UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUIMICAS

TESIS DOCTORAL

Imaging and proteomics based analysis of colorectal cancer metástasis

Análisis de la metástasis del cáncer colorectal mediante proteómica y microscopía

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Guillermo Solís Fernández

Directores

Rodrigo Barderas Manchado
Ana Guzmán Aránguez
Johan Hofkens

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS

TESIS DOCTORAL

Imaging and proteomics based analysis of colorectal cancer metastasis
Análisis de la metástasis del cáncer colorectal mediante proteómica y microscopía

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Guillermo Solís Fernández

DIRECTORES

Rodrigo Barderas Manchado
Ana Guzmán Aránguez
Johan Hofkens

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS

TESIS DOCTORAL

Imaging and proteomics based analysis of colorectal cancer metastasis
Análisis de la metástasis del cáncer colorectal mediante proteómica y microscopía

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Guillermo Solís Fernández

DIRECTORES

Rodrigo Barderas Manchado
Ana Guzmán Aránguez
Johan Hofkens

Programa de doctorado en Bioquímica, Biología Molecular y Biomedicina

Acknowledgements

Acknowledgements

First of all, I would like to thank the economic support that has made possible the consecution of this thesis. I would like to thank and acknowledge the Faculty of Science for the Science@Leuven schorlaship that allowed me to come to Leuven to follow my master studies. Without this initial step none of the wonderful things that have come after would have been possible. Then I would like to thank the FWO for considering the PhD project proposal we presented almost five years and a half ago to be worth funding. Also, I would like to thank the FWO for the long-stay abroad travel grant that I was granted and that was a great financial support for a project that has required so much travelling. Finally, I would like to thank my supervisor, Johan Hofkens for those extra months of funding, finishing this PhD would have way more complicated.

However, as everybody knows, money does not bring happiness, and I would like to thank now all those persons that have supported me throughout these last years. I apologize upfront in case I forget anybody, five years between two labs and too many Madrid-Brussels flights have taken a toll on this poor head of mine.

To the members of my examination and supervisory committee, thanks a lot, organizing a defence with 11 members could have been a complete nightmare and it was only possible with your collaboration. Thanks for all your feedback that has helped to improve this thesis, and thanks for your nice comments about me and my work, after all the work and time invested in this project it felt great to have that recognition.

To my supervisors, Johan, Susana, Rodrigo and Ana, thanks to all of you, you have always been there when I have needed your help. Johan, you encouraged me to apply for the FWO and accepted to embark in that crazy idea that was doing a PhD between two universities. You have always given what I needed whenever I needed, and I will always consider following the advanced fluorescence course as one of the best decisions that I could take during my master studies. Susana, although you were

not initially in the supervisory team, you were always there when I needed help. You always counted on me for supervising students and helped me feel like a valuable researcher. Seeing your lab grow and being part of that process was amazing and I am sure you will be an incredible professor.

Ana, aceptaste entrar a formar parte de esta locura que ha sido la tesis en cotutela con todo lo que ello conllevaba y sin ti no habría sido posible, si ahora estoy escribiendo esto es gracias a ti también. Rodrigo, no creo que sea capaz de expresar bien lo agradecido que estoy por estos años. Ya cuando empecé a hacer el TFG, sin tener nada que conmigo, estuviste ahí intentando echarme un cable. Me dijiste que cuando volviera del máster, hablásemos, que algo se podría hacer. Casi 6 años después de aquello el algo ha cogido forma y sin tu ayuda habría sido imposible. Me cuesta pensar en alguien que pudiera haberme guiado mejor en todo este proceso, con todos los experimentos que no han salido y la dificultad que implicaba tener una tesis entre dos países, muchísimas gracias, de verdad.

