CENTRO INTERNACIONAL DE ESTUDOS
DE DOUTORAMENTO E AVANZADOS
DA USC (CIEDUS)

TESIS DE DOCTORADO

LIQUID BIOPSY: CIRCULATING TUMOR CELLS UTILITY IN PRECISION MEDICINE

Tais Pereira Veiga

ESCUELA DE DOCTORADO INTERNACIONAL

PROGRAMA DE DOCTORADO EN MEDICINA MOLECULAR

SANTIAGO DE COMPOSTELA

2019

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DEL AUTOR DE LA TESIS

Liquid biopsy: circulating tumor cells utility in precision medicine

Dña. Tais Pereira Veiga

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento, y declaro que:

1) *La tesis abarca los resultados de la elaboración de mi trabajo.*
2) *En su caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.*
3) *La tesis es la versión definitiva presentada para su defensa y coincide con la versión enviada en formato electrónico.*
4) *Confirmo que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.*

En Santiago de Compostela, 22 de Febrero de 2019

Fdo. Tais Pereira Veiga

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DEL DIRECTOR / TUTOR DE LA TESIS

Liquid biopsy: circulating tumor cells utility in precision medicine

Dña. Clotilde Costa Nogueira,

Dña. Laura Muinelo Romay,

D. Rafael López López,

INFORMAN:

Que la presente tesis, corresponde con el trabajo realizado por Dña. ***Tais Pereira Veiga****, bajo mi dirección, y autorizo su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como director de ésta no incurre en las causas de abstención establecidas en Ley 40/2015.*

En Santiago de Compostela, 22 de Febrero de 2019

Fdo. Clotilde Costa Nogueira
(Directora)

Fdo. Laura Muinelo Romay
(Directora)

Fdo. D. Rafael López López

Tais Pereira Veiga

(Tutor)

I, Tais Pereira Veiga, author of this thesis, have no conflict of interest to declare.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

The studies described in this thesis were performed within the framework of the Roche-CHUS Joint Unit at the University Hospital Complex of Santiago de Compostela, Spain.

The work in this thesis was financially supported by the Galician Agency of Innovation (GAIN), Consellería de Economía, Empleo e Industria, Roche Pharma and by the InveNNta Project co-financed by the European Union through the Operational Programme for Cross-border Cooperation Spain-Portugal.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A los pacientes y sus familias

A mi abuelo

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Those who can imagine anything, can create the impossible

A.T.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

INDEX

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

INDEX

ABBREVIATIONS AND ACRONYMS

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ABBREVIATIONS AND ACRONYMS

AJCC	American Joint Committee on Cancer
ASCO	American Society of Clinical Oncology
CDX	CTCs-Derived Xenograft
cfDNA	Cell Free DNA
CKs	Cytokeratins
CTC	Circulating Tumour Cell
ctDNA	Circulating Tumour DNA
DAPI	Diamino-2-Phenylindole
ddPCR	Digital Droplet PCR
DFS	Disease Free Survival
dPCR	Digital PCR
EMT	Epithelial-to-Mesenchymal Transition
EpCAM	Epithelial Cell Adhesion Molecule
ER	Estrogen Receptor
ESMO	European Society for Medical Oncology
FDA	Food and Drug Administration
FFPE	Formalin-Fixed Paraffin-Embedded
GATK	Genome Analysis Tool Kit
GO	Gene Ontology
HER2	Human Epidermal Growth Factor Receptor 2
IHC	Immunohistochemistry
MBC	Metastatic Breast Cancer
MICs	Metastasis-Initiating Cells
miRNAs	microRNAs
NSG	Non-Obese Diabetic Scid Gamma mice
NUDE	NMRI-Foxn1nu/un mice
OS	Overall Survival
PBMCs	Peripheral Blood Mononuclear Cell Fraction
PBS-BSA	Phosphate Buffered Saline – Bovine Serum Albumin
PDX	Patient-Derived Xenograft
PFS	Progression Free Survival
PR	Progesterone Receptor
qPCR	Quantitative PCR
SD	Standard Deviation
SNPs	Single Nucleotide Polymorphisms
TNBC	Triple Negative Breast Cancer
TNM	Tumour Node Metastasis Staging System
UICC	Union for International Cancer Control
V0	Visit 0
V1	Visit 1
V2	Visit 2
V3	Visit 3

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

RESUMEN

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

RESUMEN

La "biopsia líquida" es un término de reciente introducción en el campo de la oncología. Hace referencia a la detección de células tumorales, ácidos nucleicos libres y exosomas en fluidos corporales de pacientes con cáncer, siendo la sangre periférica el fluido más frecuente. Se considera un sistema de diagnóstico mínimamente invasivo que permite analizar de manera repetitiva y en tiempo real la evolución dinámica de los tumores. La biopsia líquida mejora a las biopsias convencionales de tejido a nivel de invasividad, representación de la heterogeneidad tumoral y descripción de la evolución clonal durante la resistencia al tratamiento y la diseminación metastásica. Por lo tanto, la integración de la información obtenida mediante el análisis de biopsias de tejido con la resultante del análisis de biopsia líquida, permite a los médicos adoptar acciones clínicas relevantes. Estas acciones incluyen el diagnóstico temprano, la estratificación, el pronóstico, la anticipación y la predicción de las respuestas terapéuticas durante el seguimiento de la enfermedad. Todo ello permite la implementación de la llamada medicina de precisión.

La investigación en biopsia líquida se ha centrado principalmente en células tumorales circulantes (CTCs), pero cada vez más, también crece el interés por el ADN tumoral circulante (ctDNA), los microRNAs (miRNAs) y los exosomas circulantes asociados con cáncer. Las CTCs son células tumorales que se liberan al torrente sanguíneo desde el tumor primario y/o las metástasis. Los mecanismos por los cuales estas células se liberan al torrente sanguíneo aún están en debate. Algunos estudios apuntan a un proceso de invasión activa de células con mayor potencial migratorio, como resultado de la transición epitelio-mesénquima (EMT), y otros a una liberación pasiva de células individualizadas o agrupaciones de células tumorales resultantes de una vasculatura tumoral dañada.

Las CTCs son extremadamente escasas y la mayoría de los pacientes metastásicos tienen entre 1 y 10 células por cada 10 ml de

sangre total, donde se engloban miles de millones de células sanguíneas. Además, el torrente sanguíneo es un entorno hostil para las células tumorales epiteliales y, como resultado, la vida media de las CTCs en la sangre es de entre una y dos horas y media. Por lo tanto, la mayoría de las células recuperadas se encuentran en diferentes etapas de apoptosis, esto, junto con su escasez, hace que su estudio sea especialmente difícil. Los principales desafíos en el campo son identificar el subconjunto de CTCs capaces de iniciar una lesión metastásica y su contribución al proceso de metastásis.

Las CTCs se han definido como células nucleadas, morfológicamente heterogéneas, negativas para marcadores de células sanguíneas (CD45) y positivas para citoqueratinas (CKs). Por ello, la detección estándar de CTCs ha sido a través de marcadores epiteliales como la Molécula de Adhesión Celular Epitelial (EpCAM) y CKs, considerando que estos marcadores no se expresan en las células sanguíneas ni en las células endoteliales. Sin embargo, las células tumorales epiteliales pueden sufrir EMT, que da como resultado una expresión reducida de marcadores epiteliales y una inducción de un fenotipo más mesenquimal. La visión actual es que las CTCs podrían tener un fenotipo EMT intermedio y dinámico, expresando conjuntamente marcadores epiteliales y mesenquimales, que sería el que tendría mayor plasticidad para adaptarse a las condiciones presentes en los sitios secundarios de metástasis. Las CTCs también difieren de las células hematopoyéticas en su morfología y su capacidad de deformación. Las células cancerosas epiteliales son más rígidas y más grandes que los leucocitos, sin embargo, las células que han sufrido EMT son más deformables. Además, hay estudios que apoyan la existencia de CTCs de distintos tamaños. Este conocimiento limitado de las propiedades físicas y biológicas de las CTCs impide el desarrollo de un sistema universal para la detección y el análisis de las mismas, lo que dificulta su traslación a la práctica clínica.

Los métodos de aislamiento de CTCs se pueden clasificar en tres categorías según sus propiedades físicas, biológicas o una combinación de ambas. La mayoría de los ensayos comparten un primer paso de enriquecimiento de la muestra que aumenta la

proporción de CTCs en relación con las células sanguíneas. A continuación pueden ser detectadas con diferentes aproximaciones. Dentro de las técnicas de aislamiento basadas en propiedades físicas, las más comunes son la microfiltración, la microfluídica, la dielectroforesis y el gradiente de densidad. En función de sus propiedades biológicas, como la expresión de proteínas de superficie, las CTCs pueden enriquecerse positiva o negativamente. En esta categoría podemos encontrar el sistema CellSearch, que en la actualidad es la única tecnología comercial aprobada por la *Food and Drug Administration* (FDA) para la evaluación del pronóstico de pacientes. El CellSearch selecciona las células por su expresión de EpCAM y CKs (CK8/18/19), y la ausencia de CD45. Por lo tanto, la muestra aislada sólo se enriquece en CTCs epiteliales, con la consiguiente pérdida de fenotipos más mesenquimales o *stem*. También se pueden combinar propiedades tanto biológicas como físicas. Por ejemplo, el sistema RosetteSep combina un cóctel de anticuerpos que agrega las células hematopoyéticas y una centrifugación con gradiente de densidad. Esta técnica permite el enriquecimiento negativo de una fracción de CTCs viables que se puede utilizar para estudios *in vivo*, *ex vivo* o para análisis posteriores.

El estudio de CTCs tiene un gran potencial como fuente de información sobre dianas terapéuticas y resistencia a terapia en pacientes con cáncer. La información más básica que podemos obtener de las CTCs es su contaje, pero también pueden estudiarse mediante la genómica, la transcriptómica y la proteómica. Para el recuento de CTCs, se ha establecido como mal pronóstico la detección de más de 5 CTCs en cáncer de mama y próstata, o más de 3 CTCs en el cáncer colorrectal por 7,5 ml de sangre. Sin embargo, actualmente ninguna guía clínica recoge el uso del contaje de CTCs para la toma decisiones clínicas. Con respecto a la transcriptómica, el estudio más relevante se ha llevado a cabo en cáncer de próstata, analizando la expresión de la isoforma v7 del receptor de andrógenos, que proporciona información sobre sensibilidad y resistencia a quimioterapia. En cuanto al estudio del genoma, las mutaciones más relevantes analizadas en CTCs afectan a los genes *EGFR*, *KRAS*, *PIK3CA*, *AR* y *BRAF*. Finalmente, a nivel proteico se han

caracterizado proteínas con potencial interés terapéutico como el regulador inmunitario PDL1, o los receptores de estrógeno (ER) y progesterona (PR) y el oncogén HER2 en cáncer de mama. Los análisis de CTCs permiten mediciones precisas y en tiempo real de la heterogeneidad tumoral y de las poblaciones subclonales resultantes de la presión selectiva causada por los diferentes tratamientos. Además, se podrían aplicar en la práctica clínica para monitorizar la enfermedad en tiempo real, mejorar la estratificación de los pacientes y facilitar el cambio de terapia en base a la expresión de biomarcadores específicos y al contaje de CTCs. En la actualidad, varios ensayos clínicos en fase III, principalmente en pacientes con cáncer de mama, están explorando cómo las CTCs pueden contribuir a una evaluación temprana de los efectos de las terapias en cáncer.

En los últimos años, el término biopsia líquida también se ha extendido al análisis del ctDNA, que está constituido por fragmentos de DNA liberados por las células del tumor primario, CTCs lisadas, micrometástasis o metástasis abiertas a la sangre. En un futuro próximo, es probable que el estudio de CTCs y ctDNA sea complementario, ya que tanto los análisis de CTCs como los de ctDNA reflejan aspectos biológicamente diferentes de la enfermedad. Las CTCs permiten un análisis molecular del tumor tanto a nivel global como a nivel de célula individualizada, mientras que los análisis de ctDNA proporcionan una imagen global del perfil genético del tumor.

Los modelos preclínicos basados en CTCs tienen un gran potencial para la investigación básica y preclínica en cáncer, ya que nos proporcionan información sobre el proceso de diseminación metastásico. Hasta ahora, tres grupos han descrito el establecimiento de cultivos a largo plazo de CTCs derivados de pacientes y varios más han descrito el crecimiento *in vitro* de estas células a corto plazo. Sin embargo, se trata de una tarea complicada porque además de su baja proporción, muchas CTCs tienen una capacidad de proliferación limitada y se vuelven inviables después de pocas divisiones celulares. Los modelos *in vivo* también son muy interesantes en el campo de la oncología para el descubrimiento de biomarcadores, la comprensión

de los mecanismos de resistencia a fármacos y el desarrollo de nuevas terapias. Varios artículos han descrito el potencial de las CTCs para generar modelos preclínicos de xenoinjertos derivados de CTCs (CDXs) que coincidan con los tumores de los pacientes de los que han derivado. Sin embargo, solo tres estudios, uno en cáncer de mama luminal y dos en cáncer de pulmón, pudieron establecer un tumor en CDXs.

El cáncer de mama es la principal causa de muerte relacionada con cáncer en mujeres y tiene una tasa de incidencia más alta que cualquier otro tipo tumoral, siendo la enfermedad metastásica la responsable de la mayoría de las muertes en estas pacientes. Se trata de una enfermedad heterogénea que se ha clasificado en tres grupos terapéuticos básicos, basados en la expresión de diferentes marcadores (receptores hormonales, Ki67 y HER2), diferencias en la morfología e implicaciones clínicas. Estos tres grupos son el subtipo luminal (que engloba al luminal A y el luminal B), el HER2 sobreexpresado y el subtipo triple negativo (TNBC). Los pacientes con TNBC tienen una supervivencia específica corta y un pronóstico desfavorable, y su tratamiento recomendado es la quimioterapia sistémica.

Las estrategias de tratamiento estándar para el cáncer de mama metastásico (MBC) se basan en el uso de quimioterapia, terapia dirigida a HER2 y terapia endocrina. Además han llegado a la práctica clínica nuevos tipos de medicamentos dirigidos a cambios específicos en las células tumorales, tanto solos como en diferentes combinaciones. Sin embargo, la heterogeneidad tumoral y la aparición de resistencias plantean un gran problema a la hora de seleccionar terapias De manera tradicional, las decisiones de tratamiento en MBC se basan en las características del tumor primario. Aun así, esta opción tiene limitaciones, ya que en algunos pacientes la biopsia de tejido no es posible o han pasado muchos años desde la biopsia hasta la recaída. Además, una sola biopsia de tejido es insuficiente para representar cambios evolutivos en el tumor y los diversos mecanismos de resistencia que conducen a poblaciones clonales individuales en metástasis en progresión.

Las CTCs han sido objeto de estudio en cáncer de mama desde hace más de 10 años, explorando su potencial como biomarcadores con factor pronóstico, marcadores de propagación de micrometástasis y como una valiosa forma de biopsia líquida y herramienta para la selección de terapias. Sin embargo, muchas preguntas siguen sin respuesta. Todavía se desconoce cuál es la heterogeneidad real de las CTCs, cuánta concordancia existe entre las CTCs y el tumor primario o cuál es su estado de EMT; razón por la cual existe un interés creciente en la caracterización fenotípica de CTCs en MBC.

El objetivo de esta tesis fue profundizar en la biología de las CTCs y evaluar su utilidad como herramienta de monitorización en pacientes con cáncer. Con este objetivo, aislamos y analizamos las CTCs de pacientes con cáncer para estudiar su expresión genética e identificar biomarcadores asociados con la resistencia a la terapia. Además, queríamos evaluar la utilidad clínica y el valor pronóstico de los marcadores identificados y su relación con los parámetros clínicos de los pacientes. Asimismo, teníamos el objetivo de desarrollar y caracterizar modelos preclínicos *in vivo* para el estudio de la biología de las CTCs y validar la biopsia líquida como una herramienta de medicina personalizada, teniendo en cuenta la optimización de nuevas tecnologías para el aislamiento y la caracterización de CTCs.

Para realizar este estudio, recogimos muestras de sangre de controles y de tres cohortes de pacientes (MBC, TNBC y cáncer colorrectal) en diferentes puntos de la enfermedad. Además, en algunos pacientes pudimos obtener muestras emparejadas de parafinas del tumor primario. Enumeramos las CTCs con CellSearch. Aislamos las CTCs de la cohorte de MBC mediante un enriquecimiento negativo con RosetteSep y de la cohorte de TNBC mediante el enriquecimiento positivo de células EpCAM+ con *beads* inmunomagnéticas. Se realizó un estudio de expresión génica longitudinal por qPCR en la cohorte de MBC. Además, establecimos un CDX a partir de un caso de TNBC y realizamos estudios de inmunohistoquímica, RNA-seq y qPCR en este paciente. Finalmente, validamos con muestras de sangre de pacientes con cáncer colorrectal metastásico un nuevo dispositivo de microfluídica con mayor

rendimiento que CellSearch. Este dispositivo, denominado "chip CROSS" fue diseñado por el INL (International Iberian Nanotechnology Laboratory). Finalmente, optimizamos el análisis molecular posterior en el chip CROSS mediante inmunofluorescencia y PCR digital (ddPCR).

Para estudiar la importancia clínica de las CTCs en pacientes con MBC, estas se aislaron y se caracterizaron molecularmente evitando el sesgo introducido por la selección de marcadores. Utilizando esta aproximación se enriquecen todos los fenotipos de CTCs, hasta la fecha, solo un estudio ha utilizado un enfoque similar. Además, analizamos de forma conjunta las muestras de los diferentes subtipos moleculares de cáncer de mama para detectar biomarcadores relacionados con la progresión tumoral o con un interés clínico independientemente del subtipo. Nuestros resultados mostraron que los pacientes con ≥ 5 CTCs tuvieron una supervivencia gobal (OS) más corta considerando el momento del diagnóstico metastásico. Además, el recuento de ≥ 5 CTCs en pacientes después de un ciclo de tratamiento también se asoció con peor OS, pero con una significación estadística mayor. Con respecto a la distribución de CTCs en diferentes subtipos, encontramos una mayor frecuencia en el subconjunto luminal. Observamos una tasa de concordancia entre la expresión de *ERBB2* en las CTCs y el estado de HER2 en el tumor primario similar a la descrita en otros artículos. Sin embargo, encontramos mayor detección de expresión de marcadores epiteliales, EMT y *stem* en CTCs que estudios anteriores. Nuestros análisis moleculares revelaron que la expresión de genes específicos epiteliales (*EpCAM*, *KRT19*) o relacionados con el cáncer de mama (*ERBB2*) se asociaba con la presencia de ≥ 5 CTCs antes del tratamiento. Además, existe una asociación con la expresión de un marcador epitelial (*CDH1*) en pacientes con una o más CTCs, después de la quimioterapia. Por lo tanto, nuestros datos sugieren que la metodología de enriquecimiento negativo permite la detección de marcadores específicos de CTCs y cáncer de mama, incluso en casos no positivos mediante CellSearch, lo que demuestra que este enfoque podría mejorar algunas de las limitaciones de CellSearch.

El análisis de correlación entre la expresión génica y los subtipos de cáncer de mama mostró que, al contrario que *ALDH1A1*, la expresión de *KRT19* es mayor en el subtipo luminal A y menor en el subtipo TNBC, lo que sugiere que ALDH1 podría ser útil para la detección de CTCs, en los que la detección de células EpCAM+ es limitada. Por lo tanto, combinando el análisis de ambos marcadores, podríamos identificar CTCs en los diferentes subtipos de cáncer de mama. Además, encontramos que los tumores primarios con distinto estado de receptores hormonales conducen a CTCs con diferentes perfiles de expresión. Los tumores primarios negativos para el PR se correlacionaron con la expresión de *BCL11*, *KRT5* y *RB1* en CTCs antes del tratamiento; mientras que los tumores ER positivos se correlacionaron con la expresión de *KRT19* y *RB1* antes del tratamiento y *GDF15*, *CDH1* y *CD36* después de un ciclo de quimioterapia. En este contexto, la alta expresión de diferentes genes en CTCs podría estar asociada con un comportamiento más agresivo y con la resistencia a terapia endocrina. A continuación, estudiamos la correlación del perfil de expresión génica de las muestras con la evolución clínica de los pacientes. Si bien el análisis de marcadores de expresión del tejido de los tumores primarios no se asoció con el desenlace clínico de los pacientes, se encontró asociación entre la expresión de *ERBB2*, *PALB2* y *MYC* en CTCs antes del tratamiento y un peor pronóstico, lo que subraya el potencial del análisis de CTCs para evaluar el pronóstico de los pacientes. Además, identificamos asociación entre la expresión de *MYC* y *CDK4* (una diana, junto con *CDK6*, de inhibidores de CDK4/6 como Palbociclib, Ribociclib o Abemaciclib) después del tratamiento y una peor evolución de la enfermedad. Estos resultados resaltan la importancia de la monitorización de la evolución del tumor durante la terapia mediante el análisis molecular de las CTCs, ofreciendo nuevas perspectivas a los oncólogos para la puesta en práctica de terapias dirigidas. Aunque ninguno de los pacientes de la cohorte de MBC fue tratado con inhibidores de CDK4/6, esto abre nuevas futuras vías de investigación.

En nuestro análisis encontramos una firma $EpCAM^{high}VIM^{low}$ que fue capaz de predecir un peor resultado clínico con mayor

significación estadística que la expresión de *EpCAM* o *VIM* por si solos. Además, cuando incluimos la expresión de *ALDH1A1* en la firma, *EpCAM*^{high}*VIM*^{low}*ALDH1A1*^{high}, su potencial de predicción mejoró. Estos datos sugieren que un estado epitelial-*stem* de las CTCs puede dar lugar a una enfermedad más agresiva. En conjunto, nuestros resultados enfatizan la importancia del método de detección elegido para el aislamiento de las CTCs, ya que algunos métodos pueden subestimar u obviar ciertas subpoblaciones de CTCs que podrían poseer roles relevantes.

Describimos por primera vez la generación de un modelo de CDX de un paciente con TNBC, lo que demuestra que las CTCs de este caso clínico son tumorigénicas y permiten el desarrollo de un sistema *in vivo* que permite obtener una mejor comprensión de la biología del tumor en este subtipo de cáncer. Tanto el análisis histológico como el estudio de expresión mediante qPCR y RNA-seq confirmaron la semejanza de CDX con el tumor primario del paciente. Analizamos diferentes muestras de tejido tumoral, de tres pases de ratones, y sangre del paciente a lo largo del tiempo, realizando un seguimiento molecular de la enfermedad. Detectamos cambios moleculares entre todas las muestras, lo que respalda la relevancia de la monitorización mediante biopsia líquida como una herramienta valiosa para comprender la evolución del tumor. Además, el análisis de los CDX nos permitió identificar mecanismos moleculares clave involucrados en el desarrollo de TNBC que podrían representar dianas terapéuticas relevantes. El análisis de ontología génica señaló la vía WNT como el principal proceso de señalización subyacente con regulación al alza en todas las muestras analizadas. Los análisis de RNA-seq llevaron a la identificación de genes altamente expresados en todas las muestras tumorales, sugieriendo su relevancia en la progresión tumoral de este paciente. Cinco genes seleccionados se analizaron adicionalmente en CTCs aisladas de este paciente y en CTCs de la cohorte de pacientes TNBC. Este análisis demostró que *AURKB*, *HIST1H4A1*, *MELK* y *PCDHA8* podrían ser empleados para la detección de la presencia de CTCs y, por lo tanto, pueden ser valiosos como indicadores de la diseminación del tumor. Entre estos genes, encontramos que altos niveles de expresión de *MELK* en las CTCs de la cohorte TNBC se

asociaron con tasas más bajas de OS y supervivencia libre de progresión (PFS). Por lo tanto, nuestro trabajo marca un hito que señala a *MELK* como un posible marcador de supervivencia detectado por biopsia líquida y también como una posible diana terapéutica para la que incluso ya existen inhibidores. Con el desarrollo de un modelo de CDX fuimos capaces de integrar el análisis de CTCs, de muestras de tejido, la generación de CDX y la tecnología RNA-Seq como una estrategia valiosa para profundizar en la biología de TNBC, brindando a los clínicos nuevas posibles dianas terapéuticas, así como posibles marcadores que podrían mejorar el manejo clínico de estos pacientes.

La heterogeneidad molecular de las CTCs y sus implicaciones clínicas hacen necesario mejorar los métodos de aislamiento, permitiendo la maximización de la detección de dichas células y, por lo tanto, su caracterización. En este estudio validamos una nueva tecnología de aislamiento por microfluídica llamada CROSS chip. Este sistema se testó y comparó con CellSearch utilizando como prueba de concepto una cohorte de pacientes con cáncer colorrectal metastásico. A continuación, las células aisladas con el CROSS chip fueron analizadas por ddPCR para detectar la presencia de una mutación específica del gen *APC* frecuente en pacientes con cáncer colorrectal, para confirmar su origen maligno y validar la capacidad de caracterización molecular posterior con este sistema. Teniendo en cuenta los resultados obtenidos en este estudio comparativo con una cohorte pequeña de pacientes con cáncer colorrectal metastásico y debido a que el CROSS chip tiene una alta sensibilidad, propusimos un valor referencia mayor que el del CellSearch ($\geq$ 7 CTC / 7.5 ml de sangre total). Este nuevo valor de referencia permitió la estratificación de pacientes en 2 poblaciones definidas con diferente OS. Sin embargo, se son necesarios estudios adicionales sobre cohortes más grandes de pacientes, que incluyan diferentes tipos de tumores, para poder valorar la relevancia clínica de este método para la monitorización y la caracterización de pacientes metastásicos.

La biopsia líquida ofrece alternativas de análisis en tumores que no son fáciles de biopsiar y permite volver a clasificar en distintos estadios a los pacientes y analizar molecularmente las metástasis.

Además, el diagnóstico mediante biopsia líquida puede servir como un control en tiempo real del estado del tumor, que permitiría adaptar la terapia a las necesidades individuales del paciente con cáncer. En este sentido, nuestros resultados demuestran que el análisis de CTCs puede proporcionar información clínicamente relevante y respalda la importancia de la monitorización mediante biopsia líquida como una herramienta valiosa para comprender la evolución del tumor. Sin embargo, la falta de conocimiento respecto a la biología dinámica de las CTCs podría obstaculizar la interpretación de los resultados clínicos. Nuestros estudios resaltan la necesidad de la caracterización de las CTCs más allá del contaje, para poder proporcionar a los pacientes una medicina más precisa y personalizada. En este contexto, abordamos diferentes enfoques (aislamiento por tamaño, aislamiento magnético, enriquecimiento negativo y el establecimiento de modelos preclínicos) que nos permitieron realizar una aproximación a las CTCs desde diferentes perspectivas, mejorando el rendimiento de los resultados obtenidos hasta el momento con otras tecnologías e identificando biomarcadores con potencial traslación a la práctica clínica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

SUMMARY

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

SUMMARY

The term “liquid biopsy” was introduced in oncology several years ago. It makes reference to the detection of tumour cells, cell-free nucleic acids and exosomes in body fluids from cancer patients, being the most frequent peripheral blood. It is considered a minimal invasive diagnostic system that allows a repetitive, real-time questioning of the dynamic evolution of tumours. Liquid biopsy surpass conventional tissue biopsies in terms of invasiveness, representation of tumour heterogeneity and description of clonal evolution during therapy resistance and metastatic dissemination. Hence, integrating the information obtained by standard tissue biopsies analysis with liquid biopsy allows the clinicians to adopt relevant medical actions. These actions include early diagnosis, staging, prognosis, anticipation and prediction of therapy responses during the follow-up of the disease, enabling the implementation of the so-called precision medicine.

Liquid biopsy applications had focused mainly on circulating tumour cells (CTCs) but recently it has been broadened to circulating tumour DNA (ctDNA), microRNAs (miRNAs) and cancer associated exosomes. CTCs are tumour cells that are released into the blood from the primary tumour and/or metastatic sites. The mechanisms by which these cells are released into the bloodstream are still under debate, with some studies pointing to active invasion of cells with increased migratory potential, as a result of epithelial-to-mesenchymal transition (EMT), and others to a release by passive shedding of individual cells or tumour cell clusters resulting from compromised tumour vasculature.

CTCs are extremely rare and most metastatic patients have as few as 1 to 10 cells per 10 mL of whole blood, which englobes billions of blood cells. In addition, bloodstream is a harsh environment for epithelial tumour cells and, as a result, CTCs life-span in the blood is described to be between one and two and a half hours. Therefore, most of the recovered cells are at different stages of apoptosis, this, together with their scarcity, makes their study especially difficult. So, major challenges in the field are identifying the subset of CTCs capable of

initiating a metastatic lesion and their contribution to the metastatic process.

CTCs have been defined as nucleated cells, morphologically heterogeneous, negative for blood cell markers (CD45) and positive for cytokeratin (CKs). Thus, the standard detection of CTCs has been through epithelial markers like the Epithelial Cell Adhesion Molecule (EpCAM) and CKs, considering that these markers are not expressed on the surrounding blood cells neither on the endothelial cells. However, epithelial tumour cells can undergo EMT that results in a reduced expression of epithelial markers and an induction of a more mesenchymal phenotype. The current picture is that CTCs might have an intermediate and dynamic EMT phenotype, co-expressing both epithelial and mesenchymal markers, and this phenotype might have the highest plasticity to adapt to the conditions present in secondary sites. CTCs also differ from hematopoietic cells on their morphology and their deformability. Epithelial cancer cells are reported to be stiffer and larger than leukocytes, however, cells undergoing EMT are more deformable and CTCs of various sizes have been reported. This limited understanding of both physical and biological properties of CTCs is preventing the development of a universal system for CTCs detection and analysis, hampering its translation into the clinical practice.

The isolation methods of CTCs can be classified in three different categories, depending on their physical or biological properties, or a combination of both. Most of the assays share a first step of enrichment of the sample that increases the yield of CTCs in relation with blood cells. Then, CTCs can be detected by different approaches. Within the isolation techniques based on physical properties, the most common are microfiltration, microfluidics, dielectrophoresis and density gradient. Based on their biological properties, like the expression of surface protein markers, CTCs can be positively or negatively enriched. In this category we can find the CellSearch System, which nowadays is the only commercial technology approved by the Food and Drug Administration for prognostic purposes. CellSearch selects cells by their EpCAM and CKs (CK8/18/19)

expression, and the absence of CD45. Thus, the isolated sample is only enriched in epithelial CTCs, with the consequent loss of more mesenchymal or stem phenotypes. In addition, both biological and physical properties can be combined. This can be achieved for example with RosetteSep, a system that combines an antibody cocktail that crosslinks the hematopoietic cells and a density gradient centrifugation. This technique allows the negative enrichment of a viable CTCs fraction that can be used for *in vivo* or *ex vivo* studies or for downstream analysis.

The study of CTCs offers a mine of information on therapeutic targets and resistance to therapy in cancer patients. The most elemental information we can get from CTCs is enumeration but they can also be approached with genomics, transcriptomics and proteomics. For the enumeration of CTCs, traditional criteria has established more than 5, for breast and prostate cancer, or more than 3 CTCs, in colorectal cancer, per 7.5 mL of blood as bad prognosis. However, CTCs enumeration is currently not advised in clinical guidelines for any clinical decision. Regarding transcriptomics, the most relevant study has been on prostate cancer, with the mRNA expression of ARv7, which provides information about drug sensitivity and resistance. On the genomic approach the most relevant mutations are on the *EGFR*, *KRAS*, *PIK3CA*, *AR* and *BRAF* genes. Finally, at a protein level, the object of study have been proteins with potential therapeutic interest like the immune checkpoint regulator PDL1, or the estrogen (ER) and progesterone (PR) receptors and the HER2 oncogene in breast cancer. CTCs analyses allow precise, real-time measurements of cancer heterogeneity and the subclonal populations resulted from the selective pressure caused by the different treatments. In addition, they could be translated into the clinical practice through real-time monitoring, stratification of patients, and therapy switch based on CTCs counts and CTCs biomarker expression. At the present time, several phase III clinical trials, mostly in breast cancer patients, are exploring how CTCs can contribute to an early assessment of therapy effects.

In the last few years, term liquid biopsy has also been extended to the analysis of ctDNA, which is constituted from fragments of DNA derived from primary tumours, lysed CTCs, micrometastases or overt metastases into the blood. In the near future, CTCs and ctDNA technologies are likely to be synergistic, as both CTCs and ctDNA analyses reflect biologically different aspects of the disease. CTCs allow a molecular analysis of the tumour both at bulk or single cell level while ctDNA analyses provide a global picture of the genetic status of the disease.

CTCs-based preclinical models have a great potential for basic and preclinical cancer research as they provide us with information of the metastatic dissemination process. So far, three groups have reported patient-derived CTCs long-term cultures and several more have described *in vitro* short-term growth. Nevertheless, it is a complicated task because in addition to their low numbers, many CTCs have limited proliferation ability and they become non-viable after a few cell divisions. *In vivo* models are also very interesting on the oncology field in terms of the discovery of biomarkers, the understanding of drug resistance mechanisms and the development of new therapies. Several reports have described the potential of CTCs to generate preclinical models matched to individual patient's tumours on CTCs-derived xenografts (CDX). However, only three studies, one in luminal breast cancer and two in lung cancer, were able to establish a tumour on CDXs.

Breast cancer is the leading cancer-related cause of death in women and has a higher incidence rate than any other cancer, being the metastatic disease the responsible for the majority of deaths in these patients. Breast cancer is a heterogeneous disease that has been classified in three basic therapeutic groups based in the expression of different markers (hormonal receptors, Ki67 and HER2), differences in morphology and their clinical implications. These three groups are luminal (which englobes luminal A and luminal B), HER2 overexpression and triple-negative breast subtype (TNBC). TNBC patients have a short disease-specific survival and poor prognosis and their recommended therapy is systemic chemotherapy.

The standard treatment strategies for metastatic breast cancer (MBC) are based in the use of chemotherapy, HER2-targeted therapy and endocrine therapy. In addition, new types of drugs that target specific changes in cancer cells have reached the clinical practice both alone or as part of different treatment combinations. However, tumour heterogeneity and the emergence of resistance poses a large problem for therapeutic strategies. Traditionally, treatment decisions in MBC are based on the characteristics of the primary tumour. Still, this strategy has limitations as in some patients tissue biopsy is either not possible or it has been done many year before the patient relapses. Besides, a single tissue biopsy may be insufficient to represent evolutionary changes in the tumour and the diverse resistance mechanisms driving individual clonal populations of progressing metastases.

CTCs have been studied in breast cancer for more than 10 years, exploring their potential as biomarkers for breast cancer as prognosis factors, markers of micrometastasis spread and as a tool for liquid biopsy and therapy selection. However, so far, many questions remain unanswered, like the heterogeneity of CTCs, how much concordance exists between CTCs and primary tumour or which is their EMT state; being the reason why there is an increasing interest on the phenotypic characterisation of CTCs in MBC.

In this context, the objective of this thesis was to delve into the biology of CTCs through liquid biopsy and to evaluate their usefulness as a monitoring tool for cancer patients. To this aim, we isolated and analysed CTCs from cancer patients in order to study their gene expression and to identify biomarkers associated with resistance to therapy. In addition, we wanted to evaluate the clinical utility and prognostic value of the identified markers and its relationship with patients' clinical parameters. Besides, we had the objective of developing and characterising *in vivo* preclinical models for the study of CTCs biology and to validate liquid biopsy as a tool for personalised medicine, taking into account the optimisation and validation of new technologies for the isolation and characterisation of CTCs.

