desenvolupar els tres objectius específics d'aquesta tesi doctoral. La figura 36 il·lustra la relació entre els objectius, les alteracions moleculars analitzades i els materials i mètodes emprats.

	ALTERACIÓ MOLECULAR	MOSTRES	MÈTODES	MATERIALS I MÈTODES	CAPÍTOL 3
OE1: ALTERACIONS A NIVELL DE VARIANTS SOMÀTIQUES	VS EP	16 CEE + 8 CSE	TS, IHQ, SS		
OE2: ALTERACIONS DEL NÚMERO DE CÒPIES DELS GENS	ASNC EP	12 CEE + 12 CSE	MLPA, IHQ		
OE3: HETEROGENEÏTAT INTRATUMORAL EN EL CSE	ASNC VS EP	3 REGIONS DE 8 CSE	MLPA TS, SS IHQ, SISH		

VS: variants somàtiques, EP: expressió proteïnes, ASNC: alteració somàtica número de còpies, CEE: carcinoma endometriode d'endometri, CSE: carcinoma serós d'endometri, TS: targeted sequencing, MLPA: multiplex ligase-dependent probe amplification, SS: Seqüenciació Sanger, IHQ: immunohistoquímica, SISH: silver in-situ hybridization

Figura 36. Relació entre objectius, i materials i mètodes

1 MOSTRES

Les mostres utilitzades per al desenvolupament d'aquesta tesi doctoral han estat biòpsies tumorals i no tumorals provinents de pacients diagnosticades de carcinoma endometriode d'endometri (CEE) o carcinoma serós d'endometri (CSE), i tractades a l'Hospital Universitari Arnau de Vilanova de Lleida (HUAV). Cadascuna de les mostres, després de la cirurgia, va ser immediatament registrada, congelada i emmagatzemada a -80°C al banc de tumors de l'IRBLleida. A la vegada aquestes biòpsies van ser diagnosticades al departament d'Anatomia Patològica de l'HUAV seguint les recomanacions de la OMS (Organització Mundial de la Salut), i els tumors van ser estadificats i gradats seguint el sistema FIGO (taula 7).^{49,50} De cada biòpsia es va avaluar la qualitat tissular en seccions d'hematoxilina-eosina, i a la vegada, es va quantificar el percentatge tumoral, la necrosi i la porció no tumoral. Prèviament a les anàlisis realitzades les pacients van firmar els consentiments informats oportuns, i els estudis també van estar aprovats pel Comitè Ètic d'Investigació Clínica (CEIC) de l'Hospital.

Per assolir cadascun dels objectius específics proposats s'han utilitzat diferents conjunts de mostres, d'aquesta manera tenim la següent distribució de mostres en funció de l'objectiu:

- Per a l'objectiu específic 1 s'han emprat 16 mostres de CEE, 8 mostres de CSE i els respectius teixits no tumorals de cadascuna d'elles.
- Per a l'objectiu específic 2 s'han emprat 12 mostres de CEE, 12 mostres de CSE, i els respectius teixits no tumorals de cadascuna d'elles.
- Per a l'objectiu específic 3 s'han emprat 8 mostres de CSE, i tres porcions de tumor de cadascun dels carcinomes, en total 24 mostres, i els respectius teixits no tumorals de cadascuna d'elles.

Taula 7. Característiques histopatològiques de totes les mostres utilitzades en funció de cadascun dels objectius

OBJECTIU ESPECÍFIC 1				OBJECTIU ESPECÍFIC 2				OBJECTIU ESPECÍFIC 3			
TIPUS				TIPUS				TIPUS			
CAS	HISTOLÒGIC	GRAU	ESTADI	CAS	HISTOLÒGIC	GRAU	ESTADI	CAS	HISTOLÒGIC	GRAU	ESTADI
1	CEE	II	IB	1	CEE	IB	I	1	CSE	III	IA
2	CEE	III	II	2	CEE	IB	II	2	CSE	III	IIIA
3	CEE	II	IB	3	CEE	IA	II	3	CSE	III	IB
4	CEE	II	IB	4	CEE	IB	II	4	CSE	III	IA
5	CEE	II	IA	5	CEE	IA	I	5	CSE	III	IA
6	CEE	III	IIIA	6	CEE	IA	I	6	CSE	III	IIIB
7	CEE	III	IB	7	CEE	IB	I	7	CSE	III	IB
8	CEE	II	IB	8	CEE	IA	I	8	CSE	III	II
9	CEE	III	IB	9	CEE	IB	III				
10	CEE	I	IB	10	CEE	IA	I				
11	CEE	III	IB	11	CEE	IB	III				
12	CEE	III	IB	12	CEE	IA	I				
13	CEE	III	IB	13	CSE	III	IA				
14	CEE	II	II	14	CSE	III	IA				
15	CEE	I	IB	15	CSE	III	IIIA				
16	CEE	III	IB	16	CSE	III	IA				
17	CSE	III	IA	17	CSE	III	IIA				
18	CSE	III	IIIA	18	CSE	III	IB				
19	CSE	III	IB	19	CSE	III	IA				
20	CSE	III	IA	20	CSE	III	IIIB				
21	CSE	III	IA	21	CSE	III	IB				
22	CSE	III	IIIB	22	CSE	III	IIIB				
23	CSE	III	IB	23	CSE	III	II				
24	CSE	III	II	24	CSE	III	IB				

2 EXTRACCIÓ D'ADN

De cadascuna de les mostres tumorals i no tumorals congelades a -80 °C amb el compost OCT (*Optimum Cutting Temperature*) es va extreure ADN mitjançant el Kit DNeasy and Tissue Kit (Qiagen) seguint les instruccions del fabricant. Un cop extret l'ADN, es va determinar la quantitat i la seva puresa mitjançant l'ús de l'equip Nanodrop ND-2000 (Thermo Fisher Scientific). A més a més, la solució d'ADN de les mostres que van ser emprades per l'anàlisi de *targeted sequencing* (NGS) es va quantificar utilitzant el kit Qubit dsADN HS Assay (Invitrogen, Thermo Fisher Scientific) amb l'equip Qubit Fluorometer 1.0 (Invitrogen, Life Technologies).

3 MULTIPLEX LIGASE-DEPENDENT PROBE AMPLIFICATION (MLPA)

En el primer capítol s'ha descrit el funcionament general de la tècnica MLPA, així que en aquest apartat s'especifiquen els kits utilitzats i el protocol específic per desenvolupar-la.

Els kits SALSA MLPA utilitzats van ser el P171-A2 *gain*, el P172-B1 *gain*, el P173-A2 *gain* i el P294-A1 *tumor-loss* de MRC-Holland (Amsterdam, The Netherlands). Amb aquests kits es va avaluar l'estat del número de còpies d'un total de 106 gens (taula 8) en els objectius específics 2 i 3.

Per realitzar aquesta tècnica es van seguir les recomanacions del fabricant. Tot i això, a continuació es descriu breument el protocol:

A. DESNATURALITZACIÓ I HIBRIDACIÓ

Rotular un tub de 0.2 µl per cada mostra, afegir 5µl d'ADN i desnaturalitzar al termociclador durant 15 min. a 98°C. A continuació preparar la mix d'hibridació (MLPA *Buffer* y MLPA *probemix*), afegir-la a l'ADN desnaturalitzat, incubar 1 min a 95°C i a continuació hibridar entre 16-20h a 60°C.

B. LLIGACIÓ

Preparar la mix de lligació (aigua, *Ligase Buffer A*, *Ligase Buffer B* i *Ligase-65*), afegir-la a la mostra hibridada i lligarla a 54°C durant 15 min. Pausar a 20°C

C. AMPLIFICACIÓ

Preparar la mix d'amplificació (aigua, *Salsa pcr primer mix* i *Salsa polymerase*), afegir-la a la mostra lligada i procedir amb el programa d'amplificació: 35 cicles: 30 seg. 95°C; 30 seg. 60°C; 60 seg. 72°C i incubar 20 min. 72°C. Pausar a 15 °C.

D. ANÀLISI DE FRAGMENTS

Anàlisi de fragments mitjançant una electroforesis capil·lar amb l'equip 3130 *Genetic Analyzer* (Applied Biosystems) i el *software* GeneMapper®4.1 (Applied Biosystems).

E. ANÀLISI DE LES DADES I INTERPRETACIÓ DELS RESULTATS

Avaluació dels electroferogrames i dels patrons de pics, càlcul dels DQ (*Dosage Quocient*) després de realitzar la normalització intra i inter mostra, i assignació de l'estat del número de còpies (DQ inferior a 0,7 = pèrdua o deleció, DQ entre 0,7 i 1,4 = normal o diploide, DQ superior a 1,4 = guany, duplicació o amplificació).

Taula 8. Llistat dels 106 gens analitzats per MLPA amb les seves corresponents localitzacions cromosòmiques.

MLPA			
GEN	LOC. CROM.	GEN	LOC. CROM.
<i>NRAS</i>	1p13.2	<i>WT1</i>	11p13
<i>BCAS2</i>	1p13.3	<i>CCND1</i>	11q13
<i>BCAR3</i>	1p22.1	<i>EMS</i>	11q13
<i>PSMB4</i>	1q21.3	<i>FGF3</i>	11q13
<i>MDM4</i>	1q32.1	<i>GSTP1</i>	11q13
<i>MYCL1</i>	1p34.2	<i>RELA</i>	11q13
<i>KIF1B</i>	1p36.23	<i>BRMS1</i>	11q13.2
<i>CAMTA1</i>	1p36.31	<i>FGF4</i>	11q13.3
<i>CHD5</i>	1p36.31	<i>BIRC2</i>	11q22
<i>PRDM16</i>	1p36.32	<i>BIRC3</i>	11q22
<i>DVLI</i>	1p36.33	<i>MMP7</i>	11q22.3
<i>NTRK1</i>	1q23.1	<i>BCL2L14</i>	12p13.2
<i>CENPF</i>	1q32.3	<i>CCND2</i>	12p13.3
<i>PIK3C2B</i>	1q32.1	<i>CDK4</i>	12q14
<i>MYCN</i>	2p24.1	<i>MDM2</i>	12q14.3
<i>BCL2L11</i>	2q13	<i>CYP27B1</i>	12q14.1
<i>ERBB4</i>	2q33	<i>CCNA1</i>	13q12.3
<i>IGFBP2</i>	2q35	<i>BRCA2</i>	13q12.3
<i>IGFBP5</i>	2q35	<i>RB1</i>	13q14.2
<i>FHIT</i>	3p14.2	<i>DLEU1</i>	13q14.3
<i>VHL</i>	3p25.3	<i>MIRN15A</i>	13q14.3
<i>PIK3CA</i>	3q26.32	<i>IRS2</i>	13q34
<i>EVII</i>	3q26	<i>AKT1</i>	14q32.33
<i>BCL6</i>	3q27	<i>BCL2A1</i>	15q24.3
<i>PDGFRA</i>	4q11	<i>NTRK3</i>	15q25.3
<i>TERT</i>	5p25	<i>IGF1R</i>	15q26.3
<i>BIRC1</i>	5q13.2	<i>TSC2</i>	16p13.3
<i>APC</i>	5q22	<i>TOM1L2</i>	17p11.2
<i>PDGFRB</i>	5q32	<i>TP53</i>	17p13.1
<i>HMGA1</i>	6p21.3	<i>NF1</i>	17q11.2
<i>CCND3</i>	6p21.1	<i>ERBB2</i>	17q21.1
<i>NFKBIE</i>	6p21.1	<i>IGFBP4</i>	17q21.2
<i>SERPINB9</i>	6p25.2	<i>BRCA1</i>	17q21.31
<i>ESR1</i>	6q25.1	<i>PPM1D</i>	17q23.2
<i>EGFR</i>	7p12	<i>BIRC5</i>	17q25.3
<i>CDK6</i>	7q21.3	<i>BCL2</i>	18q21.33
<i>MET</i>	7q31	<i>SMAD4</i>	18q21.1
<i>BRAF</i>	7q34	<i>SERPINB2</i>	18q21.33
<i>MOS</i>	8q11	<i>SERPINB7</i>	18q21.33
<i>RNF139</i>	8q24.13	<i>FKBP8</i>	19p13.11
<i>FGFR1</i>	8p11.2	<i>STK11</i>	19p13.3
<i>MYBL1</i>	8q21	<i>CCNE1</i>	19q12
<i>PTK2</i>	8q24	<i>BCL2L1</i>	20q11.1
<i>MYC</i>	8q24.12	<i>MYBL2</i>	20q12
<i>PTP4A3</i>	8q24.3	<i>PTPN1</i>	20q13.13
<i>CDKN2A</i>	9p21	<i>BCAS1</i>	20q13.2
<i>CDKN2B</i>	9p21	<i>AURKA</i>	20q13.31
<i>JAK2</i>	9p24.1	<i>GNAS</i>	20q13.32
<i>NTRK2</i>	9q21.33	<i>UCKL1</i>	20q13.33
<i>PTCH</i>	9q22.3	<i>RUNX1</i>	21q22.12
<i>TSC1</i>	9q34	<i>BCL2L13</i>	22q11
<i>PTEN</i>	10q23.3	<i>SMARCB1</i>	22q11.23
		<i>FAM123B</i>	Xq11.1
		<i>BIRC4</i>	Xq25

4 APLICACIÓ TARGETED SEQUENCING (NGS)

En el primer capítol s'ha fet una descripció detallada del funcionament de la tecnologia NGS i l'aplicació de *targeted sequencing*. En aquest apartat es detallen més concetament els passos realitzats per desenvolupar aquesta anàlisi en els objectius específics 1 i 3.

A. OBTENCIÓ DEL PANELL CUSTOMITZAT DE GENS

Amb l'ús de l'algoritme per al disseny de panells de gens (Illumina) es va fer el disseny d'un panell customitzat de 40 gens basant-nos en els resultats moleculars de diferents estudis realitzats en càncer d'endometri.^{59,94,95,147} A la taula 9 es mostra el llistat dels gens seleccionats relacionats amb la seva localització cromosòmica i amb les funcions cel·lulars en què cadascun d'ells està implicat. El disseny va incloure un total de 1333 amplicons de 250 pb de longitud, que cobrien el 98,59% d'exons, introns i zones no traduïdes de cadascun dels gens seleccionats.

Capítol 3

Taula 9. Llistat dels 40 gens analitzats mitjançant l'aplicació de *targeted sequencing* (NGS), amb la localització cromosòmica, la descripció i les vies de senyalització en que estan implicats.

GEN	LOC. CROM.	DESCRIPCIÓ	VIA SENYALITZACIÓ
<i>ABCC9</i>	chr12	ATP binding cassette subfamily C member 9	Cicle cel·lular
<i>ATR</i>	chr3	ATR Serine/Threonine Kinase	
<i>CCND1</i>	chr11	Cyclin D1	
<i>CSDE1</i>	chr1	Cold Shock Domain Containing E1	
<i>EP300</i>	chr22	E1A Binding Protein P300	
<i>MKI67</i>	chr10	Marker Of Proliferation Ki-67	
<i>PPP2R1A</i>	chr19	Protein Phosphatase 2 Scaffold Subunit A alpha	
<i>TP53</i>	chr17	Tumor Protein P53	
<i>WRAP53</i>	chr17	WD Repeat Containing Antisense to TP53	
<i>ARID1A</i>	chr1	AT-rich interaction domain 1A	Organització cromatina
<i>ARID5B</i>	chr10	AT-Rich Interaction Domain 5B	
<i>CHD4</i>	chr12	Chromodomain Helicase ADN Binding Protein 4	
<i>KMT2B</i>	chr19	Lysine Methyltransferase 2B	
<i>SMARCA4</i>	chr19	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4	
<i>CDH19</i>	chr18	Cadherin 19	Comunicació cèl·lula-cèl·lula
<i>COL11A1</i>	chr1	Collagen Type XI Alpha 1 Chain	
<i>CTNNB1</i>	chr3	Catenin Beta 1	Via WNT
<i>ERBB2</i>	chr17	Erb-B2 Receptor Tyrosine Kinase 2	Via PI3K/AKT/mTOR
<i>PIK3CA</i>	chr3	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	
<i>PIK3R1</i>	chr5	Phosphoinositide-3-Kinase Regulatory Subunit 1	
<i>PTEN</i>	chr10	Phosphatase And Tensin Homolog	
<i>TAF1</i>	chrX	TATA-Box Binding Protein Associated Factor 1	
<i>FGFR2</i>	chr10	Fibroblast Growth Factor Receptor 2	Via RTK/RAS/MAPK
<i>KRAS</i>	chr12	KRAS Proto-Oncogene, GTPase	
<i>MAP3K4</i>	chr6	Mitogen-Activated Protein Kinase Kinase Kinase 4	
<i>NRAS</i>	chr1	NRAS Proto-Oncogene, GTPase	
<i>BCOR</i>	chrX	BCL6 Corepressor	Divers
<i>CSMD3</i>	chr8	CUB And Sushi Multiple Domains 3	
<i>CTCF</i>	chr16	CCCTC-Binding Factor	
<i>FBXW7</i>	chr4	F-Box And WD Repeat Domain Containing 7	
<i>FOXA2</i>	chr20	Forkhead Box A2	
<i>KLLN</i>	chr10	Killin, P53 Regulated ADN Replication Inhibitor	
<i>PGAP3</i>	chr17	Post-GPI Attachment To Proteins 3	
<i>PRPF18</i>	chr10	Pre-MARN Processing Factor 18	
<i>RPL22</i>	chr1	Ribosomal Protein L22	
<i>SCAARN11</i>	chr12	Small Cajal Body-Specific ARN 11	
<i>SIN3A</i>	chr15	SIN3 Transcription Regulator Family Member A	
<i>SPOP</i>	chr17	Speckle Type BTB/POZ Protein	
<i>TSPYL2</i>	chrX	TSPY Like 2	
<i>USP36</i>	chr17	Ubiquitin Specific Peptidase 36	

B. PREPARACIÓ I OBTENCIÓ DE LLIBRERIES

Les llibreries es van crear a partir de 100 ng d'ADN genòmic de les mostres tumorals i no tumorals utilitzant el Kit *TruSeq Custom Amplicon* (TSCA) (Illumina) seguint les instruccions detallades pel fabricant.

Per cada *run* de seqüència es va produir una llibreria comú que contenia la combinació de les llibreries individuals de 16 mostres. Amb el kit TSCA la creació de la llibreria es basa en el sistema d'amplicons (figura 37), on l'ADN genòmic de cadascuna de les mostres s'hibrida amb el *pool* customitzat d'oligonucleòtids (CAT *probes*) específics per poder cobrir els gens seleccionats. Després de la eliminació dels oligonucleòtids no hibridats, es produeix un procés d'extensió-lligació que permet connectar els oligonucleòtids hibridats als extrems 3' i 5' complementaris a la regió d'interès. Una ADN polimerasa sintetitza tot el fragment des de l'oligonucleòtid unit a l'extrem 3' fins a l'unit al 5', i seguidament es produeix un lligament a l'extrem de 5' amb l'ús d'una lligasa d'ADN. D'aquesta manera s'obtenen fragments que contenen les regions d'interès flanquejades per seqüències necessàries per a la posterior amplificació. En l'amplificació els productes resultants de l'extensió-lligació s'amplifiquen amb l'ús de *primers* que, a més a més, afegiran als fragments anteriors les seqüències índex (i5, i7), necessàries per marcar cadascuna de les mostres, i els adaptadors (P5 i P7) requerits per la generació de *clusters* en l'equip MiSeq. El següent pas és purificar els productes resultants de l'amplificació de la resta de components, i fer una comprovació amb gel d'agarosa per assegurar que la llibreria ha estat amplificada amb èxit. A continuació s'ha de quantificar cadascuna de les llibreries amb el kit Qubit dsADN HS (Invitrogen, Thermo Fisher Scientific) en el fluoròmetre Qubit 1.0 (Invitrogen, Life Technologies) per tal de poder realitzar la normalització de cadascuna d'elles a 10nM i finalment fer la mescla de totes les llibreries (*pooling*) amb concentracions equitatives de cadascuna.

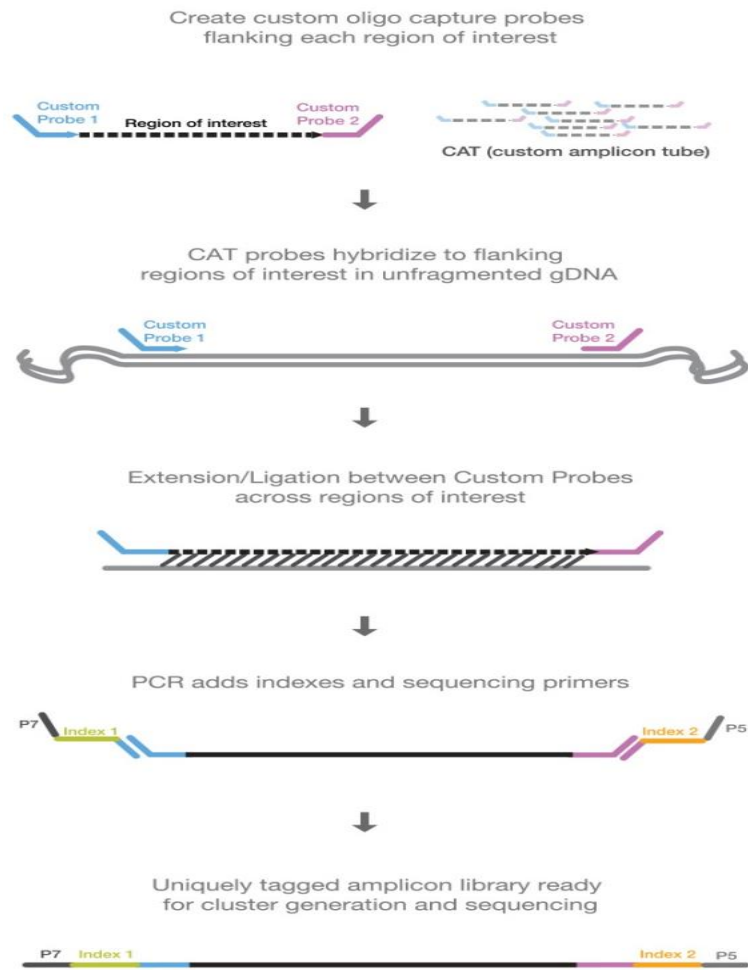


Figura 37. Flux de treball per la preparació de la llibreria TSCA (Font: <https://emea.illumina.com/>)

C. SEQÜENCIACIÓ

La seqüenciació de la llibreria comú (*pooled library*) es va realitzar en un equip MiSeq (Illumina) amb l'ús del kit Miseq *Reagent v3 600 cycles* (Illumina).

D. ANÀLISI BIOINFORMÀTICA I IDENTIFICACIÓ DE VARIANTS SOMÀTIQUES

L'anàlisi bioinformàtica es va dur a terme amb la col·laboració de l'empresa Sistemas Genómicos S.L. (Espanya). L'anàlisi va incloure els següents punts: l'anàlisi de qualitat de les seqüències, l'alineament de les lectures contra el genoma humà de referència (en el nostre cas el GRCh38/hg38), el control de qualitat dels mapatges, l'anàlisi de les profunditats de lectura (*read depth*), la detecció de variants, l'anotació, i la validació i interpretació dels resultats.

Amb la detecció de variants obtenim tot un llistat de variants, és a dir, nucleòtids que són diferents al del genoma de referència, que han passat els filtres de qualitat primaris. L'anotació ens recopila i dona, per cadascuna de les variants, tota la informació existent a les bases de dades públiques per tal de poder classificar les variants. A l'annex I es mostren tots els camps de l'anotació utilitzats en el nostre estudi. Per poder fer una classificació de variants més exhaustiva i adequada als nostres objectius, vam procedir a realitzar un filtrat específic de les variants prèviament anotades. Per aconseguir-ho es va fer un primer filtrat de variants analitzant de manera aparellada la mostra tumoral i no tumoral, per tal de descartar els polimorfismes i les variants germinals, i consegüentment, quedar-nos només amb les variants potencialment somàtiques. Per això vam seleccionar aquelles variants de la mostra tumoral que no estiguessin presents a la mostra no tumoral i tinguessin una freqüència al·lèlica (*variant allelic frequency* (VAF)) superior a 0,15 i una profunditat de lectura superior a 100X. A posteriori les variants somàtiques seleccionades van ser altre cop filtrades en funció de l'efecte de la variant sobre la isoforma canònica del transcrit, exclouent les que afectaven a les isoformes no canòniques i les variants que produïen canvis sinònims, és a dir, no hi havia canvi de l'aminoàcid. Per restringir més l'anàlisi, les variants van ser seleccionades tenint en compte les freqüències poblacionals detallades a les bases de dades públiques de l'annex I, per tal de descartar variants conegudes amb una freqüència d'al·lel menor (*minor allele frequency* (MAF)) superior a l'1%.

Finalment, les variants somàtiques definitives les vam classificar, d'una banda, tenint en compte la guia de recomanació de la classificació de les variants de l'ACMG (*American College of Medical Genetics*)¹⁴⁸; i d'altra banda, tenint en compte l'efecte que produïen sobre el transcrit. Tenint en compte les recomanacions de l'ACMG en el nostre estudi vam categoritzar les variants en quatre

Capítol 3

grups en funció de la informació de la base de dades COSMIC (*Catalog of Somatic Mutations in Cancer*) i les dades de diferents predictors de patogenicitat com Polyphen2, Sift, Condel i Fathmm. El grup 1 contenia les variants patogèniques (*pathogenic variants (PV)*) que eren aquelles variants reportades prèviament a la base de dades COSMIC, associades amb la malaltia i avaluades com deletèries pels diferents predictors de patogenicitat; el 2 contenia les variants probablement patogèniques (*likely pathogenic variants (LPV)*) que eren variants reportades a COSMIC, amb alguna evidència d'associació amb la malaltia, però sense informació sobre la predicció de patogenicitat; el 3 incloïa les variants de significat incert (*variant of unknown significance (VUS)*) que eren les variants que no havien estat prèviament reportades ni com a patogèniques ni com a benignes, i tampoc es coneix amb certesa el grau de patogenicitat; i el 4 que incloïa les variants probablement benignes o neutres (*likely benign or neutral variants (LBNV)*) que eren les que havien estat anotades a COSMIC però avaluades com neutres o benignes segons els predictors de patogenicitat.

Tenint en compte l'efecte de la variant sobre el transcrit les vam classificar en 6 categories d'acord amb la categorització de les variants del projecte *Sequence Ontology* a través de la plataforma *Open Targets* ¹⁴⁹. Les 6 categories van ser les següents: *missense*, és una variant que causa el canvi d'una o més bases, donant lloc a una seqüència d'aminoàcids diferent però on es conserva la longitud; *frameshift*, és una variant que provoca una interrupció de la pauta de lectura de la traducció de la proteïna, perquè el nombre de nucleòtids inserits o eliminats no és un múltiple de tres; *stop gained*, és una variant mitjançant la qual es canvia almenys una base d'un codó, donant lloc a un codó stop prematur, donant lloc a una proteïna més curta; *inframe deletion*, és una variant causada per una deleció, donant lloc a una seqüència d'aminoàcids diferents però on es conserva la longitud de la proteïna; *splice acceptor*, és una variant on s'ha canviat la regió de 2 bases a l'extrem 3' d'un intró modificant la regió d'*splicing*; i *protein altering*, és una variant de la seqüència, produïda per una indel (inserció-delecció) que es preveu que canviï la proteïna codificada.

5 SEQÜENCIACIÓ SANGER

La seqüenciació Sanger es considera la tècnica de seqüenciació de primera generació, sent la que s'ha utilitzat des del 1977 i que encara es manté en vigència, tot i l'adveniment des del 2005 de la tecnologia de segona generació (NGS). Es basa en una primera ampliació de fragments concrets (normalment exons) dels gens d'interès mitjançant una PCR, i després d'uns passos de purificació d'aquestes regions amplificades, es tracta de determinar la seqüència específica de nucleòtids d'aquest fragment. La reacció de seqüenciació té com a fonament l'ús d'uns nucleòtids especials anomenats dideoxinucleòtids que el que fan és que un cop incorporats en el procés de síntesi d'un fragment d'ADN no es pugui continuar la síntesi i es produeixi una terminació de l'elongació de la cadena (*chain-termination*) del determinat fragment. En aquesta reacció es subministra una quantitat equilibrada de nucleòtids normals i de dideoxucleòtids. Després de la precipitació dels fragments resultants de la seqüenciació, es desnaturalitzen amb formamida i es carreguen en un analitzador genètic, per tal de fer una electroforesi capil·lar i que es puguin anar llegint de manera automatitzada. En el nostre cas per fer la reacció de seqüència hem utilitzat el kit *BigDye Terminator* v3.1 (Applied Biosystems, Thermo Fisher Scientific) i per córrer els fragments el 3130 *Genetic Analyzer* (Applied Biosystems).

En aquesta tesi la seqüenciació Sanger s'ha utilitzat, d'una banda per fer l'estudi mutacional del gen *POLE* de tots els casos utilitzats per a l'objectiu específic 1, de 4 casos de CSE de l'objectiu específic 2, i d'altra banda per validar variants detectades als gens *TP53*, *ARID1A*, *PTEN*, i *PIK3CA* en les anàlisis mutacionals mitjançant *targeted sequencing* en els objectius específics 1 i 3. Els primers utilitzats per l'amplificació dels exons específics de tots els gens descrits anteriorment es detallen a la taula 10.

Taula 10. Llistat dels *primers* utilitzats per amplificar diferents exons dels gens *POLE*, *TP53*, *ARID1A*, *PTEN* i *PIK3CA*.

	Forward 5'-3'	Reverse 5'-3'
<i>POLE</i>		
Exó 9	TGCTTATTTTGTCCCCACAG	TACTTCCCAGAAGCCACCTG
Exó 11	GCAGACCTCTGACTGCTGTG	CCTAAGTCGACATGGGAAGC
Exó 13	TCTGTTCTCATTCTCCTCCAG	CGGGATGTGGCTTACGTG
Exó 14	ACCCTGGGCTTTGATTTTT	CACCTCCATTGAGTCCAGT
<i>TP53</i>		
Exó 4	GTCCAGATGAAGCTCCCAGA	ACTGACCGTGCAAGTCACAG
Exó 5	CACCTGTGCCCTGACTTTCA	AACCAGCCCTGTCGTCTCT
Exó 6	GTGCAGCTGTGGGTTGATT	CTTAACCCCTCCTCCAGAG
Exó 7	TTGGGCCTGTGTTATCTCCT	GTGTGCAGGGTGGCAAGT
Exó 8	GCCTCTTGCTTCTTTTTCC	GCTTCTTGCTGCTTGCTT
Exó 10	ACTTCTCCCCCTCCTCTGTT	GAAGGCAGGATGAGAATGGA
<i>ARID1A</i>		
Exó 2	GGCCATCAAAGCTCAGGTTA	GGGTACCCATGTCCTTGCT
Exó 3	CCTCTCAGCCTCCATACTCC	GAGACTGGGGCTGAGGATAC
Exó 12	CCTCGACTCCTTTGGTTTGG	GATCTTGGGCTGGGACTTCT
Exó 18	TCTCCATTCTGCACTCTGG	TGTCTTTCATTGTGAGCCGC
Exó 20	CATGCAGAACCCACCCTTTG	TAAGTTCTCCACACACGCAC
<i>PTEN</i>		
Exó 2	TTTCAGATATTTCTTTCCTTA	AACAAGAATATAAAACATCAA
Exó 5	TTTTCTTATTCTGAGGTTATC	TAAATATGCACATATCATTAC
Exó 6	TTGGCTTCTCTTTTTTTCTG	ACATGGAAGGATGAGAATTC
Exó 7	CCTGTGAAATAAATACTGGTATG	CTCCCAATGAAAGTAAAGTACA
Exó 8	ACACATCACATACATAACAAGTC	GTGCAGATAATGACAAGGAATA
<i>PIK3CA</i>		
Exó 2	CCCCTCCATCAACTTCTTCAA	AGAAAGGGACAACAGTTAAGCT
Exó 9	GGGAAAATGACAAAGAACAGC	ACATGCTGAGATCAGCCAAAT
Exó 20	TGAGCAAGAGGCTTTGGAGT	GGTCTTTCCTGCTGAGAGT

6 IMMUNOHISTOQUÍMICA (IHQ)

La immunohistoquímica (IHQ) neix a partir de la immunofluorescència. És una tècnica que permet detectar in situ components cel·lulars i extracel·lulars (antígens), mitjançant anticossos específics utilitzant sistemes de detecció enzimàtics. La fixació i la posterior inclusió de les mostres en parafina permet transformar el material biològic semi-líquid en un de sòlid, homogeni i dur, i d'aquesta manera es poden realitzar talls fins que preserven l'estructura.

6.1 IHQ DE P53, PTEN, ARID1A, MSH6, I PMS2

En el cas de l'objectiu específic 1 es va utilitzar la IHQ per testar diferents marcadors moleculars (p53, PMS2, MSH6) per tal de poder classificar els CEE i els CSE segons el panell alternatiu del TCGA (*surrogate TCGA*) juntament amb la seqüenciació del gen *POLE* per Sanger. També es va utilitzar la IHQ per validar l'efecte de les variants detectades en els gens *TP53*, *PTEN* i *ARID1A* sobre l'expressió de la proteïna. En el cas de l'objectiu específic 2 es va utilitzar per testar el marcador molecular p53.

La IHQ es va realitzar sobre seccions dels blocs de parafina de 3 µm, després de ser estufades durant una hora a 65°C. A continuació, es va realitzar un pretractament de desparafinació i rehidratació del teixit. El següent pas va ser el desemascament antigènic, que permet recuperar i exposar antigens emascarats durant la preparació de la mostra. Així s'incuben les preparacions amb un tampó comercial que conté Tris/EDTA 50x a pH 9, en un mòdul de pretractament PT-LINK (Dako) durant 20 minuts a 95°C. Després es bloqueja la peroxidasa endògena de la mostra incubant els portaobjectes amb una solució d'H₂O₂ al 3% per reduir el soroll de fons que s'obté amb el cromogen emprat. Seguidament, es realitzen tres rentats amb PBS (*Phosphate-Buffered Saline*) i s'incuba la mostra amb l'anticòs primari durant 30 minuts a temperatura ambient. Posteriorment, es realitzen tres rentats més i s'incuba la mostra amb l'anticòs secundari. Per acabar s'incuba amb el cromogen diaminobenzimida (DAB, Dako-Agilent) per rebel·lar la reacció, i es realitza una contra-tinció amb hematoxilina per visualitzar els nuclis de les cèl·lules del teixit analitzat. Les mostres van ser testades pels següents marcadors moleculars: p53, PTEN, ARID1A, MSH6, i PMS2. Els anticossos utilitzats, clons, fonts i dilucions es mostren a la taula 11.

Taula 11. Relació dels anticossos utilitzats per la tècnica d'immunohistoquímica

ANTICOS	CLON	FONT	DILUCIÓ
p53	DO7	DAKO	*RTU
ARID1A	EPR13501-73	DAKO	1/500
PTEN	6H2.1	DAKO	1/100
MSH6	EP49	DAKO	*RTU
PMS2	EP51	DAKO	*RTU

*RTU: Ready to use

Els criteris per avaluar els 5 marcadors van ser els següents: p53 va ser avaluat amb patró mutat si hi havia una tinció forta en més del 80% de les cèl·lules tumorals, o si hi havia una tinció completament negativa, respecte l'avaluació amb patró wild type en la resta de casos; *PTEN*, ARID1A, MSH6 i PMS2 van ser avaluats com positius, si hi havia qualsevol grau de tinció, i negatiu si hi havia absència de tinció. D'aquesta manera cadascuna de les mostres va ser classificada molecularment d'acord amb el panell alternatiu al TCGA (seqüenciació *POLE*, i test per IHQ dels marcadors p53, MSH6 i PMS2). Així els casos van ser classificats en cinc categories:

- 1) els casos amb inestabilitat de microsatèl·lits o hipermutats (MSI) que inclouen els casos negatius per MSH6 i PMS2.
- 2) els casos tipus-serosos (*serous-like*) que inclouen els casos amb un patró mutat per p53.
- 3) els casos *POLE*-mutats, on també es van incloure els casos amb mutacions al gen *POLE* encara que fossin MSI o p53 mutat.
- 4) els casos amb un perfil mutacional no específic (NSMP, *No Specific Mutation Profile*) que inclouen casos que no s'han pogut classificar en cap dels grups anteriors.
- 5) els casos amb la presència de més d'un marcador alternatiu (MSI i p53) van ser classificats com doble mutats.

6.2 IHQ DE ERBB2

En el cas de l'objectiu específic 3, pel que fa als resultats d' MLPA, es van validar per IHQ i SISH (*Silver In Situ Hybridization*) aquells casos de CSE amb alteració al gen *ERBB2*. Tant la IHQ com el SISH es van realitzar sobre seccions de 2 µm de tres blocs de parafina diferents de cada cas de carcinoma serós d'endometri. Per determinar l'estat del gen *ERBB2* per IHQ es va utilitzar l'Hercep test (Dako-Agilent) seguint les instruccions del fabricant. La valoració per IHQ es va fer seguint les recomanacions ASCO (*American Society of Clinical Oncology*)/CAP (*College of American Pathologists*) 2018 per al càncer de mama.¹⁵⁰ Els casos *ERBB2* negatius van ser valorats per IHQ com 0 o 1+, i els casos *ERBB2* positius van ser valorats per IHQ com 3+, o 2+ amb amplificació gènica confirmada per SISH.

7 SILVER IN SITU HYBRIDIZATION (SISH)

La hibridació in situ és una tècnica que consisteix en hibridar directament sobre el teixit regions específiques de l'ADN amb un oligonucleòtid complementari (sonda). Aquest marcatge pot ser fluorescent (FISH), cromogènic (CISH) o en plata (SISH).

En el cas de l'objectiu específic 3 es va utilitzar la tècnica SISH per validar els resultats d'MLPA per al gen *ERBB2*. Es va utilitzar un sistema automatitzat de dos colors, on el gen *ERBB2* es tenyia amb negre amb el precipitat de plata, i el *CEP17* (centròmer 17) en vermell (Roche). Els passos del protocol del SISH inclouen una desparafinació, un tractament amb citrat, una incubació amb ISH proteasa, l'addició i hibridació de les sondes *ERBB2* i *CEP1*, la incubació amb plata i la incubació amb hematoxilina com a contrast. La valoració per SISH es va fer seguint les recomanacions ASCO (*American Society of Clinical Oncology*)/CAP (*College of American Pathologists*) 2018 per al càncer de mama.¹⁵⁰ Els casos *ERBB2* negatius van ser definits per un resultat IHQ de 0 o 1+, i els casos *ERBB2* positius van ser definits per un resultat d'IHQ de 3+, o 2+ amb amplificació gènica confirmada per SISH.

8 ANÀLISI BIOESTADÍSTICA

En l'objectiu específic 1 es van realitzar les següents anàlisis:

A) un estudi estadístic per determinar l'associació entre els gens alterats a nivell de variants somàtiques i el tipus de càncer d'endometri (CEE, CSE). Per a això, primer es van determinar les freqüències absolutes i relatives d'alteració per cada gen, i per cada subconjunt de mostres. Posteriorment, es van obtenir les *odds ratio* i els p-valors corresponents per determinar la força i la significació de l'associació. El llindar del valor de significació va ser establert en un p-valor=0,05.

B) una anàlisi de doble *cluster* jeràrquic amb distància euclídea per veure l'agrupació automàtica de les mostres en funció de la similitud de les alteracions de cadascuna d'elles. Per representar aquesta agrupació i determinar la capacitat del perfil genètic en discriminar entre tipus histològic es van construir dos *heatmaps*. El *heatmap* 1 que incloïa tots els gens alterats, i el 2 que incloïa només els gens alterats estadísticament significatius.

Per a aquestes dos anàlisis es va utilitzar el software estadístic R (www.r-project.org) i es va establir el llindar del valor de significació com p-valor=0,05.

En l'objectiu específic 2, es van realitzar les mateixes anàlisis que en l'objectiu específic 1 per determinar l'associació entre les ASNC dels gens i el tipus de càncer d'endometri (CEE, CSE), l'agrupació automàtica de les mostres en funció de les similituds de les alteracions de cadascuna d'elles, i la capacitat discriminativa del perfil de gens. En aquest cas l'agrupació es va representar construint un total de 6 *heatmaps*: el *heatmap* 1 que incloïa tots els gens alterats, el 2 tots els gens estadísticament significatius, el 3 els gens detectats amb guanys, el 4 els gens estadísticament significatius que presentaven guanys, el 5 tots els gens que presentaven pèrdua, i el 6 els gens significatius amb pèrdua.

En l'objectiu específic 3, es va realitzar una anàlisi estadística descriptiva calculant les freqüències globals d'alteració, i les freqüències d'heterogeneïtat per cada gen i pel total de mostres alterades, tant pel que fa a l'alteració en el número de còpies com a les variants somàtiques. A més a més, es van establir els criteris d'heterogeneïtat per definir quan un determinat gen i mostra es presentaven en heterogeneïtat intratumoral, tant a nivell d'alteració en el número de còpies com de variants somàtiques.

A continuació es detallen aquests criteris:

1) Establiment del criteri d'heterogeneïtat intratumoral en l'anàlisi, mitjançant MLPA, de l'alteració en el número de còpies

Per establir els criteris descrits a continuació ens vam basar en el treball de Dopierala et al ¹⁵¹. Per determinar quan un gen presentava heterogeneïtat intratumoral es van calcular tres paràmetres en base als valors DQ de les tres regions tumorals analitzades per gen i per mostra. Els tres paràmetres (taula 12) van ser a) el ADQ (*average dosage quotient*) que és la mitjana dels tres valors de DQ, b) el SDDQ (*standard deviation dosage quotient*) que és la desviació estàndard dels tres valors de DQ, i c) el DDQ (*difference between ADQ+SDDQ i ADQ-SDDQ*) que és la diferència entre ADQ+SDDQ i ADQ-SDDQ. L'anàlisi d'aquests paràmetres en gens amb dotació diploide ens va permetre determinar la seva variació estadística. Conseqüentment, un gen va ser considerat significativament heterogeni si es complien totes les condicions següents: (1) que el SDDQ fos superior a 0,15, (2) que el DDQ fos superior a 0,3, i (3) que els valors DQ de les tres regions tumorals pertanyessin a diferents categories (pèrdua, normal, guany).

Finalment, per considerar que un cas presentava heterogeneïtat intratumoral s'havien de comptabilitzar més de tres gens en heterogeneïtat.

Taula 12. Fórmules per calcular els tres paràmetres necessaris per determinar la presència d'heterogeneïtat tumoral en un gen.

MOSTRA 1						
GEN	1	2	3	ADQ	SDDQ	DDQ
A	DQ A 11	DQ A 12	DQ A 13	ADQA= (DQA11+DQA12+DQA13)/ 3	SDDQA= $\sqrt{((DQA11-ADQA)^2+(DQA12-ADQA)^2+(DQA13-ADQA)^2)/2}$	DDQA= (ADQA+SDDQA)- (ADQA-SDDQA)
...
N	DQ N 11	DQ N 12	DQ N 13	ADQN= (DQN11+DQN12+DQN13)/ 3	SDDQN= $\sqrt{((DQN11-ADQN)^2+(DQN12-ADQN)^2+(DQN13-ADQN)^2)/2}$	DDQN= (ADQN+SDDQN) - (ADQN-SDDQN)

2) Establiment del criteri d'heterogeneïtat intratumoral en l'anàlisi, mitjançant *targeted sequencing*, de les variants somàtiques

En aquest cas es va determinar la presència o absència d'una determinada variant somàtica en cadascuna de les tres porcions de tumor. Si no hi havia concordança quant a la variant entre les tres porcions es va considerar que aquesta es trobava en heterogeneïtat en el determinat cas.

CAPÍTOL 4

RESULTATS

En aquest capítol es presenten els resultats d'aquesta tesi doctoral. S'han estructurat en 3 apartats que corresponen als 3 objectius específics de la tesi. Tot i que no es tracta d'una tesi amb format d'articles científics, els resultats corresponents als objectius específics 1 i 3 han estat publicats en revistes indexades dintre el JCR-SCI (figura 38). Aquests articles es presenten a l'annex II. Moltes de les taules i figures utilitzades en aquest capítol han estat extretes d'aquests treballs, i per tant, es presenten en el seu format original.

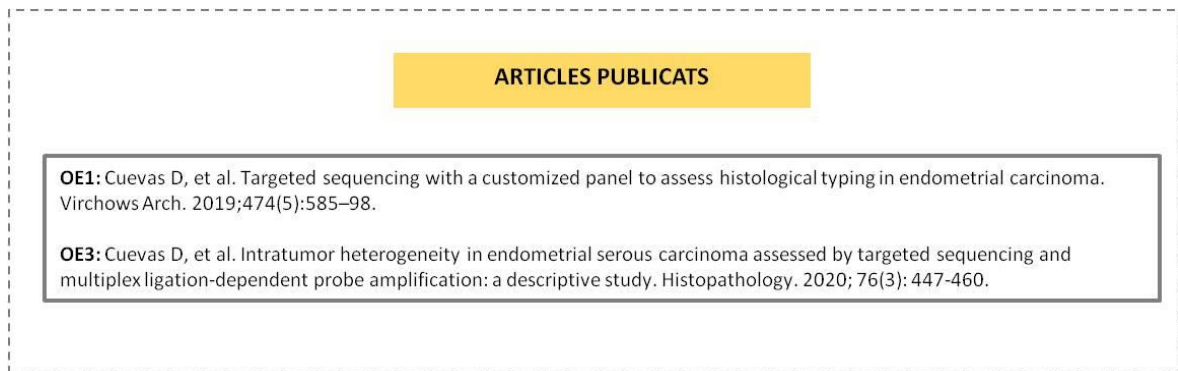


Figura 38. Articles publicats.

1 ALTERACIONS GENÈTIQUES A NIVELL DE VARIANTS SOMÀTIQUES

En aquest apartat es detallen els resultats obtinguts en el marc de l'objectiu específic 1: analitzar les alteracions de diferents gens a nivell de variants somàtiques (VS) mitjançant una anàlisi de *targeted sequencing* amb la tecnologia NGS per tal d'establir un perfil de gens que permetin classificar el càncer d'endometri. Aquestes alteracions s'han determinat en mostres pertanyents als dos tipus més freqüents de càncer d'endometri (CEE i CSE).

1.1 MÈTRIQUES I ESTADÍSTIQUES DE L'ANÀLISI DE *TARGETED SEQUENCING* (NGS)

Durant el procés de seqüenciació es van anar avaluant tots els paràmetres de qualitat per tal d'assegurar-nos una correcta identificació de les variants somàtiques. Així es van obtenir valors adequats de la densitat de *clusters*, de profunditat de lectura (*read depth*), entre molts altres. Les figures 39, 40 i 41 mostren l'exemple de les dades de qualitat d'una determinada mostra, com són la distribució de la mitjana de les qualitats de les bases (*base calling*) en funció de la posició dintre de les lectures, la distribució de la mitjana del percentatge de cadascun dels nucleòtids per cada posició dintre de les lectures, i la distribució de la mitjana de les qualitats sobre totes les seqüències. Per totes les mostres es van obtenir unes dades de qualitat molt similars.

A la figura 39 s'observa que les lectures que tenen entre 1 i 200 parells de bases tenen uns adequats valors de qualitat (Q superior a 30), i és a partir de les lectures de més de 200 parells de bases que es comença a perdre aquesta qualitat (Q inferior a 30). Aquesta pèrdua de qualitat es inherent al a la pròpia tècnica de seqüenciació. Aquests valors ens demostren que hem tingut una seqüenciació adequada per continuar amb les següents anàlisis.

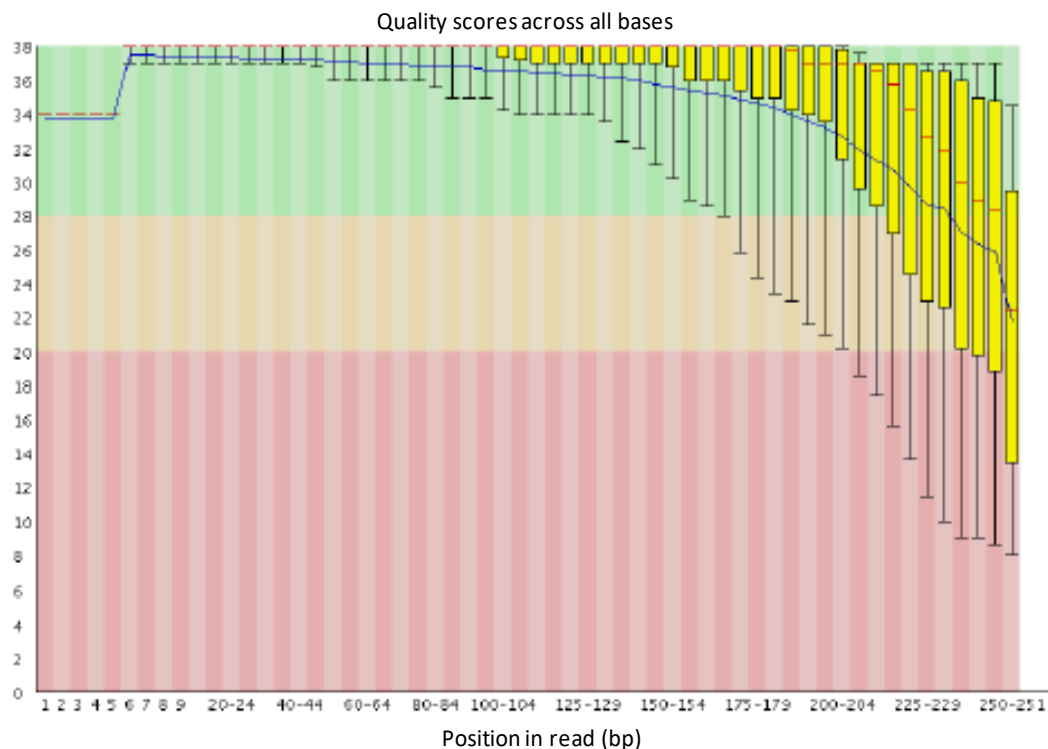


Figura 39. Puntuacions de qualitat (Q, *Phred Score*) (eix y) all llarg de totes les posicions dels fragments seqüenciats (eix x).

Capítol 4

A la figura 40 s'observa una distribució equitativa dels nucleòtids i homogeneïtat d'aquesta distribució al llarg de totes les posicions de les lectures. Això, és el que s'espera en un procés de seqüenciació adequat.

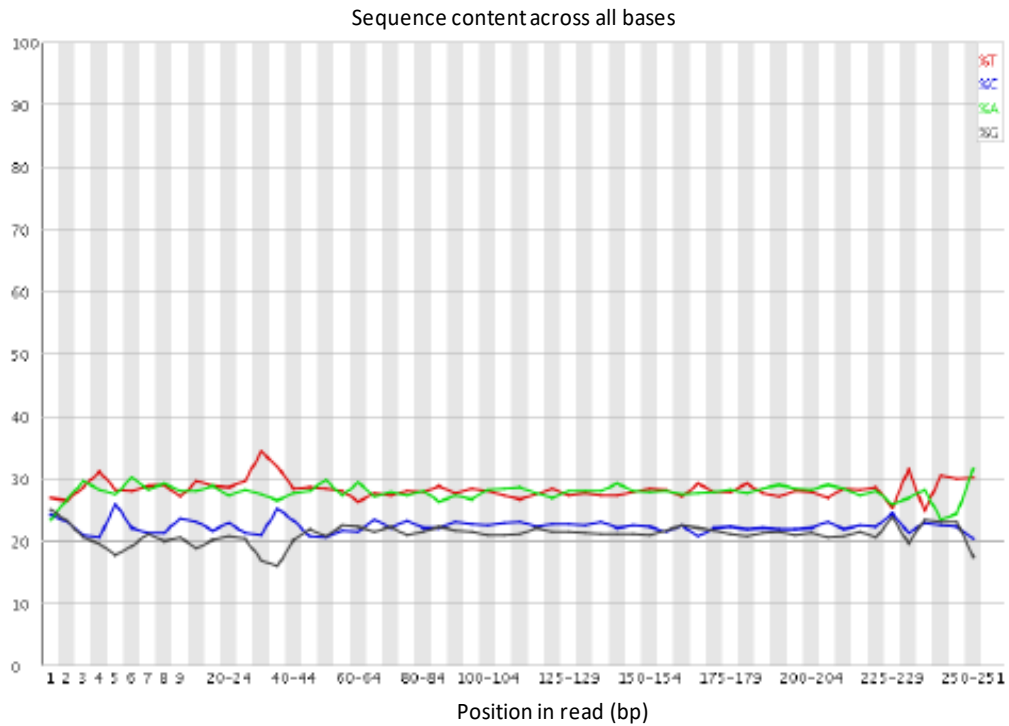


Figura 40. Contingut dels 4 nucleòtids (A, T, C, G) en percentatge (eix y) al llarg de totes les posicions (pb, *base pair*) dels fragments seqüenciats (eix x).

A la figura 41 s'observa que a partir de les 50000 lectures ja s'obtenen uns valors de qualitat mitjans superiors al Q= 30. Això indica que un alt percentatge de les nostres lectures tenen una elevada qualitat, i que no s'observa cap rang de lectures amb *scores* de qualitat baixos.

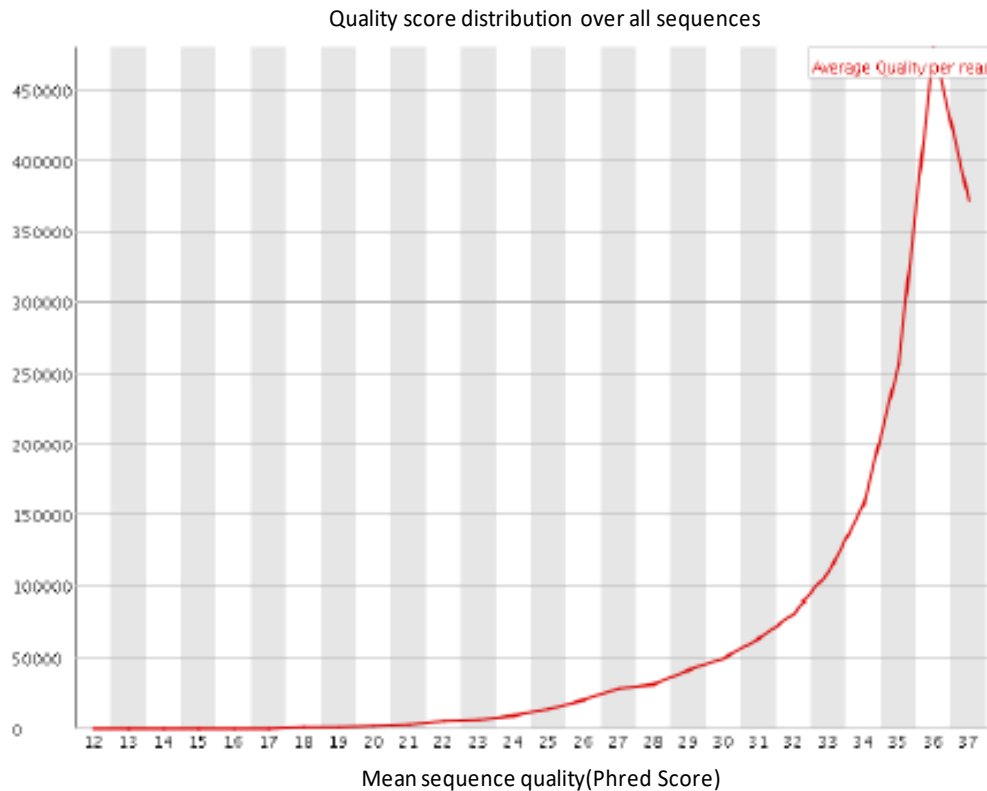


Figura 41. Distribució de les puntuacions de qualitat de totes les lectures generades.