To all the people in the MIP division that have shared these 5 years with me, Laurens, Aline, Marisa, Rafa, Bea, Hongbo, Elizabete, Michiel, you always made me feel welcome when I was around. To Charlotte and Quinten, my students who became later colleagues, I learned from you while I was your supervisor and I kept learning from you when you kept progressing, I wish you the best. To Indra, although your driving skills nearly got me killed (together with Boris) I know this would not have been nearly as fun if you would not have been around. With you around I always felt like I had a good friend and I'm really gonna miss your car problem-stories. To my favourite crazy Italian, Monica, thanks a lot, your energy was contagious and the arenberg Orchestra concerts were some of the funniest nights I have had in my life, although you still owe a true Italian carbonara dinner. To Johannes, the paperwork man, without you I

probably would not have been able to finish many of the insane amount of bureaucracy I had to go through, not only that, you were always a good friend. To Boris, the French speaking Belgian with the best English accent I've ever met, the programming genius, I have so many good memories with you, the La Trappe ladder in Veldhoven, your visit to Madrid or the Tour Louis, even thinking about setting a company together. You are great and I am looking forward to your PhD defense, leave one spot for me over there in Sweden.

Finally, I would like to thank all the technical staff at the MIP, without you the whole unit would collapse. Greet, thanks a lot, I am not sure if there has been anybody that has asked you for more PO's for omnia than me. Ann, way more than a secretary, you have always cared about me and having you around while being in a foreign country help to feel protected and a bit more like home. Rik, the microscope saver, despite my leica-curse you have always been there whenever I needed help in that nightmare that the SP8 can be. The patience you have with all the users and the amount of work you do for us is something that should be acknowledged with a guiness world record. Last but not least, Carine, the day you leave the MIP the whole unit will collapse, I honestly do not know how we would survive without you. You always know where everything is and whenever something goes wrong the first question is always, have you asked Carine? Thanks a lot.

A toda la gente del Carlos III, Lucía, Pablo, Lidia, ir hasta Majadahonda sería aún más rollo sin vosotros por allí. A las huevitos, Sofía y Eli, y el pollito, Edu, por ser unos liantes que han conseguido que por primera vez salgamos de fiesta desde que llegamos al Carlos III. Edu, tu eres el más liante de todos. A Maripaz, Maripili para los amigos, madre de loros, muchas gracias, eres una tía como pocas y tienes la colección de anécdotas más flipante del mundo. A Ana y María, mis compañeras de PF, han sido muchas horas juntos, muchos experimentos, que no salen,

marrones, muchos (demasiados) flasks en cultivos, unos cuantos bollitos y por encima de todo un barco que sigue flotando y no termina de hundirse. Por todas las veces que me habéis dicho por enésima vez cómo iba el protocolo de la extracción de RNA o dónde encontrar alguna cosa, machinimas gracias.

A mis amigos de toda la vida, a los que me aguantan cuando doy la chapa y con los que he bebido y me he divertido para un par de vidas ya probablemente, sin quedar y vernos todas las semanas para soltar un poco de tensión y desconectar un poco estoy seguro que habría terminado con la cabeza ida.

A mis bioquímicos y bioquímicos adjuntos, sois los mejores si os agradeciera uno a uno a todos por todas las cosas buenas que me habéis dado estos años estos agradecimientos serían aún más largos de lo que ya son, menos mal que nos decimos con bastante frecuencia que nos queremos mucho y eso me quita algo de responsabilidad ahora. Cris, la mejor futura mamá del mundo, te quiero mucho y vas a ser la mejor profe de universidad que la UCM ha visto nunca, muchas gracias por todo. Rodri, muchas gracias porque aunque estemos en desacuerdo de vez en cuando, sigues siendo mi compinche favorito para ir a andar con cerveza y hablar, e incluso sin cerveza también.

A mi familia postiza, Uwe y Maricruz, siempre habéis sido como unos padres en el extranjero, siempre nos habéis cuidado a Carmen y a mí, cuando estábamos los dos en Bélgica y cuando ya me quede yo sólo a mi también. Gracias porque vuestra casa siempre ha estado abierta, incluso cuando más falta ha hecho, como la vez que me dejé las llaves dentro de casa. A mi familia, mis padres y mi hermano, porque no podría haber tenido más suerte en la vida, me habéis apoyado en cada decisión que he tomado y habéis sufrido con cada papeleo, cada rechazo, cada paso extra que había que dar como lo sufría yo y siempre habéis estado

ahí. Estoy tremendamente orgulloso de vosotros y no cambiaría ni un minuto de los que hemos pasado juntos.