To perform this study we collected blood samples from controls and three cohorts of patients (MBC, TNBC and colorectal cancer) at different time-points of the disease. In addition, in some patients we were able to obtain paired formalin-fixed paraffin-embedded samples from the primary tumour. We enumerated CTCs with CellSearch. We isolated CTCs by a negative enrichment with RosetteSep on the MBC cohort and by positive enrichment of EpCAM+ cells with immunomagnetic beads on the TNBC cohort. We performed a prospective longitudinal gene expression study by qPCR on the MBC cohort. In addition, we established a CDX from a TNBC case and we performed immunohistochemistry, RNA-seq and qPCR studies in this patient. Finally, we validated with blood samples from metastatic colorectal cancer patients a new microfluidic device with higher yield than CellSearch. This device, named "CROSS chip" was designed by the INL (International Iberian Nanotechnology Laboratory). Finally, we optimised molecular downstream analysis in the CROSS chip by immunofluorescence and by ddPCR.

We aimed to isolate CTCs, avoiding the bias introduced by marker selection, and also to characterise them at a molecular level, before and after treatment, in order to study their clinical significance in MBC patients. With this approach, all CTCs phenotypes are enriched, and to this date, only one study has used a similar approach. In addition, we analysed the different molecular subtypes of breast cancer collectively in order to detect biomarkers related to tumour progression or with a clinical interest in a subtype independent manner. Our results showed that patients with ≥ 5 CTCs had a shorter overall survival (OS) considering the metastatic diagnose time point. Interestingly, the enumeration of ≥ 5 CTCs in patients after one cycle of treatment was also associated with worst OS but with better significance. Regarding the distribution of CTCs in different subtypes, we found a higher frequency in the luminal subset. We found a similar rate of concordance to previous reports between *ERBB2* expression on CTCs and HER2 status on the primary tumour. However, we reported a higher detection rate on epithelial, EMT and stem markers expression on CTCs that previous studies. Our molecular analyses revealed that the expression of specific epithelial (*EpCAM*, *KRT19*) or

breast cancer related genes (*ERBB2*) was associated with the presence of ≥ 5 CTCs before therapy. Furthermore, we found association in the expression of one epithelial marker (*CDH1*) in patients with one or more CTCs after chemotherapy. Thus, our data suggest that the CTCs negative enrichment methodology allows the detection of specific markers of CTCs and breast cancer, even in CellSearch non positive cases, demonstrating that this approach might overcome some of the CellSearch limitations.

Correlation analysis between gene expression and breast cancer subtypes showed that, contrary to *ALDH1A1*, the expression of *KRT19* was higher in luminal A subtype and lower in TNBC subtype, suggesting that ALDH1 might be useful for the detection of CTCs in the cases in which the detection of EpCAM+ cells is limited. Thus, combining the analysis of both markers we could identify CTCs in all the different breast cancer subtypes.

In addition, we found that primary tumours with different hormonal receptor characterisation lead to CTCs with different expression profiles. PR negative primary tumours were correlated with the expression of *BCL11A*, *KRT5* and *RB1* in CTCs before treatment; while ER positive tumours were correlated with *KRT19* and *RB1* expression before treatment and *GDF15*, *CDH1* and *CD36* after one cycle of chemotherapy. In this context, high expression of different genes in CTCs could be associated with a more aggressive behaviour and resistance to endocrine therapy. Next, we studied the correlation of the gene expression profile of the samples with the outcome of the patients. While primary tumours tissue expression did not associate with the outcome of the patients, we found an association between the expression of *ERBB2*, *PALB2* and *MYC* on CTCs before treatment, with a worse prognosis of the patients, which remarks the potential of CTCs analysis for patients' prognosis. Moreover, we identified an association between the expression of *MYC* and *CDK4* (a target, together with *CDK6*, of CDK4/6 inhibitors such as Palbociclib, Ribociclib or Abemaciclib) after treatment and a worse prognosis of the patients. These results highlight the importance of the tumour evolution monitoring during treatment by molecular

analysis of CTCs, offering new perspectives to clinicians for targeted therapies. Although none of the patients from the MBC cohort was treated with CDK4/6 inhibitors, this opens new avenues for research.

In our analysis we have found an *EpCAM*^{high}*VIM*^{low} signature which was able to predict worst outcome with better significance than *EpCAM* or *VIM* expression alone. In addition, when we included *ALDH1A1* expression in the signature, *EpCAM*^{high}*VIM*^{low}*ALDH1A1*^{high}, its outcome prediction potential was improved. These data suggest that an epithelial-stem state of the CTCs may give rise to a more aggressive disease. Altogether, our results emphasize the importance of the chosen detection method for the isolation of CTCs, as some methods may underestimate or neglect certain subpopulations of CTCs with putative relevant roles.

We were also able to describe for the first time the generation of a CDX mice model from a TNBC patient, demonstrating that CTCs from a TNBC patient are tumorigenic and constitute an attractive *in vivo* system to gain a better understanding of tumour biology in this cancer subtype. Both histological, gene expression and RNA-seq analysis confirmed the resemblance of the CDX with the patient's primary tumour. We analysed different tumour tissue samples, from three mice passages, and blood from the patient over time, performing a molecular tracking of the disease. We detected molecular changes among all the samples, further supporting the relevance of liquid biopsy monitoring as a valuable tool for understanding tumour evolution. Moreover, the analysis of the CDXs allowed us to identify key molecular mechanisms involved in TNBC development that could represent relevant therapeutic targets. Gene ontology analysis pointed to the WNT pathway as the main underlying signalling process up-regulated in all analysed samples. Comprehensive RNA-seq data analyses led to the identification of highly expressed genes on all tumour tissue samples, suggesting their relevance in tumour progression in this patient. Five selected genes were further analysed in CTCs isolated from this patient and in CTCs from a TNBC patient cohort. This analysis demonstrated that *AURKB*, *HIST1H4A1*, *MELK* and *PCDHA8* could be potentially used to detect the presence of

CTCs and therefore, valuable as indicators of tumour dissemination. Among these five genes, we found that high expression levels of *MELK* in CTCs from the TNBC cohort were associated with lower OS and progression free survival (PFS) rates. Therefore our work marks a milestone pointing *MELK* as a potential survival marker detected by liquid biopsy and also a potential therapeutic target, with the additional value of the existence of active *MELK* inhibitors. Thus, with the development of a CDX mouse model, we were able to integrate CTCs analysis, tissue samples, CDXs generation and RNA-seq technology as a valuable strategy to delve into TNBC biology, providing clinicians with new potential therapeutic targets and markers that could improve the clinical management of these patients.

The discovery of the molecular heterogeneity of CTCs and its clinical implications highlight the need of improvement of the CTCs isolation methods, which will allow the maximisation of CTCs detection and thereby their further characterisation. In this context, we validated a new microfluidic technology called the CROSS chip. This system was tested and compared with CellSearch using as a proof of concept a cohort of metastatic colorectal cancer patients. Lastly, cells isolated using the CROSS chip device were screened by ddPCR for the presence of a specific mutation of the *APC* gene, highly frequent in colorectal cancer patients, to confirm their malignant origin and to validate the capability of downstream molecular characterisation with this system. Considering the results obtained in this comparative study with a small metastatic colorectal cancer cohort, due to the higher sensitivity of the CROSS chip, we suggested a higher cut off value than CellSearch for bad prognosis ($\geq$ 7 CTCs/7.5 ml of whole blood). This new cut off allowed the stratification of patients in 2 defined populations with OS differences. However, further studies on larger cohorts of patients, including different tumour types, are required to clarify the clinical relevance of this method for metastatic patients monitoring and characterisation.

To summarize, liquid biopsy offers a significant opportunity in tumours that are not easy to biopsy and for the restaging and molecular analysis of metastasis. In addition, liquid biopsy diagnosis

can serve as a real-time monitoring of tumour status that could tailor the therapy to the individual need of the cancer patient. In this sense, our results prove that CTCs analysis can provide clinically important information, further supporting the relevance of liquid biopsy monitoring as a valuable tool for understanding tumour evolution. Nevertheless, interpretation of the clinical results might be hampered by the fact that the dynamic biology of CTC is still widely unknown. Our studies highlight the need of CTCs characterisation besides enumeration to provide a more accurate and personalised medicine to the patients. In this context, we tried different approaches (size isolation, magnetic isolation, negative enrichment and the establishment of preclinical models) that allowed us to make an approximation to CTCs from different sides, improving the yield of results obtained so far with other technologies and identifying cell markers that could be translated into the clinical practice.

INTRODUCTION

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

INTRODUCTION

1. LIQUID BIOPSY

The term "liquid biopsy" was adopted in the oncology field on 2013[1], to refer to the detection of tumour cells, cell-free nucleic acids and exosomes in peripheral blood and other body fluids from cancer patients. It is considered one of the most advanced minimal invasive diagnostic systems that enables clinically relevant actions and the potential implementation of precision medicine[2].

Until recently, liquid biopsy applications had focused mainly on circulating tumour cells (CTCs), but nowadays the perspective has been broadened to circulating tumour DNA (ctDNA), microRNAs (miRNAs) and cancer associated exosomes[2,3]. This wider view of liquid biopsy provides new potential applications for the development of multi-marker diagnostic, prognostic and therapeutic signatures[4].

The main advantage of liquid biopsy is that it allows a repetitive, real-time questioning of the dynamic evolution of tumours; avoiding some key limitations of conventional tissue biopsies like invasive tumour sampling, under-representation of tumour heterogeneity and poor description of clonal evolution during therapy resistance and metastatic dissemination[2]. Thus, the information gathered through liquid biopsy allows clinicians to complement standard tissue biopsies analysis so they can adopt relevant medical actions such as early diagnosis, staging, prognosis, anticipation and prediction of therapy responses during the follow-up of the disease, enabling the implementation of the so-called precision medicine.

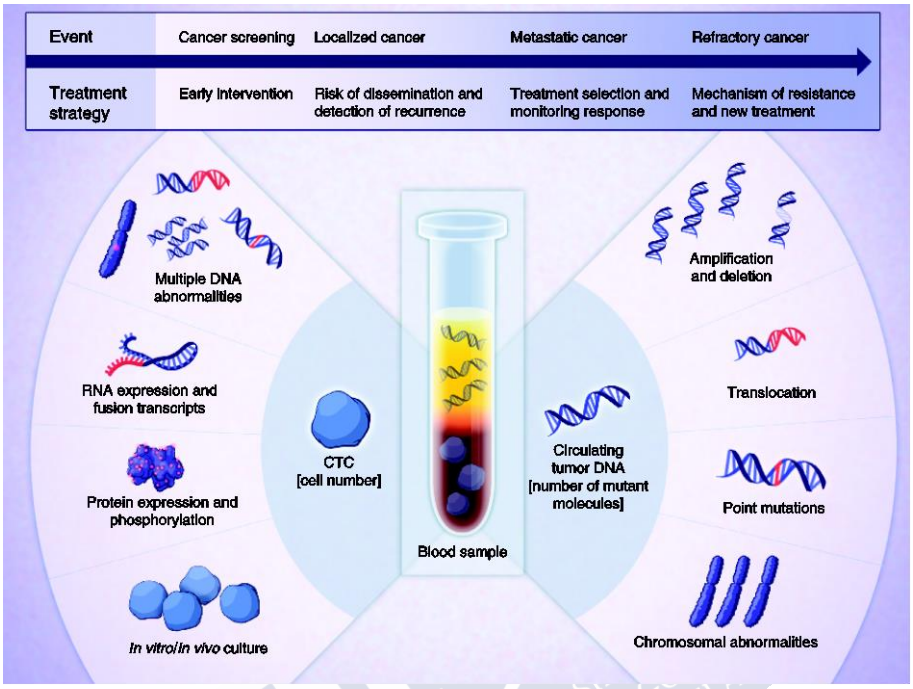

Figure 1. Clinical applications of liquid biopsy in blood for cancer care (*Haber and Velculescu, 2014*, with permission of the American Association for Cancer Research[5]).

1.1. Circulating tumour cells

CTCs were detected for the first time in 1869 when Thomas Ashworth described cells in the blood that appeared similar to those observed in the tumour while carrying out the autopsy of a patient with widespread breast cancer[6]. However, technical challenges posed by CTCs detection have been limiting progress until recently[5].

CTCs are tumour cells that are released into the blood from the primary tumour and/or metastatic sites. The mechanisms by which these cells are released into the bloodstream are still a matter of controversy within the field, and several mechanisms with different amounts of supporting evidence have been proposed[7]. The propagation of tumour cells starts early with preneoplastic lesions,

sometimes even before the formation of apparent primary tumours[8,9]. This may involve both active invasion of cells with increased migratory potential, as a result of epithelial-to-mesenchymal transition (EMT), and also passive shedding of individual cells or tumour cell clusters resulting from compromised tumour vasculature. Furthermore, the intravasation of CTCs to distant organ sites can also be promoted by their association with activated platelets through the formation of heteroaggregates that could promote their arrest by the endothelium, contributing to the metastasis formation[10,11]. In addition, studies with mice models have pointed that this migration of metastatic cells in circulation might also be dependent upon gradients of chemokines that could direct tumour cells through the vasculature[12].

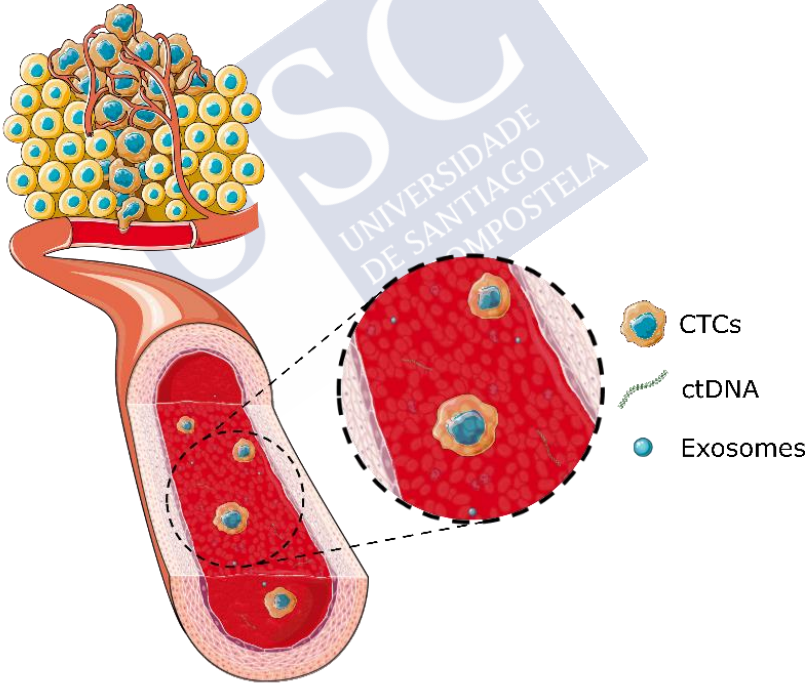

Figure 2. Simplified image of tumour cells entering the bloodstream and CTCs, ctDNA and exosomes travelling in the bloodstream.

In any case, the bloodstream is a harsh environment for epithelial tumour cells. As a result, the life-span of CTCs in the blood is described to be short, between one and two and a half hours[13], and the clearance of surviving CTCs occurs through their extravasation into secondary organs. Nevertheless, it is possible to detect CTCs in some of these patients months or years after primary tumour resection, supporting the hypothesis of the recirculation of CTCs from secondary metastatic sites into the blood[14].

1.1.1. Characteristics and challenges of CTCs

The biology of CTCs holds the key for understanding and targeting the process of blood-borne metastasis, and also, CTCs can be a surrogate marker of the tumour status that allows early-detection and applications in diagnosis, treatment, monitoring and prognosis[5].

CTCs can be approached at single-cell or at bulk-cell level, allowing the acquisition of valuable morphologic information through imaging analysis; and the evaluation of cell functional status and genetic alterations through functional and molecular studies. However, CTCs are extremely rare among the abundance of normal blood cells. As a matter of fact, in most patients they can be as few as 1 to 10 cells per 10 mL of whole blood, which englobes billions of blood cells[12]. This scarcity makes their study especially challenging. In addition, the inherent heterogeneity of the tumours gives rise to CTCs with distinct morphological and phenotypic features. Besides, as only a small percentage of CTCs have developed mechanism to avoid anoikis (apoptosis triggered by lack of correct cell–extracellular matrix attachment), most of them are at different stages of apoptosis[15]. Therefore, one of the major challenges in the field is the identification of the subset of CTCs capable of initiating a metastatic lesion and their contribution to the metastatic process itself[5].

Over the last 10 years, and in most of the current assays still in use, the standard detection of CTCs has been through epithelial markers like the Epithelial Cell Adhesion Molecule (EpCAM) and cytokeratins (CKs), since these markers are not expressed on the surrounding blood cells neither on the endothelial cells[12]. Reports have defined cells of epithelial origin in blood as morphologically

heterogeneous, nucleus positive, CD45 negative and CK positive[16]. These circulating epithelial cells can be detected in patients with metastatic and organ confined tumours, whereas only few of them are observed in healthy controls[17].

However, epithelial tumour cells can undergo EMT that results in a reduced expression of epithelial markers and an induction of a more mesenchymal phenotype. This EMT process also leads to an increased cell plasticity and capacity for migration and invasion, as well as a resistance to the before mentioned anoikis, which are attributes required for CTCs survival and dissemination. The present view, based on evidence published recently, is that CTCs might have an intermediate and dynamic EMT phenotype, co-expressing both epithelial and mesenchymal markers, and this phenotype might have the highest plasticity to adapt to the conditions present in secondary sites[18].

Besides differences on biological properties, CTCs also differ from hematopoietic cells on their morphology and their deformability. Epithelial cancer cells are reported to be larger (diameter range from 12-25 μm) than leukocytes (5-10 μm), however, CTCs of various sizes have been identified[19]. Regarding the cellular deformability, tumour cells are stiffer than hematopoietic cells but CTCs that are capable of undergoing EMT might also be as deformable as leukocytes[12,19].

To summarise, this limited understanding of both physical and biological properties of CTCs is preventing us from developing a universal system for CTCs detection and analysis, hampering its translation into the clinical practice.

1.1.2. CTCs clusters

CTCs clusters are aggregations that can range from two cells to large microemboli with more than 50 cells. CTCs clusters are rare events found in the circulation of patients with tumours of different origins, and are described to have 23-50 fold increased metastatic potential when compared with single CTCs[20].

Aceto and colleagues demonstrated in 2014 that CTCs clusters arise not from intravascular aggregation but from clumps of tumour cells with oligoclonal origin. However, just as it happens with single

CTCs they were unable to determine if the intravasation into the blood was the result of an active invasive process or, on the contrary, it was the result of passive shedding. In this study, plakoglobin, a cell junction component, was found highly differentially expressed in clusters compared with single cells in breast cancer samples, suggesting its role in the cohesion of the CTCs clusters in circulation, contributing to their metastatic capability[20]. Recently, it has also been described that CTCs clusters are able to shape the DNA methylome, promoting stemness and metastasis[21].

1.1.3. CTCs isolation methods

The first isolation methods of CTCs implicated the use of a manual or an automatic micromanipulator after their detection by immunocytochemistry or immunofluorescent staining. Nevertheless, in the last few years, a lot of different isolation strategies have emerged. All these strategies share the challenge of sorting the few CTCs present in the sample without damaging or losing them, being able to purify the CTCs efficiently but without contamination with leukocytes, and finally, correctly identifying CTCs based on unique immunophenotypes, cytopathologic or molecular genetic features[22].

The technological development for the recovery of CTCs is the bottleneck step that has been hampering the implementation of CTCs analyses into the clinical practice, due to the ignorance of the phenotypic or intracytoplasmatic characteristics of CTCs and the extremely low abundance of these cells in the blood. The purpose is to be able to recover large representative cancer cells populations so they can be identified, enumerated and molecular characterised. And, even though the field has undergone a big explosion of techniques, there is neither a gold-standard technique nor a single approach that allows the recovery of the total amount of CTCs present in a sample or that isolates those CTCs at a single step[23].

Most of the CTCs assays start with a first step of enrichment of the sample that increases the concentration of CTCs in relation with blood cells. These assays are usually performed from a tube of 7.5 mL of blood, so they rely on the frequency of CTCs that can be found on

this limited volume of blood. Recently, a new approach to overcome this issue has been proposed, the "diagnostic leukapheresis", which consists on the isolation by apheresis of the mononuclear cell fraction, which is believed to contain the majority of CTCs[24,25]. However, the use of this technique is still limited by the lack of technology able to process such a high amount of cells.

After a first enrichment step of the sample, CTCs can be detected by different approaches. CTCs isolation techniques fall broadly within three different classes, depending on their physical properties, their biological properties or a combination of both. There are innumerable technological approaches within these categories, at different stages of development, from "proof of concept" using spiked cancer cell lines into the blood, to more advanced testing with blood specimens from patients with different types of cancer[5,22,23].

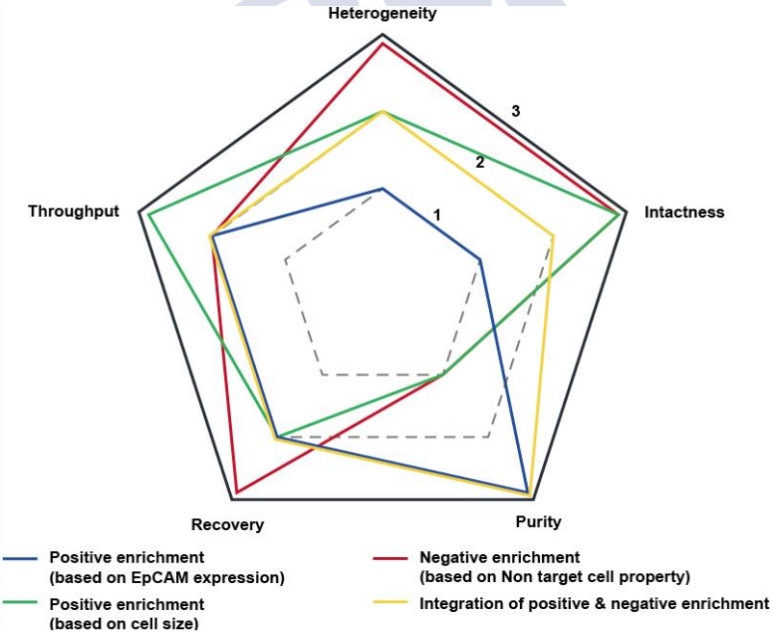

Figure 3. Comparison of different methods for CTCs isolation, taking into account five performance categories: heterogeneity, intactness and purity of the isolated cells, and recovery rate and throughput of the technology. Scale: 1-3, where 3 represents the highest score (Gwak and colleagues, 2018, with permission of Creative Commons[26]).

1.1.3.1. Isolation methods based on physical properties

Among the isolation techniques based on physical properties, the most common are: microfiltration (based on different size), microfluidics (based on deformability and size), dielectrophoresis (based on electrical charges) and density gradient (based on density)[5,22].

Size-based filtering approaches had been explored since the 1960s and they take advantage of the size difference between CTCs and hematopoietic cells. Size-based methods may be particularly useful in cancer types associated with larger tumour cells (like large cell lung carcinoma), as they can provide a relatively simple way to assess CTCs burden[5,19].

The first sized-based test developed for CTCs recovery was the **ISET®** technology (ISET: Isolation by SizE of Tumour cells), by Rarecells. It allows direct filtration of peripheral blood and the isolation of CTCs by sieving the sample through vertical filtration with a calibrated membrane with 8 μm diameter cylindrical pores[27].

Currently, the most extended technology for CTCs isolation through microfluidics is the **Parsortix Cell Separation System**, a semi-automated system developed by ANGLE, capable of capturing and harvesting rare cells from body fluids such as blood, urine, bone marrow or ascites. The Parsortix Cell Separation System isolation principle is based both on the size and the deformability of the cells, enabling the system to capture different rare cell types like both epithelial and mesenchymal cancer cell phenotypes[28,29].

Other size-based isolation technologies that have been commercialised are **ScreenCell**, **Vortex** and **VyCAP microsieve**. In addition, VyCAP integrates an inverted microscope that allows the *in situ* visualisation of the cells and the recovery of CTCs at single cell level.

Sized-based isolation methods have the main advantage of easy use and little manipulation of the samples, which is crucial for avoiding the loss of cells. On the other hand, filtering large volumes of

cells through a static filter creates significant hemodynamic stress on the cells, which may have an impact on their integrity. Furthermore, measurements of CTCs isolated using other parameters reveal considerable heterogeneity in the size of CTCs, even on those derived from an individual patient. Therefore, smaller CTCs[30] or tumour cell fragments may not be detected using these methodologies. Additionally, CTCs that have undergone EMT are smaller and more deformable and they can squeeze through narrow constrictions resulting in their loss during the recovery. In any case, the main problem of size-based methods relies on the purity of the recovered sample, due to their overlap in size between CTCs and hematopoietic cells[19,22,31].

Besides size and deformability, CTCs can be isolated based on their electrophoretic properties. **DEPArray™ System** (Menarini-Silicon Biosystems) allows a single cell level isolation by dielectrophoresis. This technology is described to be able to recover up to 80% of cancer cells avoiding lymphocyte contamination but it has a limitation in the number of cells that can be recovered and it needs a first enrichment step[22,32].

Finally, CTCs can also be isolated from whole blood based on their density properties with the use of a density gradient medium. The recovered sample can be placed on microscopy slides for pathology analysis[33]. However, this method is highly unspecific as CTCs are found in their majority in the PBMCs layer, so this technique is used mainly as a first enrichment step rather than an isolation method itself.

1.1.3.2. Isolation methods based on biological properties

Based on their biologic properties, like the expression of surface protein markers, CTCs can be positively or negatively enriched. Inside this category most of the methods are based on immunoaffinity. However, there is a lack of a constitutive membrane antigen or a specific phenotype. For many years EpCAM has been used as a CTCs panmarker and many technologies have relied on this antigen to quantify CTCs. However, with this approach EpCAM- or EpCAM low CTCs are lost. Moreover, EpCAM expression has been detected

in patients with benign breast conditions, which could lead to potential false positive cases[34]. In addition, negative enrichment can be performed by targeting blood cells with the blood marker CD45, however, it was demonstrated that a circulating CD45 negative population exists also in peripheral blood of healthy donors, so this phenotypic characteristic may not be exclusive of CTCs[35].

The **CellSearch System** (Menarini-Silicon Biosystems) is the only commercial technology approved by the Food and Drug Administration (FDA) for prognostic purposes. It was cleared for breast and colorectal cancer on 2004[34,36], and for prostate cancer in 2008[37]. CellSearch, was designed for the enumeration of CTCs from 7.5 mL of blood. It makes use of magnetically tagged antibodies (ferrofluids) against EpCAM. EpCAM positive cells are then separated with the use of a magnetic field and incubated with antibodies that allow their identification by their expression of CKs (CK8/18/19) and the absence of CD45[16].

However, CellSearch, as it has been designed to select CTCs by their EpCAM expression, enriches the isolated sample only in epithelial CTCs, losing the more mesenchymal or stem phenotypes, being the reason why many groups are exploring techniques based on other approaches.

For instance, **Adnatest** (Qiagen) allows the separation of CTCs with a combination of different antibodies conjugated to magnetic beads. CTCs are then analysed via quantitative PCR (qPCR) assay for tumour-associated transcripts[38].

Other strategies include the selection of CD45 negative cells, by depleting CD45 positive cells, preceded or not by a red blood cell lysis. This technique based on specific antibodies that can be coupled to magnetic beads, can be performed with different technologies such as **MACsSystem** (Miltenyi), **Dynabeads** (Thermo Fisher) or **EasySep** (STEMCELL Technologies). The rationale behind this strategy is that leukocyte cell markers are well characterised and remain invariant, while cancer cells may express multiple and different markers[5,22].

Another approach is **high-throughput microscopic scanning** of blood samples depleted of red blood cells and plated onto cytospin slides. This design is unbiased by cell size in the initial selection of CTCs but it depends on the detection of tumour markers to identify CTCs. Besides, molecular characterisation of unenriched cancer cells within blood populations is remarkably challenging[5].

1.1.3.3. Isolation methods based on a combination of physical and biological properties

It has also been proposed that the combination of methods based on different properties can be helpful for positive or negative CTCs enrichment as each technique has its limitations and none is robust enough to be considered the best one. Some groups have focused on comparing and combining different isolation technologies in order to find the most suitable for each cancer type and the following downstream analysis[33,39,40].

The **IsoFlux System** (Fluxion Biosciences Inc) combines the use of immunomagnetic beads that can be coupled to the antibody of choice by the user, with a microfluidic device equipped with a magnetic field. This system allows the recovery of the target cells either on buffer lysis for downstream molecular analysis, or on a microscope slide for pathology studies[41].

Another example of this combination of approaches is **NanoVelcro chip**. It is a microfluidic chip combined with cell-affinity substrates, in which CTCs are immobilised with agent-coated nanostructured substrates. NanoVelcro last generation of thermo-responsive chips allows CTCs release and it is also able to purify CTCs with well-preserved RNA transcripts. So far, it has been proven on spiked blood samples of lung cancer[22,42].

Another strategy is the use of **RosetteSep™ System** (STEMCELL Technologies), an antibody cocktail that crosslinks the hematopoietic cells, in combination with a density gradient centrifugation. This technique allows the negative enrichment of a viable CTCs fraction that can be used for *in vivo* or *ex vivo* studies or for downstream analyses[43].

1.1.4. CTCs analysis

The study of CTCs offers a mine of information on therapeutic targets and resistance to therapy in cancer patients. The most basic information we can get from CTCs is enumeration but they can also be approached with genomics, transcriptomics and proteomics[44].

1.1.4.1. CTCs analysis based on enumeration

For the **enumeration of CTCs**, traditional criteria have been established through immunocytochemistry, by their positive expression of CKs and their absence of CD45[17]. CTCs enumeration has been proved to have prognosis value by different studies from breast, prostate and colorectal cancer[37,45]. In addition, the follow-up of patients through CTCs enumeration provides further information to standard imaging studies, to identify responding and non-responding patients[46]. Other promising stains that have been published include Ki67/PSA and PSA/PSMA in prostate cancer patients, and the ER/BCL2/HER2 in breast cancer patients[47,48]. The field is moving towards the use of cancer type-specific panels, which will provide valuable information for the monitoring of the tumour status and to guide therapeutic decisions.

One of the critical issues to establish the clinical value of CTCs enumeration was the selection of an appropriate threshold value. Since Cristofanilli and colleagues published in 2004 their landmark article on the CTCs enumeration in metastatic breast cancer (MBC), a cut off of 5 CTCs per 7.5 mL of blood was established for poor prognosis[34]. Furthermore, in this study they demonstrated that CTCs counts before treatment were an independent predictor of progression-free survival (PFS) and overall survival (OS) in MBC patients[34]. Yet, in the light of the results from the clinical trial SWOG0500, Bidard and colleagues recommend that the thresholds for clinical validity should be distinguished from those intended for clinical utility, as they do not always can be translated into the clinical practice. Thus, CTCs enumeration is currently not advised in any of the American Society of Clinical Oncology (ASCO) or European Society for Medical Oncology (ESMO) guidelines for any clinical decision in any tumour

type[49,50]. However, efforts are being made to unify and standardise reproducible quantification of therapeutic marker expression across different analysis platforms, like the ACCEPT Software, which is an image analysis package for the automated CTC classification, enumeration and phenotyping[51].

1.1.4.2. CTCs analysis based on their molecular characterization

Regarding **transcriptomics**, the studies from Antonarakis and colleagues, on prostate cancer, demonstrated the clinical value of the mRNA analysis of CTCs, gathering information about drug sensitivity and resistance, through the mRNA expression of ARv7 (Androgen receptor variant 7)[52,53]. miRNAs have also emerged as diagnostic markers and targets for cancer treatment[54]. Gasch and colleagues described a protocol combining *in situ* hybridisation with the CellSearch system, enabling clinical research of the heterogeneity of miRNAs between different CTCs in patients with breast, prostate, or colorectal cancer[55]. However, RNA-based expression studies in CTCs have the drawback that, except for EDTA, most of the preservatives used in peripheral blood collection tubes interfere with the analysis. Besides, the time to perform the analysis is also crucial and the samples must be processed within two hours after blood acquisition[56].

The **genomic** approach allows the research of mutations in genes encoding therapeutic targets and signalling proteins downstream of the target that can affect the efficacy of targeted drugs. The most relevant mutations have been the ones found on *EGFR* in lung cancer and *KRAS* in colorectal cancer due to their implication in the resistance to anti-EGFR therapies. Also *PIK3CA* mutations in the case of breast cancer for its relation to the resistance to HER2-targeting therapies, and the alterations in the *AR* gene that can result in cells that are refractory to androgen blockade in prostate cancer. Likewise, in the case of melanomas, *BRAF* mutations are important predictors of sensitivity to BRAF-directed therapies. Besides, tumour-specific translocations, like *EML4–ALK* in non–small cell lung cancer and *TMPRSS2–ERG* in prostate cancer, are also being studied. From a

clinical point of view, this approach is likely to be one of the most immediate applications[5,55].

At a **protein level**, several studies have dug into proteins with potential therapeutic interest like the ER and PR receptors and HER2 oncogene, which are key targets in breast cancer; PSA and PSMA, which play a role in the AR regulation on prostate cancer; the proliferation maker Ki67; the immune checkpoint regulator PDL1 and apoptosis and DNA-repair related proteins[49,55]. Paoletti and colleagues have developed a multiparameter CTC-Endocrine Therapy Index that combines the enumeration of CTCs and the expression of ER, BCL2, HER2, and Ki67. This index, which is being evaluated on an ongoing prospective clinical trial, may predict resistance to endocrine therapy in patients with HER2-positive MBC[48].

Overall, CTCs analysis could potentially provide really valuable insight and great depth of knowledge by allowing the examination of the complete cell, the RNA, the detection of diagnostic proteins, as well as DNA-based genotyping. Most relevant, with the evolution and refining of single-cell technologies, CTCs analyses will permit precise, real-time measurements of cancer heterogeneity and the subclonal populations resulted from the selective pressure caused by the different treatments. Nevertheless, CTCs studies will only become of extensive use when the new technologies currently in ongoing development and testing, achieve commercial and broad availability for the cancer research and clinical community[5]. In addition, in order to implement the study of CTCs into the clinical routine, standardisation of pre-analytical conditions and protocols must be established[56].

1.1.5. Clinical relevance of the study of CTCs

CTCs quantification and characterisation could be translated into the clinical practice through real-time monitoring, stratification of patients, and therapy switch based on CTCs count and CTCs biomarker expression.

CTCs have shown to be of prognostic significance in patients with different solid tumours. Thus, evidence shows that the

determination of CTCs counts before or after initial surgery in non-metastatic patients is a reliable indicator of an unfavourable prognosis, furthermore, in primary breast cancer, the detection of CTCs, both before and after adjuvant chemotherapy, was linked to an increased risk of relapse[57]. Further studies have demonstrated significant correlations between CTCs counts and metastatic relapse in other tumours like oesophageal, colorectal, liver and bladder cancer[45,58–60].

The quantification and characterisation of CTCs have also allowed stratification and therapeutic intervention based on liquid biopsy. Several phase III clinical trials are exploring this approach, studying how CTCs can contribute to an early assessment of therapy effects. These clinical trials are being performed mostly in breast cancer patients.

The multicenter SUCCESS study has explored the relevance of CTCs at the time of primary diagnosis regarding their prognostic relevance to follow-up care. They found that the presence of CTCs two years after chemotherapy was associated with decreased OS and disease free survival (DFS). Based on their results, surveillance strategies for breast cancer survivors based on CTCs biomarkers could anticipate tumour relapses[61–63].