Quant a les mètriques obtingudes de l'anàlisi de totes les mostres, es va obtenir per mostra una mitjana de 2270521 lectures (*total reads*) i de 1563X de profunditat de lectura (*read depth*). La cobertura (*coverage*) mitjana de les lectures sobre el genoma de referència GRCh38 / hg38 va ser del 97,64%. Cal dir que tant els valors de nombre de lectures com de profunditat de lectura van ser molt similars entre les mostres tumorals i no tumorals. A la taula 13, es detallen els valors d'aquestes mètriques per cadascuna de les mostres analitzades.

Taula 13. Detall dels valors de número total de lectures, de la mitjana de profunditat de lectura i del percentatge mitjà de cobertura per cadascuna de les mostres analitzades.

SAMPLE	TOTAL READS	MEAN READ DEPTH	% HIGH QUALITY MAPPED READS
1-NT	3487474	2862,27	98,31
1-T	3407288	2774,02	97,87
2-NT	3550442	2902,57	98,15
2-T	3265326	2654,93	97,92
3-NT	597512	474,68	95,89
3-T	666952	532,58	96,17
4-NT	4003316	3271,38	98,04
4-T	3421968	2794,06	97,94
5-NT	3139500	2556,25	97,85
5-T	3266798	2666,54	98,06
6-NT	595286	475,53	96,28
6-T	579432	457,94	95,36
7-NT	3137584	2553,74	97,86
7-T	3335370	2716,47	97,94
8-NT	730934	585,33	96,61
8-T	602492	478,97	96,13
9-NT	685580	548,91	96,4
9-T	920658	739,85	96,66
10-NT	3462116	2825,75	98,1
10-T	2552294	2081,3	98,12
11-NT	471706	378,99	96,65
11-T	498354	400,36	96,77
12-NT	545164	438,83	96,78
12-T	487756	389,22	96,26
13-NT	3465816	2822,89	97,98
13-T	3169558	2582,17	97,93
14-NT	3020668	2461,82	97,96
14-T	3494936	2849,24	97,96
15-NT	482576	388,63	96,89
15-T	607962	489,72	96,93
16-NT	492558	395,83	96,73
16-T	419832	336,54	96,6
17-NT	2712498	1426,73	98,97
17-T	2659436	1408,4	99,26
18-NT	2231814	1119,64	97,12
18-T	2355950	1187	97,19
19-NT	1662798	852,21	98,41
19-T	1699924	867,51	98,05
20-NT	2148976	1093,73	98,09
20-T	1618364	831,69	98,52
21-NT	4962612	2609,67	98,83
21-T	3151148	1654,86	98,98
22-NT	2456272	1294,76	99,14
22-T	3191244	1677,26	98,89
23-NT	6906944	3642,52	99,11
23-T	4894570	2580,91	99,16
24-NT	2050068	1048,45	98,45
24-T	1717170	871,85	97,63

NT: no tumoral; T: tumoral

1.2 DESCRIPCIÓ DE LES VARIANTS SOMÀTIQUES I GENS ALTERATS

Les variants somàtiques es van identificar i classificar seguint els criteris descrits a l'apartat de materials i mètodes. En tot el conjunt de mostres analitzades (n=24) es van identificar un total de 164 variants somàtiques, observant una mitjana de 7 variants per mostra (rang de 1 a 23). A la taula 14 es descriuen totes les variants detectades en relació a cada mostra (*sample_id*) i es detalla de cadascuna d'elles el gen implicat (*gene symbol*), la freqüència al·lèlica (*tumor vaf*), la profunditat de lectura (*tumor depth*), la nomenclatura de la variant a nivell genòmic (*HGVSc*), la nomenclatura de la variant a nivell de la proteïna (*HGVSp*), l'efecte de la variant sobre el transcrit (*variant effect*), i si ha estat reportada o no a la base de dades COSMIC (*COSMIC data base*).

Taula 14. Descripció de totes les variants somàtiques detectades per cadascuna de les mostres analitzades

SAMPLE ID	GENE SYMBOL	TUMOR VAF	TUMOR DEPTH	HGVS_c_NAME	HGVS_p_NAME	VARIANT EFFECT	COSMIC DATABASE
1	ARID1A	0.435	3710	c.3216delA	p.Lys1072AsnfsTer21	frameshift	YES
	ARID1A	0.451	1114	c.6707G>C	p.Arg2236Pro	missense	YES
	BCOR	0.443	451	c.4376A>G	p.Asn1459Ser	missense	YES
	EP300	0.379	3919	c.2746T>A	p.Ser916Thr	missense	YES
	FOXA2	0.472	397	c.643C>T	p.His215Tyr	missense	YES
	KMT2B	0.267	498	c.1250T>C	p.Leu417Pro	missense	YES
	MKI67	0.213	7670	c.5455C>A	p.Leu1819Ile	missense	NO
	PIK3CA	0.263	4748	c.3140A>G	p.His1047Arg	missense	YES
	PTEN	0.881	2652	c.994A>T	p.Lys332Ter	stop_gained	YES
RPL22	0.3	1790	c.44delA	p.Lys15ArgfsTer5	frameshift	YES	
2	KRAS	0.412	3013	c.37G>T	p.Gly13Cys	missense	YES
	MKI67	0.319	3839	c.6725G>A	p.Ser2242Asn	missense	YES
	PPP2R1A	0.432	4929	c.536C>G	p.Pro179Arg	missense	YES
	PTEN	0.688	1343	c.929A>G	p.Asp310Gly	missense	YES
	TP53	0.8	1113	c.524G>A	p.Arg175His	missense	YES
3	ATR	0.264	197	c.2320delA	p.Ile774TyrfsTer5	frameshift	YES
	BCOR	0.229	262	c.248G>A	p.Arg83Gln	missense	NO
	FBXW7	0.287	1302	c.1513C>T	p.Arg505Cys	missense	YES
	PTEN	0.243	551	c.491dupA	p.Val166SerfsTer14	frameshift	YES
	PTEN	0.314	551	c.697C>T	p.Arg233Ter	stop_gained	YES
	RPL22	0.265	310	c.44delA	p.Lys15ArgfsTer5	frameshift	YES
4	CHD4	0.418	2165	c.3328G>A	p.Asp1110Asn	missense	NO
	CTCF	0.383	2358	c.859delA	p.Ser287AlafsTer23	frameshift	NO
	FGFR2	0.385	2946	c.755C>G	p.Ser252Trp	missense	YES
	PIK3RI	0.32	2586	c.2105C>T	p.Ser702Phe	missense	YES
	PTEN	0.382	2863	c.388C>T	p.Arg130Ter	stop_gained	YES
5	CTNNB1	0.414	762	c.134C>T	p.Ser45Phe	missense	YES
	PIK3CA	0.455	582	c.1636C>A	p.Gln546Lys	missense	YES
	PTEN	0.881	2659	c.388C>G	p.Arg130Gly	missense	YES
6	ARID1A	0.859	497	c.3281delA	p.Lys1094SerfsTer67	frameshift	NO
	BCOR	0.427	626	c.695C>T	p.Pro232Leu	missense	NO
	EP300	0.415	585	c.1701delT	p.Arg568GlyfsTer33	frameshift	NO
	FGFR2	0.388	1260	c.208A>C	p.Ile70Leu	missense	NO
	KMT2B	0.361	1197	c.1067delA	p.Lys356SerfsTer31	frameshift	NO
	PIK3CA	0.4	370	c.1634A>C	p.Glu545Ala	missense	YES
	PTEN	0.549	488	c.376G>A	p.Ala126Thr	missense	YES
	PTEN	0.291	729	c.800delA	p.Lys267ArgfsTer9	frameshift	YES
	RPL22	0.881	303	c.44delA	p.Lys15ArgfsTer5	frameshift	YES
	SIN3A	0.342	237	c.2006T>C	p.Leu669Pro	missense	YES
	TP53	0.401	519	c.743G>A	p.Arg248Gln	missense	YES
TP53	0.395	401	c.455C>T	p.Pro152Leu	missense	YES	
7	ARID1A	0.163	192	c.4910G>A	p.Arg1637Gln	missense	YES
	COL11A1	0.168	502	c.3893C>T	p.Pro1298Leu	missense	NO
	KMT2B	0.259	536	c.1250T>C	p.Leu417Pro	missense	YES
	PPP2R1A	0.188	320	c.685G>A	p.Val229Met	missense	YES
	SIN3A	0.159	472	c.2444T>C	p.Ile815Thr	missense	NO
	SMARCA4	0.152	2494	c.28G>A	p.Gly10Arg	missense	YES
	TP53	0.151	2175	c.523C>T	p.Arg175Cys	missense	YES
8	ARID1A	0.208	308	c.1650dupC	p.Tyr551LeufsTer72	frameshift	YES
	CTNNB1	0.182	176	c.121A>G	p.Thr41Ala	missense	YES
	FGFR2	0.212	578	c.755C>T	p.Ser252Leu	missense	YES
	FOXA2	0.269	104	c.646T>G	p.Ser216Ala	missense	NO
	KMT2B	0.214	482	c.3057delA	p.Gly1020AlafsTer4	frameshift	YES
	MKI67	0.223	467	c.9414T>A	p.Asp3138Glu	missense	NO
	PIK3RI	0.279	1111	c.1727_1729del3pb	p.Thr576del	inframe_deletion	YES
	PTEN	0.152	184	c.955_958del4pb	p.Thr319Ter	frameshift	YES

Taula 14. Continuació

SAMPLE ID	GENE SYMBOL	TUMOR VAF	TUMOR DEPTH	HGVSc_NAME	HGVSp_NAME	VARIANT EFFECT	COSMIC DATABASE
9	ARID1A	0.416	974	c.5548delG	p.Asp1850ThrfsTer33	frameshift	YES
	ARID5B	0.444	895	c.1336_1363del28pb	p.His446LysfsTer24	frameshift	YES
	ATR	0.288	1429	c.6932_6934del3pb	p.Lys2311del	inframe_deletion	YES
	CHD4	0.43	734	c.3313C>T	p.Arg1105Trp	missense	YES
	COL11A1	0.328	546	c.5273C>T	p.Ala1758Val	missense	YES
	EP300	0.212	199	c.2992T>C	p.Ser998Pro	missense	NO
	FBXW7	0.375	1590	c.1513C>G	p.Arg505Gly	missense	YES
	MKI67	0.429	1761	c.3161C>G	p.Thr1054Arg	missense	NO
	NRAS	0.407	481	c.182A>G	p.Gln61Arg	missense	YES
	PPP2R1A	0.398	869	c.1586C>T	p.Ala529Val	missense	NO
	PTEN	0.422	638	c.388C>T	p.Arg130Ter	stop_gained	YES
	PTEN	0.434	671	c.518G>A	p.Arg173His	missense	YES
TP53	0.406	1038	c.731G>A	p.Gly244Asp	missense	YES	
USP36	0.505	191	c.2524G>A	p.Gly842Arg	missense	YES	
10	ARID1A	0.22	1846	c.1897C>T	p.Gln633Ter	stop_gained	YES
	ATR	0.313	1258	c.3402delT	p.Phe1134LeufsTer6	frameshift	YES
	CTCF	0.251	3445	c.1369C>T	p.Arg457Ter	stop_gained	YES
	NRAS	0.232	3543	c.35G>C	p.Gly12Ala	missense	YES
	PTEN	0.24	3101	c.133delG	p.Val45TyrfsTer9	frameshift	YES
	PTEN	0.198	3664	c.775delC	p.His259ThrfsTer7	frameshift	YES
	SIN3A	0.234	687	c.3157_3165del9pb	p.Lys1053_Glu1055del	inframe_deletion	NO
11	ARID1A	0.326	466	c.3281A>G	p.Lys1094Arg	missense	YES
	ARID1A	0.318	601	c.5548dupG	p.Asp1850GlyfsTer4	frameshift	YES
	ATR	0.313	1258	c.3402delT	p.Phe1134LeufsTer6	frameshift	YES
	CDH19	0.177	588	c.1974G>T	p.Glu658Asp	missense	NO
	CHD4	0.33	197	c.2998C>T	p.Arg1000Ter	stop_gained	YES
	CHD4	0.296	1185	c.2543T>C	p.Leu848Pro	missense	NO
	COL11A1	0.362	338	c.2063G>A	p.Gly688Glu	missense	NO
	CSMD3	0.261	406	c.6959G>A	p.Gly2320Asp	missense	YES
	CTNNB1	0.16	346	c.931A>T	p.Ser311Cys	missense	NO
	EP300	0.354	952	c.4195G>T	p.Asp1399Tyr	missense	YES
	KMT2B	0.388	134	c.8135G>A	p.Arg2712Gln	missense	YES
	MKI67	0.174	1871	c.4792C>T	p.Arg1598Ter	stop_gained	YES
	PIK3CA	0.249	672	c.113G>A	p.Arg38His	missense	YES
	PIK3CA	0.336	290	c.3073A>G	p.Thr1025Ala	missense	YES
	PPP2R1A	0.336	652	c.548G>A	p.Arg183Gln	missense	YES
	PRPF18	0.279	524	c.755A>G	p.Asp252Gly	missense	NO
	PTEN	0.287	342	c.469dupG	p.Glu157GlyfsTer23	frameshift	YES
	PTEN	0.252	155	c.968dupA	p.Asn323LysfsTer2	frameshift	YES
SIN3A	0.279	161	c.384A>G	p.Asp129Gly	missense	NO	
SMARCA4	0.188	580	c.749G>T	p.Ser250Ile	missense	YES	
TAF1	0.315	1011	c.3605G>A	p.Arg1202His	missense	YES	
TP53	0.218	883	c.400T>C	p.Phe134Leu	missense	YES	
USP36	0.313	264	c.1264C>T	p.Arg422Ter	stop_gained	YES	
12	ARID1A	0.292	137	c.1207C>T	p.Gln403Ter	stop_gained	YES
	CTCF	0.592	198	c.1061C>T	p.Ser354Phe	missense	YES
	PIK3CA	0.347	248	c.2521A>G	p.Ile841Val	missense	YES
	PIK3R1	0.311	713	c.1600C>G	p.Arg534Gly	missense	YES
	PIK3R1	0.336	917	c.1723_1731del9pb	p.Lys575_Arg577del	inframe_deletion	YES
	PTEN	0.319	376	c.388C>T	p.Arg130Ter	stop_gained	YES
	PTEN	0.374	374	c.389G>A	p.Arg130Gln	missense	YES
13	ARID1A	0.29	3745	c.3281dupA	p.Gln1095AlafsTer10	frameshift	YES
	CHD4	0.292	3888	c.417_419del3pb	p.Glu139del	inframe_deletion	YES
	FGFR2	0.302	2584	c.755C>G	p.Ser252Trp	missense	YES
	KMT2B	0.305	1548	c.521delC	p.Pro174GlnfsTer20	frameshift	YES
	PIK3CA	0.279	2684	c.1030G>A	p.Val344Met	missense	YES
	PIK3R1	0.26	1595	c.1701A>C	p.Lys567Asn	missense	NO
	PIK3R1	0.278	2105	c.1943T>A	p.Val648Asp	missense	NO
	PTEN	0.54	3216	c.723dupT	p.Glu242Ter	frameshift	YES
RPL22	0.338	2470	c.44delA	p.Lys15ArgfsTer5	frameshift	YES	

Taula 14 . Continuació

SAMPLE ID	GENE SYMBOL	TUMOR VAF	TUMOR DEPTH	HGVSc_NAME	HGVSp_NAME	VARIANT EFFECT	COSMIC DATABASE
14	ARID1A	0.8	103	c.4891C>T	p.Gln1631Ter	stop_gained	YES
	ARID5B	0.154	6395	c.1489dupA	p.Ile497AsnfsTer31	frameshift	NO
	CTCF	0.387	1245	c.1022G>C	p.Arg341Pro	missense	NO
	CTNNB1	0.317	894	c.133T>C	p.Ser45Pro	missense	YES
	KMT2B	0.21	557	c.1250T>C	p.Leu417Pro	missense	YES
	PIK3CA	0.333	160	c.2675C>T	p.Ala892Val	missense	NO
	PIK3CA	0.333	160	c.2695C>T	p.Arg899Cys	missense	NO
	PIK3R1	0.206	6487	c.1708_1709insGAT	p.Leu570delinsArgPhe	protein_altering	YES
PTEN	0.415	2728	c.491delA	p.Lys164ArgfsTer3	frameshift	YES	
15	ARID1A	0.162	370	c.304A>G	p.Asn102Asp	missense	NO
	ARID1A	0.174	230	c.1120C>G	p.Leu374Val	missense	NO
	KRAS	0.17	693	c.35G>T	p.Gly12Val	missense	YES
	PTEN	0.257	420	c.724G>T	p.Glu242Ter	stop_gained	YES
16	ATR	0.25	238	c.7501A>G	p.Lys2501Glu	missense	NO
	CHD4	0.247	526	c.2629C>T	p.Arg877Trp	missense	YES
	CTCF	0.162	174	c.1186delA	p.Arg396GlyfsTer3	frameshift	NO
	MAP3K4	0.224	286	c.479delA	p.Asn160MetfsTer8	frameshift	YES
	MKI67	0.156	276	c.1928G>T	p.Cys643Phe	missense	NO
	PIK3CA	0.27	244	c.3062A>G	p.Tyr1021Cys	missense	YES
	PIK3R1	0.197	375	c.1924C>T	p.Arg642Ter	stop_gained	YES
	PTEN	0.209	498	c.97_99del3pb	p.Ile33del	inframe_deletion	YES
	PTEN	0.26	146	c.955_958del4pb	p.Thr319Ter	frameshift	YES
	RPL22	0.213	207	c.44delA	p.Lys15ArgfsTer5	frameshift	YES
	TAF1	0.207	242	c.2591G>A	p.Arg864Gln	missense	YES
TP53	0.15	746	c.412G>A	p.Ala138Thr	missense	YES	
17	PIK3CA	0.95	1985	c.344G>C	p.Arg115Pro	missense	YES
	TP53	0.889	1658	c.586C>T	p.Arg196Ter	stop_gained	YES
18	CCND1	0.903	648	c.686delT	p.Leu229ProfsTer8	frameshift	NO
	MAP3K4	0.863	1397	c.2207delT	p.Ile736LysfsTer24	frameshift	NO
	PIK3CA	0.531	1811	c.316G>C	p.Gly106Arg	missense	YES
	PIK3CA	0.53	1810	c.333G>C	p.Lys111Asn	missense	YES
	TP53	0.726	782	c.821T>C	p.Val274Ala	missense	YES
19	PPP2R1A	0.251	1633	c.536C>G	p.Pro179Arg	missense	YES
	TP53	0.31	300	c.1024_1034del11pb	p.Arg342Ter	frameshift	YES
20	CDH19	0.227	387	c.1820_1821delTC	p.Ile607AsnfsTer22	frameshift	NO
	PIK3CA	0.233	2246	c.263G>A	p.Arg88Gln	missense	YES
	PPP2R1A	0.517	1601	c.536C>G	p.Pro179Arg	missense	YES
	TAF1	0.882	537	c.408dupC	p.Ser137LeufsTer4	frameshift	NO
	TP53	0.788	1039	c.724T>C	p.Cys242Arg	missense	YES
21	CSMD3	0.392	917	c.1552C>A	p.Leu518Ile	missense	NO
	ERBB2	0.344	1276	c.722C>T	p.Ala241Val	missense	YES
22	TP53	0.299	1358	c.376-1G>A		splice_acceptor	YES
	FOXA2	0.222	694	c.903dupC	p.Glu302ArgfsTer66	frameshift	NO
23	TP53	0.352	3082	c.536A>T	p.His179Leu	missense	YES
	TP53	0.376	1437	c.329G>T	p.Arg110Leu	missense	YES
24	MKI67	0.21	1876	c.4892G>A	p.Arg1631Gln	missense	YES
	PIK3CA	0.495	503	c.353G>A	p.Gly118Asp	missense	YES
	TP53	0.968	2465	c.701A>G	p.Tyr234Cys	missense	YES

Les variants somàtiques es van classificar de dues maneres, a) en funció del seu efecte sobre el transcrit com *missense* (62%), *frameshift* (24%), *stop gained* (9%), *inframe deletion* (4%), *splice acceptor* (1%) i *protein altering* (1%) (figura 42, A); i b) en funció del seu grau de patogenicitat com variants patogèniques (PV) (51%), variants probablement patogèniques (LPV) (21%), variants de significació desconeguda (VUS) (24%) i variants probablement benignes o neutres (LBNV) (4%) (Figura 42- B, i figura 43).

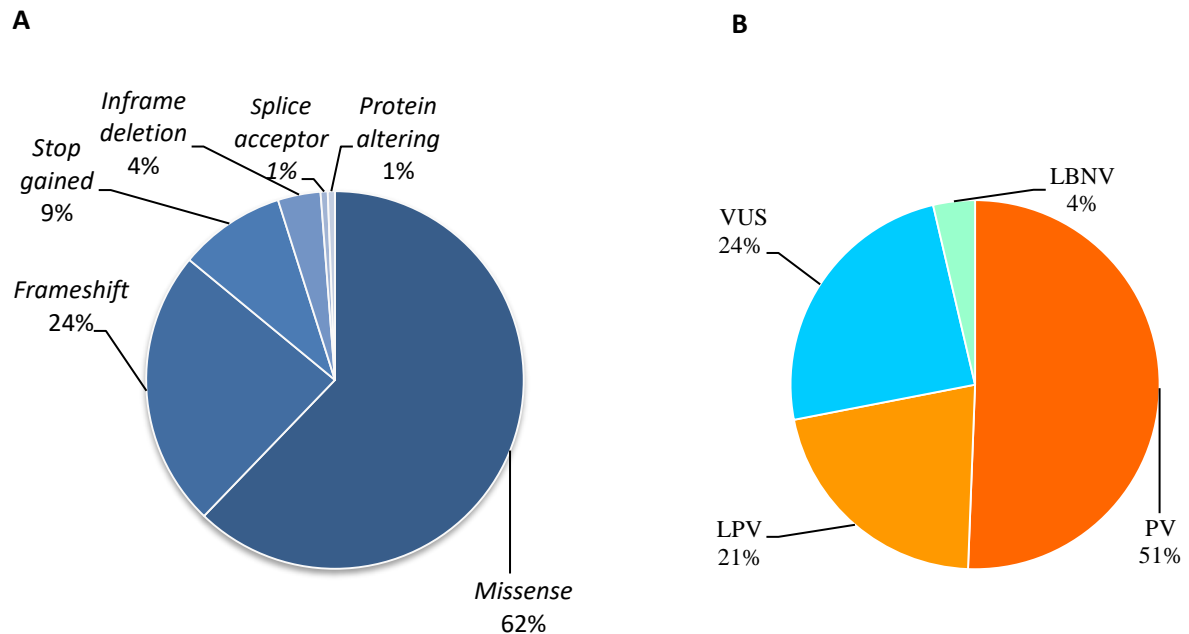


Figura 42. Representació gràfica de les freqüències de les variants somàtiques detectades en funció de l'efecte sobre el transcrit (A) i del seu grau de patogenicitat (B).

Quant als gens alterats, es va observar una mitjana de 6 per mostra. Els gens més freqüentment alterats van ser *PTEN* (63%), *TP53* (54%), *PIK3CA* (50%), *ARID1A* (46%), *MKI67* (29%), *KMT2B* (29%), *PPP2R1A* (25%), *ATR* (21%), *CHD4* (21%), *CTCF* (21%) i *RPL22* (21%) (figura 43). Les vies de senyalització més alterades, tenint en compte els gens alterats amb freqüències superiors al 40%, van ser la via PI3K/AKT/mTOR (*PTEN* i *PIK3CA*), la via de regulació de cicle cel·lular (*TP53*), i la via implicada en la organització de la cromatina (*ARID1A*).

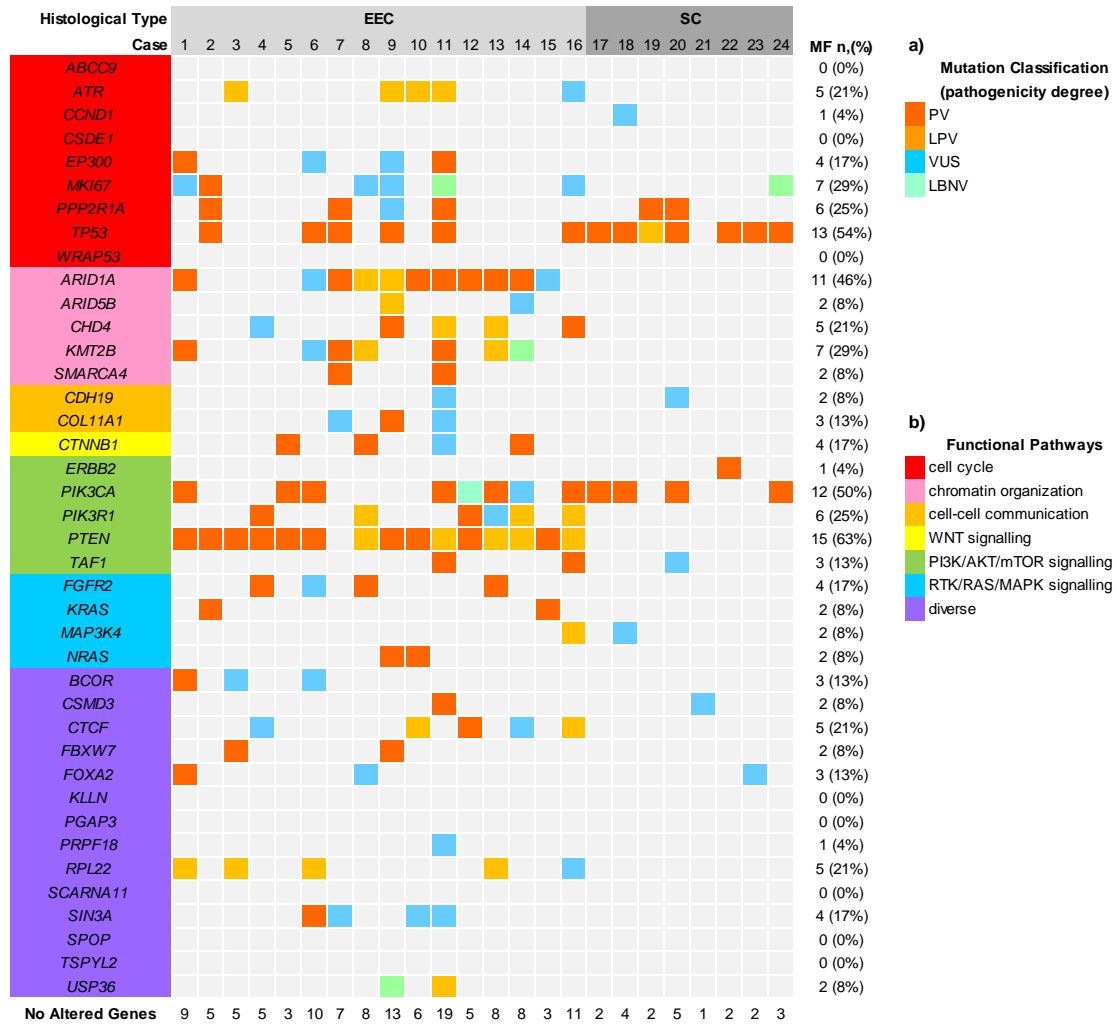


Figura 43. Distribució de les variants somàtiques detectades segons els casos i els gens analitzats. La part superior de l'eix x mostra els casos agrupats en funció del tipus histològic (CEE (EEC), CSE (SC)) i a la part inferior, el número de gens alterats per cas. A la part esquerra de l'eix y es mostren tots els gens analitzats i a la part dreta es mostra la freqüència d'alteració de cada gen (MF). Els quadres de colors indiquen el següent: a) els tipus de variants detectades segons la classificació pel seu grau de patogenicitat, així el taronja fort mostra les PV, el taronja clar les LPV, el blau les VUS i el verd les LBNV; i b) els tipus de vies que estan associades a cadascun dels gens analitzats, així el vermell indica la via cicle cel·lular, en rosa la de l'organització de la cromatina, en taronja clar la de comunicació cel·lular, en groc la de senyalització Wnt, en verd la de senyalització PI3K/AKT/mTOR, en blau la de senyalització RTK/RAS/MAPK, i en lila gens associats a diferents vies menys prevalents.

1.3 ANÀLISI D'ASSOCIACIÓ ENTRE GENS I TIPUS HISTOLÒGIC

Pel que fa a la distribució de les variants en funció del tipus de càncer d'endometri, al grup de CEE es va observar una mitjana de 9 variants per mostra (rang 3 a 23), i al grup de CSE una mitjana de 3 variants per mostra (rang 1 a 5). Pel que fa als gens alterats, en el grup CEE es va observar una mitjana de 8 (rang de 3 a 19) per mostra, i al grup CSE de 3 per mostra (rang de 1 a 5).

Els gens més freqüentment alterats en els CEE van ser *PTEN* (93,75%), *ARID1A* (68,75%), *PIK3CA* (50%), *KMT2B* (43,75%) i *TP53* (37,5%). En el grup de CSE van ser *TP53* (87,5%), *PIK3CA* (50%) i *PPP2R1A* (25%). L'anàlisi estadística d'associació entre gens alterats i tipus histològic va determinar que els gens significativament alterats (p -valor $<0,05$) d'una manera diferencial entre els dos tipus histològics eren *PTEN*, *ARID1A*, *KMT2B* i *TP53*. Observant les *odds-ratio* i els *p-value* veiem que l'alteració en els gens *PTEN*, *ARID1A*, *KMT2B* es va associar amb el tipus CEE, i l'alteració de *TP53* amb el tipus CSE. La taula 15 mostra el llistat de gens alterats ($n=32$) amb el nombre de casos alterats i les freqüències d'alteració, i els valors de l'anàlisi estadística. Dels 40 gens analitzats, en 8 d'ells no es va detectar cap variant (*ABCC9*, *CSDE1*, *KLLN*, *PGAP3*, *SCAARN11*, *SPOP*, *TSPYL2* i *WRAP53*).

Taula 15. Descripció per cada gen alterat (*gene symbol*) del número total de casos alterats (n), freqüència d'alteració (%), *odds-ratio* i p-valor (*p-value*) en funció del tipus histològic, CEE (EEC) i CSE (SC). L'asterisc indica els gens estadísticament significatius (*p-value* < 0.05).

GENE SYMBOL	MUTATED CASES, n (%)		ODDS-RATIO		
	EEC (16 cases)	SC (8 cases)	EEC vs SC	SC vs EEC	p-value
<i>PTEN</i> *	15 (93.75%)	0 (0%)	60.00		0.00001*
<i>ARID1A</i> *	11 (68.75%)	0 (0%)	14.67		0.002*
<i>TP53</i> *	6 (37.5%)	7 (87.5%)		9.62	0.03*
<i>KMT2B</i> *	7 (43.75%)	0 (0%)	5.60		0.03*
<i>PIK3R1</i>	6 (37.5%)	0 (0%)	4.36		0.06
<i>ATR</i>	5 (31.25%)	0 (0%)	3.33		0.1
<i>CHD4</i>	5 (31.25%)	0 (0%)	3.33		0.1
<i>CTCF</i>	5 (31.25%)	0 (0%)	3.33		0.1
<i>RPL22</i>	5 (31.25%)	0 (0%)	3.33		0.1
<i>CTNNB1</i>	4 (25%)	0 (0%)	2.46		0.17
<i>EP300</i>	4 (25%)	0 (0%)	2.46		0.17
<i>FGFR2</i>	4 (25%)	0 (0%)	2.46		0.17
<i>SIN3A</i>	4 (25%)	0 (0%)	2.46		0.17
<i>MKI67</i>	6 (37.5%)	1 (12.5%)	3.66		0.25
<i>BCOR</i>	3 (18.75%)	0 (0%)	1.71		0.28
<i>COL11A1</i>	3 (18.75%)	0 (0%)	1.71		0.28
<i>CCND1</i>	0 (0%)	1 (12.5%)		2.00	0.33
<i>ERBB2</i>	0 (0%)	1 (12.5%)		2.00	0.33
<i>FBXW7</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>KRAS</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>NRAS</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>SMARCA4</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>USP36</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>CDH19</i>	1 (6.25%)	1 (12.5%)		2.07	0.67
<i>CSMD3</i>	1 (6.25%)	1 (12.5%)		2.07	0.67
<i>MAP3K4</i>	1 (6.25%)	1 (12.5%)		2.07	0.67
<i>ARID5B</i>	1 (6.25%)	0 (0%)	1.01		0.67
<i>PRPF18</i>	1 (6.25%)	0 (0%)	1.01		0.67
<i>FOXA2</i>	3 (18.75%)	1 (12.5%)	1.48	0.67	0.76
<i>TAF1</i>	2 (12.5%)	1 (12.5%)		1.05	0.97
<i>PPP2R1A</i>	4 (25%)	2 (25%)	1.00	1.00	0.98
<i>PIK3CA</i>	8 (50%)	4 (50%)	1.00	1.00	1

1.4 IMPORTÀNCIA DE LES VARIANTS SOMÀTIQUES DETECTADES

Mitjançant l'eina *Mutation Mapper* (www.cbioportal.org) es van identificar quins eren els dominis de la proteïna més recurrentment afectats dels gens més freqüentment alterats. Aquesta eina ens va permetre localitzar les variants somàtiques detectades en els dominis proteïcs segons la base de dades de Pfam i, per tant, valorar la rellevància funcional de les variants detectades. La figura 44 mostra els *hotspots* més recurrents dels gens *PTEN*, *TP53*, *PIK3CA* i *ARID1A*.

El 45% (10/22) de les variants detectades al gen *PTEN* es van localitzar en el domini C2, i el 31,8% (7/22) en el domini DSPc, sent R130 el *hotspot* més freqüent (71,4%, 5/7). El 72,8% (16/22) de les variants de *PTEN* generaven una proteïna truncada i el 22% (5/22) restant eren *missense*. De les 16 variants detectades, 10 van ser classificades com variants patogèniques (PV) i les 6 restants com probablement patogèniques (LPV).

En el cas del gen *TP53*, el 86% (13/15) de les variants es localitzaven en el domini d'unió a l'ADN. A diferència de *PTEN*, el 80% (12/15) d'aquestes eren *missense* i només dues variants produïen una proteïna truncada (13%). De les 13 variants detectades 12 van ser classificades com PV i 1 com LPV.

Curiosament, vam observar un mapeig diferencial de les mutacions de *PIK3CA* depenent del tipus de tumor. En les mostres de CSE les variants del gen *PIK3CA* es localitzaven al final de l'exó 2 i al començament de l'exó 3. En canvi, al grup CEE estaven àmpliament distribuïdes al llarg de tot el gen. En aquest últim grup es van observar variants als *hotspots* típics de l'exó 9 i 20, amb unes freqüències del 25% i el 37%, respectivament. De les 12 variants detectades, 10 van ser classificades com PV, 1 com a variant de significació desconeguda (VUS), i 1 com a probablement benigna o neutra (LBNV).

En el cas d'*ARID1A*, les variants estaven distribuïdes al llarg de tot el gen, sense detectar cap *hotspot* recurrent, el 64,3% eren variants que generaven una proteïna truncada i el 35,7% eren *missense*. De les 11 variants detectades, 7 van ser classificades com PV, 2 com LPV, i 2 com VUS.

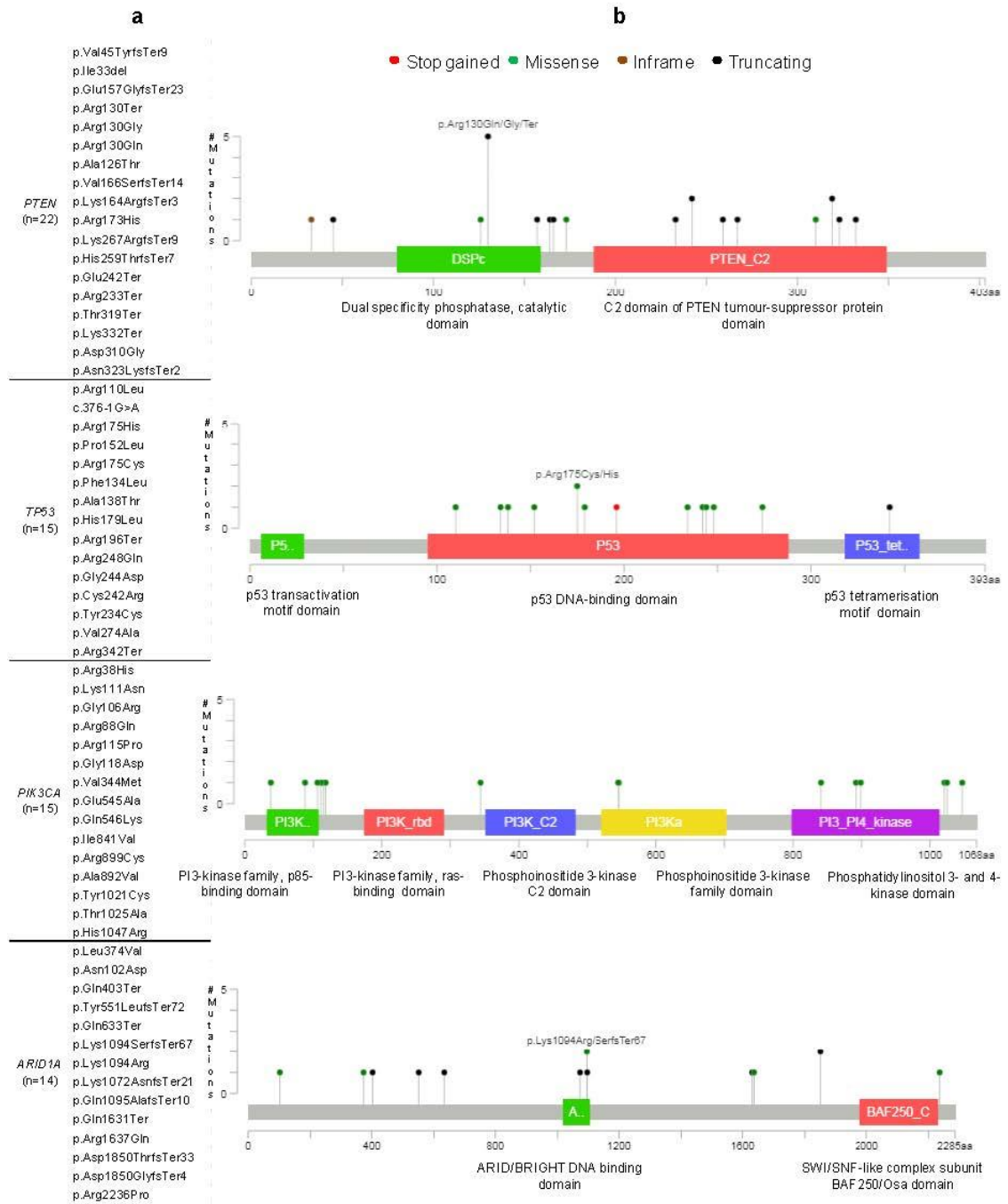


Figura 44. a) llista de les variants somàtiques identificades en els gens *PTEN*, *TP53*, *PIK3CA* i *ARID1A*, b) localització de cadascuna de les variants sobre la seqüència linial de la proteïna on es mostren els dominis Pfam i s'indiquen en vermell les variants *stop gained*, en verd les *missense*, en marró les *inframe* i en negre les *truncating*. L'eix Y indica el nombre de variants i l'eix x representa la seqüència de la proteïna amb els nombres dels aminoàcids.

Les variants del gen *KMT2B* no es van localitzar en dominis coneguts de la proteïna. El 57% d'elles van ser *missense*, on la mutació L417P representava el 75%, i el 43% restant corresponia a variants *frameshift*. De les 7 variants detectades, 3 van ser classificades com PV, 2 com LPV, 1 com VUS i 1 com LBNV.

Totes les variants detectades en el gen *PPP2R1A* eren *missense*, sent la P179R la més observada (50%, 3/6). De les 6 variants detectades, 5 van ser classificades com PV, i 1 com VUS.

Cal esmentar, que en els gens *PTEN*, *ARID1A*, *PIK3CA* i *TP53* es van trobar casos amb la coexistència de més d'una mutació en el mateix gen.

Segons les dades del TCGA, les cinc mostres amb alteració al gen *RPL22* van presentar la variant p.Lys15ArgfsTer5 i quatre d'elles presentaven inestabilitat de microsatèl·lits (MSI).

1.5 ANÀLISI DE DOBLE *CLUSTER* JERÀRQUIC

L'anàlisi de doble *cluster* jeràrquic, descrit als materials i mètodes, va agrupar les mostres, de manera automàtica, en funció del perfil d'alteracions de cadascuna d'elles.

Es van obtenir dos grups clarament diferenciats que estaven caracteritzats per perfils genètics específics i que es correlacionaven cadascun d'ells amb una histologia determinada. Així, el perfil genètic corresponent als casos CEE estava caracteritzat per la presència d'alteracions als gens *PTEN*, *ARID1A* i *KMT2B*; i el dels casos CSE per la presència d'alteracions al gen *TP53* i absència d'alteracions als gens *PTEN*, *ARID1A* i *KMT2B* (figura 45). Aquests resultats confirmen que la nostra anàlisi de *targeted sequencing* (NGS) té capacitat per discriminar entre els casos CEE i CSE, i es pot utilitzar com una eina complementària al diagnòstic histològic per tal d'arribar a un diagnòstic final més objectiu i precís.

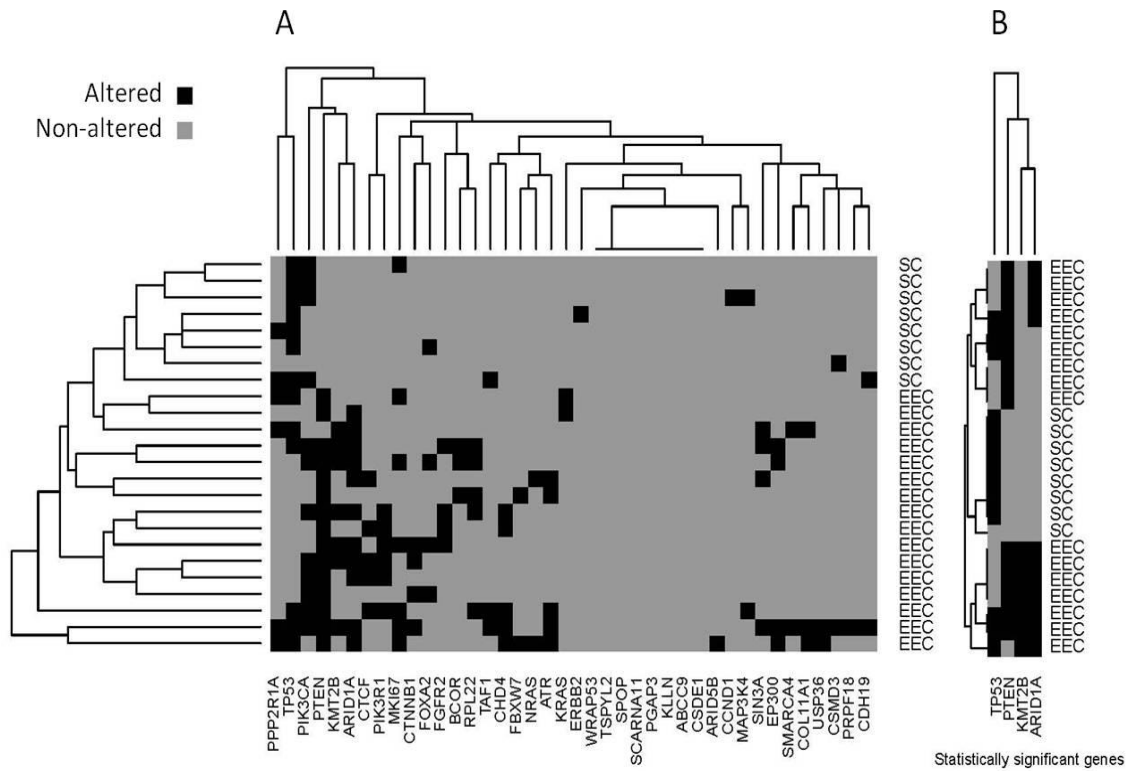


Figura 45. Representació dels dos *heatmaps* obtinguts que il·lustren la presència (*altered* en negre) o no presència (*non-altered* en gris) de variants, tenint en compte per una banda, els 32 gens alterats en les 24 mostres analitzades (A), i per altra, només els 4 gens alterats que són estadísticament significatius (B).

1.6 VALIDACIÓ MITJANÇANT SEQÜENCIACIÓ SANGER

L'anàlisi de seqüenciació Sanger dels gens *TP53*, *PTEN* i *ARID1A* va permetre la identificació de les mateixes variants detectades amb l'anàlisi de *targeted sequencing* (NGS) (taula 16).

A més, l'anàlisi de variants somàtiques del gen *POLE* mitjançant la seqüenciació Sanger va demostrar que tots els casos eren *wild type*, excepte un cas de CEE (cas 11) que va mostrar la variant patogènica p.Y458N situada a l'exó 14 (figura 46). Aquesta anàlisi també va permetre fer la classificació molecular de tots els casos amb el sistema alternatiu al TCGA.

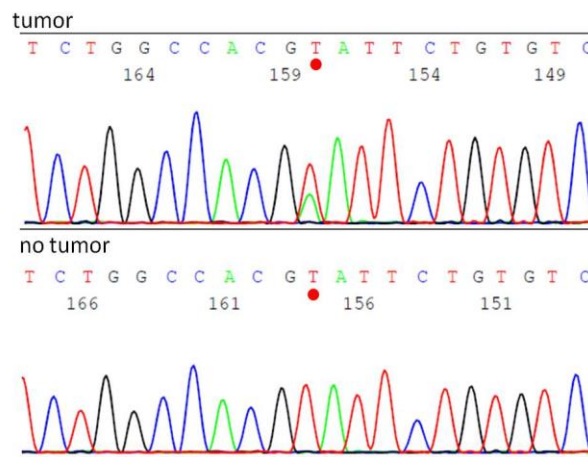


Figura 46. Seqüències obtingudes de l'exó 14 de gen *POLE* de la mostra tumoral i no tumoral del cas 11, on s'observa que la mostra tumoral presenta la variant (indicada amb cercle vermell).

1.7 VALIDACIÓ MITJANÇANT IMMUNOHISTOQUÍMICA

Mitjançant la tècnica immunohistoquímica (IHQ) es va realitzar la validació de l'efecte de les variants dels gens *TP53*, *PTEN* i *ARID1A* (taula 16) sobre l'expressió de la proteïna. En tots els casos es va obtenir una bona correlació entre la immunohistoquímica, l'anàlisi de *targeted sequencing* (NGS) i la seqüenciació Sanger. Dotze (92%) dels 13 tumors amb *TP53* alterat van donar lloc a un patró mutat per IHQ. En el cas de *PTEN* es va observar una tinció negativa en 13 (87%) dels 15 casos alterats. La tinció negativa d'*ARID1A* es va observar en 10 (83%) dels 12 tumors alterats.

Taula 16. Correlació dels resultats obtinguts mitjançant l'anàlisi de *targeted sequencing*, la seqüenciació Sanger i la immunohistoquímica (IHC) del gens *ARID1A*, *PTEN*, i *TP53* per cadascuna de les mostres (case).

CASE	ARID1A			PTEN			TP53		
	TARGETED	SANGER	IHC	TARGETED	SANGER	IHC	TARGETED	SANGER	IHC
1	K1072NfsTer21 R2236P	K1072NfsTer21 R2236P	NS	K332*	K332*	NS	WT	ND	"WT" P
2	WT	ND	PS	D310G	D310G	NS	R175H	R175H	"M" P
3	WT	ND	PS	R233*	R233*	NS	WT	ND	"WT" P
4	WT	ND	PS	R130*	R130*	NS	WT	ND	"WT" P
5	WT	ND	PS	R130G	R130G	PS	WT	ND	"WT" P
6	K1094SfsTer67	K1094SfsTer67	NS	A126T K267RfsTer9	A126T K267RfsTer9	NS	R248Q P152L	R248Q P152L	"M" P
7	R1637Q	R1637Q	PS	WT	ND	PS	R175C	R175C	"M" P
8	Y551LfsTer72	Y551LfsTer72	NS	T319*	T319*	NS	WT	ND	"WT" P
9	D1850TfsTer4	D1850TfsTer4	NS	R130* R173H	R130* R173H	NS	G244D	G244D	"WT" P
10	Q633*	ND	NS	V45YfsTer9 H259TfsTer7	V45YfsTer9 H259TfsTer7	NS	WT	ND	"WT" P
11	K1094R D1850GfsTer4	K1094R D1850GfsTer4	PS	E157GfsTer23 N323KfsTer2	E157GfsTer23 N323KfsTer2	NS	F134L	F134L	"M" P
12	Q403*	Q403*	NS	R130* R130Q	R130* R130Q	PS	WT	ND	"WT" P
13	Q1095AfsTer10	Q1095AfsTer10	NS	E242*	E242*	NS	WT	ND	"WT" P
14	Q1631*	Q1631*	NS	K164RfsTer3	ND	NS	WT	ND	"WT" P
15	N102D L374V	ND	NS	E242*	E242*	NS	WT	ND	"WT" P
16	#P373R #Q372SfsTer19	ND	NS	I33del T319*	I33del T319*	NS	A138T	A138T	"M" P
17	WT	ND	PS	WT	ND	PS	R196*	R196*	"M" P
18	WT	ND	PS	WT	ND	PS	V274A	V274A	"M" P
19	WT	ND	PS	WT	ND	PS	R342*	R342*	"M" P
20	WT	ND	PS	WT	ND	PS	C242R	C242R	"M" P
21	WT	ND	PS	WT	ND	NV	WT	ND	"WT" P
22	WT	ND	PS	WT	ND	NV	splice acceptor c.376-1G>A	splice acceptor c.376-1G>A	"M" P
23	WT	ND	PS	WT	ND	NS	H179L R110L	H179L R110L	"M" P
24	WT	ND	PS	WT	ND	NV	Y234C	Y234C	"M" P

indica la detecció de la variant per sota del llindar 100X de profunditat de lectura (cas 16, 4X); *indica que l'efecte de la variant és la generació d'un codó stop. NS (*negative staining*) tinció negativa; PS (*positive staining*) tinció positiva; NV (*not valuable*) no valorable; "WT" P (*"wild type" pattern*) patró wild type; "M" P (*"mutated" pattern*); ND (*not done*) no realitzat.

La figura 47, mostra l'exemple d'un cas CEE doble mutat (cas 6), que també presentava variants a *PTEN* i *ARID1A*. En aquest exemple es pot observar la bona correlació entre 3 tècniques, la seqüenciació Sanger, la *targeted sequencing* i la immunohistoquímica.

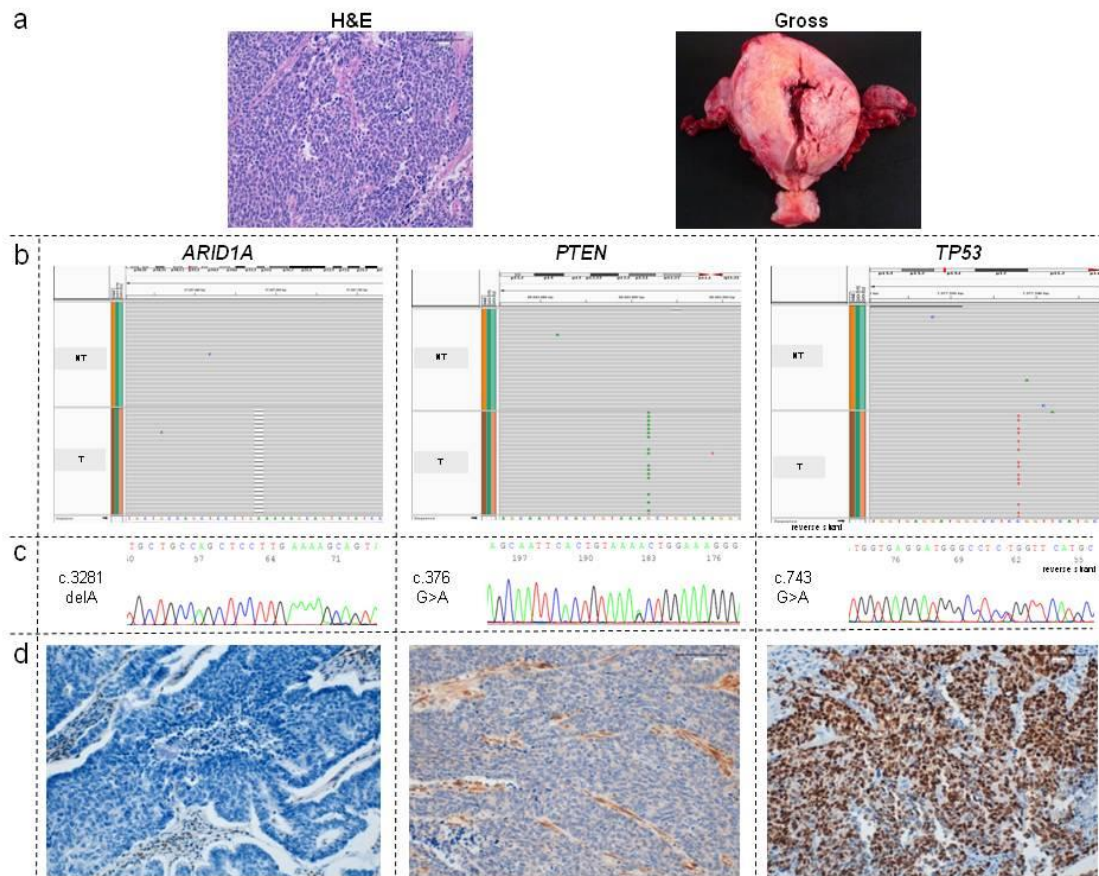


Figura 47. Composició que il·lustra la integració de l'anàlisi macroscòpica, microscòpica i molecular d'un CEE de grau 3 (cas 6). a) Aspecte microscòpic ($\times 20$) i macroscòpic; b) visualització mitjançant l'eina l'IGV de les variants detectades als gens *ARID1A* (c.3281delA, p.K1094SfsTer67), *PTEN* (c.376G>A, p.A126T) i *TP53* (c.743 G>A, R248Q); c) seqüències Sanger de les variants detallades en el punt b); i d) resultat immunohistoquímic de l'anàlisi de les proteïnes *ARID1A*, *PTEN* i *TP53* ($\times 20$).