Leyre, gracias, soy consciente de que el proceso de este último año es parte de tu trabajo, pero como todo, un trabajo se puede hacer bien y se puede hacer mal. En ningún momento durante este año me he sentido incómodo y aunque ha habido días que me he quedado más descolocado sé que siempre ha sido buscando el que yo estuviera mejor. No sé si habría sido igual con otra persona, pero la verdad que estoy muy agradecido de haber dado desde el principio con una persona que ha entendido las cosas que pasaban por mi cabeza y que me ha ayudado a que yo las entendiera mejor. Muchas gracias.

Por último, a ti, mi amor, Carmen, la persona que más me ha soportado y más ha aguantado conmigo durante no sólo estos 5 años de tesis si no todo lo que vino antes. Empezamos pensando que lo nuestro duraría poco y que acabaríamos cada uno en un país y de repente pasamos dos años increíbles viviendo juntos y descubriéndonos a 1600km de casa. Tomé la decisión de embarcarme en esta cosa de ir y venir entre países cada pocos meses durante cuatro años y decidiste seguir a mi lado, has aguantado horas de conversaciones por Skype y pasar de darnos los buenos días con un beso a hacerlo a través de una pantalla. Me has apoyado cuando me parecía que todo se me venía encima y cuando me he quedado roto y no podía más, da igual lo lejos o lo cerca que estuviéramos siempre has estado ahí para mi. Eres lo mejor que me podría haber pasado en la vida y esta tesis es tan mía como es tuya porque sin ti no habría aguantado. Te quiero.

TABLE OF CONTENTS

SUMMARY

Cancer is the second most common cause of death, only preceded by cardiovascular diseases. The main reason behind the high mortality rates of cancer is the appearance of metastasis, the dissemination and colonisation of secondary tissues by cells from the primary tumour. In particular, in the case of colorectal cancer (CRC) the survival rates of patients diagnosed, drop drastically if diagnosis happens at later stages when metastasis has occurred. A similar trend can be observed for most of the cancers commonly diagnosed. When metastasis occurs, treatment relies heavily on the use of chemotherapy, even if the original tumour mass is surgically removed, the metastases will still survive. It is thus paramount to find new tools that improve early diagnosis so that detection can occur before the tumour spreads. In this context, the main focus of this thesis has been the understanding and characterisation of CRC metastasis to find new diagnostic markers that can be used in the clinic. We have used isogenic cell lines, that share the same genetic background but have different metastatic capacities, to define the proteome of CRC metastasis *in vitro*.

The first part of the thesis consists of the proteomic analysis of five isogenic cell lines of CRC. The KM12 model (that comprises the non-metastatic KM12C cells and the metastatic KM12SM and KM12L4a cells) and the SW480/SW620 model. The KM12 recapitulates the metastasis of CRC towards liver (KM12SM) and liver/lung (KM12L4a). On the other hand, SW620 cells are the lymph node metastatic counterpart of SW480 cells. Using tandem mass tag (TMT) multiplexed isotopic labelling and subcellular fractionation we have analysed the protein content of 6 different subcellular compartments (including secreted proteins). Among the proteins we found to be altered, some of them were already known to be related to metastasis, as it would be the case for MUC5AC, while others had not yet been associated with it, as BAIAP2 or GLG1. We validated the results obtained using both western blot and immunofluorescence. Through western blot we could not only confirm the

differences that we had detected in the mass spectrometry analysis but also observe that some of these differences would have been overlooked if a whole cell approach would have been followed. Immunofluorescence analysis confirmed that the alterations in protein levels and subcellular localisation were also visible in cells in native conditions. Finally, the evaluation of two of the altered markers, BAIAP2 and GLG1, revealed very interesting results in terms of their application for diagnosis or prognosis. For BAIAP2, we observed that the switch in its localisation from the cytosol to the membrane, as seen in metastatic cell lines, correlates with poor patient survival. On the other hand, GLG1 plasma levels were significantly higher in CRC patients than in controls and they were able to discriminate between metastatic CRC and control patients.