The previously mentioned SWOG0500 clinical trial, on MBC patients treated with first line chemotherapy, concluded that CTCs counts have prognostic significance but an early switch to a different cytotoxic therapy was not effective in prolonging OS in patients with persistently elevated CTCs after one cycle of chemotherapy, suggesting the need for more effective treatments than standard chemotherapy in this population[55]. However, Georgoulias and colleagues, in a randomised phase II study in patients with early breast cancer, indicated that trastuzumab decreases the incidence of clinical relapses in patients with chemotherapy-resistant CK19 mRNA-positive CTCs[64].

The STIC CTC METABREAST is another ongoing clinical trial studying the value of baseline CTCs in luminal MBC to determine first-line treatment, in order to stratify the patients to chemotherapy or

hormonal therapy. In this trial, breast cancer patients with more than 5 CTCs counts in 7.5 mL blood receive chemotherapy, while patients with 5 or less CTCs receive endocrine therapy as the first-line treatment[65,66].

CTCs are also investigated as a surrogate for tumour biology, based on *HER2* expression and amplification. The DETECT III study is focused on patients with MBC with up to three chemotherapy lines; all the patients must be HER2 negative by traditional biopsy but have at least one HER2 positive CTC in 7.5 mL of blood. Patients are then randomised between standard therapy (chemotherapy or endocrine therapy) and standard therapy plus lapatinib, an EGFR/HER2 tyrosine kinase inhibitor. Similarly, the CirCe01 trial explores a CTCs-based management of chemotherapy in advance MBC patients. CirCe01 study uses the HER2/CEP17 ratio measurement by fluorescence *in situ* hybridisation for HER2 amplification assessment. Patients with HER2 amplification on CTCs receive chemotherapy in combination with an anti-HER2 drug[55,61].

Finally, besides all the ongoing clinical trials on breast cancer, the study VISNU-1 analyses the value of first-line triplet chemotherapy (FOLFOXIRI-bevacizumab) versus doublet chemotherapy (FOLFOX-bevacizumab) in metastatic colorectal cancer patients with 3 or more CTCs at baseline[61].

In early stages of the disease, a meta-analysis to assess the clinical validity of CTCs detection as a prognostic marker was recently published by Bidard and colleagues in non-MBC patients treated by neoadjuvant chemotherapy. This study revealed that CTCs counting is an independent and quantitative prognostic factor in these patients and it could complement the current prognostic models based on tumour characteristics and response to therapy[67].

1.2. Cell-free circulating tumour nucleic acids

In the last few years, term liquid biopsy has also been extended to the analysis of cell-free circulating tumour nucleic acids.

CtDNA is constituted from fragments of DNA derived from primary tumours, lysed CTCs, micrometastases or overt metastases into the blood. Non-malignant host cells also release cell free DNA (cfDNA), which dilutes the ctDNA in patients with cancer, especially on certain conditions including inflammation, exercise or tissue injury. The abundance of ctDNA in the cfDNA fraction in patients with early stage tumours is described to be approximately 10-fold lower than in patients with more advanced disease. The range in the ctDNA levels is not well understood and it is thought to be affected by tumour burden, stage, cellular turnover, accessibility to the circulation and factors affecting blood volume[68].

Studying ctDNA is technically challenging, not only for its low abundance among variable amounts of cfDNA, but also because it is typically fragmented to 160 to 180 bp in length, corresponding to nucleosome-protected DNA observed in apoptotic cells, and it has a short half-life of 16 minutes[69]. Besides sensitivity, the specificity in the identification of clinical valuable mutations is also an issue, as cancer-associated mutations occur with increasing age even in individuals who never develop cancer during their lifetime[70].

Due to these challenges, the analysis of circulating tumour nucleic acids has been mostly focused on ctDNA mutations; nevertheless, other molecular targets have been explored. Tumour-specific methylation analysis can also be used as markers of ctDNA presence, being the most frequent alterations DNA methylation at specific promoter regions and specific DNA hypermethylation of tumour suppressors. Other molecular targets present in the blood are circulating free miRNAs. miRNAs are abundant in several extracellular body fluids, where they are protected and stabilised by exosome-like structures and small intraluminal vesicles that can be produced by different cells, included cancer cells[2].

CtDNA allows the study of different somatic alterations (point mutations and structural alterations like copy-number changes and chromosome rearrangements) that are directly derived from an individual tumour. CtDNA analyses have its translation into the clinic through real time monitoring of patients, to track clonal evolution and

targeted drug responses, and the stratification of patients based of their mutational status. Thus, ctDNA mutations can be used to identify potentially actionable changes affecting driver genes, such as *EGFR*, *KRAS*, *NRAS* and *BRAF* (to detect anti-EGFR acquired resistance); to select targeted therapies like *PIK3CA* (for monitoring paclitaxel resistance in breast cancer); and to detect residual disease or monitor tumour levels during therapy[5,71,72]. Recently, the first ctDNA test for *EGFR* mutations in non-small cell lung cancer has been approved. This test allows the stratification of patients based on their *EGFR* status and it is an important step toward clinical implementation of liquid biopsy[73]. Despite these advances, an exciting challenge will be the shift from analysis of patients in advanced stages with high loads of ctDNA to early-stage patients who are treated with curative intent[55].

For ctDNA analysis, the gold standard technologies are qPCR and digital PCR (dPCR); however, other technologies have been proposed, such as PCR-single strand conformation polymorphism (PCR-SSCP), multiplex dPCR, allele-specific qPCR, whole genome sequencing (WGS), cancer personalised profiling deep sequencing (Capp-Seq), methylation-specific PCR (MS-PCR), the Discrimination of Rare EpiAlleles by Melt qPCR (DREAMing), bidirectional pyrophosphorolysis-activated polymerisation (bi-PAP) and tagged-amplicon deep sequencing (TAm-Seq)[2,3].

For the detection and analysis of point mutations as biomarkers in ctDNA, highly sensitive and specific methods have been developed. The common strategy is to quantify the number of reactions containing wild-type or mutant PCR product, which can be achieved by dPCR analysis. Within dPCR techniques we can find BEAMing, which combines water-in-oil emulsion PCR with magnetic beads to allow single-molecule PCR reactions that then can be analysed using flow cytometry. Other technologies recently developed for this kind of analyses are droplet digital PCR (ddPCR) and micro-fluidic systems for parallel PCR reactions (Fluidigm). To establish copy-number changes and chromosome rearrangements WGS is the best available option[2,5,74].

1.3. CTCs versus ctDNA

In the near future, as clinical decisions become increasingly dependent on real-time monitoring of tumour status, CTCs and ctDNA technologies are likely to be synergistic.

Both CTCs and ctDNA analyses reflect biologically different aspects of the disease, becoming essential components of cancer management due to their capacity to capture the heterogeneity across tumour sites and the evolution of tumour cells and mutations[75].

ctDNA analyses provide a global picture of the genetic status of the disease while CTCs allow a molecular analysis of the tumour both at bulk or single cell level. ctDNA could be used to monitor cancer patients during treatment or remission, gathering real-time molecular information to monitor treatment response and relapse, with a higher sensitivity than CTCs. However, CTCs, as the responsible entities of metastasis formation, hold invaluable information about the intrinsic biology of the tumours and their dominant clones. Thus, they could be used to test and to guide drug therapy once there is evidence of therapeutic failure or disease recurrence[76].

However, the information obtained from these two liquid biopsy biomarkers, CTCs and ctDNA, is different, complementary, and depends on the context of use. Hopefully, it will complement tissue biopsies as diagnostic procedures and make cancer treatment more precise[55,75].

2. PRECLINICAL MODELS FOR THE STUDY OF CTCS

CTCs based preclinical models have a great potential for basic and preclinical cancer research as they provide us with information of the metastatic dissemination process. Thus, they may constitute more accurate and sustainable disease models than the previously introduced adherent membrane cultures and they can be used to form CTC-derived tumours[77]. Although tumour cell lines studies have allowed the gaining of knowledge in cancer research, cancer cell lines have some limitations as they do not always recapitulate closely the

studied disease[78]. In addition, the tendency towards precision medicine has resulted in an increased interest in adapting *in vitro* tumour models for patient-specific therapies, clinical management, and assessment of metastatic potential[79].

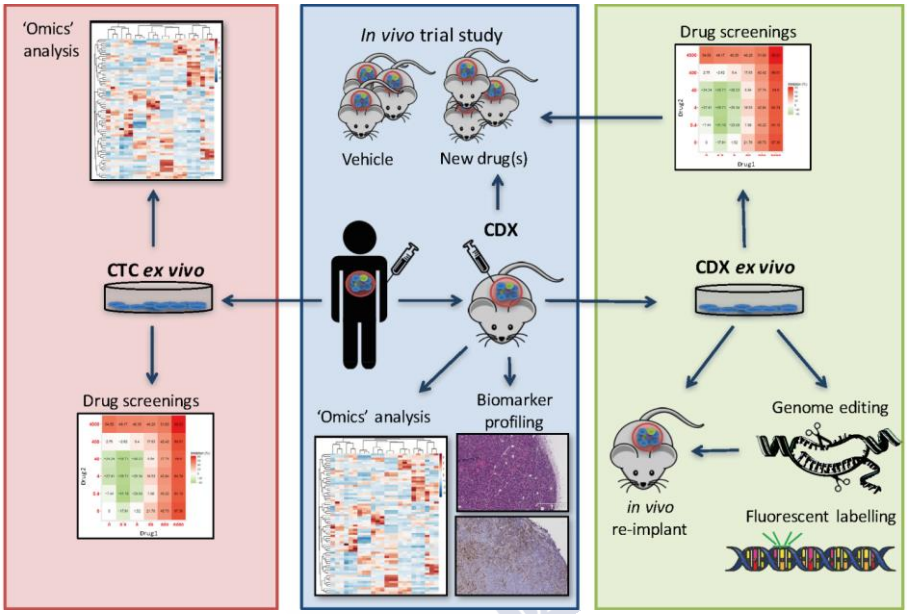

Figure 4. Diagram of the current preclinical use of CTCs (Lallo and colleagues, 2017, with permission of Pioneer Bioscience Publishing Company).

2.1. *In vitro* models: CTCs cell culture

The *in vitro* culture and maintenance of CTCs represents a great tool for patient personalised treatment that allows the testing of their tumorigenic properties, as well as their sensibility to different drugs. Nevertheless, it is a complicated task because in addition to their low numbers, many CTCs have limited proliferation ability and they become non-viable after a few cell divisions.

Various techniques have been used in order to optimise CTCs culture proliferation, including hypoxic and non-adherent conditions.

So far, three groups have reported patient-derived CTCs long-term cultures and several more have described *in vitro* short-term growth. The first long-term CTCs culture was established from a metastatic luminal subtype breast cancer. CTCs were maintained for more than six months, under serum free and non-adherent culture conditions. This study enabled the discovery of a potential signature in CTCs competent for brain metastasis[80,81]. Another report described the first CTCs-derived permanent cells line isolated from the blood of a colorectal cancer patient whose CTCs have been cultured for more than 1 year. This study was able to obtain several cell lines from different time points of the disease progression of this patient[82]. Finally, a third study by Gao and colleagues described the generation of a long term prostate cancer organoid cell line from a patient with a high count of CTCs (>100 CTCs/8 mL of blood)[83]. These established cell lines are invaluable tools for the functional research of the biology of CTCs. Thus, they report that cultured CTCs were very similar to captured CTCs; however they share a very low success rate (less than a 8% and a 2%, respectively on the two first studies), and the requirement of high CTCs counts, that can only be found on patients with a high tumour burden[81,82]. In addition, short-term CTCs cultures have been described from blood from head and neck, breast, prostate and gastric cancer patients[84–87].

Finally, two recent studies have reported the use of microfluidic technologies to establish CTCs cultures. One of this studies described the obtaining of short-term CTCs cultures from blood samples of early stage lung cancer patients. In order to facilitate CTCs expansion, the authors introduced tumour associated fibroblasts as a 3D co-culture condition to reproduce tumour microenvironment[88]. Another study reports the use of a microfluidics-based culture approach, with blood cells as a co-culture, to develop CTCs clusters, from patients with locally advanced cancer, which then can be used for drug screening[85].

So far, it has been challenging to implement the methodologies described in these studies into routine clinical procedures due to the low efficiency rate of the methods described and the prolonged periods required for cell line establishment. However, with further

development, all of these methodologies have the potential use for patient-specific drug susceptibility testing and mutational cancer profiles studies[23].

2.2. *In vivo* models: CTCs patient derived xenografts

In vivo models have a great potential for basic and preclinical cancer research directed to the discovery of biomarkers, the understanding of drug resistance mechanisms and the development of new therapies. In fact, efforts are being made for the establishment of patient-derived xenograft (PDX) as a useful *in vivo* model system to study the biology of human tumours and metastases. The main interest in PDX-derived tumours comes from the resemblance of their morphologies, architectures and molecular signatures with those of the original tumours[89].

Usually, both tumour molecular characterisation and PDX generation are based on tissue biopsies from the primary tumour. However, several reports described that CTCs have the potential to generate preclinical models matched to individual patient's tumours on CTCs-derived xenografts (CDX) in immunosuppressed mice, demonstrating the tumorigenic potential and the feasibility to expand these cells *in vivo*.

Pretlow and colleagues were the first to report the formation of xenografts from carcinoma cells taken directly from the peripheral blood of patients. They detected lung metastasis in 15% of the nude mice which were previously injected with peripheral blood samples depleted from red blood cells and plasma, from metastatic prostate and colorectal cancer patients[90].

Bacelli and colleagues, in a study with 110 luminal breast cancer patients, described the need of at least 1000 CTCs enumerated by CellSearch in order to establish a CDX. In their study, mice receiving at least 1,109 CTCs developed multiple bone, lung and liver metastases but no primary tumour. Characterisation of these CDX allowed them to report the existence and phenotype of metastasis-

initiating cells (MICs) among CTCs. This MICs population had *EpCAM*, *CD44*, *CD47* and *MET* expression[91].

Rossi and colleagues developed a xenograft assay in NOD/SCID mice with EpCAM+ cells present in peripheral blood from metastatic prostate (n=6) and breast (n=2) cancer patients. None of the CTCs-injected mice developed clinical evidence of tumour neither at the injection, nor at secondary sites, but they were able to recover human CTCs from the peripheral blood of all mice. This allowed them to describe that the EpCAM+ fraction of CTCs retains migratory capacity[92].

In the previously described study of long-term luminal breast cancer CTCs culture, Yu and colleagues tested the carcinogenic capacity of the CTCs lines by injecting 20,000 cells into the mammary fat pad of immunosuppressed Non-obese diabetic Scid Gamma mice (NSG). They reported that CTC-derived tumours shared histological and immunohistochemical features with their matched primary patient's tumour[80].

Hodgkinson and colleagues injected CTCs from serial samples of patients with either chemosensitive or chemorefractory small cell lung cancer into NSG mice. CDXs were established in patients with more than 400 CTCs by CellSearch. They demonstrated that CTCs from small cell lung cancer are tumorigenic and the resultant CDXs mirrored the donor patient's response to chemotherapy. Genomic analysis of isolated CTCs also revealed similarity to the corresponding patient tumour. Thus, molecular analysis of CDXs via serial blood sampling could facilitate delivery of personalised medicine for small cell lung cancer[77,93]. In addition, another study from this group described the establishment of a CDX from a non-small-cell lung cancer dying patient with non-detectable CTCs by CellSearch. In this study they were able to represent an end of life disease model; they enriched CTCs with the RosetteSept and injected them in NSG mice[94].

Finally, Vishnoi and colleagues were able to recapitulate *in vivo* the asymptomatic progression of metastatic melanoma. To do that, NSG mice received an intracardiac injection of the peripheral blood

mononuclear cell fraction (PBMCs) depleted from CD45+,CD34+, CD73+, CD90+and CD105+ cells, from stage III/IV patients. Transcriptomic analysis of *ex vivo* bone marrow-resident tumour cells from these CDXs allowed the identification of a new molecular target, *USP7*, to prevent progression in these patients[95].

To sum up, CDXs can be complementary to tumour biopsies and be a source of tumour material for research purposes. CDXs offer an opportunity to generate models for those patients that either cannot undergo surgery or an alternative invasive procedure, or lack a primary tumour. Furthermore, CDXs can be derived from CTCs collected at different time points during patient's follow-up, allowing the establishment of paired models that recapitulate the patient's tumour evolution[55]. Nevertheless, the low efficiency and the time and cost of establishing tumours, prevents the use of animal models as a realistic way to monitor cancer progression routinely for personalised cancer management.

3. BREAST CANCER

3.1. Epidemiology, aetiology and histologic classification of breast cancer

Breast carcinoma, with more than two million new cases per year worldwide, is the leading cancer-related cause of death in women and has a higher incidence rate than any other cancer[96]. The incidence rate reaches 43.3 cases per every 100000 women worldwide, but in industrialised countries, this incidence increases up to 80-90 cases. Even though the mortality rate has been declining since 1991, it still represents 14.7% of all the cancer-related deaths[97,98], being the metastatic disease the responsible for the majority of deaths in these patients[99].

The aetiology of the vast majority of breast cancer cases is unknown. However, several risk factors have been described. These risk factors include female gender, increasing patient age, family

history of breast cancer at a young age, early menarche, late menopause, older age at first childbirth, prolonged hormone replacement therapy, previous exposure to therapeutic chest wall irradiation, benign proliferative breast disease, increased mammographic breast density and genetic mutations such as the ones found in the *BRCA1* and *BRCA2* genes. However, except for female gender and increasing patient age, these risk factors are only associated with a minority of breast tumours[98].

Breast cancer is comprised of multiple subtypes classified according to different markers, differences in morphology and their clinical implications. Classical immunohistochemistry (IHC) markers such as estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2), together with traditional clinicopathological variables including proliferation index (Ki67), tumour size, tumour grade and nodal involvement are conventionally used for patient prognosis and management[100].

Clinically, this heterogeneous disease has been classified in three basic therapeutic groups. **luminal**, or ER-positive, breast cancer subtype englobes luminal A and luminal B. Luminal A is characterised by the expression of hormonal receptors but the absence of HER2, and a low proliferation index [ER+ | PR+/- | HER2- | Ki67-], while luminal B has a high proliferation index, expression of hormonal receptors and can also express HER2 [ER+ | PR+/- | HER2+/- | Ki67+]. These ER+ patients, which represent around two thirds of the total, are considered to have a better prognosis[101]. **HER2 overexpression subtype** is characterised by the absence of hormonal receptors and the overexpression of HER2 [ER- | PR- | HER2+]. Finally, **triple-negative** breast cancer (TNBC), also known as Basal-like subtype, does not express any of the before mentioned markers [ER- | PR- | HER2-]. This subtype leads to a short disease-specific survival and poor prognosis and the recommended therapy is systemic chemotherapy[102–104]. Breast cancer is also subjected to different metastatic patters depending on the subtypes, meanwhile all the subtypes tend to spread to brain and bone, TNBC, besides the brain, tends to establish lung and distant nodal metastases[105].

Histologically, proliferative disorders in the breast are limited to the lobular and ductal epithelium, where a spectrum of proliferative abnormalities including hyperplasia, atypical hyperplasia, *in situ* carcinoma and invasive carcinoma can be found. Approximately 85% to 90% of invasive carcinomas are ductal in origin[98].

However, the trend nowadays is the implementation of the molecular analysis to complement information from ICH. A good example of this combined approach is the study conducted by Sørlie and colleagues. They reported a distinctive 'molecular portrait' of breast cancer using 456 cDNA clones that allowed them to classify tumours into five intrinsic subtypes with distinct clinical outcomes. These subtypes included luminal A, luminal B, HER2 overexpression, basal and normal-like tumours. The rationale of this classification is that the differences underlying the gene expression patterns among cancer subtypes reflect the fundamental differences of the tumours at the molecular level, revealing differences among these subtypes in terms of incidence, survival and response to treatment[106].

3.2. Staging of the breast cancer

Several cancer-staging systems are used worldwide. The most clinically useful staging system is the tumour, node and metastasis (TNM) staging system developed by the American Joint Committee on Cancer (AJCC) in collaboration with the Union for International Cancer Control (UICC), referred as the AJCC-TNM staging system[107].

This system classifies cancers by the size and extent of the primary tumour (T), involvement of regional lymph nodes (N), and the presence or absence of distant metastases (M), supplemented in recent years by evidence-based prognostic and predictive factors. Primary tumour categories range from T0 to T4 depending on the existence, the size and the extent of the tumour. Regional lymph node categories range from N0 to N3 according to the existence and extension of regional nodes invasion. And finally, distant metastases categories can be M1 or M0 whether distant metastases are present or not[97,107].

In 2010, the AJCC implemented a revision of the 7th edition of the AJCC Cancer Staging Manual, incorporating changes to the TNM staging system. These changes included a new category of M0(i+) disease referring to tumour cells microscopically detectable in bone marrow, circulating blood (such as CTCs), or found incidentally in other tissues, not exceeding 0.2 mm, in patients who have no signs or symptoms of metastasis[97,107].

Imaging studies allow assessment of the tumour's size, location, and relationship to normal anatomic structures, as well as the existence of nodal and/or distant metastatic disease. However, imaging has some limitations, such as in the detection of micrometastasis. Computed tomography and magnetic resonance imaging are the most commonly used imaging modalities, although positron emission tomography, ultrasound and plain film radiography also play relevant roles in some clinical cases[97].

3.3. Treatment of metastatic breast cancer

The selection of MBC treatment is currently based primarily on clinical and pathological factors, supplemented with hormone receptor and HER2 status. Breast cancers with different histopathological and biological features exhibit distinct behaviours that lead to diverse treatment responses and should be approached with different therapeutic strategies.

The standard treatment strategies for MBC are based in the use of chemotherapy (being the most common taxanes and anthracyclines), HER2-targeted therapy and endocrine therapy. However, as cancer knowledge has evolved, new types of drugs that target specific changes in cancer cells have reached the clinical practice both alone or as part of different treatment combinations.

Newer targeted therapies that have been or are being evaluated, alone and in combination with each other and with traditional chemotherapy. The most relevant are polymerase 1 (PARP1) inhibitors[108,109], phosphatidylinositol 3-kinase (PI3K) inhibitors[110] and

CDK4/6 inhibitors[111]. In addition, angiogenesis inhibitors, tyrosine kinase inhibitors, inhibitors of mammalian target of rapamycin (mTOR), poly(ADP-ribose), insulin-like growth factor 1 receptor (IGF-1R) inhibitors, proteasome inhibitors and others are o have been also studied[97,112]. Moreover, the demonstration of modest single-agent activity of PD-L1 and PD-1 antibodies in breast cancer patients has generated hope that this type of cancer can be treated with immunotherapy[113,114].

Furthermore, and adding a new level of complexity to the already complex subject of breast tumour response to treatment, it is now well recognised that tumours are evolving entities that exhibit both intratumour and intertumour heterogeneity. This heterogeneity represents a large problem for therapeutic strategies. Patients with a similar type or grade of breast cancer may present different responses to therapy or long-term outcomes or even make switches between subtypes. Despite the advances and success of some therapies, the emergence of resistance represents one of the greatest current clinical challenges in the treatment of breast cancer.

3.3.1. Resistance to endocrine therapy

Anti-estrogen therapies block the effect of estrogen at the receptor level (selective estrogen receptor modulators or down-regulators) or inhibit the estrogen production (aromatase inhibitors)[115].

Luminal cancers, as previously mentioned, have a better prognosis than other types of breast cancer and are sensitive to anti-estrogen therapies. Nevertheless, despite the high sensitivity of luminal tumours to endocrine therapy, 30–50% of early breast cancer patients will later relapse[116], being resistance to therapy and distant metastases the main causes of death in these patients. In 15–20% of the cases, the resistance is associated with the activation of an ER-independent proliferation mechanism (such as *PIK3CA*, *mTOR* or *ERBB2*) that can be associated with a phenotypic change in cells, from ER+ to ER−[116].

In addition, these cancers have a tendency to stay dormant, and metastasis can be triggered up to 20 years after diagnosis. However,

resistance to this therapy is thought to be a progressive, step-wise process[115].

The value of this endocrine therapy relies on its lower toxicity and better quality of life when compared to chemotherapy, which is more toxic and has only modest benefits for many patients with ER-positive breast cancer. It is therefore critical to discover ways to extend endocrine therapy benefit and to monitor therapeutic resistance whenever possible[117].

3.3.2. Resistance to standard chemotherapy

Specific biological processes and distinct genetic pathways are associated with prognosis and sensitivity to chemotherapy and targeted agents in different subtypes of MBC. Conventional chemotherapies are initially effective in controlling tumour growth through the targeting of proliferating cells. However, subpopulations of cells with tumorigenic potential are intrinsically resistant to this type of therapy. In this case, the relative proportion of cells in residual tumours with tumorigenic properties would be expected to increase after treatment. Some of the alterations that occur in these cells involve changes on regulatory pathways of the cell cycle such as cyclins, CDKs or RB1[112].

In the past two decades, significant progress has been achieved in understanding drug resistance in breast cancer, involving drug efflux, alterations in DNA repair pathways, suppression of apoptosis as well as EMT and cell plasticity. However, more effective therapeutic targets and novel biomarkers are still urgently needed to refine the therapeutic strategies and to improve the OS of MBC patients.

In any case, highly specific biomarkers for predicting therapeutic resistance have not yet been identified.

3.4. Potential role of CTCs in metastatic breast cancer

Traditionally, treatment decisions in MBC are based on the characteristics of the primary tumour but a single tissue biopsy may be

insufficient to represent evolutionary changes in the tumour and the diverse resistance mechanisms driving individual clonal populations of progressing metastases. In addition, in some patients, tissue biopsy is either not possible or has been done many year before patient relapses[118].

CTCs, through liquid biopsy, as previously described, can be sampled and characterised repeatedly during the course of the disease in order to monitor treatment response and disease progression. In addition, cancer cells in circulation can complete the understanding of the metastatic cascade[119].

CTCs have been studied in breast cancer for more than 10 years, and, as previously mentioned on the section *Clinical relevance of the study of CTCs*, the majority of the phase III trials on CTCs are currently ongoing on breast cancer patients. Studies have explored their potential as biomarkers for breast cancer, not only as prognosis factors and markers of micrometastasis spread through their detection and enumeration (≥ 5 CTCs/7.5 mL of blood); but also, as a tool for liquid biopsy and therapy selection. However, so far, many questions remain unanswered, like the heterogeneity of CTCs, how much concordance exists between CTCs and primary tumour or which is their EMT state; being the reason why there is an increasing interest on the phenotypic characterisation of CTCs in MBC.

OBJECTIVES

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

OBJECTIVES

The objective of this thesis is to delve into the biology of CTCs through liquid biopsy, and evaluate their usefulness as a monitoring tool for cancer patients. To this aim we have proposed the following specific objectives:

• To isolate and analyse CTCs from cancer patients:

- To analyse the gene expression and identify biomarkers associated with prognosis.

- To evaluate the clinical utility and prognostic value of the identified markers and its relationship with the patients' clinical parameters.

• To develop and characterise preclinical models for the study of CTCs biology and the validation of liquid biopsy as a tool for personalised medicine:

- To achieve *in vivo* modelling studies of CTCs isolated from MBC patients for molecular characterisation.

- To validate clinically the identified markers.

• To optimise and validate new technologies for the isolation and characterisation of CTCs.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

METHODS

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

METHODS

1. CLINICAL SAMPLES

1.1. Metastatic breast cancer patient cohort: all breast cancer subtypes

To perform a prospective longitudinal expression analysis study, blood samples from MBC patients were obtained after informed patient consent and following the approval and recommendations of the Ethics Committee of Galicia (code approval: 2015/772) at the Oncology Department of the University Hospital Complex of Santiago de Compostela, Spain. The recruitment was performed from February 2016 to December 2017, and the follow-up until May 2018.

Two 7.5 mL EDTA-coated vacutainers blood tubes from MBC patients (n= 20, median age 53 years) were obtained during the routine analytical test once the patient was diagnosed with metastatic disease before treatment (visit 1, V1), after treatment (visit 2, V2) and after progression (if it takes place) (visit 3, V3) for each patient, and were processed within two hour after withdrawal (Figure 5). In addition, one CellSave preservative tube (Menarini-Silicon Biosystems) was collected in parallel at each visit for CTCs enumeration. Samples of 12 patients from this cohort, at V1, were also used to try to establish a CDX mice model.

In 8 patients, a formalin-fixed paraffin-embedded (FFPE) sample from normal tissue and the primary tumour (visit 0, V0) were provided by the Pathology Service and the Biobank of the University Hospital Complex of Santiago de Compostela (PT17/0015/0002), integrated in the Spanish National Biobanks Network (Figure 5).

In addition, one 7.5 mL EDTA-coated vacutainer blood tube from 6 female healthy donors was collected for PBMCs expression analysis.

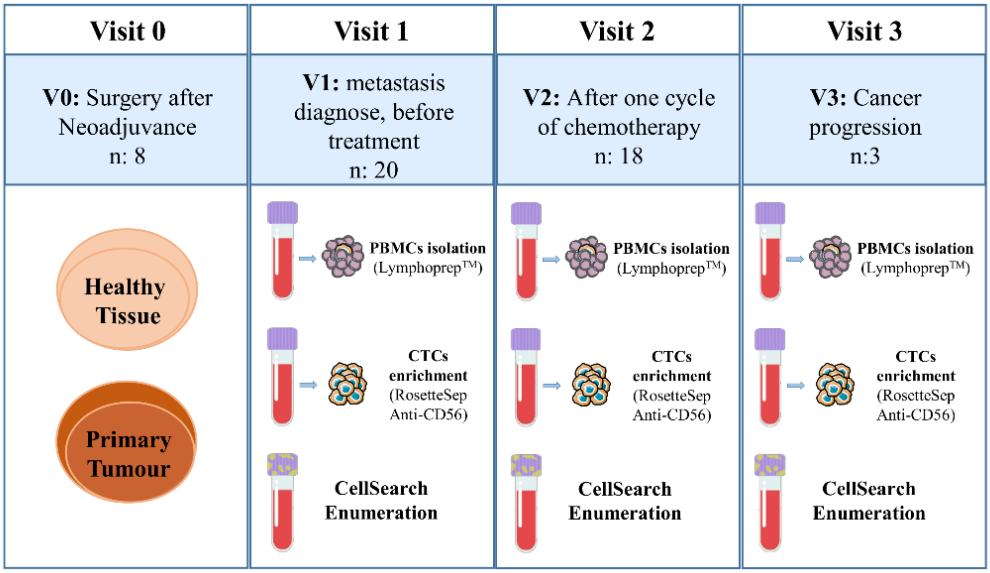

Figure 5. Scheme of the clinical sampling performed on the MBC patient cohort.

Patients' clinicopathological characteristics are summarised in Table I. Overall, 45% of the collected samples correspond to luminal breast cancer subtype, 40% to TNBC and 15% HER2 overexpressed subtype. All samples were anonymised an encoded before the analysis.

Table I. Clinical parameters	
Variable	**n(%)**
Age	
> 50 years	14 (70%)
≤ 50 years	6 (30%)
Tumour Stage	
IV	20 (100%)
ER status	
Positive	8 (40%)
Negative	12 (60%)
PR status	
Positive	7 (35%)
Negative	13 (65%)
HER2 status	
Positive	6 (30%)
Negative	14 (70%)
Metastasis location	
Bone & Visceral	10 (50%)
Visceral	9 (45%)
Unknown	1 (5%)
Total	**20 (100%)**

Table I. Clinical parameters of the MBC patient cohort.

1.2. Triple negative breast cancer patient cohort

For the establishment of a CDX mouse model and its posterior analysis and validation, a total of 32 patients (median age 58.5 years) diagnosed of TNBC at the Oncology Department of the University Hospital Complex of Santiago de Compostela were included in the study. In addition, 22 age matched healthy controls were also included. All participants signed an informed consent specifically approved for this study by the Galician Investigation Ethical Committee (code of approval: 2013/462). The recruitment was

performed from July 2013 to December 2017, and the follow-up until January 2018.

At diagnostic of the metastatic disease and before treatment (visit 1, V1), two tubes (7.5 mL) of peripheral blood were obtained from each patient: one EDTA-coated vacutainer tube for CTCs enrichment and characterisation and one CellSave preservative tube for CTCs enumeration.

In the presence of a clear progression of the disease (visit 2, V2), a second set of samples were obtained from patient #20, a 43 years old woman with high CTCs counting by CellSearch (969 CTCs/7,5 mL). Additionally, another EDTA-coated blood tube (7.5 mL) was obtained at this time point for CDX generation from this patient (time line of the collection of samples can be found in Figure 6).

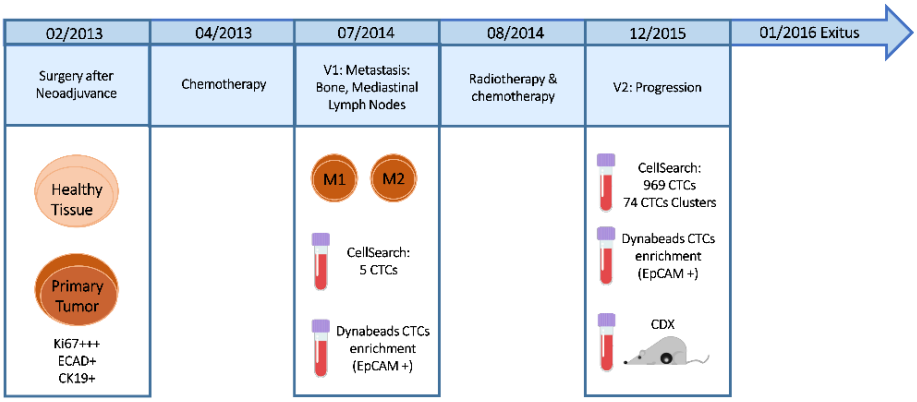

Figure 6. Patient #20 timeline and clinical case presentation, including the different samples analysed in our study. M1: Metastasis site 1. M2: Metastasis site 2. V1: Visit 1; V2: Visit 2.

FFPE biopsies from this patient's primary tumour (obtained at baseline, before surgery and chemotherapy) and metastasis after disease progression were provided by the Pathology Service and the Biobank of Hospital Complex of Santiago de Compostela (PT17/0015/0002). These samples were processed following standard

operating procedures with the appropriate approval of the Ethical and Scientific Committees.

Clinicopathological characteristics of this cohor of patients are summarized in Table II. All of the samples were anonymized and encoded before the analysis.

Table II. Clinical parameters	
Variable	**n(%)**
Tumour Stage	
III	9 (28%)
IV	23 (72%)
Histology	
Ductal	29 (91%)
Visceral	1 (3%)
Metaplastic	2 (6%)
Status at sample date	
PD	8 (25%)
No PD	24 (75%)
Previous chemotherapy	
Yes	13 (41%)
No	19 (59%)
Previous surgery	
Yes	23 (72%)
No	9 (28%)
Metastasis location	
Bone & Visceral	4 (17%)
Visceral	17 (74%)
Unknown	2 (9%)
Total	**32 (100%)**

Table II. Summary of the clinical parameters of the TNBC cohort. PD: progression disease.

1.3. Colorectal cancer patient cohort

For the biological validation of a microfluidic device (CROSS chip), metastatic colorectal cancer patients, (n=9, median age 72.44 years), were recruited at the Oncology Department from the University Hospital Complex of Santiago de Compostela, from November 2015 until October 2016.

One 7.5 mL EDTA-coated tube and another 7.5 mL CellSave preservative tube were collected after informed consent and following the approval and recommendations of the Galician Investigation Ethics Committee (code of approval: 2014/126). For control purposes, peripheral blood from two healthy donors was collected in EDTA-coated tubes after informed consent.

Clinicopathological characteristics of the patients are summarised in Table III. All the samples were anonymised and encoded before the analysis.