1.8 CLASSIFICACIÓ MOLECULAR AMB EL SISTEMA ALTERNATIU AL TCGA

Per tal de realitzar la classificació molecular de tots els casos amb el sistema alternatiu al TCGA es va fer ús de la tècnica IHQ per les proteïnes p53, MSH6 i PMS2, i la seqüenciació del gen *POLE* (taula 17). Com hem descrit a l'apartat de material i mètodes, els casos van ser classificats en cinc categories:

- 1) els casos amb inestabilitat de microsatèl·lits o hipermutats (MSI) que inclouen els casos negatius per MSH6 i PMS2.
- 2) els casos tipus-serosos (*serous-like*) que inclouen els casos amb un patró mutat per p53.
- 3) els casos *POLE*-mutats, on també es van incloure els casos amb mutacions al gen *POLE* encara que fossin MSI o p53 mutat.
- 4) els casos amb un perfil mutacional no específic (NSMP, *No Specific Mutation Profile*) que inclouen casos que no s'han pogut classificar en cap dels grups anteriors,
- 5) els casos amb la presència de més d'un marcador alternatiu (MSI i p53) van ser classificats com doble mutats.

Pel que fa als resultats de l'expressió de MSH6 i PMS2, es va obtenir una tinció negativa en 1 cas per MSH6, i en 7 casos per PMS2. D'aquesta manera en 8 casos, els quals eren CEE, es va detectar una deficiència del sistema de reparació de l'ADN (MMRd o MSI). En 2 d'aquests casos, va haver-hi la coexistència de MSI i defecte de p53, així van ser categoritzats com doble mutats. Els casos MSI tenien un rang de 7 a 19 gens alterats (mitjana 10; mediana 9). Els casos MSS, amb les proteïnes reparadores estables, tenien un rang d'1 a 6 gens alterats (mitjana 3,7; mediana 4). Set casos van ser classificats com NSMP, 6 eren CEE i un era CSE (presentava patró *wild type* per p53 i tampoc tenia cap variant somàtica). Es van categoritzar 8 casos com *serous-like*, els quals presentaven un patró mutat de p53 per IHQ. Curiosament, 1 d'aquests casos corresponia a CEE. L'únic cas mutat per *POLE* (cas 11) va ser un tumor CEE de grau 3, que va mostrar una tinció negativa pel PMS2, així com el major nombre de variants (19), inclosos els gens *TP53*, *PTEN* i *ARID1A*.

Taula 17. Detall de la classificació molecular de tots els casos analitzats (*case*), a més a més de l'edat de les pacients (*age*), el tipus histològic (*histological type*), el grau (*grade*), l'estadi (*stage*), i l'estat del seguiment de la malaltia (*follow up*).

CASE	HISTOLOGICAL			MOLECULAR		FOLLOW UP
	AGE	TYPE	GRADE	STAGE	CLASSIFICATION	
1	86	EEC	II	IB	MSI (Hypermutated)	NED
2	91	EEC	III	II	Serous-like	DOD
3	94	EEC	II	IB	MSI (Hypermutated)	NED
4	91	EEC	II	IB	NSMP	NED
5	83	EEC	II	IA	NSMP	NED
6	60	EEC	III	IIIA	Double mutated*	NED
7	67	EEC	III	IB	Serous-like	NED
8	76	EEC	II	IB	MSI (Hypermutated)	NED
9	77	EEC	III	IB	Double mutated*	NED
10	86	EEC	I	IB	NSMP	NED
11	73	EEC	III	IB	POLE-mutated	NED
12	77	EEC	III	IB	NSMP	NED
13	77	EEC	III	IB	MSI (Hypermutated)	NED
14	77	EEC	II	II	NSMP	NED
15	78	EEC	I	IB	NSMP	DDD
16	59	EEC	III	IB	Double mutated*	NED
17	86	SC	III	IA	Serous-like	DOD
18	89	SC	III	IIIA	Serous-like	DOD
19	85	SC	III	IB	Serous-like	AWOD
20	76	SC	III	IA	Serous-like	DOD
21	81	SC	III	IA	NSMP	DOD
22	87	SC	III	IIIB	Serous-like	DOD
23	93	SC	III	IB	Serous-like	DOD
24	96	SC	III	II	Serous-like	DOD

*indica els casos doble mutats (*double mutated*); MSI(*hypermutated*), tumors amb inestabilitat de microsatèl·lits; NSMP (*no specific molecular profile*), tumors amb un patró mutacional no específic; EEC (*endometrial endometrioid carcinoma*), CEE; SC (*serous carcinoma*), CSE; NED (*no evidence of disease*), no evidència de la malaltia; DOD (*dead of disease*), mort per la malaltia; DDD (*dead of different disease*), mort per causes diferents; AWOD (*alive with other disease*), viu amb altres malalties.

1.9 INTEGRACIÓ DE RESULTATS: CORRELACIÓ ENTRE LA CLASSIFICACIÓ HISTOLÒGICA, LA CLASSIFICACIÓ MOLECULAR ALTERNATIVA, I EL PERFIL GENÈTIC

En aquest últim apartat es mostra la integració de tots els resultats obtinguts per tal de visualitzar d'una manera general la correlació entre els diferents sistemes de classificació utilitzats. La taula 18 ens detalla per cada cas la classificació histològica, la molecular alternativa i la classificació en funció del perfil genètic. Veiem que la classificació mitjançant el perfil genètic, tenint en compte els 4 gens significatius, concorda amb la classificació histològica. On hi ha discrepàncies és entre la classificació alternativa del TCGA i el perfil genètic. Els casos discrepants són el 2 i 7, que amb la classificació alternativa són *serous-like* i el perfil genètic els classifica com un CEE; i el 21, que amb la classificació alternativa és un NSMP i el perfil genètic el classifica com un CSE.

Taula 18. Detall de la classificació histològica (*histological type*), de la classificació alternativa al TCGA (*molecular classification (surrogate)*) i de la classificació mitjançant el perfil genètic (*classification by genetic profile*) per cadascun dels casos analitzats.

CASE	HISTOLOGICAL TYPE	MOLECULAR CLASSIFICATION (SURROGATE)	CLASSIFICATION BY GENETIC PROFILE				Cluster Analysis Groups
			<i>ARID1A</i>	<i>KMT2B</i>	<i>PTEN</i>	<i>TP53</i>	
1	EEC	MSI (Hypermutated)	Altered	Altered	Altered	No altered	EEC
2	EEC	Serous-like	No altered	No altered	Altered	Altered	EEC
3	EEC	MSI (Hypermutated)	No altered	No altered	Altered	No altered	EEC
4	EEC	NSMP	No altered	No altered	Altered	No altered	EEC
5	EEC	NSMP	No altered	No altered	Altered	No altered	EEC
6	EEC	Double mutated*	Altered	Altered	Altered	Altered	EEC
7	EEC	Serous-like	Altered	Altered	No altered	Altered	EEC
8	EEC	MSI (Hypermutated)	Altered	Altered	Altered	No altered	EEC
9	EEC	Double mutated*	Altered	No altered	Altered	Altered	EEC
10	EEC	NSMP	Altered	No altered	Altered	No altered	EEC
11	EEC	POLE-mutated	Altered	Altered	Altered	Altered	EEC
12	EEC	NSMP	Altered	No altered	Altered	No altered	EEC
13	EEC	MSI (Hypermutated)	Altered	Altered	Altered	No altered	EEC
14	EEC	NSMP	Altered	Altered	Altered	No altered	EEC
15	EEC	NSMP	Altered	No altered	Altered	No altered	EEC
16	EEC	Double mutated*	No altered	No altered	Altered	Altered	EEC
17	SC	Serous-like	No altered	No altered	No altered	Altered	SC
18	SC	Serous-like	No altered	No altered	No altered	Altered	SC
19	SC	Serous-like	No altered	No altered	No altered	Altered	SC
20	SC	Serous-like	No altered	No altered	No altered	Altered	SC
21	SC	NSMP	No altered	No altered	No altered	No altered	SC
22	SC	Serous-like	No altered	No altered	No altered	Altered	SC
23	SC	Serous-like	No altered	No altered	No altered	Altered	SC
24	SC	Serous-like	No altered	No altered	No altered	Altered	SC

EEC (*endometrial endometrioid carcinoma*), CEE ; SC (*serous carcinoma*), CSE; *indica els casos doble mutats (*double mutated*); MSI(*hypermutated*), tumors amb inestabilitat de microsatèl·lits; NSMP (*no specific molecular profile*), tumors amb un patró mutacional no específic.

2 ALTERACIONS SOMÀTIQUES EN EL NÚMERO DE CÒPIES DELS GENS

En aquest apartat es detallen els resultats obtinguts en el marc de l'objectiu específic 2: analitzar les alteracions somàtiques en el número de còpies de gens (ASNC) mitjançant la tècnica MLPA per tal d'establir un perfil de gens que permeti classificar el càncer d'endometri. Aquestes alteracions s'han determinat en mostres pertanyents als dos tipus més freqüents de càncer d'endometri (CEE i el CSE).

2.1 DESCRIPCIÓ DE LES ALTERACIONS SOMÀTIQUES EN EL NÚMERO DE CÒPIES DELS GENS

Com s'ha descrit a la introducció, l'estat del número de còpies dels gens es va classificar en tres categories: 1) pèrdua o deleció (quan el DQ (*Dosage Quocient*) per una sonda era inferior a 0,7), 2) normal o diploide (quan el DQ per una sonda estava entre 0,7 i 1,4), i 3) guany, duplicació o amplificació (quan el DQ per una sonda era superior a 1,4). La figura 48 exemplifica com es van obtenir els valors DQ d'una mostra amb l'ús d'un dels Kits MLPA. D'una banda, es mostren els pics obtinguts en l'anàlisi de fragments corresponents a cadascuna de les sondes amplificades d'un determinat kit MLPA, tant per la mostra tumoral com per la mostra no tumoral (control), i d'altra banda es mostra el gràfic amb els valors DQ obtinguts per cadascuna de les mostres, després de la normalització intra i inter mostra. Als gràfics estan marcades aquelles sondes que han donat uns valors alterats, amb groc està marcada una sonda amb pèrdua (*BIRC4*) i amb blau les sondes amb guany (*ERBB2*).

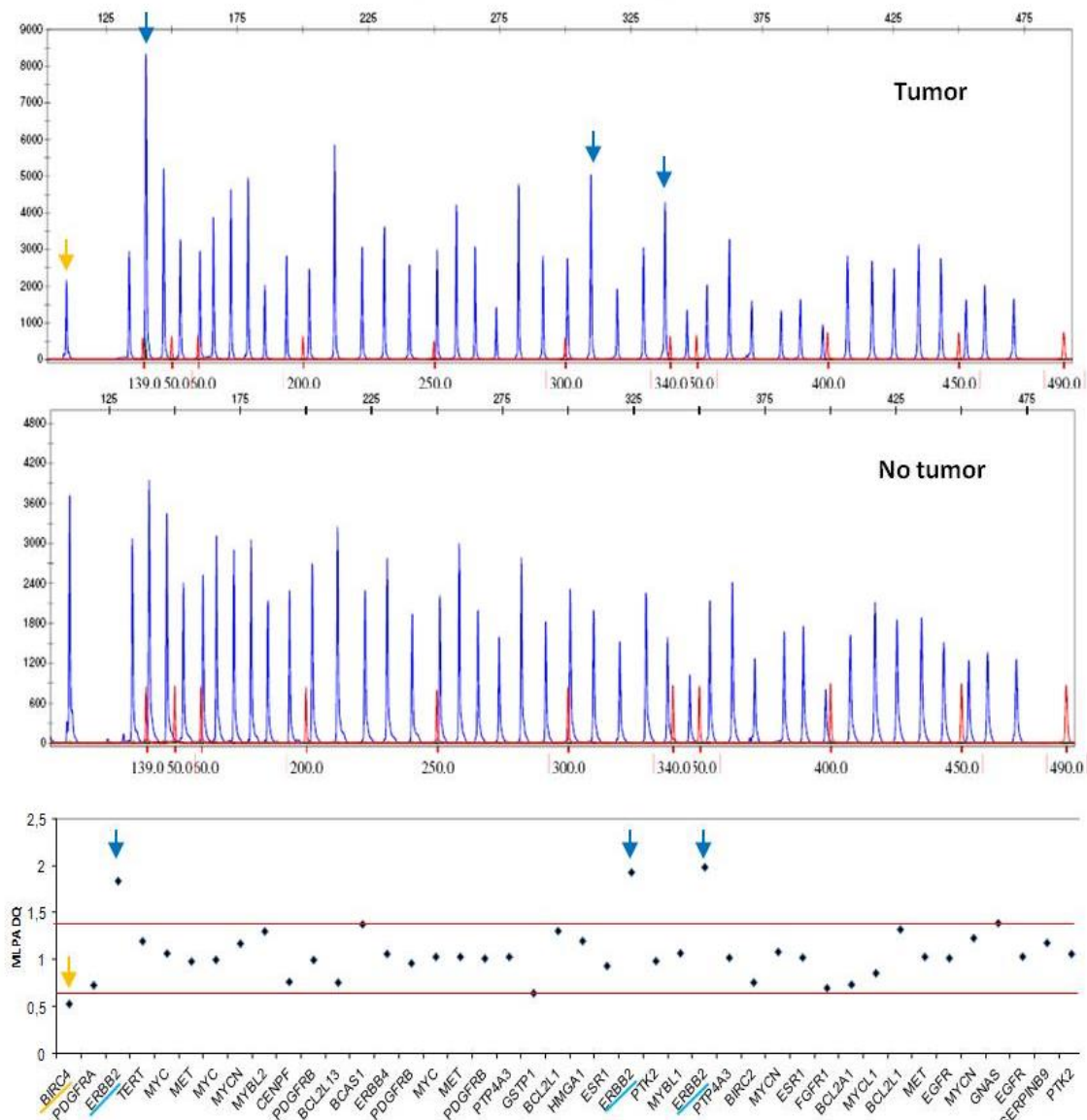


Figura 48. Exemple de l'obtenció dels pics corresponents a cadascuna de les sondes incloses al Kit SALSA P171-A2 *gain* mitjançant l'electroforesi capil·lar, tant de la mostra tumor com de la no tumor, i de la seva transformació a valors DQ en el gràfic de sota. En aquest gràfic l'eix y indica els valors de DQ MLPA, on l'interval normal es troba entre 0,7-1,4 (rang definit per les línies vermelles). Les fletxes blaves indiquen sondes detectades amb guanys i les grogues amb pèrdua.

La figura 49 mostra la distribució dels valors de DQ obtinguts en totes les mostres de CEE i de CSE per tots els gens analitzats. Amb la comparació dels dos gràfics es poden constatar les diferències quant al perfil dels valors DQ, i consegüentment, d'ASNC entre els dos subconjunts de mostres, posant de manifest que el grup dels CSE presenten molta més freqüència d'ASNC, sobretot a nivell de guanys.

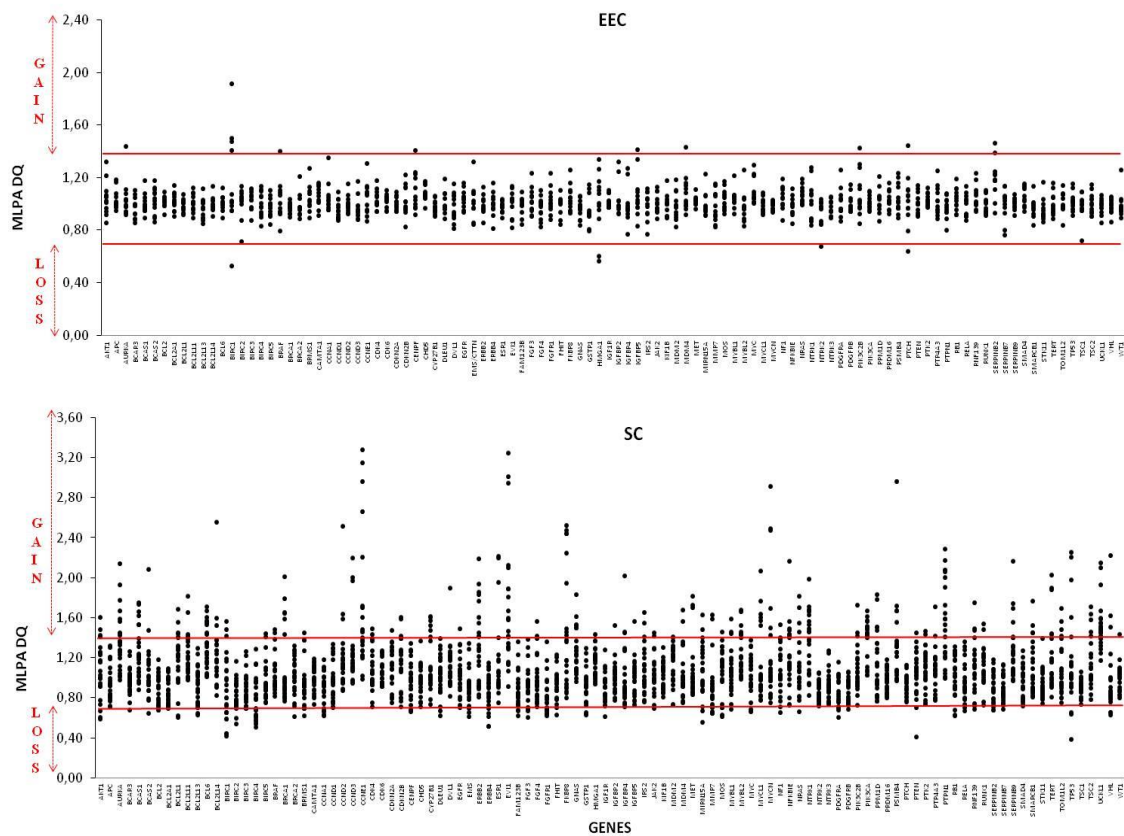


Figura 49. Distribució dels valors de DQ obtinguts en totes les mostres de CEE (EEC) i de CSE (SC) per tots els gens analitzats (eix x). L'eix y indica els valors de DQ MLPA, on l'interval normal es troba entre 0,7-1,4 (rang definit per les línies vermelles), el de guany (*gain*) és el superior a 1,4 i el de pèrdua (*loss*) és l'inferior a 0,7.

Capítol 4

En el grup de CEE es van detectar ASNC en 10 (83%) dels 12 casos estudiats però amb unes freqüències d'alteració (FA) molt baixes. D'aquests 10 casos, en 5 (50%) es va detectar només un únic gen alterat (0.9%), en 2 casos 2 gens alterats, en altres 2 casos tres gens alterats i en 1 cas 4 gens alterats. La mitjana de gens alterats per mostra va ser de 1,6, amb un rang que oscil·lava entre 0 a 4. La mitjana de gens amb guanys va ser de 1,1 (rang de 0 a 3), i la de gens amb pèrdues vas ser de 0,5 (rang de 0 a 2). A diferència del grup de CEE, en el grup de CSE tots els casos analitzats presentaven ASNC i freqüències d'alteració més elevades. Quatre casos (33,3%) presentaven freqüències d'alteració similar al dels casos CEE, oscil·lant entre el 0,9 i el 2%. Els altres 8 casos (66,6%) presentaven unes FA que anaven de l'11 al 31%. La mitjana de gens alterats per mostra va ser de 14, amb un rang que oscil·lava entre 1 i 33. La mitjana de gens amb guanys va ser de 10,8 (rang de 1 a 22), i la de gens amb pèrdues vas ser de 3,2 (rang de 0 a 11). (figura 50).

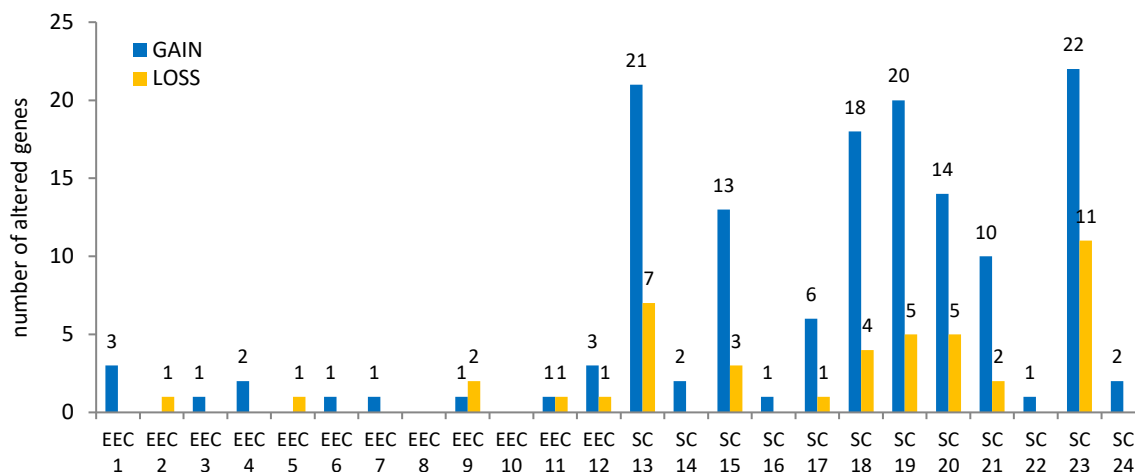


Figura 50. Número de gens alterats, a nivell de guanys (blau) i de pèrdues (groc), per cadascuna de les mostres analitzades CEE (EEC) i CSE (SC).

2.2 GENS ALTERATS I LOCALITZACIONS CROMOSÒMIQUES IMPLICADES

De tot el conjunt de mostres tumorals analitzades (n=24), es van detectar un total de 80 (75%) gens alterats, dels quals 63 (79%) presentaven guanys i 28 (35%) pèrdues. D'entre aquests n'hi havia 11 que depenent de la mostra presentaven guanys o pèrdues.

Al grup de CEE es van trobar alteracions en un total de 12 gens (11,3%), 9 van ser alterats amb guanys i 5 amb pèrdues. Cal dir que en tres d'aquests gens es van detectar guanys o pèrdues depenent de la mostra. Els gens amb guanys més freqüentment alterats van ser *BIRC1* (36%), *SERPIN2* (17%), *HMGA1* (17%), *PTCH* (9%), i amb pèrdua *HMGA1* (17%), *BIRC1* (13%) i *PTCH* (9%). Al grup de CSE, es van observar un total de 75 gens alterats (70,75%), 59 d'aquests (78,6%) presentaven guanys i 25 (33,3%) presentaven pèrdues. Cal dir que en 10 d'aquests gens es van detectar guanys o pèrdues depenent de la mostra. Els gens amb guanys més freqüentment alterats van ser *CCNE1* (58%), *UCKL1* (58%), *EVI1* (50%), *PIK3CA* (50%), *NTRK1* (42%), *PTPN1* (42%), *AURKA* (42%), *BCL6* i *FKBP8* (33%). Els gens més freqüents amb pèrdues van ser *BIRC4* (42%), *BIRC1* (36%), *FGFR1* (33%), i *BIRC2*, *CCNA1*, *ERBB4* amb un 17 % respectivament. Les localitzacions cromosòmiques més freqüents dels gens amb guanys van ser al cromosoma 19, la 19q12 (*CCNE1*) i 19p13.11 (*FKBP8*), al 20, la 20q13.13 (*PTPN1*) i 20q13.33 (*UCKL1*), al 3, la 3q26 (*EVI1*), 3q26.32 (*PIK3CA*) i 3q27 (*BCL6*), i a l'1, la 1q23.1 (*NTRK1*). Pel que fa als gens amb pèrdua les localitzacions van ser al cromosoma X, la Xq25 (*BIRC4*) i al 8, la 8p11.2 (*FGFR1*). (figura 51).

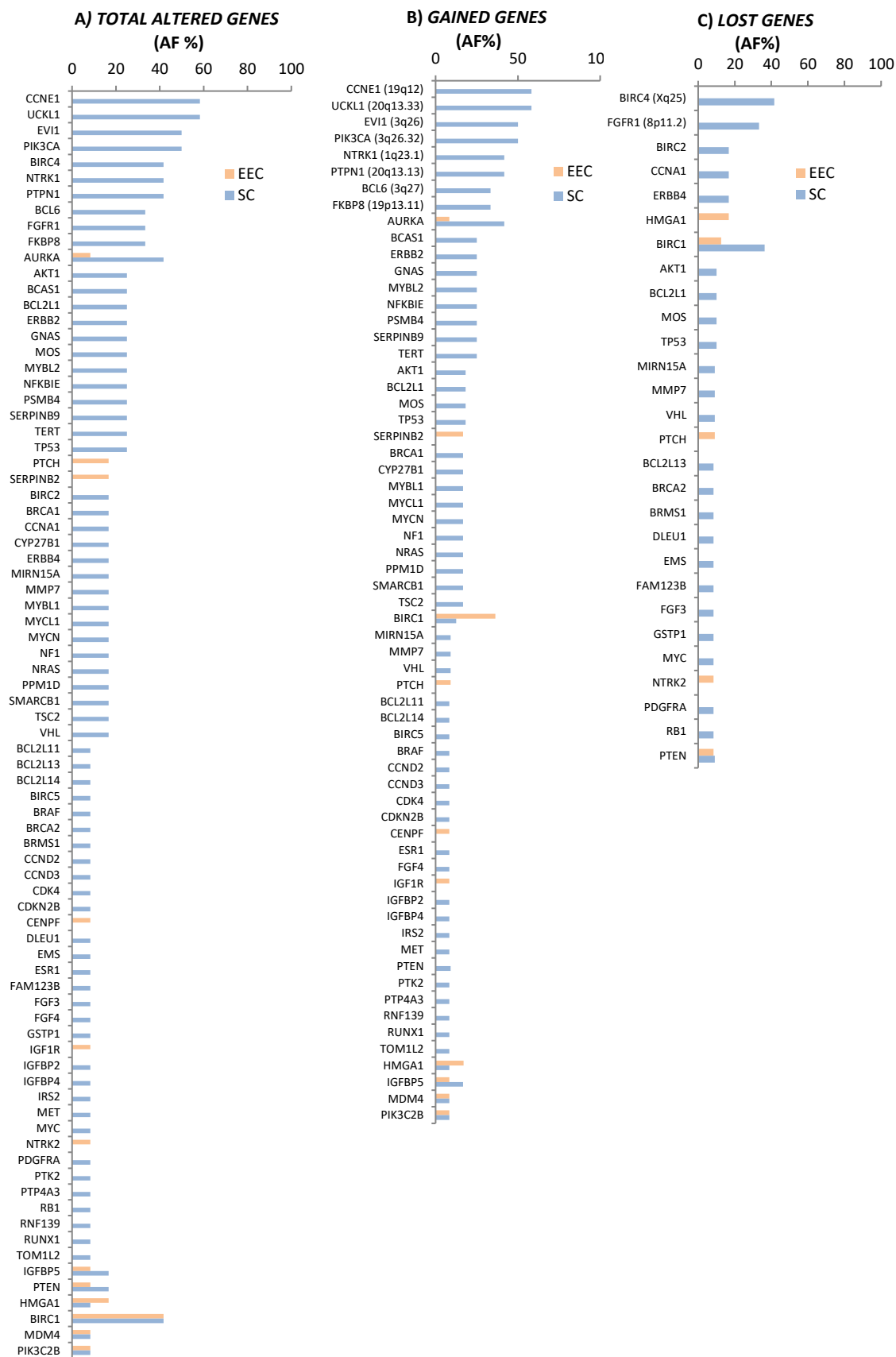


Figura 51. A) Freqüències d'alteració de tots els gens per cada subconjunt de mostres; B) freqüències dels gens amb guanys ; i C) freqüències dels gens amb pèrdues.

2.3 ANÀLISI D'ASSOCIACIÓ ENTRE GENS I TIPUS HISTOLÒGIC

L'anàlisi estadística d'associació entre gens i tipus histològic va determinar que els gens significativament alterats (p -valor $<0,05$) d'una manera diferencial entre els dos tipus histològics eren *CCNE1*, *UCKL1*, *EVI1*, *PIK3CA*, *NTRK1*, *PTPN1*, *BCL6* i *FKBP8*, *BIRC4* i *FGFR1* (taula 19). Els vuit primers van ser detectats a nivell de guanys, i els dos últims a nivell de pèrdues. És de remarcar que aquests gens no es van trobar alterats en cap de les mostres del tipus CEE, i per tant l'alteració en aquests gens va estar associada només al CSE.

Taula 19. Descripció per a cada gen significativament alterat (*gene symbol*) del número total de casos alterats (n), freqüència d'alteració (%), ODDS-RATIO i p-valor (*p-value*) en funció del tipus histològic, CEE (EEC) i CSE (SC).

	GENE SYMBOL	ALTERED CASES (n, %)		ODDS-RATIO	
		EEC (12 cases)	SC (12 cases)	CEE vs SC	SC vs CEE
GAIN	<i>CCNE1</i> *	0 (0%)	7 (58.33%)	14.00	0.002*
	<i>UCKL1</i> *	0 (0%)	7 (58.33%)	14.00	0.002*
	<i>EVI1</i> *	0 (0%)	6 (50%)	10.29	0.007*
	<i>PIK3CA</i> *	0 (0%)	6 (50%)	10.29	0.007*
	<i>NTRK1</i>	0 (0%)	5 (41.67%)	7.50	0.02*
	<i>PTPN1</i> *	0 (0%)	5 (41.67%)	7.50	0.02*
	<i>BCL6</i> *	0 (0%)	4 (33.33%)	5.33	0.047*
	<i>FKBP8</i> *	0 (0%)	4 (33.33%)	5.33	0.047*
LOSS	<i>BIRC4</i> *	0 (0%)	5 (41.67%)	7.50	0.02*
	<i>FGFR1</i> *	0 (0%)	4 (33.33%)	5.33	0.047*

2.4 ANÀLISI DE DOBLE *CLUSTER* JERÀRQUIC

L'anàlisi de doble *cluster* jeràrquic, descrit als materials i mètodes, va agrupar les mostres en funció de les alteracions de cadascuna d'elles.

De manera global, tenint en compte els 10 gens significativament alterats, l'anàlisi va agrupar les mostres en dos grans grups, el grup 1 (G1) que contenia tots els casos de CEE (n=12) més 5 casos de CSE, i el grup 2 (G2) que incloïa 7 casos de CSE (figura 52). El G1 es caracteritzava per tenir entre cap i 2 gens significatius alterats, i el G2, amb només casos CSE, presentava entre 5 i 8 gens significatius alterats.

Els casos del grup 1 es van reagrupar en 3 subgrups, A, B i C. El B corresponia a tots els casos CEE i a dos casos CSE que presentaven només un gen alterat, i no era cap dels 10 més significatius. L'A corresponia a un cas CSE (cas 17) que presentava alteracions a gens estadísticament significatius com *FKBP8* i *UCKL1* a més d'altres gens alterats (un total de 7). Finalment el C corresponia a dos casos CSE, que estaven molts propers al grup 2, que presentaven també una baixa freqüència d'alteració amb només un gen alterat, però en aquest cas el gen era *CCNE1*, que va resultar ser un dels gens amb més poder estadístic (*p-value* 0.002). Amb aquesta agrupació, observem que hi ha tres casos de CSE que s'ens agrupen en unes categories intermitges (G1A i G1C) entre els típics CEE (G1B) i els típics CSE (G2). (figura 52).

Aquests resultats demostren que la nostra anàlisi d'ASNC mitjançant MLPA permet identificar un subgrup de casos amb histologia serosa que genotípicament s'agrupen amb els casos d'histologia endometriode, i per tant, suggereixen que aquesta anàlisi es pot utilitzar com una eina complementària al diagnòstic histològic per tal d'arribar a un diagnòstic final més objectiu i precís.

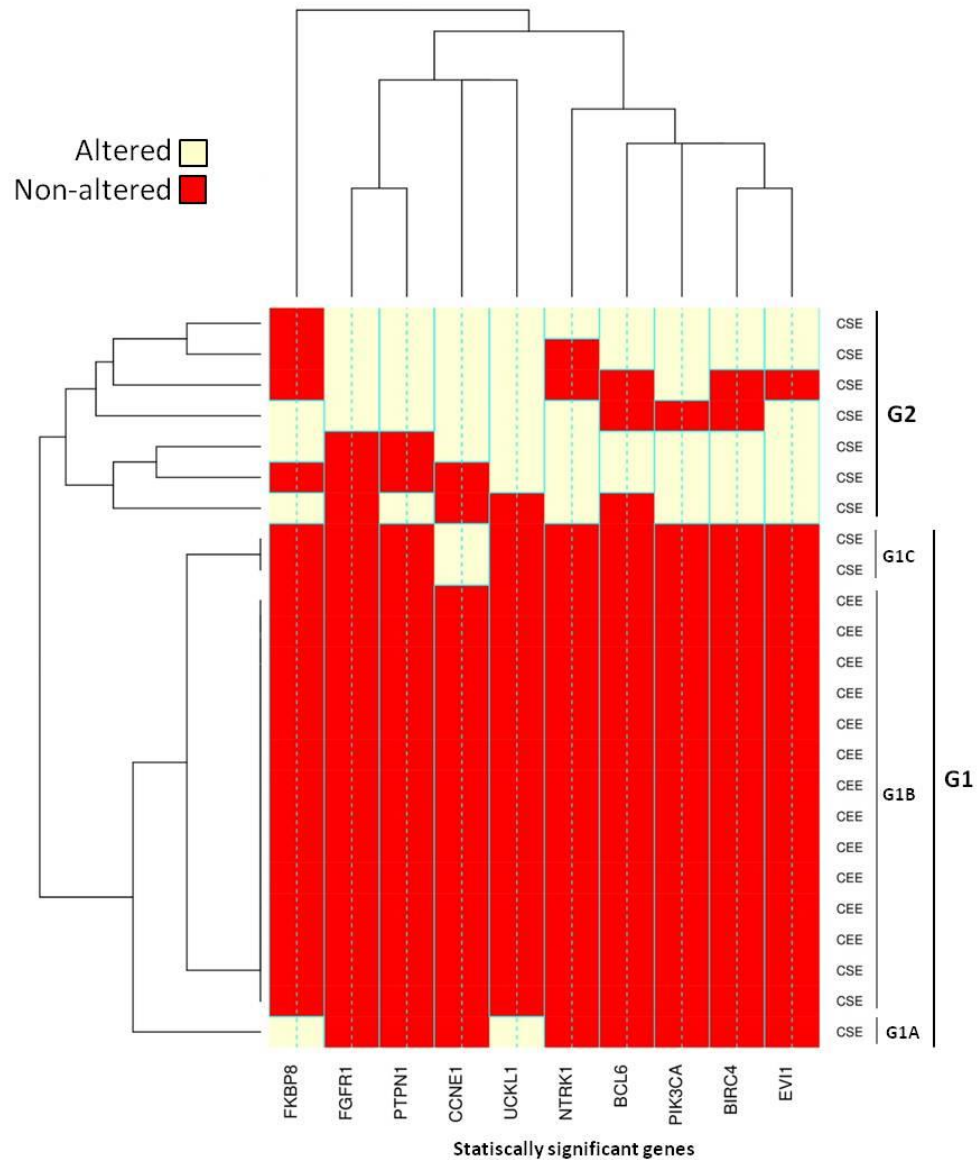


Figura 52. Heatmap que il·lustra la presència (*altered*, groc clar) o no presència d'alteració (*non-altered*, vermell) dels 10 gens significativament alterats (eix x) en les 24 mostres analitzades (eix y). S'indiquen els diferents grups: G1 (G1A, G1B, G1C) i G2.

2.5 SEQÜENCIACIÓ SANGER GEN *POLE*

Es va realitzar l'anàlisi mutacional del gen *POLE* dels 4 casos (14, 16, 22 i 24) dels CSE que presentaven unes baixes freqüències d'ASNC. Un d'aquests casos, el 14, va presentar la variant somàtica patològica p.P286R (c.857C>G) situada a l'exó 9, confirmant que era un CEAG amb un potencial bon pronòstic. (figura 53) .

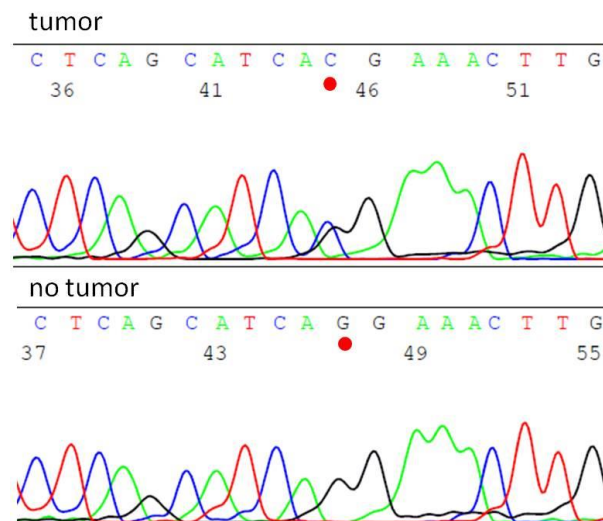


Figura 53. Seqüències obtingudes de l'exó 9 de gen *POLE* de la mostra tumoral i no tumoral del cas 14, on s'observa que la mostra tumoral presenta la variant c.857C>G (indicada amb cercle vermell).

2.6 ANÀLISI IMMUNOHISTOQUÍMICA DE P53

Ja que segons el sistema de classificació alternatiu del TCGA, s'utilitza la IHQ de p53 per classificar els casos com CN *high/serous-like*, en tots els casos d'aquest objectiu es va fer la determinació de p53 per veure l'associació entre el patró IHQ de p53 i els valors d'ASNC.

En global es va obtenir una freqüència d'alteració de p53 del 42% (10/24). Dels 14 casos amb patró p53 wild type, 13 casos (93%) presentaven un baix número d'ASNC, 11 corresponien a CEE i 2 a CSE; i l'altre cas (7%) era un CSE que presentava un alt número d'ASNC. Dels 10 casos amb patró mutat de p53, 3 (30%) eren casos que presentaven un baix número d'ASNC, 1 era un CEE i 2 eren CSE, i 7 (70%) eren CSE que presentaven un alt número d'ASNC.

2.7 INTEGRACIÓ DE RESULTATS: CORRELACIÓ ENTRE LA CLASSIFICACIÓ HISTOLÒGICA, L'EXPRESSIÓ DE P53, EL NÚMERO D'ASNC I EL PERFIL GENÈTIC

En aquest últim apartat es mostra la integració de tots els resultats obtinguts per tal de visualitzar d'una manera general la correlació entre els diferents sistemes de classificació utilitzats. La taula 20 detalla per cada cas com va ser la classificació histològica, l'expressió de p53, la freqüència d'ASNC i la classificació en funció del perfil genètic. Per als casos CEE la concordança es va definir en presentar histologia de CEE, patró *wild type* de p53 (entès com mètode per captar els casos amb ASNC), baix número d'ASNC, i pertanyer al grup perfil molecular G1B. Per als casos CSE la concordança es va definir en presentar histologia de CSE, patró mutat de p53, elevat número d'ASNC i pertanyer al grup molecular G2.

Dins els casos CEE veiem que hi ha una concordança del 92% entre les 4 categories. Només hi ha un cas, cas 4, on no hi ha concordança entre el resultat de p53, que era mutat, i la histologia d'endometrioides i l'assignació al grup G1B. Aquest cas també va ser analitzat a l'objectiu 1, cas 16, i allí es va confirmar que era un cas doble mutat, amb presència de MSI i d'alteració a *TP53*, i el perfil genètic el va classificar com a CEE amb alteració a *PTEN* i *TP53*. En aquests casos l'alteració de *TP53* no s'enten com un event *driver* sinó com un *passenger*, i no ha de ser classificat com un *serous-like* ja que el que preval és l'estat MSI i per tant està ben classificat amb el perfil genètic.

Dins els casos CSE, s'ha observat una concordança del 50% (6/12) entre les 4 categories. Els 6 casos amb discordança entre les 4 categories els podem subdividir en 4 grups, el que inclouria el cas 19, el que inclouria els casos 14 i 24, el que inclouria els casos 16 i 22, i el que inclouria el cas 17. En el cas 19, no es correlaciona ni la histologia CSE, ni l'alt número d'ASNC, ni l'assignació al grup G2, amb el patró *wild type* de p53. Curiosament, aquest cas també va ser analitzat a l'objectiu 1, cas 21, on es va confirmar que no presentava cap variant somàtica per *TP53* i va ser classificat pel perfil genètic com un CSE. Per tant, en aquest cas el perfil genètic d'ASNC el classificaria correctament com un *serous-like* del grup 4 del TCGA, i no erròniament com un *CN-low* o *NSMP*. En els casos 14 i 24, no es correlaciona la histologia de CSE amb el patró *wild type* de p53, el baix número d'ASNC i l'agrupació al grup G1B. Concretament el cas 14 es va confirmar que era *POLE* mutat i per tant es classificaria dintre el grup 1 del TCGA (*POLE ultramutated*), i el cas 24 es va confirmar que no presentava inestabilitat de microsatèl·lits i per tant s'hauria de classificar dins el grup 3 del TCGA (*CN-low* o *NSMP*). En aquests 2 casos demostrem que el perfil molecular d'ASNC ens ajudaria a classificar-los més adientment. En els casos 16 i 22, no es correlaciona ni la histologia de CSE, ni el patró mutat de p53 amb el baix número d'ASNC. No obstant això, aquests

Capítol 4

són dos casos molt particulars ja que tot i presentar baix número d'ASNC, la única alteració que presenten és al gen *CCNE1*, que és un dels dos gens amb més associació estadística amb el CSE, i per això amb el perfil molecular s'agrupen en un grup intermig el G1C que està entre el G1B i el G2. I per últim el cas 17, que presenta una histologia de CSE, que es classifica en un altre grup intermig, el G1A, que presenta un número mitjà d'ASNC (n=7) i patró mutat de p53.

Taula 20. Detall de la classificació histològica (*histological type*), de la IHQ de p53, de la freqüència d'ASNC (*SCNA frequency*) i de la classificació mitjançant el perfil genètic (*classification by genetic profile*) per cadascun dels casos analitzats.

CASE	HISTOLOGICAL TYPE	SCNA p53 (frequency, %)	CLASSIFICATION BY GENETIC PROFILE											Cluster Analysis Groups	
			<i>BCL6</i>	<i>CCNE1</i>	<i>EV11</i>	<i>FKBP8</i>	<i>NTRK1</i>	<i>PIK3CA</i>	<i>PTPN1</i>	<i>UCL1</i>	<i>BIRC4</i>	<i>FGFR1</i>			
1	EEC	wt	3	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
2	EEC	wt	1	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
3	EEC	wt	1	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
4	EEC	ab	2	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
5	EEC	wt	1	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
6	EEC	wt	1	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
7	EEC	wt	1	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
8	EEC	wt	0	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
9	EEC	wt	3	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
10	EEC	wt	0	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
11	EEC	wt	2	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
12	EEC	wt	4	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
13	SC	ab	26	Normal	Gain	Normal	Normal	Normal	Gain	Gain	Gain	Normal	Loss	Normal	G2
14	SC	wt	2	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B
15	SC	ab	15	Gain	Gain	Gain	Gain	Gain	Gain	Normal	Gain	Loss	Normal	Normal	G2
16	SC	ab	1	Normal	Gain	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1C
17	SC	ab	7	Normal	Normal	Normal	Gain	Normal	Normal	Normal	Gain	Normal	Normal	Normal	G1A
18	SC	ab	21	Gain	Gain	Gain	Normal	Gain	Gain	Gain	Gain	Loss	Loss	Normal	G2
19	SC	wt	24	Gain	Normal	Gain	Normal	Gain	Gain	Normal	Gain	Loss	Normal	Normal	G2
20	SC	ab	18	Gain	Gain	Gain	Normal	Normal	Gain	Gain	Gain	Loss	Loss	Normal	G2
21	SC	ab	11	Normal	Gain	Gain	Gain	Gain	Normal	Gain	Gain	Normal	Loss	Normal	G2
22	SC	ab	1	Normal	Gain	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1C
23	SC	ab	31	Normal	Normal	Gain	Gain	Gain	Gain	Gain	Normal	Loss	Normal	Normal	G2
24	SC	wt	2	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	G1B

EEC (*endometrial endometrioid carcinoma*), CEE; SC (*serous carcinoma*), CSE; wt, ("wild type" pattern) patró wild type; ab ("abnormal" pattern) patró mutat.

3 HETEROGENEÏTAT INTRATUMORAL (HIT) EN EL CARCINOMA SERÓS D'ENDOMETRI

En aquest apartat es detallen els resultats obtinguts en el marc de l'objectiu específic 3: estudiar l'heterogeneïtat intratumoral (HIT) en mostres de carcinoma serós d'endometri a diferents nivells d'alteració molecular: (a) a nivell de les alteracions somàtiques en el número de còpies dels gens (ASNC) mitjançant la tècnica MLPA, i (b) a nivell de variants somàtiques (VS) mitjançant l'aplicació de *targeted sequencing* amb la tecnologia NGS.

3.1. HETEROGENEÏTAT INTRATUMORAL A NIVELL D'ASNC

3.1.1. DESCRIPCIÓ DE LES ASNC I GENS ALTERATS

Mitjançant la tècnica MLPA es va analitzar l'alteració somàtica en el número de còpies de 106 gens en 8 casos de CSE analitzant tres regions tumorals per cada cas. La figura 55 mostra els resultats globals d'aquesta anàlisi on es detalla la distribució dels valors DQ (*Dosage Quocient*) obtinguts per cadascuna de les regions tumorals analitzades ($n=24(8 \times 3)$) i per cadascun dels gens analitzats. De manera general, s'observa que hi ha un elevat percentatge de gens que presenten alteracions a nivell de guanys (valors de DQ superiors a 1,4). D'aquests els que presenten les freqüències d'alteració més elevades són *UCKL1*, *ERBB2*, *EVI1*, *PIK3CA*, *CCNE1*, *PTPN1*, *NTRK1*, *MYBL2*, *AURKA*, *TP53*, *BCL6* i *FKBP8*. En aquestes mostres s'observa una menor freqüència de gens alterats amb pèrdues, tot i això, cal remarcar els tres gens que presenten les freqüències més elevades, que són *BIRC4*, *BIRC1* i *FGFR1*.

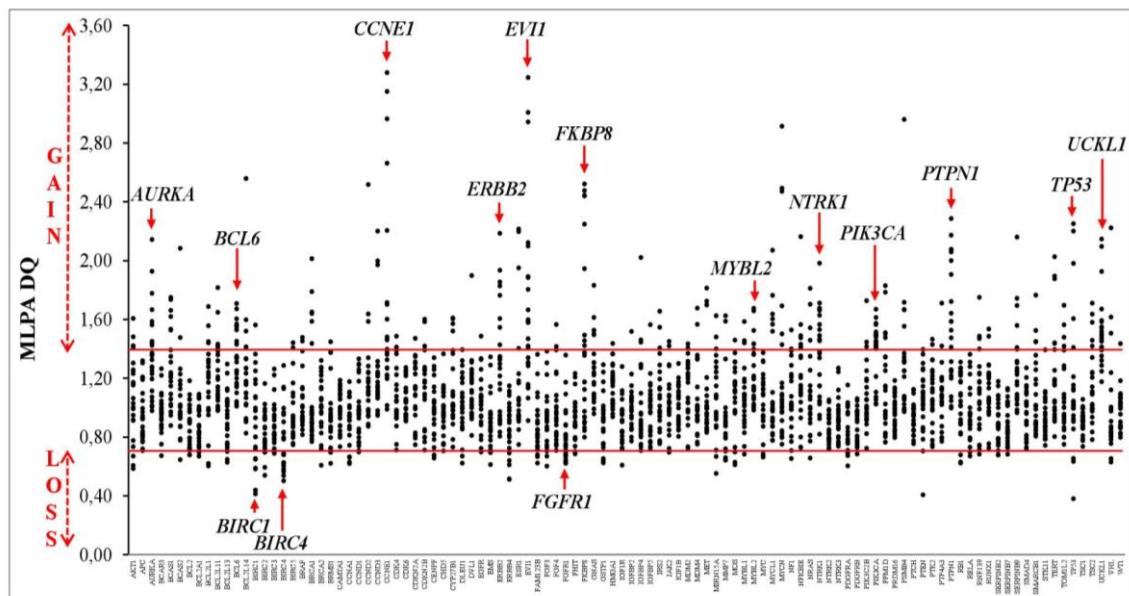


Figura 55. Perfil de les ASNC de les 24 regions tumorals (8 casos CSE per tres regions analitzades de cadascun) per cadascun dels gens analitzats. L'eix y indica els valors de MLPA DQ, on el rang normal està entre 0,7 i 1,4 (rang definit per les línies vermelles). L'eix x mostra tots els gens analitzats ($n = 106$). Les fletxes vermelles indiquen els gens amb unes freqüències d'alteració superior al 50%.

De manera global, en tots els casos (n=8) es van observar ASNC. De mitjana es van observar 22 gens alterats per cas, amb un rang que oscil·lava entre 8 i 36. D'aquests 22, es va observar una mitjana d'11 gens amb guanys (rang de 2 a 22) i de 4 amb pèrdues (rang de 1 a 11) (figura 56).

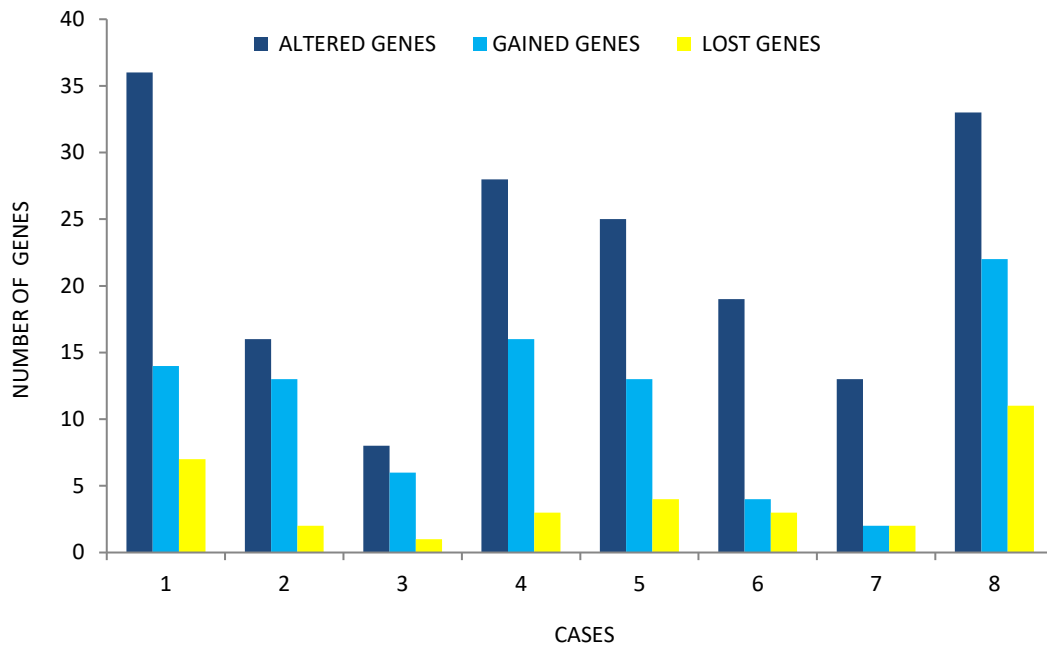


Figura 56. Distribució del número de gens alterats (eix y), indicant el número total de gens alterats (blau fort), el número de gens amb guanys (blau clar) i el número de gens amb pèrdua (groc), per cada cas (eix x).

Capítol 4

Dels 106 loci cromosòmics analitzats es van detectar ASNC en 82 d'ells (77%), en el 67% dels quals es van detectar guanys, en el 22% pèrdues i en l'11% restant es van detectar pèrdues i guanys depenent de la mostra (figura 57).

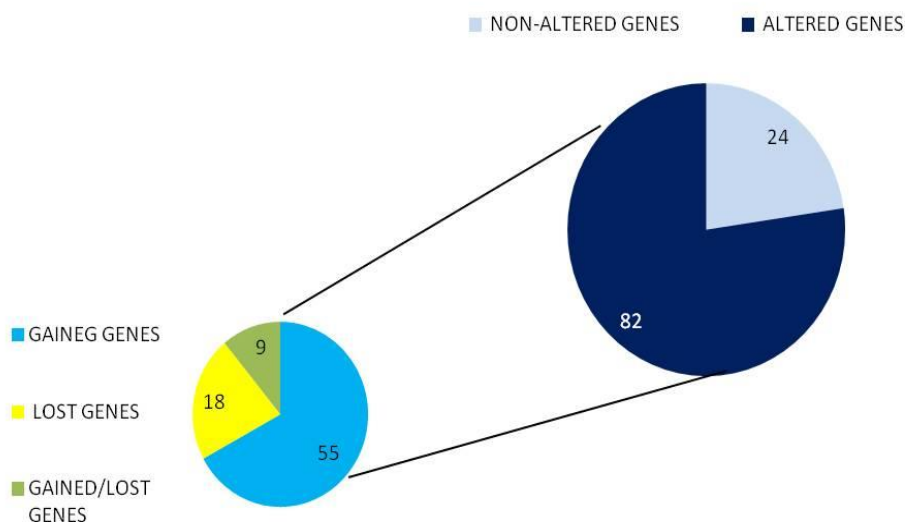


Figura 57. Gràfic que representa el número de gens alterats i no alterats, indicant també el número dels gens amb guanys, pèrdues o gens amb guanys i pèrdues depenent de la mostra.

Per cada gen es va quantificar la freqüència d'alteració (AF), que era el resultat del quocient entre el nombre de casos alterats respecte al nombre total de casos analitzats ($n=8$). Els gens més freqüentment alterats amb guanys van ser *UCKL1* (88%), *ERBB2*, *EVI1* i *PIK3CA* (75%), *CCNE1*, *PTPN1* i *NTRK1* (63%), i *MYBL2*, *AURKA*, *BCL6* i *FKBP8* (50%). Les regions on es troben localitzats aquests gens són les següents: 20q12 – q13.3 (*UCKL1*, *PTPN1*, *MYBL2* i *AURKA*), 3q26 –q27 (*EVI1*, *PIK3CA* i *BCL6*), 17q21.1 (*ERBB2*), 19q12 (*CCNE1*), 19p13.11 (*FKBP8*) i 1q23.1 (*NTRK1*). Els gens més freqüentment alterats amb pèrdues van ser *BIRC4* (63%), *BIRC1* (50%), *FGFR1* (50%) i *ERBB4* (38%), localitzats a les regions Xq25, 5q13.2, 8p11.2 i 2q33, respectivament (figura 58). En particular, en el cas del gen *TP53* es van detectar pèrdues (25%) o guanys (25%) depenent de la mostra. Entre tots aquests gens, cal destacar que *PIK3CA*, *CCNE1*, *BIRC1* i *FGFR1* es van trobar alterats en totes les regions del tumor analitzades sense presentar heterogeneïtat intratumoral.