The second part of the thesis was focused on in the translation of our proteomic pipelines to other cell culture modalities. In particular we were interested in understanding how the extracellular matrix and culture dimensionality influence cancer cell behaviour. As a first step to evaluate how suitable our analysis was for 3D cultured cells, in collaboration with the group of Prof. Dr. Kouwer we studied the differentiation of adipose derived stem cells in matrices with different mechanical and chemical properties. The results we obtained were in agreement with their previous results and proved the tremendous impact that 3D culture can have in cell differentiation. One of the most influential factors for cell differentiation in 3D was the presence of RGD, a tripeptide commonly involved in integrin binding. Consequently, we decided to evaluate how ligands for other integrins would affect this process. We found that one peptide, recognised by integrin $\alpha_5\beta_1$ was able to induce the spreading of adipose derived stem cells much faster than any other matrix evaluated. Then, once we had confirmed that our analysis could be used for 3D culture cells we decided to continue and investigate the more complex KM12C/KM12SM model of CRC metastasis. We encapsulated CRC cells in different matrices and compared protein expression in each of these conditions and with cell

cultured in a monolayer. Interestingly, we could observe that 3D culture induces non-metastatic KM12C cells to become more similar to KM12SM cells, which would indicate that they acquire a more metastatic phenotype when they are cultured in 3D. Finally, we focused on the interplay of KM12C and KM12SM cells with cancer associated fibroblasts, another key player in the metastatic process. Our data showed that KM12C and KM12SM interact differently with cancer associated fibroblasts, forming different cellular structures.

In the last section of results, we studied the role in CRC metastasis of one of the proteins that had been previously found to be overexpressed in metastatic KM12SM cells, the Aryl hydrocarbon receptor-interacting protein (AIP). Through gain-of-function experiments we were able to demonstrate that AIP induced a drastic change in the behaviour of KM12C cells that displayed a more aggressive and metastatic like phenotype. Cells displayed higher invasive, adhesive and migration capacities when AIP was overexpressed. AIP also triggered changes in several signalling pathways, including AKT, JNK and SRC and EMT related mediators that were all altered in AIP overexpressing cells. Proteomic comparison of controls and cells ectopically expressing AIP revealed a series of proteins to be up and downregulated with 60 proteins being commonly altered between KM12C and KM12SM. Western blot analysis of some of them validated the proteomic results. In addition, through immunofluorescence analysis we could also detect changes in the levels and localisation of E-Cadherin, ZO-1 and Cadherin-17. The latter had been previously associated with metastasis and our results indicate that AIP overexpression induces cadherin-17 translocation to the cell membrane. Finally, *in vivo* experiments confirmed that the observed increase in metastatic capacities was not limited to *in vitro* studies. KM12C cells ectopically expressing AIP were able to reach the liver upon intrasplenic injection while control cells could not. Survival of mice injected with AIP overexpressing cells was significantly reduced. Altogether, our results

demonstrate that AIP can have a central role in the process of CRC metastasis.

To sum up, the results presented in this thesis show the clear potential that proteomic approaches can have for the understanding of not only CRC metastases but other biological questions as adipose derived stem cell differentiation. Some of the results shown here have clear potential for the application in the clinic and will be further explored in the future.