Table III. Clinical parameters	
Variable	**n(%)**
Gender	
Men	6 (66.6%)
Women	3 (33.3%)
Tumour Stage	
IV	9 (100%)
Localization	
Colon	4 (44.44)
Sigma	3 (33.33)
Recto	2 (22.2)
Lines of treatment	
Two lines	6 (66.6%)
Three lines	3 (33.3%)
Metastasis location	
Liver	9 (100%)
Lungs	2 (22.2%)
Lymph nodes	1 (11.1%)
Total	**9 (100%)**

Table III. Summary of the clinical parameters of the colorectal cancer cohort.

2. CTCs ENUMERATION

2.1. Human blood samples

One CellSave preservative tube (Menarini-Silicon Biosystems) from each patient was obtained in parallel at each withdrawal and was analysed for CTCs enumeration by the CellSearch System, using the

CellSearch Epithelial Circulating Tumour Cell Kit (Menarini-Silicon Biosystems).

This analysis was performed by the Liquid Biopsy Analysis Unit of the University Hospital Complex of Santiago de Compostela. Briefly, this automated process consists on the incubation of the mononuclear cells phase, obtained from 7.5 mL of peripheral whole blood, with ferrofluid nanoparticles coated with antibodies anti-EpCAM. CTCs are then magnetically separated and labelled with anti-CKs antibodies (CK8/18/19) conjugated with phycoeritrin, with anti-CD45 antibodies conjugated with allophycocyanin and with 4,6-diamino-2-phenylindole (DAPI). Next, CellTracks Analyzer (Menarini-Silicon Biosystems) is used to acquire digital images of the three different fluorescent dyes which then are reviewed by trained operators to determine the CTCs count.

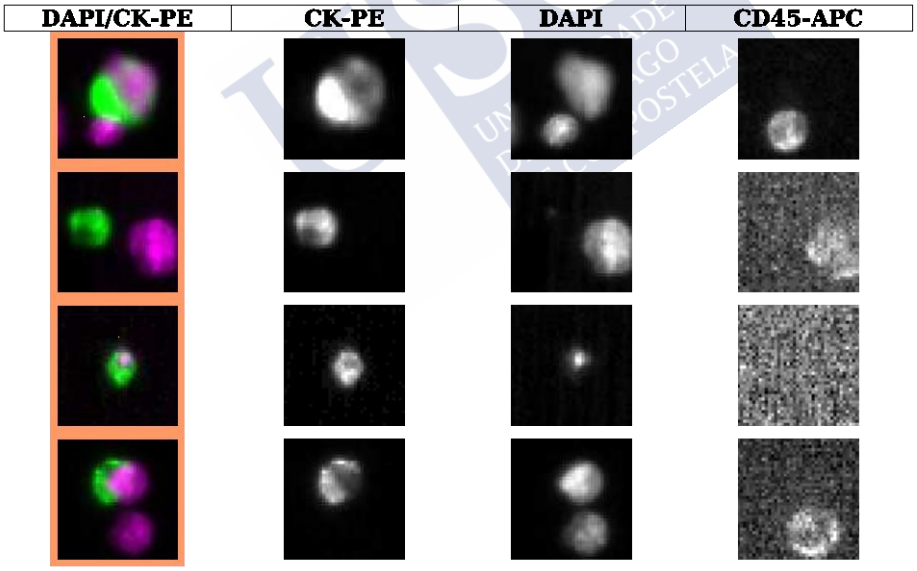

Figure 7. Extract of CellSearch report from patient #20 with tumour cell candidates that are positive for CK (green) and DAPI (pink) but negative for CD45.

2.2. Mice blood samples

For the CTCs enumeration on mice blood samples, blood was drown by cardiac puncture and collected in 4 mL EDTA tubes. CTCs enumeration was carried out using the CellSearch System from 150 µl of mouse blood mixed with 25 µl of human healthy blood (essential for the correct autofocus of the sample).

The volume-scaled protocol for isolation and immunostaining was performed manually using CellSearch Epithelial Circulating Tumour Cell Kit as described in Methods 2.1.

3. CTCs ENRICHMENT

3.1. Negative enrichment

One EDTA tube from the MBC cohort was used to isolate CTCs by negative selection using the RosetteSep™ CTC Enrichment Cocktail Containing Anti-CD56 (STEMCELL Technologies).

This cocktail is designed to enrich CTCs from fresh whole blood by negative selection. Unwanted cells are targeted for removal with tetrameric antibody complexes that recognise different blood cell populations (Table IV).

Antigen	Target
CD3	T cells
CD14	Macrophagues/monocytes and granulocytes
CD16	T cells, dendritic cells, NK cells, macrophagues/monocytes and granulocytes
CD19	B cells and dendritic cells
CD38	T cells, B cells, dendritic cells, NK cells, macrophagues/monocytes and granulocytes
CD45	T cells, B cells, dendritic cells, NK cells, macrophagues/monocytes and granulocytes
CD56	T cells and NK cells
CD61	Macrophagues/monocytes, platelets and endothelial cells
CD66b	Granulocytes
Glycophorin A	Red blood cells

Table IV. List of all the antigens and its targets included in RosetteSep™ CTC Enrichment Cocktail Containing Anti-CD56.

To enrich CTCs, 10 mL of fresh whole blood were incubated for 20 min with 500 µL of RosetteSep™, at room temperature. Then, blood was diluted to 20 mL with phosphate buffered saline containing a 2% of bovine serum albumin (PBS-BSA 2%). Next, sample was placed carefully on a SepMate™ tube (STEMCELL Technologies) containing 15 mL of gradient density medium Lymphoprep™ (STEMCELL Technologies) previously warmed at 37ºC. SepMate™ tubes were then centrifuged for 20 min, at 1200g, without brake. With this procedure, the unwanted cells are pelleted along with the red blood cells. The purified tumour cells are present as a highly enriched population in the interface between the plasma and the density gradient medium (Figure 8). This interface was recovered into another tube and centrifuged twice at 1200g with 20 mL of PBS-BSA 2% to remove traces of the gradient density medium.

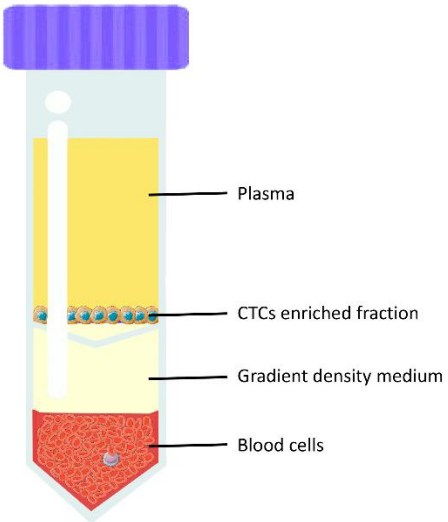

Figure 8. Schematic drawing of the different fractions obtained with the negative enrichment approach described above.

Immunoisolated cells were placed in RNAlater™ Solution (Invitrogen, ThermoFisher Scientific) and kept at -80ºC until further analysis.

3.2. Positive enrichment

For the analysis of the CTCs from the TNBC cohort, included patient #20, one EDTA tube was used for EpCAM+ isolation with CELLection™ Epithelial Enrich Dynabeads (Thermo Fisher Scientific), following manufacturer's instructions.

This kit allows the isolation of CTCs by incubating whole blood with superparamagnetic polymer beads coated with anti-EpCAM that are then specifically separated by a magnet.

The isolated CTCs were diluted in 100 µl of RNAlater (Ambion) and stored at -80ºC until RNA extraction.

4. PBMCS ISOLATION

One EDTA tube from each patient and control of the MBC cohort was used for PBMCs isolation by density gradient centrifugation with Lymphoprep™ medium in SepMate™ tubes. In this case, whole blood was diluted to 20 mL with PBS-BSA 2% and placed directly on SepMate™ tubes, containing 15 mL of Lymphoprep™ previously warmed at 37ºC, which are centrifuged for 10 min, at 1200g, without brake. PBMCs were recovered into another tube and centrifuged twice at 1200g with 20 mL of PBS-BSA 2%.

In addition, PBMCs from the collected controls of the TNBC cohort were isolated from EDTA tubes with CELLection™ Epithelial Enrich Dynabeads following manufacturer's instructions to serve as a control sample for this cohort.

Finally, the recovered PBMCs were placed in RNAlater™ Solution and kept at -80ºC until further analysis.

5. MICE EXPERIMENTS

Mice experimental protocols were approved by the Ethical Committee of the University of Santiago de Compostela (code of approval: 15010/2015/001). Mice were held in the animal facility at the Centre for Research in Molecular Medicine and Chronic Diseases (CIMUS, Santiago de Compostela, ES150780275701) and given food and water *ad libitum*, in accordance with CIMUS guidelines.

One NMRI-Foxn1nu/un mouse (NUDE) was obtained from Janvier Lab (France) and Scid Beige mice were obtained from the Barcelona Biomedical Research Park (PRBB, Barcelona). After mice arrival, at least one week of acclimation was considered.

For CDX establishment, isolation of PBMCs was performed by density gradient centrifugation protocol as described in Methods 4 on patient #20 from the TNBC cohort and, in addition, from 12 more

patients from the MBC cohort (CTCs account range: 0-483/7.5 mL) which were selected by the clinicians by their high tumour burden.

The recovered cells were diluted 1:2 with Matrigel Matrix (Corning) and implanted subcutaneously into a NUDE mouse in the case of patient #20 and into Scid Beige mice in the case of the other 12 patients. After cell injection, the mice were followed up weekly for tumour development using XenoLight RediJect 2-DG-750 (Perkin Elmer) by Xenogen IVIS 200 system.

For the mice monitoring, 100 µl of the reagent were injected intraperitoneally and the fluorescence was read 3 hours later. Two months after cell injection, a macroscopic tumour was observed on the patient #20 mouse, and three months later mouse was euthanised due to ethical reasons and the tumour was collected (CDX1). A piece of this tumour explant (25%) was implanted subcutaneously into a Scid Beige mouse (CDX2), and another piece (25%) was mechanically disaggregated and cultured in RPMI medium (Sigma Aldrich) in suspension. After 13 days of culture, cells were collected and injected into the mammary fat pad of a Scid Beige mouse (CDX2M). Tumour growth was monitored by *in vivo* image weekly; CDX2 was euthanised two months after cell injection, and CDX2M after three and a half months. CDX-derived tumour tissue fragments were collected from necropsied animals into RNAlater and stored at −80°C. On the other 12 injected mice tumour growth was observed in 9 cases and animals were sacrificed between 2.5 and 8 months after injection. The other three mice were sacrificed 8 months after sample injection without tumour development. Mice tumours were included in paraffin for pathology studies and in RNAlater for molecular analyses.

6. IMMUNOHISTOCHEMISTRY

Mice FFPE samples were analysed by IHC by the Oncologic Pathology Group at the Department of Pathology and Molecular Genetics from the Arnau de Vilanova University Hospital, (University of Lleida).

FFPE mice tissue blocks were sectioned at 3 µm, dried at 65ºC for one hour 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 50x Tris/EDTA buffer, pH 9 and endogenous peroxidase was blocked. The primary antibodies used are listed on Table V. After overnight incubation of the primary antibody, the reaction was visualized with Biotin-SP-AffiniPure Goat Anti-Rabbit IgG (dilution 1:200, Jackson ImmunoResearch), and Streptavidin (1:400, Agilent Technologies-DAKO) using diaminobenzidine chromogen as a substrate. Sections were counterstained with haematoxylin. Appropriate negative controls including no primary antibody were also tested.

The Ki67 percentage of the samples was automatically measured with an ACIS® III Instrument (DAKO). A percentage ≥ 14% was considered as high expression.

Target	Dilution	Reference
KI67	1:100	clone SP6, ABCAM
Wide Spectrum Cytokeratin	1:100	Polyclonal, ABCAM
CD45	1:100	clone EP322Y, ABCAM
ER	1:100	clone SP1, ABCAM
PR	1:100	clone YR85, ABCAM
ECAD	1:100	clone EP700Y, ABCAM
ALDH1A1	1:100	Polyclonal, ABCAM
NCAD	1:100	clone EPR1792Y, Merk Millipore
SNAI1	1:100	clone H-130, Santa Cruz Biotechnology
EpCAM	1:50	clone H70, Santa Cruz Biotechnology
c-erbB-2	1:100	Polyclonal, Agilent Technologies-DAKO

Table V. List of all the primary antibodies included in the IHC analysis.

7. NUCLEIC ACIDS EXTRACTION

7.1. RNA extraction

To perform the extraction of RNA from the negative enriched CTCs and the PBMCs from the MBC cohort, AllPrep DNA/RNA Mini Kit (Qiagen) was used following the manufacturer's protocol. In addition, CDX tumours were disaggregated with Tissuelyser (Qiagen) and RNA was extracted with the same extraction kit.

RNA from FFPE tissue sections was extracted with the miRNeasy FFPE kit (Qiagen) according manufacturer's instructions.

Total RNA from magnetically isolated samples from the TNBC patient cohort, was extracted with the QIAmp viral RNA mini kit (Qiagen).

7.2. DNA extraction

Extraction of genomic DNA from colorectal cancer CTCs retained in the microfluidic devices was performed using AllPrep DNA/RNA Mini Kit (Qiagen). Firstly, cells were lysed upon injection of a lysis buffer (Buffer RLT) in the CROSS chip at 80 µl/min using a syringe pump, followed by 5 min incubation and a second injection of the same buffer at 250 µL/min to collect all cell content. Subsequent steps were performed according to the manufacturer's recommendations.

Quantification of the extracted genomic DNA was performed with the Quantifluor ONE dsDNA System using Quantus Fluorometer (Promega).

8. cDNA SYNTHESIS AND PREAMPLIFICATION

In the CTCs, FFPE and PBMCs samples from the MBC cohort, as well as in the CDXs tumours, 11 µl of RNA were retrotranscribed into

cDNA using the SuperScript III (ThermoFisher Scientific) according manufacturer's protocol.

On the FFPE and the magnetically isolated samples from the TNBC patient cohort, cDNA was synthesised with MulV retrotranscriptase (Applied Biosystems) following manufacturer's instructions.

Due to the low recovery of RNA in CTCs and FFPE samples, cDNA from this samples were preamplified with 14 reaction cycles with Taqman Preamp Master Mix (ThermoFisher Scientific) containing a pool with all the TaqMan probes (Applied Biosystems) that will be included in the following qPCR step for each set of samples.

9. QUANTITATIVE PCR

For the MBC patient cohort expression analysis assay, cDNA expression from CTCs, PBMCs and FFPE samples was analysed on a LightCycler 480 II (Roche Diagnostics) with TaqMan Gene Expression Master Mix (Applied Biosystems) and TaqMan probes for a customised panel of 25 genes (listed in Table VI). PCR was performed as follows: denaturation at 95ºC for 10 min, and 40 cycles of amplification at 95ºC for 10 seconds, 60ºC for 10 seconds and 75ºC for 10 seconds, with fluorescence acquisition at 60ºC. *B2M* was used as a reference gene. After housekeeping normalisation, expression data from CTCs was relativised to the PBMCs corresponding transcripts for each patient sample.

cDNA from magnetically isolated samples was analysed following the same procedure for a customised panel of 19 genes (Table VI). In this case, expression values for each gene were normalised to *GAPDH*, and then referred to *CD45* as a marker of non-specific isolation.

cDNA from the CDXs and the FFPE samples from patient #20 was analysed also following the same procedure for a panel of 13

genes in the FFPE tissue and 23 genes in the CDX tumours (Table VII). *GAPDH* and *B2M* were used as reference genes in the FFPE samples and CDX tumours respectively.

Gene name	Taqman Assay	Functional Gene Grouping
ALDH1A1	Hs00946916_m1	Stem
B2M	Hs00187842_m1	Housekeeping gene
BCL11A1	Hs01093197_m1	Stem
BCL2	Hs00608023_m1	Apoptosis
CCND1	Hs00765553_m1	Proliferation/cell cycle regulation
CD36	Hs00354519_m1	Cell metabolism
CDH1	Hs00170423_m1	Epithelial
CDK4	Hs01565683_g1	Proliferation/cell cycle regulation
CTNNB1	Hs00355049_m1	Cell adhesion/gene transcription
E2F4	Hs00608098_m1	Proliferation/cell cycle regulation
EpCAM	Hs00158980_m1	Epithelial
ERBB2	Hs01001580_m1	Breast cancer associated
ESR1	Hs01046816_m1	Breast cancer associated
FAS	Hs00163653_m1	Apoptosis
GDF15	Hs00171132_m1	Proliferation/cell cycle regulation
KRT5	Hs00361185_m1	Epithelial
KRT19	Hs00761767_s1	Epithelial
MYC	Hs00153408_m1	Oncogene
MYCL	Hs00420495_m1	Oncogene
PALPB2	Hs00226617_m1	Breast cancer associated
PROM1	Hs01009257_m1	Stem
PTPRC	Hs04189704_m1	Blood cell
RB1	Hs01078066_m1	Proliferation/cell cycle regulation
SNAI1	Hs00195591_m1	EMT
VIM	Hs00958116_m1	EMT
ZEB1	Hs00232783_m1	EMT

Table VI. Gene expression custom panel assay for the study of CTCs on the MBC cohort.

Gene name	Taqman Assay	Functional Gene Grouping	Analysed in
ALDH1A1	Hs00946916_m1	Stem	CDX, CTCs
ALDH2	Hs01007998_m1	Stem	CDX, FFPE, CTCs
AR	Hs00171172_m1	Hormonal receptor	CDX, FFPE, CTCs
B2M	Hs00187842_m1	Housekeeping gene	CDX
BCL11A1	Hs01093197_m1	Stem	CDX, FFPE, CTCs
CCND1	Hs00765553_m1	Proliferation/cell cycle regulation	CDX, CTCs
CD44	Hs01075861_m1	Stem	CDX, FFPE, CTCs
CD49f	Hs01041011_m1	Stem	CDX, FFPE, CTCs
CDH1	Hs00170423_m1	Epithelial	CDX, FFPE, CTCs
CTNNB1	Hs00355049_m1	Cell adhesion/gene transcription	CDX, CTCs
EGFR	Hs01076090_m1	Epithelial	CDX, FFPE, CTCs
EpCAM	Hs00158980_m1	Epithelial	CDX, FFPE, CTCs
GAPDH	Hs99999905_m1	Housekeeping gene	FFPE, CTCs
KDR	Hs00911700_m1	Endothelial	CDX, CTCs
KRT19	Hs00761767_s1	Epithelial	CDX, CTCs
PECAM1	Hs01065282_m1	Endothelial	CDX, CTCs
PI3KCA	Hs00907957_m1	Oncogen	CDX, CTCs
PROM1	Hs01009257_m1	Stem	CDX, CTCs
PTPRC	Hs04189704_m1	Blood cell	CDX, CTCs
TDGF1	Hs02339497_g1	Tumour growth	CDX, FFPE, CTCs
TWIST1	Hs01675818_s1	EMT	CDX
SNAI1	Hs00195591_m1	EMT	CDX, FFPE, CTCs
VIM	Hs00958116_m1	EMT	CDX, FFPE, CTCs
ZEB1	Hs00232783_m1	EMT	CDX, CTCs
AURKB	Hs00945858_g1	Validation assay - Mitosis regulator	TNBC cohort CTCs
HIST1H4A1	Hs01924141_s1	Validation assay - Transcriptation regulator	TNBC cohort CTCs
MELK	Hs01106438_m1	Validation assay - Mitosis regulator	TNBC cohort CTCs
MYCL	Hs00420495_m1	Validation assay - Oncogene	TNBC cohort CTCs
PCDHA8	Hs00560506_s1	Validation assay - Cell Adhesion	TNBC cohort CTCs

Table VII. Gene expression panel and the corresponding samples where it was analysed.

10. RNA-SEQUENCING ANALYSIS

For the RNA-seq analysis of patient #20 tumour and the corresponding CDXs tumours, samples were barcoded and prepared for sequencing at the Wellcome Trust Centre for Human Genetics, Oxford, where 75 bp paired-end reads were obtained on an Illumina HiSeq 4000. The raw data was deposited in the NCBI's Sequence Read Archive under accession number PRJNA464335.

The quality of the sequencing output was assessed using FastQC v.0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Quality filtering and removal of residual adaptor sequences was conducted on read pairs using Trimmomatic v.0.35[120]. Specifically, Illumina adaptors were clipped from the reads, leading and trailing bases with a Phred score < 20 were removed and the read trimmed if a sliding window average Phred score over four bases was less than 20. Only reads where both paired-end reads had a length greater than 36 bp post-filtering were retained.

Filtered reads were mapped to the human genome (GRCh38.p10) and the mouse genome (GRCm38.p5) using STAR v.2.5.2b[121], the maximum number of mismatches for each read pair was set to 10 % of trimmed read length, and minimum and maximum intron lengths were set to 20 bases and 1 Mb respectively. For those reads aligning to both the human and the mouse genome, only those with a higher mapping quality in human were retained for further analysis.

Paired-reads uniquely mapped to the human genome were counted and assigned to genes using FeatureCounts[122], included in the SourceForge Subread package v.1.5.0. Only reads with both ends mapped to the same gene were retained. Gene count data was used to estimate differential gene expression using the Bioconductor packages DESeq 2 v.3.4[123]. Samples were hierarchically clustered according to gene read counts after a variance stabilising transformation, using Euclidean as the distance measure and complete-linkage as the agglomeration method (R package flashClust[124]). Heatmaps of gene expression were created using the R package gplots v3.0.1 heatmap.2 function, using read counts after regularised log transformation (DESeq2[123]).

The Genome Analysis Toolkit (GATK) framework was used for variant calling. Duplicated reads were removed using Picard v.1.128 (http://broadinstitute.github.io/picard). GATK[125] was used to remove sequences overhanging into the intronic regions, to reassign STAR mapping qualities to default values and to perform base quality score recalibration. Somatic single nucleotide polymorphisms (SNPs) and

Indels were identified using Mutect2, which combines the original MuTect[126] and HaplotypeCaller [125]. The mutations were annotated using Variant Effect Predictor (Ensembl version 90[127]).

Enrichment analysis of Gene Ontology (GO) terms was done upon uploading selected probe sets identifiers into GSEA (Gene Set Enrichment Analysis) and Panther web tools. Venn Diagrams were drawn with VENNY 2.1 software.

11. PATIENTS' BLOOD SAMPLE PROCESSING IN THE CROSS CHIP

A microfluidic device named "CROSS chip" was developed by the INL International Iberian Nanotechnology Laboratory. This device was designed to split the blood equally in 4 different modules (Figure 9A). Each module is able to process 1 ml of whole blood and contains a set of pre-filters with 120 μm gaps to prevent large clumps or debris from clogging the setup (Figure 9B). Across the middle section of each module, a single row of 700 anisotropic micropillars with diameter 25μm and spaced 5 μm constitutes the cell filtering area (Figure 9C). The gap size, geometry and aspect ratio were chosen to allow blood cells to deform and gently flow through, while retaining larger or more rigid cells in the filter. Cells can be retrieved from this system by inverting the flow.

After technical validation of the device with spiked blood samples by the INL International Iberian Nanotechnology Laboratory, eight blood samples from metastatic colorectal cancer patients were tested in the CROSS chip. Each tube containing 7.5 ml of whole blood was divided in half. Thus, 3.75 ml of blood were processed in each of two CROSS chips and injected at 80 μl/min in the CROSS chip with a syringe pump (New Era Pump Systems, Inc.). Trapped cells were rinsed with PBS-BSA 2% (Sigma Aldrich) and fixed with 4% paraformaldehyde (Sigma Aldrich) for 15 min at room temperature. Devices were rinsed once again with PBS-BSA 2% and stored at 4ºC until further analysis.

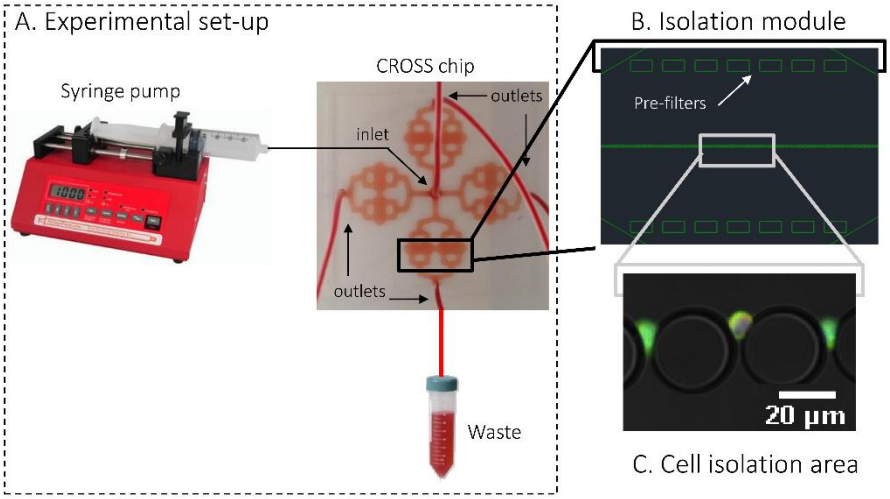

Figure 9. (A) Experimental set-up for CTCs isolation using the CROSS chip. (B) Each chip displays 4 modules containing a set of pre-filters with 120 µm gaps. (C) Across the middle section of each module, a single row of 25 µm anisotropic micropillars spaced 5 µm constitutes the cell filtering area.

11.1. Immunofluorescence and CTCs enumeration in the CROSS chip

Isolated cells from patient samples were permeabilised with 0.25% Triton X-100 solution (Sigma Aldrich) and fluorescently labelled inside the microfluidic device with anti-pan CK-FITC (clone C-11, recognises human CK 4,5,6,8,10,13, and 18, Sigma; dilution 1:100), anti-Vimentin eFluor 570 (eBioscience, dilution 1:50) and anti-CD45-Cy5 (Abcam; dilution 1:25) antibodies for one hour. DAPI was used as a nuclear marker. Fluorescence microscopy analysis of the trapped cells was performed using a plan fluor 20x objective (Nikon) coupled to a fluorescence-adapted inverted Nikon-MA 200 microscope (Nikon) Only DAPI+/CK+/CD45- cells were considered for CTCs enumeration, whereas DAPI+/CK-/CD45+ represented the leukocytes population.

CTCs quantification was performed adding the number of cells isolated in the 2 CROSS chips used for each analysis, with blind scoring and by 3 different examiners. The ability of the CROSS device to isolate epithelial-mesenchymal or mesenchymal-epithelial transitioning CTCs was evaluated by confirming the presence of DAPI+/Vim+/CD45- cells.

12. ddPCR ANALYSIS

Genomic DNA from colorectal cancer CTCs retained in the microfluidic chips was analysed for the absolute quantification of *APC* mutations by ddPCR analysis (QX200™ Droplet Digital™ PCR System, Bio-Rad) at the Universitat Autònoma de Barcelona Scientific Technical Services (Barcelona).

Prior to quantification, samples were digested with the restiction enzyme HaeIII (Sigma-Aldrich) and pre-amplified with Sso Advanced Preamp Supermix (Bio-Rad) following manufacturer's recommendations. ddPCR experiments were performed using probes dHsaCP2500509 and dHsaCP2500508 for detecting the *APC* mutation p.R1450* (COSM13127). The droplets were quantified using the Bio-Rad Quantasoft software. Two replicates per sample were performed.

13. STATISTICAL ANALYSIS

Statistical analysis was performed using IBM SPSS Statistics for Macintosh, Version 22.0 (SPSS Inc.), GraphPad Prism 6.01 software (GraphPad Softwares Inc.) and R Studio Version R-3.5.0. The Wilcoxon signed-rank test was used to compare CTCs longitudinal enumeration and the performance of CellSearch System versus the CROSS chip. Assuming PBMCs contamination in the enriched fraction of CTCs, the expression of the autologous PBMCs was used as a normalizer. Differences on CTCs enumeration and expression among subtypes were compared with Kruskal-Wallis test. Fisher test and Mann-Whitney test were used to study the association between

enumeration and gene expression of the CTCs and gene expression differences between patients and controls. Correlations between gene expression and clinical data were tested by chi2. PFS and OS were visualised using Kaplan-Meier plots and tested by log-rank test. Only p values < 0.05 were considered statistically significant.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

RESULTS

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

RESULTS

1. CTCs EXPRESSION PROFILING FOR METASTATIC BREAST CANCER MONITORING

1.1. Longitudinal CTCs enumeration: prognostic role in metastatic breast cancer patients

Blood samples from 20 MBC patients at different time points of the disease were analysed: diagnose of metastasis (V1, n=20), after first-cycle of therapy (V2, n=18) and after patients' progression (V3, n=3). 45% of the collected samples were luminal breast cancer subtype, 40% were TNBC and 15% were HER2 overexpressed subtype. In total, 41 blood samples were collected and analysed by CellSearch for CTCs enumeration.

At V1, 70% of the patients shown CTCs detection (≥ 1 CTCs) and 40 % were CTCs+ for the predefined cut off of ≥ 5 CTCs (mean= 69.85, range= 0-445). After the first cycle of treatment the percentage of patients with ≥ 5 CTCs decreased to 22% (mean=35.9, range= 0-484), and all of the samples suffered a reduction, in different grade, on their CTCs counting, being this reduction statistically significant among V1 and V2 (Figure 10, Wilcoxon test, p= 0.041). In addition, of the collected patients in V1, two deceased before V2 and three patients keep ≥ 5 CTCs in V2. In V3, two out of the three patients that progressed shown ≥ 5 CTCs, one of the samples depicted a high increase in the CTCs account from 121 to 233 CTCs (mean=83, range= 2-233) (Figure 10).

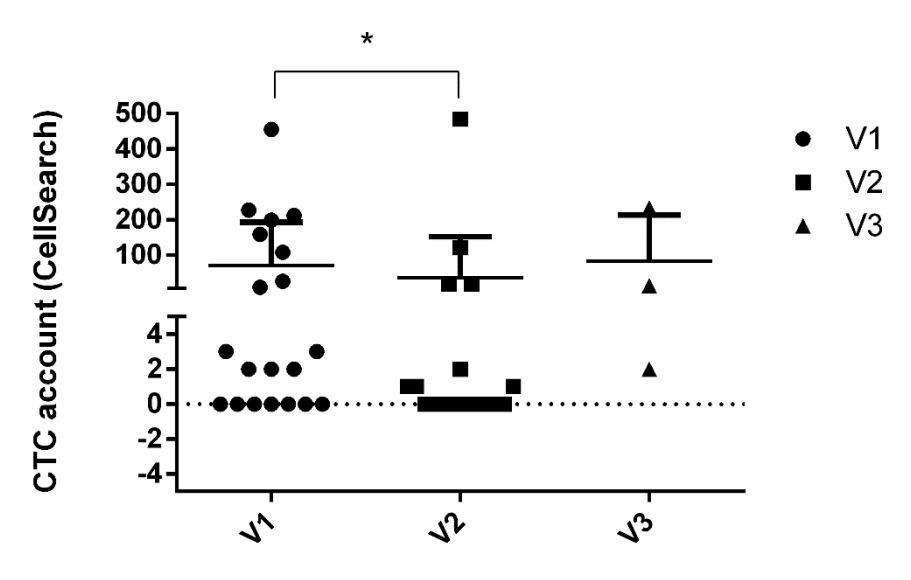

Figure 10. Longitudinal CTCs enumeration on the MBC patient cohort, (V1 mean=69.85, range= 0-445, V2 mean=35.9, range= 0-484, V3 mean= 83, range= 2-233; Wilcoxon test, p= 0.041).

Next, we performed the analysis of CTCs enumeration data by molecular subtypes to check whether CTCs detection was more frequent in any subtype. This analysis revealed that at V1, CTCs detection by CellSearch was mainly in luminal and HER2 patients, while CTCs detection in TNBC patients was rare, statistical differences were found between luminal B and TNBC at V1 (Kruskal-Wallis, p= 0.005) [luminal A (99.5 ± 140.71), luminal B (131.57 ±170.33); HER2 (37.66 ± 60.91); TNBC (20.5 ± 55.97)] (Figure 11).

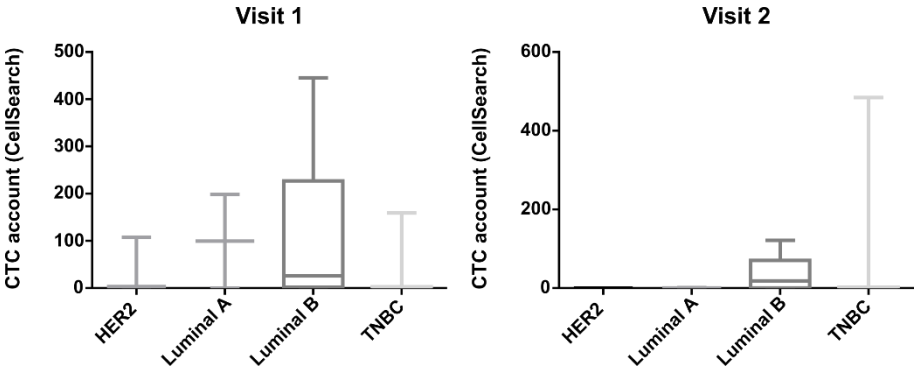

Figure 11. Boxplot representing CTCs enumeration on the different molecular subtypes of breast cancer across the V1 and V2.

In order to check the prognostic value of CTCs enumeration in our cohort, we performed a survival analysis considering the cut off of ≥ 5 CTCs. Patients with ≥ 5 CTCs in V1 shown to be significantly associated with a shorter OS (Figure 12A, 111 days, p= 0.029), although no differences were found in PFS (Figure 12B). Interestingly, in V2, after a first cycle of therapy, patients with ≥ 5 CTCs had both shorter PFS and OS (49.5 days, p= 0.027 and 35.5 days, p= 0.002, respectively, by log-rank test) (Figure 12C, 12D). There were not enough samples at the progression point (V3) to perform a conclusive survival analysis.

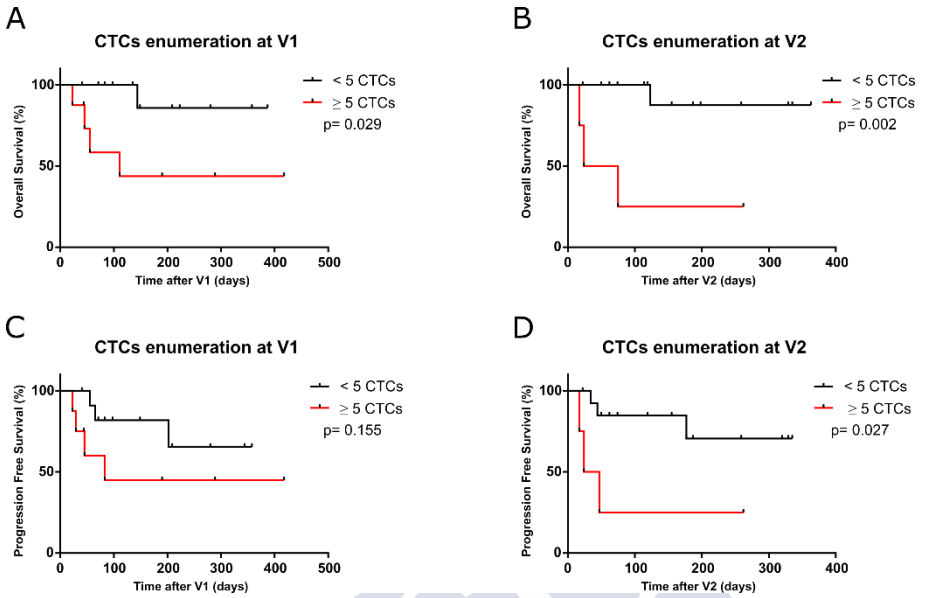

Figure 12. Estimates of probabilities for OS (A) and PFS (C) at V1 (111 days, p= 0.029 and p= 0.155) and V2 (B,D) (49.5 days, p= 0.002 and 35.5 days, p= 0.027) in MBC patients with ≥5 CTCs per 7.5 ml of blood.