CHR. LOC.	GENE	SAMPLES																								AF(%)	HF(%)			
		1			2			3			4			5			6			7			8							
		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3					
20q13.33	<i>UCKL1</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	88	14
17q21.1	<i>ERBB2</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	75	67
3q26	<i>EVII</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	75	33
3q26.32	<i>PIK3CA</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	75	0
5q13.2	<i>BIRC1</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	63	0
Xq25	<i>BIRC4</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	63	40
19q12	<i>CCNE1</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	63	0
1q23.1	<i>NTRK1</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	63	20
20q13.13	<i>PTPNI</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	63	40
20q13.31	<i>AURKA</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	50	50
3q27	<i>BCL6</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	50	25
8p11.2	<i>FGFR1</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	50	0
19p13.11	<i>FKBP8</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	50	25
20q12	<i>MYBL2</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	50	75
17p13.1	<i>TP53</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	50	50
14q32.33	<i>AKT1</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	38	33
20q13.2	<i>BCAS1</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	38	67
20q11.1	<i>BCL2L1</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	38	0
2q33	<i>ERBB4</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	38	67
20q13.32	<i>GNAS</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	38	67
6p21.1	<i>NFKBIE</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	38	33
1q21.3	<i>PSMB4</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	38	33
6p25.2	<i>SERPINB9</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	38	33
5p25	<i>TERT</i>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	38	67

Figura 58. Distribució de les ASNC detectades en els gens més freqüentment alterats. L'eix x mostra cadascuna de les regions tumorals analitzades. A la part esquerra de l'eix y s'indica la localització cromosòmica associada a cada gen (*Chr. Loc.*), i a la part dreta de l'eix y, s'indiquen les freqüències d'alteració (AF) i d'heterogeneïtat (HF). Els quadres blaus indiquen guany, els grocs pèrdues, els blancs una dotació normal i els rectangles negres indiquen les mostres on s'han detectat les ASNC en heterogeneïtat.

3.1.2. ANÀLISI DE L'HETEROGENEÏTAT INTRATUMORAL

Pel que fa a l'anàlisi de l'HIT es va observar que cinc (62,5%) dels vuit casos presentaven heterogeneïtat intratumoral de manera significativa (veure criteris heterogeneïtat, material i mètodes). D'entre els gens alterats, de manera global, es va observar una mitjana de 7 gens heterogenis per cas, amb un rang que anava de l'1 als 15. D'aquests, es va observar una mitjana de 7 gens heterogenis amb guanys i de 1 amb pèrdues. La figura 59 mostra la distribució del nombre de gens heterogenis respecte el nombre total de gens alterats per cada cas.

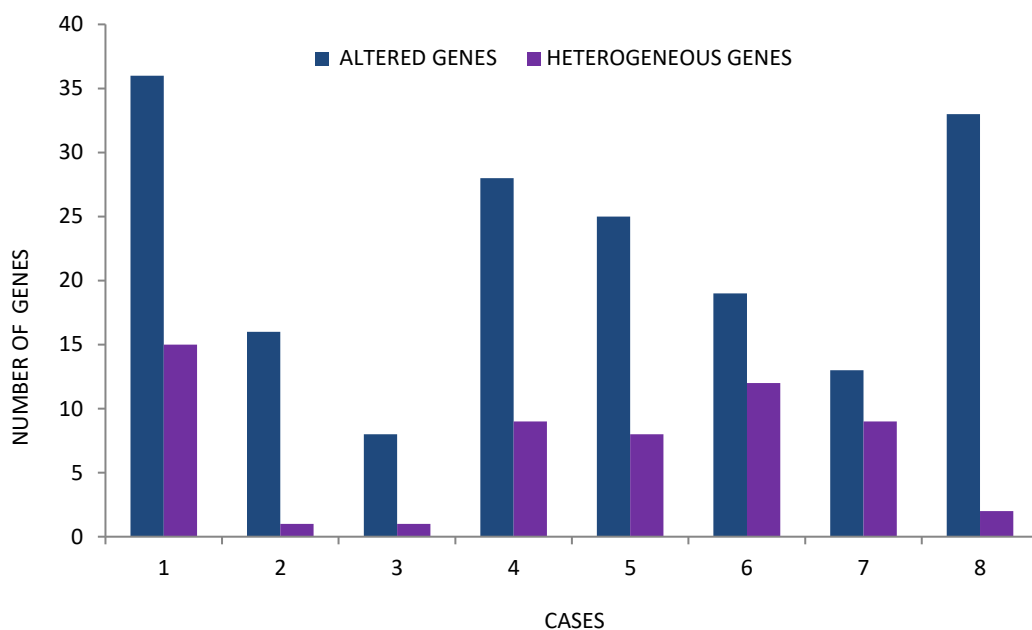


Figura 59. Distribució del nombre total de gens alterats (blau fort) i de gens heterogenis (lila fort) per cada cas analitzat.

En els 5 casos que presentaven HIT de mitjana es van detectar 10 gens heterogenis per cas (rang de 8 a 14), dels quals el 90% mostrava guanys i el 10% pèrdues. La figura 60 mostra els nombre de gens heterogenis per cas.

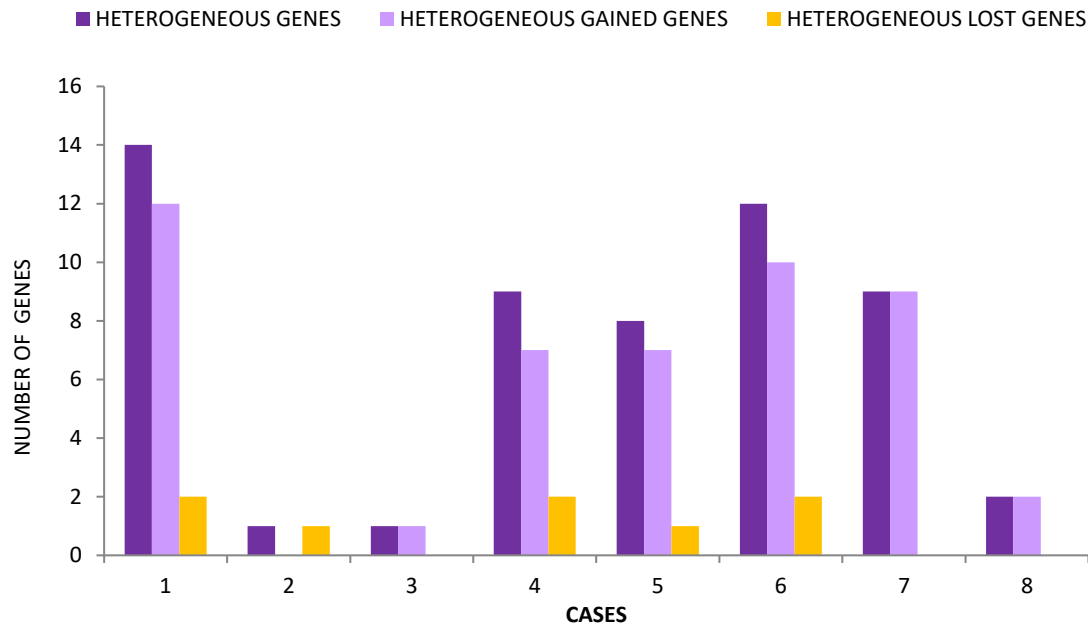


Figura 60. Distribució del nombre total de gens heterogenis (lila fort), del nombre de gens heterogenis amb guanys (lila clar) i del nombre de gens heterogenis amb pèrdua (groc) per cada cas analitzat.

D'entre aquests 5 casos, el cas 1 és el que presenta més gens en heterogeneïtat. Aquest cas s'il·lustra a la figura 61 on es mostra la distribució dels valors DQ de cadascun dels gens analitzats per cada regió de tumor, i s'indiquen els 14 gens heterogenis. D'aquests cal destacar el gen *ERBB2* que és una potencial diana terapèutica.

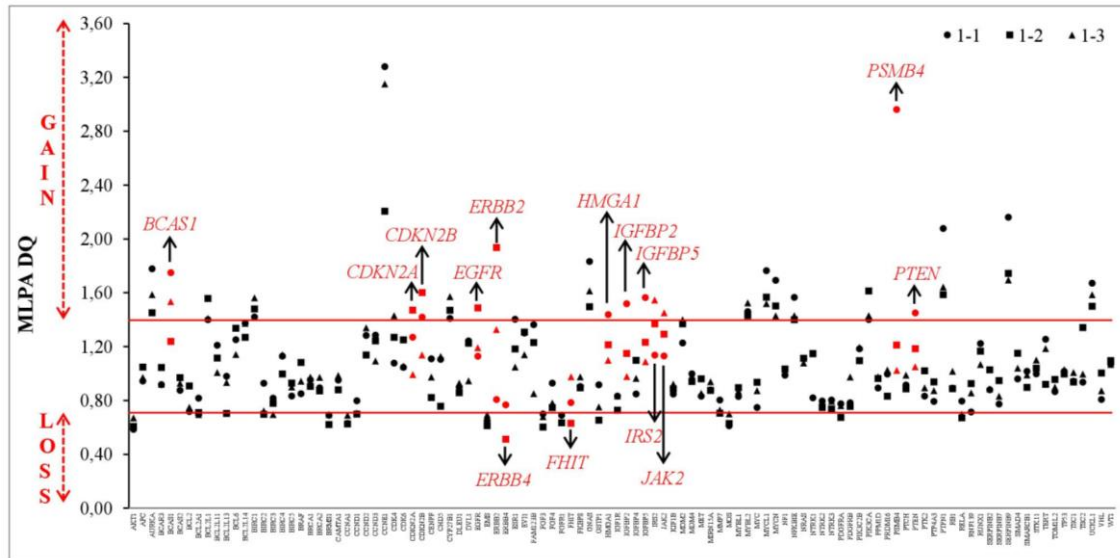


Figura 61. Exemple d'un cas heterogeni, el cas 1. L'eix y mostra els valors DQ on el rang normal està entre 0,7 i 1,4 (rang definit per les línies vermelles). L'eix x mostra tots els gens analitzats (n = 106). Les fletxes negres indiquen els gens heterogenis en base als criteris d'heterogeneïtat descrits a l'apartat de material i mètodes.

Dels 82 gens alterats es va observar que 40 d'ells presentaven heterogeneïtat. D'entre aquests gens, 31 van ser detectats amb guanys, 5 amb pèrdua i 4 amb guany/pèrdua depenent de la mostra. Cal esmentar que hi ha haver 16 gens que es van observar sempre en heterogeneïtat.

Per a cada gen es va calcular la freqüència d'heterogeneïtat (HF), que era el resultat del quocient entre el nombre de casos que presentaven heterogeneïtat respecte al total de casos alterats (figura 58). Els gens que presentaven més heterogeneïtat a nivell de guanys van ser: *MYBL2* (75%), *ERBB2* (67%), *AURKA* (50%), *PTPN1* (40%) i *EVI1* (33%); i a nivell de pèrdues van ser *ERBB4* (67%) i *BIRC4* (40 %). El gen *TP53*, que es va detectar amb guanys i pèrdues depenent del cas, també es va observar en heterogeneïtat amb un HF del 50%.

3.1.2.1 VALIDACIÓ DE L'HETEROGENEÏTAT INTRATUMORAL DEL GEN *ERBB2*

Amb la tècnica MLPA, es va observar una freqüència d'alteració (AF) del gen *ERBB2* a nivell d'ASNC del 75% (6 casos alterats de 8 analitzats), i una freqüència d'heterogeneïtat (HF) del 67% (4 casos amb heterogeneïtat de 6 casos alterats). Per la seva rellevància terapèutica es va validar la presència d'heterogeneïtat intratumoral del gen *ERBB2* en els casos 1, 3, 5 i 7 mitjançant les tècniques d'IHQ i de SISH. En tres dels casos (75%) (1, 3 i 7) es va observar concordança de resultats entre les tres tècniques utilitzades (MLPA, IHQ i SISH) (figures 62, 63, 64). En canvi, en el cas 5 només es va observar presència d'heterogeneïtat mitjançant la MLPA. Amb la IHQ i el SISH no es va observar alteració en aquest gen (Figura 65). Amb aquests resultats es va poder demostrar l'heterogeneïtat intratumoral a nivell d'ASNC del gen *ERBB2* en un 75% dels casos.

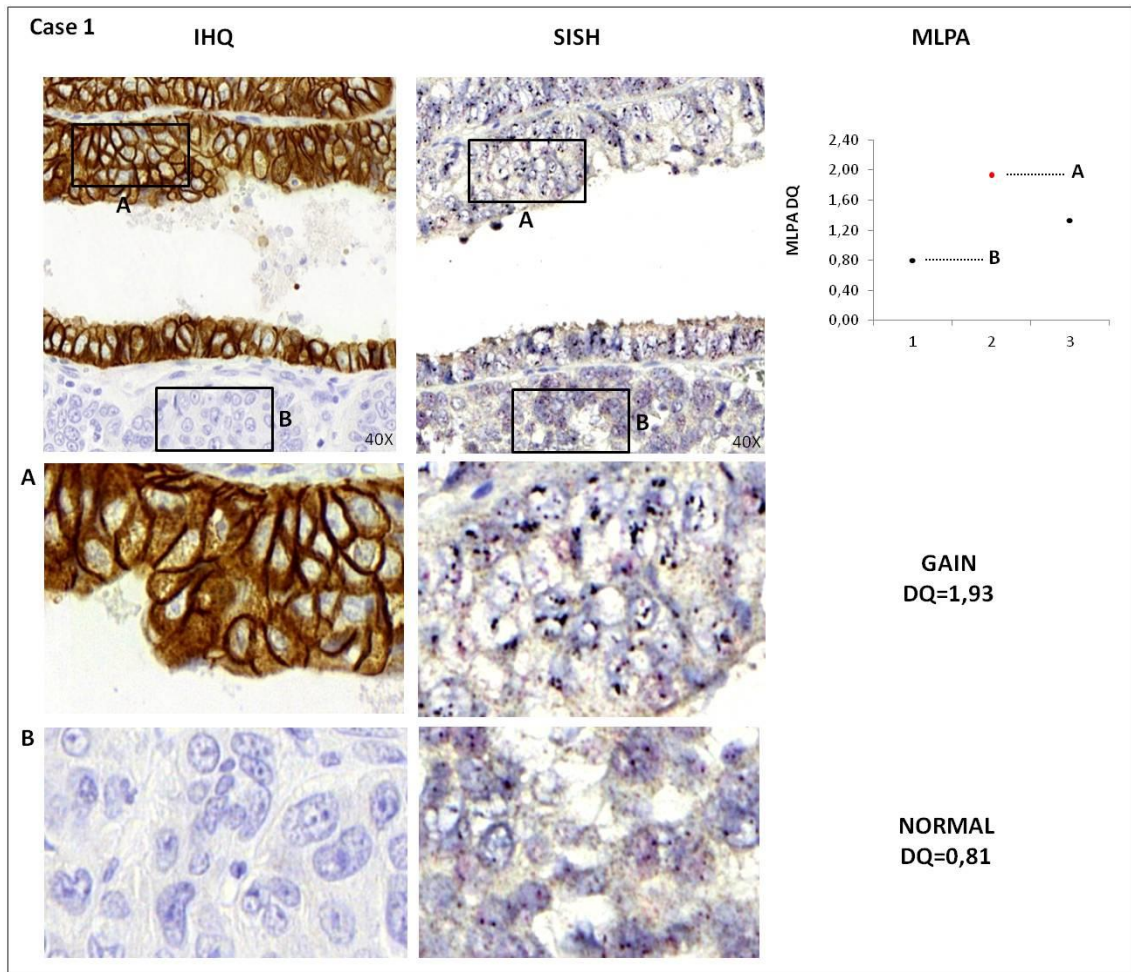


Figura 62. Imatge que mostra la presència d'heterogeneïtat intratumoral del gen *ERBB2* en el cas 1 mitjançant les tècniques d'IHQ, SISH i MLPA. A) Visualització de l'àrea tumoral que presenta guany d'*ERBB2* per IHQ i SISH, i que està en concordança amb el resultat de guany per MLPA. B) Visualització de l'àrea tumoral que presenta un estat normal d'*ERBB2* i que està en concordança amb el resultat normal per MLPA.

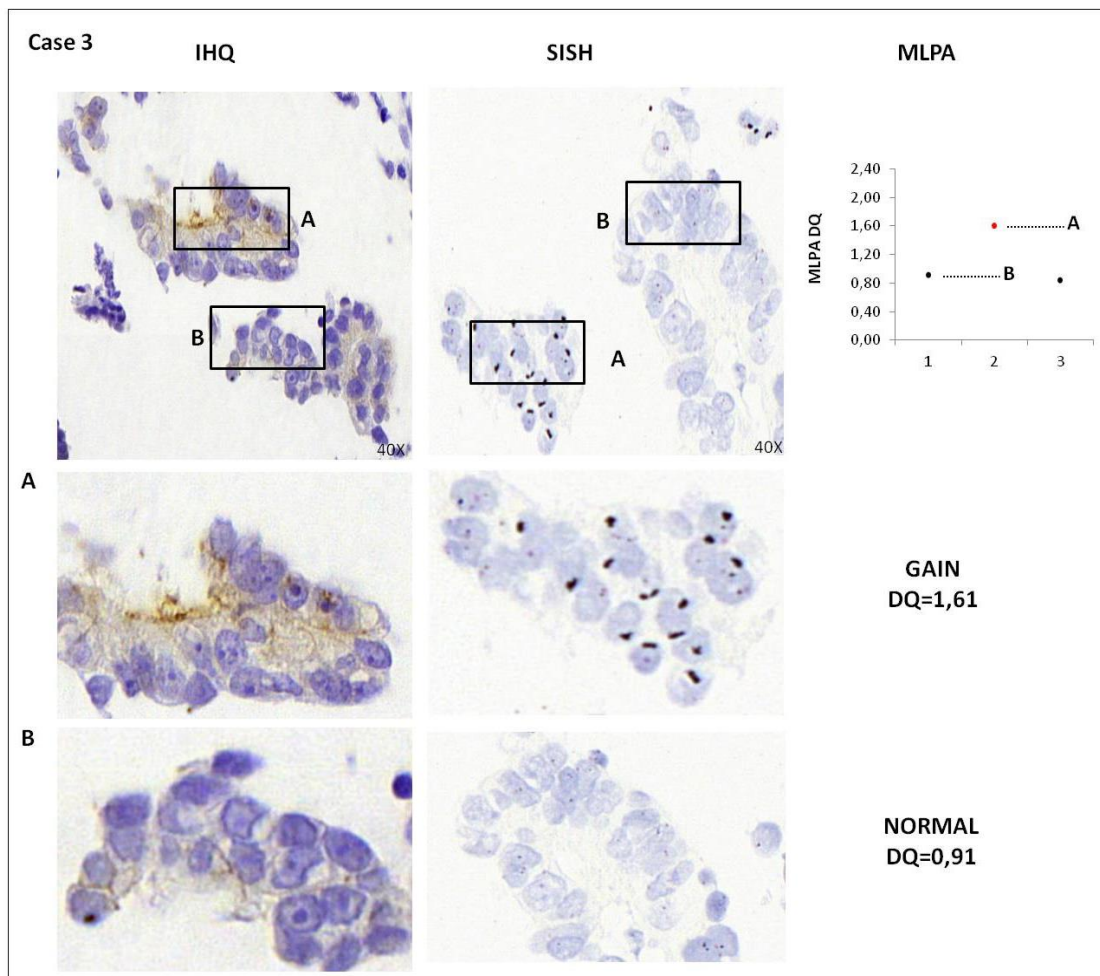


Figura 63. Imatge que mostra la presència d'heterogeneïtat intratumoral del gen *ERBB2* en el cas 3 mitjançant les tècniques d'IHQ, SISH i MLPA. A) Visualització de l'àrea tumoral que presenta guany d'*ERBB2* per IHQ i SISH, i que està en concordança amb el resultat de guany per MLPA. B) Visualització de l'àrea tumoral que presenta un estat normal d'*ERBB2* i que està en concordança amb el resultat normal per MLPA.

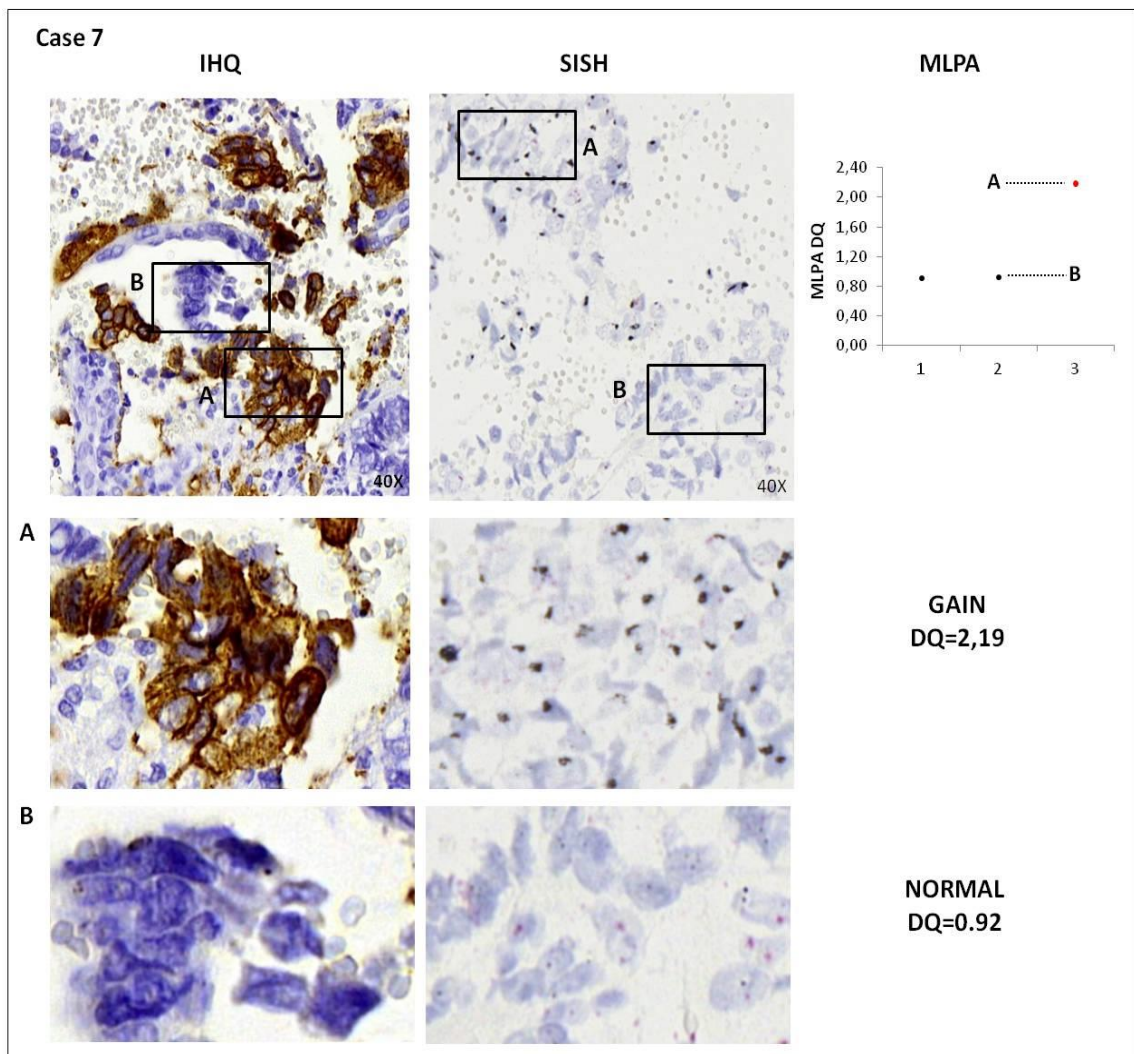


Figura 64. Imatge que mostra la presència d'heterogeneïtat intratumoral del gen *ERBB2* en el cas 7 mitjançant les tècniques d'IHQ, SISH i MLPA. A) Visualització de l'àrea tumoral que presenta guany d'*ERBB2* per IHQ i SISH, i que està en concordança amb el resultat de guany per MLPA. B) Visualització de l'àrea tumoral que presenta un estat normal d'*ERBB2* i que està en concordança amb el resultat normal per MLPA.

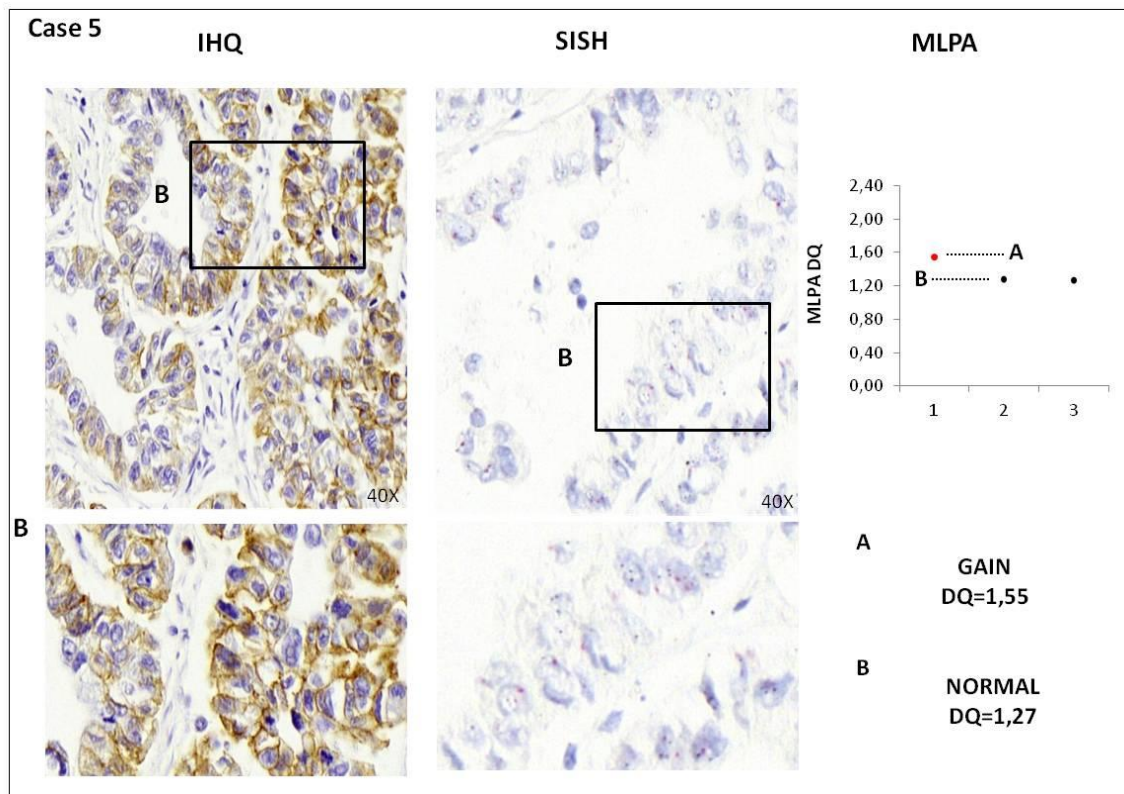


Figura 65. Imatge que mostra la presència d'heterogeneïtat intratumoral del gen *ERBB2* en el cas 5 mitjançant MLPA. A) Regió tumoral observada amb guany *d'ERBB2* per MLPA. B) Visualització de l'àrea tumoral que presenta un estat normal *d'ERBB2* per IHQ i SISH, i que està en concordança amb el resultat normal per MLPA.

3.2 HETEROGENEÏTAT INTRATUMORAL A NIVELL DE VARIANTS SOMÀTIQUES

3.2.1 DESCRIPCIÓ DEL LES VARIANTS SOMÀTIQUES I GENS ALTERATS

Durant el procés de seqüenciació amb la tecnologia NGS es van anar avaluant tots els paràmetres de qualitat per tal d'assegurar-nos una correcta identificació de les variants somàtiques. Així es van obtenir valors adequats de la densitat de *clusters*, de profunditat de lectura, entre molts altres. Per totes les mostres es van obtenir unes dades de qualitat molt similars. Quant a les mètriques de NGS obtingudes de totes les mostres, es va obtenir una mitjana per mostra de 2 825 883 lectures, i de 1470X de profunditat de lectura. La cobertura mitjana de les lectures sobre el genoma de referència GRCh38 / hg38 va ser del 98,38%. Cal dir que tant els valors de nombre de lectures com de profunditat de lectura van ser molt similars entre les mostres no tumorals i les tumorals. A la taula 21, es detallen els valors d'aquestes mètriques per cadascuna de les mostres analitzades.

Taula 21. Detall dels valors de número total de lectures, de la mitjana de profunditat de lectura i del percentatge mitjà de cobertura per cadascuna de les mostres analitzades.

SAMPLE	TOTAL READS	MEAN READ DEPTH	% HIGH QUALITY MAPPED READS
1-NT	2712498	1366,73	98,97
1.1-T	2659436	1408,4	99,26
1.2-T	2840022	1402,1	99,21
1.3-T	3188830	1550,87	98,98
2-NT	2231814	1119,64	97,12
2.1-T	2355950	1187	97,19
2.2-T	1746264	872,45	96,87
2.3-T	1612624	790,81	95,84
3-NT	1662798	852,21	98,41
3.1-T	1699924	867,51	98,05
3.2-T	2051564	1041,64	97,60
3.3-T	1738914	897,12	98,20
4-NT	2148976	1093,73	98,09
4.1-T	1618364	831,69	98,52
4.2-T	1646842	843,17	98,25
4.3-T	1687670	867,12	98,52
5-NT	4962612	2509,67	98,83
5.1-T	3151148	1454,86	98,98
5.2-T	2890850	1321,68	99,00
5.3-T	3224582	1592,12	98,73
6-NT	2456272	1294,76	99,14
6.1-T	3191244	1577,26	98,89
6.2-T	5971550	3032,33	98,67
6.3-T	3420826	1801,94	99,07
7-NT	6906944	3470,52	99,11
7.1-T	4894570	2580,91	99,16
7.2-T	3399234	1692,85	99,04
7.3-T	5452866	2679,02	99,14
8-NT	2050068	1048,45	98,45
8.1-T	1717170	871,85	97,63
8.2-T	1554736	800,36	97,95
8.3-T	1581082	799,48	97,48

NT: no tumoral; T: tumoral

Les variants somàtiques es van identificar i classificar seguint els criteris descrits a l'apartat de materials i mètodes. En les 8 mostres analitzades es van identificar un total de 23 variants somàtiques que afectaven a 11 gens. El nombre mitjà de variants somàtiques per mostra va ser de 3, amb un rang que oscil·lava entre 1 i 5. Aquestes variants, segons el seu efecte sobre el transcrit, van ser classificades com *missense* (n = 15), *frameshift* (n = 6), *stop gained* (n = 6), *inframe deletion* (n = 1) i *splice acceptor* (n = 1). A la taula 22 es descriuen totes les variants detectades en relació a cada mostra i es detalla de cadascuna d'elles el gen implicat (*gene symbol*), la freqüència al·lèlica (*tumor vaf*), la profunditat de lectura (*tumor depth*), la nomenclatura de la variant a nivell genòmic (*HGVSc*), la nomenclatura de la variant a nivell de la proteïna (*HGVSp*), l'efecte de la variant sobre el transcrit (*variant effect*), i si ha estat reportada o no a la base de dades COSMIC (*COSMIC data base*).

Taula 22. Descripció de totes les variants somàtiques detectades per cadascuna de les mostres analitzades

SAMPLE ID	GENE SYMBOL	TUMOR VAF	TUMOR DEPTH	HGVS_c_NAME	HGVS_p_NAME	VARIANT EFFECT	COSMIC DATABASE
1	<i>PIK3CA</i>	0.95	1985	c.344G>C	p.Arg115Pro	missense	YES
	<i>TP53</i>	0.889	1658	c.586C>T	p.Arg196Ter	stop_gained	YES
2	<i>CCND1</i>	0.903	648	c.686delT	p.Leu229ProfsTer8	frameshift	NO
	<i>MAP3K4</i>	0.863	1397	c.2207delT	p.Ile736LysfsTer24	frameshift	NO
	<i>PIK3CA</i>	0.531	1811	c.316G>C	p.Gly106Arg	missense	YES
	<i>PIK3CA</i>	0.53	1810	c.333G>C	p.Lys111Asn	missense	YES
	<i>TP53</i>	0.726	782	c.821T>C	p.Val274Ala	missense	YES
3	<i>PPP2R1A</i>	0.251	1633	c.536C>G	p.Pro179Arg	missense	YES
	<i>TP53</i>	0.31	300	c.1024_1034del11pb	p.Arg342Ter	frameshift	YES
4	<i>CDH19</i>	0.227	387	c.1820_1821delTC	p.Ile607AsnfsTer22	frameshift	NO
	<i>PIK3CA</i>	0.233	2246	c.263G>A	p.Arg88Gln	missense	YES
	<i>PPP2R1A</i>	0.517	1601	c.536C>G	p.Pro179Arg	missense	YES
	<i>TAF1</i>	0.882	537	c.408dupC	p.Ser137LeufsTer4	frameshift	NO
	<i>TP53</i>	0.788	1039	c.724T>C	p.Cys242Arg	missense	YES
5	<i>CSMD3</i>	0.392	917	c.1552C>A	p.Leu518Ile	missense	NO
6	<i>ERBB2</i>	0.344	1276	c.722C>T	p.Ala241Val	missense	YES
	<i>TP53</i>	0.299	1358	c.376-1G>A		splice_acceptor	YES
7	<i>FOXA2</i>	0.222	694	c.903dupC	p.Glu302ArgfsTer66	frameshift	NO
	<i>TP53</i>	0.352	3082	c.536A>T	p.His179Leu	missense	YES
	<i>TP53</i>	0.376	1437	c.329G>T	p.Arg110Leu	missense	YES
8	<i>MK167</i>	0.21	1876	c.4892G>A	p.Arg1631Gln	missense	YES
	<i>PIK3CA</i>	0.495	503	c.353G>A	p.Gly118Asp	missense	YES
	<i>TP53</i>	0.968	2465	c.701A>G	p.Tyr234Cys	missense	YES

Per cada gen es va quantificar la freqüència d'alteració (AF), que era el resultat del quocient entre el nombre de casos alterats respecte al nombre total de casos analitzats (n=8). Els gens més freqüentment alterats van ser *TP53* (88%), *PIK3CA* (50%) i *PPP2R1A* (25%) (figura 66). Totes les variants somàtiques que afectaven aquests gens van ser classificades com patogèniques. A més a més, aquestes variants es van detectar en les tres regions tumorals, per la qual cosa aquestes alteracions no estaven afectades per l'HIT. Hi va haver dos casos on es va detectar una variant d'*ERBB2* i una de *CSMD3* també de manera homogènia.

De les variants detectades al gen *TP53* el 12 % van ser classificades com *stop gained* i el 88% com *missense*. Aquestes variants estaven localitzades als exons 4, 5, 6, 7, 8 i 10. Concretament, els *hotspots* de la proteïna afectats per aquestes variants van ser el R110, H179, R196, Y234, C242, V274, R342. Addicionalment, es va detectar una variant que afectava la regió d'*splicing* de l'intró 4 (c. 376-1G>A). Pel que fa al gen *PIK3CA*, en tres casos es van observar variants *missense* a l'exó 2 i en un cas en el primer codó de l'exó 3. Cal remarcar que en cap mostra es va detectar cap variant als *hotspot* típics dels exons 9 i 20 d'aquest gen. En referència al gen *PPP2R1A*, en un dels casos es va identificar la variant recurrent *missense* P179R.

GENE	SAMPLES																								AF (%)	HF (%)
	1			2			3			4			5			6			7			8				
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3		
<i>TP53</i>	R196*			V274A			R342X			C242R						c.376-1G>A			H179L, R110L			Y234C			88	0
<i>PIK3CA</i>	R115P			G106R, K111N						R88Q												G118D			50	0
<i>PPP2R1A</i>							P179R			P179R															25	0
<i>ERBB2</i>																A241V									13	0
<i>MAP3K4</i>				T736X																					13	100
<i>CCND1</i>				L229X																					13	100
<i>TAF1</i>													H136X												13	100
<i>CDH19</i>										I607X															13	100
<i>C5MD3</i>													L518I												13	0
<i>FOXA2</i>																			E302X						13	100
<i>MKI67</i>																						R1631Q, R1631C			13	100

Figura 66. Distribució de les variants somàtiques detectades. L'eix x mostra cadascuna de les regions tumorals analitzades. A la part esquerra de l'eix y s'indiquen els gens, i a la part dreta de l'eix y, s'indiquen les freqüències d'alteració (AF) i d'heterogeneïtat (HF). Els rectangles negres indiquen les mostres on s'han detectat les variants en heterogeneïtat.

3.2.1.1 VALIDACIÓ VARIANTS SOMÀTIQUES PER SEQÜENCIACIÓ SANGER

Mitjançant la seqüenciació Sanger es van validar les variants identificades per l'anàlisi de NGS als gens més freqüentment alterats, *TP53* i *PIK3CA*. Les figures 67, 68 i 69 mostren l'exemple de la validació de 3 d'aquestes variants somàtiques.

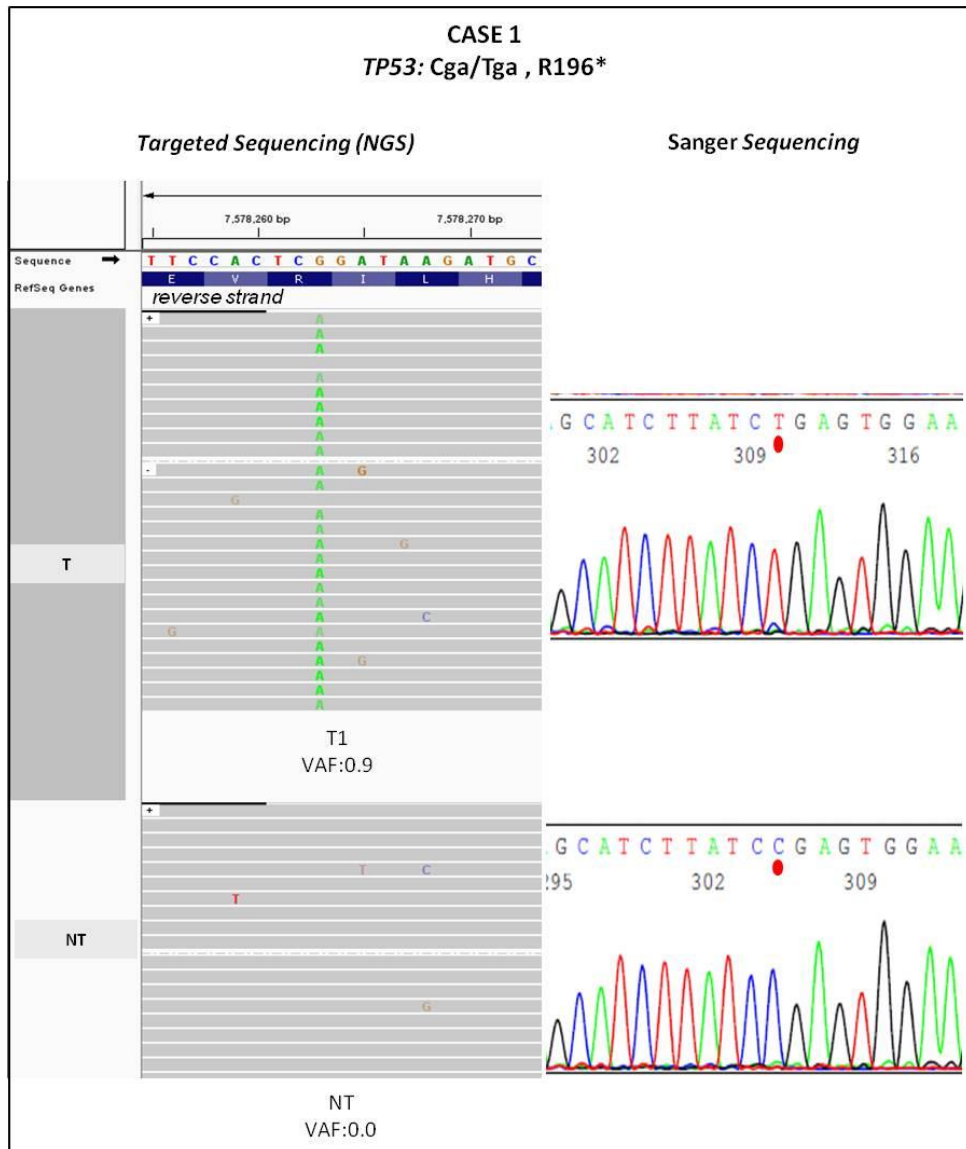


Figura 67. Composició que il·lustra la validació per seqüenciació Sanger de la variant R196* del gen *TP53* detectada mitjançant *targeted sequencing* (NGS) al cas 1. A l'esquerra de la imatge es mostra la visualització de la variant mitjançant l'IGV (la seqüència és la *reverse* 3'-5'), on T indica la mostra tumoral, NT la mostra no tumoral, i VAF la freqüència al·lèlica de la variant. A la dreta es mostren les seqüències obtingudes amb la seqüenciació Sanger (la seqüència és la *forward* 5'-3'), on el punt vermell indica la localització de la variant.

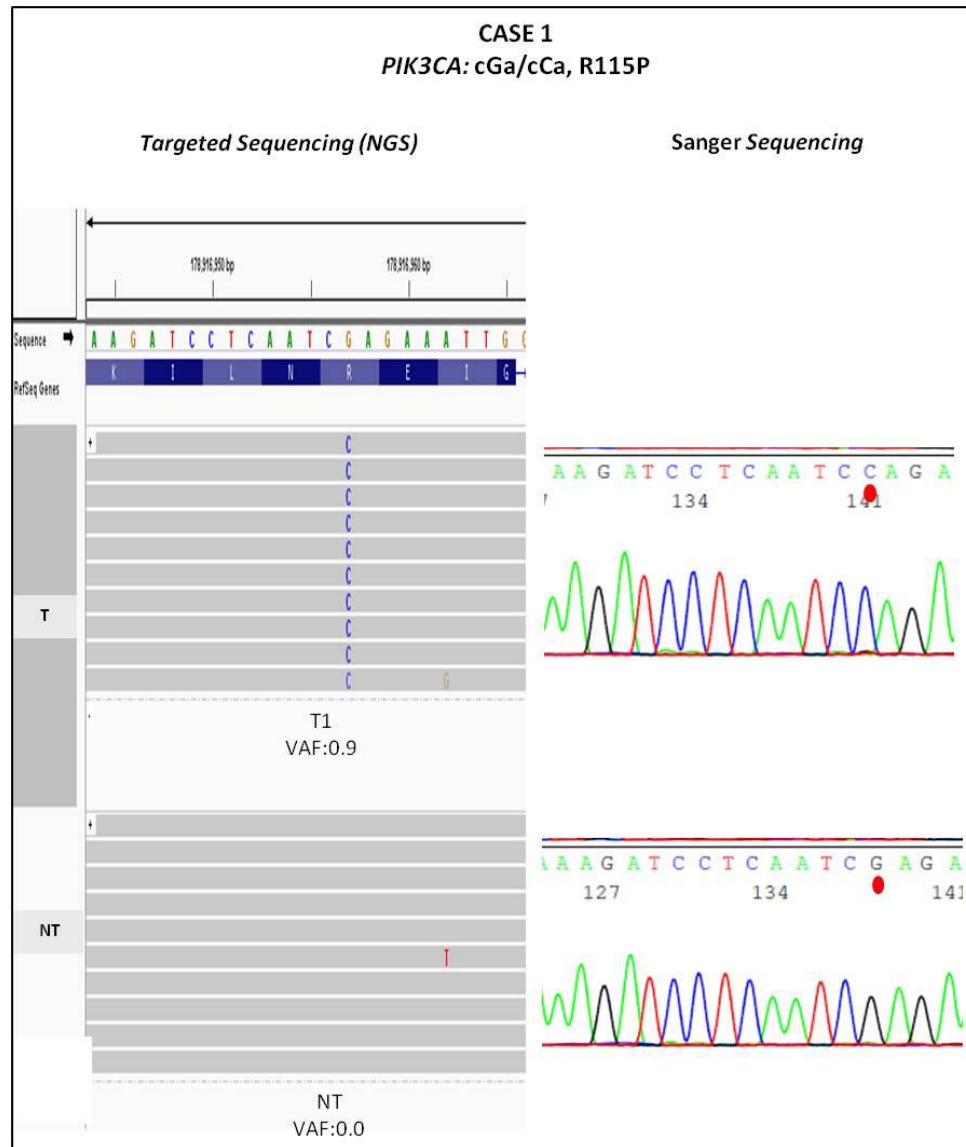


Figura 68. Composició que il·lustra la validació per seqüenciació Sanger de la variant R115P del gen *PIK3CA* detectada mitjançant *targeted sequencing* (NGS) al cas 1. A l'esquerra de la imatge es mostra la visualització de la variant mitjançant l'IGV, on T indica la mostra tumoral, NT la mostra no tumoral, i VAF la freqüència al·lèlica de la variant. A la dreta es mostren les seqüències obtingudes amb la seqüenciació Sanger, on el punt vermell indica la localització de la variant.

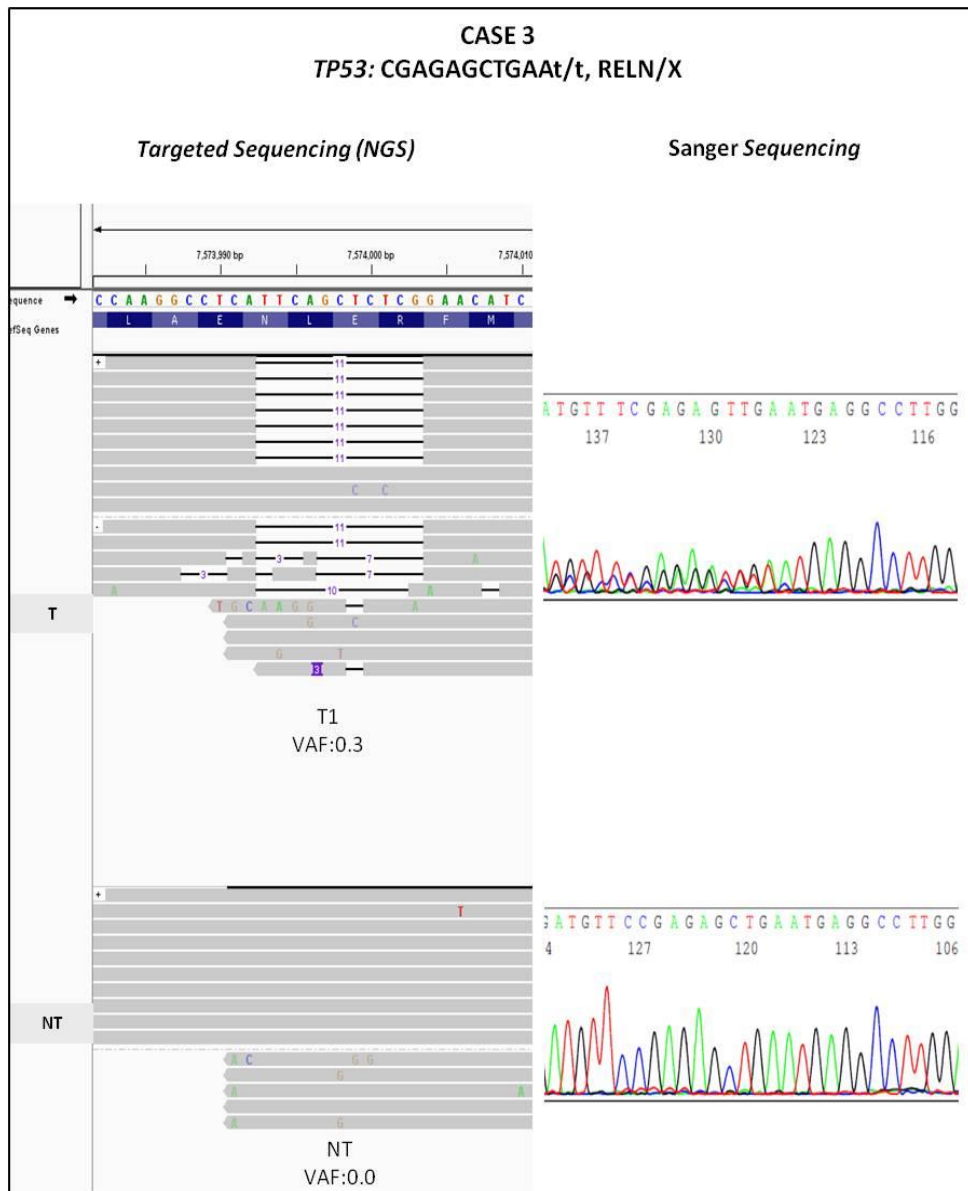


Figura 69. Composició que il·lustra la validació per seqüenciació Sanger de la variant RELN/X del gen *TP53* detectada mitjançant *targeted sequencing* (NGS) al cas 3. A l'esquerra de la imatge es mostra la visualització de la variant mitjançant l'IGV (la seqüència és la *reverse* 3'-5'), on T indica la mostra tumoral, NT la mostra no tumoral, i VAF la freqüència al·lèlica de la variant. A la dreta es mostren les seqüències obtingudes amb la seqüenciació Sanger (la seqüència és la *forward* 5'-3').

3.2.2 ANÀLISI DE L'HETEROGENEÏTAT INTRATUMORAL

Per a cada gen es va calcular la freqüència d'heterogeneïtat (HF), que era el resultat del quocient entre el nombre de casos que presentaven heterogeneïtat respecte al total de casos alterats (n=8) (figura 66). Els gens que van mostrar heterogeneïtat van ser *MAP3K4*, *CCND1*, *TAF1*, *CDH19*, *FOXA2* i *MKI67*, amb unes HF del 13%. Les variants observades en aquests gens van ser classificades com a variants de significació desconeguda (VUS).

CAPÍTOL 5

DISCUSSIÓ

El càncer d'endometri és el càncer ginecològic més comú en els països desenvolupats, i en estudis recents s'ha observat que la seva prevalència està augmentant.⁴⁷ La introducció d'aquesta tesi descriu l'efecte que ha tingut la revolució tecnològica dels últims anys, sobretot amb la irrupció de la tecnologia NGS, en el coneixement molecular més profund del càncer d'endometri, i detalla els coneixements actuals d'aquest càncer. Tota aquesta informació ha permès confirmar que aquest càncer està constituït per un grup molt heterogeni de tumors, no només a nivell histològic, sinó a nivell molecular, i també a nivell dels propis tumors individuals. Això ha posat de manifest l'heterogeneïtat d'aquest tumor i la necessitat d'establir sistemes d'estratificació més acurats per poder predir els casos d'alt risc.

S'ha passat de classificar el CE d'una manera senzilla i dualística (1983), amb el tipus I i II, a una classificació molecular (2013), proposada pel TCGA, que ha categoritzat el CE en quatre grups moleculars amb informació pronòstica: 1) POLE *ultramutated*, 2) MSI *hypermuted*, 3) CN-low i 4) CN-high/*serous-like*).^{55,59} En aquesta classificació els grups 1, 2 i 3 estan representats majoritàriament per carcinomes d'histologia endometrioides (CEE), i el grup 4 majoritàriament per carcinomes serosos (CSE). Segons la classificació histològica, els CEE de grau 1 i 2 solen associar-se amb un pronòstic favorable, en canvi els tumors CSE s'associen a un pitjor pronòstic.^{152, 153} El risc de recurrència està al voltant d'un 20% en els casos de CEE, i en un 50% en els casos CSE. El programa SEER (*Surveillance, Epidemiology, and End Results*) va comparar els tumors de tipus CSE i CEE de grau 3 i va confirmar que aquests tipus tumorals representaven entre el 10% i el 15% dels CE, respectivament. Tot i això, són responsables d'entre el 39% i el 27% de les morts per càncer d'endometri, respectivament.¹⁵⁴ Recentment, també s'ha generat més coneixement sobre l'entitat dels carcinomes d'endometri d'alt grau (CEAG).^{58,155} S'ha demostrat que aquest és un grup peculiar, molt heterogeni, i que inclou la representació dels quatre grups moleculars del TCGA, incloent el subgrup de tumors *serous-like*, caracteritzats per tenir el pronòstic més advers. A més a més, en els darrers anys també s'ha posat en coneixement la importància que pot jugar l'heterogeneïtat intratumoral en els comportaments més agressius dels tumors i en la resistència als tractaments, com seria el cas del CSE.²⁵

L'existència dels carcinomes d'endometri d'alt grau (CEAG) demostra que el CE és complex i reflexa que en certs casos es difícil poder donar un diagnòstic segur. Alguns d'aquests casos (30%) no presenten característiques histològiques típiques i els resultats immunohistoquímics tampoc són conclouents, d'aquesta manera és difícil classificar-los adientment.^{57,155-157} A més a més, també s'ha observat que hi ha un percentatge de casos (20%) amb histologia de baix grau que s'acaben comportant com casos d'alt risc.¹⁵⁸ Tot i que la reproductibilitat de la classificació histològica és de

Capítol 5

moderada a excel·lent (0,62-0,87) en els casos de baix grau, en els tumors d'alt grau com els CEE grau 3 o els CSE existeix una baixa reproductibilitat inter-observador.⁵⁷

En l'actualitat, hi ha certa controvèrsia sobre la possibilitat d'incloure en la pràctica clínica diària la classificació molecular, i que aquesta substitueixi la classificació histològica tradicional. Un dels punts a favor de la classificació molecular és que permet fer una classificació objectiva, sobrepassant l'inadequat acord que hi ha entre els observadors en la classificació histològica i la dificultat del diagnòstic diferencial en un subconjunt de casos de CE. En aquest sentit, la incorporació a la pràctica clínica del sistema alternatiu al TCGA, com seria l'algoritme ProMisE, pot comportar avantatges importants com la identificació de tumors tipus *serous-like/CN-high* i els *POLE* mutats, els quals presenten pronòstics totalment contraposats, sent de mal i molt bon pronòstic, respectivament.⁶²

Tot i això, la classificació molecular encara presenta reptes i necessita més refinament per resoldre problemes encara no resolts com els casos ambigus que presenten solapament d'histologies, els casos amb perfils moleculars no específics (NSMP), i els casos que poden pertanyer a diferents grups moleculars (*multiple classifiers*).⁷⁶ Aquests últims casos, un 2-3% segons Stello et al i Talhouk et al ^{61,62}, poden presentar a la vegada mutacions a *POLE* i MSI, mutacions a *TP53* i MSI i mutacions a *POLE* i *TP53*. Respecte a això, s'ha descrit que les mutacions del gen *TP53* poden ser un efecte secundari del fenotip mutat dels casos *POLE* i MSI. D'acord amb això seria recomanable, a l'hora de fer la classificació molecular del CE, primer determinar l'estat mutacional del gen *POLE*, ja que a més a més, s'ha vist que aquests pacients presenten pronòstics molt favorables.⁶³

Amb el descrit anteriorment es posa en evidència que el diagnòstic diferencial encara és difícil en un conjunt de casos, que la classificació molecular encara té espais buits per resoldre-ls, i que conseqüentment aquests casos són difícils de classificar adientment en grups de risc per al posterior maneig quirúrgic i terapèutic. També s'ha manifestat la importància de l'heterogeneïtat intratumoral en el comportament agressiu dels tumors.