RESUMEN

Tras las enfermedades cardiovasculares, la segunda causa de muerte en países desarrollados es el cáncer. El principal factor detrás de la alta mortalidad del cáncer es la aparición de metástasis, el proceso por el cual células provenientes de la masa tumoral original diseminan y son capaces de colonizar tejidos distantes. En el caso del cáncer colorrectal (CCR), las tasas de supervivencia de los pacientes caen de manera drástica si el diagnóstico de la enfermedad se produce en estadios tardíos en los que ya haya aparecido metástasis. Esta tendencia se puede observar también en la mayoría de los cánceres diagnosticados en la actualidad. Uno de los principales problemas de la metástasis es que una vez aparece, el tratamiento del cáncer se vuelve complejo, siendo la quimioterapia la principal forma de tratamiento ya que, aunque se elimine quirúrgicamente la masa tumoral original, los nichos metastáticos sobrevivirán a la operación. Es por tanto necesario el desarrollo de nuevas técnicas de diagnóstico que permitan mejorar la detección temprana, antes de que el tumor se extienda. En este contexto, el principal objetivo de esta tesis ha sido entender y caracterizar la metástasis del CCR para encontrar marcadores de diagnóstico que puedan trasladarse a la práctica clínica. Para definir el proteoma de la metástasis de CCR hemos empleado una serie de líneas celulares isogénicas, que comparten la misma carga genética, con diferente potencial metastático.

La primera parte de esta tesis está centrada en la caracterización proteómica de cinco líneas isogénicas de CCR. Por un lado el modelo KM12 (compuesto por las células no metastáticas KM12C y las células metastáticas KM12SM y KM12L4a) y por otro el modelo SW480/SW620. El modelo KM12 recapitula la metástasis de CCR hacia hígado (KM12SM) y hacia hígado/pulmón (KM12L4a). El modelo de las células SW480 y SWS620 recapitula la metástasis de CCR hacia nódulo linfático. Para realizar el análisis proteómico de las muestras primero se realizó el fraccionamiento subcelular del extracto proteico de las cinco líneas

celulares. Las diferentes fracciones fueron marcadas utilizando el marcaje TMT 11-plex para su posterior análisis mediante espectrometría de masas. Entre las proteínas alteradas algunas de ellas ya habían sido descritas previamente como asociadas a metástasis, como es el caso de MUC5AC. Los resultados obtenidos fueron validados mediante western blot e inmunofluorescencia. La validación permitió, no solo confirmar los resultados obtenidos del análisis de espectrometría de masas sino que además nos permitió demostrar que alguno de los cambios detectados habría pasado desapercibido de no haber realizado el análisis a nivel subcelular. Por otro lado, la inmunofluorescencia nos permitió confirmar que los resultados derivados del western blot y la proteómica eran representativos de la situación real en células en condiciones nativas. Por último, los experimentos adicionales realizados con BAIAP2 y GLG1 demostraron el potencial de ambos marcadores para su uso en clínica. En el caso de BAIAP2, pudimos observar que el cambio en la localización desde el citoplasma hacia la membrana nuclear, tal y como ocurría en las células metastáticas, correlacionaba con una disminución en la supervivencia de los pacientes de CCR. Por su parte, encontramos que en el caso de GLG1 sus niveles séricos eran significativamente más altos en los pacientes de CCR comparado con los controles y eran suficientes para discriminar entre pacientes con CCR metastático y la población control.

La segunda parte de los resultados de la tesis ha estado enfocada a determinar la validez de nuestro método de análisis para el estudio de células y cultivos realizados en 3 dimensiones, además de tratar de entender los procesos de diferenciación celular en estas condiciones. Como primer paso hemos realizado el estudio de la diferenciación de células madre derivadas de tejido adiposo en colaboración con el grupo del profesor Paul Kouwer. Los resultados obtenidos coinciden con los resultados preliminares que ya habían obtenido en el grupo del profesor Kouwer, y proporcionaban una posible explicación para las diferencias