1.2. Association study between enumeration by CellSearch and gene expression analysis on negative enriched CTCs

In parallel to CTCs enumeration by CellSearch, an expression analysis study with a customised panel of 25 selected genes (epithelial, mesenchymal, stemness, proliferation, EMT, cellular metabolism, apoptosis, oncogenes and breast related genes, described in Methods Table V) was performed. To achieve this aim, one EDTA tube was used to isolate CTCs through a negative enrichment approach with Rosettesep^{TM} System, (STEMCELL Technologies) and a second EDTA tube was used in parallel for the isolation and expression analysis of PBMCs.

PBMCs gene expression pattern in our cohort of patients was highly heterogeneous for several of the analysed genes such as *E2F4*, *EpCAM*, *GDF15*, *KRT19*, *PROM1* and *ZEB1* (Figure 13). In addition,

when being compared with a pool of six female healthy controls, PBMCs from patients showed an abnormal patron of *EpCAM*, *GDF15*, *KRT19* and *SNAI1* expression. Thus, in this study, the expression of CTCs from each patient was calculated relative to the autologous PBMCs expression, minimising the bias from inter-patient heterogeneity.

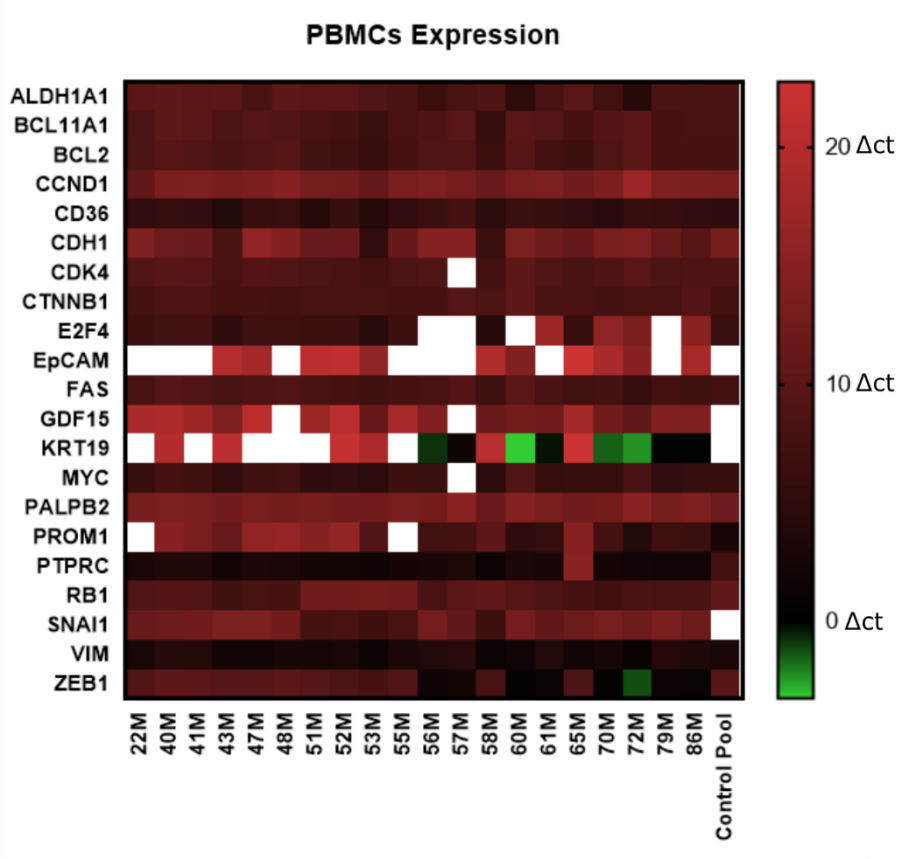

Figure 13. Heatmap illustrating the qPCR expression levels of patients' PBMCs of the custom panel of genes analysed at V1 (ΔCt, lower levels in red, higher levels in green, non-expression in white). The expression of a pool of 6 controls was included in the last lane. *B2M* was used for normalisation.

An association analysis between CTCs enumeration by CellSearch and the relative expression of the panel of genes analysed

in CTCs was performed. For this study, median value was used to define high /low levels of each marker and perform the contingency analysis. This analysis revealed a correlation at V1 between CTCs detection by CellSearch (≥ 1 CTCs) and the overexpression of epithelial (*EpCAM*, *KRT5*), proliferation (*GDF15*) and breast cancer associated (*ERBB2*, *ESR1*) genes (Fisher test, p= 0.019). However, if the detection of CTCs by CellSearch was considering the pre-established cut off of ≥ 5, this association was established with the overexpression of *EpCAM*, *KRT19* and *ERRB2* (Fisher test, p= 0.006). In addition, at V2, after 1 cycle of treatment, the expression of the epithelial marker *CDH1* was associated with the detection of ≥ 1 CTCs (Fisher test, p= 0.016); and the cell cycle gene *CCND1* associated with the presence of ≥ 5 CTCs (Fisher test, p= 0.043). Besides, expression of *CCND1* was found also in samples with no detection of CTCs (Mann-Whitney test with mean values, p>0.05).

1.3. CTCs gene expression in metastatic breast cancer patients

In total, CTCs were analysed in 20 samples from patients at V1, 18 at V2 and in 3 patients at V3. As pictured in Figure 14, gene expression was highly consistent across all time points tested (fold change ≥ 1.5 was considered positive expression), however, some differences were identified. In all the visits, at least one epithelial marker was detected in all the patients, being *CDH1* the most commonly expressed gene (95%, 95% and 100% respectively).

Interestingly, at V3 the presence of epithelial markers was higher compared with the other visits. Regarding the EMT markers, their expression was highly homogeneous between all the visits and *SNAI1* was the most frequently expressed (80%, 83% and 64%, respectively). From the stem marker panel, we observed a slight decrease on the expression of *ALDH1A1* at V2 (from 45% to 38%) but its expression increased again at V3 (66.5%). Finally, at least one breast cancer associated maker was detected in 60% of the patients at V1, 72% at V2 and 100% of the patients at V3.

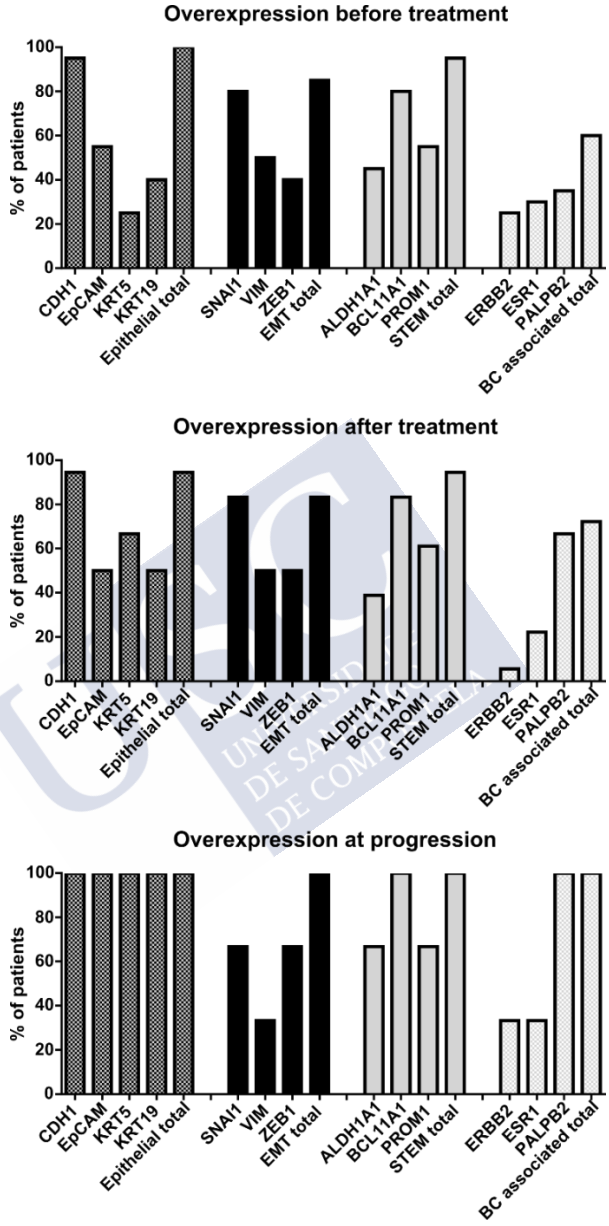

Figure 14. Percentages of gene expression in the patients for the epithelial, EMT, stem and breast cancer (BC) associated genes in the different time points of the disease.

We compared the pathology results for HER2 and ER expression in the primary tumour with the expression of *ERBB2* and *ER* in the CTCs, at the different visits (Figure 15). We found concordance on the HER2 status in 70% of the patients at V1, 55.5% at V2 and 66% at V3. Interestingly, we observed one case (patient #58M) where HER2 expression was lost at metastasis diagnosis and after treatment but was acquired again at progression. In a similar way, four patients with HER2- tumours had *ERBB2*+ CTCs. Regarding *ER* expression, we detected concordance with the primary tumour in 65% of the patients at V1, 66.6% at V2 and 100% at V3. Just like *ERBB2*, *ER* expression showed a dynamic pattern of expression in the different time points of the disease.

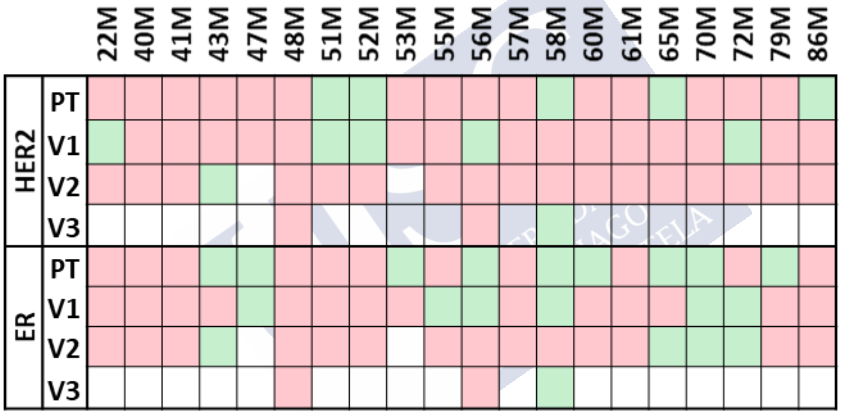

Figure 15. HER2 and ER expression evolution on the primary tumour and the CTCs at the different time points of the disease. PT: primary tumour, V1: Visit 1, V2: Visit 2, V3: Visit 3 (positive expression in green, negative in red, deceased or NA patients in white).

1.4. Gene expression analysis correlation with clinical data

First, we performed a correlation analysis to investigate whether the expression of the selected genes was related with any of the different breast cancer subtypes. Results showed that only *KRT19* and *ALDH1A1* median expression had a correlation with the different breast cancer molecular subtypes. Thus, *ALDH1A1* showed

statistically lower levels of expression in HER2 and luminal A patients and higher expression in luminal B and TNBC patients (Kruskal-Wallis test, p= 0.036), while *KRT19* showed higher expression levels in luminal patients and lower in HER2 and TNBC subtypes (Kruskal-Wallis test, p= 0.008). On Figure 16 a comparative between mean values is depicted.

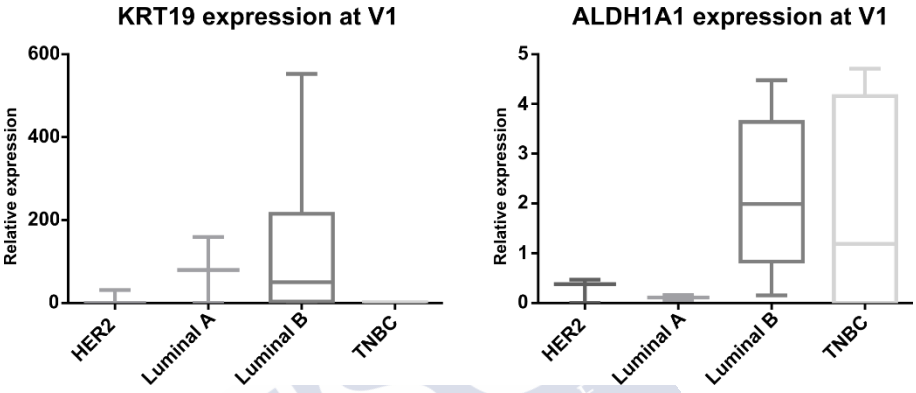

Figure 16. Boxplot of means for *KRT19* and *ALDH1A1* expression across the different subtypes.

Next, in order to investigate if the expression of any of the analysed markers was associated with clinical characteristics, a correlation analysis was performed considering the median values as a cut off to define high/low levels of expression (Spearman correlation). A higher tumour grade at cancer diagnosis correlated positively with high expression of *KRT5* in CTCs at V1 (p= 0.024) while the expression of *E2F4* had a negative correlation (p= 0.024). At V2, CDK4 also had a negative correlation (p= 0.011) with the tumour grade. Regarding the TNM staging system, we found that a greater tumour size correlated with a high expression of *PROM1* in CTCs after one cycle of treatment (p= 0.025).

Last, we studied the correlation between the hormonal receptors and HER2 status on the primary tumour of the patients and the gene expression panel. Patients who were PR- showed a median expression value higher than those PR+ for *BCL11A*, *KRT15* and *RB1* genes (p= 0.019) at V1. Patients who were ER+ had a correlation with *KRT19*

expression in CTCs at V1 (p= 0.002), which is concordant with a higher expression of *KRT19* in luminal subtypes. In addition, also at V1, high *RB1* expression in CTCs correlated with ER+ expression in the primary tumour (p= 0.025). After one cycle of treatment (V2), high expression of *GDF15*, *CDH1* and *CD36* correlated positively with ER+ primary tumours (p= 0.016). HER2 status on the primary tumour did not correlate with any marker from the gene expression panel.

1.5. CTCs gene expression analysis and patient prognosis

To identify markers with prognostic value, we next performed a survival analysis. For that, we considered the median value of expression as the threshold to determine high or low expression. We didn't identify any association between the primary tumour tissue expression of the FFPE samples and the outcome of the patients. Nevertheless, on the CTCs enriched fraction, results showed that at V1, high expression levels of *MYC*, *PALB2* or *ERBB2* were able to discriminate patients with poor outcome. Thus, patients whose CTCs had high expression levels of *MYC* or *ERBB2* showed a shorter OS (144 days, log-rank test, p= 0.006 and 0.020 respectively). In addition, patients with high expression of PALB2 had a poorer outcome, both for OS and PFS (144 days, p= 0.027 and 74 days p= 0.024, by log-rank test). (Figure 17).

When we studied the CTCs expression levels after one cycle of treatment (V2), we found that *MYC* expression had prognostic value. In Figure 18 is depicted the Kaplan-Meier curve, showing that high expression of *MYC* leads to a shorter OS (123 days, p= 0.016) and PFS (177 days, p= 0.05). As the oncogene *MYC* has been related with resistance to anti-estrogen therapy, we extended the survival analysis considering only patients diagnosed with luminal breast cancer. In this new analysis, high expression of *MYC* was correlated with a shorter OS also in V1 (111days, p= 0.046).

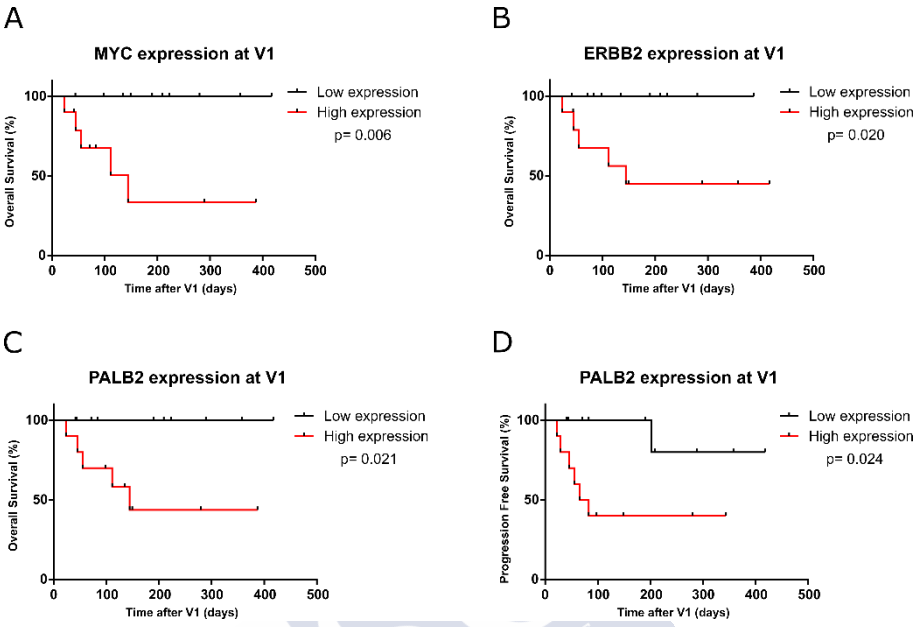

Figure 17. Kaplan-Meier plots for OS according to *MYC* (A) (144 days, p= 0.006), *ERBB2* (B) (144 days, p= 0.02) and *PALB2* (C) (144 days, p= 0.0021) expression levels; and PFS according to *PALPB2* (D) (74 days, p= 0.024) expression levels, at V1.

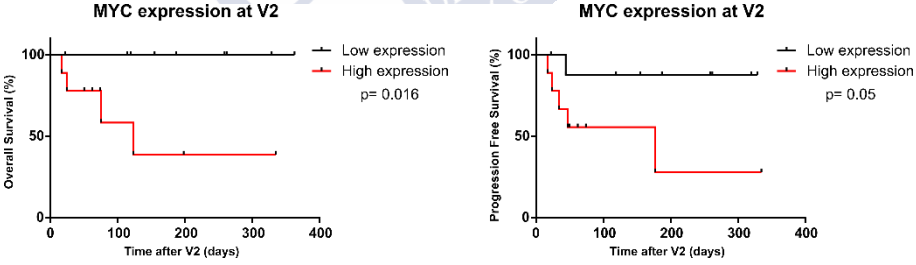

Figure 18. *MYC* expression after one cycle of treatment. Kaplan-Meier plots for OS and PFS according to *MYC* (123 days, p= 0.016 and 177 days, p= 0.05) expression levels at V2.

We next examine whether the expression of *CDK4* (a cyclin dependent kinase required for cell cycle entry) had an impact in patient´s outcome. *CDK4* is a target, together with *CDK6*, of CDK4/6 inhibitors. It is worth to mention that none of the analysed patients had

been treated with this therapy at the moment of collection of the samples. Results showed that after on cycle of treatment high expression of *CDK4* was associated with shorter OS (123 days, p= 0.032) (Figure 19).

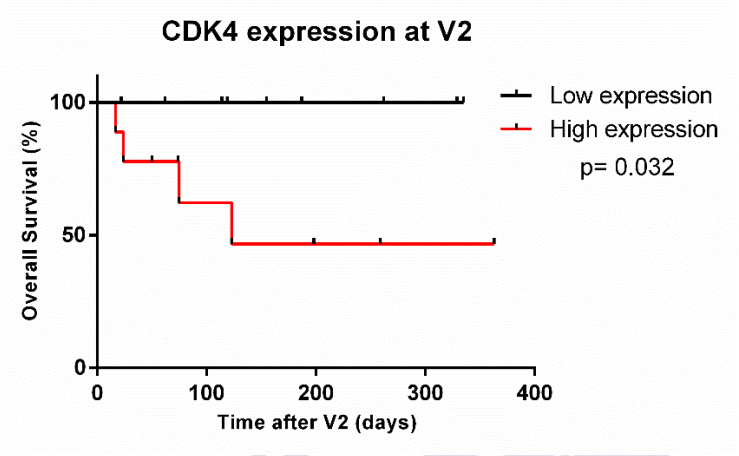

Figure 19. Kaplan-Meier plot for OS according to CDK4 (123 days, p= 0.032) expression levels at V2.

1.6. EMT prognostic value of a CTCs negative enriched population

The isolation of CTCs was performed using a negative enrichment protocol (Methods 3.1) in order to recover a wider CTCs population, without being biased by any antigen such as EpCAM for an epithelial fraction. High expression levels of *EpCAM* could discriminate patients with worse prognosis but without statistical significance ($p > 0.05$) (Figure 20A). In addition, *VIM*, an EMT related gene, was analysed in order to know its prognostic value. Unexpectedly, we found that a high expression of *VIM* lead to a better outcome in the MBC patients analysed (p= 0.021) (Figure 20B). Next, we established a signature considering both expression data, that is, high expression of *EpCAM* combined with low expression of *VIM*. This $EpCAM^{high}VIM^{low}$ signature was able to predict both shorter OS

(83 days, p= 0.006) and PFS (64 days, p= 0.032) at V1, with better significance than both markers separately (Figure 20C & D).

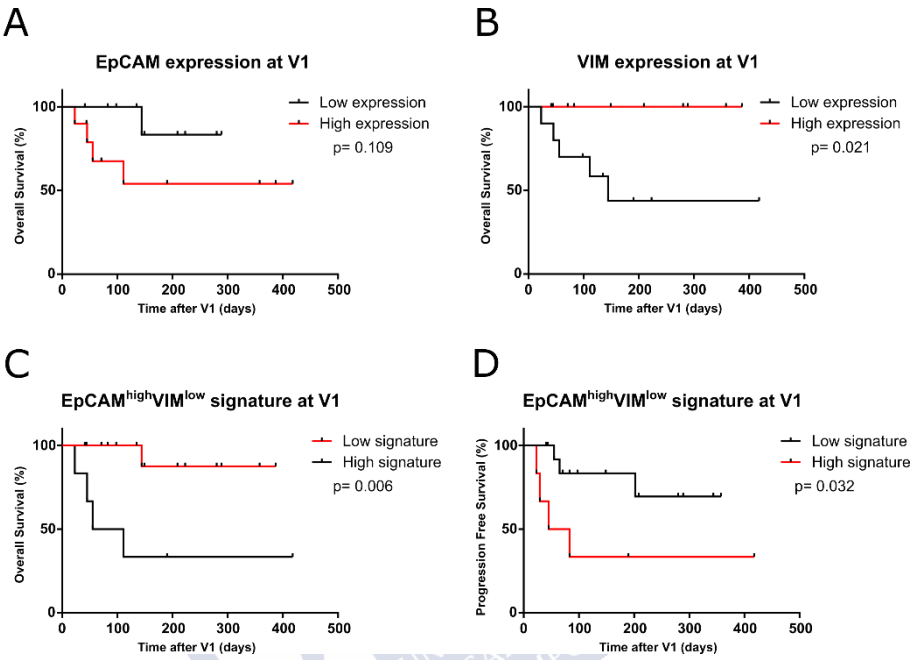

Figure 20. Kaplan-Meier plot for OS according to (A) *VIM* (p= 0.021) and (B) *EpCAM* (p>0.05) expression levels, and plot for (C) OS and (D) PFS for the *EpCAM*^{high}*VIM*^{low} signature (83 days, p= 0.006 and 64 days, p= 0.032, respectively).

If we took into account the CTCs with stem phenotype by including high *ALDH1A1* expression in the signature, *EpCAM*^{high}*VIM*^{low}*ALDH1A1*^{high}, we were also able to discriminate those patients with poor outcome improving the statistical value of the analysis (OS, 45 days, p<0.0001) (Figure 21). *ALDH1A1* expression alone was not able to discriminate patients with poor prognosis (OS, p= 0.12).

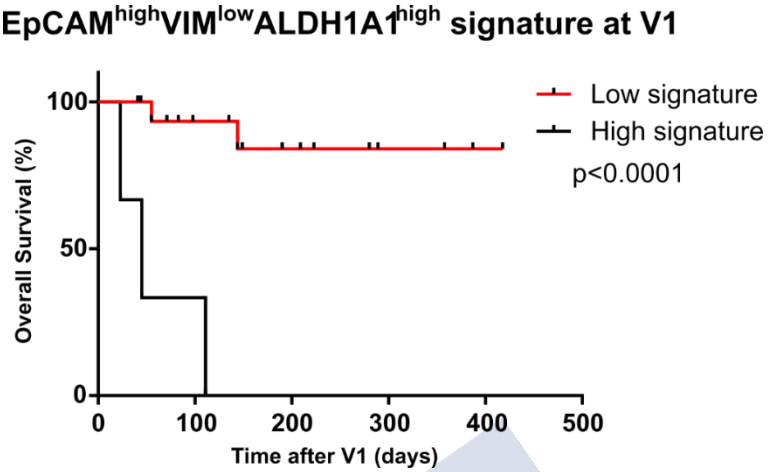

Figure 21. Kaplan-Meier plot for OS of the *EpCAM*^{high}*VIM*^{low}*ALDH1A1*^{high} signature (45 days, p<0.0001).

2. CTCS-DERIVED XENOGRAFT DEVELOPMENT IN A TRIPLE NEGATIVE BREAST CANCER CASE

2.1. CTCs-derived xenograft establishment

Patient #20 with advanced disease, from the TNBC patient cohort, was selected for CDX generation due to its high CTCs count after CellSearch analysis (969 CTCs/7,5 mL). In addition, twelve more samples from the MBC patient cohort, identified by the clinicians by their high tumour burden, were injected into immunocompromised mice. These samples showed smaller CTCs counts, ranging from 0 to 483 CTCs. From the MBC cohort, nine mice were sacrificed after tumour development (between 2.5 and 8 months after injection) and the pathology analysis revealed that all the generated tumours were diffuse large B-cell lymphomas. The other three mice from this cohort were sacrificed 8 months after sample injection without tumour development. The mouse injected with CTCs from patient #20 developed a carcinoma within 3 months after injection and is establishment is described in the next section.

2.2. CTCs-derived xenograft development from a triple negative breast cancer patient is representative of primary tumour

In order to establish the TNBC CDX, the PBMCs fraction from the V2 (with low percentage of CD45+ cells) of patient #20 was injected subcutaneously in an immunocompromised NUDE mouse. At this time point of the course of the disease, the CellSearch enumeration was 969 CTCs, including 74 CTCs clusters (ranging from 2 to 7 cells) (Methods, Figure 6).

Five months after injection, the mouse was sacrificed and the tumour was removed (CDX1) (Figure 22A). Part of the tumour was subsequently passaged to 2 Scid Beige mice (CDX2, CDX2M) (Figure 22B-D) and monitored over time. In the orthotopic xenograph (CDX2M) blood was collected at mouse sacrifice and analysed by CellSearch detecting a CTCs cluster (Figure 22C), which revealed the invasiveness potential of these tumour cells.

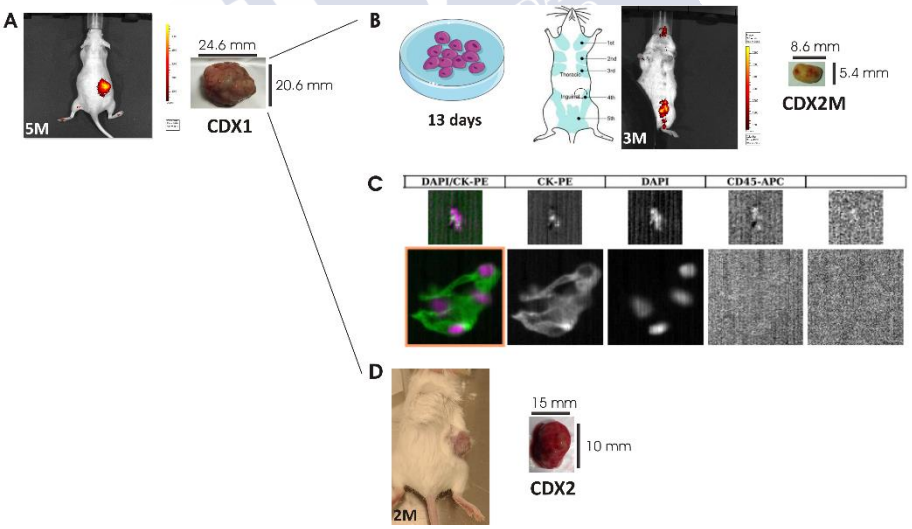

Figure 22. CDX generation. (A) Tumour growth evaluated by *in vivo* imaging 5 months after CTCs injection and tumour development of CDX1. (B) Orthotopic injection of tumour cells disaggregated from CDX1 leads to tumour growth in the mammary fad pat (CDX2M). *In vivo* image of tumour tracking using 2-DG-750. (C) CTCs cluster obtained after processing CDX2M mouse blood using

CellSearch. (D) Macroscopic image of CDX growth after subcutaneous implantation of part of CDX1 tumour (passage 2: CDX2).

The histopathological analysis revealed that the three CDX tumours were poorly differentiated carcinoma specimens with high proliferative activity (high Ki67 expression: 40% in CDX1, CDX2 and CDX2M). This analysis matched with the primary tumour molecular features: negative for CD45, ER and PR expression and positive for pan-CKs and CDH1 (Figure 23). Positive expression of EpCAM, N-CAD, ALDH1A1 and SNAI1 was also observed (Figure 23).

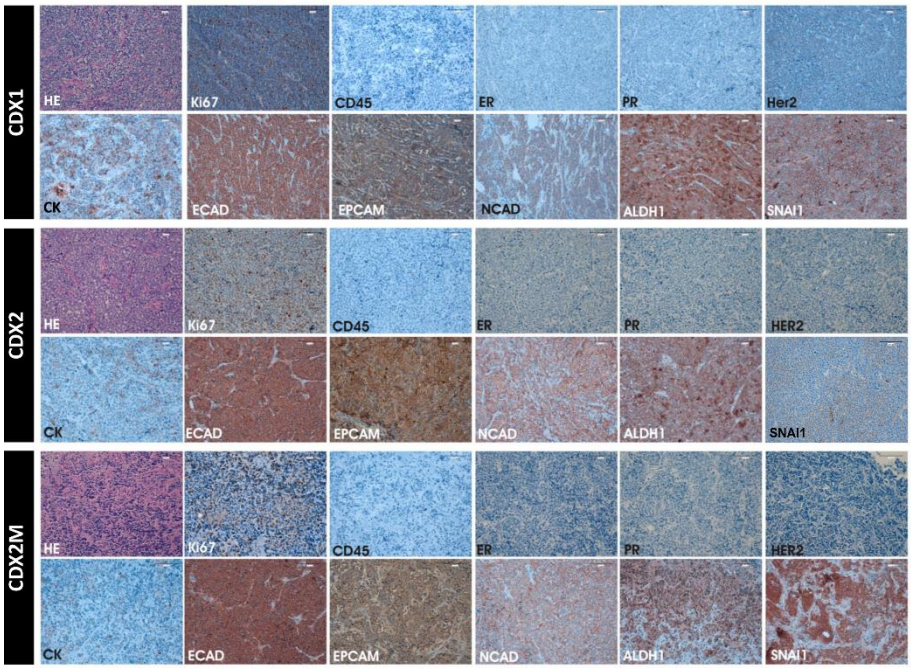

Figure 23. CDX histopathological characterization. Histological characterization of paraffin-embedded from CDX1, CDX2 and CDX2M samples. Haematoxilin-eosin staining and IHC analysis of indicated markers (Scale bar: 100 µm).

Gene expression profile of CDX samples was performed by qPCR to determine similarity among tissue and CTCs samples and to check if tumour passaging modified the genomic profile (Figure 24). Most of the analysed genes did not change through the passages but it was

observed a decrease in *AR*, *CDH1*, *EGFR* and *CCND1* expression rate on the second generation mice. An increase in the expression of *VIM* and a decrease in the expression of *CRIPTO1* were observed in CDX2M and CDX2 respectively.

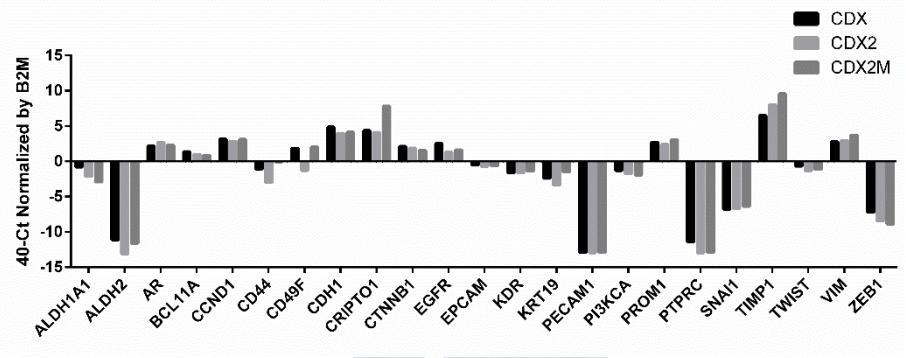

Figure 24. CDX Molecular Characterization. Gene expression analysis of CDX1, CDX2 and CDX2M samples by qPCR. 40-ct normalized by *B2M* is represented.

2.3. Molecular profiling of CTCs during tumour evolution

Due to the establishment of a CDX mouse model we aimed a deeper analysis of patient #20. For that purpose, a longitudinal study was carried out throughout the disease evolution (Methods, Figure 6), collecting blood samples from two different time points of the disease to perform CTCs characterisation by qPCR.

This analysis included a panel of genes related with epithelial (*EpCAM*, *CDH1*, *EGFR*, *CRIPTO1*), mesenchymal (*VIM*, *SNAI1*) and stem cell (*ALDH1A1*, *CD49f*) features, together with tumour progression associated genes (*CD44*, *BCL11A* or *AR*).

CTCs enumeration by CellSearch was 5 CTCs and 969 CTCs/7,5 mL in V1 and V2, respectively. Figure 25 shows the expression level of the analysed genes in both CTCs and tumour tissues (T: primary tumour, M: lymph node metastasis). *VIM* and *CD44* showed high expression in all samples while *EGFR* and *AR* had low levels of expression. CTCs expressed higher levels of *CRIPTO1* than tissue

samples, while the contrary occurs for *SNAI1* expression. *CD49f*, *EpCAM* and *CDH1* increased their expression levels in V2 when compared with V1. However, when comparing CTCs with tissue samples, *EpCAM* expression was higher in primary tumour while *CDH1* expression in V2 was comparable with metastatic samples.

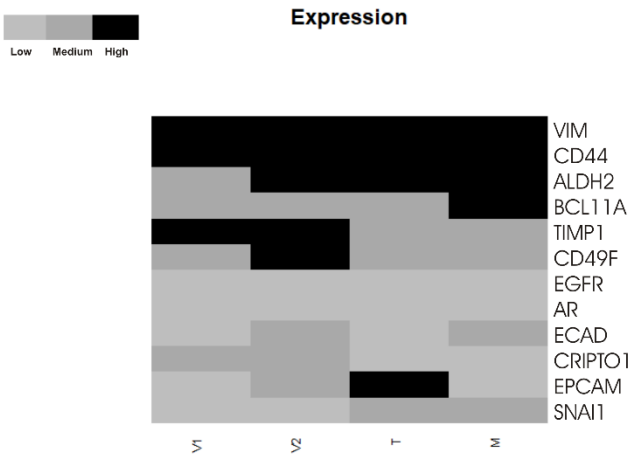

Figure 25. Gene expression analysis. Heatmap depicting qPCR expression levels of a panel of genes implicated in TNBC biology analysed in CTCs isolated at Visit 1 (V1, 5 CTCs counted by CellSearch) and Visit 2 (V2, 969 CTCs counted by CellSearch), primary tumour (T) and metastatic tissue (M). Expression levels were determined based on 33.3 and 66.6 percentiles [Low level (clear grey) account for 0 to 33.3 percentiles; medium level (medium grey) for 33.3 to 66.6 and high level (black) for 66.6 to 100 percentile].