Per aquests motius, amb els dos primers objectius específics (OE1 i OE2) d'aquesta tesi s'ha tingut interès en determinar en els tipus més freqüents de càncer d'endometri, CEE i CSE, diverses alteracions moleculars que ens permetin obtenir uns perfils genètics específics i que, conseqüentment, ens permetin classificar d'una manera més precisa i objectiva aquests dos tipus tumorals. Les tecnologies escollides per realitzar les anàlisis de les alteracions genètiques han estat: per una banda, l'aplicació de *targeted sequencing* amb la tecnologia NGS per determinar els

gens alterats respecte a variants somàtiques; i per altra banda, la tècnica MLPA per determinar les alteracions en el número de còpies dels gens. Finalment, amb el tercer objectiu específic (OE3) ens hem interessat en conèixer el paper que juga l'heterogeneïtat intratumoral (HIT) en el carcinoma serós d'endometri que és el subtipus amb el comportament més agressiu dintre el CE. En aquest hem analitzat l'efecte de l'HIT a dos nivells d'alteració molecular, a nivell de l'alteració dels gens determinant variants somàtiques i de l'alteració en el número de còpies dels gens amb les mateixes tecnologies emprades en els dos primers objectius.

El desenvolupament del càncer es produeix per l'adquisició d'alteracions genètiques somàtiques com substitucions, insercions o delecions de nucleòtids, translocacions, i ASNC entre d'altres.⁴ En el CE s'ha vist que les alteracions moleculars implicades en el desenvolupament dels CEE són diferents de les del CSE.^{81,159,160} De manera general, els CEE mostren inestabilitat de microsatèl·lits, així com mutacions als gens *PTEN*, *KRAS* i *CTNNB1*; en canvi, els CSE presenten mutacions al gen *TP53* i alteracions en el número de còpies dels gens (ASNC). Les mutacions als gens *PIK3CA* i *ARID1A* s'han identificat en ambdós tipus de tumor. L'estudi TCGA⁵⁹ al 2013 va confirmar aquest perfil mutacional diferencial entre els CEE i els CSE. Els gens més freqüentment alterats als CEE van ser *PTEN* (77%), *PIK3CA* (53%), *PIK3R1* (37%), *CTNNB1* (36%), *ARID1A* (35%), *KRAS* (24%), *CTCF* (20%), *RPL22* (12%) , *TP53* (11%), *FGFR2* (11%) i *ARID5B* (11%). Els gens més freqüentment alterats al CSE van ser *TP53* (90,7%), *PIK3CA* (41,9%), *FBXW7* (30,2%), *PPP2R1A* (36,6%), *CHD4* (16,3%), *CSMD3* (11,6%) i *COL11A1* (11,6%). Pel que fa a les ASNC que afecten al CSE, s'ha descrit en diferents estudis que els segments cromosòmics més freqüentment alterats són els que inclouen gens com *ERBB2*, *CCNE1*, *PIK3CA*, *EV1*, *MYC*, *CHD4*, *TP53* i *LPR1B*. Els 6 primers estan afectats per petites amplifícacions focals a nivell de duplicacions, i els dos últims per pèrdues.^{94,95}

En l'actualitat, l'ús de la tecnologia NGS amb l'aplicació *targeted sequencing* és una de les millors opcions per analitzar la implicació de diferents gens, a nivell de la identificació de variants somàtiques, en la tumorigènesis d'un determinat càncer, com per exemple el càncer d'endometri, que es sap que és molt heterogeni molecularment.³¹ Diversos autors han utilitzat aquesta tecnologia d'alt rendiment per estudiar el CE. En aquest sentit, Chang et al.¹⁶¹ va avaluar deu casos de CE utilitzant el panell integral de càncer de Nimblegen que inclou 578 gens relacionats amb el càncer. Aquest estudi va revelar 120 variants en 99 gens, 21 dels quals es van incloure posteriorment al panell de l'Oncomine (Thermo Fisher Scientific). Recentment, s'ha utilitzat l'aplicació de *targeted sequencing* (NGS) amb un panell customitzat de gens per identificar la càrrega mutacional (*mutational load*) en casos de CE amb MSI.¹⁶² McConechy et al.¹⁶³ també va

realitzar una anàlisi de *targeted sequencing* de 393 casos de CE seqüenciant nou gens (*ARID1A*, *PPP2R1A*, *PTEN*, *PIK3CA*, *KRAS*, *CTNNB1*, *TP53*, *BRAF* i *PPP2R5C*) i va suggerir que aquest panell podia ser útil com a complement a la classificació histològica del CE. En un altre estudi es va determinar que la majoria dels diagnòstics microscòpics eren concordants amb els genotips identificats, i a la vegada, es va observar que existien un conjunt de diagnòstics incompatibles amb els genotips identificats.¹⁶⁴

En relació amb aquests estudis, en el primer objectiu específic d'aquesta tesi s'ha determinat, en un total de 24 mostres de CE (16 CEE i 8 CSE), les alteracions a nivell de variants somàtiques mitjançant l'aplicació de *targeted sequencing* (NGS) utilitzant un panell personalitzat de 40 gens implicats en la carcinogènesi del CE segons el TCGA i altres estudis.^{59,94,95,147} Cal dir, que aquest ha estat un panell específic que no s'ha emprat prèviament en cap estudi. Posteriorment, amb l'anàlisi estadística dels resultats s'ha obtingut un perfil genètic que ens ha permès discriminar entre tipus histològic (CEE, CSE), i per tant, classificar els casos d'una manera objectiva mitjançant una anàlisi molecular.

Els 24 casos inclosos en l'estudi estaven prèviament ben caracteritzats histològicament, i amb la seva classificació molecular alternativa es va confirmar que hi havia una representació dels quatre grups moleculars del TCGA. Tres casos van ser especialment interessants; el cas 9 que va ser considerat MSI (inestabilitat de microsatèl·lits) per la pèrdua de l'expressió dels gens reparadors, i p53 *wild type* per IHQ però amb una variant somàtica per *TP53*, tant detectada per NGS com per seqüenciació Sanger; el cas 11 que va ser un tumor mutat per *POLE* que va mostrar un patró mutat de p53 per IHQ; i el cas 21 que era un CSE que tenia un patró *wild type* de p53 per IHQ, però altrament presentava les característiques histològiques típiques d'un CSE. Els casos restants eren més o menys clàssics de CEE i CSE.

Amb la nostra anàlisi de NGS es van identificar un total de 164 variants somàtiques (102 *missense*, 39 *frameshift*, 15 *stop gained*, 1 *splice acceptor*, 1 *protein altering* i 6 *inframe deletions*). El 72% d'elles corresponien a variants patogèniques (PV) o probablement patogèniques (LPV). Els gens més freqüentment alterats en tot el conjunt de casos van ser *PTEN* (63%), *TP53* (54%), *PIK3CA* (50%), *ARID1A* (46%), *MKI67* (29%), *KMT2B* (29%), *PPP2R1A* (25%), *ATR* (21%), *CHD4* (21%), *CTCF* (21%) i *RPL22* (21%). En el subgrup dels 16 casos de CEE els gens més freqüentment afectats van ser *PTEN* (93,7%), *ARID1A* (68,7%), *PIK3CA* (50%) i *KMT2B* (43,7%), mentre que en els 8 tumors de CSE van ser *TP53* (87,5%), *PIK3CA* (50%) i *PPP2R1A* (25%). Al comparar les nostres dades amb els resultats de l'estudi del TCGA, tot i que els gens i les freqüències d'alteració són similars, podem

remarcar que nosaltres tenim més percentatge de casos CEE amb alteracions a *PTEN*, i que no hem observat alteracions del gen *FBXW7* en els casos CSE. Una de les raons d'aquesta diferència pot ser la limitació en el número de mostres del nostre estudi.

L'anàlisi estadística de doble *cluster* jeràrquic de totes les alteracions va demostrar que el nostre enfoc personalitzat, pel que fa al panell de gens escollits per seqüenciar amb l'aplicació de *targeted sequencing* (NGS), permetia una agrupació correcta de tots els tumors en les dues categories histològiques. I conseqüentment, es va definir un perfil genètic específic associat a cadascuna de les histologies, d'una banda, l'existència de variants en els gens *PTEN*, *ARID1A* i *KMT2B* era de CEE, i d'altra banda, l'absència de *PTEN*, *ARID1A* i *KMT2B* en presència de variants de *TP53* era diagnòstica de CSE. Les alteracions als gens *TP53* i *PIK3CA* van estar presents en els dos tipus de tumors amb freqüències similars.

Les alteracions en els gens *TP53*, *PTEN* i *ARID1A* identificades amb l'anàlisi de NGS van ser validades mitjançant la seqüenciació Sanger i la immunohistoquímica. Els resultats d'aquesta anàlisi de validació van mostrar una bona correlació entre les tècniques utilitzades. En general, les alteracions al gen *TP53* es van associar amb un patró mutat per IHQ (només un tumor amb variant patogènica a *TP53* va mostrar un patró de tipus *wild type* per IHQ). Els casos amb alteracions als gens *ARID1A* i *PTEN* van mostrar una concordança amb una immunotinció negativa en el 83% i el 86%, respectivament. Curiosament, va haver una bona correlació entre els casos de CEE MSI, valorats per IHQ per a MSH6 i PMS2, i el nombre de variants somàtiques detectades, tot i que, el nostre enfoc no és l'eina correcta per avaluar la càrrega mutacional dels tumors. L'únic cas *POLE* alterat (cas 11), va ser un CEE de grau 3 que va mostrar una expressió negativa de MSH6, un patró mutat de p53, tinció de *PTEN* negativa i el major nombre de variants somàtiques detectades, incloent una a *TP53* i una a *PTEN*. Un dels casos de CEE de grau 3 (cas 6), classificat com a doble mutat va ser categoritzat correctament amb la nostra anàlisi de NGS. La sèrie de casos també va incloure un CSE que tenia una tinció *wild type* per p53, però que presentava característiques microscòpiques típiques de CSE (cas 21). La nostra anàlisi va classificar aquest cas correctament com un CSE, i va confirmar la manca de variants somàtiques al gen *TP53*. Aquest cas il·lustra el fet que hi ha un subconjunt de CSE (9% segons TCGA) que no mostra variants a *TP53* ni patró mutat de p53 per IHQ. Curiosament, aquest cas es classificaria erròniament com un càncer d'endometri del grup NSMP, basat en el patró *wild type* de p53 mitjançant l'aproximació alternativa del TCGA, mentre que el perfil genètic confirma que el diagnòstic correcte és d'un CSE. A més a més, la seqüenciació de l'exoma (dades no mostrades en aquesta tesi) d'aquest cas va confirmar un elevat nombre d'ASNC, per tant, el tumor va ser considerat com un CSE. Hi va haver un altre

tumor (cas 9) que va ser classificat erròniament amb el sistema alternatiu del TCGA en el grup de tumors MSI, ja que per IHQ tenia expressió negativa de les proteïnes reparadores i patró *wild type* de p53. Aquest cas en el nostre estudi va mostrar una variant a *TP53*, que també va ser confirmada per la seqüenciació Sanger demostrant que, efectivament, es tractava d'un tumor doble mutat. En general, el nostre estudi amb la tecnologia NGS va demostrar que el sistema alternatiu del TCGA va classificar erròniament 2 dels 24 tumors d'aquesta sèrie de casos.

En resum, els resultats del primer objectiu específic d'aquesta tesi han demostrat la idoneïtat de la nostra anàlisi de NGS amb un panell personalitzat de gens com a eina molecular addicional per confirmar la classificació histològica del càncer d'endometri. Aquesta estratègia sembla interessant com una eina per classificar tumors amb troballes microscòpiques inusuals i ambigües, així com tumors amb resultats d'IHQ inesperats, com ara CSE sense alteracions de *TP53*, tal com s'il·lustra en un cas d'aquesta sèrie. No obstant, es requereixen estudis posteriors amb una sèrie més gran de casos que contingui casos consecutius per validar aquest enfocament en casos ambigus amb característiques combinades o mixtes de CEE i CSE.

En el segon objectiu s'ha utilitzat la tècnica MLPA per determinar les alteracions del número de còpies de 106 gens implicats en el procés carcinogènic en 24 mostres de càncer d'endometri (12 CEE i 12 CSE). Posteriorment, amb l'anàlisi estadística dels resultats s'ha obtingut un perfil genètic que ha permès, d'una banda, discriminar de manera global els CEE dels CSE, i d'altra demostrar l'existència d'un subconjunt de casos amb característiques histològiques típiques de CSE que presenten uns patrons d'ASNC que no concorden amb el fenotip.

Les ASNC són events moleculars que alteren el genoma fent que aquest canviï la seva dosi gènica normal. Aquesta dosi tant es pot modificar a nivell de pèrdues com de guanys. Les ASNC es poden categoritzar en diferents graus en funció de les longituds dels segments gènics implicats. Així tenim les ASNC àmplies (*broad*) que engloben cromosomes sencers (aneuploïdies) o segments de cromosomes, i les ASNC focals que engloben segments de longitud inferior a un braç cromosòmic (*arm-level*) o només un o pocs gens (*gene-level*) que de vegades són intragèniques (petites amplificacions o duplicacions dintre un únic gen).¹⁶⁵

Per analitzar aquests events moleculars es poden utilitzar diferents tecnologies com la citometria de flux per determinar aneuploïdies¹⁶⁶ i els *microarrays* de SNP (*Single Nucleotide Polymorphism*) o de CGH (*Comparative Genomic Hybridization*) amb conjunció de l'algoritme GISTIC (*Genomic Identification of Significant Targets in Cancer*) per tal d'identificar regions cromosòmiques recurrents amb ASNC.^{40,41}

En diversos estudis s'ha observat que les ASNC tenen un paper molt crític en poder activar oncogens o inactivar gens supressors tumorals, així és important poder identificar específicament quins són els gens afectats per aquestes alteracions, i així saber la implicació biològica i fenotípica d'aquests events.⁴⁰ Una de les tècniques que permeten la identificació directa de gens afectats per ASNC és la MLPA. Aquesta és una tecnologia d'alt rendiment, que té l'avantatge de ser més econòmica i menys costosa que les altres tecnologies àmpliament utilitzades per les anàlisis d'ASNC, com els *arrays* de SNP i els de CGH.^{42,43,167} Amb la MLPA s'ha avaluat l'estat d'ASNC en diferents tumors com el melanoma uveal, el càncer gàstric, el glioblastoma i el càncer de mama.

167-171

Al llarg dels últims anys, s'han descrit molts tumors amb comportaments agressius afectats per ASNC com els carcinomes escamosos de cap i coll, de pròstata, de pulmó, tumors que afecten els òrgans reproductors com el carcinoma serós d'ovari, d'endometri, i els tumors de mama *basal-like*.⁴⁰ S'ha vist que diferents tipus de càncer varien entre ells pel seu balanç d'ASNC respecte variants somàtiques, així el càncer de pròstata, el carcinoma serós d'ovari d'alt grau, i el carcinoma serós d'endometri presenten altes taxes d'ASNC comparat amb les taxes de variants somàtiques. A la vegada també s'ha demostrat que les ASNC tenen valor pronòstic en un ampli ventall de tumors, i recentment s'ha definit el concepte de càrrega d'ASNC (alteració somàtica número de còpies) com un factor pronòstic associat a recurrència i morts en múltiples càncers.

172,173

En el càncer d'endometri les ASNC, a nivell d'aneuploidia, es van començar a estudiar l'any 1959 amb l'estudi d'Atkins.¹⁷⁴ Fins a l'actualitat hi ha hagut diferents estudis que han avaluat el valor pronòstic de l'aneuploidia en aquest càncer.¹⁵⁸ En aquests estudis l'aneuploidia s'ha correlacionat amb la invasió linfovascular i amb l'afectació dels ganglis limfàtics, i també s'ha demostrat que té un efecte pronòstic advers en casos CEE amb estabilitat de microsatèl·lits (MSS), i que és un marcador pronòstic independent en CSE.¹⁷⁵⁻¹⁷⁷ Un estudi recent ha demostrat que l'aneuploidia en el CE es correlaciona amb CEAG, de tipus CSE i amb unes pitjors taxes de supervivència (PFS, *Patient Free Survival*).¹⁶⁶ El TCGA⁵⁹ però ja va fer l'estudi d'ASNC en el CE utilitzant la tecnologia de *microarrays* de SNP (Affymetrix) i l'algoritme GISTIC. En aquest gran estudi el CE es va subdividir en quatre *clusters* de diferents graus d'ASNC. Els tres primers *clusters* estaven compostos majoritàriament per carcinomes d'histologia endometriode (97%), i tenien un molt baix percentatge d'ASNC. En canvi, el *cluster* 4 es va definir com el grup amb un alt percentatge d'ASNC. Aquest grup incloïa gairebé tots els carcinomes serosos (94%), un 62% de casos amb histologia mixta, i un 12% de casos amb histologia endometriode. Els tumors d'aquest grup

Capítol 5

estaven caracteritzats per amplificacions focals del gen *MYC* (8q24.12), *ERBB2* (17q12), i *CCNE1* (19q12), *FGFR3* (4p16.3) i *SOX17* (8q11.23). Aquests tumors també estaven caracteritzats per presentar mutacions al gen *TP53* (90%), poca MSI (6%) i molt baixa freqüència de mutacions de *PTEN* (11%) comparat amb els CEE (84%). A més a més, aquests tumors tenien un pitjor pronòstic comparat amb els dels *clusters* amb carcinomes majoritàriament endometrioides. En relació amb aquests resultats, l'estudi de Zhao i Kuhn^{94,95} també van demostrar que els CSE presentaven un alt percentatge d'ASNC. En aquests estudis la determinació d'aquestes alteracions es va realitzar mitjançant la seqüenciació de l'exoma, i *microarrays* de SNP i l'algoritme GISTIC, respectivament. De les regions més freqüentment alterades a nivell de guanys, cal destacar amplificacions focals del cromosoma 17 que conté el gen *ERBB2* (44%), llargues duplicacions de la regió 3q26-27 que conté a *PIK3CA* (60%) i *EVI1* (22%), duplicacions petites al cromosoma 19 que conté *CCNE1* (48%), amplificació d'un llarg segment del cromosoma 8 que conté *MYC* (44%), i amplificacions d'un segment del cromosoma 12 que inclou *CHD4* (28%). De les regions més alterades a nivell de pèrdues cal destacar la regió del cromosoma 17 que inclou el gen *TP53* (44%), la del cromosoma 7 amb *MGAM* (35%) i la del cromosoma 16 amb *CDH13* (35%).

El nostre estudi d'MLPA va confirmar que els CSE estan caracteritzats per un alt nombre d'ASNC en comparació als CEE. En el grup de CEE (n=12) es van obtenir unes freqüències globals d'ASNC que oscil·laven entre l'1 i el 4%, en canvi la majoria de CSE (66%) mostraven unes freqüències de l'11 al 31%. L'anàlisi estadística del perfil genètic d'ASNC obtingut mitjançant MLPA va identificar un total de 10 gens alterats que presentaven una associació significativa amb el tipus histològic on cap d'aquests gens va ser detectat en el grup CEE. Dels 10 gens alterats, 8 corresponien a gens afectats per guanys: *CCNE1*, *UCKL1*, *EVI1*, *PIK3CA*, *NTRK1*, *PTPN1*, *BCL6* i *FKBP8*, i dos d'ells afectats per pèrdues: *BIRC4* i *FGFR1*. A diferència d'altres estudis descrits anteriorment^{94,95} en el nostre estudi vam observar freqüències més baixes d'amplificacions d'*ERBB2* (25%), i vam observar gens alterats amb freqüències elevades que no havien estat descrits prèviament com *UCKL1*, *NTRK1*, *PTPN1*, *FKBP8*, *BIRC4* i *FGFR1*.

L'anàlisi estadística de doble *cluster* jeràrquic, tenint en compte els 10 gens significativament alterats, va agrupar les mostres en dos grups, el grup 1 (G1) que es corresponia amb tots els casos CEE (n=12) més 5 casos que eren CSE, i el grup 2 (G2) que es corresponia amb casos CSE (n=7). Amb aquesta anàlisi destaquem els 5 casos (14, 16, 17, 22 i 24) amb histologia de CSE que no es van classificar en el grup G2, típic dels CSE, i per tant es posa en dubte la classificació histològica d'aquests casos. D'aquest 5 casos, el 14 i 24, són els que s'agrupaven al grup G1B amb tota la resta de CEE. Es van caracteritzar per tenir un patró *wild type* de p53, un baix número d'ASNC

sense presentar alterat cap dels gens significatius. L'estudi mutacional del gen *POLE* va confirmar que el cas 14 era *POLE* mutat, i es classificaria al grup 1 del TCGA (*POLE ultramutated*). El cas 24 es va confirmar que tampoc tenia inestabilitat de microsatèl·lits i per tant s'hauria de classificar com un cas *CN-low* o *NSMP*. Amb aquests 2 casos demostrem que el perfil molecular d'ASNC ens ajudaria a classificar-los més adientment. En els casos 16 i 22, inclosos al grup intermig G1C, tampoc es correlaciona la histologia de CSE, ni el patró mutat de p53 amb el baix número d'ASNC. No obstant això, aquests són dos casos molt particulars ja que, tot i presentar baix número d'ASNC, la única alteració que presenten és al gen *CCNE1*, que és un dels dos gens amb més associació estadística amb el CSE. En aquests casos es pot suggerir la importància de la determinació de l'alteració de la *CCNE1* juntament amb la IHQ de p53 per poder els classificar els casos en el grup *serous-like* del TCGA. I per últim el cas 17, que presenta una histologia de CSE, que es classifica en un altre grup intermig, el G1A, que presenta un número mitjà d'ASNC (n=7) i patró mutat de p53.

A més a més d'aquests 5 casos, també cal destacar dos casos més, el cas 4 i 19, que no es correlacionava el patró d'IHQ de p53 amb la histologia i l'agrupació del perfil molecular. El cas 4 era un CEE classificat al grup G1B però que presentava un patró mutat de p53. Aquest cas també va ser analitzat a l'objectiu 1, cas 16, i allí es va confirmar que era un cas doble mutat, amb presència de MSI i d'alteració a *TP53*, i el perfil genètic el va classificar com a CEE amb alteració a *PTEN* i *TP53*. En aquests casos l'alteració de *TP53* no s'enten com un event *driver* sinó com un *passenger*, i no ha de ser classificat com un *serous-like* ja que el que preval és l'estat MSI i per tant està ben classificat amb el perfil genètic tant de variants somàtiques com d'ASNC. I per acabar, el cas 19, que presentava una histologia de CSE i va ser classificat al grup G2 però que presentava un patró *wild type* de p53. Aquest cas també va ser analitzat a l'objectiu 1, cas 21, on es va confirmar que no presentava cap variant somàtica per *TP53* i va ser classificat pel perfil genètic com un CSE. En aquest cas el perfil genètic d'ASNC i de variants somàtiques també ens permetria una correcta classificació d'aquest cas com *serous-like* enlloc d'un cas *NSMP*.

En resum, els resultats del segon objectiu d'aquesta tesi han descrit que de manera general els CSE presenten moltes més ASNC que els CEE. Tot i això, també s'ha presentat que dins els carcinomes amb histologia serosa, hi ha un percentatge de casos (42%) que presenten unes característiques genètiques que no es corresponen amb el fenotip, i per tant en aquests casos és molt important l'ajuda del perfil genètic d'ASNC per poder realitzar una classificació més correcta. A més a més, també han suggerit l'ús de la combinació de la determinació de p53 per IHQ i del número de còpies de la *CCNE1* per classificar els casos dins el grup *serous-like*. No obstant això, es

requereixen estudis posteriors amb una sèrie més gran de casos que contingui casos consecutius per poder estudiar d'una manera més profunda aquells casos amb histologia serosa o ambigua que potser molecularment no serien classificats com d'alt risc.

Després de les troballes del primer i segon objectiu específics de la tesi, es va creure interessant avaluar el paper de l'heterogeneïtat intratumoral (HIT) en el CSE. Per abordar aquesta anàlisi es van escollir els 8 casos de CSE de l'objectiu 2 que presentaven un nombre més elevat d'ASNC. De cadascun d'ells es van seleccionar tres regions de tumor a l'atzar i cada porció es va analitzar mitjançant *targeted sequencing* (NGS) i MLPA per tal d'identificar la implicació de l'HIT tant a nivell de variants somàtiques com d'ASNC.

Com s'ha descrit anteriorment, el CSE és un dels tipus histològics més agressius dintre el CE.⁵⁴ Conseqüentment, en aquest tipus de tumor és important poder proporcionar evidències científiques per millorar els criteris diagnòstics, pronòstics i de predicció de resposta a les teràpies. En l'última dècada, s'han realitzat grans esforços per conèixer els esdeveniments moleculars *drivers* que estan implicats en la tumorigènesi del CSE. Així, tant altres estudis com els resultats obtinguts en aquesta tesi (OE1 i OE2) han revelat les característiques moleculars específiques del CSE com l'alta proporció d'ASNC i l'alteració del gen *TP53*.^{59,94,95,178}

En els últims anys s'ha descrit la influència de l'HIT en una gran proporció de tumor sòlids, i s'ha proposat com un factor addicional implicat en la progressió tumoral i en la supervivència dels pacients.^{22,25,26} Concretament, en el CSE la implicació de l'HIT encara no ha estat plenament avaluada. L'HIT s'entèn com la coexistència de diferents subpoblacions de cèl·lules tumorals en un únic tumor que difereixen entre elles per les seves alteracions moleculars que els hi donen característiques genotípiques i fenotípiques específiques. Es creu que l'HIT resulta d'un procés d'evolució ramificada (*branched evolution*), on a mesura que el tumor progressa es van generant diferents poblacions cel·lulars subclonals a partir d'una o poques poblacions clonals fundadores.^{20,179,180} Conseqüentment, diferents regions d'un mateix tumor poden presentar diferents característiques moleculars. L'HIT es presenta com un gran repte, tant per al maneig global dels pacients amb càncer com per avaluar el paper de fàrmacs en teràpies dirigides, degut a la diferent distribució espacial de les alteracions subclonals dels gens o proteïnes diana. L'arribada de les tecnologies de NGS ha permès donar més coneixement sobre l'extensió i la prevalença de l'HIT, tanmateix, les seves implicacions clíniques i biològiques encara no estan plenament establertes. S'ha demostrat que l'heterogeneïtat espacial i temporal estan implicades en l'adaptació dels

tumors a les variacions del microambient tumoral i, alhora, que aquest microambient pot oferir una pressió selectiva per determinar aquesta HIT en el càncer.

Amb l'anàlisi de *targeted sequencing* (NGS) de les 8 mostres, els gens més freqüentment alterats van ser *TP53* (88%), *PIK3CA* (50%) i *PPP2R1A* (25%), tal com ha estat descrit en altres estudis.^{59,94,95,147} A més a més, vam trobar una concordança del 100% d'aquestes alteracions entre les tres regions del tumor analitzades. L'estudi de Kuhn⁹⁴ també van trobar aquests gens alterats en el carcinoma serós intraepitelial; per tant, la identificació d'aquestes alteracions poden tenir implicacions clíniques ja que es poden detectar a l'etapa preinvasiva utilitzant tècniques d'alt rendiment com la NGS. A diferència d'altres estudis, nosaltres no vam observar alteracions al gen *FBXW7*. Aquest gen s'ha descrit que està implicat en l'augment dels nivells de proteïna ciclina E i en la facilitació de la progressió del cicle cel·lular. L'alta prevalença i ubiqüitat de les alteracions del gen *TP53* reforça la importància de p53 en el desenvolupament i progressió del CSE, i remarca l'evidència de l'existència d'un clon fundador amb aquestes alteracions.

A nivell de les ASNC, es va observar una alta presència d'aquestes alteracions en tots els casos. Els valors de DQ (*Dosage Quocient*) ens van suggerir que les ASNC detectades eren a nivell de duplicacions dels gens implicats. Interessantment, el nostre estudi va mostrar que en el CSE l'HIT estava caracteritzada genèticament per ASNC. D'acord amb altres estudis, vam observar guanys a nivell de duplicacions en gens com *CCNE1* (53%), *ERBB2* (75%) i *PIK3CA* (75%).^{59,94,95} En aquest punt cal remarcar que en l'objectiu específic 2 vam trobar que el gen *ERBB2* no era un dels gens més freqüentment alterats (25%), en canvi, en aquest estudi vam veure que la freqüència d'aquest gen va augmentar (75%) a causa d'identificar-lo en heterogeneïtat. Els gens més freqüentment alterats a nivell de guanys que no van estar afectats per HIT, és a dir, que les alteracions es van detectar homogèniament distribuïdes en les tres regions del tumor, van ser *PIK3CA* i *CCNE1*. Aquest fet suggereix que les alteracions en aquests gens, igual que l'alteració de *TP53*, actuen com events *drivers* i són presents en el clon fundador. En l'objectiu específic 2 ja vam mostrar la significació de les alteracions del gen *CCNE1* i *PIK3CA*, i juntament amb les dades anteriors, fa que aquests gens puguin ser considerats com un potencial marcador d'alteracions a nivell d'ASNC, i per tant un possible event molecular per classificar el grup CN-high/serous-like. A més d'aquests gens, també vam detectar altres gens en guanys a nivell de duplicacions com *UCKL1*, *EVI1*, *PTPN1*, *NTRK1*, *MYBL2*, *AURKA*, *BCL6* i *FKBP8*.

En el nostre estudi les regions cromosòmiques més prevalents van ser en el cromosoma 20, els segments 20q12 – q13.3, i en el cromosoma 3, el segments 3q26 – q27. El segment 20q12 – q13.3 conté el gen més freqüentment alterat, *UCKL1*, que és una uridin cinasa. Aquesta cinasa catalitza la fosforilació d'uridina a UMP, que és un pas limitant en la síntesi de l'ADN i l'ARN. La sobreexpressió d'aquesta proteïna ha estat implicada en l'activació de la progressió tumoral, i el seu efecte s'ha vist activat per la sobreexpressió d'altres proteïnes com la *CCNE1*, també implicada en la carcinogènesi del CSE.¹⁸¹ El segment 3q26– regió q27, que conté gens com *EVI1*, *BCL6* i *PIK3CA*, ja ha estat prèviament reportat com una regió freqüentment amplificada en el CSE.⁹⁴ Un estudi recent ha analitzat l'efecte d'ASNC en una multitud de càncers (*pan-cancer analysis*) i ha demostrat que els càncers de llinatges similars com el càncer d'endometri, el d'ovari i el de mama, presenten taxes similars d'amplificacions i delecions en regions específiques afectades per ASNC.⁴⁰ Un altre gran estudi, centrat en l'anàlisi dels patrons d'ASNC en diferents tipus de tumors, va revelar que el càncer d'ovari, de mama i el CSE comparteixen uns perfils similars caracteritzat per alteracions del gen *TP53*, per un enriquiment de regions amb ASNC que contenen gens implicats en la via de la glicòlisi, que és un dels *hallmarks* del càncer.^{4,91}

Recentment s'ha descrit un fenomen molecular, les duplicacions en tàndem (TD) o fenotip TDP (TDP, *Tandem Duplication Phenotype*), que pot ser l'explicació de les altes taxes de duplicació de regions específiques del genoma en el CSE. Menghi et al.^{90,182} va proposar la TDP com un cromotip (*chromotype*) del càncer, que es considerat com un fenomen genòmic molt prevalent en el càncer i que està caracteritzat per un enriquiment de segments amb duplicacions en tàndem al llarg de tot el genoma. Aquest fenomen s'ha observat en el càncer de mama, en el d'ovari, i en el d'endometri. Aquest estudi també suggereix un escenari mecanicista per a la inducció TDP, on es diu que l'estrés de replicació produït per les amplificacions de *CCNE1* i la presència de les mutacions de *TP53* pot produir la formació de TD. Un altre *pan-cancer* anàlisi d'ASNC en diferents tumors també va proposar que les duplicacions del genoma sencer (WGD, *Whole Genome Duplications*) es correlacionaven amb alteracions de *TP53*, amplificacions de la *CCNE1* i amb mutacions al gen *PPP2R1A*.⁴⁰

Segons els nostres resultats, els guanys del gen *ERBB2* semblen ser una de les alteracions moleculars més freqüentment afectades per l'heterogeneïtat intratumoral. Aquesta troballa podria suggerir el probable origen subclonal d'aquesta alteració i la raó de la baixa eficàcia de les teràpies anti-ERBB2. L'elevada prevalència d'alteracions d'*ERBB2* en el CSE ens porta a proposar que *ERBB2* podria ser una bona diana terapèutica en el CSE. En el càncer de mama l'heterogeneïtat intratumoral d'*ERBB2* ha estat documentada en una freqüència de l'11% al 40%.

^{183,184} Els resultats d'alguns estudis sobre la teràpia dirigida contra ERBB2 en el CSE han mostrat dades contraposades ¹⁸⁵, no obstant això, recentment, Fader et al ¹⁸⁶ han obtingut resultats prometedors d'aquesta teràpia en el CSE, sobretot fent una adequada selecció de les pacients. Cal esmentar que, altres autors també han demostrat la marcada heterogeneïtat d'*ERBB2* mitjançant les tècniques d'IHQ i FISH. ^{187,188} En el nostre estudi també vam detectar altres gens distribuïts heterogèniament com *MYBL2*, *AURKA*, *PTPN1* i *EVI1*.

Un de les determinacions importants d'aquest objectiu ha estat que les alteracions de *TP53* no s'han vist afectades per l'HIT. En aquest sentit, podem suggerir que l'HIT no té un impacte negatiu en la classificació molecular del càncer d'endometri. Cal remarcar, que en aquesta sèrie de casos, un dels tumors (cas 5) no presentava alteració de *TP53* ni per *targeted sequencing* (NGS), ni per seqüenciació Sanger ni per IHQ. Segons les dades del TCGA, existeix al voltant del 9% de CSE que, tot i no presentar alteracions de *TP53*, es classifiquen com CSE. En un altre projecte de recerca, es va realitzar l'anàlisi de l'exoma d'aquest cas (dades no mostrades) i es va confirmar que aquest tumor presentava ASNC, de la mateixa manera que ho hem demostrat nosaltres per MLPA. A més a més, aquest cas en l'objectiu específic 1 es va classificar amb el perfil genètic com un CSE.

Aquest tercer objectiu té algunes limitacions com que es va limitar l'anàlisi només de diferents regions del tumor seleccionades a l'atzar, que no es van incloure mostres de metàstasis ni recidives, i que es va analitzar un nombre reduït de casos. Totes aquestes limitacions poden explicar les diferents freqüències en ASNC en comparació amb altres estudis.

En resum, els resultats del tercer objectiu específic d'aquesta tesi han mostrat que el CSE, per una banda, presenta unes alteracions moleculars clonals com les variants somàtiques al gen *TP53* i guanys en els gens *CCNE1* i *PIK3CA*, i per altra banda, que la seva heterogeneïtat intratumoral està caracteritzada principalment per les ASNC dels gens. A més a més, es remarca que un dels gens principalment afectats per aquesta heterogeneïtat és el gen *ERBB2*, que és una diana terapèutica àmpliament utilitzada.

Les dades globals obtingudes d'aquesta tesi han demostrat que el nostre estudi personalitzat de NGS es pot utilitzar com una eina complementària al diagnòstic histològic per tal d'arribar a un diagnòstic final més objectiu i precís, que l'anàlisi de les ASNC mitjançant la MLPA i la detecció de p53 per IHQ són eines que ajuden a poder classificar millor casos de CSE dubtosos, i que el CSE presenta HIT a nivell d'ASNC afectant gens diana com l'*ERBB2*.

CAPÍTOL 6
CONCLUSIONS

Primera. El nostre estudi personalitzat de seqüenciació de nova generació (NGS), amb el panell de 40 gens, és útil com a eina molecular addicional per confirmar la classificació histològica del càncer d'endometri.

Segona. L'existència de variants en els gens *PTEN*, *ARID1A* i *KMT2B* és diagnòstica del carcinoma endometrioides d'endometri.

Tercera. L'absència de variants en els gens *PTEN*, *ARID1A* i *KMT2B* en presència de variants de *TP53* és diagnòstica del carcinoma serós d'endometri.

Quarta. El carcinoma serós d'endometri presenta un elevat nombre d'alteracions somàtiques en el número de còpies dels gens en comparació al carcinoma endometrioides d'endometri.

Quinta. L'anàlisi de les alteracions somàtiques en el número de còpies dels gens mitjançant la tècnica *Multiple Ligase-dependent Probe Amplification* (MLPA) permet realitzar una classificació més adient dels casos dubtosos de carcinoma serós d'endometri.

Sexta. La combinació de la determinació de l'expressió de p53 per immunohistoquímica i la determinació de les alteracions somàtiques en el número de còpies del gen *CCNE1* és una estratègia adequada per classificar d'una manera més encertada el carcinoma serós d'endometri.

Setena. El carcinoma serós d'endometri presenta alteracions moleculars clonals com les variants somàtiques al gen *TP53* i guanys en els gens *CCNE1* i *PIK3CA*.

Vuitena. L'heterogeneïtat intratumoral en el carcinoma serós d'endometri està caracteritzada per les alteracions somàtiques en el número de còpies dels gens.

Novena. En el carcinoma serós d'endometri el gen *ERBB2* és un dels gens principalment afectats per l'heterogeneïtat intratumoral.

BIBLIOGRAFIA

1. Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bettio, M., *et al.* Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Eur. J. Cancer* **103**, 356–387 (2018).
2. Sociedad Española Oncología Médica. Las Cifras del cancer en España 2018. (2018).
3. Massagué, J. Evolución y metástasis del cáncer. *Sebbm* **160**, 22–25 (2009).
4. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
5. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
6. Roy, N. K. Cancer - an Overview and Molecular Alterations in Cancer. *Fusion Genes and Cancer* 1–15 (2012).
7. Salaverry, O. Cancer etymology and its historical curious course. *Rev. Peru. Med. Exp. Salud Publica* **30**, 137–41 (2013).
8. Sudhakar, A. History of Cancer, Ancient and Modern Treatment Methods. *J Cancer Sci Ther.* **1**, 1–4 (2010).
9. Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153**, 17–37 (2013).
10. MacConaill, L. E. & Garraway, L. A. Clinical implications of the cancer genome. *J. Clin. Oncol.* **28**, 5219–5228 (2010).
11. Sanger, F., Nicklen, S., & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467 (1977).
12. Luckey, J. A., Drossman, H., Kostichka, A. J., Mead, D. A., D’Cunha, J. N., Norris, T. B. *et al.* High speed DNA sequencing by capillary electrophoresis. *Nucl Acids Res* **18**, 4417–4421 (1990).
13. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679 (1986).
14. Mccarthy, A. Third generation DNA sequencing: Pacific biosciences’ single molecule real time technology. *Chem. Biol.* **17**, 675–676 (2010).
15. Moorthie, S., Mattocks, C. J. & Wright, C. F. Review of massively parallel DNA sequencing technologies. *Hugo J.* **5**, 1–12 (2011).
16. Michael, L. M. Sequencing technologies — the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
17. Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S. *et al.* Third-generation sequencing and the future of genomics. *bioRxiv* 048603 (2016).
18. Nowell, P. C. The clonal evolution of tumor cell populations. *Science (80-)*. **194**, 23–28 (1976).

Bibliografia

19. Zapata, L., Pich, O., Serrano, L., Kondrashov, F. A., Ossowski, S. & Schaefer, M.H. Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome. *Genome Biol.* **19**, 1–17 (2018).
20. Barber, L. J., Davies, M. N. & Gerlinger, M. Dissecting cancer evolution at the macro-heterogeneity and micro-heterogeneity scale. *Curr. Opin. Genet. Dev.* **30**, 1–6 (2015).
21. Greaves, M. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
22. Hu, Z., Sun, R., & Curtis, C. A population genetics perspective on the determinants of intra-tumor heterogeneity. *Biochim Biophys Acta* **1867**, 109–126 (2017).
23. Podlaha, O., Riester, M., Subhajyoti, D., & Michor, F. Evolution of the cancer genome. *Trends Genet* **28**, 155–163 (2012).
24. Gerlinger, M., McGranahan, N., Dewhurst, S. M., Burrell, R. A., Tomlinson, I. & Swanton, C. Cancer: Evolution Within a Lifetime. *Annu. Rev. Genet.* **48**, 215–236 (2014).
25. Andor, N., Graham, T.A., Jansen, M., Xia, L.C., Aktipis, C.A., Petritsch, C. *et al.* Pan-cancer analysis of the extent and consequences of intra-tumor heterogeneity. *Nat Med.* **22**, 105–113 (2016).
26. Morris, L.G., Riaz, N., Desrichard, A., Şenbabaoğlu, Y., Hakimi, A.A., Makarov, V., Reis-Filho, J.S. *et al.* Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget.* **22**(1), 105-13 (2016).
27. Caiado, F., Silva-Santos, B. & Norell, H. Intra-tumour heterogeneity – going beyond genetics. *FEBS J.* **283**, 2245–2258 (2016).
28. Lipinski, K.A., Barber, L.J., Davies, M.N., Ashenden, M., Sottoriva, A. & Gerlinger, M..Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends in Cancer* **2**, 49–63 (2016).
29. Normand, R. & Yanai, I. An introduction to high-throughput sequencing experiments: Design and bioinformatics analysis. *Methods Mol. Biol.* **1038**, 1–26 (2013).
30. Sharif, A. Principles of Next generation sequencing. (2019).
31. Meldrum, C., Doyle, M. a & Tothill, R. W. Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective. *Clin. Biochem. Rev.* **32**, 177–195 (2011).
32. Mertes, F., Elsharawy, A., Sauer, S., van Helvoort, J.M., van der Zaag, P.J., Franke, A. *et al.* Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief. Funct. Genomics* **10**, 374–386 (2011).
33. Jennings, L.J., Arcila, M.E., Corless, C., Kamel-Reid, S., Lubin, I.M., Pfeifer, J. *et al.* Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J. Mol. Diagnostics* **19**, 341–365 (2017).
34. Mardis, E. R. Next-Generation Sequencing Platforms. *Annu. Rev. Anal. Chem.* **6**, 287–303 (2013).

35. Only, U. Illumina sequencing introduction. *Illumina Seq. Introd.* 1–8 (2017). doi:http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf
36. Low, L. & Tammi, M. T. Introduction to Next Generation Sequencing Technologies. *Bioinformatics* 1–21 (2017).
37. Oliver, G. R., Hart, S. N. & Klee, E. W. Bioinformatics for clinical next generation sequencing. *Clin. Chem.* **61**, 124–135 (2015).
38. Koboldt, D. C., Ding, L., Mardis, E. R. & Wilson, R. K. Challenges of sequencing human genomes. *Brief. Bioinform.* **11**, 484–498 (2010).
39. Ewing B, G. P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
40. Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
41. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhi, R., & Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
42. Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., & Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30**, e57 (2002).
43. Homig-Holzel, C. & Savola, S. Multiplex ligation-dependent probe amplification (MLPA) in tumor diagnostics and prognostics. *Diagn.Mol.Pathol.* **21**, 189–206 (2012).
44. Porth & Grossman. Fisiopatología: Alteraciones de la salud. Conceptos básicos. **9a edn**, (2014).
45. Ross & Pawlina. Histología. Texto y Atlas. Correlación con biología celular y molecular. **2015 7ª edn.**,
46. Planaguma J. tesis: anàlisi de l'expressió gènica diferencial en el càncer d'Endometri.
47. Lortet-Tieulent, J., Ferlay, J., Bray, F. & Jemal, A. International Patterns and Trends in Endometrial Cancer Incidence, 1978–2013. *JNCI J. Natl. Cancer Inst.* **110**, 1–8 (2017).
48. Morice, P., Leary, A., Creutzberg, C., Abu-Rustum, N., & Darai, E. Endometrial cancer. *Lancet* **387**, 1094–1108 (2016).
49. Kurman, R.J., Carcangiu, M.L., Herrington, C.S. & Young, R.H. WHO classification of tumours of the female reproductive organs. WHO classification of tumours. *IARC Press. Lyon* (2014).
50. Pecorelli, S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obs.* **105**, 103–104 (2009).

Bibliografia

51. Amant, F., Mirza, M. R., Koskas, M. & Creutzberg, C. L. Cancer of the corpus uteri. *Int. J. Gynecol. Obstet.* **143**, 37–50 (2018).
52. Conlon, N., Leitao, M.M., Abu-Rustum, N.R. & Soslow, R. Grading uterine endometrioid carcinoma: a proposal that binary is best. *Am J Surg Pathol.* **38**, 1583–1587 (2014).
53. Soslow, R. A., Tornos, C., Park, K.J., Malpica, A., Matias-Guiu, X., Oliva, E. *et al.* Endometrial Carcinoma Diagnosis. *Int. J. Gynecol. Pathol.* **38**, S64–S74 (2018).
54. Gatus, S. & Matias-Guiu, X. Practical issues in the diagnosis of serous carcinoma of the endometrium. *Mod. Pathol.* **29**, S45–S58 (2016).
55. Bokhman JV. Two pathogenetic types of endometrial carcinoma. *Gynecol. Oncol.* **15**, 10–17 (1983).
56. Murali, R., Soslow, R. A. & Weigelt, B. Classification of endometrial carcinoma: More than two types. *Lancet Oncol.* **15**, e268–e278 (2014).
57. Gilks, C.B., Oliva, E. & Soslow, R. Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma. *Am. J. Surg. Pathol.* **37**, 874–881 (2013).
58. Murali, R., Davidson, B., Fadare, O., Carlson, J.A., Crum, C.P., Gilks, C.B. *et al.* High-grade Endometrial Carcinomas. *Int. J. Gynecol. Pathol.* **38**, S40–S63 (2019).
59. The cancer Genome Atlas Network. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
60. Piulats, J.M., Guerra, E., Gil-Martín, M., Roman-Canal, B., Gatus, S., Sanz-Pamplona, R. *et al.* Molecular approaches for classifying endometrial carcinoma. *Gynecol Oncol* **145**, 200–207 (2017).
61. Stelloo, E., Nout, R.A., Osse, E.M., Jürgenliemk-Schulz, I.J., Jobsen, J.J., Lutgens, L.C. *et al.* Improved risk assessment by integrating molecular and clinicopathological factors in early-stage endometrial cancer-combined analysis of the PORTEC cohorts. *Clin. Cancer Res.* **22**, 4215–4224 (2016).
62. Talhouk, A., McConechy, M.K., Leung, S., Li-Chang, H.H., Kwon, J.S., Melnyk, N., *et al.* A clinically applicable molecular-based classification for endometrial cancers. *Br. J. Cancer* **113**, 299–310 (2015).
63. Talhouk, A. & McAlpine, J. N. New classification of endometrial cancers: the development and potential applications of genomic-based classification in research and clinical care. *Gynecol. Oncol. Res. Pract.* **3**, 14 (2016).
64. Bosse, T., Nout, R.A., McAlpine, J.N., McConechy, M.K., Britton, H., Hussein, Y.R. *et al.* Molecular Classification of Grade 3 Endometrioid Endometrial Cancers Identifies Distinct Prognostic Subgroups. *Am. J. Surg. Pathol.* **42**, 561–568 (2018).
65. Singh, N., Hirschowitz, L., Zaino, R., Alvarado-Cabrero, I., Duggan, M.A., Ali-Fehmi, R. *et al.* Pathologic Prognostic Factors in Endometrial Carcinoma (Other Than Tumor Type and Grade). *Int. J. Gynecol. Pathol.* **38**, S93–S113 (2019).

66. Bendifallah, S., Canlorbe, G., Raimond, E., Hudry, D., Coutant, D., Graesslin, O. *et al.* A clue towards improving the European Society of Medical Oncology risk group classification in apparent early stage endometrial cancer? Impact of lymphovascular space invasion. *Br. J. Cancer* **110**, 2640–2646 (2014).
67. Shinbrot, E., Henninger, E.E., Weinhold, N., Covington, K.R., Göksenin, A.Y., Schultz, N. *et al.* Exonucleasemutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res.* **24**, 1740–1750 (2014).
68. Rayner, E., van Gool, I.C., Palles, C., Kearsley, S.E., Bosse, T., Tomlinson, I. *et al.* A panoply of errors: polymerase proofreading domain mutations in cancer. *Nat Rev Cancer* **16**, 71–81 (2016).
69. Temko, D., Van Gool, I.C., Rayner, E., Glaire, M., Makino, S., Brown, M. *et al.* Somatic POLE exonuclease domain mutations are early events in sporadic endometrial and colorectal carcinogenesis, determining driver mutational landscape, clonal neoantigen burden and immune response. *J. Pathol.* **245**, 283–296 (2018).
70. Glaire, M.A., Brown, M., Church, D.N. & Tomlinson, I. Cancer predisposition syndromes: lessons for truly precision medicine. *J. Pathol.* **241**, 226–235 (2016)
71. Hussein, Y.R., Weigelt, B., Levine, D.A., Schoolmeester, J.K., Dao, L.N., Balzer, B.L. *et al.* Clinicopathological analysis of endometrial carcinomas harboring somatic POLE exonuclease domain mutations. *Mod Pathol* **28**, 505–514 (2015).
72. Church, D.N., Briggs, S.E., Palles, C., Domingo, E., Kearsley, S.J., Grimes, J.M. *et al.* DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum Mol Genet* **22**, 2820–2828 (2013).
73. McConechy, M.K., Talhouk, A., Leung, S., Chiu, D., Yang, W., Senz, J. *et al.* Endometrial carcinomas with POLE exonuclease domain mutations have a favorable prognosis. *McConechy MK, Talhouk A, Leung S, al.* **22**, 2865–2873 (2016).
74. Meng, B., Hoang, L.N., McIntyre, J.B., Duggan, M.A., Nelson, G.S., Lee, C.H. *et al.* POLE exonuclease domain mutation predicts long progression-free survival in grade 3 endometrioid carcinoma of the endometrium. *Gynecol Oncol* **134**, 15–19 (2014).
75. Howitt, B.E., Shukla, S.A., Sholl, L.M., Ritterhouse, L.L., Watkins, J.C., Rodig, S. *et al.* Association of polymerase e-mutated and microsatellite-instable endometrial cancers with neoantigen load, number of tumor-infiltrating lymphocytes, and expression of PD-1 and PD-L1. *JAMA Oncol* **1**, 1319–1323 (2015).
76. León-Castillo, A., Gilvazquez, E., Nout, R., Smit, V.T., McAlpine, J.N., McConechy, M., *et al.* Clinicopathological and molecular characterisation of ‘multiple-classifier’ endometrial carcinomas. *J. Pathol.* (2019).
77. Thibodeau, S.N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science (80-)*. **260**, 816–819 (1993).
78. Popat, S., Hubner, R. & Houlston, R. Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol* **23**, 609–618 (2005).

Bibliografia

79. Sinicrope, F.A., Foster, N.R., Thibodeau, S.N., Marsoni, S., Monges, G., Labianca, R. *et al.* DNA mismatch repair status and colon cancer recurrence and survival in clinical trials of 5-fluorouracil-based adjuvant therapy. *J Natl Cancer Inst* **103**, 863–875 (2011).
80. Thibodeau, S.N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Nature* **260**, 816–819 (1993).
81. Matias-Guiu, X. & Prat, J. Molecular pathology of endometrial cancer. *Mol. Surg. Pathol.* **62**, 111-23(2013).
82. McMeekin, D. S., Trichtler, D.L., Cohn, D.E., Mutch, D.G., Lankes, H.A., Geller, M.A. *et al.* Clinicopathologic significance of mismatch repair defects in endometrial cancer: An NRG oncology/gynecologic oncology group study. *J. Clin. Oncol.* **34**, 3062–3068 (2016).
83. Petitjean, A., Achatz, M., Borresen-Dale, A., Hainaut, P. & Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene* **26**, 2157–65 (2007).
84. Tan, E. & Cheok, C. Bringing p53 into the Clinic. *J Cancer Sci Ther.* **6**, 363–9 (2014).
85. López, I., Oliveira, L.P., Tucci, P., Álvarez-Valín, F., Coudry, R. & Marin, M. Different mutation profiles associated to P53 accumulation in colorectal cancer. *Gene* **499**, 81–7 (2012).
86. Schultheis, A.M., Martelotto, L.G., De Filippo, M.R., Piscuglio, S., Ng, C.K., Hussein, Y.R. *et al.* TP53 Mutational Spectrum in Endometrioid and Serous Endometrial Cancers. *Int J Gynecol Pathol* **35**, 289–300 (2016).
87. Murnyák, B. & Hortobágyi, T. Immunohistochemical correlates of TP53 somatic mutations in cancer. *Oncotarget* **7**, 64910-64920 (2016).
88. Muller, P.A. & Vousden, K.H. p53 mutations in cancer. *Nat Cell Biol* **15**, 2–8 (2013).
89. Smith, J. C. & Sheltzer, J. M. Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. *Elife* **7**, 1–26 (2018).
90. Menghi, F., Barthel, F.P., Yadav, V., Tang, M., Ji, B., Tang, Z., Carter, G.W. *et al.* The Tandem Duplicator Phenotype Is a Prevalent Genome-Wide Cancer Configuration Driven by Distinct Gene Mutations. *Cancer Cell* **34**, 197-210.e5 (2018).
91. Graham, N. A., Minasyan, A., Lomova, A., Cass, A., Balanis, N.G., Friedman, M. *et al.* Recurrent patterns of DNA copy number alterations in tumors reflect metabolic selection pressures. *Mol. Syst. Biol.* **13**, 914 (2017).
92. O’Hara, A. J. & Bell, D. W. The genomics and genetics of endometrial cancer. *Adv. Genomics Genet.* **2012**, 33–47 (2012).
93. Marenne, G. Statistical Methods To Combine Snp and Cnv Information in Genome-Wide Association Studies : an Application To Bladder Cancer. (2012).