observadas entre las células encapsuladas en cada una de las matrices evaluadas, además de demostrar el tremendo impacto que las condiciones de cultivo en 3D pueden tener en el proceso de diferenciación. Una de las propiedades que observamos fue más crítica fue la presencia de ligandos de integrina. Decidimos, por tanto, evaluar el efecto que otros ligandos de integrina podrían tener en la diferenciación de las células madre derivadas de tejido adiposo. De los diferentes ligandos estudiados encontramos que un péptido en particular, reconocido por la integrina $\alpha_5\beta_1$, era capaz de inducir la rápida diferenciación de las células madre y cambios morfológicos muy marcados. Tras confirmar la validez de nuestros protocolos para el análisis del proteoma de células cultivadas en 3D, procedimos a cultivar las células KM12C y KM12SM del modelo de metástasis de CCR para investigar el efecto de la matriz en el proceso metastático. La comparativa entre células de CCR cultivadas en monocapa o encapsuladas reveló que las células KM12C se vuelven más similares a las metastáticas KM12SM al ser cultivadas en 3D. Por último, investigamos la relación entre las células KM12C y KM12SM y los fibroblastos asociados al tumor que también forman parte del microambiente tumoral. Mediante microscopía de fluorescencia pudimos observar que las células KM12C, no metastáticas, se distribuyen de manera diferente (más homogénea) con respecto a los fibroblastos que las células KM12SM (más segregadas).

El último bloque de resultados se centra en el estudio de una de las proteínas que habíamos encontrado como regulada al alza en un trabajo anterior, la proteína que interactúa con el receptor de aril hidrocarburos (AIP, por sus siglas en inglés). Para llevar a cabo dicho análisis indujimos la sobreexpresión de AIP tanto en las células KM12C como en las KM12SM. Mediante ensayos funcionales pudimos detectar un aumento drástico en las capacidades metastáticas (invasión, adhesión, migración y formación de colonias) de ambas líneas celulares al sobreexpresar AIP, aunque el cambio fue particularmente notable para las células KM12C.

Además la sobreexpresión de AIP también provocó cambios en los niveles de una serie de proteínas implicadas en señalización celular (como AKT, JNK o SRC) y en mediadores de la transición epitelio-mesénquima. Los cambios de los mediadores de la transición epitelio-mesénquima no se limitaron sólo a nivel de expresión, por inmunofluorescencia también pudimos detectar cambios en la localización de E-Cadherina y ZO-1. Para hacer una evaluación más detallada de las proteínas alteradas, realizamos la caracterización proteómica de las células control y aquellas que sobreexpresaban AIP. De dicha caracterización pudimos obtener una serie de proteínas cuyos niveles estaban alterados tanto en las células KM12C como en las KM12SM. Por último, para evaluar si el incremento en la capacidad metastática de las células KM12C tenía también efecto *in vivo* realizamos una serie de experimentos en ratones que confirmaron el claro incremento en las propiedades tumorigénicas y metastáticas de las células KM12C al expresar AIP de manera ectópica.

En conclusión, los resultados presentados en este trabajo constituyen una clara demostración del potencial que las técnicas proteómicas tienen. No sólo orientados al campo de la oncoproteómica de cara a descubrir nuevas dianas terapéuticas o marcadores de diagnóstico, sino también como una herramienta más a la hora de responder preguntas biológicas complejas.

SAMENVATTING

Kanker is de tweede meest voorkomende doodsoorzaak, alleen voorafgegaan door hart- en vaatziekten. De voornaamste oorzaak van het hoge sterftecijfer bij kanker is het optreden van metastasen, de verspreiding en kolonisatie van secundaire weefsels door cellen van de primaire tumor. In het bijzonder bij colorectale kanker (CRC) daalt de overlevingskans van de gediagnosticeerde patiënten drastisch wanneer de diagnose in een later stadium wordt gesteld en er metastase is opgetreden. Een soortgelijke tendens kan worden waargenomen bij de meeste vormen van kanker die vaak worden gediagnosticeerd. Wanneer metastase optreedt, is de behandeling sterk afhankelijk van het gebruik van chemotherapie; zelfs wanneer de oorspronkelijke tumormassa operatief wordt verwijderd, zullen de metastasen blijven voortbestaan. Het is dus van het grootste belang nieuwe instrumenten te vinden die de vroegtijdige diagnose verbeteren, zodat opsporing kan plaatsvinden voordat de tumor zich verspreidt. In deze context was de hoofdfocus van dit proefschrift het begrijpen en karakteriseren van CRC metastase om nieuwe diagnostische markers te vinden die in de kliniek gebruikt kunnen worden. We hebben gebruik gemaakt van isogene cellijnen, die dezelfde genetische achtergrond hebben maar verschillende metastatische capaciteiten, om het proteoom van CRC metastase *in vitro* te bepalen.