2.4. WNT pathway role in tumour progression in a CTCs-derived xenograft case

RNA-seq analysis was performed to determine whether CDX samples shared molecular characteristics with patient samples and to identify potential pathways involved in tumour progression in TNBC. For that, normal tissue (N), primary tumour (PT) and two different metastatic sites (M1 and M2: lymph nodes) from FFPE samples and CDX samples (fresh tissue) were included.

First, primary tumour was compared with healthy tissue and, after removing lower expressed genes (<100 reads), genes showing a log2 Fold Change |2| were selected as specifically altered due to tumour transformation. These genes behaviour in all samples was interpreted using a heatmap (Figure 26A), and a principal component analysis (PCA) was carried out to understand how the tumour samples clustered based on their differences with the control (Figure 26B). CDXs samples grouped together near the primary tumour.

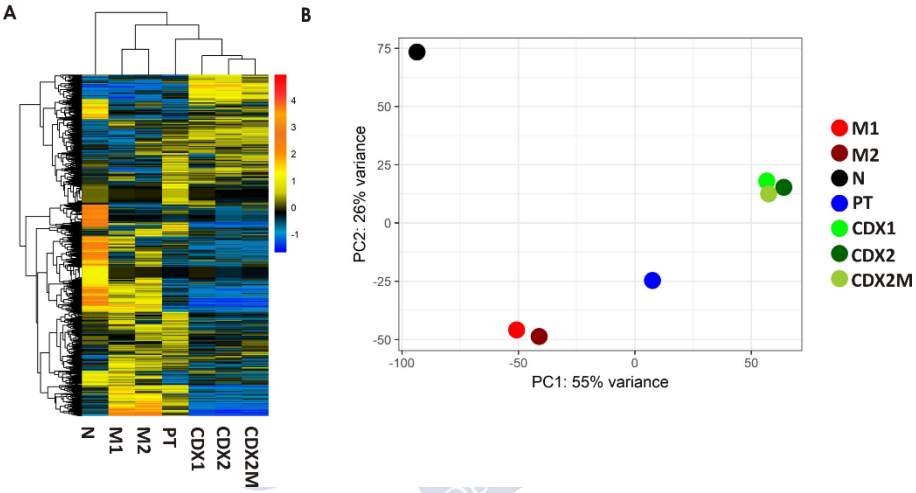

Figure 26. RNA sequencing analysis. (A) Clustered Heatmap depicting RNA-normalized expression levels for genes Fold Change |2| to normal sample. (B) PCA of indicate samples: normal tissue (N, in black); primary tumour (PT, in blue); metastatic sites (M1 and M2, in red and brown respectively); and CDX1, CDX2 and CDX2M (in green).

A set of 3401 up-regulated genes and 2372 down-regulated genes were obtained from this analysis and represented through Venn diagrams. These graphs, which show those genes shared among samples, are represented in Figure 27. Thus, CDXs tissues have 1080 up-regulated genes in common among the three tumours (CDX1, CDX2 and CDX2M). M1 and M2 showed 1235 (36.3%) common up-regulated genes. Then, those common genes were compared between CDXs and metastases, finding 433 (23%) mutual genes (Figure 27A). We also found 1706 (71.9%) common down-regulated genes in CDX

tumours, while 1079 (45.5%) down-regulated genes were shared between metastatic samples. Comparative analysis of common genes inferred 823 (41.9%) mutual genes (Figure 27C).

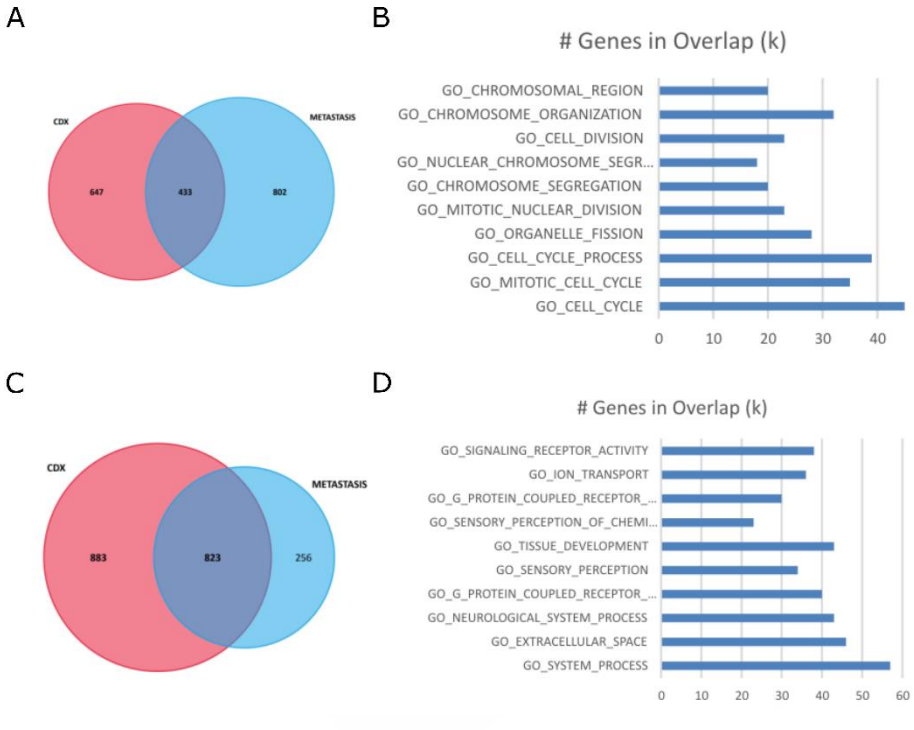

Figure 27. Venn Diagram analysis. (A) Overlapping up-regulated common genes between CDXs and metastases. (B) GO analysis denoting the main pathways associated to those common up-regulated genes (from Venn diagram A). (C) Venn Diagram showing common down-regulated genes between CDXs and metastases. (D) GO analysis denoting the main pathways associated to those common down regulated genes (from diagram C).

Amongst up-regulated genes in both metastasis and CDX (433), cell cycle genes were strongly represented (Figure 27B), while down-regulated genes (823) accounted for different general cell functions such as system process or tissue development (Figure 27D).

Next, we performed a GO analysis. We included the up-regulated genes from all samples compared to normal tissue. Amongst those

genes, WNT signalling was the main pathway involved in the biology of all samples (Figure 28).

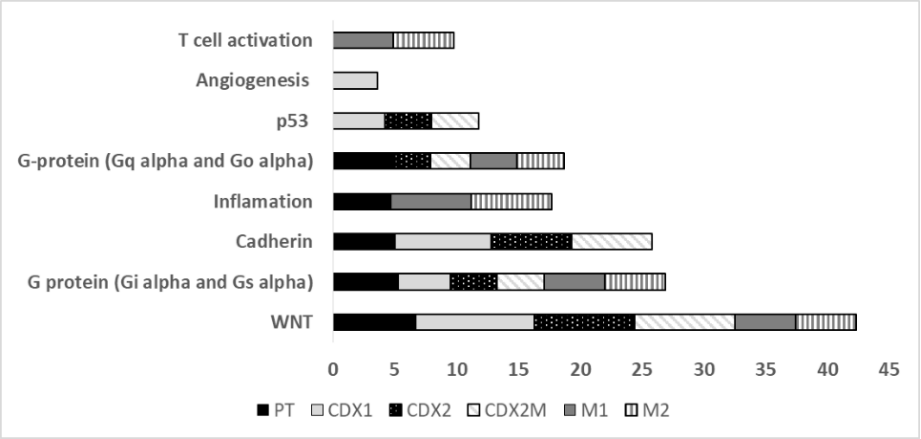

Figure 28. GO analysis denoting the main pathways associated to the indicated samples (Primary tumour (PT), CDXs (CDX1, CDX2 and CDX2M) and metastases (M1 and M2). Pathways corresponding names: WNT signalling pathway (P00057), Heterotrimeric G-protein signalling pathway-Gi alpha and Gs alpha mediated pathway (P00026), Cadherin signalling pathway (P00012), Inflammation mediated by chemokine and cytokine signalling pathway (P00031), Heterotrimeric G-protein signalling pathway-Gq alpha and Go alpha mediated pathway (P00027), p53 pathway (P00059), Angiogenesis (P00005), T cell activation (P00053).

Further, GO analysis was performed considering only the common genes between the three CDX, the two metastatic samples and both. Again, WNT pathway was the most represented followed by Cadherin signalling pathway (Figure 29). Regarding down-regulated genes, GnRH, inflammation and WNT pathway stood out from the others.

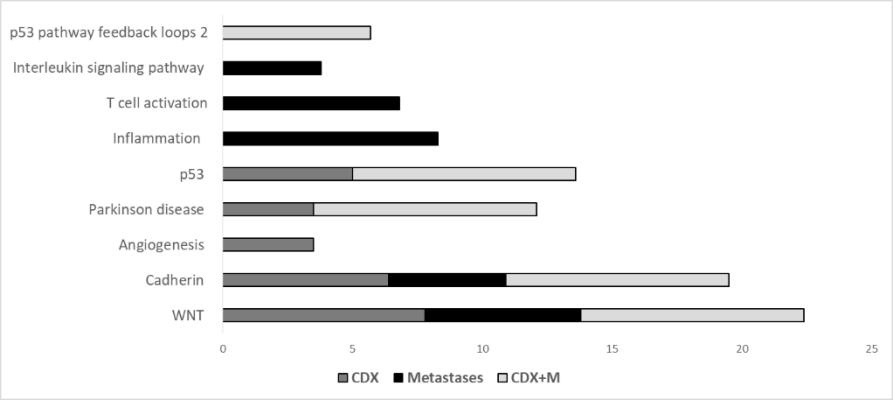

Figure 29. GO analysis. Gene ontology analysis denoting the main pathways associated to CDXs common genes (CDX, dark grey), metastases common genes (Metastases, black) and shared genes among CDX and metastasis (CDX+M, in light grey). Pathways corresponding names: WNT signalling pathway (P00057), Cadherin signalling pathway (P00012), Angiogenesis (P00005), Parkinson disease (P00049), p53 pathway (P00059), Inflammation mediated by chemokine and cytokine signalling pathway (P00031), T cell activation (P00053), Interleukin activation pathway (P00036) and p53 pathway feedback loop 2 (P04398).

Finally, SNPs analysis demonstrated that the three CDXs were very homogeneous since they shared 68.4% of the deleterious polymorphisms found (Figure 30A). This group of common SNPs was selected for comparison with the other samples, showing important tumour heterogeneity among them, especially marked between M1 and M2, which only share 1.4% (Figure 30B). While these metastases only share 3 SNPs with the primary tumour, the CDXs samples have 13 SNPs in common with it. There was only one deleterious mutation shared by all the analysed samples, which was situated in the Cyclin I gene (CCNI, g.77058527A>G) (Figure 30C).

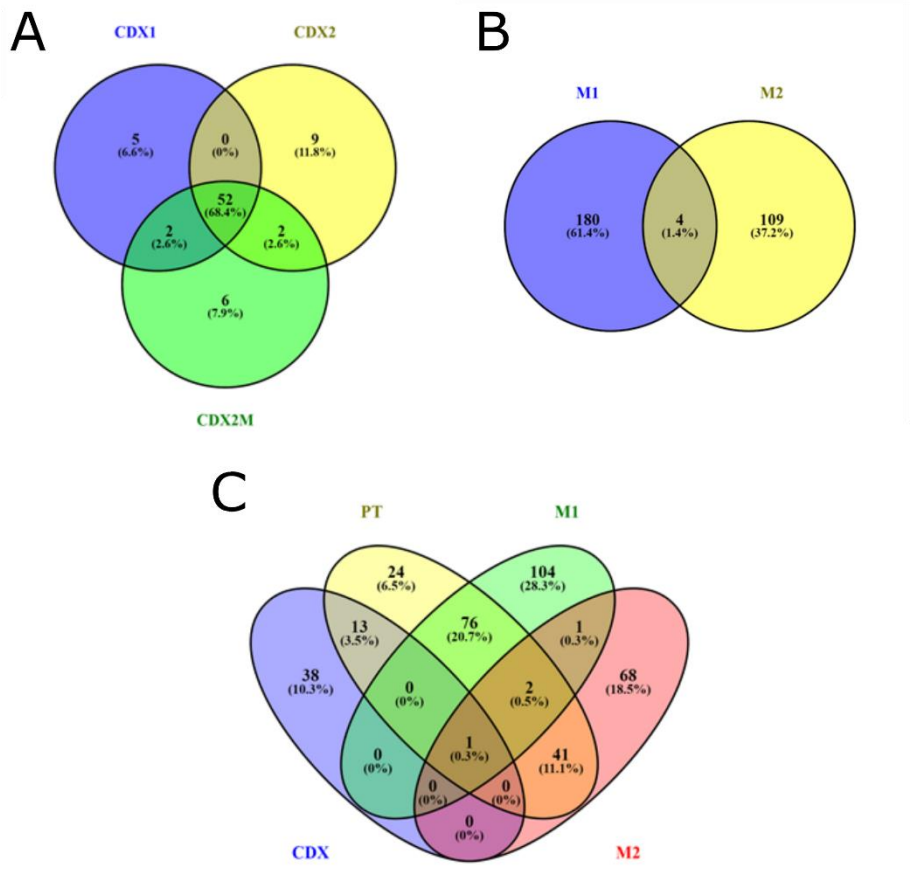

Figure 30. SNPs Analysis. Venn Diagrams showing common SNPs among samples: CDXs (A), metastasis (B) and common SNPs between CDX and primary tumour and metastatic samples (C).

2.5. MELK: a prognostic marker for triple negative breast cancer identified from CTCs-derived xenograft molecular characterization

A panel, from the total of differentially expressed genes, was selected for a further validation based on their representation of the

main pathways identified by GO analysis (Figure 28 & 29), their expression level in RNA-seq analysis and their involvement in breast cancer development. Genes were analysed by qPCR in the CTCs population isolated from the V1 and V2 in patient #20. Gene expression of 5 selected genes (*AURKB*, *HIST1H4A1*, *MELK*, *MYCL* and *PCDHA8*) was detected in both sampling points and further analysed in CTCs from the cohort of TNBC patients (n=32, which includes patient #20) and healthy donors (n=22). The *AURKB*, *HIST1H4A1*, *MELK* and *PCDHA8* genes were more expressed in patients than in controls ($p < 0.05$) (Figure 31) demonstrating their presence in CTCs.

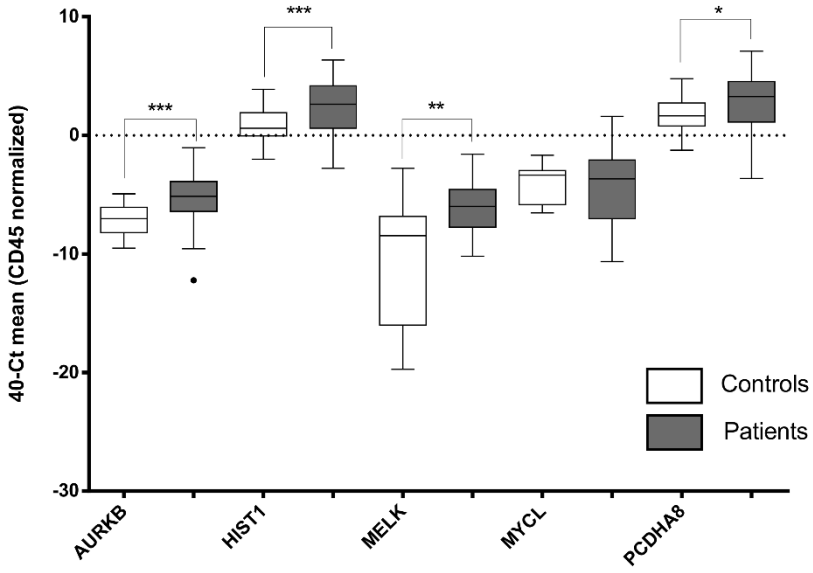

Figure 31. Gene expression levels of *AURKB*, *HIST1H4A1*(HIST1), *MELK*, *MYCL* and *PCDHA8* genes in a cohort of 32 patients (grey) and 22 controls (white) analysed by qPCR.

In addition, we explored the prognostic potential of our CTCs markers by Kaplan–Meier survival analysis (Table VIII) and we found that *MELK* overexpression was statistically associated with shorter OS and PFS rates (Figure 32, Table VIII). Besides, although differences

were not statistically significant, patients with high levels of *AURKB*, *HIST1H4A1* and *PCDHA8* showed also lower survival rates.

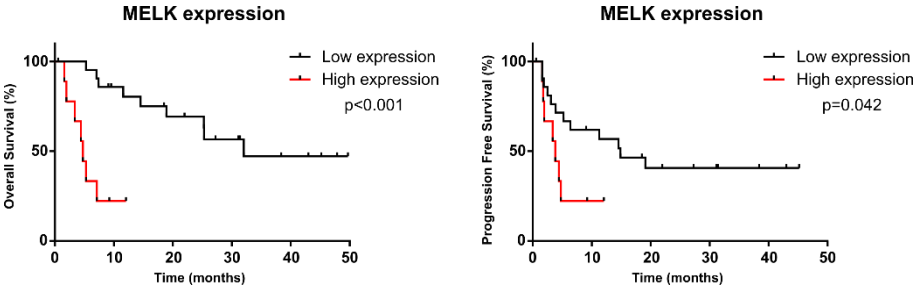

Figure 32. Survival curves for OS and PFS according to the *MELK* expression levels in the cohort of 32 TNBC patients (5.85-33.1 months, p<0.001 and 5.09-22.8 months, p= 0.042).

Table II. Survival data analysis					
Marker	n	Overall survival (OS)		Progression free survival (PFS)	
		mean (95% CI)	p	mean (95% CI)	p
AURKB					
<p70	23	28,74 (20,09 - 37,39)	0,265	20,31 (11,94 -28,68)	0,54
>p70	9	16,12 (8,60 -23,64)		10,02 (4,19 - 15,85)	
HIST1					
<p70	23	28,94 (20,33 - 37,55)	0,223	20,76 (12,43 - 29,09)	0,437
>p70	9	16,12 (8,60 - 23,64)		10,02 (4,19 - 15,85)	
MELK					
<p70	23	33,10 (25,11 - 41,08)	<0,001	22,80 (14,47 - 31,13)	0,042
>p70	9	5,85 (3,45 - 8,35)		5,09 (2,56 - 7,62)	
MYCL					
<p70	23	34,96 (26,25 - 43,68)	<0,001	25,48 (16,70 - 34,25)	0,003
>p70	9	10,36 (3,88 - 16,83)		5,69 (1,89 -9,48)	
PCDHA8					
<p70	23	27,24 (18,63 - 35,85)	0,801	19,61 (11,51 - 27,70)	0,726
>p70	9	19,83 (11,91 - 27,75)		13,69 (4,72 - 22,65)	

Table VIII. Prognosis value of validated markers in CTCs: *AURKB*, *HIST1H4A1* (HIST1), *MELK* and *PCDHA8* in the TNBC patient cohort (marker expression values were grouped according to a 70 percentile value).

3. VALIDATION OF A NEW MICROFLUIDIC CELL FILTER FOR CTCS ISOLATION WITH UNPROCESSED WHOLE BLOOD

3.1. Comparative analysis: Isolation of CTCs by CROSS chip versus CellSearch

A microfluidic filter device, the CROSS chip, was designed and developed by the INL International Iberian Nanotechnology Laboratory aiming a rapid and unbiased isolation of CTCs. Considering its reported good performance in spiking experiments by the INL, we tested its pre-clinical functioning. To that aim, 7.5 ml blood samples from metastatic colorectal cancer patients were collected, split in half, loaded in two syringes, and run simultaneously in two CROSS chips. In parallel, another set of 7.5 ml blood samples from the same individuals were collected simultaneously and subjected to CellSearch test.

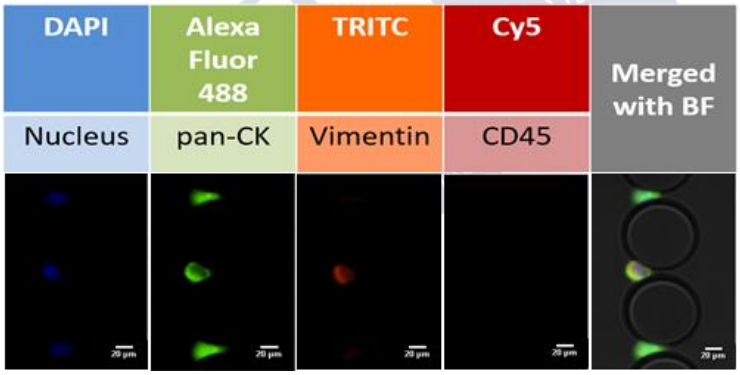

Figure 33. Immunofluorescence image of CTCs from colorectal cancer patients retained in the CROSS chip.

Immunofluorescence staining was used to identify captured CTCs in the CROSS chips, by detecting nucleated, morphologically intact DAPI+/CK+/CD45- cells (Figure 33). Cells which were positive for VIM and negative for CD45, as well as CK+/CD45- cell clusters were also observed retained in the CROSS device, but not considered for CTCs enumeration. Seven out of nine patient samples analysed

showed ≥ 3 CTCs/7.5 ml of whole blood (mean value = 20.28 ± 14.3) by the CROSS chip. In contrast, none of the patients scored ≥ 3 CTCs/7.5 ml of whole blood by CellSearch (Wilcoxon test, p= 0.0039) (Figure 34). No CTCs were detected in the blood of two healthy donors using the CROSS chip.

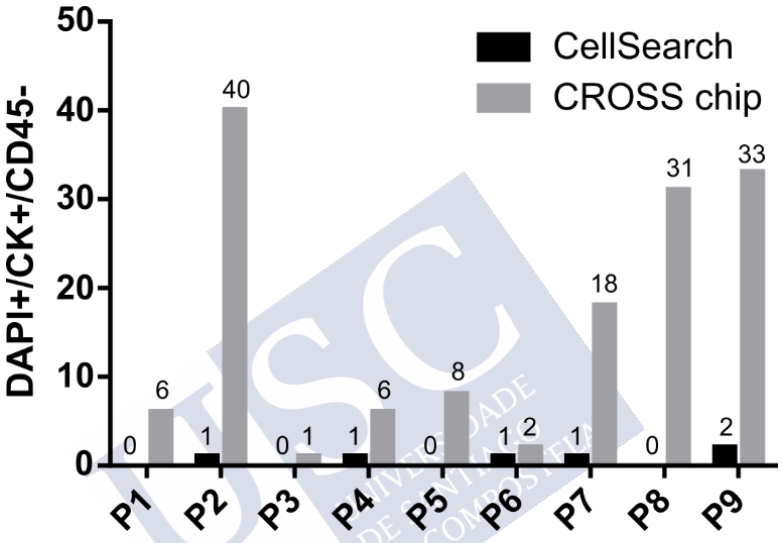

Figure 34. Comparative bar chart demonstrating the enumeration of DAPI+/CK+/CD45- cells (CTCs) using the CellSearch System versus the CROSS chip, for the nine patients included in this validation assay.

3.2. Detection of *APC* mutations by ddPCR in CTCs isolated with the CROSS chip

In order to evaluate the origin of the cells isolated using the CROSS chip and to assess its capability to perform downstream analyses, CTCs were screened for the most common DNA mutation of the *APC* gene (c.4348C>T), which is highly frequent in colorectal cancer patients. Due to the limited amount of starting genetic material available, this analysis was performed by ddPCR. This *APC* mutation was found in 7 out of the 9 patients analysed, which confirmed the tumour origin of the cells isolated by the CROSS chip (Figure 35).

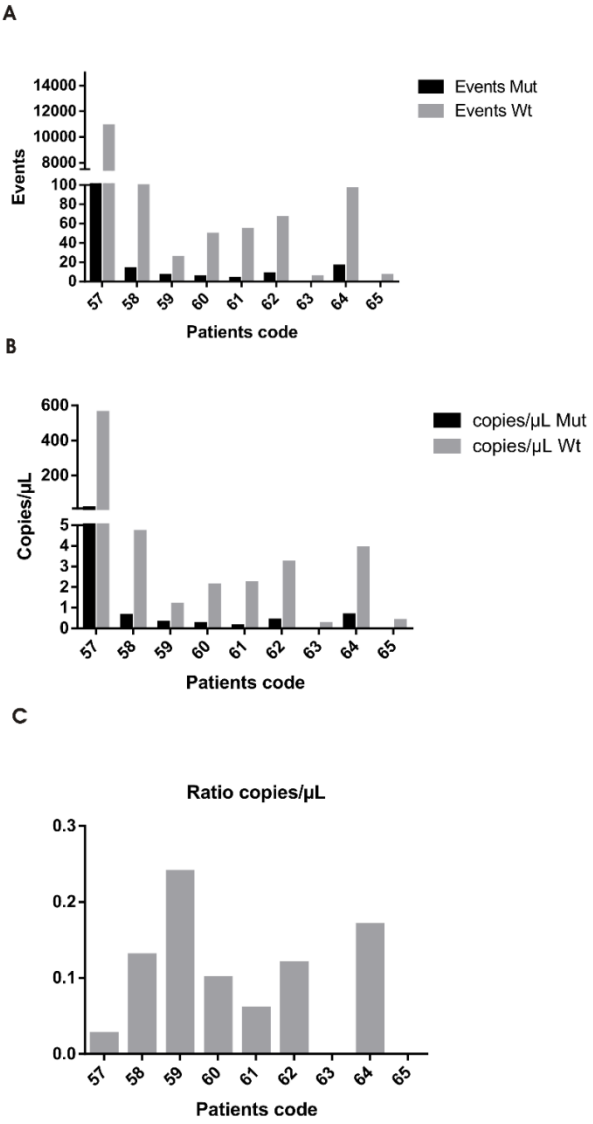

Figure 35. ddPCR analysis for *APC* mutation c.4348C>T for the 9 analysed patients. A. Number of positive events from QuantaSoft Version 1.7.4.0917 for wild type (WT) and Mutant (MUT) *APC* gene. B. Copies/µL for mutant and wild type *APC* gene. C. Ratio of copies /µL mutant to wild type, mean: 0.09 ± 0.08. Patients' codes P7 and P9 did not present *APC* mutation.

3.3. Clinical data correlation and overall survival

The number of CTCs enumerated by the CellSearch were less than 3 CTCs/7.5 ml of whole blood for all samples analysed (Figure 34), below the pre-established cut off for colorectal cancer using the CellSearch technology. Considering these data, patients could not be divided in different prognostic groups and all were classified as having good prognosis.

However, with the CROSS chip, the CTCs number obtained was higher in every patient and thus a possible correlation between CTCs enumeration and disease prognosis was investigated. Patients were grouped in good or bad prognosis according to the number of isolated CTCs by the microfluidic device and using the cut off value defined by CellSearch (< or ≥ 3 CTCs/7.5 ml of whole blood respectively). As illustrated in Figure 36, and according to a Kaplan Meier analysis, a trend for shorter PFS was observed for patients with ≥ 3 CTCs/7.5 ml of whole blood than those with < 3 CTCs/7.5 mL blood, although it was not statistically significant (p =0.381).

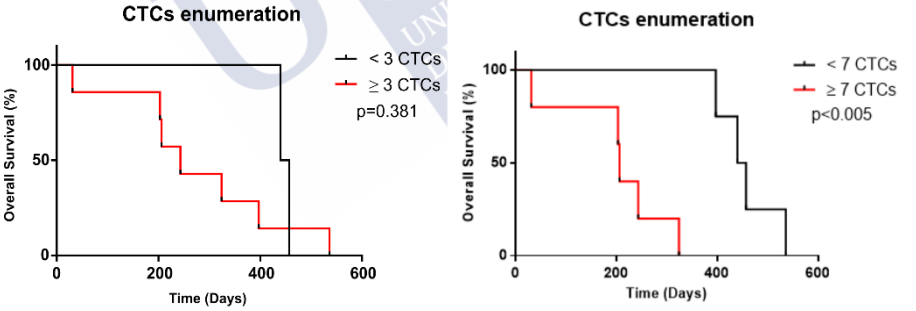

Figure 36. Kaplan-Meier plot of OS based on CTCs isolation with CROSS chip. In the left side, cut off ≥ 3 CTCs (p= 0.381); on the right side cut off ≥7 CTCs (212 days, p<0.005).

Remarkably, defining an alternative cut off of ≥ 7 CTCs/7.5 ml of whole blood, the CROSS chip was able to discriminate patients with good prognosis from those facing an unfavourable outcome (CTCs ≥ 7) (p= 0.005), with a greater survival of 212 days (Figure 36).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DISCUSSION

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DISCUSSION

The interest in liquid biopsy, as a tool for cancer monitoring prior to and/or during therapy and even in early disease detection, has grown considerably in the last years, since peripheral blood sampling is easy and can be repeated when needed. In this context, CTCs can provide valuable information for the clinical management of cancer patients. However, CTCs detection is hampered by their molecular heterogeneity and low ratio in the peripheral blood.

CellSearch is the only FDA-approved method for CTCs quantification, with a proved prognostic value in a number of different epithelial cancers[16]. However, it is based on the expression of epithelial markers (EpCAM, CK8, CK18 and CK19). Therefore, the results obtained with this methodology have limitations derived from the inability to isolate different CTCs phenotypes. Besides, the existence of tumour heterogeneity and the phenotypic changes promoted during the EMT process that allows the dissemination from the primary tumour and metastasis, hinders the selection of appropriate CTCs markers. Furthermore, several studies have reported that EpCAM expression is decreased in some aggressive breast cancer cell lines (*i.e.* SK-BR-7, MDA-MB-231, BT549), suggesting that EpCAM-based CTCs detection may be insufficient[128,129].

We aimed to isolate CTCs, avoiding the bias introduced by marker selection, and also to characterise them at a molecular level, before and after treatment, in order to study their clinical significance in MBC patients. For that purpose, we used a gradient density centrifugation in combination with a negative enrichment protocol. With this approach, it is possible to enrich the samples in all CTCs phenotypes, including epithelial CTCs but also CTCs in a more mesenchymal or stem state[130]. In addition, we analysed the different molecular subtypes of breast cancer collectively in order to detect biomarkers related to tumour progression or with a clinical interest in a subtype independent manner.

For CTCs analyses, 20 patients with MBC were recruited. Samples were analysed using CellSearch for CTCs enumeration, and in parallel, the molecular expression of a custom panel of different markers was analysed by qPCR. CTCs enumeration and their molecular features were correlated with patient clinical parameters and outcomes.

Our results showed that patients with ≥ 5 CTCs had a shorter OS considering the metastatic diagnose time point (before treatment, V1) matching and further supporting previously described reports[34]. Interestingly, the enumeration of ≥ 5 CTCs in patients after one cycle of treatment (V2) was also significantly associated with worst OS. Although, several studies have evaluated the prognostic value of CTCs enumeration in breast cancer patients[34,131,132], only a small number of them have explored the prognostic relevance of CTCs numbers before and after chemotherapy describing that the presence of persisting CTCs after chemotherapy was associated with worse outcome and that a decrease in the CTCs count during the follow-up is an early marker of individual response[57,133].

In this study we found that just a 45% of the analysed patients showed ≥ 5 CTCs (counted by CellSearch) before treatment. This is a slightly lower percentage than the previously described by Lianidou and colleagues (50-70%)[134]. This discrepancy could be explained by the differences within cohorts in subtype proportions. More specifically because our cohort included a high proportion of TNBC cases, which usually present lower EpCAM+ CTCs counts due to their mesenchymal or stem features[135].

Regarding the overall distribution of CTCs in different subtypes, our data revealed that a larger proportion of patients in the luminal subset were CTCs+ by CellSearch, compared with the other subtypes of tumours. These results match with what was previously described by Giordano and colleagues[136].

Besides the prognostic impact of CTCs enumeration, the molecular characterisation of these cells offers new perspectives that can increase the understanding of CTCs biology, and could enable

better clinical decisions, which nowadays are not possible through enumeration. In this sense, some studies have explored CTCs on breast cancer by using molecular methods[137–144]. However, in a small number of these studies, researchers have evaluated the characteristics of CTCs after therapy[57,142,145–147], and, to our knowledge, only one has used a negative enrichment approach comparable with ours[148].

To avoid background expression from leukocyte in the CTCs enriched fraction, we verified the heterogeneity that exists in the PBMCs among patients regarding gene expression values. In fact, previous studies have reported specific gene expression changes between breast cancer patients PBMCs[149]. Thus, CTCs expression values were normalised by their autologous PBMCs.

Similarly to previous studies, our results showed a 70% concordance between *ERBB2* expression on CTCs and HER2 status on the primary tumour[150]. However, we reported a higher detection rate on epithelial, EMT and stem markers expression on CTCs, compared with previous articles[139,141,148,151] that could be due to the isolation method of choice. In addition, comparably with Aaltonen and colleagues, we observed discordances in the *ER* expression between the patient's solid tumour and CTCs[152], which emphasises the dynamic nature of tumours and the need for real-time monitoring in cancer patients.

Our molecular analyses revealed that the expression of specific epithelial (*EpCAM*, *KRT19*) or breast cancer related genes (*ERBB2*) was associated with the presence of ≥ 5 CTCs at V1. Furthermore, in patients with one or more CTCs, after one cycle of chemotherapy, we found association with the expression of *CDH1*, the gene encoding for E-Cadherin, a commonly used marker by the pathologists to distinguish between lobular from ductal carcinomas[153]. In addition, we found an association between the overexpression of *CCND1* and the detection of ≥ 5 CTCs at V2. *CCND1* oncogenic capacity has long been established in breast cancer, and its overexpression in transgenic mammary tissues has been linked with mammary hyperplasia and tumours[154]. Thus, our data suggest that the CTCs negative enrichment methodology allows the detection of specific markers of CTCs and

breast cancer, even in CellSearch non positive cases, demonstrating that this approach might overcome some of the CellSearch limitations.

Correlation analysis between gene expression and breast cancer subtypes showed that the expression of *KRT19* was higher in luminal A subtype and lower in TNBC subtype. These results can be explained by the fact that *KRT19* encodes for CK19 protein, an epithelial marker which can be downregulated during the EMT process[155], hence it is lower expressed in the TNBC subtype[156]. Furthermore, Bredemeier and colleagues have described *KRT19* as a powerful marker to identify CTCs[147]. However, *ALDH1A1* expression, a marker for stem-like CTCs, was higher on the luminal B and TNBC subtype. *ALDH1A1* has been suggested to characterise a more aggressive population of CTCs that might be associated with therapy failure[144], so, although our patient cohort is relatively small, these results suggest that ALDH1 might be useful for the detection of CTCs in the cases in which the detection of EpCAM+ cells is limited. Thus, combining the analysis of both markers we could identify CTCs in all the different breast cancer subtypes.

Analysing clinical data more comprehensibly, we found an association between a higher tumour grade at diagnosis, and an overexpression of *KRT5* in CTCs at V1, which is the gene that encodes CK5. CK5 breast cancer cells have enhanced mammosphere forming potential and are endocrine and chemotherapy resistant in MBC[157]. Albeit, *KTR5* expression has not been reported yet on CTCs.