94. Kuhn, E., Wu, R.C., Guan, B., Wu, G., Zhang, J., Wang, Y. *et al.* Identification of molecular pathway aberrations in uterine serous carcinoma by genome-wide analyses. *J. Natl. Cancer Inst.* **104**, 1503–1513 (2012).
95. Zhao, S., Choi, M., Overton, J.D., Bellone, S., Roque, D.M., Cocco, E., *et al.* Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc. Natl. Acad. Sci.* **110**, 2916–2921 (2013).
96. El-Sahwi, K. S., Schwartz, P. E. & Santin, A. D. Development of targeted therapy in uterine serous carcinoma, a biologically aggressive variant of endometrial cancer. *Expert Rev. Anticancer Ther.* **12**, 41–49 (2012).
97. Morrison, C., Zanagnolo, V., Ramirez, N., Cohn, D.E., Kelbick, N., Copeland, L. *et al.* HER-2 is an independent prognostic factor in endometrial cancer: association with outcome in a large cohort of surgically staged patients. *J Clin Oncol* **24**, 2376–2385 (2006).
98. Lopez, S., Zeybek, B. & Santin, A. D. Targeting Her2/neu in uterine serous carcinoma: A paradigm shift in management. *Oncotarget* **9**, 36652–36653 (2018).
99. Menderes, G., Lopez, S., Han, C., Altwerger, G., Gysler, S., Varughese, J., Schwartz, P.E. *et al.* Mechanisms of resistance to HER2-targeted therapies in HER2-amplified uterine serous carcinoma, and strategies to overcome it. *Discov Med.* **26**, 39–50 (2018).
100. Kuhn, E., Bahadirli-Talbott, A. & Shih, I. M. Frequent CCNE1 amplification in endometrial intraepithelial carcinoma and uterine serous carcinoma. *Mod. Pathol.* **27**, 1014–1019 (2014).
101. Geisen, C. & Moroy, T. The oncogenic activity of cyclin E is not confined to Cdk2 activation alone but relies on several other, distinct functions of the protein. *J Biol Chem* **277**, 39909–39918 (2002).
102. Donnellan, R. & Chetty, R. Cyclin E in human cancers. *FASEB J* **13**, 773–780 (1999).
103. Engelman, J.A., Luo, J. & Cantley, L. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nat. Rev. Genet.* **7**, 606–619 (2006).
104. Holst, F., Werner, H.M.J., Mjøs, S., Hoivik, E.A., Kusonmano, K., Wik, E. *et al.* PIK3CA amplification associates with aggressive phenotype but not markers of AKT-MTOR signaling in endometrial carcinoma. **25**, 334–345 (2019).
105. Salvesen, H. B., Carter, S.L., Mannelqvist, M., Dutt, A., Getz, G., Stefansson, I.M. *et al.* Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation. *Proc. Natl. Acad. Sci.* **106**, 4834–4839 (2009).
106. Goyama, S. & Kurokawa, M. Evi-1 as a critical regulator of leukemic cells. *Int J Hematol* **91**, 753–757 (2010).
107. Yates, L. R. & Campbell, P. J. Europe PMC Funders Group. **13**, 795–806 (2013).
108. Bard-Chapeau, E.A., Gunaratne, J., Kumar, P., Chua, B.Q., Muller, J., Bard, F.A. *et al.* EVI1 oncoprotein interacts with a large and complex network of proteins and integrates signals through protein phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, (2013).

Bibliografia

109. Bauckman, K., Campla, C. & Nanjundan, M. Dysregulated TGF β Signaling in Ovarian Cancer. In *Ovarian Cancer-Basic Science Perspective*. *Intech open* **2**, 121–138 (2012).
110. Osterberg, L., Levan, K., Partheen, K., Delle, U., Olsson, B., Sundfeldt, K. *et al.* Potential predictive markers of chemotherapy resistance in stage III ovarian serous carcinomas. *BMC Cancer* **9**, 368 (2009).
111. Ma, H., Li, Y., Wang, X., Wu, H. Qi, G., Li, R. *et al.* PBK, targeted by EVI1, promotes metastasis and confers cisplatin resistance through inducing autophagy in high-grade serous ovarian carcinoma. *Cell Death Dis.* **10**, (2019).
112. Fields, A., Justilien, V. & Murray, N. The chromosome 3q26 OncCassette: a multigenic driver of human cancer. *Adv Biol Regul* **60**, 47–63 (2016).
113. Meric-Bernstam, F. & Gonzalez-Angulo, A.M. Targeting the mTOR signaling network for cancer therapy. *J Clin Oncol* **27**, 2278–2287 (2009).
114. Li, X., Wu, C., Chen, N., Gu, H., Yen, A., Cao, L. *et al.* PI3K/Akt/mTOR signaling pathway and targeted therapy for glioblastoma. *Oncotarget* **7**, 33440–33450 (2016).
115. Oda K, Stokoe D, Taketani Y & McCormick, F. High frequency of coexistent mutations of PIK3CA and PTEN genes in endometrial carcinoma. *Cancer Res.* **65**, 10669–10673 (2005).
116. Rudd, M.L., Price, J.C., Fogoros, S., Fogoros, S., Godwin, A.K., Sgroi, D.C., Merino, M.J., *et al.* A unique spectrum of somatic PIK3CA (p110 α) mutations within primary endometrial carcinomas. *Clin. Cancer Res.* **17**, 1331–1340 (2011).
117. Urick, M.E., Rudd, M.L., Godwin, A.K, Sgroi, D., Merino, M., Bell, D.W. PIK3R1 (p85 α) is somatically mutated at high frequency in primary endometrial cancer. *Cancer Res.* **71**, 4061–4067 (2011).
118. Maxwell, G.L., Risinger, J.I., Gumbs, C., Shaw, H., Bentley, R.C., Barret, J.C. *et al.* Mutation of the PTEN tumor suppressor gene in endometrial hyperplasias. *Cancer Res.* **58**, 2500–2503 (1998).
119. Salvesen, H.B., Stefansson, I., Kalvenes, M.B., Das, S. & Akslen L.A. Loss of PTEN expression is associated with metastatic disease in patients with endometrial carcinoma. *Cancer* **94**, 2185–2191 (2002).
120. Hayes, M.P., Wang, H., Espinal-Witter, R., Douglas, W., Solomon, G.J., Baker, S.J. *et al.* PIK3CA and PTEN mutations in uterine endometrioid carcinoma and complex atypical hyperplasia. *Clin. Cancer Res.* **12**, 5932–5935 (2006).
121. Barandier, C., Ming, X.F. & Yang, Z. Small G proteins as novel therapeutic targets in cardiovascular medicine. *News Physiol Sci* **18**, 18–22 (2003).
122. Enomoto, T., Inoue, M., Perantoni, A.O., Buzard, G.S., Miki, H., Tanizawa, O. *et al.* K-ras activation in premalignant and malignant epithelial lesions of the human uterus. *Cancer Res.* **51**, 5308–5318 (1991).
123. Lagarda, H., Catusus, L., Arguelles, R., Matias-Guiu, X. & Prat J. K-ras mutations in endometrial carcinomas with microsatellite instability. *J Pathol* **193**, 193–199 (2001).

124. Sasaki, H., Nishii, H., Takahashi, H., Tada, A., Furusato, M., Terashima, Y., Siegal, G.P. *et al.* Mutation of the K-ras protooncogene in human endometrial hyperplasia and carcinoma. *Cancer Res.* **53**, 1906–1910 (1993).
125. Chaffer, C.L., Dopheide, B., Savagner, P., Thompson, E.W. & Williams, E.D. Aberrant fibroblast growth factor receptor signaling in bladder and other cancers. *Differentiation* **75**, 831–842 (2007).
126. Dutt, A., Salvesen, H.B., Chen, T.H., Ramos, A.H., Onofrio, R.C., Hatton, C. *et al.* Drug-sensitive FGFR2 mutations in endometrial carcinoma. *Proc Natl Acad Sci U S A.* **105**, 8713–8717 (2008).
127. Pollock, P.M., Gartside, M.G., Dejeza, L.C., Powell, M.A., Mallon, M.A., Davies, H. *et al.* Frequent activating FGFR2 mutations in endometrial carcinomas parallel germline mutations associated with craniosynostosis and skeletal dysplasia syndromes. *Oncogene* **26**, 7158–7162 (2007).
128. Byron, S.A., Gartside, M.G., Wellens, C.L., Mallon, M.A., Keenan, J.B., Powell, M.A. *et al.* Inhibition of activated fibroblast growth factor receptor 2 in endometrial cancer cells induces cell death despite PTEN abrogation. *Cancer Res.* **68**, 6902–6907 (2008).
129. Guan, B., Wang, T.L. & Shih, I. ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. *Cancer Res.* **71**, 6718–6727 (2011).
130. Wiegand, K.C., Lee, A.F., Al-Agha, O.M., Chow, C., Kalloger, S.E., Scott, D.W. *et al.* Loss of BAF250a (ARID1A) is frequent in high-grade endometrial carcinomas. *J Pathol* **224**, 328–333 (2011).
131. Guan, B., Mao, T.L., Panuganti, P.K., Kuhn, E., Kurman, R.J., Maeda, D. *et al.* Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. *Am J Surg Pathol.* **35**, 625–632 (2011).
132. Eichhorn, P.J., Creighton, M.P. & Bernards, R. Protein phosphatase 2A regulatory subunits and cancer. *Biochim Biophys Acta* **1795**, 1–15 (2009).
133. McConechy, M.K., Anglesio, M.S., Kalloger, S.E., Yang, W., Senz, J., Chow, C. *et al.* Subtype-specific mutation of PPP2R1A in endometrial and ovarian carcinomas. *J Pathol.* **223**, 567–573 (2011).
134. Shih, IeM., Panuganti, P.K., Kuo, K.T., Mao, T.L., Kuhn, E., Jones, S. *et al.* Somatic mutations of PPP2R1A in ovarian and uterine carcinomas. *Am J Pathol.* **178**, 1442–1447 (2011).
135. Welcker, M. & Clurman, B. E. FBW7 ubiquitin ligase: A tumour suppressor at the crossroads of cell division, growth and differentiation. *Nat Rev Cancer* **8**, 83–93 (2008).
136. Altevogt, P., Doberstein, K. & Fogel, M. L1CAM in human cancer. *Int. J. Cancer* **138**, 1565–1576 (2016).
137. Kiefel, H., Bondong, S., Hazin, J., Ridinger, J., Schirmer, U., Riedle, S. *et al.* L1CAM: a major driver for tumor cell invasion and motility. *Cell Adh. Migr.* **6**, 374–384 (2012).

Bibliografia

138. Chen, J., Gao, F. & Liu, N. L1CAM promotes epithelial to mesenchymal transition and formation of cancer initiating cells in human endometrial cancer. *Exp. Ther. Med.* **15**, 2792–2797 (2018).
139. McAlpine, J., Leon-Castillo, A. & Bosse, T. The rise of a novel classification system for endometrial carcinoma; integration of molecular subclasses. *J. Pathol.* **244**, 538–549 (2018).
140. Ganesh, K., Basnet, H., Kaygusuz, Y., Laughney, A.M., He, L., Sharma, R. *et al.* L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer. *Nat. Cancer* **1**, (2020).
141. Kommoss, F., Karnezis, A.N., Kommoss, F., Talhouk, A., Taran, F.A., Staebler, A. *et al.* L1cam further stratifies endometrial carcinoma patients with no specific molecular risk profile. *Br. J. Cancer* **119**, 480–486 (2018).
142. Machin, P., Catusus, L., Pons, C., Muñoz, J., Matias-Guiu, X. & Prat, J. CTNNB1 mutations and beta-catenin expression in endometrial carcinomas. *Hum Pathol* **33**, 206–212 (2002).
143. Schlosshauer, P.W., Pirog, E.C., Levine, R.L. & Ellenson, L.H. Mutational analysis of the CTNNB1 and APC genes in uterine endometrioid carcinoma. *Mod Pathol* **13**, 1066–1071 (2000).
144. Charo, L. M. & Plaxe, S. C. Recent advances in endometrial cancer: a review of key clinical trials from 2015 to 2019. *F1000Research* **8**, 1–13 (2019).
145. Eritja, N., Yeramian, A., Chen, B.J., Llobet-Navas, D., Ortega, E., Colas, E. *et al.* Endometrial Carcinoma: Specific Targeted Pathways. *Adv Exp Med Bio* **943**, 149–207 (2017).
146. Mitamura, T., Dong, P., Ihira, K., Kudo, M. & Watari, H. Molecular-targeted therapies and precision medicine for endometrial cancer. *Jpn. J. Clin. Oncol.* **49**, 108–120 (2019).
147. Le Gallo, M. & Bell, D. W. The emerging genomic landscape of endometrial cancer. *Clin. Chem.* **60**, 98–110 (2014).
148. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J. *et al.* Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
149. Carvalho-Silva, D., Pierleoni, A., Pignatelli, M., Ong, C., Fumis, L., Karamanis, N., *et al.* Open Targets Platform: New developments and updates two years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2019).
150. Wolff, A.C. Hammond MEH, Allison KH, *et al.* Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer. *J. Clin. Oncol.* **36**, 2105–2122 (2018).
151. Dopierala, J., Damato, B. E., Lake, S. L., Taktak, A. F. G. & Coupland, S. E. Genetic heterogeneity in uveal melanoma assessed by multiplex ligation-dependent probe amplification. *Invest. Ophthalmol. Vis. Sci.* **51**, 4898–905 (2010).

152. Matias-Guiu, X. & Davidson, B. Prognostic biomarkers in endometrial and ovarian carcinoma. *Virchows Arch.* **464**, 315–31 (2014).
153. Clarke, B.A. & Gilks, C.B. Endometrial carcinoma: controversies in histopathological assessment of grade and tumor cell type. *J. Clin. Pathol.* **63**, 410–415 (2010).
154. Hamilton, C. A., Cheung, M.K., Osann, K., Chen, L., Teng, N.N., Longacre, T.A. *et al.* Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers. *Br. J. Cancer* **94**, 642–646 (2006).
155. Soslow, R. A. High-grade endometrial carcinomas - strategies for typing. *Histopathology* **62**, 89–110 (2013).
156. Bartosch C., Lopes, J.M & Oliva, E. Endometrial carcinomas: a review emphasizing overlapping and distinctive morphological and immunohistochemical features. *Adv. Anat. Pathol* **18**, 415–437 (2011).
157. Garg, K. & Soslow, R.A. Strategies for distinguishing low-grade endometrioid and serous carcinomas of endometrium. *Adv. Anat. Pathol* **19**, 1–10 (2012).
158. Mauland, K.K., Wik, E., & Salvesen, H.B. Clinical value of DNA content assessment in endometrial cancer. *Cytom. B Clin Cytom.* **86**, 154–163 (2014).
159. Yeramian, A., Moreno-Bueno, G., Dolcet, X., Catusus, L., Abal, M., Colas E. *et al.* Endometrial carcinoma: molecular alterations involved in tumors development and progression. *Oncogene* **32**, 403–413 (2013).
160. Llobet, D., Pallares, J., Yeramian, A., Santacana, M., Eritja, N., Velasco, A. *et al.* Molecular pathology of endometrial carcinoma: practical aspects from the diagnostic and therapeutic viewpoints. *J. Clin. Pathol.* **62**, 777–785 (2009).
161. Chang, Y.S., Huang, H.D., Yeh, K.T. & Chan, J. Genetic alterations in endometrial cancer by targeted next-generation sequencing. *Exp Mol Pathol.* *Exp Mol Pathol* **100**, 8–12 (2016).
162. Lee, P.J., McNulty, S., Duncavage, E.J. & Heusel, J.W. Clinical Targeted Next-Generation Sequencing Shows Increased Mutational Load in Endometrioid-type Endometrial Adenocarcinoma With Deficient DNA Mismatch Repair. *Int. J. Gynecol. Pathol.* **37**, 581–589 (2017).
163. Mcconehey, M. K., Ding, J., Cheang, M.C., Wiegand, K., Senz, J., Tone, A. *et al.* Use of mutation profiles to refine classification of endometrial carcinomas. *J Pathol* **228**, 20–30 (2014).
164. Hoang, L.N., McConechy, M.K., Köbel, M., Han, G., Rouzbahman, M., Davidson, B. *et al.* Histotype-genotype correlation in 36 high-grade endometrial carcinomas. *Am J Surg Pathol.* **37**, 1421–1423 (2013).
165. Beroukhi, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).

Bibliografia

166. Proctor, L., Pradhan, M., Leung, S., Cheng, A., Lee, C.H., Soslow, R.A. *et al.* Assessment of DNA Ploidy in the ProMisE molecular subgroups of endometrial cancer. *Gynecol. Oncol.* **146**, 596–602 (2017).
167. Zhang, D., Wang, Z., Luo, Y., Xu, Y., Liu, Y., Yang, W. *et al.* Analysis of DNA copy number aberrations by multiple ligation-dependent probe amplification on 50 intestinal type gastric cancers. *J Surg Oncol.* **103**, 124–132 (2011).
168. Baykara, O., Bakir, B., Buyru, N., Kaynak, K. & Dalay, N. Amplification of chromosome 8 genes in lung cancer. *J. Cancer* **6**, 270–275 (2015).
169. Bunyan, D. J., Eccles, D.M., Sillibourne, J., Wilkins, E., Thomas, N.S., Shea-Simonds, J. *et al.* Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification. *Br. J. Cancer* **91**, 1155–1159 (2004).
170. Caines, R., Cochrane, A., Kelaini, S., Vila-Gonzalez, M., Yang, C., Eleftheriadou, M. *et al.* Cluster analysis of multiplex ligation-dependent probe amplification data in choroidal melanoma. *Mol. Vis.* **21**, 1–11 (2015).
171. Beckner, M. E., Pollack, I. F., Nordberg, M. L. & Hamilton, R. L. Glioblastomas with copy number gains in EGFR and RNF139 show increased expressions of carbonic anhydrase genes transformed by ENO1. *BBA Clin.* **5**, 1–15 (2016).
172. Hieronymus, H., Murali, R., Tin, A., Yadav, K., Abida, W., Moller, H. *et al.* Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. *Elife* **7**, 1–18 (2018).
173. Danielsen, H.E., Pradhan, M. & Novelli, M. Revisiting tumour aneuploidy - the place of ploidy assessment in the molecular era. *Nat. Rev. Clin. Oncol.* **13**, 291–304. (2016).
174. Atkin, N. B., Richards, B.M. & Ross, A.J. The Deoxyribonucleic Acid Content of Carcinoma of the Uterus: an Assessment of its Possible Significance in Relation to Histopathology and Clinical Course, Based on Data from 165 Cases. *Br. J. Cancer* **13**, 773–787 (1959).
175. Njølstad, T.S., Trovik, J., Hveem, T.S., Kjæreng, M.L., Kildal, W., Pradhan, M. *et al.* DNA ploidy in curettage specimens identifies high-risk patients and lymph node metastasis in endometrial cancer. *Br. J. Cancer* **112**, 1656–1664 (2015).
176. Bilbao-Sieyro, C., Ramirez, R., Rodriguez-Gonzalez, G., Falcon, O., Leon, L., Torres, S. *et al.* Microsatellite instability and ploidy status define three categories with distinctive prognostic impact in endometrioid endometrial cancer,5 (15) (2014 Aug 15) . *Oncotarget* **5**, 6206–6217 (2014).
177. Pradhan, M., Davidson, B., Abeler, V. M., Danielsen, H.E., Trope, C.G., Kristensen, G.B. *et al.* DNA ploidy may be a prognostic marker in stage I and II serous adenocarcinoma of the endometrium. *Virchows Arch* **461**, 291–298 (2012).
178. Cuevas, D., Valls, J., Gatus, S., Roman-Canal, B., Estaran, E., Dorca, E. *et al.* Targeted sequencing with a customized panel to assess histological typing in endometrial carcinoma. *Virchows Arch* **474(5)**, 585–598 (2019).

179. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: A looking glass for cancer? *Nat. Rev. Cancer* **12**, 323–334 (2012).
180. Gerlinger, M., Rowan, A.J., Horswell, S., Math, M., Larkin, J., Endesfelder, D. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
181. Gullickson, G., Ambrose, E.C., Hoover, R.G. & Kornbluth, J. Uridine Cytidine Kinase Like-1 Enhances Tumor Cell Proliferation and Mediates Protection from Natural Killer-Mediated Killing. *Int. J. Immunol. Immunother.* **3**, 1–8 (2016).
182. Menghi, F., Inaki, K., Woo, X., Kumar, P.A., Grzeda, K.R., Malhotra, A. *et al.* The tandem duplicator phenotype as a distinct genomic configuration in cancer. *Proc. Natl. Acad. Sci.* **113**, E2373–E2382 (2016).
183. Allison, K.H., Dintzis, S.M. & Schmidt R.A. Frequency of HER2 heterogeneity by fluorescence in situ hybridization according to CAP expert panel recommendations: time for a new look at how to report heterogeneity. *Am J Clin Pathol.* **136**, 864–871 (2011).
184. Chang, M.C., Malowany, J.I., Mazurkiewicz, J. & Wood, M. Genetic heterogeneity' in HER2/neu testing by fluorescence in situ hybridization: a study of 2,522 cases. *Mod. Pathol.* **25**, 683–688 (2012).
185. Fleming, G. F., Sill, M.W., Darcy, K.M., McMeekin, D.S., Thigpen, J.T., Adler, L.M. *et al.* Phase II Trial of Trastuzumab in Women with Advanced or Recurrent, HER2-Positive Endometrial Carcinoma: a Gynecologic Oncology Group Study. **116**, 15–20 (2010).
186. Fader, A. N., Roque, D.M., Siegel, E., Buza, N., Hui, P., Abdelghany, O. *et al.* Randomized Phase II trial of carboplatin-paclitaxel versus carboplatin-paclitaxel-trastuzumab in uterine serous carcinomas that overexpress human epidermal growth factor receptor 2/neu. *J. Clin. Oncol.* **36**, 2044–2051 (2018).
187. Buza, N., English, D. P., Santin, A. D. & Hui, P. Toward standard HER2 testing of endometrial serous carcinoma: 4-year experience at a large academic center and recommendations for clinical practice. *Mod. Pathol.* **26**, 1605–1612 (2013).
188. Buza, N. & Hui, P. Marked heterogeneity of HER2/NEU gene amplification in endometrial serous carcinoma. *Genes Chromosom. Cancer* **52**, 1178–1186 (2013).

ANNEX

Annex I: Descripció dels diferents camps utilitzats en l'anotació i la seva descripció

CAMPS DE L'ANOTACIÓ_DESCRIPCIÓ	
HGNC_symbol: HGNC gene nomenclature.	1KG_AFR_MAF: Non-reference allele frequency of existing variant in 1000 Genomes combined African population.
HGNC_ID: HGNC gene identifier	1KG_AFR_MA: Non-reference allele of existing variant in 1000 Genomes combined African population.
Gene: Ensembl stable ID of affected gene.	1KG_AMR_MAF: Non-reference allele frequency of existing variant in 1000 Genomes combined American population.
Gene_description: brief gene description.	1KG_AMR_MA: Non-reference allele of existing variant in 1000 Genomes combined Asian population.
Chr: Chromosome name.	1KG_ASN_MAF: Non-reference allele frequency of existing variant in 1000 Genomes combined Asian population.
Strand: The ADN strand on which the transcript/feature lies.	1KG_ASN_MA: Non-reference allele of existing variant in 1000 Genomes combined Asian population.
Pos: Genomic position.	1KG_EUR_MAF: Non-reference allele frequency of existing variant in 1000 Genomes combined European population.
Ref_Allele: Allele in the reference sequence.	1KG_EUR_MA: Non-reference allele of existing variant in 1000 Genomes combined European population.
Var_Allele: Allele observed in the sample.	1KG_EAS_MAF: Non-reference allele frequency of existing variant in 1000 Genomes combined East Asian population.
Sample_Genotype: Sample genotype assuming a diploid genome.	1KG_EAS_MA: Non-reference allele of existing variant in 1000 Genomes combined East Asian population.
Homo_ref: homozygous for the reference allele (Variant Frequency for SNVs (VFSNVs): 0; Variant Frequency for indels (VFindels): 0)	1KG_SAS_MAF: Non-reference allele frequency of existing variant in 1000 Genomes combined East Asian population.
P_Homo_ref: probable homozygous for the reference allele. (VFSNVs: >0 and <=0.12; VFindels: Not applicable)	1KG_SAS_MA: Non-reference allele of existing variant in 1000 Genomes combined South Asian population.
UNC_Hetero: uncertain heterozygous (VFSNVs: >0.12 and <0.35; VFindels: >0 y <0.3)	Conservation_score: The GERP (Genomic Evolutionary Rate Profiling) value is the estimation of conserving a position on the evolutionary scale by means of multiple alignments. The higher the value, the more conserved the position, so the impact of the variant on the transcript can be more dramatic. This value can be useful to prioritise candidate variants in Mendelian genetic diseases.
P_Hetero: probable heterozygous (VFSNVs: >=0.35 and <0.65; VFindels: >=0.3 y <0.6)	Grantham_distance: For missense variants, Grantham distance is a value that indicates the physico-chemical differences between amino acids.
UNC_Homo: uncertain homozygous for the variant (VFSNVs: >=0.65 and <0.85; VFindels: Not applicable)	Condel_prediction: Effect of the variant on the protein, predicted by the Condel algorithm.
P_Homo_var: probable homozygous for the variant allele (VFSNVs: >=0.85 and <=1; VFindels: >=0.6 y <=1)	SIFT_prediction: Effect of the variant on the protein, predicted by the SIFT algorithm.
Sample_Depth: Coverage value in that position.	PolyPhen_prediction: Effect of the variant on the protein, predicted by the Polyphen2 algorithm.
Sample_Ratio_Var/Depth: Frequency of the variant allele regarding the total number of sequences.	SWISSPROT: UniProtKB/Swiss-Prot identifier of protein product.
HGVSc_name: Variant identification code at cADN level.	TREMBL: UniProtKB/TrEMBL identifier of protein product.

CAMPS DE L'ANOTACIÓ_DESCRIPCIÓ

HGVSp_name: Variant identification code at protein level.	UNIPARC: UniParc identifier of protein product.
Variant_effect: Predicted effect of the variant on the transcript.	IMPACT: The impact modifier for the consequence type.
Feature_ID: Ensembl stable ID of transcript/feature.	TSL: Transcript support level (1 to 5).
RefSeq_ID: RefSeq accessions.	Intron: Intron number.
CCDS_ID: The CCDS identifier for this transcript.	Exon: Exon number.
Canonical_isoform: A flag indicating if the transcript is denoted as the canonical transcript for this gene.	Amino_acids: Amino acid change (only given if the variant affects the protein-coding sequence).
HGMD_accession: HGMD mutation accession number.	Codons: Codon change (the alternative codons with the variant base in upper case).
HGMD_variant_class: HGMD mutation variant class.	Feature_type: Feature type
HGMD_HGVS: HGMD mutation HGVS.	InterPro_IDs: InterPro IDs
HGMD_disease: HGMD mutation related disease.	InterPro_descriptions: InterPro Descriptions
HGMD_PUBMED: HGMD mutation related Pubmed IDs publications.	Affected_prot_domains: The source and the identifier of any overlapping protein domain.
PUBMED: Pubmed IDs of publications that cite existing variant.	Regulatory_Motif_name: The source and the identifier of a transcription factor binding profile (TFBP), aligned in this position.
ClinVar_ID: ClinVar ID (variant-phenotype association).	Regulatory_Motif_position: The relative position of the variant in the aligned TFBP.
ClinVar_Clinical_Sig: ClinVar clinical significance.	Regulatory_High_Inf_Pos: It indicates if the variant falls in a high information position of a TFBP.
ClinVar_Sig_Level: ClinVar significance level (1 to 5)	Regulatory_Motif_Score_Change: The difference in the motif score between the reference sequences and the TFBP variant.
Existing_variation:	cADN_start: cADN start position.
MIM_ID: OMIM gene identifier.	cADN_end: cADN end position.
MIM_MORBID_IDs: OMIM gene-phenotype association.	CDS_start: CDS start position.
Phenotype_association: Variant-Phenotype association (Ensembl-OMIM association).	CDS_end: CDS end position.
ESP_EA_MA: Non-reference allele of existing variant in NHLBI-ESP European American population.	Protein_start: Protein start position.
ESP_EA_MAF: Non-reference allele frequency of existing variant in NHLBI-ESP European American population.	Protein_end: Protein end position.
ESP_AA_MA: Non-reference allele of existing variant in NHLBI-ESP African American population.	Strand_Bias: Strand bias
ESP_AA_MAF: Non-reference allele frequency of existing variant in NHLBI-ESP African American population.	BIOTYPE: Biotype of feature/transcript or regulatory feature.
1KG_GMAF: Non-reference allele frequency of existing variant in 1000 Genomes.	Flanking_sequence: Flanking Sequence. 5' at 3', 50 bases on each side.
1KG_GMA: Non-reference allele of global existing variant in 1000 Genomes.	

Annex II: Publicacions

- **Targeted sequencing with a customized panel to assess histological typing in endometrial carcinoma**

Cuevas D, Valls J, Gatus S, Roman-Canal B, Estaran E, Dorca E, Santacana M, Vaquero M, Eritja N, Velasco A, Matias-Guiu X. *Virchows Arch.* 2019 May;474(5):585-598. doi: 10.1007/s00428-018-02516-2. Epub 2019 Feb 1.

- **Intratumour heterogeneity in endometrial serous carcinoma assessed by targeted sequencing and multiplex ligation-dependent amplification: a descriptive study**

Cuevas D, Velasco A, Vaquero M, Santacana M, Gatus S, Eritja N, Estaran E, Matias-Guiu X. *Histopathology.* 2020 Feb;76(3):447-460. doi: 10.1111/his.14001. Epub 2020 Jan 13.



Targeted sequencing with a customized panel to assess histological typing in endometrial carcinoma

Dolors Cuevas¹ · Joan Valls¹ · Sònia Gatius¹ · Berta Roman-Canal² · Elena Estaran¹ · Eduard Dorca² · Maria Santacana¹ · Marta Vaquero¹ · Núria Eritja¹ · Ana Velasco¹ · Xavier Matias-Guiu^{1,2}

Received: 22 November 2018 / Revised: 19 December 2018 / Accepted: 20 December 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The two most frequent types of endometrial cancer (EC) are endometrioid (EEC) and serous carcinomas (SC). Differential diagnosis between them is not always easy. A subset of endometrial cancers shows misleading microscopical features, which cause problems in differential diagnosis, and may be a good scenario for next-generation sequencing. Previous studies have assessed the usefulness of targeted sequencing with panels of generic cancer-associated genes in EC histological typing. Based on the analysis of TCGA (The Cancer Genome Atlas), EEC and SC have different mutational profiles. In this proof of principle study, we have performed targeted sequencing analysis with a customized panel, based on the TCGA mutational profile of EEC and SC, in a series of 24 tumors (16 EEC and 8 SC). Our panel comprised coding and non-coding sequences of the following genes: *ABCC9*, *ARID1A*, *ARID5B*, *ATR*, *BCOR*, *CCND1*, *CDH19*, *CHD4*, *COL11A1*, *CSDE1*, *CSMD3*, *CTCF*, *CTNNB1*, *EP300*, *ERBB2*, *FBXW7*, *FGFR2*, *FOXA2*, *KLLN*, *KMT2B*, *KRAS*, *MAP3K4*, *MKI67*, *NRAS*, *PGAP3*, *PIK3CA*, *PIK3R1*, *PPP2R1A*, *PRPF18*, *PTEN*, *RPL22*, *SCARNA11*, *SIN3A*, *SMARCA4*, *SPOP*, *TAF1*, *TP53*, *TSPYL2*, *USP36*, and *WRAP53*. Targeted sequencing validation by Sanger sequencing and immunohistochemistry was performed in a group of genes. *POLE* mutation status was assessed by Sanger sequencing. The most mutated genes were *PTEN* (93.7%), *ARID1A* (68.7%), *PIK3CA* (50%), and *KMT2B* (43.7%) for EEC, and *TP53* (87.5%), *PIK3CA* (50%), and *PPP2R1A* (25%) for SC. Our panel allowed correct classification of all tumors in the two categories (EEC, SC). Coexistence of mutations in *PTEN*, *ARID1A*, and *KMT2B* was diagnostic of EEC. On the other hand, absence of *PTEN*, *ARID1A*, and *KMT2B* mutations in the presence of *TP53* mutation was diagnostic of SC. This proof of concept study demonstrates the suitability of targeted sequencing with a customized endometrial cancer gene panel as an additional tool for confirming histological typing.

Keywords Endometrial carcinoma · Histological type · Endometrioid · Serous · Biomarker · Targeted sequencing

Introduction

In western countries, endometrial carcinoma (EC) is the most common cancer of the female genital tract. Prognosis is quite good for low-grade, early-stage tumors. However, clinical outcome is not good for some tumor types and advanced-stage

tumors. There are different histological types of EC. The two most common types of tumors are endometrioid carcinomas (EEC) and serous carcinomas (SC). Histological typing is a good prognostic indicator in EC [1–4].

Histological classification of EC is easy in the vast majority of cases. However, diagnosis may be difficult in some high-grade tumors, which sometimes show combined or mixed features [5–7]. Furthermore, some EEC may exhibit papillary architecture (characteristic of SC), and also some SC may show a predominant glandular arrangement (typical of EEC). Finally, each of them (EEC or SC) may show a predominant solid pattern. There is a poor inter-observer agreement in assessing histological type in this subset of tumors, even among experts [8].

Immunohistochemistry (IHC) has some value in histological typing. Some immunohistochemical markers, such as p53,

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00428-018-02516-2>) contains supplementary material, which is available to authorized users.

✉ Xavier Matias-Guiu
fjmatiasguiu.lleida.ics@gencat.cat

¹ Hospital Universitari Arnau de Vilanova, Universitat de Lleida, IRBLLEIDA, CIBERONC, Lleida, Spain

² Hospital Universitari de Bellvitge, IDIBELL, Barcelona, Spain

p16, PTEN, and ARID1A present a distinct immunostaining pattern between different types of tumors, but there are many exceptions, particularly in high-grade EEC. TCGA performed an integrating genomic, transcriptomic, and proteomic characterization of EC, based on array and sequencing technologies [9]. In this study, the mutation profile of EEC and SC was different. The most frequently mutated genes in EEC were *PTEN* (77.7%), *PIK3CA* (53.1%), *PIK3R1* (37.1%), *CTNNB1* (36.6%), *ARID1A* (35.4%), *KRAS* (24.6%), and *CTCF* (20.6%), while the most frequently mutated genes in SC were *TP53* (90.7%), *PIK3CA* (41.9%), *FBXW7* (30.2%), *PPP2R1A* (36.6%), and *CHD4* (16.3%).

The aim of the study was to check the usefulness of targeted sequencing analysis in assessing histological type in EC. To this purpose, we designed a specific customized gene panel, based on the sequencing profile of EEC and SC, according to TCGA results. The panel was tested in a series of 24 tumors (16 EEC and 8 SC), well characterized at the pathological level.

Materials and methods

Case selection

Twenty four tumors and their corresponding matched normal tissues were selected from the frozen tumor tissue bank of IRBLLEIDA. Sixteen corresponded to EEC and 8 to SC. All tumors were diagnosed at the Department of Pathology of Hospital Arnau de Vilanova of Lleida, by following the most recent World Health Organization (WHO) criteria. They were surgically staged and graded according to the International Federation of Gynecology and Obstetrics (FIGO) staging and grading systems. Table 1 shows the main clinical and pathological features, including the TCGA-based molecular classification of each of them with the surrogate markers (*POLE*, p53, MSH-6, PMS-2). Mean follow up period was 76 months (range 24 to 108). A specific informed consent was used. Matched formalin-fixed, paraffin-embedded tissue blocks were used for immunohistochemical validation.

Samples

For each case, after surgery, tumor and normal tissue were obtained and immediately frozen and stored at -80°C at the Tissue Bank of IRBLLEIDA. Tumor and normal tissue quality was assessed in hematoxylin and eosin (H&E) sections. The percentage volume of tumor, necrosis, and non-tumor tissue was quantified in each tissue block. DNA was purified using DNeasy Blood and Tissue Kit (Qiagen) according to manufacturer's instructions. For all samples, DNA purity was determined using a Nanodrop ND-1000 (Thermo Fisher Scientific)

and for targeted next-generation sequencing, DNA samples were further quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific), performed on the Qubit Fluorometer 1.0 (Invitrogen, Life Technologies).

Targeted sequencing analysis

Targeted sequencing analysis was performed to detect mutations in 40 genes that have been previously found to be recurrently mutated in endometrial carcinoma by the TCGA [9]. These genes were grouped in seven signaling pathways (Table 2) as cell cycle, chromatin organization, cell-cell communication, Wnt signaling, PI3K/AKT/mTOR signaling, RTK/RAS/MAPK signaling, and diverse which includes genes with different functionalities. Our TruSeq genetic panel design included 1333 amplicons (250 bp in length) that covered 98.59% of exons, introns, and untranslated regions of the selected genes. Custom amplicon library was created using 100 ng of genomic DNA from tumor and non-tumor specimens according to manufacturer's instructions (TruSeq Custom Amplicon (TSCA) library preparation guide). In brief, a custom pool containing specific oligos to targeted regions of interest was hybridized to genomic DNA samples. After removal of unbound oligos, an extension-ligation step was performed. Resulting products were amplified using primers that add multiplexing index sequences for sample tagging as well as common adapters required for cluster generation. Before pooling, normalization was performed by quantifying individual libraries using the Qubit fluorometer and then pooled based on equal concentrations. Pooled TruSeq libraries were sequenced using Miseq Reagent kit v3 600 cycles (Illumina).

Bioinformatic analysis

Bioinformatic analysis was performed by Sistemas Genómicos S.L. (Spain). Sistemas Genómicos (SG) pipeline included fastq files processing, base calling, quality score assignment, read alignment against the human reference genome version GRCh38/hg38, quality control of mapping quality, coverage analysis, variant calling, and variant annotation. After completion of the pipeline, manual inspection of the variant list was performed. With the paired normal-tumor analysis, we selected as potential somatic mutations those variants with a VAF (variant allelic frequency) above 0.15 and a depth above 100 in the tumor tissue and not present in the normal tissue. Somatic gene mutations were further filtered based on the canonical isoforms and the variant effect on the transcript, excluding the non-canonical isoforms and synonymous variants. To

Table 1 Clinical and pathological features, including TCGA-based molecular classification of the tumors by surrogate markers

Case	Age	Histological type	Grade	Stage	Molecular classification	Follow-up
1	86	EEC	II	IB	MSI (hypermutated)	NED
2	91	EEC	III	II	Serous-like	DOD
3	94	EEC	II	IB	MSI (hypermutated)	NED
4	91	EEC	II	IB	NSMP	NED
5	83	EEC	II	IA	NSMP	NED
6	60	EEC	III	IIIA	Double mutated ^a	NED
7	67	EEC	III	IB	Serous-like	NED
8	76	EEC	II	IB	MSI (hypermutated)	NED
9	77	EEC	III	IB	MSI (hypermutated)	NED
10	86	EEC	I	IB	NSMP	NED
11	73	EEC	III	IB	POLE-mutated	NED
12	77	EEC	III	IB	NSMP	NED
13	77	EEC	III	IB	MSI (hypermutated)	NED
14	77	EEC	II	II	NSMP	NED
15	78	EEC	I	IB	NSMP	DDD
16	59	EEC	III	IB	Double mutated ^a	NED
17	86	SC	III	IA	Serous-like	DOD
18	89	SC	III	IIIA	Serous-like	DOD
19	85	SC	III	IB	Serous-like	AWOD
20	76	SC	III	IA	Serous-like	DOD
21	81	SC	III	IA	NSMP	DOD
22	87	SC	III	IIIB	Serous-like	DOD
23	93	SC	III	IB	Serous-like	DOD
24	96	SC	III	II	Serous-like	DOD

^a Indicates double mutated cases characterized by both p53 “mutated” IHC pattern and MSI. *MSI (hypermutated)*, microsatellite unstable; *NSMP*, tumors with no specific molecular profile; *EEC*, endometrial endometrioid carcinoma; *SC*, serous carcinoma; *NED*, no evidence of disease; *DOD*, dead of disease; *DDD*, dead of different disease; *AWOD*, alive with other disease

restrict the analysis, mutations were also selected based on their frequency on population databases (NHLBI-ESP and 1000 Genomes Project) to discard known variants with a MAF (minor allele frequency) higher than 1%. Lastly, selected somatic variants were categorized into four according to their pathogenicity degree. Pathogenic variant (PV) category included variants previously reported on COSMIC database (Catalog of Somatic Mutations in Cancer), associated with disease and evaluated as deleterious with different pathogenicity predictors (Polyphen2, Sift, Condel, and Fathmm). Likely pathogenic variants (LPV) were variants reported on COSMIC with some evidence of disease association, but with absent data on pathogenicity prediction. Variants of unknown significance (VUS) included variants that had not been previously reported either as a disease associated or as a normal variant, but whose pathogenicity could not be established with certainty. And the likely benign or neutral variants (LBNV) were variants annotated on COSMIC but evaluated as neutral or benign through pathogenicity predictors.

Sanger sequencing validation and *POLE* mutation analysis

A total of 46 variants identified by targeted sequencing analysis were validated by Sanger sequencing. These variants comprise hotspots of the following genes: *TP53*, *PTEN*, and *ARID1A*. We also performed a *POLE* exons 9, 11, 13, and 14 Sanger sequencing analysis in all samples. Sequencing reactions were carried out with BigDye Terminator Kit v3.1 (Applied Biosystems, Thermo Fisher Scientific) and run on Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems).

Immunohistochemistry testing

Formalin-fixed, paraffin-embedded tissue blocks were sectioned at a thickness of 3 μ m, dried for 1 h at 65° before pre-treatment procedure of deparaffinization, rehydration, and epitope retrieval in the Pre-Treatment Module, PT-LINK (DAKO) at 95 °C for 20 min in 50×

Table 2 List of the 40 genes included in the customized panel

Gene symbol	Chr.	Gene target (bps)	Gene description	Related pathway
<i>ABCC9</i>	chr12	11,265	ATP binding cassette subfamily C member 9	Cell cycle
<i>ATR</i>	chr3	16,885	ATR serine/threonine kinase	
<i>CCND1</i>	chr11	1689	Cyclin D1	
<i>CSDE1</i>	chr1	421	Cold shock domain containing E1	
<i>EP300</i>	chr22	12,740	E1A binding protein P300	
<i>MKI67</i>	chr10	11,934	Marker of proliferation Ki-67	
<i>PPP2R1A</i>	chr19	4384	Protein Phosphatase 2 scaffold subunit A alpha	
<i>TP53</i>	chr17	3327	Tumors protein P53	
<i>WRAP53</i>	chr17	273	WD repeat containing antisense to TP53	
<i>ARID1A</i>	chr1	10,823	AT-rich interaction domain 1A	Chromatin organization
<i>ARID5B</i>	chr10	6419	AT-rich interaction domain 5B	
<i>CHD4</i>	chr12	12,781	Chromodomain helicase DNA-binding protein 4	
<i>KMT2B</i>	chr19	13,342	Lysine methyltransferase 2B	
<i>SMARCA4</i>	chr19	11,274	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 4	
<i>CDH19</i>	chr18	4550	Cadherin 19	Cell-cell communication
<i>COL11A1</i>	chr1	18,812	Collagen type XI alpha 1 chain	
<i>CTNNB1</i>	chr3	5164	Catenin beta 1	WNT signaling
<i>ERBB2</i>	chr17	9822	Erb-B2 receptor tyrosine kinase 2	PI3K/AKT/mTOR signaling
<i>PIK3CA</i>	chr3	7060	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	
<i>PIK3R1</i>	chr5	6625	Phosphoinositide-3-kinase regulatory subunit 1	
<i>PTEN</i>	chr10	3974	Phosphatase and tensin homolog	
<i>TAF1</i>	chrX	11,840	TATA-box binding protein associated factor 1	
<i>FGFR2</i>	chr10	7522	Fibroblast growth factor receptor 2	RTK/RAS/MAPK signaling
<i>KRAS</i>	chr12	1615	KRAS proto-oncogene, GTPase	
<i>MAP3K4</i>	chr6	9345	Mitogen-activated protein kinase kinase kinase 4	
<i>NRAS</i>	chr1	1606	NRAS proto-oncogene, GTPase	
<i>BCOR</i>	chrX	8733	BCL6 corepressor	Diverse
<i>CSMD3</i>	chr8	26,363	CUB and sushi multiple domains 3	
<i>CTCF</i>	chr16	4977	CCCTC-binding factor	
<i>FBXW7</i>	chr4	5905	F-box and WD repeat domain containing 7	
<i>FOXA2</i>	chr20	2303	Forkhead box A2	
<i>KLLN</i>	chr10	1265	Killin, P53 regulated DNA replication inhibitor	
<i>PGAP3</i>	chr17	1170	Post-GPI attachment to proteins 3	
<i>PRPF18</i>	chr10	2835	Pre-mRNA processing factor 18	
<i>RPL22</i>	chr1	947	Ribosomal protein L22	
<i>SCARNA11</i>	chr12	867	Small Cajal body-specific RNA 11	
<i>SIN3A</i>	chr15	8292	SIN3 transcription regulator family member A	
<i>SPOP</i>	chr17	3542	Speckle type BTB/POZ protein	
<i>TSPYL2</i>	chrX	3192	TSPY like 2	
<i>USP36</i>	chr17	6645	Ubiquitin specific peptidase 36	

In table are reported gene symbol, chromosome localization (Chr.), gene target region analyzed in base pairs (bps), gene description, and gene-related functional pathway

Tris/EDTA buffer, pH 9. Before staining the sections, endogenous peroxidase was blocked. Samples were subjected to IHC for p53, PTEN, ARID1A, MSH6, and PMS2. Antibody, clone, source, and dilution are shown

in Table 3. After incubation, the reaction was visualized with the EnVision FLEX Detection Kit (DAKO) using diaminobenzidine chromogen as a substrate. Sections were counterstained with hematoxylin. Appropriate

Table 3 Antibody, clone, source, and dilution

Antibody	Clone	Source	Dilution
p53	DO7	DAKO	RTU
ARID1A	EPR13501-73	DAKO	1/500
PTEN	6H2.1	DAKO	1/100
MSH6	EP49	DAKO	RTU
PMS2	EP51	DAKO	RTU

RTU, ready to use

negative controls were also tested. Scoring criteria for the five biomarkers were the following: p53 was scored as “mutated” pattern (strong staining in more than 80% of tumor cells or complete negative staining) versus “wild type” pattern; and PTEN, ARID1A, MSH6, and PMS2 were scored as positive (any staining) versus negative (absence of staining).

TCGA-based molecular classification by surrogate markers

Each sample was molecularly classified (Table 1) according to TCGA-based surrogate (*POLE* Sanger sequencing and immunohistochemical testing of p53, MSH-6, and PMS-2). Four categories were established: (1) microsatellite unstable (MSI) (hypermuted) class including cases with negative staining for MSH-6 or PMS-2, (2) serous-like category including cases with just p53 “mutated” IHC pattern, (3) NSMP class grouping tumors with no specific molecular profile, and (4) *POLE*-mutated category (ultramutated). Tumors with more than one surrogate marker (MSI and p53) were classified as double mutated. Tumors with *POLE* mutations were classified as *POLE*-mutated regardless of the presence of MSI or p53 “mutated” IHC pattern.

Statistical analysis

Absolute and relative (percentage) presence of mutations for each gene was computed in EEC and SC cases, computing the odds ratio (with small sample adjustment when no mutation was detected in EEC and/or SC cases) and its corresponding *p* value to assess the strength and significance of the association. To depict the ability of genes to discriminate between EEC and SC cases, two heatmaps were produced (one with all-altered genes and other with significant genes), using double hierarchical cluster analysis with Euclidean distance. All analyses were made using the open-source R statistical software (www.r-project.org). Threshold for significance level was set at $\alpha = 0.05$.

Results

Mutation profile by targeted sequencing

Targeted sequencing analysis produced a mean of 2,270,521 reads, 97.64% of which mapped on reference genome version GRCh38/hg38 with high quality without PCR duplicates. Mean coverage depth was of 1563× per sample, being average depth and reads of normal and tumor samples similar.

After applying the previously described filters, our analysis identified a total of 164 somatic variants (Supplementary Table 1). Somatic variants were classified according to variant effect on transcripts as missense (62%), frameshift (24%), stop gained (9%), splice acceptor (1%), protein-altering (1%), and inframe deletions (4%). These variants were also classified depending on their pathogenicity degree: 51% of pathogenic variants (PV), 21% of likely pathogenic variants (LPV), 24% of variants of unknown significance (VUS), and 4% of likely benign or neutral variants (LBNV) (Fig. 1).

In the EEC group, we found a mean of 7.8 variants per sample (range 3 to 19), and in the SC, a mean of 2.6 variants per sample (range 1 to 5). Table 4 shows the genes more frequently altered in EEC and SC. No mutations were found in *ABCC9*, *CSDE1*, *KLLN*, *PGAP3*, *SCARNA11*, *SPOP*, *TSPYL2*, and *WRAP53* genes.

Recurrent mutations affecting protein domains in most frequently mutated genes were identified using the Mutation Mapper (www.cbioportal.org). This tool allowed us to locate mutations in protein domains according to Pfam database and thus, to assess the relevance of detected variants. Most likely mutational hotspots in *PTEN*, *TP53*, *PIK3CA*, and *ARID1A* genes are shown in Fig. 2. The 45% (10/22) of *PTEN*-detected variants were located in the C2 domain and 31.8% (7/22) in the DPSc domain, being R130 the most frequent hotspot (71.4%, 5/7). The 72.8% (16/22) of *PTEN* variants generated a truncated protein, and the 22% (5/22) of the remaining variants were missense. In the case of *TP53* gene, the 86% (13/15) of mutations occurred in the p53 DNA-binding domain. Unlike *PTEN*, 80% (12/15) of *TP53* mutations were missense and just two variants built a truncated protein (13%). Interestingly, we found a different mapping of *PIK3CA* mutations depending on tumor type. SC samples had *PIK3CA* gene mutations located at the end of exon 2 and at the beginning of exon 3. Instead, EEC group harbored *PIK3CA* variants widely distributed. These cases showed those typical exon 9 and 20 hotspots at 25% and 37% frequencies, respectively. For *ARID1A* gene, mutations were distributed throughout the entire gene, without finding any specific hotspot, 64.3% were truncating mutations and 35.7% were missense mutations.

KMT2B gene mutations were not mapped in known protein domains. The 57% were missense variants where L417P mutation represented 75% of them, and the remaining 43%

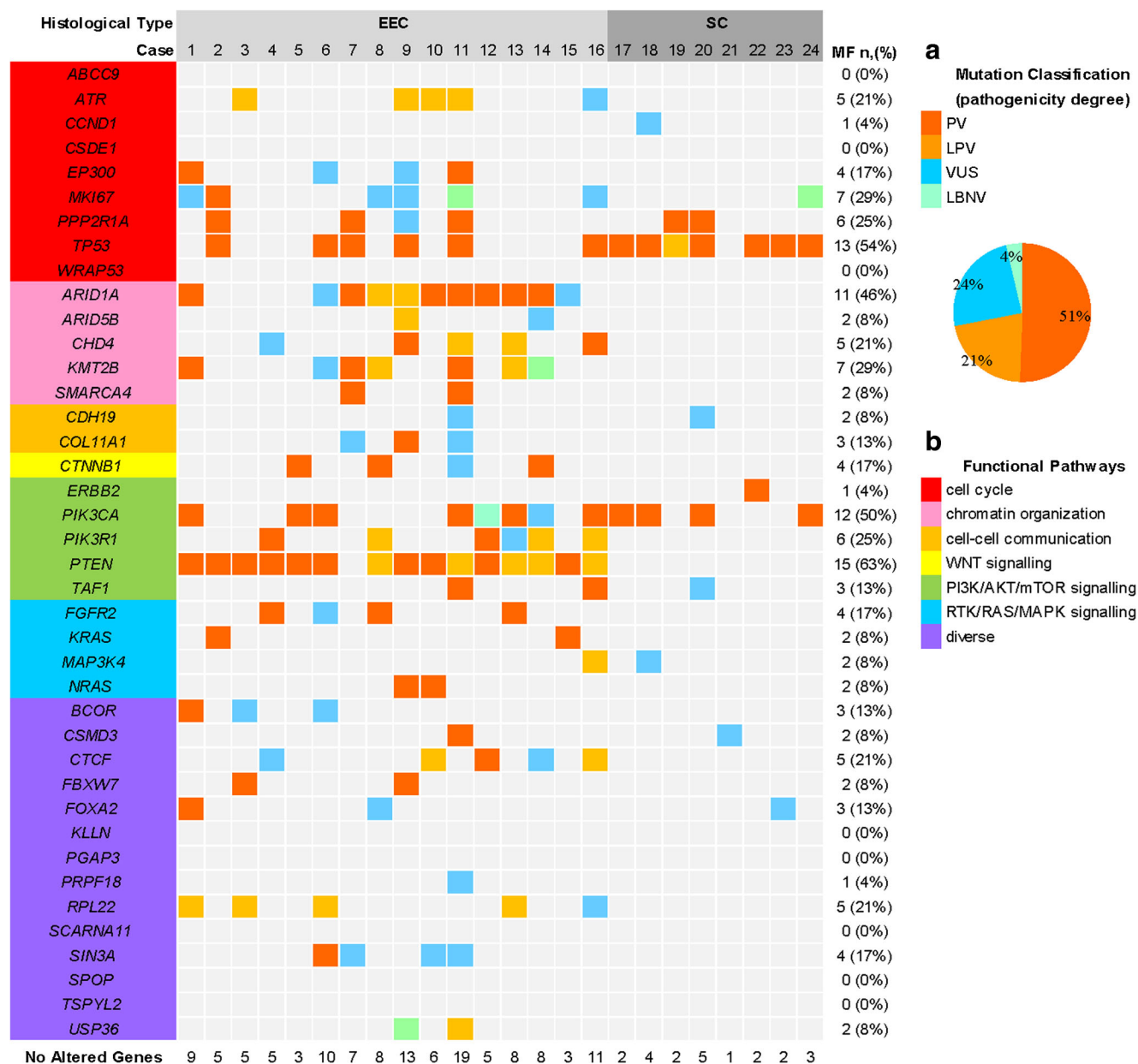


Fig. 1 Distribution of detected somatic mutations according to cases and analyzed genes. At the top of the *x*-axis, cases are grouped by histological type (EEC, endometrioid carcinomas; SC, serous carcinomas), and at the bottom, the number of altered genes per case is shown. At the right of the *y*-axis, mutation frequency (MF (*n* (%))) per gene is shown. Colored boxes were drawn to indicate: **a** The mutation classification according to pathogenicity degree: PV (pathogenic variants) in strong orange, LPV (likely pathogenic variants) in light orange, VUS (variants of unknown

significance) in blue, and LBNV (likely benign or neutral variants) in green. Pie chart below shows somatic mutation percentage based on the pathogenicity degree classification; and to indicate: **b** Functional pathways associated to analyzed genes: cell cycle in red, chromatin organization in pink, cell-cell communication in light orange, Wnt signaling in yellow, PI3K/AKT/mTOR signaling in green, RTK/RAS/MAPK signaling in blue, and diverse in purple

corresponded to frameshift mutations. All mutations detected in *PPP2R1A* gene were missense; from them, P179R was the most observed (50%, 3/6). Cases with more than one mutation in the same gene were found for *PTEN*, *ARID1A*, *PIK3CA*, and *TP53*.

Most frequently altered genes in our sample set belong to the following signaling pathways: PI3K/AKT/mTOR including *PTEN* and *PIK3CA*, cell cycle including *TP53* and *PPP2R1A*, and chromatin organization with *ARID1A* and *KMT2B*.

According to TCGA reports [9], all five samples with *RPL22* mutation harbored p.Lys15ArgfsTer5 variant and four of them presented lack of microsatellite stability (MSI).

Histological classification by targeted sequencing

Mutational profile of the 24 tumors was concordant with histological typing in all cases. Most frequently mutated genes

Table 4 Absolute number (*n*) and mutation frequency (%) of altered genes in EEC and SC

Gene symbol	Mutated cases, <i>n</i> (%)		Odds ratio		<i>p</i> value
	EEC (16 cases)	SC (8 cases)	EEC vs SC	SC vs EEC	
<i>PTEN</i> *	15 (93.75%)	0 (0%)	60.00		0.00001*
<i>ARID1A</i> *	11 (68.75%)	0 (0%)	14.67		0.002*
<i>TP53</i> *	6 (37.5%)	7 (87.5%)		9.62	0.03*
<i>KMT2B</i> *	7 (43.75%)	0 (0%)	5.60		0.03*
<i>PIK3R1</i>	6 (37.5%)	0 (0%)	4.36		0.06
<i>ATR</i>	5 (31.25%)	0 (0%)	3.33		0.1
<i>CHD4</i>	5 (31.25%)	0 (0%)	3.33		0.1
<i>CTCF</i>	5 (31.25%)	0 (0%)	3.33		0.1
<i>RPL22</i>	5 (31.25%)	0 (0%)	3.33		0.1
<i>CTNNB1</i>	4 (25%)	0 (0%)	2.46		0.17
<i>EP300</i>	4 (25%)	0 (0%)	2.46		0.17
<i>FGFR2</i>	4 (25%)	0 (0%)	2.46		0.17
<i>SIN3A</i>	4 (25%)	0 (0%)	2.46		0.17
<i>MKI67</i>	6 (37.5%)	1 (12.5%)	3.66		0.25
<i>BCOR</i>	3 (18.75%)	0 (0%)	1.71		0.28
<i>COL11A1</i>	3 (18.75%)	0 (0%)	1.71		0.28
<i>CCND1</i>	0 (0%)	1 (12.5%)		2.00	0.33
<i>ERBB2</i>	0 (0%)	1 (12.5%)		2.00	0.33
<i>FBXW7</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>KRAS</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>NRAS</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>SMARCA4</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>USP36</i>	2 (12.5%)	0 (0%)	1.07		0.43
<i>CDH19</i>	1 (6.25%)	1 (12.5%)		2.07	0.67
<i>CSMD3</i>	1 (6.25%)	1 (12.5%)		2.07	0.67
<i>MAP3K4</i>	1 (6.25%)	1 (12.5%)		2.07	0.67
<i>ARID5B</i>	1 (6.25%)	0 (0%)	1.01		0.67
<i>PRPF18</i>	1 (6.25%)	0 (0%)	1.01		0.67
<i>FOXA2</i>	3 (18.75%)	1 (12.5%)	1.48	0.67	0.76
<i>TAF1</i>	2 (12.5%)	1 (12.5%)		1.05	0.97
<i>PPP2R1A</i>	4 (25%)	2 (25%)	1.00	1.00	0.98
<i>PIK3CA</i>	8 (50%)	4 (50%)	1.00	1.00	1

*Indicate genes associated to tumor type ($p < 0.05$)

were *PTEN* (93.7%), *ARID1A* (68.7%), *PIK3CA* (50%), and *KMT2B* (43.7%) for EEC, and *TP53* (87.5%), *PIK3CA* (50%), and *PPP2R1A* (25%) for SC (Table 4).

Our panel allowed correct tumor classification in the two categories (EEC, SC). Tumor type was found to be associated with mutations in four different genes ($p < 0.05$), being *PTEN* (OR = 60, $p = 0.00001$), *ARID1A* (OR = 14.67, $p = 0.002$), and *KMT2B* (OR = 5.6, $p = 0.03$) associated with EEC type, and *TP53* (OR = 9.62, $p = 0.03$) with SC type (Table 4). Coexistence of mutations in *PTEN*, *ARID1A*, and *KMT2B* was diagnostic of EEC. On the other hand, absence of *PTEN*, *ARID1A*, and *KMT2B* in the presence of *TP53*

mutation was diagnostic of SC. *TP53* and *PIK3CA* mutations were present in both tumor types.

Figure 3 shows the heatmaps depicting the presence of mutations in all 32 altered genes and those statistically significant, showing that cases with the same tumor type (EEC or SC) are classified correctly.

Sanger validation and *POLE* mutation analysis

Sanger sequencing analysis allowed the correct identification of the same variants found by targeted sequencing for *TP53*, *PTEN*, and *ARID1A* genes (Table 5).

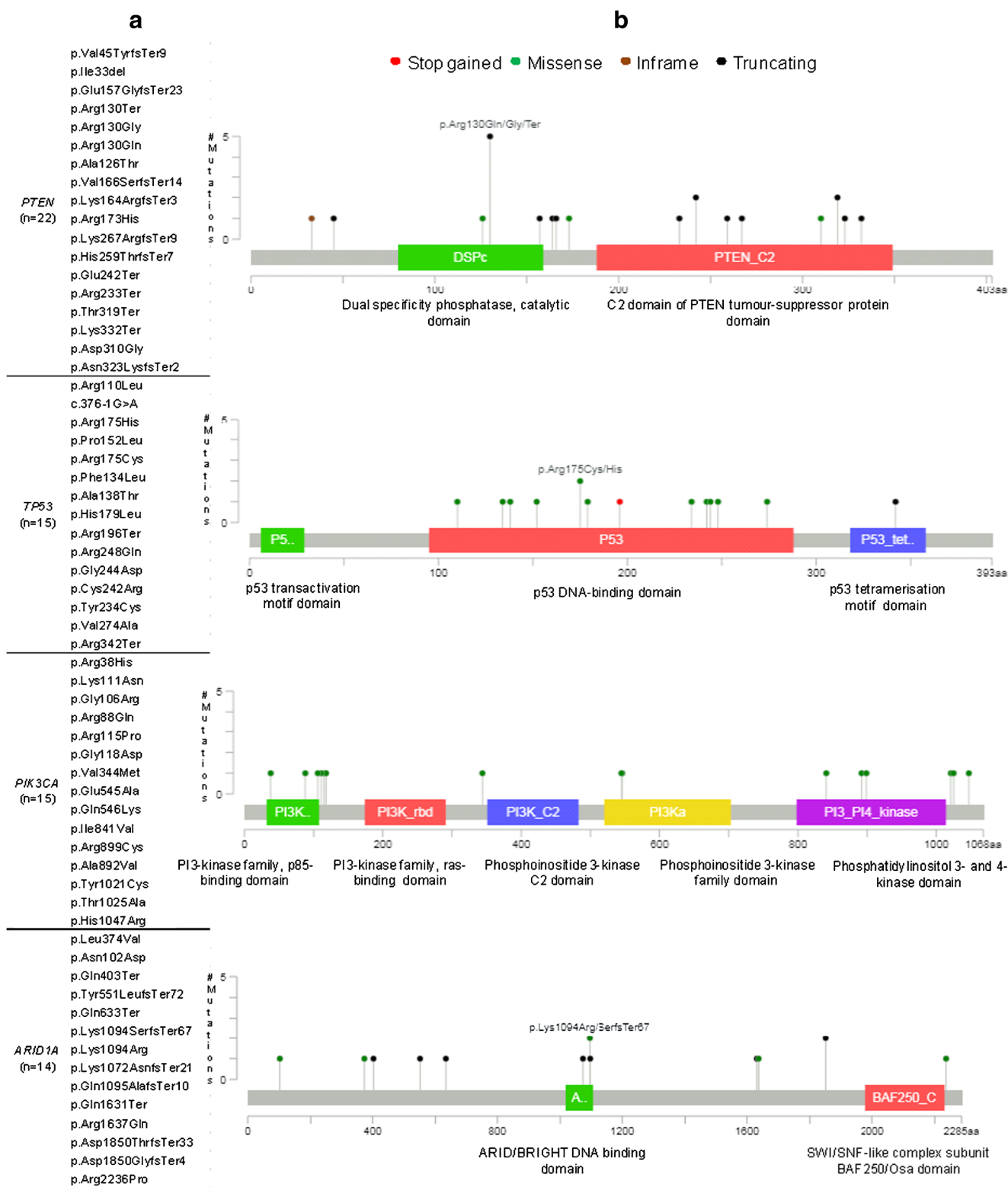


Fig. 2 a List of somatic mutations identified in *PTEN*, *TP53*, *PIK3CA*, and *ARID1A* genes. b Lollipop plots mapping detected somatic mutations on respective linear protein and its Pfam domains (Mutation Mapper

tool). The y-axis indicates the number of mutations, and the x-axis represents the protein sequence with numbers below indicating amino acid sequence positions

Additionally, *POLE* mutation analysis by Sanger sequencing demonstrated *POLE* wild-type sequence in all

but one of the cases, which showed the *POLE* gene variant p.Y458N.

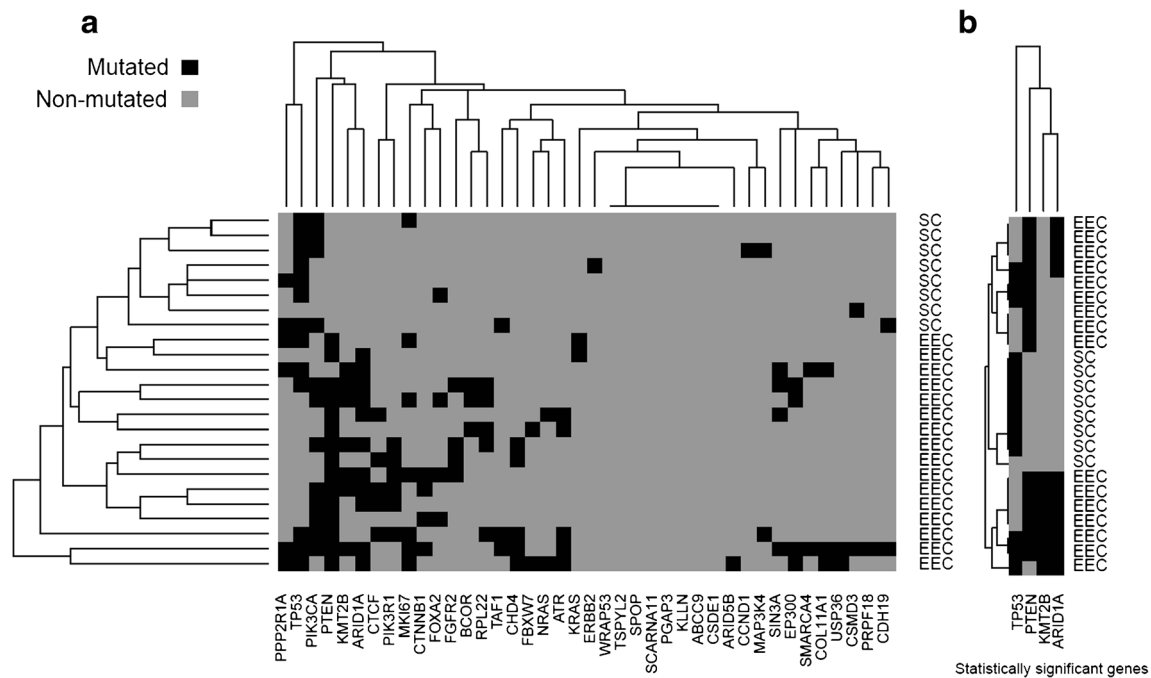


Fig. 3 **a** Heatmap depicting presence of mutations in all 32 altered genes (columns) for EEC and SC cases (rows). **b** Heatmap showing the presence of mutations in four statistically significant genes (columns) for EEC

and SC cases (rows). Note that heatmaps were created using a double hierarchical cluster analysis with a Euclidean distance

Immunohistochemical validation

Immunohistochemical validation was performed for *TP53*, *PTEN*, and *ARID1A* genes (Table 5). Moreover, immunostaining for MSH6 and PMS2 was done. Correlation between immunohistochemistry and sequencing analysis was obtained in the vast majority of cases. Twelve of the 13 *TP53* gene-mutated tumors resulted in a “mutated” p53 IHC pattern. *PTEN* negative staining was seen in 13 of 15 mutated tumors. Negative staining for *ARID1A* was observed in 10 of the 12 *ARID1A*-mutated tumors. Negative immunostaining for MSH6 was obtained in 2 cases, while negativity for PMS2 was detected in 7 cases. Overall mismatch repair deficiency was detected in 9 cases; all of them are EEC; and in 2 cases, mismatch repair deficiency coexisted with p53 “mutated” IHC pattern. Cases with mismatch repair deficiency had a range of 7 to 19 mutations (mean 10, median 9). Cases without mismatch repair deficiency had a range of 1 to 6 mutations (mean 3.7, median 4). The only one *POLE*-mutated (case 11) was an EEC grade 3 tumor which showed negative staining for PMS2 as well as the highest number of mutations (19) including *TP53*, *PTEN*, and *ARID1A* genes.

Discussion

Histological type is considered an important biologic predictor in EC [1–3]. EEC grades 1 and 2 tumors (EEC 1 and 2) are usually associated with a favorable outcome, while prognosis

for SC is worse. EEC grade 3 (EEC 3) is a heterogeneous group of tumors, with an adverse prognosis subgroup [10]. Recurrence occurs in about 20% of EEC and 50% of SC. SC and EEC 3 tumors had been compared using Surveillance, Epidemiology, and End Results (SEER) program data from 1988 to 2001. They represented 10% and 15% of EC, respectively, but accounted for 39% and 27% of endometrial cancer deaths, respectively [11]. Moreover, assessment of histological type determines the extent of initial surgical procedure and subsequent use of adjuvant therapy [4]. Although there is moderate to excellent (0.62–0.87) reproducibility in histological typing, inter-observer agreement is worse in high-grade tumors [8].

EEC 1 and 2 are usually composed of cells arranged in a glandular pattern of growth which is reminiscent of that of the proliferative endometrium, while EEC 3 has a predominant solid arrangement, and serous carcinomas have a complex architectural pattern with papillae and cellular budding [4]. However, SC with prominent glandular pattern is frequently mistaken as EEC 1 and 2; on the other hand, EEC with papillary pattern is sometimes interpreted as SC [4–7]. Differential diagnosis is difficult in a subset of cases.

The Cancer Genome Atlas (TCGA) has provided an integrated genomic characterization of EC and has proposed a new molecular classification in four groups of tumors [9]. Group 1 encompasses EEC with 7% of somatic inactivating mutations in *POLE* exonuclease (ultramutated), associated with good prognosis. They have very high mutation rate (232×106 mutations/Mb) and are characterized by mutations

Table 5 *ARID1A*, *PTEN*, and *TP53* mutation validation analysis by the Sanger sequencing and immunohistochemistry

Case	<i>ARID1A</i>			<i>PTEN</i>			<i>TP53</i>		
	Targeted	Sanger	IHC	Targeted	Sanger	IHC	Targeted	Sanger	IHC
1	K1072NfsTer21 R2236P	K1072NfsTer21 R2236P	NS	K332*	K332*	NS	WT	ND	“WT”P
2	WT	ND	PS	D310G	D310G	NS	R175H	R175H	“M”P
3	WT	ND	PS	R233*	R233*	NS	WT	ND	“WT”P
4	WT	ND	PS	R130*	R130*	NS	WT	ND	“WT”P
5	WT	ND	PS	R130G	R130G	PS	WT	ND	“WT”P
6	K1094SfsTer67	K1094SfsTer67	NS	A126T K267RfsTer9	A126T K267RfsTer9	NS	R248Q P152L	R248Q P152L	“M”P
7	R1637Q	R1637Q	PS	WT	ND	PS	R175C	R175C	“M”P
8	Y551LfsTer72	Y551LfsTer72	NS	T319*	T319*	NS	WT	ND	“WT”P
9	D1850TfsTer4	D1850TfsTer4	NS	R130* R173H	R130* R173H	NS	G244D	G244D	“WT”P
10	Q633*	ND	NS	V45YfsTer9 H259TfsTer7	V45YfsTer9 H259TfsTer7	NS	WT	ND	“WT”P
11	K1094R D1850GfsTer4	K1094R D1850GfsTer4	PS	E157GfsTer23 N323KfsTer2	E157GfsTer23 N323KfsTer2	NS	F134 L	F134 L	“M”P
12	Q403*	Q403*	NS	R130* R130Q	R130* R130Q	PS	WT	ND	“WT”P
13	Q1095AfsTer10	Q1095AfsTer10	NS	E242*	E242*	NS	WT	ND	“WT”P
14	Q1631*	Q1631*	NS	K164RfsTer3	ND	NS	WT	ND	“WT”P
15	N102D L374 V	ND	NS	E242*	E242*	NS	WT	ND	“WT”P
16	P373R [#] Q372SfsTer19 [#]	ND	NS	I33del T319*	I33del T319*	NS	A138T	A138T	“M”P
17	WT	ND	PS	WT	ND	PS	R196*	R196*	“M”P
18	WT	ND	PS	WT	ND	PS	V274A	V274A	“M”P
19	WT	ND	PS	WT	ND	PS	R342*	R342*	“M”P
20	WT	ND	PS	WT	ND	PS	C242R	C242R	“M”P
21	WT	ND	PS	WT	ND	NV	WT	ND	“WT”P
22	WT	ND	PS	WT	ND	NV	Splice acceptor c.376-1G>A	Splice acceptor c.376-1G>A	“M”P
23	WT	ND	PS	WT	ND	NS	H179L R110L	H179L R110L	“M”P
24	WT	ND	PS	WT	ND	NV	Y234C	Y234C	“M”P

[#] Indicate *ARID1A* mutations below the 100× tumor depth filter (case 16 tumor depth = 4); *Indicate that the consequence of the mutation at protein level is a termination codon. *NS*, negative staining; *PS*, positive staining; *NV*, not valuable; “*WT*”*P*, “wild type” pattern; “*M*”*P*, “mutated” pattern; *ND*, not done

in *PTEN*, *PIK3CA*, *PIK3R1*, *FBXW7*, *ARID1A*, *KRAS*, and *ARID5B*. Group 2 includes EEC with microsatellite instability (hypermuted), frequently associated with *MLH-1* promoter hypermethylation (28%). They have high mutation rate (18×10^6 mutations/Mb) and show mutations in *PTEN*, *RPL22*, *KRAS*, *PIK3CA*, *PIK3R1*, and *ARID1A*. Group 3 tumors include EEC with low copy-number alterations (39%), also called EC with no specific molecular profile. They show low mutation rate (2.9×10^6 mutations/Mb) including mutations in *PTEN*, *CTNBN1*, *PIK3CA*, *PIK3R1*, and *ARID1A*. Group 2 and group 3 both show similar progression-free survival rates. Finally, group 4 called serous-like or copy-number high (26%) show frequent *TP53* mutations and worse prognosis and is predominantly composed of most (but not all) SC, but also

some EEC (many EEC 3, but also some EEC 1 and 2). They have low mutation rate (2.3×10^6 mutations/Mb), with frequent mutations in *TP53*, *PPP2R1A*, and *PIK3CA* and show chromosomal instability, with recurrent gene amplifications for *MYC*, *ERBB2*, *CCNE1*, *FGFR3*, and *SOX17*.

Although molecular classification is already important to predict prognosis, histological typing remains the gold standard for tumor stratification. Combination of POLE mutational analysis and IHC testing of p53 and mismatch repair proteins (PMS-2 and MSH-6) has been proposed as a surrogate strategy to classify EC into the four TCGA groups [10, 12, 13]. There is some controversy regarding the possibility that the molecular classification would replace the traditional histological typing in daily practice. The main reason in favor of

molecular classification is the poor inter-observer agreement in histological typing in a subset of EC. Incorporation of TCGA-based surrogate classification into clinical practice may carry important advantages. Identification of *POLE*-mutated and serous-like/high copy-number tumors may lead to a much better clinical management of these patients. Molecular classification still needs further validation to solve some unresolved issues. For example, a subset of *POLE*-mutated and microsatellite unstable tumors show also p53 “mutated” IHC pattern and *TP53* mutation (“double positive”). These cases raise the differential diagnosis with serous-like/high copy-number EC. Moreover, the current TCGA-based surrogate approaches are not perfect. Nowadays, p53 immunostaining is used to identify the serous-like or copy-number high group. However, the correlation between p53 immunostaining and *TP53* mutation is not perfect. Using p53 immunostaining as a surrogate, *TP53* wild-type serous-like tumors are incorrectly classified as EC with no specific molecular profile. Finally, there are some preliminary evidences suggesting that, even in the serous-like group, the pattern of metastasis may be different, depending on histological type. The most probable scenario for the future is the coexistence of these classification systems, traditional histological typing, and molecular classification [14]. Based on that idea, it remains interesting to improve the tools for distinguishing EEC from SC.

Several IHC biomarkers have been shown to have a differential immunostaining pattern in EEC, including EEC 3, as opposed to SC. Some of these IHC markers were previously found to be differentially expressed between EEC and SC by cDNA analysis [15–17]. They include estrogen receptor, *PTEN*, *HER2*, claudins 3 and 4, *Nrf2*, p53, p16, *FOLR-1*, *HMG2*, cyclin E, *IMP2*, and *IMP3* [18–23]. Several authors have attempted the use of panels of antibodies to help in diagnosis and prognosis [24–26]. There is no single or antibody combination profile which can discriminate between EEC and SC due to a degree of overlap. However, in daily practice, p53, p16, *IMP2*, and *IMP3* are helpful tools.

Molecular alterations involved in the development of EEC are different from those of SC [27, 28]. EEC shows microsatellite instability, as well as mutations in *PTEN*, *KRAS*, and *CTNNB1* genes whereas SC exhibits alterations of *TP53*, widespread loss of heterozygosity, as reflected by chromosomal instability as well as other molecular alterations (*STK15*, p16, E-cadherin, and C-erbB2). *PIK3CA* and *ARID1A* mutations have been identified in both tumor types. Recently [9], TCGA study confirmed a different mutational profile between EEC and SC. The most frequently mutated genes in EEC were *PTEN* (77%), *PIK3CA* (53%), *PIK3R1* (37%), *CTNNB1* (36%), *ARID1A* (35%), *KRAS* (24%), *CTCF* (20%), *RPL22* (12%), *TP53* (11%), *FGFR2* (11%), and *ARID5B* (11%). The most frequently mutated genes in SC were *TP53* (90.7%), *PIK3CA* (41.9%), *FBXW7* (30.2%), *PPP2R1A* (36.6%), *CHD4* (16.3%), *CSMD3* (11.6%), and *COL11A1* (11.6%) [9].

Several authors have used targeting sequencing in EC. Chang et al. [29] assessed ten cases of EC using the Nimble Gen Comprehensive Cancer Panel that includes 578 cancer-related genes. This survey revealed 120 variants in 99 genes, 21 of which were included in the OncoPrint Cancer Research Panel. Mass spectrometric-based mutation analysis was also used in several studies [30]. Lately, a panel of cancer-related genes was used to identify increased mutational load in EC with deficient DNA mismatch repair [31]. McConechy et al. [32] performed target enrichment sequencing on 393 EC, by sequencing nine genes (*ARID1A*, *PPP2R1A*, *PTEN*, *PIK3CA*, *KRAS*, *CTNNB1*, *TP53*, *BRAF*, and *PPP2R5C*) and suggest that this nine-gene panel may be useful as an adjunct to the morphological classification of EC. The authors found that majority of microscopical diagnoses were concordant with the identified genotype. However they noticed a subset of genotype-incompatible diagnoses [33].

In the present study, we used targeting sequencing as a tool for assessing histological typing. We decide to use a customized gene panel that included most frequently mutated genes in EEC and SC, according to TCGA. This unique panel has never been tested before in the literature for histological typing in EC. In this proof of concept study, we selected 24 well-characterized cases at microscopical level which included the four TCGA molecular groups. Three cases were especially interesting; case 9 was a MSI with p53 “wild type” IHC pattern but with a *TP53* detected mutation; case 11 was a *POLE*-mutated tumor that showed a p53 “mutated” IHC pattern; and case 21 was a SC that had a “wild type” p53 IHC pattern but otherwise typical features of SC. Remaining cases were more or less classical EEC and SC.

The targeted sequencing analysis produced a mean of 2,270,521 reads. We identified a total of 164 somatic variants (102 missense, 39 frameshift, 15 stop gained, 1 splice acceptor, 1 protein altering, and 6 inframe deletions), and 72% of them were either pathogenic or likely pathogenic variants. The mutational profile of the 24 tumors was concordant with the pathologic diagnosis in all cases. Most mutated genes in the 16 EEC cases were *PTEN* (93.7%), *ARID1A* (68.7%), *PIK3CA* (50%), and *KMT2B* (43.7%) whereas the 8 SC tumors showed mutations in *TP53* (87.5%), *PIK3CA* (50%), and *PPP2R1A* (25%). Our customized gene panel allowed correct classification of all tumors in the two categories. Coexistence of mutations in *PTEN*, *ARID1A*, and *KMT2B* genes was diagnostic of EEC. On the other hand, absence of *PTEN*, *ARID1A*, and *KMT2B* in the presence of *TP53* mutation was diagnostic of SC. *TP53* and *PIK3CA* mutations were present in both tumor types.

Targeted sequencing results were validated by the Sanger sequencing and immunohistochemistry for *TP53*, *PTEN*, and *ARID1A* genes, showing a nice correlation between validation techniques. In general, *TP53* gene

mutations were associated with a “mutated” p53 IHC pattern (only one *TP53* mutated tumor showed “wild type” p53 pattern by IHC). Cases with *ARID1A* and *PTEN* gene mutations exhibited concordance with negative immunostaining in 83% and 86%, respectively. Mismatch repair deficiency, by IHC, was restricted to EEC cases. Interestingly, there was a nice correlation between the presence of mismatch repair deficiency, as assessed by IHC for MSH6 and PMS2, and the number of mutations, although it is clear that our approach is not the correct tool to assess tumor mutation load. A *POLE*-mutated, EEC grade 3 showed negative expression of MSH6, “mutated” IHC pattern for p53, negative PTEN staining, and the highest number of mutations, including one in *TP53* and one in *PTEN* (case 11). A high-grade EC double mutated (case 6) was correctly classified as EEC 3 by targeted sequencing (Fig. 4). The series included a SC that had a “wild type” IHC p53 staining

pattern, but otherwise typical microscopical features of SC (case 21). Targeted sequencing correctly classified the tumors as SC, and confirmed the lack of *TP53* mutations. This case illustrates the fact that there is a subset of SC (9% according to TCGA) that does not show either *TP53* mutations or “mutated” p53 IHC pattern. Interestingly, this case would be misclassified as EC with no specific molecular profile, based on the “wild type” IHC pattern for p53 by the surrogate TCGA approach, while targeted sequencing confirms that the correct diagnosis is SC. Exome sequencing of this case (data not shown) confirmed a high number of copy-number alterations and, thus, the tumor was a serous-like EC. There was an additional tumor (case 9) that was erroneously classified by the TCGA-based surrogate in the group of microsatellite unstable tumors, because of negative mismatch repair and “wild type” p53 IHC pattern. It showed a *TP53* mutation, either by targeted and Sanger

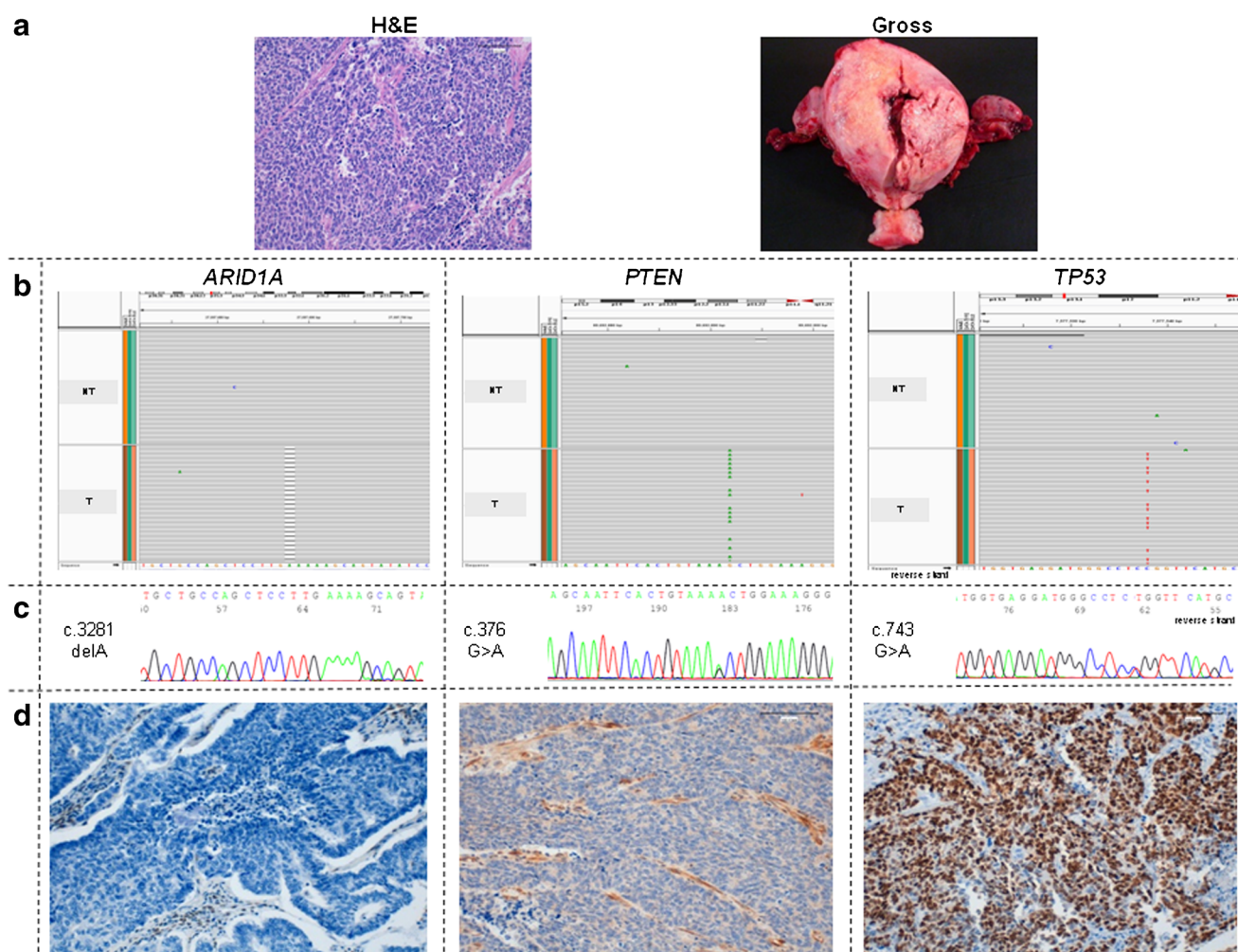


Fig. 4 Photomic depiction to illustrate integration of anatomic and molecular pathology of a FIGO EEC grade 3 (case 6). Microscopic ($\times 20$) and gross appearance (a); *ARID1A* (c.3281delA, p.K1094SfsTer67), *PTEN* (c.376G > A, p.A126T), and *TP53* (c.743 G > A, R248Q)

mutations detected by targeted sequencing and visualized through IGV (Integrative Genomic Viewer) tool (b) and detected by the Sanger sequencing (c); *ARID1A*, *PTEN*, and *TP53* immunohistochemistry ($\times 20$) (d)

sequencing demonstrating that indeed was a really double mutated tumor. Overall, targeted sequencing analysis showed that TCGA-based surrogate misclassified 2 of the 24 tumors in this series of cases.

In summary, this proof of concept study demonstrates the suitability of targeted sequencing with a customized endometrial cancer gene panel as an additional tool for confirming histological typing. The strategy seems interesting as a tool to classify tumors with unusual microscopical findings, as well as tumors with unexpected IHC features, such as SC without *TP53* alterations, as illustrated in one case of this series. Nevertheless, further studies with a larger series containing consecutive cases are required to validate this approach in equivocal cases with combined or mixed features between EEC and SC.

Author's contribution Dolores Cuevas: collected samples, collected clinical patient data, performed experiments, analyzed data, participated in study design, and wrote, edited, and reviewed the manuscript.

Joan Valls: performed statistical analysis and wrote part of the manuscript.

Sònia Gatius: collected clinical and histological patient data, assisted in histopathological examinations, and wrote part of the manuscript.

Berta Roman-Canal: reviewed the manuscript.

Elena Estaran: assisted in histopathological examinations.

Eduard Dorca: assisted in histopathological examinations.

Maria Santacana: performed immunohistochemical experiments and wrote part of the manuscript.

Marta Vaquero: assisted in carrying out experiments and reviewed the manuscript.

Ana Velasco: participated in study design, assisted in carrying out experiments, and reviewed the manuscript.

Núria Eritja: assisted in data collection and reviewed the manuscript.

Xavier Matias-Guiu: conceived and designed the study and wrote, edited, and reviewed the manuscript.

Funding The research team was supported by grants from the ISCIII (PI16/00692, CB16/12/00231) and Generalitat de Catalunya 2014SGR138. Tumor samples were obtained with the support of Xarxa Catalana de Bancs de Tumors and Plataforma de Biobancos ISCIII (PT 17/0015/0027).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Prat J (2004) Prognostic parameters of endometrial carcinoma. *Hum Pathol* 35:649–662
- Matias-Guiu X, Davidson B (2014) Prognostic biomarkers in endometrial and ovarian carcinoma. *Virchows Arch* 464:315–331
- Salvesen HB, Haldorsen IS, Trovik J (2012) Markers for individualised therapy in endometrial carcinoma. *Lancet Oncol* 13:e353–e361
- Clarke BA, Gilks CB (2010) Endometrial carcinoma: controversies in histopathological assessment of grade and tumor cell type. *J Clin Pathol* 63:410–415
- Bartosch C, Manuel LJ, Oliva E (2011) Endometrial carcinomas: a review emphasizing overlapping and distinctive morphological and immunohistochemical features. *Adv Anat Pathol* 18:415–437
- Soslow RA (2013) High-grade endometrial carcinomas - strategies for typing. *Histopathology* 62:89–110
- Garg K, Soslow RA (2012) Strategies for distinguishing low-grade endometrioid and serous carcinomas of endometrium. *Adv Anat Pathol* 19:1–10
- Gilks CB, Oliva E, Soslow RA (2013) Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma. *Am J Surg Pathol* 37:874–881
- Cancer Genome Atlas Research Network (2013) Integrated genomic characterization of endometrial carcinoma. *Nature* 497:67–73
- Bosse T, Nout RA, McAlpine JN, McConechy MK, Britton H, Hussein YR, Gonzalez C, Ganesan R, Steele JC, Harrison BT, Oliva E, Vidal A, Matias-Guiu X, Abu-Rustum NR, Levine DA, Gilks CB, Soslow RA (2018) Molecular classification of grade 3 endometrioid endometrial cancers identifies distinct prognostic subgroups. *Am J Surg Pathol* 42:561–568
- Hamilton CA, Cheung MK, Osann K, Chen L, Teng NN, Longacre TA, Powell MA, Hendrickson MR, Kapp DS, Chan JK (2006) Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers. *Br J Cancer* 94:642–646
- Talhok A, McConechy MK, Leung S, Li-Chang HH, Kwon JS, Melnyk N, Yang W, Senz J, Boyd N, Kamezis AN, Huntsman DG, Gilks CB, McAlpine JN (2015) A clinically applicable molecular-based classification for endometrial cancers. *Br J Cancer* 113:299–310
- Stelloo E, Nout RA, Osse EM, Jürgenliemk-Schulz JJ, Jobsen JJ, Lutgens LC, van der Steen-Banasik EM, Nijman HW, Putter H, Bosse T, Creutzberg CL, Smit VT (2016) Improved risk assessment by integrating molecular and clinicopathological factors in early-stage endometrial cancer-combined analysis of the PORTEC cohorts. *Clin Cancer Res* 22:4215–4224
- Piulats JM, Guerra E, Gil-Martín M, Roman-Canal B, Gatius S, Sanz-Pamplona R, Velasco A, Vidal A, Matias-Guiu X (2017) Molecular approaches for classifying endometrial carcinoma. *Gynecol Oncol* 145:200–207
- Moreno-Bueno G, Sanchez-Estevéz C, Cassia R, Rodríguez-Perales S, Diaz-Uriarte R, Dominguez O, Hardisson D, Andujar M, Prat J, Matias-Guiu X, Cigudosa JC, Palacios J (2003) Differential gene expression profile in endometrioid and nonendometrioid endometrial carcinoma: STK15 is frequently overexpressed and amplified in nonendometrioid carcinomas. *Cancer Res* 63:5697–5702
- Risinger JJ, Maxwell GL, Chandramouli GV, Jazaeri A, Aprelikova O, Patterson T, Berchuck A, Barrett JC (2003) Microarray analysis reveals distinct gene expression profiles among different histologic types of endometrial cancer. *Cancer Res* 63:6–11
- Alvarez T, Miller E, Duska L, Oliva E (2012) Molecular profile of grade 3 endometrioid endometrial carcinoma: is it a type I or type II endometrial carcinoma? *Am J Surg Pathol* 36:753–761
- Wik E, Raeder MB, Krakstad C, Trovik J, Birkeland E, Hoivik EA, Mjos S, Werner HM, Mannelqvist M, Stefansson IM, Oyan AM, Kalland KH, Akslen LA, Salvesen HB (2013) Lack of estrogen receptor-alpha is associated with epithelial-mesenchymal transition and PI3K alterations in endometrial carcinoma. *Clin Cancer Res* 19:1094–1105

19. Morrison C, Zanagnolo V, Ramirez N, Cohn DE, Kelbick N, Copeland L, Maxwell GL, Fowler JM (2006) HER-2 is an independent prognostic factor in endometrial cancer: association with outcome in a large cohort of surgically staged patients. *J Clin Oncol* 24:2376–2385
20. Chen N, Yi X, Abushahin N, Pang S, Zhang D, Kong B, Zheng W (2010) Nrf2 expression in endometrial serous carcinomas and its precancers. *Int J Clin Exp Pathol* 4:85–96
21. Alkushi A, Lim P, Coldman A, Huntsman D, Miller D, Gilks CB (2004) Interpretation of p53 immunoreactivity in endometrial carcinoma: establishing a clinically relevant cut-off level. *Int J Gynecol Pathol* 23:129–137
22. Yemelyanova A, Ji H, Shih I, Wang TL, Wu LS, Ronnett BM (2009) Utility of p16 expression for distinction of uterine serous carcinomas from endometrial endometrioid and endocervical adenocarcinomas: immunohistochemical analysis of 201 cases. *Am J Surg Pathol* 33:1504–1514
23. McCluggage WG, Connolly LE, McBride HA, Kalloger S, Gilks CB (2012) HMGA2 is commonly expressed in uterine serous carcinomas and is a useful adjunct to diagnosis. *Histopathology* 60:547–553
24. Alkushi A, Clarke BA, Akbari M, Makretsov N, Lim P, Miller D, Magliocco A, Coldman A, van de Rijn M, Huntsman D (2007) Identification of prognostically relevant and reproducible subsets of endometrial adenocarcinoma based on clustering analysis of immunostaining data. *Mod Pathol* 20:1156–1165
25. Reid-Nicholson M, Iyengar P, Hummer AJ, Linkov I, Asher M, Soslow RA (2006) Immunophenotypic diversity of endometrial adenocarcinomas: implications for differential diagnosis. *Mod Pathol* 19:1091–1100
26. Santacana M, Maiques O, Valls J, Gatus S, Abó AI, López-García MÁ, Mota A, Reventós J, Moreno-Bueno G, Palacios J, Bartosch C, Dolcet X, Matias-Guiu X (2014) A 9-protein biomarker molecular signature for predicting histologic type in endometrial carcinoma by immunohistochemistry. *Hum Pathol* 45:2394–2403
27. Yeramian A, Moreno-Bueno G, Dolcet X, Catusus L, Abal M, Colas E, Reventos J, Palacios J, Prat J, Matias-Guiu X (2013) Endometrial carcinoma: molecular alterations involved in tumors development and progression. *Oncogene* 32:403–413
28. Llobet D, Pallares J, Yeramian A, Santacana M, Eritja N, Velasco A, Dolcet X, Matias-Guiu X (2009) Molecular pathology of endometrial carcinoma: practical aspects from the diagnostic and therapeutic viewpoints. *J Clin Pathol* 62:777–785
29. Chang YS, Huang HD, Yeh KT, Chang JG (2016) Genetic alterations in endometrial cancer by targeted next-generation sequencing. *Exp Mol Pathol* 100:8–12
30. Krakstad C, Birkeland E, Seidel D, Kusonmano K, Petersen K, Mjøs S, Hoivik EA, Wik E, Halle MK, Øyan AM, Kalland KH, Werner HM, Trovik J, Salvesen H (2012) High-throughput mutation profiling of primary and metastatic endometrial cancers identifies KRAS, FGFR2 and PIK3CA to be frequently mutated. *PLoS One* 7:e52795
31. Lee PJ, McNulty S, Duncavage EJ, Heusel JW, Hagemann IS (2018) Clinical targeted next-generation sequencing shows increased mutational load in endometrioid-type endometrial adenocarcinoma with deficient DNA mismatch repair. *Int J Gynecol Pathol* 37:581–589
32. McConechy MK, Ding J, Cheang MC, Wiegand KC, Senz J, Tone AA, Yang W, Prentice LM, Tse K, Zeng T (2012) Use of mutation profiles to refine the classification of endometrial carcinomas. *J Pathol* 228:20–30
33. Hoang LN, McConechy MK, Köbel M, Han G, Rouzbahman M, Davidson B, Irving J, Ali RH, Leung S, McAlpine JN, Oliva E, Nucci MR, Soslow RA, Huntsman DG, Gilks CB, Lee CH (2013) Histotype-genotype correlation in 36 high-grade endometrial carcinomas. *Am J Surg Pathol* 37:1421–1423

Intratumour heterogeneity in endometrial serous carcinoma assessed by targeted sequencing and multiplex ligation-dependent probe amplification: a descriptive study

Dolors Cuevas,¹  Ana Velasco,¹ Marta Vaquero,¹ Maria Santacana,¹ Sònia Gatius,¹ Núria Eritja,¹ Elena Estaran¹ & Xavier Matias-Guiu^{1,2}

¹Hospital Universitari Arnau de Vilanova, Universitat de Lleida, IRBLleida, Ciberonc, Lleida, and ²Hospital Universitari de Bellvitge, Idibell, Barcelona, Spain

Date of submission 12 July 2019

Accepted for publication 20 September 2019

Published online Article Accepted 24 September 2019

Cuevas D, Velasco A, Vaquero M, Santacana M, Gatius S, Eritja N, Estaran E & Matias-Guiu X

(2020) *Histopathology* 76, 447–460. <https://doi.org/10.1111/his.14001>

Intratumour heterogeneity in endometrial serous carcinoma assessed by targeted sequencing and multiplex ligation-dependent probe amplification: a descriptive study

Aims: Endometrial serous carcinoma (ESC) represents the most aggressive subtype of endometrial carcinoma (EC). According to The Cancer Genome Atlas (TCGA), ESC exhibits a genomic profile characterised by frequent *TP53* mutations and somatic copy-number alterations (SCNA). Several studies have suggested the role of intratumour heterogeneity (ITH) in tumour progression and therapy resistance, highlighting ITH as a challenge for personalised medicine. ITH is described as the co-existence of clonal and subclonal cellular populations within a single tumour. To date, the extent and prevalence of ITH in ESC have not been fully evaluated. The aim of this study was to address ITH analysis in ESC. We performed a descriptive integrated molecular approach using targeted sequencing and multiplex ligation-dependent probe amplification (MLPA) to identify mutations and SCNA patterns, respectively.

Methods and results: Eight ESC were examined, selecting three tumour regions per case and their corresponding normal tissue. For targeted sequencing a gene panel of 40 genes based on TCGA and other survey data was performed. For MLPA different probe mixes were used to detect SCNA in 106 genes. Analysis of mutations and SCNA were performed in each sample and comparative analysis of the three tumour regions was also conducted. Targeted sequencing showed that mutations in *TP53*, *PIK3CA* and *PPP2R1A* were ubiquitous in all tumour regions. Moreover, MLPA results demonstrated a high frequency of SCNA, according to the already known presence of genomic instability in ESC. Unlike the homogeneous distribution of somatic mutations, SCNA exhibited ITH affecting targetable genes such as *ERBB2*.

Conclusions: Our study suggests that somatic gene copy-number alterations are the main source of ITH in ESC.

Keywords: chromosomal instability, endometrial neoplasms, gene dosage, intratumour heterogeneity, mutation, precision medicine

Address for correspondence: X Matias-Guiu, Hospital Universitari Arnau de Vilanova, Avenida Rovira Roure, 80, 25198 Lleida, Spain. e-mail: fjmatiasguiu.lleida.ics@gencat.cat

© 2019 John Wiley & Sons Ltd.

Introduction

Endometrial serous carcinoma (ESC) represents approximately 10% of all endometrial carcinoma (EC); however, it constitutes the most clinically aggressive subtype, accounting for a disproportionate number of deaths among women affected. This

aggressive behaviour is due to the tumour often being widely disseminated at presentation and highly resistant to conventional chemotherapy.^{1–3} The Cancer Genome Atlas (TCGA), using multiple sequencing-based technologies, revealed that ESC exhibited a unique genomic profile among endometrial cancers, characterised by frequent *TP53* mutations and extensive somatic copy-number alterations (SCNA).⁴ It is well known that genomic instability and mutations are enabling characteristics that allow the acquisition of cancer hallmark capabilities.⁵ In a recent study, a shared signature in breast, lung, ovarian and ESC was observed. This signature was characterised by *TP53* mutations, high numbers of genomic breakpoints and SCNA of genes from the core glycolysis pathway, which is an emerging hallmark of human tumours.⁶

Recently, intratumour heterogeneity (ITH) has been proposed as another key factor for tumour progression and patient survival. It is also reported that ITH is a feature of the vast majority of solid tumours. Therefore, ITH is currently considered a great challenge to personalised therapies and cancer management.^{7–9} ITH is believed to result from a process of branching evolution, and it is understood to be the coexistence of cancer cell subpopulations that differ in their molecular alterations which gives them specific genetic and phenotypical features. Thus, spatially separated areas within tumours may exhibit diverse molecular features.^{10–12}

To investigate ESC intratumour heterogeneity we examined the contribution of molecular alterations such as mutations and SCNA on 24 samples from eight cases of ESC. In all samples we performed targeted sequencing and multiplex ligation-dependent probe amplification (MLPA) analysis to assess the mutational profile and the SCNA patterns, respectively. For targeted sequencing we used a specific customised gene panel that included the 40 most frequently mutated genes in EC, according to TCGA⁴ and other survey data.^{13–15} We have previously demonstrated the suitability of this targeted sequencing analysis as a useful tool for confirming histological typing in EC.¹⁶ For SCNA analysis we used the MLPA assay to detect copy-number changes in 106 specific genes that have been shown to be implicated in the development, progression and response to therapy in many types of tumour. MLPA is a high-throughput method, which is easier and faster than the widely used comparative genomic hybridisation technology. It is based on hybridisation of multiple probes to a sample DNA and further amplification using only one polymerase chain reaction (PCR)

primer pair, enabling simultaneous amplification of different loci in a single reaction. This semiquantitative technique determines the relative gene copy-number between tumour and non-tumour regions from the same specimen.^{17–19}

Previous molecular studies have focused on the genomic features of ESC^{4,13,15} and on ITH analysis of different tumour types^{7–9}; however, the extent and prevalence of intratumour heterogeneity in this highly aggressive carcinoma have not been fully evaluated. Thus, the purpose of this study was to elucidate the ITH features of ESC by using an integrative molecular approach examining somatic mutations and gene somatic copy-number alterations, emphasising prospective targetable gene pathways.

Materials and methods

SAMPLES AND DNA EXTRACTION

The present study was performed on 32 frozen tissue samples from eight ESCs. For each case, three tumour regions and one matched normal tissue were sampled and immediately frozen and stored at -80°C at the frozen tumour tissue bank of IRBLleida as part of a prospective study to assess differences in distinct areas of the primary tumour. Metastatic tissue was not included. All tumours were diagnosed at the Department of Pathology of Hospital Arnau de Vilanova of Lleida, according to the most recent World Health Organisation (WHO) criteria.²⁰ They were surgically staged and graded according to the International Federation of Gynaecology and Obstetrics (FIGO) staging and grading systems.²¹ The study was approved by the local Ethical Committee and specific informed consent was used. Tumour and normal tissue quality were assessed in haematoxylin and eosin (H&E) sections. The percentage of tumour, tumour necrosis and non-tumour tissue was quantified (volume in two dimensions) in each tissue block. Table 1 shows patient characteristics and pathological features of the tumours. The mean follow-up period was 54 months (range = 21–112). Matched formalin-fixed, paraffin-embedded tissue blocks were used for immunohistochemical (IHC) and silver *in-situ* hybridisation (SISH) validation. DNA was purified using DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions. For all samples, DNA purity was assessed using a Nanodrop ND-1000 (Thermo Fisher Scientific, Fremont, CA, USA). Additionally, for targeted sequencing, DNA samples were further quantified using the Qubit dsDNA HS Assay Kit (Invitrogen/Thermo Fisher Scientific) performed

Table 1. Clinical and pathological features

Case	Age (years)	Histological type	Grade	Stage	Follow-up	Follow-up time (months)
1	86	SC	III	IA	DOD	84
2	89	SC	III	IIIA	DOD	48
3	85	SC	III	IB	AWOD	108
4	76	SC	III	IA	DOD	48
5	81	SC	III	IA	DOD	36
6	87	SC	III	IIIB	DOD	48
7	93	SC	III	IB	DOD	36
8	96	SC	III	II	DOD	36

SC, Serous carcinoma; DOD, Dead of disease; AWOD, Alive with other disease.

on the Qubit Fluorometer 1.0 (Invitrogen/Life Technologies, Carlsbad, CA, USA).

MLPA

MLPA procedure

MLPA analysis was performed according to the manufacturer's instructions using four different kits from MRC-Holland (Amsterdam, the Netherlands). A list of 106 chromosomal loci (Table 2) was analysed using the following SALSA MLPA kits: P171-A2 gain-1, P172-B1 gain-2, P173-A2 gain-3 and P294-A1 tumour-loss. Genes included in these probe mixes are known human genes that have often been detected with an increased copy-number in one or more types of tumour. These kits include multiple MLPA probes for different chromosomal loci plus quality control probes to check for adequate quality of test DNA and efficient ligation. Each MLPA probe consists of two oligonucleotides that must bind to adjacent target DNA sequences to be ligated. Upon ligation, a single molecule is formed and can be amplified during PCR. Following PCR, amplified probes are separated by capillary electrophoresis on a 3130 Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA) and analysed using GeneMapper® software version 4.1 (Applied Biosystems).

MLPA data analysis

MLPA assay determines relative gene copy-number changes in a semiquantitative manner, comparing MLPA peak patterns of tumour and non-tumour samples. Three non-tumour samples were used as reference samples in each MLPA assay. Fragment analysis data were received in GeneMapper® software, and peak areas, as a measure of peak intensity for each

probe, were imported in an adapted version of an Excel spreadsheet designed by Genycell Biotech España SL (GBE, Granada, Spain). First, intrasample normalisation was performed by dividing the peak area of each probe by the peak area of every other probe in that sample, thus creating as many ratios per probe as there are other probes. Subsequently, the median of all these produced ratios per probe was taken by acquiring the probe's normalisation constant for each probe and sample. Secondly, intersample comparison was performed by dividing the normalisation constant of each probe in a given sample by the average normalisation constant of that probe in the reference samples, giving a final ratio for a probe that is called the dosage quotient (DQ). Finally, somatic copy-number status of a specific chromosomal locus was categorised, as suggested by GBE, as loss (deletion) if DQ was <0.7 ; normal (diploid) if DQ was between 0.7 and 1.4; and gain (duplication or amplification) if DQ was >1.4 . Thus, altered loci were those with a somatic copy-number status of loss or gain.