In addition, we found that primary tumours with different hormonal receptor characterisation lead to CTCs with different expression profiles. If we consider that PR positivity on the primary tumour predicts sensitivity to endocrine therapy, we could associate the high expression of *BCL11A*, *KRT5* and *RB1* with resistance to this therapy. In fact, *BCL11A* has been identified in aggressive subtypes of breast cancer, and an overexpression of *BCL11A* drives the development and progression of TNBC[158]. Regarding CK5, Dairkee and colleagues reported for the first time the possible poor survival or early recurrence associated with the expression of CK5 in tumour cells in 1987[159], and although the functional role of the CK such as CK5,

CK14 or CK17 is still unknown, it is clear that their expression is associated with poor prognosis. In the case of *RB1*, the retinoblastoma susceptibility gene, although it was the first tumour suppressor gene to be molecularly defined, its protein product, pRB, has recently been linked with cell-to-cell and cell-to-extracellular matrix interactions[160]; therefore its high expression in CTCs is in concordance with a more aggressive behaviour, hence poor outcome. Regarding the high *RB1* expression observed in ER+ patients, it is well-known that pRB is fundamental for ESR1 activity, and its loss decreases the expression of ESR1 protein[161]. Thus, a high expression of *RB1* in CTCs could predict endocrine therapy response in ER+ patients. Concerning *GDF15*, although its signalling pathway is poorly understood and the "canonical" pathway is unknown, it has been reported that exogenously added *GDF15* induced the formation of tumour spheres in primary cancer cells derived from luminal breast cancer tissues[162]. The authors suggested that *GDF15+* cells represented cells with similar features to cancer stem cells, therefore, CTCs with high expression of *GDF15* after therapy could also be representing a subpopulation with stem cell characteristics in ER + patients.

Next, we studied the correlation of the gene expression profile of the samples with the outcome of the patients. We did not find any association between the primary tumour tissue expression and the outcome of the patients. Nevertheless, we found an association between the expression of *ERBB2*, *PALB2* and *MYC*, on the enriched fraction of CTCs before treatment, with a worse prognosis of the patients, which remarks the potential of CTCs analysis for patients' prognosis. *ERBB2* is also commonly referred as *HER2*; HER2 protein is overexpressed in 20% of breast cancers but overall it is expressed above the healthy breast tissue level in 60% of breast cancers[163]. *HER2* expression in breast cancer CTCs is one of the most extensively studied markers[138,139,143,146,148,151,152,164] and it has been associated with poor prognosis[143]. In addition, discrepancies in *HER2* amplification between CTCs and the primary tumour have already been reported[51,143,151,165]. , demonstrating the additional information that the circulating tumour population can provide to monitor tumour evolution.

Regarding *PALB2*, (Partner And Localizer of BRCA2) mutations on this gene have been associated with increased risk of breast cancer[166], besides, its overexpression in tissue has recently been linked to a worse outcome in patients with advanced breast cancer[167]. Here, we report similar results on CTCs, allowing real-time monitoring of *PALB2* status in the patients.

In the case of *MYC*, it encodes the oncoprotein c-MYC and its overexpression is associated with poor clinical outcome in breast cancer patients[168]. Little is known regarding its expression profile in CTCs; nonetheless, *MYC* inhibitors have been proposed for targeting cancer stem cells in drug-resistant TNBC. In addition, *MYC* has been related with anti-estrogen therapy resistance[169].Thus, here we report that the tracking of *MYC* expression in CTCs is feasible and it might be of interest in breast cancer patients.

When we studied the correlation of the gene expression profile of CTCs with the outcome of the patients after treatment, we identified an association between the expression of *MYC* and *CDK4* with a worse prognosis of the patients. *CDK4* is a cyclin dependent kinase required for cell cycle entry. *CDK4* is a target, together with *CDK6*, of CDK4/6 inhibitors such as Palbociclib, Ribociclib or Abemaciclib[170]. To our knowledge, *CDK4* expression on CTCs has not been reported yet, but our results suggest that monitoring *CDK4* status on CTCs could stratify patients to better or worse prognosis. None of the patients from our cohort of study was receiving anti-CDK4/6 targeted therapy at the time of collection of the samples, therefore, it will be interesting to study the expression levels of *CDK4* in CTCs from patients treated with Palbociclib, Ribociclib or Abemaciclib, before and after treatment.

Altogether, these results highlight the importance of the tumour evolution monitoring during treatment by molecular analysis of CTCs, which may offer new perspectives to clinicians for targeted therapies, allowing them to select or adapt the therapy as early as possible to the clonal tumour evolution.

Unexpectedly, in our analysis we have found that a high expression of *VIM*, a mesenchymal marker, was correlated with a better outcome on the patients. We established an *EpCAM*^{high}*VIM*^{low} signature which was able to predict worst outcome with better significance than *EpCAM* or *VIM* expression alone. Polioudaki and colleagues described that CTCs undergoing EMT acquired mesenchymal morphology, which is associated with full or partial CK19 replacement by Vimentin, which is encoded by the gene *VIM*[155]. The EMT status of CTCs has been a matter of controversy, with some studies pointing to an association between tumour cells with partial EMT state and a worse outcome, when compared with cells which have undergone complete EMT[18,171–173]; in addition, it has been postulated that EMT is not sufficient for metastasis in a number of cancer types[174–176]. However, some of the studies made on CTCs may be biased by the isolation method used, which is mostly based on a selection marker or on a combination of several markers. Interestingly, Markiewicz and colleagues, with a negative enrichment approach based only on anti-CD45 magnetic beads, found that an EMT subtype of CTCs did not have any significantly impact on the survival of early breast cancer patients[148]. Furthermore, a recent study by de Wit and colleagues on breast and prostate cancer patients, reported that, contrary to EpCAM^{high} CTCs, the presence of EpCAM^{low} CTCs in those patients had no relation with OS[177]. In addition, when we included *ALDH1A1* expression in the signature, *EpCAM*^{high}*VIM*^{low}*ALDH1A1*^{high}, its outcome prediction potential was improved. Besides its association with therapy failure, *ALDH1A1* expression in tissue samples has been related with poor prognosis in different breast cancer subtypes[178–181], and its expression in CTCs has also been linked with worse outcome in breast cancer patients[139,142,144,148,164]. Furthermore, Kasimir-Bauer and colleagues suggested that the use of novel agents to attack breast cancer stem cells, like salinomycin and a new synthetic curcumin analogue against ALDH1, could be promising in patients with ALDH1+ CTCs[143].

Our data suggest that an epithelial-stem state of the CTCs may give rise to a more aggressive disease; however, the analysis of CTCs at a single cell level could allow the precise determination of the

epithelial/mesenchymal/stem state of each individual cell, providing better insight into CTCs heterogeneity and its significance in patients' prognosis. In any case, when analysing CTCs in bulk, the chosen detection method for the isolation of CTCs is decisive as some methods may underestimate or neglect some subpopulations of CTCs with putative relevant roles.

Due to the limited number of CTCs positive samples in our study, and the relatively short follow-up of patients, further research with a larger sample size is required in order to confirm the clinical significance of our findings. All in all, our promising results corroborate that CTCs phenotypes can provide clinically important information regarding patients' survival and their stratification into specific clinical studies. The CTCs population is a heterogeneous one with different subpopulations that could lead to different prognosis. Thus, it is important to take all these populations into account by selecting an appropriate isolation method that ensures a representation of all CTCs, like a negative enrichment.

To complement the study of CTCs expression by qPCR, we explored a different approach to delve into the biology of CTCs. We attempt the generation of CDX mice models from breast cancer CTCs. Similarly to other studies[91,93], only a sample with a high count of CTCs, more than 900 EpCAM+ CK+ CTCs (determined by CellSearch) per tube of blood, gave rise to the establishment of a CDX.

We were able to describe for the first time the generation of a CDX mice model from a TNBC patient and contrary to previously described[182], we demonstrate that CTCs from a TNBC patient are tumorigenic and constitute an attractive *in vivo* system to gain a better understanding of tumour biology in this cancer subtype.

Among breast cancer tumours, TNBC subtype is the most challenging due to its aggressive nature, high metastatic potential and lack of targeted therapies[183,184]. Therefore, there is an urgent need to understand the underlying mechanisms involved in TNBC tumours

development, in order to improve the clinical management of these patients.

Histological analysis revealed that the obtained CDX matched with the patient tumour phenotype. Likewise, RNA-Seq analysis also demonstrated that CDX had a common origin with the tumour samples from the patient, confirming that CTCs isolated from blood were tumorigenic, since they were able to reproduce a tumour in a mouse model and subsequent passages. CDX tumours also had metastatic potential, evidenced by the detection of CTCs clusters in mouse blood.

The analysis of different tumour tissue samples, from the three mice passages, and blood from patient #20 over time, allowed us to perform a molecular tracking of the disease. We detected molecular changes among the primary tumour and the metastasis sites, as well as among CTCs (V1 and V2), further supporting the relevance of liquid biopsy monitoring as a valuable tool for understanding tumour evolution.

In addition, when CDX and CTCs were compared at V2, there was a concordance in the gene expression of *BCL11A*, *CD49f*, *CRIPTO1*, *CDH1* and *VIM*. The expression of this set of genes, that includes stem and mesenchymal genes, is similar to that observed in patients with TNBC which, as previously described, despite being an epithelial tumour type it has mesenchymal characteristics[185]. However, since CTCs were obtained using an EpCAM positive enrichment protocol and CDX samples came from a population of CTCs that contained EpCAM+ and EpCAM– CTCs after *in vivo* development, the expression level of the analysed genes is not entirely comparable among CTCs and CDXs samples.

Moreover, the analysis of the CDXs allowed us to identify key molecular mechanisms involved in TNBC development that could represent relevant therapeutic targets. GO analysis pointed to the WNT pathway as the main underlying signalling process up-regulated in all analysed samples. WNT has been associated previously with a higher risk of metastasis and worse prognosis in TNBC patients[186,187].

In addition, *in vitro* studies have shown that the WNT pathway is preferentially activated in TNBC subtypes and may represent a possible therapeutic target to treat this subtype of cancers[188]. Notably, WNT/β-catenin status identifies patients who are most likely to develop lung and brain metastases[189]. In our data, β-catenin (*CTNNB1*) expression is increased in CTCs isolated from a more advanced status of the disease V2, when compared to V1. Moreover, in the three CDX generated, this increase in expression is higher than two-fold when they are compared with healthy tissue. In addition to the WNT pathway, other up-regulated common genes among CDX samples and metastases samples have roles mainly involved in cell cycle regulation, in accordance with the high proliferative activity required to form new tumour locations.

Comprehensive RNA-seq data analyses led to the identification of highly expressed genes on all tumour tissue samples, suggesting their relevance in tumour progression in patient #20. Five genes, selected by their representation on the main pathways identified by GO analysis, were further analysed in CTCs isolated from this patient and in CTCs from a TNBC patient cohort formed by 32 cases. This analysis demonstrated that *AURKB*, *HIST1H4A1*, *MELK* and *PCDHA8* could be potentially used to detect the presence of CTCs, and therefore valuable as indicators of tumour dissemination.

These aforementioned genes have distinct roles in breast cancer development, including the TNBC subtype. *AURKB* is a mitosis-related serine/threonine kinase that is overexpressed in various tumour types such as TNBC[103,190]. Elevated *AURKB* expression contributes to chemoresistance and predicts poor prognosis in breast cancer patients[191], which has led to the development of AURKB inhibitors as anticancer drugs[192,193]. In this sense, the monitoring of AURKB expression in CTCs from TNBC patients could be a good alternative as a tool to study the suitability of this targeted therapy.

Other analysed gene, *HIST1H4A* is a histone cluster member of the H4 family and plays a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability. Lai and colleagues verified that acetylation of histone H4 induces cell

apoptosis and growth arrest by inhibiting AKT signalling in hepatocellular carcinoma[194], however, little is known regarding its role in other tumour types, including TNBC.

PCDHA8 gene is a member of the protocadherin alpha gene cluster. These neuronal proteins are cadherin-like cell adhesion proteins and have a role in the establishment and functioning of the cell-cell connections that take place in the brain[195]. However, its role *in vivo*, gene regulation or its cellular function, have still to be discerned. PCDHA8 is also described to be involved in the WNT pathway and has been suggested as a potential marker for the prediction of breast cancer classification and staging, since it belongs to a set of genes identified through computational analysis of tissue-based gene expression data to identify possible gene signatures and markers of blood or urine proteins[196]. We found this gene expressed in CTCs isolated from TNBC patients' blood samples, however, little is known about its involvement in tumour development or metastasis. Novak and colleagues identified that the *PCDHA* family presents aberrant hypermethylation in breast cancer, being the overall decrease in the expression of these genes correlated with the increase in the CpG islands methylation of *PCDHA* cluster. This could be due to the disruption of the function of the transcription factors and the regulators of genes involved in the control of their expression[197]. Our results in CTCs support these evidences since we found a reduction in *PCDHA8* expression as the disease progressed, probably due to the methylation process.

Finally, *MELK* has been described as an important kinase for the developmental process and has been implicated in mitotic progression, proliferation, apoptosis, differentiation, stem cell phenotypes, and tumorigenesis[198–201]. This gene has been associated with various types of cancer, particularly aggressive malignancies, including TNBC[202–205]. In fact, *MELK* has been described as one of 22 kinases overexpressed in TNBC when compared with other breast cancer subtypes, and this finding has been functionally validated *in vitro*[203]. A study using the breast cancer data set from The Cancer Genome Atlas showed that *MELK* expression was eight-fold higher in tumours

than in normal breast tissue. Besides, *MELK* expression correlates with metastatic recurrence and increased mortality indicating that *MELK* may be predictive of MBC and OS rate[206]. We found that CTCs from TNBC patients were expressing *MELK*, and those high levels of expression are associated with lower OS and PFS rates, resulting in a difference of 27.25 and 17.7 months, respectively. Regarding liquid biopsy studies, a recent publication described *MELK* expression in spiked experiments using the TNBC cell line MDA-MB-231[207]. Supporting our findings, Fina and colleagues detected *MELK* expression in CTCs isolated from 7 MBC patients using AdnaTest EMT-1/Stem CellSelect kit[208]. Besides describing its expression, we were also able to correlate *MELK* expression with OS or PSF in CTCs from TNBC patients. Therefore this work marks a milestone pointing *MELK* as a potential survival marker detected by liquid biopsy, and also a potential therapeutic target with the additional value of already having active *MELK* inhibitors[209].

Taking into account the interesting results obtained for *AURKB*, *HIST1H4A1*, *MELK* and *PCDHA8*, further studies should focus on these molecules and their role as tumour markers as well as their implication in tumour biology and tumour dissemination of TNBC subtype.

Overall, with the development of a CDX mouse model, we were able to integrate CTCs analysis, tissue samples, CDXs generation and RNA-seq technology as a valuable strategy to delve into TNBC biology, providing clinicians with new potential therapeutic targets and markers that could improve the clinical management of these patients. We described for the first time a CDX establishment from a TNBC patient demonstrating that CTCs from these patients could be tumorigenic in mice. Although CDX generation cannot be considered a general approach to improve patient care, it has great value for translational research. Characterising the primary tumour, the metastasis and the CDX we also confirmed the relevant role of the WNT pathway in TNBC, and we identified a panel of markers that can be monitored in CTCs from these patients, providing important

information about their tumour aggressiveness and suggesting their possible role in tumour dissemination.

The molecular heterogeneity of CTCs and its clinical implications highlight the need of improvement of the CTCs isolation methods, which will allow the maximisation of CTCs detection and thereby their further characterisation. In this context, through a collaboration with the INL International Iberian Nanotechnology Laboratory, we were able to clinically validate a new microfluidic technology called the CROSS chip. This system was tested in a clinical setting and compared with CellSearch using as a proof of concept a cohort of metastatic colorectal cancer patients. Lastly, cells isolated using the CROSS chip device were screened by ddPCR for the presence of a specific mutation of the *APC* gene to confirm their malignant origin and to validate the capability of downstream molecular characterisation with this system.

The nine patients´ samples analysed using CellSearch were classified as good prognosis, since the CTC count was below the cut off $\geq$ 3 CTCs[36]. Nonetheless, four of the nine samples (44%) had CTCs only detectable by the CROSS chip, while in the other five samples, a great discrepancy was observed in CTCs enumeration, with CellSearch reporting 1-2 CTCs and the CROSS chip ranging from 2-40 CTCs (average 19.8 CTCs). These results suggest that the isolation of CTCs using the CROSS chip is more efficient and sensitive than CellSearch. Interestingly, VIM+/CD45- cells were also found retained in the CROSS device, indicating entrapment of not just epithelial-like CTCs but also cells with different phenotypes which would improve the CTCs yield. Similarly, a recent study using the Parsortix System described the isolation of mesenchymal-like prostate CTCs, whose number correlated with worse prognosis[210]. Like Parsortix, other microfluidic systems, such as the Vortex[211] and Labyrinth[212] have also reported the capture of heterogeneous CTCs subpopulations expressing epithelial, mesenchymal, EMT and/or cancer stem cell markers. The capacity of the CROSS device to isolate not only single CTCs but also CTC clusters, similarly to other systems such as

Parsortix, holds great potential as these clusters have been correlated with higher invasive capacity[20].

Considering the results obtained in this comparative study with a small metastatic colorectal cancer cohort, due to the higher sensitivity of the CROSS chip, we suggested a higher cut off value than CellSearch for bad prognosis (≥ 7 CTCs/7.5 ml of whole blood). This cut off allows the stratification of patients in 2 defined populations with OS differences higher than 200 days. In contrast, the results obtained by CellSearch were negative for all the analysed samples using the pre-established bad prognosis cut off in colorectal cancer (≥ 3 CTCs/7.5 ml of whole blood), which highlights once again the limitations of this system. However, further studies on larger cohorts of patients are required to clarify the clinical relevance of this method for metastatic patients monitoring and characterisation.

The isolated cells were further characterised to confirm their tumour origin and to assed the competence of the CROSS chip for downstream molecular analysis. To that aim, the mutational status of *APC*, a tumour suppressor gene frequently mutated in sporadic colorectal cancer (up to 60% of colorectal cancer patients)[213], was evaluated. We selected a somatic non-sense mutation with high frequency of mutation among patients population. *APC* mutations were detected by ddPCR in CROSS chip-isolated CTCs in 7 out of 9 patients, even using DNA yields as low as 0,065 ng/µl. As a matter of fact, ddPCR technology has demonstrated high sensitivity to detect clinically relevant mutations at very low concentration in liquid biopsies from patients with different malignancies[214]. Our results are in agreement with the overall frequency of *APC* mutation in colorectal cancer[215], however, false-negative results cannot be discarded due to the low amount of starting DNA. A recent study by Kong and colleagues confirmed that in all colorectal cancer patients analysed, the mutational status of *APC* in both CTCs and the primary tumour matched with 60% concordance[216]; similarly, *APC* mutations were investigated in ctDNA using the BEAMing technology and were detected in > 60% of colorectal cancer patients[217]. CTCs isolated by ScreeCell device from colorectal cancer patients have also been

screened for mutations in the *KRAS* gene using ddPCR, which were observed in 57% of the cases[218].

In summary, several CTCs isolation systems have been described and although a fraction of them have reached commercialisation as automated platforms[219], it should be considered that clinical validation is not always performed. Blood from cancer patients shows different features compared with healthy donors, in terms of density or clotting[220], influencing cell isolation performance of the technologies under investigation. Furthermore, of those studies including patient samples, not all performed a comparison with CellSearch, essential as a positive control to provide a non-biased estimation number of captured CTCs for each sample. The CROSS chip described in this study was able to isolate unfixed cells from unprocessed whole blood with higher sensitivity than the gold standard, capturing cells even in CellSearch non detectable samples. In addition, it allows downstream molecular analyses, which highlights its potential as a powerful tool for liquid biopsy studies.

All in all, liquid biopsy offers a significant opportunity in tumours that are not easy to biopsy and for the restaging and molecular analysis of metastasis. In addition, liquid biopsy diagnosis can serve as a real-time monitoring of tumour status that could tailor the therapy to the individual need of the cancer patient. In this sense, our results prove that CTCs analysis can provide clinically important information, further supporting the relevance of liquid biopsy monitoring as a valuable tool for understanding tumour evolution. Nevertheless, interpretation of the clinical results might be hampered by the fact that the dynamic biology of CTC is still widely unknown.

Our studies highlight the need of CTCs characterisation besides enumeration to provide a more accurate and personalised medicine to the patients. In this context, we tried different approaches. On one side we tested a new microfluidic device that allows downstream analysis of the trapped cells with a higher sensitivity than CellSearch. In addition, we established a preclinical model for the study of CTCs, and, although these models are not a clinical reality due to their establishment timing, its analysis can provide essential information to

delve into the biology of CTCs and to find molecular markers that can be translated into the clinical practice. Finally, we studied CTCs expression throughout the course of the metastatic disease, identifying prognosis markers for subtype independent MBC. All of these approaches, size isolation, magnetic isolation, negative enrichment and the establishment of preclinical models, allowed us to make an approximation to CTCs from different sides, improving the yield of results obtained so far with other technologies and identifying cell markers that could be translated into the clinical practice. Furthermore, different approaches have their advantages and disadvantages, and their combination allows the acquisition of a more complete and complementary information. Challenges for future research include the implementation of CTCs monitoring during systemic therapy, by the standardisation of isolation methods and analyses. The present study provides valuable information to face these challenges with a greater knowledge about most efficient and informative CTCs analysis strategies with value to manage cancer patients.

CONCLUSIONS

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CONCLUSIONS

1. The negative enrichment protocol for CTCs isolation allows the recovery of a wider spectrum of CTCs phenotypes. While CellSearch detects CTCs preferentially in luminal patients, this approach detects CTCs gene expression in all breast cancer subtypes.

2. CTCs characterisation can provide clinical relevant information regarding patients' metastatic disease monitoring and the identification of prognosis markers in metastatic breast cancer patients.

3. High expression of *MYC, PALB2* and *ERBB2* genes in CTCs from breast cancer patients is associated with poor progression free survival and/or overall survival, independently of the molecular subtype. In addition, high expression of *MYC* and *CDK4* after treatment onset is also correlated with poor patients' outcome.

4. Gene expression analyses in breast cancer patients point to epithelial-stem CTCs as more aggressive than mesenchymal CTCs. This state can be monitored with the presence of an $EpCAM^{high}VIM^{low}ALDH1A1^{high}$ signature on CTCs.

5. CDXs technology is feasible in triple negative breast cancer and represents a valuable tool to characterise key steps on tumour progression and to identify new potential therapeutic targets and prognosis markers in CTCs for patient monitoring.

6. The canonical WNT pathway is the main one involved in the triple negative breast cancer case CDX developed. Moreover, high expression of MELK in CTCs from triple negative breast cancer patients is associated with a poor outcome.

7. The CROSS chip device shows higher sensitivity than CellSearch in the detection of CTCs from a metastatic colorectal cancer patient cohort, allowing a better discrimination of patients with poorer prognosis. Furthermore, downstream analysis can be performed to achieve clinical utility with this device.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

APPENDIX

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

APPENDIX

LIST OF PUBLICATIONS

- Pereira-Veiga T, Abreu M, Robledo D, Matias-Guiu X, Santacana M, Sánchez L, Cueva J, Palacios P, Abdulkader I, López-López R, Muinelo-Romay L, Costa C. CTCs-derived xenograft development in a triple negative breast cancer case. *Int. J. Cancer* (2018). doi: 10.1002/ijc.32001. IF: 7.36.
- Ribeiro-Samy S*, Oliveira MI*, Pereira-Veiga T, Muinelo-Romay L, Carvalho S, Gaspar J, Freitas PP, López-López R, Costa C^{+}, Diéguez L^{+}. Fast and efficient microfluidic cell filter for isolation of circulating tumor cells from unprocessed whole blood of colorectal cancer patients. *Under review (Sci Rep).*
- Pereira-Veiga T, Martínez-Fernández M, Abuin C, Piñeiro R, Cebey V, Cueva J, Palacios P, López-López R, Muinelo-Romay L, Costa C. Circulating tumour cells expression for metastatic breast cancer patients monitoring. *Manuscript in preparation.*

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AGRADECIMIENTOS

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AGRADECIMIENTOS

Mis primeras palabras son para los pacientes y sus familiares, gracias por vuestro altruismo y por confiar en nuestro trabajo. Sois la brújula sin la cual esta investigación no tendría sentido.

Quiero agradecer a mi tutor, Rafa López, y a mis directoras de tesis, Clotilde Costa y Laura Muinelo, la confianza depositada en mi hace 3 años, abriéndome las puertas de su grupo y dándome la oportunidad de iniciar mi aventura científica en la Unidad Mixta. Gracias **Rafa** por tu rigurosidad, tus consejos clínicos y por ser el motor de la investigación básica aplicada a la clínica. Gracias **Laura** por tus acertados consejos y tus valiosas contribuciones. Sin ti este proyecto no hubiera visto la luz. Gracias por implicarte hasta el último momento. **Cloti**, gracias por apostar por mí y por descubrirme el mundo de las CTCs. Gracias por haber sido directora, amiga y a veces incluso madre. Por confiar en mi criterio y estar siempre disponible, por escuchar mis ideas por descabelladas que fuesen y además ayudarme a encauzarlas para darles sentido. Esta experiencia no hubiera sido la misma sin una directora como tú. Y por supuesto, gracias al pequeño Artai por prestarme a su madre hasta el final.

Gracias a todo el **Servicio de Oncología** por todo lo que me habéis enseñado de clínica y por vuestra ayuda con las bases de datos, ha sido muy enriquecedor ponerse del otro lado. A **Juan** Cueva por ayudarme a mejorar la calidad del trabajo. A **Cris** y a **Carol**, por siempre estar dispuestas a hacer de vampirillos. Ha sido un placer coincidir con todos vosotros.

Gracias a todo el grupo **ONCOMET** y en especial al **Club del Cactus**. Me da miedo poner nombres porque seguro que me dejo alguien atrás. ¡Habéis sido tantos los que me habéis ayudado durante esta etapa! Tengo la suerte de poder decir que más que compañeros me llevo muchos amigos. Gracias a **Miguel**, **Alex**, **Lore** y **Carol** por vuestros consejos y vuestra ayuda en el animalario. A **Ali** por tu ayuda con las muestras, a **Ramón** por tu carácter, tu buen humor y por tu paciencia con los números de Veridex, a **Manu** y a **Óscar** por todas

vuestros consejos, todos los debates y por cuidarme tanto en Montpellier. A **Patri** y a **Carlos** por todos los buenos ratos y por siempre tener tiempo de un ¿qué tal?, al dúo Pin y Pon, **Aitor** y **Aida** (y **Manu**), con los que siempre he pasado momentos geniales de llorar de risa. A **María** de la Fuente y las **nanogirls**, **Sandra A.**, gracias por estar siempre dispuesta a ayudar en lo que hiciera falta. Gracias al resto, por vuestros consejos y por haber estado ahí siempre que he necesitado ayuda. Gracias a todos por vuestra amistad y compañerismo.

Gracias a toda la gente del IDIS con la que he tenido la suerte de coincidir en estos años. A **Sabela**, **Marta**, **Ángel** y **Adrián**, gracias por todas las conversaciones y toda vuestra ayuda.

Gracias a todos los integrantes de la UM, desde los más nuevos a todos los que han pasado por el laboratorio. Vuestra contribución ha sido esencial para mi aprendizaje, ha sido una suerte poder estar en un grupo tan multidisciplinar. Quiero agradecer especialmente a **Gloria**, **Ana** y **Roberto** toda su disponibilidad, su paciencia, su ayuda y sus consejos. Gracias a mis compis de lab, a **Nuria**, por compartir alegrías, lágrimas, termas y paseos, a **Pablo**, por ser "el hombre tranquilo" y enseñarme a mantener la calma ante todo, e **Inés**, por tu paciencia infinita sobretodo en estos últimos meses (sí, ¡por fin se abrirá el estor!). **Carmen**, gracias por ayudarme tantísimo en todo, esta tesis también es tuya. ¡Os voy a echar mucho de menos pichones!

Quiero extender mi agradecimiento a toda la gente con la que he tenido la oportunidad de colaborar, **Lorena**, **Silvina** y **Marta** del INL, por enseñarme otro tipo de ciencia que ni me hubiera planteado. A **Diego** por tu ayuda con los análisis de RNA-seq y a **Mónica**, por todos los análisis, tu paciencia y tus consejos. A **Susana**, por todo lo que me ha enseñado de proteómica.

Thanks to all the VyCAP team, **Joska**, **Lisa**, **Michiel** and **Arjan** for their warm welcome in the Netherlands, their patience and everything they had taught me. Thanks to all the members of the **Terstappen's Lab** for their advice and support during my stay at

Twente University. All my gratitude to **Agustín**, **Joost** and **Andrea** for the fun times enjoyed at Enschede.

A lo largo de esta aventura he tenido la oportunidad de conocer a personas que se han convertido en grandes amigos. Gracias a todo el **LRT2**, a **Gau**, **Torner**, **Jesús**, **Juani**, **Vane**, **Gardenia**, **Eva M.**, **Fernando**, **Núria**, **Joan**, **Sara**, **Elena**, **Pili**, **Nonia**, **Dani**, **Ester**, **Conxi**, vosotros me metisteis el gusanillo de la ciencia y sois los mejores referentes que un científico puede tener. Me enseñasteis muchas cosas a nivel científico pero aún más a nivel personal. Gracias por ser una gran familia, por vuestros consejos, los ánimos y por hacer que no pase el tiempo. Quiero aprovechar para dar las gracias también a toda la gente con la que coincidí en el **LRT1**, especialmente a **Nadia**, **Gemma**, **Lara** y **María**, ¡cuánto he aprendido de vosotras! Gracias a **Toni**, por enseñarme mejor que nadie que significa la integridad científica.

Rebe, si hay alguien responsable de que haya llegado hasta aquí sin duda eres tú. Gracias por tu confianza, por todo tu apoyo y tus sabios consejos. Algún día retomaremos nuestro proyecto de trabajar juntas. Y por supuesto a **Marc**, nunca he visto a nadie que valore tanto el trabajo científico como tú.

A **Marta B**. y a las "**neuras**". **Izaskun**, gracias por ser mi yang (¡y por descubrirme los *bikinibreaks*!).

Gracias a la Unidad de transplante pulmonar y al laboratorio de neumo del VHIR, gracias por todo lo que me enseñasteis y por apoyarme tanto para empezar esta aventura. A **Román**, **Berta**, **Susana**, **Lola**. A mi gemelier **Marta**, por tantas cosas que no cabrían aquí!

Después de tanto traqueteo no puedo olvidarme de toda la gente con la que he compartido horas de autopista y tertulias. Gracias a los **Pacos**, **Toño**, **Ruth**, **Selina**, **Ali**, **Jandro**, **David** y **Nacho**. **Montse**, gracias por tu apoyo, por los cafés de desahogo y por tener siempre tiempo para echarme una mano. **Bea**, gracias por tus consejos (los científicos y los de repostería), por todos los ánimos y por los debates sobre cuál es el mejor croissant.

Gracias **Eli**, no importa ni el tiempo ni la distancia. T'estimo bonita.

Por supuesto, gracias a **Paula**, a **Nati** y a **Lore**, por todas las aventuras, los conciertos y el apoyo. Gracias **Pablo** por enseñarme a no rendirme nunca. Gracias **Manu** por enseñarme a ser ambiciosa. Gracias **Patri**, por enseñarme a perder el miedo, a **Juan**, a **Diego**, a **Aroa**, a **Moi** y a **Cristian**, por estar siempre ahí.

Gracias a **Ana**, **Julia** y **Arnau**, por todos los cafés, las cañas y los debates. Sois mi oráculo, personal y científico.

Gracias a mi familia por vuestro apoyo infinito. A mis padres, que siempre han puesto mi educación como prioridad en su vida, intentando que fuese feliz haciendo lo que me gusta.

Espero no haberme dejado a nadie atrás y si es así, por favor, no me lo tengáis en cuenta, gracias a todos por formar parte de este viaje.

Y por último, gracias **Dani** por esta aventura que dura ya más de una década. Gracias por tu apoyo incondicional y tu paciencia inagotable esta última temporada, sabes que no habría llegado hasta aquí sin ti. Gracias por saber cuándo necesito un paracaídas y cuando estrellarme. Gracias por volver multicolor los días grises. Gracias por poder hablar de NARS, de MSG o de EMT por igual. Esta tesis es de ambos y ahora ¡hay que celebrarla! Te quiero.

REFERENCES

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

REFERENCES

1. Alix-Panabieѐres, C. & Pantel, K. Circulating tumor cells: Liquid biopsy of cancer. *Clin. Chem.* **59,** 110–118 (2013).

2. Finotti, A. *et al.* Liquid biopsy and PCR-free ultrasensitive detection systems in oncology. *Int. J. Oncol.* **53,** 1395–1434 (2018).

3. Siravegna, G., Marsoni, S., Siena, S. & Bardelli, A. Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **14,** 531–548 (2017).

4. Cohen, A. J. D. *et al.* Detection of surgically resectable cancers with a multi- analyte blood test. *Science.* **3247,** 1–10 (2018).

5. Haber, D. A. & Velculescu, V. E. Blood-based analyses of cancer: Circulating tumor cells and circulating tumor DNA. *Cancer Discov.* **4,** 650–661 (2014).

6. Ashworth, T. A case of cancer in which cells similar to those in the tumors were seen in blood after death. *Aust Med J.* **14,** 146–147 (1869).

7. Pantel, K. & Speicher, M. R. The biology of circulating tumor cells. *Oncogene* **35,** 1216–1224 (2016).

8. Kang, Y. & Pantel, K. Tumor Cell Dissemination: Emerging Biological Insights from Animal Models and Cancer Patients. *Cancer Cell* **23,** 573–581 (2013).

9. Hosseini, H. *et al.* Early dissemination seeds metastasis in breast cancer. *Nature* (2016). doi:10.1038/nature20785

10. Gay, L. J. & Felding-Habermann, B. Contribution of platelets to tumour metastasis. *Nat. Rev. Cancer* **11,** 123–134 (2011).

11. Labelle, M., Begum, S. & Hynes, R. O. Direct signaling between platelets and cancer cells induces an epithelial-

mesenchymal-like transition and promotes metastasis. *Cancer Cell* **20,** 576–590 (2011).

12. Alix-Panabières, C., Pantel, K., Alix-Panabieres, C. & Pantel, K. Challenges in circulating tumour cell research. *Nat. Rev. Cancer* **14,** 623–631 (2014).

13. Meng, S. *et al.* Circulating Tumor Cells in Patients with Breast Cancer Dormancy. *Clin. Cancer Res.* **10,** 8152–8162 (2004).

14. van Dalum, G., Holland, L. & Terstappen, L. W. M. M. Metastasis and Circulating Tumor Cells. *EJIFCC* **23,** 87–97 (2012).

15. Strilic, B. & Offermanns, S. Intravascular Survival and Extravasation of Tumor Cells. *Cancer Cell* **32,** 282–293 (2017).

16. Andree, K. C., van Dalum, G. & Terstappen, L. W. M. M. Challenges in circulating tumor cell detection by the CellSearch system. *Mol. Oncol.* **10,** 395–407 (2016).

17. Allard, W. J. *et al.* Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin. Cancer Res.* **10,** 6897–6904 (2004).

18. Yu, M. *et al.* Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition. *Science.* **339,** 580–584 (2013).

19. Hao, S. J., Wan, Y., Xia, Y. Q., Zou, X. & Zheng, S. Y. Size-based separation methods of circulating tumor cells. *Adv. Drug Deliv. Rev.* **125,** 3–20 (2018).

20. Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158,** 1110–22 (2014).

21. Gkountela, S. *et al.* Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding. *Cell* **176,** 98–112.e14 (2019).

22. Mansilla, C., Soria, E. & Ramírez, N. The identification and isolation of CTCs: A biological Rubik's cube. *Crit. Rev. Oncol. Hematol.* **126,** 129–134 (2018).

23. Sharma, S. *et al.* Circulating tumor cell isolation, culture, and downstream molecular analysis. *Biotechnol. Adv.* **36,** 1063–1078 (2018).

24. Fischer, J. C. *et al.* Diagnostic leukapheresis enables reliable detection of circulating tumor cells of nonmetastatic cancer patients. *Proc. Natl. Acad. Sci.* **110,** 16580–16585 (2013).

25. Andree, K. C. *et al.* Toward a real liquid biopsy in metastatic breast and prostate cancer: Diagnostic LeukApheresis increases CTC yields in a European prospective multicenter study (CTCTrap). *Int. J. Cancer* **143,** 2584-2591 (2018).

26. Gwak, H. *et al.* Progress in Circulating Tumor Cell Research Using Microfluidic Devices. *Micromachines* **9,** E353 (2018).

27. Vona, G. *et al.* Isolation by size of epithelial tumor cells: A new method for the immunomorphological and molecular characterization of circulating tumor cells. *Am. J. Pathol.* **156,** 57–63 (2000).

28. Hvichia, G. E. *et al.* A novel microfluidic platform for size and deformability based separation and the subsequent molecular characterization of viable circulating tumor cells. *Int. J. Cancer* **138,** 2894–2904 (2016).

29. Miller, M. C., Robinson, P. S., Wagner, C. & O'Shannessy, D. J. The Parsortix^{TM} Cell Separation System—A versatile liquid biopsy platform. *Cytom. A.* **93,** 1234-1239 (2018).

30. Park, S. *et al.* Morphological differences between circulating tumor cells from prostate cancer patients and cultured prostate cancer cells. *PLoS One* **9,** e85264 (2014).

31. Yu, M., Stott, S., Toner, M., Maheswaran, S. & Haber, D. A. Circulating tumor cells: approaches to isolation and characterization. *J. Cell Biol.* **192,** 373–382 (2011).

32. Fabbri, F. *et al.* Detection and recovery of circulating colon cancer cells using a dielectrophoresis-based device: KRAS mutation status in pure CTCs. *Cancer Lett.* **335,** 225–231 (2013).

33. Kallergi, G., Politaki, E., Alkahtani, S., Stournaras, C. & Georgoulias, V. Evaluation of Isolation Methods for Circulating Tumor Cells (CTCs). *Cell. Physiol. Biochem.* **40,** 411–419 (2016).

34. Cristofanilli, M. *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N. Engl. J. Med.* **351,** 781–791 (2004).

35. Pantel, K. *et al.* Circulating epithelial cells in patients with benign colon diseases. *Clin. Chem.* **58,** 936–940 (2012).

36. Cohen, S. J. *et al.* Prognostic significance of circulating tumor cells in patients with metastatic colorectal cancer. *Ann. Oncol.* **20,** 1223–1229 (2009).

37. de Bono, J. S. *et al.* Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin. Cancer Res.* **14,** 6302–6309 (2008).

38. Andreopoulou, E. *et al.* Comparison of assay methods for detection of circulating tumor cells in metastatic breast cancer: AdnaGen AdnaTest BreastCancer Select/Detect versus Veridex CellSearch system. *Int. J. cancer* **130,** 1590–1597 (2012).

39. Kulasinghe, A. *et al.* Impact of label-free technologies in head and neck cancer circulating tumour cells. *Oncotarget* **7,** 71223–71234 (2016).

40. Bunger, S., Zimmermann, M. & Habermann, J. K. Diversity of assessing circulating tumor cells (CTCs) emphasizes need for

standardization: a CTC Guide to design and report trials. *Cancer Metastasis Rev.* **34,** 527–545 (2015).

41. Harb, W. *et al.* Mutational Analysis of Circulating Tumor Cells Using a Novel Microfluidic Collection Device and qPCR Assay. *Transl. Oncol.* **6,** 528–538 (2013).

42. Jan, Y. J. *et al.* NanoVelcro rare-cell assays for detection and characterization of circulating tumor cells. *Adv. Drug Deliv. Rev.* **125,** 78–93 (2018).

43. Andergassen, U., Kölbl, A. C., Mahner, S. & Jeschke, U. Real-time RT-PCR systems for CTC detection from blood samples of breast cancer and gynaecological tumour patients. *Oncol. Rep.* **35,** 1905–1915 (2016).

44. Riva, F. *et al.* Clinical applications of circulating tumor DNA and circulating tumor cells in pancreatic cancer. *Mol. Oncol.* **10,** 481–493 (2016).

45. Cohen, S. J. *et al.* Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **26,** 3213–3221 (2008).

46. Bidard, F. C. *et al.* Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. *Lancet. Oncol.* **15,** 406–414 (2014).

47. Gorges, T. M. *et al.* Heterogeneous PSMA expression on circulating tumor cells: a potential basis for stratification and monitoring of PSMA-directed therapies in prostate cancer. *Oncotarget* **7,** 34930–34941 (2016).

48. Paoletti, C. *et al.* Development of circulating tumor cell-endocrine therapy index in patients with hormone receptor-positive breast cancer. *Clin. Cancer Res.* **21,** 2487–2498 (2015).

49. Bidard, F., Proudhon, C. & Pierga, J. Circulating tumor cells in breast cancer. *Mol. Oncol.* **10,** 418–430 (2016).

50. Cardoso, F. *et al.* 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4)dagger. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **29,** 1634–1657 (2018).

51. Zeune, L. *et al.* Quantifying HER-2 expression on circulating tumor cells by ACCEPT. *PLoS One* **12,** e0186562 (2017).

52. Antonarakis, E. S. *et al.* AR-V7 and Resistance to Enzalutamide and Abiraterone in Prostate Cancer. *N. Engl. J. Med.* **371,** 1028–38 (2014).

53. Antonarakis, E. S. & Scher, H. I. Do Patients With AR-V7-Positive Prostate Cancer Benefit from Novel Hormonal Therapies? It All Depends on Definitions. *European urology* **71,** 4–6 (2017).

54. Mostert, B., Sieuwerts, A. M., Martens, J. W. M. & Sleijfer, S. Diagnostic applications of cell-free and circulating tumor cell-associated miRNAs in cancer patients. *Expert Rev. Mol. Diagn.* **11,** 259–275 (2011).

55. Alix-Panabières, C. & Pantel, K. Clinical applications of circulating tumor cells and circulating tumor DNA as liquid biopsy. *Cancer Discov.* **6,** 479–491 (2016).

56. Zavridou, M. *et al.* Evaluation of Preanalytical Conditions and Implementation of Quality Control Steps for Reliable Gene Expression and DNA Methylation Analyses in Liquid Biopsies. *Clin. Chem.* **64,** 1522–1533 (2018).

57. Rack, B. *et al.* Circulating tumor cells predict survival in early average-to-high risk breast cancer patients. *J. Natl. Cancer Inst.* **106,** dju066 (2014).

58. Reeh, M. *et al.* Circulating Tumor Cells as a Biomarker for Preoperative Prognostic Staging in Patients With Esophageal Cancer. *Ann. Surg.* **261,** 1124–1130 (2015).

59. Schulze, K. *et al.* Presence of EpCAM-positive circulating tumor cells as biomarker for systemic disease strongly correlates to survival in patients with hepatocellular carcinoma. *Int. J. Cancer* **133,** 2165–2171 (2013).

60. Gazzaniga, P. *et al.* Prognostic value of circulating tumor cells in nonmuscle invasive bladder cancer: a CellSearch analysis. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **23,** 2352–2356 (2012).

61. Cabel, L. *et al.* Circulating tumor cells: clinical validity and utility. *Int. J. Clin. Oncol.* **22,** 421–430 (2017).

62. Bauer, E. C. A. *et al.* Prevalence of circulating tumor cells in early breast cancer patients 2 and 5 years after adjuvant treatment. *Breast Cancer Res. Treat.* **171,** 571-580 (2018).

63. Trapp, E. *et al.* Presence of Circulating Tumor Cells in High-Risk Early Breast Cancer During Follow-Up and Prognosis. *J. Natl. Cancer Inst.* (2018). doi:10.1093/jnci/djy152.

64. Georgoulias, V. *et al.* Effect of front-line chemotherapy on circulating CK-19 mRNA-positive cells in patients with metastatic breast cancer. *Cancer Chemother. Pharmacol.* **74,** 1217–1225 (2014).

65. Bidard, F. C. *et al.* Clinical application of circulating tumor cells in breast cancer: overview of the current interventional trials. *Cancer Metastasis Rev.* **32,** 179–188 (2013).

66. Helissey, C. *et al.* Circulating tumor cell thresholds and survival scores in advanced metastatic breast cancer: the observational step of the CirCe01 phase III trial. *Cancer Lett.* **360,** 213–218 (2015).

67. Bidard, F.-C. *et al.* Circulating Tumor Cells in Breast Cancer Patients Treated by Neoadjuvant Chemotherapy: A Meta-analysis. *J. Natl. Cancer Inst.* **110,** 560–567 (2018).

68. Wan, J. C. M. *et al.* Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17,** 223–238 (2017).

69. Mouliere, F. *et al.* High fragmentation characterizes tumour-derived circulating DNA. *PLoS One* **6,** e23418 (2011).

70. Ignatiadis, M., Lee, M. & Jeffrey, S. S. Circulating tumor cells and circulating tumor DNA: Challenges and opportunities on the path to clinical utility. *Clin. Cancer Res.* **21,** 4786–4800 (2015).

71. O'Leary, B. *et al.* Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer. *Nat. Commun.* **9,** 896 (2018).

72. Campone, M. *et al.* Buparlisib plus fulvestrant versus placebo plus fulvestrant for postmenopausal, hormone receptor-positive, human epidermal growth factor receptor 2-negative, advanced breast cancer: Overall survival results from BELLE-2. *Eur. J. Cancer* **103,** 147–154 (2018).

73. Kwapisz, D. The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? *Ann. Transl. Med.* **5,** 46 (2017).

74. Schwarzenbach, H. & Pantel, K. Circulating DNA as biomarker in breast cancer. *Breast Cancer Res.* **17,** 136 (2015).

75. Tan, C. R. C., Zhou, L. & El-Deiry, W. S. Circulating Tumor Cells Versus Circulating Tumor DNA in Colorectal Cancer: Pros and Cons. *Curr. Colorectal Cancer Rep.* **12,** 151–161 (2016).

76. Kidess-Sigal, E. *et al.* Enumeration and targeted analysis of KRAS, BRAF and PIK3CA mutations in CTCs captured by a label-free platform: Comparison to ctDNA and tissue in metastatic colorectal cancer. *Oncotarget* **7,** 85349-85364 (2016).

77. Lallo, A., Schenk, M. W., Frese, K. K., Blackhall, F. & Dive, C. Circulating tumor cells and CDX models as a tool for preclinical drug development. *Transl. Lung Cancer Res.* **6,** 397–408 (2017).

78. Goodspeed, A., Heiser, L. M., Gray, J. W. & Costello, J. C. Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. *Mol. Cancer Res.* **14,** 3–13 (2016).

79. Katt, M. E., Placone, A. L., Wong, A. D., Xu, Z. S. & Searson, P. C. In Vitro Tumor Models: Advantages, Disadvantages, Variables, and Selecting the Right Platform. *Front. Bioeng. Biotechnol.* **4,** 12 (2016).

80. Yu, M. *et al.* Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. **345,** 216–220 (2014).

81. Zhang, L. *et al.* The identification and characterization of breast cancer CTCs competent for brain metastasis. *Sci Transl Med* **5** (2013). doi:10.1126/scitranslmed.3005109.

82. Cayrefourcq, L. *et al.* Establishment and characterization of a cell line from human Circulating colon cancer cells. *Cancer Res.* **75,** 892–901 (2015).

83. Gao, D. *et al.* Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159,** 176–187 (2014).

84. Kulasinghe, A. *et al.* Short term ex-vivo expansion of circulating head and neck tumour cells. *Oncotarget* **7,** 60101–60109 (2016).

85. Khoo, B. L. *et al.* Expansion of patient-derived circulating tumor cells from liquid biopsies using a CTC microfluidic culture device. *Nat. Protoc.* **13,** 34–58 (2018).

86. Kolostova, K. *et al.* Molecular characterization of circulating tumor cells in ovarian cancer. *Am. J. Cancer Res.* **6,** 973–980 (2016).

87. Kolostova, K. *et al.* Circulating tumor cells in localized prostate cancer: isolation, cultivation in vitro and relationship to T-stage and Gleason score. *Anticancer Res.* **34,** 3641–3646 (2014).

88. Zhang, Z. *et al.* Expansion of CTCs from early stage lung cancer patients using a microfluidic co-culture model. *Oncotarget* **5,** 12383-97 (2014).

89. Cho, S. *et al.* An Integrative Approach to Precision Cancer Medicine Using Patient-Derived Xenografts. *Mol. Cells* **39,** 77–86 (2016).

90. Pretlow, T. G. *et al.* Prostate cancer and other xenografts from cells in peripheral blood of patients. *Cancer Res.* **60,** 4033–4036 (2000).

91. Baccelli, I. *et al.* Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat. Biotechnol.* **31,** 539–44 (2013).

92. Rossi, E. *et al.* Retaining the long-survive capacity of Circulating Tumor Cells (CTCs) followed by xeno-transplantation: not only from metastatic cancer of the breast but also of prostate cancer patients. *Oncoscience* **1,** 49-56 (2013).

93. Hodgkinson, C. L. *et al.* Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. *Nat. Med.* **20,** 897–903 (2014).

94. Morrow, C. J. *et al.* Tumourigenic non-small-cell lung cancer mesenchymal circulating tumour cells: a clinical case study. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **27,** 1155–1160 (2016).

95. Vishnoi, M. *et al.* Targeting USP7 identifies a metastasis-competent state within bone marrow–resident melanoma CTCs. *Cancer Res.* **78,** 5249–5262 (2018).

96. International Agency for Research on Cancer. Breast Cancer Incidence and Mortality Statistics. **876,** 6–7 (2018).

97. Stewart, B. & Wild International Agency for Research on Cancer, WHO, C. P. (eds. . World Cancer Report (2014). doi:9283204298.

98. Miller, D. L. & Kern, F. G. *NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) Breast Cancer*. *NCCN Guidelines Version 3.2018* (2018). doi:10.1016/S1569-254X(98)80004-6.

99. Scully, O. J., Bay, B. H., Yip, G. & Yu, Y. Breast cancer metastasis. *Cancer Genomics Proteomics* **9,** 311–320 (2012).

100. Dai, X. *et al.* Breast cancer intrinsic subtype classification, clinical use and future trends. *Am. J. Cancer Res.* **5,** 2929–2943 (2015).

101. Ignatiadis, M. & Sotiriou, C. Luminal breast cancer: from biology to treatment. *Nat. Rev. Clin. Oncol.* **10,** 494–506 (2013).

102. Cejalvo, J. M. *et al.* Intrinsic subtypes and gene expression profiles in primary and metastatic breast cancer. *Cancer Res.* **77,** 2213–2221 (2017).

103. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumors. *Nature* **490,** 61–70 (2012).

104. Pantel, K. & Hayes, D. F. Disseminated breast tumour cells: Biological and clinical meaning. *Nat. Rev. Clin. Oncol.* **15,** 129–131 (2018).

105. Kennecke, H. *et al.* Metastatic behavior of breast cancer subtypes. *J. Clin. Oncol.* **28,** 3271–3277 (2010).

106. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci.* **98,** 10869–10874 (2001).

107. Amin, M. B. *et al. AJCC Cancer Staging Manual*. (2017). doi:10.1001/jama.2010.1525.

108. Robson, M. *et al.* Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *N. Engl. J. Med.* **377,** 523–533 (2017).

109. Robson, M. E. *et al.* OlympiAD final overall survival and tolerability results: Olaparib versus chemotherapy treatment of physician's choice in patients with a germline BRCA mutation and HER2-negative metastatic breast cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* (2019). doi:10.1093/annonc/mdz012

110. André, F. *et al.* LBA3_PRAlpelisib (ALP) + fulvestrant (FUL) for advanced breast cancer (ABC): Results of the phase III SOLAR-1 trial. *Ann. Oncol.* **29,** (2018).

111. Turner, N. C. *et al.* Overall Survival with Palbociclib and Fulvestrant in Advanced Breast Cancer. *N. Engl. J. Med.* **379,** 1926–1936 (2018).

112. Echavarria, I., Jerez, Y., Martin, M. & Lopez-Tarruella, S. Incorporating CDK4/6 Inhibitors in the Treatment of Advanced Luminal Breast Cancer. *Breast Care.* **12,** 296–302 (2017).

113. Vonderheide, R. H., Domchek, S. M. & Clark, A. S. Immunotherapy for Breast Cancer: What Are We Missing? *Clin. Can. Res.* **23,** 2640–2646 (2017).

114. Schmid, P. *et al.* Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *N. Engl. J. Med.* **379,** 2108–2121 (2018).

115. Fan, W. & Chang, J. Endocrine therapy resistance in breast cancer: current status, possible mechanisms and overcoming strategies. *Futur. Med. Chem.* **7,** 1511–1519 (2015).

116. Szostakowska, M., Trębińska-Stryjewska, A., Grzybowska, E. A. & Fabisiewicz, A. Resistance to endocrine therapy in breast

cancer: molecular mechanisms and future goals. *Breast Cancer Res. Treat.* (2018). doi: 10.1007/s10549-018-5023-4.

117. Mancuso, M. R. & Massarweh, S. A. Endocrine therapy and strategies to overcome therapeutic resistance in breast cancer. *Curr. Probl. Cancer* **40,** 95–105 (2016).

118. Jakabova, A. *et al.* Molecular characterization and heterogeneity of circulating tumor cells in breast cancer. *Breast Cancer Res. Treat.* **166,** 695–700 (2017).

119. Boral, D. *et al.* Molecular characterization of breast cancer CTCs associated with brain metastasis. *Nat. Commun.* **8,** 196 (2017).

120. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114-20 (2014).

121. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15-21(2013).

122. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923-30 (2014).

123. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

124. Langfelder, P. & Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. *J. Stat. Softw.* **46,** i11 (2012).

125. McKenna, A. . *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297-303 (2010).

126. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31,** 213-9 (2013).

127. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17,** 22 (2016).

128. Konigsberg, R. *et al.* Detection of EpCAM positive and negative circulating tumor cells in metastatic breast cancer patients. *Acta Oncol.*. **50,** 700–710 (2011).

129. Sieuwerts, A. M. *et al.* Anti-epithelial cell adhesion molecule antibodies and the detection of circulating normal-like breast tumor cells. *J. Natl. Cancer Inst.* **101,** 61–66 (2009).

130. Liu, Z. *et al.* Negative enrichment by immunomagnetic nanobeads for unbiased characterization of circulating tumor cells from peripheral blood of cancer patients. *J. Transl. Med.* **9,** 70 (2011).

131. Pierga, J. Y. *et al.* Circulating tumor cell detection predicts early metastatic relapse after neoadjuvant chemotherapy in large operable and locally advanced breast cancer in a phase II randomized trial. *Clin. Cancer Res.* **14,** 7004–7010 (2008).

132. Rack, B. *et al.* Prognostic Relevance of Circulating Tumor Cells in the Peripheral Blood of Primary Breast Cancer Patients. *Cancer Res.* **70,** S6-5 (2010).

133. Hayes, D. F. *et al.* Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival. *Clin. Cancer Res.* **12,** 4218–4224 (2006).

134. Lianidou, E. S. & Markou, A. Circulating tumor cells as emerging tumor biomarkers in breast cancer. *Clin. Chem. Lab. Med.* **49,** 1579–1590 (2011).

135. Mego, M. *et al.* Characterization of metastatic breast cancer patients with nondetectable circulating tumor cells. *Int. J. cancer* **129,** 417–423 (2011).

136. Giordano, A. *et al.* Circulating tumor cells in immunohistochemical subtypes of metastatic breast cancer:

lack of prediction in HER2-positive disease treated with targeted therapy. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **23,** 1144–1150 (2012).

137. Joosse, S. A., Gorges, T. M. & Pantel, K. Biology, detection, and clinical implications of circulating tumor cells. *EMBO Mol. Med.* **7,** 1–11 (2015).

138. Fehm, T. *et al.* Detection and characterization of circulating tumor cells in blood of primary breast cancer patients by RT-PCR and comparison to status of bone marrow disseminated cells. *Breast Cancer Res.* **11,** 1–9 (2009).

139. Kasimir-Bauer, S., Hoffmann, O., Wallwiener, D., Kimmig, R. & Fehm, T. Expression of stem cell and epithelial-mesenchymal transition markers in primary breast cancer patients with circulating tumor cells. *Breast Cancer Res.* **14,** R15 (2012).

140. Barrière, G., Riouallon, A., Renaudie, J., Tartary, M. & Rigaud, M. Mesenchymal characterization: Alternative to simple CTC detection in two clinical trials. *Anticancer Res.* **32,** 3363–3370 (2012).

141. Barrière, G., Riouallon, A., Renaudie, J., Tartary, M. & Rigaud, M. Mesenchymal and stemness circulating tumor cells in early breast cancer diagnosis. *BMC Cancer* **12,** 114 (2012).

142. Kasimir-Bauer, S. *et al.* Does primary neoadjuvant systemic therapy eradicate minimal residual disease? Analysis of disseminated and circulating tumor cells before and after therapy. *Breast Cancer Res.* **18,** 20 (2016).

143. Keup, C. *et al.* RNA profiles of circulating tumor cells and extracellular vesicles for therapy stratification of metastatic breast cancer patients. *Clin. Chem.* **64,** 1054–1062 (2018).

144. Aktas, B. *et al.* Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells

of metastatic breast cancer patients. *Breast Cancer Res.* **11,** 1–9 (2009).

145. Xenidis, N. *et al.* Peripheral blood circulating cytokeratin-19 mRNA-positive cells after the completion of adjuvant chemotherapy in patients with operable breast cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **14,** 849–855 (2003).

146. Tewes, M. *et al.* Molecular profiling and predictive value of circulating tumor cells in patients with metastatic breast cancer: An option for monitoring response to breast cancer related therapies. *Breast Cancer Res. Treat.* **115,** 581–590 (2009).

147. Bredemeier, M. *et al.* Gene Expression Signatures in Circulating Tumor Cells Correlate with Response to Therapy in Metastatic Breast Cancer. *Clin. Chem.* **63,** 1585–1593 (2017).

148. Markiewicz, A. *et al.* Spectrum of Epithelial-Mesenchymal Transition Phenotypes in Circulating Tumour Cells from Early Breast Cancer Patients. *Cancers.* **11,** E59 (2019).

149. Suzuki, E. *et al.* Gene expression profile of peripheral blood mononuclear cells may contribute to the identification and immunological classification of breast cancer patients. *Breast Cancer* (2018). doi:10.1007/s12282-018-0920-2.

150. Wallwiener, M. *et al.* The impact of HER2 phenotype of circulating tumor cells in metastatic breast cancer: a retrospective study in 107 patients. *BMC Cancer* **15,** 403 (2015).

151. Bredemeier, M., Edimiris, P., Tewes, M., Mach, P. & Aktas, B. Establishment of a multimarker qPCR panel for the molecular characterization of circulating tumor cells in blood samples of metastatic breast cancer patients during the course of palliative treatment. *Oncotarget* **7,** 41677–41690 (2016).

152. Aaltonen, K. E. *et al.* Molecular characterization of circulating tumor cells from patients with metastatic breast cancer reflects

evolutionary changes in gene expression under the pressure of systemic therapy. *Oncotarget* **8,** 45544–45565 (2017).

153. Canas-Marques, R. & Schnitt, S. J. E-cadherin immunohistochemistry in breast pathology: uses and pitfalls. *Histopathology* **68,** 57–69 (2016).

154. Han, E. K. *et al.* Stable overexpression of cyclin D1 in a human mammary epithelial cell line prolongs the S-phase and inhibits growth. *Oncogene* **10,** 953–961 (1995).

155. Polioudaki, H. *et al.* Variable expression levels of keratin and vimentin reveal differential EMT status of circulating tumor cells and correlation with clinical characteristics and outcome of patients with metastatic breast cancer. *BMC Cancer* **15,** 399 (2015).

156. Fujisue, M. *et al.* Clinical Significance of CK19 Negative Breast Cancer. *Cancers.* **5,** 1–11 (2012).

157. Fettig, L. M. *et al.* Cross talk between progesterone receptors and retinoic acid receptors in regulation of cytokeratin 5-positive breast cancer cells. *Oncogene* **36,** 6074–6084 (2017).

158. Khaled, W. T. *et al.* BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. *Nat. Commun.* **6,** 5987 (2015).

159. Dairkee, S. H., Ljung, B. M., Smith, H. & Hackett, A. Immunolocalization of a human basal epithelium specific keratin in benign and malignant breast disease. *Breast Cancer Res. Treat.* **10,** 11–20 (1987).

160. Engel, B. E., Cress, W. D. & Santiago-Cardona, P. G. The retinoblastoma protein: a master tumor suppressor acts as a link between cell cycle and cell adhesion. *Cell Health Cytoskelet.* **7,** 1–10 (2015).

161. Caligiuri, I., Toffoli, G., Giordano, A. & Rizzolio, F. pRb controls estrogen receptor alpha protein stability and activity. *Oncotarget* **4,** 875–883 (2013).

162. Sasahara, A. *et al.* An autocrine/paracrine circuit of growth differentiation factor (GDF) 15 has a role for maintenance of breast cancer stem-like cells. *Oncotarget* **8,** 24869–24881 (2017).

163. Tovey, S. M. *et al.* Low expression of HER2 protein in breast cancer is biologically significant. *J. Pathol.* **210,** 358–362 (2006).

164. Raimondi, C., Gradilone, A., Naso, G. & Vincenzi, B. Epithelial-mesenchymal transition and stemness features in circulating tumor cells from breast cancer patients. *Breast Cancer Res. Treat.* **130,** 449–455 (2011).

165. Pestrin, M. *et al.* Correlation of HER2 status between primary tumors and corresponding circulating tumor cells in advanced breast cancer patients. *Breast Cancer Res. Treat.* **118,** 523–530 (2009).

166. Sheikh, A. *et al.* The spectrum of genetic mutations in breast cancer. *Asian Pac. J. Cancer Prev.* **16,** 2177–2185 (2015).

167. Li, J., Li, M., Chen, P. & Ba, Q. High expression of PALB2 predicts poor prognosis in patients with advanced breast cancer. *FEBS Open Bio* **8,** 56–63 (2017).

168. Wolfer, A. *et al.* MYC regulation of a 'poor-prognosis' metastatic cancer cell state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 3698–3703 (2010).

169. McNeil, C. M. *et al.* c-Myc overexpression and endocrine resistance in breast cancer. *J. Steroid Biochem. Mol. Biol.* **102,** 147–155 (2006).

170. Kwapisz, D. Cyclin-dependent kinase 4/6 inhibitors in breast cancer: palbociclib, ribociclib, and abemaciclib. *Breast Cancer Res. Treat.* **166,** 41–54 (2017).

171. Pinto, C. A., Widodo, E., Waltham, M. & Thompson, E. W. Breast cancer stem cells and epithelial mesenchymal plasticity - Implications for chemoresistance. *Cancer Lett.* **341,** 56–62 (2013).

172. Bulfoni, M. *et al.* In patients with metastatic breast cancer the identification of circulating tumor cells in epithelial-to-mesenchymal transition is associated with a poor prognosis. *Breast Cancer Res.* **18,** 30 (2016).

173. Markiewicz, A. & Zaczek, A. J. The Landscape of Circulating Tumor Cell Research in the Context of Epithelial-Mesenchymal Transition. *Pathobiology* **84,** 264–283 (2017).

174. Lou, Y. *et al*. Epithelial-mesenchymal transition (EMT) is not sufficient for spontaneous murine breast cancer metastasis. *Dev. Dyn.* **237,** 2755–2768 (2008).

175. Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527,** 472–476 (2015).

176. Jolly, M. K., Ware, K. E., Gilja, S., Somarelli, J. A. & Levine, H. EMT and MET: necessary or permissive for metastasis? *Mol. Oncol.* **11,** 755–769 (2017).

177. de Wit, S. *et al.* EpCAMhigh and EpCAMlow circulating tumour cells in metastatic prostate and breast cancer patients. *Oncotarget* **9,** 35705–35716 (2018).

178. Charafe-Jauffret, E. *et al.* Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. *Clin. Cancer Res.* **16,** 45–55 (2010).

179. Ginestier, C. *et al.* ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1,** 555–567 (2007).

180. Ma, F. *et al.* Aldehyde dehydrogenase 1 (ALDH1) expression is an independent prognostic factor in triple negative breast cancer (TNBC). *Medicine.* **96,** e6561 (2017).

181. Yao, J., Jin, Q., Wang, X.-D., Zhu, H.-J. & Ni, Q.-C. Aldehyde dehydrogenase 1 expression is correlated with poor prognosis in breast cancer. *Medicine.* **96,** e7171 (2017).

182. Donnenberg, V. S., Huber, A., Basse, P., Rubin, J. P. & Donnenberg, A. D. Neither epithelial nor mesenchymal circulating tumor cells isolated from breast cancer patients are tumorigenic in NOD-scid Il2rg(null) mice. *NPJ breast cancer* **2,** 16004 (2016).

183. Jia, H. *et al.* Immunotherapy for triple-negative breast cancer: Existing challenges and exciting prospects. *Drug Resist. Updat.* **32,** 1–15 (2017).

184. Podo, F. *et al.* Triple-negative breast cancer: present challenges and new perspectives. *Mol. Oncol.* **4,** 209–229 (2010).

185. Zhou, S. *et al.* Differential expression and clinical significance of epithelial-mesenchymal transition markers among different histological types of triple-negative breast cancer. *J. Cancer* **9,** 604–613 (2018).

186. DiMeo, T. A. *et al.* A novel lung metastasis signature links Wnt signaling with cancer cell self-renewal and epithelial-mesenchymal transition in basal-like breast cancer. *Cancer Res.* **69,** 5364–5373 (2009).

187. Khramtsov, A. I. *et al.* Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. *Am. J. Pathol.* **176,** 2911–2920 (2010).

188. Dey, N. *et al.* Wnt signaling in triple negative breast cancer is associated with metastasis. *BMC Cancer* **13,** 537 (2013).

189. Bleckmann, A. *et al.* Nuclear LEF1/TCF4 correlate with poor prognosis but not with nuclear beta-catenin in cerebral metastasis of lung adenocarcinomas. *Clin. Exp. Metastasis* **30,** 471–482 (2013).

190. Marumoto, T., Zhang, D. & Saya, H. Aurora-A - a guardian of poles. *Nat. Rev. Cancer* **5,** 42–50 (2005).

191. Zhang, Y. *et al.* Elevated Aurora B expression contributes to chemoresistance and poor prognosis in breast cancer. *Int. J. Clin. Exp. Pathol.* **8,** 751–757 (2015).

192. Kitzen, J. J. E. M., de Jonge, M. J. A. & Verweij, J. Aurora kinase inhibitors. *Crit. Rev. Oncol. Hematol.* **73,** 99–110 (2010).

193. Keen, N. & Taylor, S. Aurora-kinase inhibitors as anticancer agents. *Nat. Rev. Cancer* **4,** 927–936 (2004).

194. Lai, J. P. *et al.* SULF1 inhibits tumor growth and potentiates the effects of histone deacetylase inhibitors in hepatocellular carcinoma. *Gastroenterology* **130,** 2130–2144 (2006).

195. Phillips, G. R. *et al.* Gamma-protocadherins are targeted to subsets of synapses and intracellular organelles in neurons. *J. Neurosci.* **23,** 5096–5104 (2003).

196. Yao, F., Zhang, C., Du, W., Liu, C. & Xu, Y. Identification of Gene-Expression Signatures and Protein Markers for Breast Cancer Grading and Staging. *PLoS One* **10,** e0138213 (2015).

197. Novak, P. *et al.* Agglomerative epigenetic aberrations are a common event in human breast cancer. *Cancer Res.* **68,** 8616–8625 (2008).

198. Pickard, M. R., Mourtada-Maarabouni, M. & Williams, G. T. Candidate tumour suppressor Fau regulates apoptosis in human

cells: an essential role for Bcl-G. *Biochim. Biophys. Acta* **1812,** 1146–1153 (2011).

199. Lin, M.-L., Park, J.-H., Nishidate, T., Nakamura, Y. & Katagiri, T. Involvement of maternal embryonic leucine zipper kinase (MELK) in mammary carcinogenesis through interaction with Bcl-G, a pro-apoptotic member of the Bcl-2 family. *Breast Cancer Res.* **9,** R17 (2007).

200. Saito, R., Nakauchi, H. & Watanabe, S. Serine/threonine kinase, Melk, regulates proliferation and glial differentiation of retinal progenitor cells. *Cancer Sci.* **103,** 42–49 (2012).

201. Joshi, K. *et al.* MELK-dependent FOXM1 phosphorylation is essential for proliferation of glioma stem cells. *Stem Cells* **31,** 1051–1063 (2013).

202. Pitner, M. K., Taliaferro, J. M., Dalby, K. N. & Bartholomeusz, C. MELK: a potential novel therapeutic target for TNBC and other aggressive malignancies. *Expert Opin. Ther. Targets* **21,** 849–859 (2017).

203. Bianchini, G. *et al.* Prognostic and therapeutic implications of distinct kinase expression patterns in different subtypes of breast cancer. *Cancer Res.* **70,** 8852–8862 (2010).

204. Speers, C. *et al.* Identification of novel kinase targets for the treatment of estrogen receptor-negative breast cancer. *Clin. Cancer Res.* **15,** 6327–6340 (2009).

205. Finetti, P. *et al.* Sixteen-kinase gene expression identifies luminal breast cancers with poor prognosis. *Cancer Res.* **68,** 767–776 (2008).

206. Wang, Y. *et al.* MELK is an oncogenic kinase essential for mitotic progression in basal-like breast cancer cells. *Elife* **3,** e01763 (2014).

207. Porras, T. B., Kaur, P., Ring, A., Schechter, N. & Lang, J. E. Challenges in using liquid biopsies for gene expression profiling. *Oncotarget* **9,** 7036–7053 (2018).

208. Fina, E. *et al.* Gene expression profiling of circulating tumor cells in breast cancer. *Clin. Chem.* **61,** 278–289 (2015).

209. Beke, L. *et al.* MELK-T1, a small-molecule inhibitor of protein kinase MELK, decreases DNA-damage tolerance in proliferating cancer cells. *Biosci. Rep.* **35,** e00267 (2015).

210. Xu, L. *et al.* The Novel Association of Circulating Tumor Cells and Circulating Megakaryocytes with Prostate Cancer Prognosis. *Clin. Cancer Res.* **23,** 5112–5122 (2017).

211. Renier, C. *et al.* Label-free isolation of prostate circulating tumor cells using Vortex microfluidic technology. *Precis. Oncol.* **1,** 15 (2017).

212. Lin, E. *et al.* High-Throughput Microfluidic Labyrinth for the Label-free Isolation of Circulating Tumor Cells. *Cell Syst.* **5,** 295–304 (2017).

213. Aghabozorgi, A. S. *et al.* Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives. *Biochimie* **157,** 64–71 (2019).

214. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6,** (2014).

215. Jass, J. R., Young, J. & Leggett, B. A. Evolution of colorectal cancer: change of pace and change of direction. *J. Gastroenterol. Hepatol.* **17,** 17–26 (2002).

216. Kong, S. L. *et al.* Molecular characterization of circulating colorectal tumor cells defines genetic signatures for individualized cancer care. *Oncotarget* **8,** 68026–68037 (2017).

217. Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 16368–16373 (2005).

218. Denis, J. A. *et al.* Droplet digital PCR of circulating tumor cells from colorectal cancer patients can predict KRAS mutations before surgery. *Mol. Oncol.* **10,** 1221–1231 (2016).

219. Cho, H. *et al.* Microfluidic technologies for circulating tumor cell isolation. *Analyst* **143,** 2936–2970 (2018).

220. Lip, G. Y. H., Chin, B. S. P. & Blann, A. D. Cancer and the prothrombotic state. *Lancet. Oncol.* **3,** 27–34 (2002).