MLPA intratumour heterogeneity criteria

To determine intratumour gene heterogeneity, we calculated three parameters based on the DQ values of the three tumour regions per locus and case: (i) ADQ: DQ average value, (ii) SDDQ: DQ standard deviation value and (iii) DDQ: difference between ADQ+SDDQ and ADQ-SDDQ. Analysis of these parameters for diploid locus allowed us to determine its statistical variation. Therefore, a locus was considered significantly heterogeneous if all the following conditions were accomplished: (i) the SDDQ value was >0.15 ; (ii) the DDQ was >0.3 ; and (iii) the DQ values of the three tumour regions belong to different DQ

Table 2. List of the 106 chromosomal loci included in the multiplex ligation-dependent probe amplification (MLPA) kits and the 40 genes included in the customised panel for targeted sequencing

MLPA				Targeted sequencing	
Gene	CHR.LOC.	Gene	CHR.LOC.	Gene	CHR
<i>NRAS</i>	1p13.2	<i>WT1</i>	11p13	<i>CSDE1</i>	chr1
<i>BCAS2</i>	1p13.3	<i>CCND1</i>	11q13	<i>ARID1A</i>	chr1
<i>BCAR3</i>	1p22.1	<i>EMS</i>	11q13	<i>COL11A1</i>	chr1
<i>PSMB4</i>	1q21.3	<i>FGF3</i>	11q13	<i>NRAS</i>	chr1
<i>MDM4</i>	1q32.1	<i>GSTP1</i>	11q13	<i>FBXW7</i>	chr1
<i>MYCL1</i>	1p34.2	<i>RELA</i>	11q13	<i>ATR</i>	chr3
<i>KIF1B</i>	1p36.23	<i>BRMS1</i>	11q13.2	<i>CTNNB1</i>	chr3
<i>CAMTA1</i>	1p36.31	<i>FGF4</i>	11q13.3	<i>PIK3CA</i>	chr3
<i>CHD5</i>	1p36.31	<i>BIRC2</i>	11q22	<i>SCARNA11</i>	chr4
<i>PRDM16</i>	1p36.32	<i>BIRC3</i>	11q22	<i>PIK3R1</i>	chr5
<i>DVL1</i>	1p36.33	<i>MMP7</i>	11q22.3	<i>MAP3K4</i>	chr6
<i>NTRK1</i>	1q23.1	<i>BCL2L14</i>	12p13.2	<i>RPL22</i>	chr8
<i>CENPF</i>	1q32.3	<i>CCND2</i>	12p13.3	<i>MKI67</i>	chr10
<i>PIK3C2B</i>	1q32.1	<i>CDK4</i>	12q14	<i>ARID5B</i>	chr10
<i>MYCN</i>	2p24.1	<i>MDM2</i>	12q14.3	<i>PTEN</i>	chr10
<i>BCL2L11</i>	2q13	<i>CYP27B1</i>	12q14.1	<i>FGFR2</i>	chr10
<i>ERBB4</i>	2q33	<i>CCNA1</i>	13q12.3	<i>SIN3A</i>	chr10
<i>IGFBP2</i>	2q35	<i>BRCA2</i>	13q12.3	<i>TSPYL2</i>	chr10
<i>IGFBP5</i>	2q35	<i>RB1</i>	13q14.2	<i>CCND1</i>	chr11
<i>FHIT</i>	3p14.2	<i>DLEU1</i>	13q14.3	<i>ABCC9</i>	chr12
<i>VHL</i>	3p25.3	<i>MIRN15A</i>	13q14.3	<i>CHD4</i>	chr12
<i>PIK3CA</i>	3q26.32	<i>IRS2</i>	13q34	<i>KRAS</i>	chr12
<i>EVI1</i>	3q26	<i>AKT1</i>	14q32.33	<i>USP36</i>	chr12
<i>BCL6</i>	3q27	<i>BCL2A1</i>	15q24.3	<i>FOXA2</i>	chr15
<i>PDGFRA</i>	4q11	<i>NTRK3</i>	15q25.3	<i>CSMD3</i>	chr16
<i>TERT</i>	5p25	<i>IGF1R</i>	15q26.3	<i>TP53</i>	chr17
<i>BIRC1</i>	5q13.2	<i>TSC2</i>	16p13.3	<i>WRAP53</i>	chr17
<i>APC</i>	5q22	<i>TOM1L2</i>	17p11.2	<i>ERBB2</i>	chr17
<i>PDGFRB</i>	5q32	<i>TP53</i>	17p13.1	<i>KLLN</i>	chr17
<i>HMGA1</i>	6p21.3	<i>NF1</i>	17q11.2	<i>PRPF18</i>	chr17
<i>CCND3</i>	6p21.1	<i>ERBB2</i>	17q21.1	<i>SPOP</i>	chr17
<i>NFKBIE</i>	6p21.1	<i>IGFBP4</i>	17q21.2	<i>CDH19</i>	chr18

Table 2. (Continued)

MLPA				Targeted sequencing	
Gene	CHR.LOC.	Gene	CHR.LOC.	Gene	CHR
<i>SERPINB9</i>	6p25.2	<i>BRCA1</i>	17q21.31	<i>PPP2R1A</i>	chr19
<i>ESR1</i>	6q25.1	<i>PPM1D</i>	17q23.2	<i>KMT2B</i>	chr19
<i>EGFR</i>	7p12	<i>BIRC5</i>	17q25.3	<i>SMARCA4</i>	chr19
<i>CDK6</i>	7q21.3	<i>BCL2</i>	18q21.33	<i>CTCF</i>	chr20
<i>MET</i>	7q31	<i>SMAD4</i>	18q21.1	<i>EP300</i>	chr22
<i>BRAF</i>	7q34	<i>SERPINB2</i>	18q21.33	<i>TAF1</i>	chrX
<i>MOS</i>	8q11	<i>SERPINB7</i>	18q21.33	<i>BCOR</i>	chrX
<i>RNF139</i>	8q24.13	<i>FKBP8</i>	19p13.11	<i>PGAP3</i>	chrX
<i>FGFR1</i>	8p11.2	<i>STK11</i>	19p13.3		
<i>MYBL1</i>	8q21	<i>CCNE1</i>	19q12		
<i>PTK2</i>	8q24	<i>BCL2L1</i>	20q11.1		
<i>MYC</i>	8q24.12	<i>MYBL2</i>	20q12		
<i>PTP4A3</i>	8q24.3	<i>PTPN1</i>	20q13.13		
<i>CDKN2A</i>	9p21	<i>BCAS1</i>	20q13.2		
<i>CDKN2B</i>	9p21	<i>AURKA</i>	20q13.31		
<i>JAK2</i>	9p24.1	<i>GNAS</i>	20q13.32		
<i>NTRK2</i>	9q21.33	<i>UCKL1</i>	20q13.33		
<i>PTCH</i>	9q22.3	<i>RUNX1</i>	21q22.12		
<i>TSC1</i>	9q34	<i>BCL2L13</i>	22q11		
<i>PTEN</i>	10q23.3	<i>SMARCB1</i>	22q11.23		
		<i>FAM123B</i>	Xq11.1		
		<i>BIRC4</i>	Xq25		

CHR.LOC, Chromosome localisation; CHR, Chromosome.

categories (loss, normal, gain). Finally, to consider a case with ITH more than three heterogeneous loci had to be found.

TARGETED SEQUENCING ANALYSIS

Targeted sequencing

Targeted sequencing analysis was performed to detect mutations in 40 genes (Table 2) that have been previously found to be recurrently mutated in endometrial carcinoma by TCGA⁴ and other surveys.^{13–15} Our custom gene panel design included 1333 amplicons [250 base pairs (bp) in length] that covered 98.59% of exons, introns and untranslated regions of

the selected genes. Custom amplicon library was created using 100 ng of genomic DNA from tumour and non-tumour specimens, according to the manufacturer's instructions [TruSeq Custom Amplicon (TSCA) library preparation guide]. In brief, a custom pool containing specific oligos to targeted regions of interest was hybridised to genomic DNA samples. After the removal of unbound oligos, an extension-ligation step was performed. The resulting products were amplified using primers that add multiplexing index sequences for sample tagging as well as common adapters required for cluster generation. Before pooling, normalisation was performed by quantifying individual libraries using the Qubit fluorometer, and then

pooled based on equal concentrations. Pooled TruSeq libraries were sequenced using Miseq Reagent kit version 3 for 600 cycles (Illumina, San Diego, CA, USA).

Bioinformatic analysis

Bioinformatic analysis was performed by Sistemas Genómicos S.L. (Valencia, Spain). The Sistemas Genómicos (SG) pipeline included fastq files processing, base calling, quality score assignment, read alignment against the human reference genome version GRCh38/hg38, quality control of mapping quality, coverage analysis, variant calling and variant annotation. After completion of the pipeline, a more accurate review of the variants list was performed. With the paired normal tumour analysis we selected as potential somatic mutations those variants with a variant allelic frequency (VAF) greater than 0.15 and a depth greater than 100 in the tumour tissue, not present in the normal tissue. Somatic gene mutations were further filtered based on the canonical isoforms and the variant effect on the transcript, excluding the non-canonical isoforms and synonymous variants. To restrict the analysis, mutations were also selected based on their frequency on population databases (NHLBI-ESP and 1000 Genomes Project) to discard known variants with a minor allele frequency (MAF) higher than 1%.

TARGETED SEQUENCING INTRATUMOUR HETEROGENEITY CRITERIA

To investigate the heterogeneously distribution of mutations, we determined the presence or absence of somatic mutations in each of the three tumour regions quantifying the percentage of mutations. Somatic mutations were assessed according to previously described methods. A case displayed intratumour heterogeneity if we detected the presence of different somatic mutations between the tumour regions.

SANGER SEQUENCING, IMMUNOHISTOCHEMISTRY AND SISH VALIDATION

A total of 10 variants identified by targeted sequencing analysis were validated by Sanger sequencing. These variants comprise hot-spots of the following genes: *TP53* and *PIK3CA*. Sequencing reactions were carried out using the BigDye Terminator Kit version 3.1 (Applied Biosystems/Thermo Fisher Scientific) and run on Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems). Samples exhibiting *ERBB2* intratumoral heterogeneity by MLPA were tested for *ERBB2* status

by IHC and SISH. Both IHC and SISH were conducted on 2- μ m thick sections from three different formalin-fixed paraffin-embedded tissue blocks from each ESC. IHC was conducted using the Hercep test (Dako, Carpinteria, CA, USA), according to the manufacturer's instructions. SISH was fulfilled using an automated dual-colour assay which stained *ERBB2* gene in black and *CEP17* in red (Roche, Basel, Switzerland). SISH protocol steps included deparaffinisation, citrate treatment, incubation with ISH protease, addition and hybridisation of *ERBB2* and *CEP17* probe, incubation with silver C, incubation with haematoxylin as contrast dye, and incubation with bluing after contrast dyeing. *ERBB2* status was scored according to guidelines set by the American Society of Clinical Oncology/College of American Pathologists (ASCO-CAP) 2018 for breast cancer.²² *ERBB2*-negative ESC were defined by an IHC score of 0 or 1+ and *ERBB2*-positive by an IHC score of 3+, or 2+, with gene amplification confirmed by SISH. Observers were blinded for results with other techniques.

Results

MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION

Somatic copy-number alterations as losses (deletions) and gains (duplications) were observed in all the cases ($n = 8$). Eighty-two of the 106 (77%) chromosomal loci analysed showed somatic copy-number changes, 67% of which were gains, 22% losses and the remaining 11% were genes harbouring losses and gains depending upon the sample (Figure 1). On average, 22 altered genes per sample were found, ranging from eight to 36. The most frequently gained genes were *UCKL1* (88%), *ERBB2*, *EVII* and *PIK3CA* (75%), *CCNE1*, *PTPN1* and *NTRK1* (63%) and *MYBL2*, *AURKA*, *BCL6* and *FKBP8* (50%). Chromosomal locations related to gained genes were 20q12–q13.3 for *UCKL1*, *PTPN1*, *MYBL2* and *AURKA*, 3q26–q27 for *EVII*, *PIK3CA* and *BCL6*, 17q21.1 for *ERBB2*, 19q12 for *CCNE1*, 19p13.11 for *FKBP8* and 1q23.1 for *NTRK1*. The genes most frequently lost were *BIRC4* (63%), *BIRC1* (50%), *FGFR1* (50%) and *ERBB4* (38%), which were located in Xq25, 5q13.2, 8p11.2 and 2q33, respectively (Figure 2). In particular, *TP53* gene was gained (25%) and lost (25%) depending on the sample. Among all these genes, it should be highlighted that *PIK3CA*, *CCNE1*, *BIRC1* and *FGFR1* were found homogeneously distributed within the tumour.

Based on the heterogeneity criteria, five (62.5%) of the eight cases exhibited significant ITH (more than

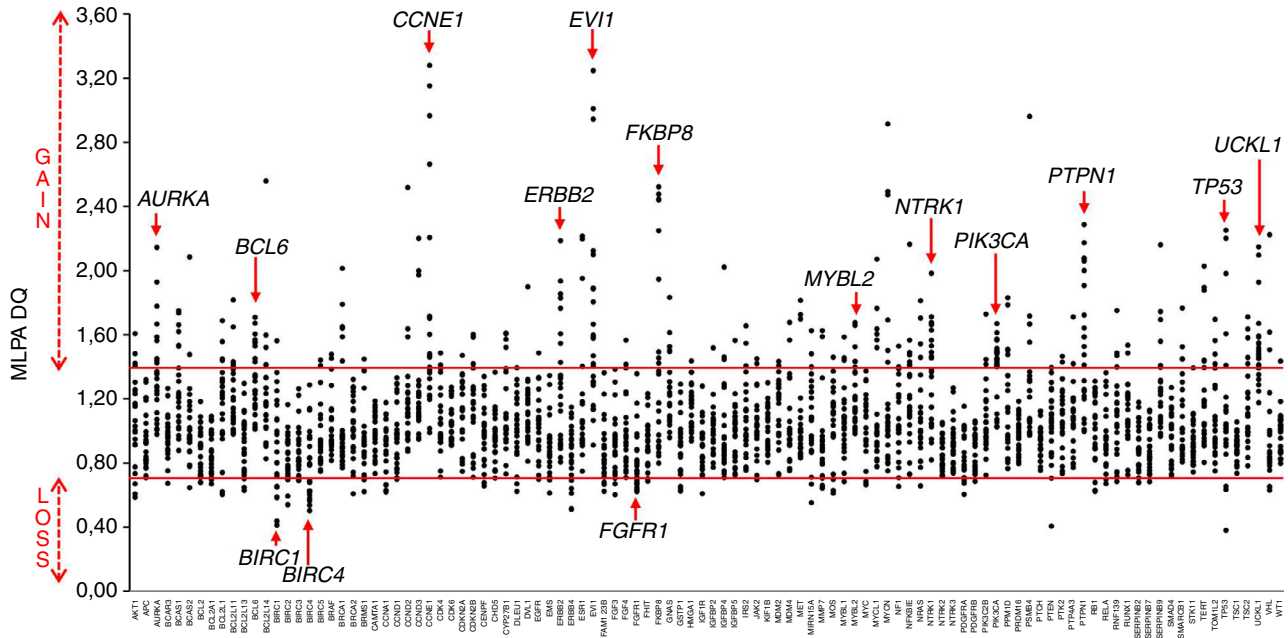


Figure 1. Somatic copy-number alteration profile of 24 endometrial serous carcinoma tumour regions for each chromosomal locus. The y -axis indicates the multiplex ligation-dependent probe amplification (MLPA) DQ values, where the normal range is between 0.7 and 1.4 (range defined by the red lines), and the x -axis shows all chromosomal loci ($n = 106$). Red arrows indicate the most frequently altered chromosomal loci (alteration frequency greater than 50%).

three heterogeneous loci). Figure 3 shows the example of a heterogeneous case (case 1). In these cases, 10 heterogeneous genes per case were detected on average (range = 8–15), of which 90% exhibited gains and the remaining 10% showed losses. ITH in chromosomal gains and losses between the three samples of each tumour was recorded, in at least one case, in 40 of the 82 altered chromosomal loci. We quantified the alteration frequency (AF) and the heterogeneity frequency (HF) for each locus. The AF was the quotient between the number of altered cases respect the total number of analyzed cases; and the HF was the quotient between the number of cases exhibiting heterogeneity respect the total of altered cases (Figure 2). The most frequently heterogeneous gained genes were, *MYBL2* (75%), *ERBB2* (67%), *AURKA* (50%), *PTPN1* (40%) and *EVI1* (33%) and lost genes were *ERBB4* (67%) and *BIRC4* (40%). The *TP53* gene, which was altered with gains and losses depending on the case, was also observed in heterogeneity with an HF of 50%.

TARGETED SEQUENCING ANALYSIS

Targeted sequencing analysis produced a mean of 2 825 883 reads, 98.38% of which mapped on reference genome version GRCh38/hg38 with high

quality without PCR duplicates. Mean coverage depth was of 1470X per sample, average depth and reads of normal and tumour samples being similar.

After applying the previously described filters, we identified somatic mutations in 11 genes with a sum of 23 somatic variants. These variants were classified as missense ($n = 15$), frameshift ($n = 6$), stop-gained ($n = 1$) and splice acceptor ($n = 1$). The average number of somatic mutations per sample was 3, ranging from 1 to 5. Among the 11 somatically mutated genes, *TP53* (88%), *PIK3CA* (50%) and *PPP2R1A* (25%) were the most frequently altered (Figure 2). All somatic variants affecting these genes were pathogenic variants, as described in the Catalogue of Somatic Mutations in Cancer (COSMIC) database, and were evaluated as deleterious with different pathogenicity predictors (Polyphen2, Sift, Condel and Fathmm). Furthermore, all these mutations were detected in the three tumour regions, thus these genes were not affected by ITH. In one case we also detected a pathogenic variant homogeneously distributed in the *ERBB2* gene. Mutations affecting *TP53* were stop-gained (12%) and missense (88%) affecting exons 4, 5, 6, 7, 8 and 10. The so-called hot-spots in *TP53* gene were R110, H179, R196, Y234, C242, V274, R342 and one splice acceptor variant in intron 4 (c.376-

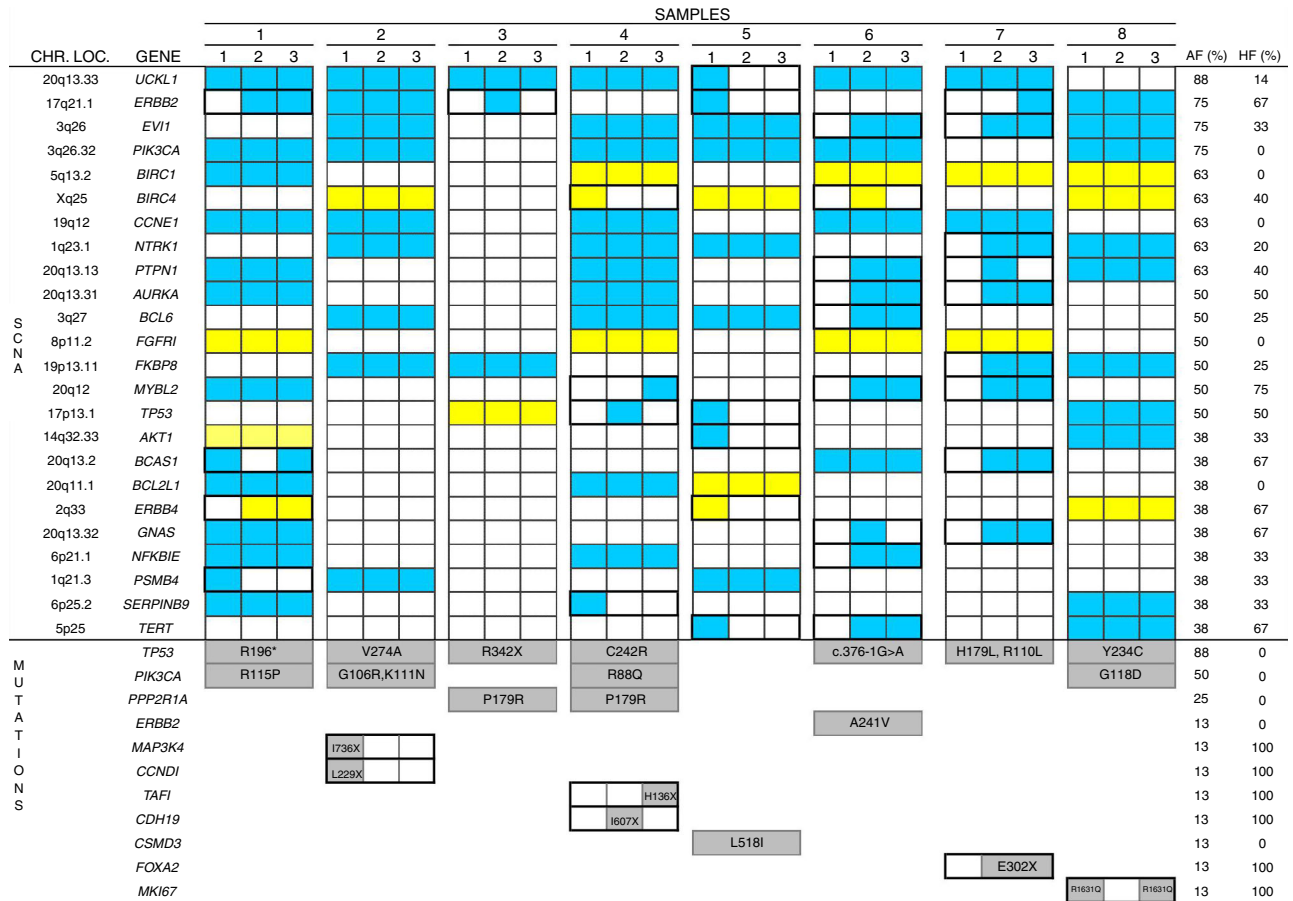


Figure 2. Distribution of the most frequently somatic copy-number alterations (SCNA) and mutations according to samples and chromosomal locus. All region samples are shown at the top of the x-axis; at the left of y-axis are shown chromosome locations and genes, and at the right the global alteration frequency and the heterogeneity frequency by chromosomal loci. In the SCNA panel, blue boxes indicate gains, yellow boxes losses, white boxes normal and black rectangles highlight the heterogeneous loci in each case, based on the heterogeneity criteria. The description of mutations per region sample are shown in the mutations panel; grey large boxes indicates that the same mutation has been observed in all region samples, and the black rectangles indicate the heterogeneous loci in each case where mutation has been observed only in any region sample.

1G>A). *PIK3CA* missense mutations were found in exon 2 in three cases and in the first codon of exon 3 in one case, and no mutations were found in exons 9 and 20. The P179R missense mutation found in *PPP2R1A* gene was a recurrent mutation found in previous studies. The remaining altered genes were *MAP3K4*, *CCND1*, *TAF1*, *CDH19*, *CSMD3*, *ERBB2*, *FOXA2* and *MKI67*, with frequencies of 13%, and their mutations were classified as variants of unknown significance (VUS), as they had not been previously reported either as a disease associated or as a normal variant, and their pathogenicity could not be established with certainty. Interestingly, all these mutations were heterogeneously distributed among the three tumour regions, thus ITH was observed.

SANGER SEQUENCING, IMMUNOHISTOCHEMISTRY AND SISH VALIDATION

Sanger sequencing analysis for *TP53* and *PIK3CA* genes validated the 10 variants identified by targeted sequencing. Using MLPA, *ERBB2* gene was altered in six of the eight analysed cases (75%), and heterogeneity, in this locus, was observed in four of the six altered cases (67%). Samples exhibiting *ERBB2* intratumoral heterogeneity by MLPA were tested for *ERBB2* status by IHC and SISH. Three of the four heterogeneous cases (75%) showed concordance between MLPA, IHC and SISH (Figures 4, 5 and 6). One case (case 5) was homogeneously negative by IHC and SISH, and by contrast was heterogeneous by MLPA (Figure 7). Thus, these techniques allowed

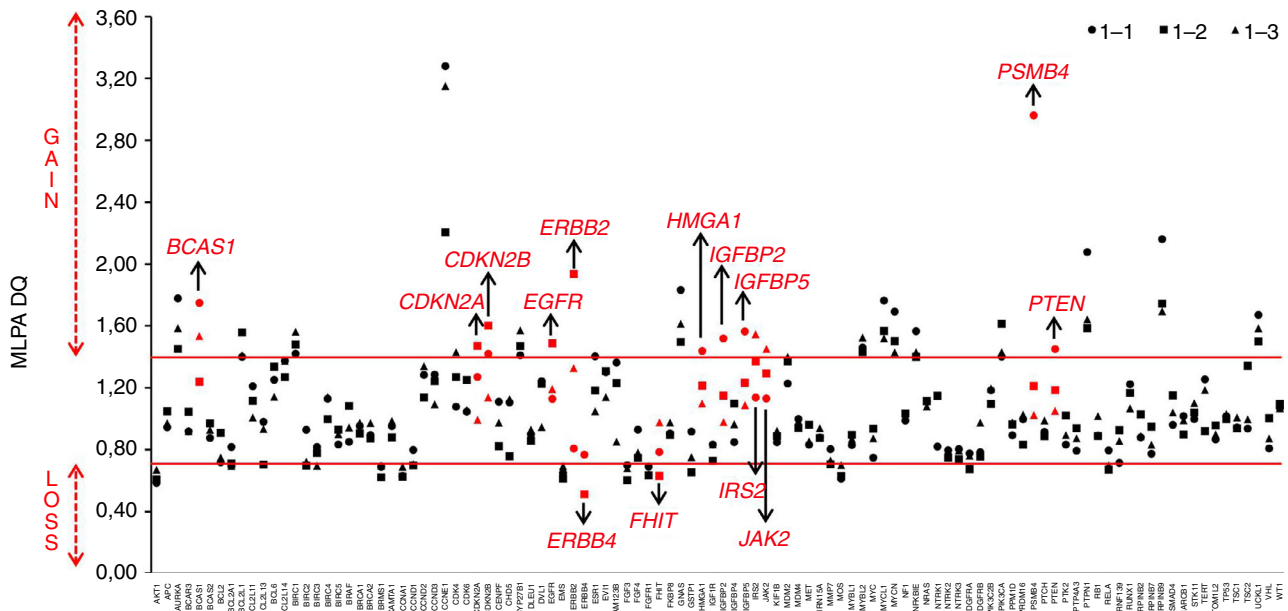


Figure 3. Example of a heterogeneous case (case 1). The *y*-axis shows the multiplex ligation-dependent probe amplification (MLPA) DQ values, where the normal range is between 0.7 and 1.4 (range defined by the red lines), and the *x*-axis shows all chromosomal loci ($n = 106$). Black arrows highlight the heterogeneous loci based on the heterogeneity criteria. [Colour figure can be viewed at wileyonlinelibrary.com]

demonstration of the *ERBB2* gene dosage intratumour heterogeneity.

Discussion

ESC is one of the most aggressive histological types of EC. It is important to provide scientific evidence to improve the criteria for diagnosis, prognosis and prediction of response to therapy. In the last decade, large efforts have been conducted to elucidate which driver molecular events are implicated in ESC tumorigenesis. These studies have revealed that the genetic landscape of ESC is characterised by a high proportion of somatic copy-number alterations, basically represented by somatic copy-number gains, and a high percentage of *TP53* mutations.^{4,13,15} Most SCNA analyses have been performed using single nucleotide polymorphism microarrays and Genomic Identification of Significant Targets in Cancer analysis to identify recurrent events, but this technology does not directly detect the specific oncogene and tumour suppressor gene (TSG) targets of driver SCNA.²³ Thus, to delineate the targeted oncogenes and TSG from the recurrent SCNA still remains a challenge and, for this reason, MLPA is a useful technique to identify SCNA in these genes.

Some reports have described the influence of ITH on different tumour types,^{7–9} but their influence on

ESC tumorigenesis has not yet been fully analysed. The present study provides a descriptive integrated molecular approach to characterise the ESC genomic landscape, taking into account the influence of intratumoural heterogeneity. It has recently been suggested that ITH can be a challenge for evaluating the role of drugs in targeted therapies, as abnormalities in target genes/proteins, post-translation modifications and/or host-response can be heterogeneously distributed among different subpopulations of cancer cells within an individual tumour. There is increasing evidence that tumours are composed of subpopulations of cells with distinct genomic and epigenomic alterations and that spatial orientation in tumours affects protein expression and host-response. With the advent of latest-generation sequencing techniques, the extent and prevalence of ITH is becoming increasingly more apparent; however, its biological and clinical implications have yet to be fully established. Spatial and temporal heterogeneity have been shown to be involved in tumour adaptation to variations in tumour microenvironment and, by the same token, the tumour microenvironment may act as a selective pressure and determine the spatial and temporal ITH of a cancer.

In a recent publication we have demonstrated that a customised panel, based on the mutational profile of EC in TCGA and other surveys, was a useful tool

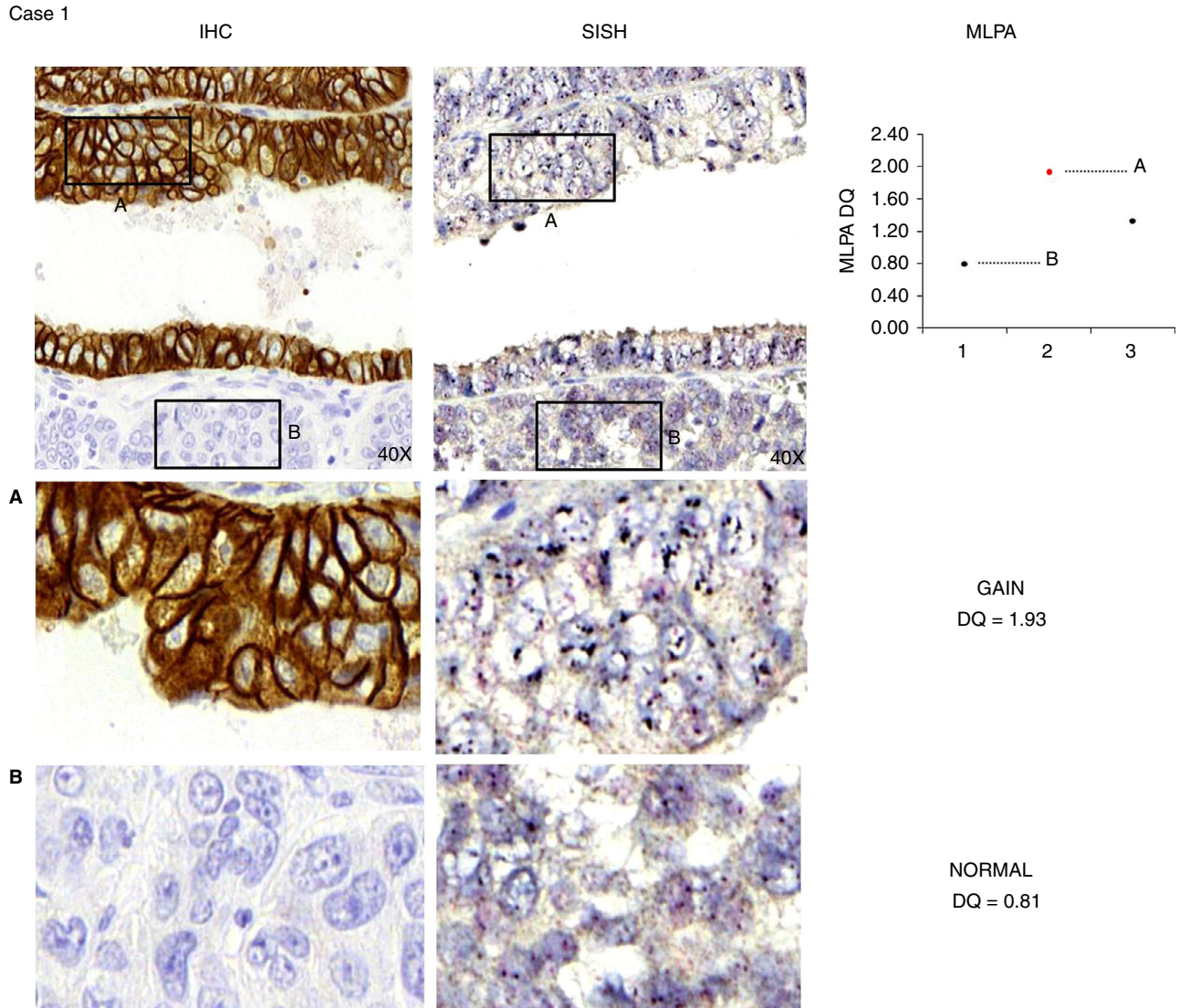


Figure 4. Representative case (case 1) harbouring *ERBB2* intratumoural heterogeneity by immunohistochemistry (IHC), silver *in-situ* hybridisation (SISH) and multiplex ligation-dependent probe amplification (MLPA). A, Magnification of the tumoural area that exhibits an *ERBB2* gained status by IHC and SISH, in concordance with a gained status by MLPA. B, Magnification of the tumoural area that exhibits an *ERBB2* normal status by IHC and SISH, in concordance with a normal status by MLPA. [Colour figure can be viewed at wileyonlinelibrary.com]

for histological typing in EC.¹⁶ The vast majority of endometrioid carcinomas exhibit *PTEN* and/or *ARID1A* mutations, while *TP53* and *PIK3CA* mutations were found in both endometrioid and ESC. Presence of *TP53* mutation and absence of *PTEN* and *ARID1A* mutation was characteristic of ESC. Using this customised panel, we observed frequent *TP53*, *PIK3CA* and *PPP2R1A* mutations, as previously reported, and furthermore we found high concordance of these gene mutations among the three tumour regions analysed. A previous study¹³ also found these altered genes in serous endometrial intraepithelial carcinoma; therefore the identification

of these alterations may have clinical implications, as they may be detected at the preinvasive stage using high-resolution techniques. Unlike other studies, we did not observe mutations in the *FBXW7* gene, the functions of which have been implicated in increasing the levels of cyclin E protein and facilitating cell cycle progression. The high prevalence and ubiquity of *TP53* mutations reinforces the importance of *TP53* gene in ESC development and progression and reinforces evidence for the existence of a common driver ancestor clone.

At gene copy-number level, we observed a high percentage of SCNA as duplications. Interestingly, our

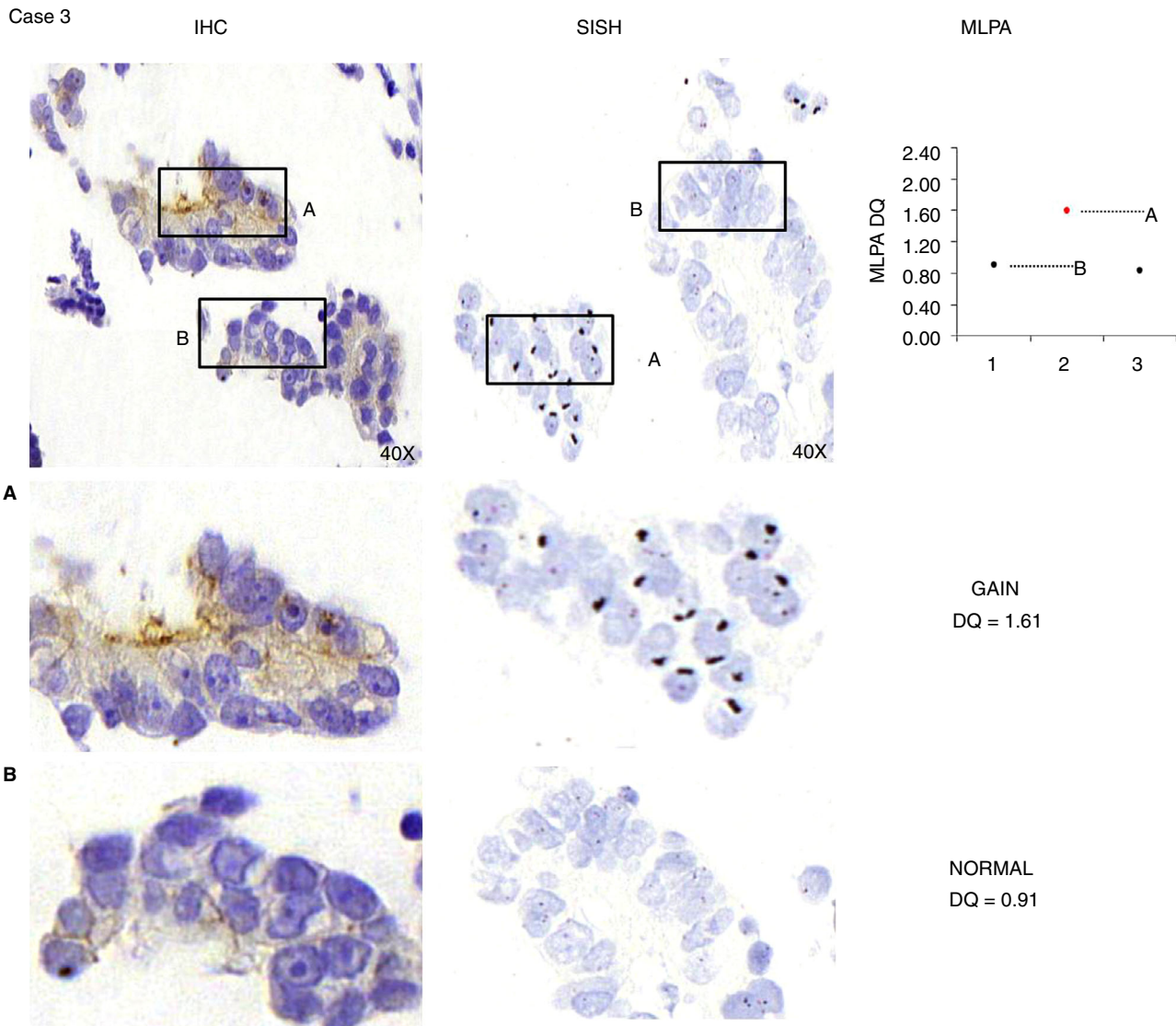


Figure 5. Representative case (case 3) harbouring *ERBB2* intratumoural heterogeneity by immunohistochemistry (IHC), silver *in-situ* hybridisation (SISH) and multiplex ligation-dependent probe amplification (MLPA). A, Magnification of the tumoural area that exhibits an *ERBB2* gained status by IHC and SISH, in concordance with a gained status by MLPA. B, Magnification of the tumoural area that exhibits an *ERBB2* normal status by IHC and SISH, in concordance with a normal status by MLPA. [Colour figure can be viewed at wileyonlinelibrary.com]

analysis indicates that SCNA are the major source of intratumour heterogeneity in ESC. In accordance with other studies, we observed gains (duplications) in genes such as *CCNE1*, *ERBB2* and *PIK3CA*.^{4,13,15} The most frequently gained genes homogeneously distributed within ESC were *PIK3CA* and *CCNE1*, suggesting that copy-number changes affecting these genes also behave as a common clonal driver event. In addition to these genes, we also detected other frequently gained genes *UCKL1*, *EV11*, *PTPN1*, *NTRK1*, *MYBL2*, *AURKA*, *BCL6* and *FKBP8*. Interestingly, the most common chromosomal locations where these genes were located were the

segments 20q12–q13.3 and 3q26–q27. Segment 20q12–q13.3 contains the most frequently gained gene, *UCKL1*, which is a uridine kinase. This kinase catalyses the phosphorylation of uridine to UMP, a rate-limiting step in DNA and RNA synthesis. This protein had been implicated in tumour progression and mediating protection from natural killer-mediated killing.²⁴ The 3q26–q27 region, which contain genes as *EV11*, *BCL6* and *PIK3CA*, had been previously reported as a significantly amplified peak region in ESC.¹³ A recent SCNA pan cancer analysis²³ has shown that tissue types of similar lineages (endometrial, ovarian, and breast cancer) have

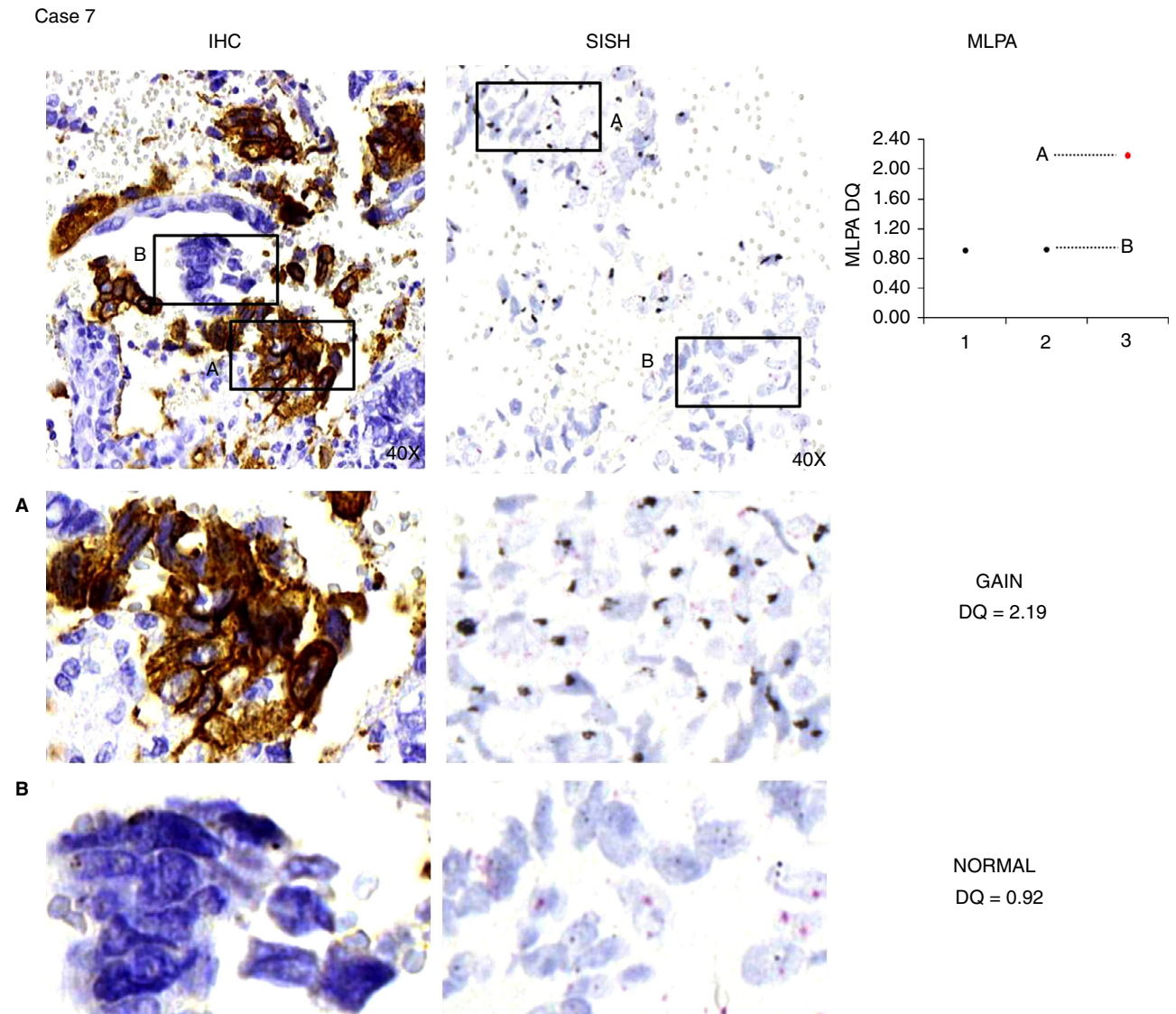


Figure 6. Representative case (case 7) harbouring *ERBB2* intratumoural heterogeneity by immunohistochemistry (IHC), silver *in-situ* hybridisation (SISH) and multiplex ligation-dependent probe amplification (MLPA). A, Magnification of the tumoural area that exhibits an *ERBB2* gained status by IHC and SISH, in concordance with a gained status by MLPA. B, Magnification of the tumoural area that exhibits an *ERBB2* normal status by IHC and SISH, in concordance with a normal status by MLPA. [Colour figure can be viewed at wileyonlinelibrary.com]

similar rates of amplifications and deletions in peak SCNA regions. Another large study, focusing upon analysis of SCNA patterns across cancer types, revealed that ovarian, breast and ESC share a similar signature profile characterised by enriched *TP53* mutations, a higher number of genomic break-points and SCNA of genes from the core glycolysis pathway, which is an emerging hallmark of human tumours.^{5,6} The tandem duplication phenotype (TDP) may be an explanation for the high duplication rates of specific genome regions on ESC. Menghi *et al.*²⁵ proposed TDP as a cancer chromotype, which is considered a prevalent genomewide cancer

configuration characterised by an enrichment of head-to-tail somatic segmental tandem duplications (TD). The TDP has been observed in breast, ovarian and endometrial cancer. This study suggests a mechanistic scenario for the TDP induction, where the excessive replicative stress produced by *CCNE1* amplification and the presence of *TP53* mutations may enhance TD formation. A large pan-cancer analysis of SCNA also suggested that whole genome duplications were correlated with *TP53* mutations and *CCNE1* amplifications, and with mutations on the *PPP2R1A* gene.²³ According to our results, *ERBB2* gains seem to be one of the most frequent

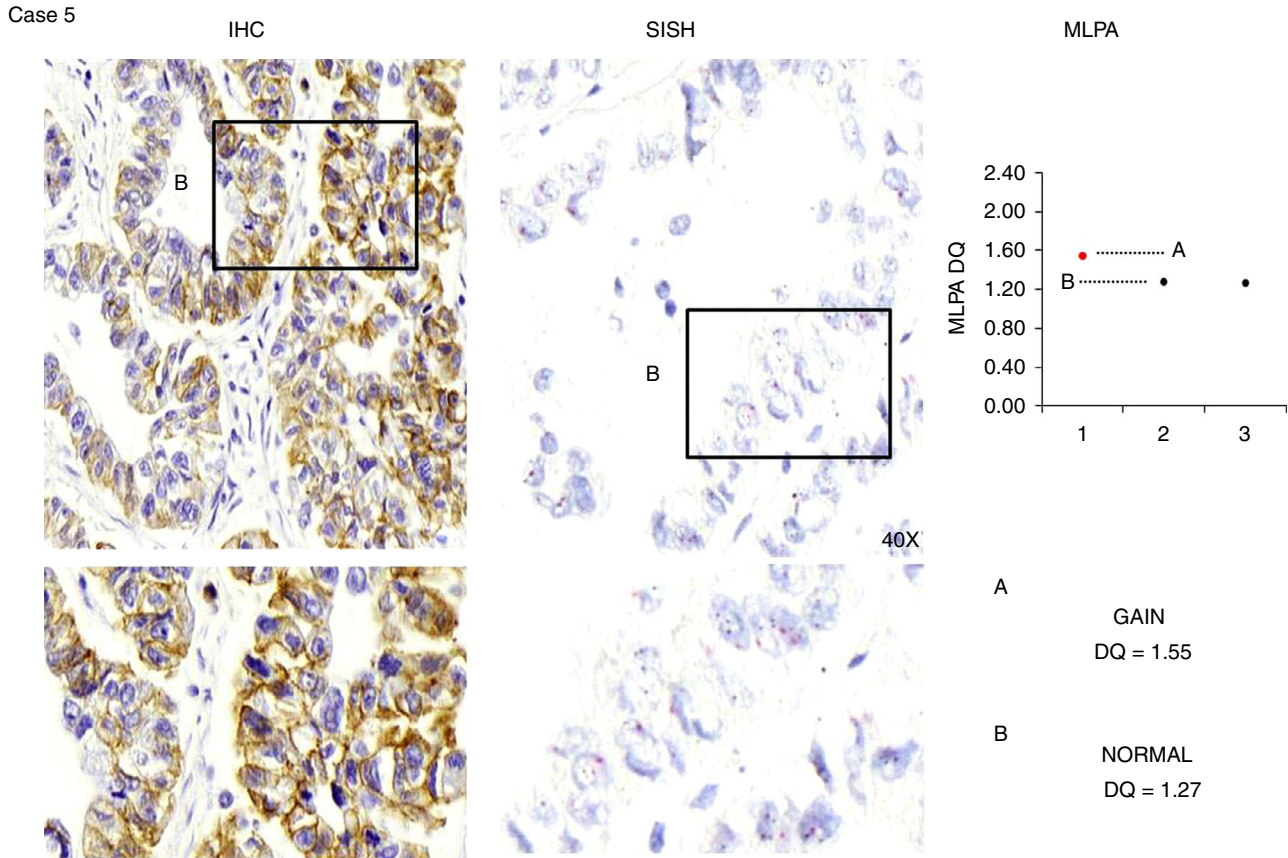


Figure 7. Representative case (case 5) harbouring *ERBB2* intratumoral heterogeneity by multiplex ligation-dependent probe amplification (MLPA). A, Tumoural area showing an *ERBB2* gained status by MLPA. B, Magnification of the tumoural area that exhibits an *ERBB2* normal status by IHC and SISH, in concordance with a normal status by MLPA. [Colour figure can be viewed at wileyonlinelibrary.com]

molecular alterations affected by intratumoral heterogeneity. This finding could suggest a probable subclonal origin of this alteration and also a role for anti-*ERBB2* therapy failure. The frequent presence of *ERBB2* alterations in ESC leads to the suggestion that *ERBB2* could be a useful target for therapy in ESC. In the case of breast cancer, *ERBB2* intratumour heterogeneity has been documented to range from 11% to 40%.^{26,27} Results regarding targeted therapy against *ERBB2* in ESC have shown controversial results,²⁸ however, Fader *et al.*²⁹ have obtained promising results. The findings of the present study may help in improving these results by providing data that may have an impact in appropriate selection of patients.^{30,31} It is worth mention that marked heterogeneity of *ERBB2* in ESC has already been detected by IHC and FISH by other authors.^{32,33} Other heterogeneously gained genes were *MYBL2*, *AURKA*, *PTPN1* and *EVII*.

Interestingly, *TP53* mutation was not found to display ITH, which implies that ITH may not negatively influence molecular classification of endometrial cancer. One of the tumours (case 5) did not show a

TP53 mutation. The tumour also exhibited a 'wild-type' p53 pattern by IHC. According to TCGA, 9% of ESC exhibit wild-type *TP53* expression, which means that negativity for p53 does not exclude the diagnosis of ESC. This case was also subjected to exome sequencing (data not shown), which confirmed that the tumour had high copy-number alterations, and targeted sequencing analysis showed the mutation profile of serous carcinoma.¹⁶

This study was limited to different parts of the primary tumour and did not include metastasis or recurrences, which is a limitation in addressing the importance of tumour heterogeneity regarding progression. Another limitation was the small number of analysed cases, which may explain the different frequencies on SCNA in comparison with other studies.

Acknowledgements

This research was funded by Generalitat de Catalunya, grant number 2017SGR1368; by CIBERONC,

grant number CB16/12/00231; and by Asociación Española Contra el Cáncer (AECC), Grupos Estables Clínicos, grant number GCTRA1804MATI. Tissue samples were obtained with the support of Xarxa catalana de bancs de Tumor and ISCIII platform of Biobancs (PT 17/0015/0027).

Conflicts of interest

The authors declare that they have no conflicts of interest.

References

- Hamilton CA, Cheung MK, Osann K *et al.* Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers. *Br. J. Cancer* 2006; **94**: 642–646.
- Del Carmen MG, Birrer M, Schorge JO. Uterine papillary serous cancer: a review of the literature. *Gynecol. Oncol.* 2012; **127**: 651–661.
- Sagae S, Susumu N, Viswanathan AN *et al.* Gynecologic cancer intergroup (GIG) consensus review for uterine serous carcinoma. *Int. J. Gynecol. Cancer* 2014; **24**: S83–S89.
- The Cancer Genome Atlas Network. Integrated genomic characterization of endometrial carcinoma. *Nature* 2013; **497**: 67–73.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; **144**: 646–674.
- Graham NA, Minasyan A, Lomova A *et al.* Recurrent patterns of DNA copy number alterations in tumors reflect metabolic selection pressures. *Mol. Syst. Biol.* 2017; **13**: 914.
- Morris L, Riaz N, Desrichard A *et al.* Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget* 2016; **7**: 109–126.
- Hu Z, Sun R, Curtis C. A population genetics perspective on the determinants of intra-tumor heterogeneity. *Biochim. Biophys. Acta* 2017; **1867**: 109–126.
- Andor N, Graham TA, Jansen M *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* 2016; **22**: 105–113.
- Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* 2012; **12**: 323–334.
- Gerlinger M, Rowan AJ, Horswell S *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 2012; **366**: 883–892.
- Barber LJ, Davies MN, Gerlinger M. Dissecting cancer evolution at the macro-heterogeneity and micro-heterogeneity scale. *Curr. Opin. Genet. Dev.* 2015; **30**: 1–6.
- Kuhn E, Wu RC, Guan B *et al.* Identification of molecular pathway aberrations in uterine serous carcinoma by genome-wide analyses. *J. Natl Cancer Inst.* 2012; **104**: 1503–1513.
- Le Gallo M, Bell DW. The emerging genomic landscape of endometrial cancer. *Clin. Chem.* 2014; **60**: 98–110.
- Zhao S, Choi M, Overton JD *et al.* Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc. Natl Acad. Sci. USA* 2013; **110**: 2916–2921.
- Cuevas D, Valls J, Gatiús S *et al.* Targeted sequencing with a customized panel to assess histological typing in endometrial carcinoma. *Virchows Arch.* 2019; **474**: 585–598.
- Schouten JP, Schouten JP, McElgunn CJ *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 2002; **30**: e57.
- Homig-Holzel C, Savola S. Multiplex ligation-dependent probe amplification (MLPA) in tumor diagnostics and prognostics. *Diagn. Mol. Pathol.* 2012; **21**: 189–206.
- Dopierala J, Damato BE, Lake SL *et al.* Genetic heterogeneity in uveal melanoma assessed by multiplex ligation-dependent probe amplification. *Invest. Ophthalmol. Vis. Sci.* 2010; **51**: 4898–4905.
- Kurman RJ, Carcangiu ML, Herrington CS, Young RH. *WHO classification of tumours of the female reproductive organs*. Lyon: IARC Press, 2014.
- Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int. J. Gynaecol. Obstet.* 2009; **105**: 103–104.
- Wolff AC, Hammond MEH, Allison KH *et al.* Human epidermal growth factor receptor 2 testing in breast cancer. *J. Clin. Oncol.* 2018; **36**: 2105–2122.
- Zack TI, Schumacher SE, Carter SL *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* 2013; **45**: 1134–1140.
- Littlefield MJ, Teboul I, Voloshyna I *et al.* Uridine cytidine kinase like-1 enhances tumor cell proliferation and mediates protection from natural killer-mediated killing. *Int. J. Immunol. Immunother.* 2014; **1**: 1–8.
- Menghi F, Inaki K, Woo X *et al.* The tandem duplicator phenotype as a distinct genomic configuration in cancer. *Proc. Natl Acad. Sci. USA* 2016; **113**: E2373–E2382.
- Allison KH, Dintzis SMSR. Frequency of HER2 heterogeneity by fluorescence in situ hybridization according to CAP expert panel recommendations: time for a new look at how to report heterogeneity. *Am. J. Clin. Pathol.* 2011; **136**: 864–871.
- Chang MC, Malowany JI, Mazurkiewicz JWM. Genetic heterogeneity in HER2/neu testing by fluorescence in situ hybridization: a study of 2,522 cases. *Mod. Pathol.* 2012; **25**: 683–688.
- Fleming GF, Sill MW, Darcy KM *et al.* Phase II Trial of trastuzumab in women with advanced or recurrent, HER2-positive endometrial carcinoma: a Gynecologic Oncology Group study. *Gynecol. Oncol.* 2010; **116**: 15–20.
- Fader AN, Roque DM, Siegel E *et al.* Randomized phase II trial of carboplatin-paclitaxel versus carboplatin-paclitaxel-trastuzumab in uterine serous carcinomas that overexpress human epidermal growth factor receptor 2/neu. *J. Clin. Oncol.* 2018; **36**: 2044–2051.
- Lopez S, Zeybek B, Santin AD. Targeting Her2/neu in uterine serous carcinoma: a paradigm shift in management. *Oncotarget* 2018; **9**: 36652–36653.
- Menderes G, Lopez S, Han C *et al.* Mechanisms of resistance to HER2-targeted therapies in HER2-amplified uterine serous carcinoma, and strategies to overcome it. *Discov. Med.* 2018; **26**: 39–50.
- Buza N, English DP, Santin AD *et al.* Toward standard HER2 testing of endometrial serous carcinoma: 4-year experience at a large academic center and recommendations for clinical practice. *Mod. Pathol.* 2013; **26**: 1605–1612.
- Buza N, Hui P. Marked heterogeneity of HER2/NEU gene amplification in endometrial serous carcinoma. *Genes Chromosomes Cancer* 2013; **52**: 1178–1186.