Het eerste deel van het proefschrift bestaat uit de proteomische analyse van vijf isogene cellijnen van CRC. Het KM12 model (dat bestaat uit de niet-metastatische KM12C cellen en de metastatische KM12SM en KM12L4a cellen) en het SW480/SW620 model. Het KM12-model bootst de metastasering van CRC naar lever (KM12SM) en lever/long (KM12L4a) na. Anderzijds zijn SW620-cellen de lymfekliermetastaserende tegenhanger van SW480-cellen. Met behulp van TMT-multiplexed isotopische labeling en subcellulaire fractionering hebben we de eiwitinhoud van 6 verschillende subcellulaire compartimenten (inclusief gesecreteerde eiwitten) geanalyseerd. Van de eiwitten die wij gewijzigd vonden, was van sommige reeds bekend dat zij

verband houden met metastase, zoals het geval zou zijn voor MUC5AC, terwijl andere daar nog niet mee in verband waren gebracht, zoals BAIAP2. Wij hebben de verkregen resultaten gevalideerd met behulp van zowel western blot als immunofluorescentie. Met behulp van western blot konden wij niet alleen de verschillen bevestigen die wij bij de massaspectrometrie-analyse hadden geconstateerd, maar ook constateren dat sommige van deze verschillen over het hoofd zouden zijn gezien als een hele-cel-benadering zou zijn gevolgd. Immunofluorescentieanalyse bevestigde dat de veranderingen in eiwitniveaus en subcellulaire lokalisatie ook zichtbaar waren in cellen in natuurlijke omstandigheden. Tenslotte bracht de evaluatie van twee van de veranderde markers, BAIAP2 en GLG1, zeer interessante resultaten aan het licht wat betreft hun toepassing voor diagnose. Voor BAIAP2 hebben wij geconstateerd dat de verschuiving in zijn lokalisatie van het cytosol naar het membraan, zoals gezien in metastatische cellijnen, correleert met een slechte overleving van de patiënt. Anderzijds waren de GLG1 plasmaniveaus significant hoger bij CRC-patiënten dan bij controles en konden zij onderscheid maken tussen metastatische CRC en controlepatiënten.

Het tweede deel van het proefschrift was gericht op de vertaling van onze proteomische pijplijnen naar andere celkweekmodaliteiten. In het bijzonder waren we geïnteresseerd om te begrijpen hoe de extra cellulaire matrix en kweekdimensies het gedrag van kankercellen beïnvloeden. Als eerste stap om te evalueren hoe geschikt onze analyse was voor 3D gekweekte cellen, bestudeerden we in samenwerking met de groep van Prof. Dr. Kouwer de differentiatie van adipose afgeleide stamcellen in matrices met verschillende mechanische en chemische eigenschappen. De resultaten die we verkregen waren in overeenstemming met hun eerdere resultaten en bewezen de enorme impact die 3D kweek kan hebben op celdifferentiatie. Een van de meest invloedrijke factoren voor celdifferentiatie in 3D was de aanwezigheid van RGD, een tripeptide dat

gewoonlijk betrokken is bij integrinebinding. Daarom besloten wij na te gaan hoe liganden voor andere integrines dit proces zouden beïnvloeden. We ontdekten dat één peptide, herkend door integrine alpha-beta, in staat was om de verspreiding van vetafgeleide stamcellen veel sneller te induceren dan enige andere geëvalueerde matrix. Toen we eenmaal hadden bevestigd dat onze analyse kon worden gebruikt voor 3D kweekcellen, besloten we verder te gaan en het complexere KM12C/KM12SM model van CRC metastase te onderzoeken. Wij kapselden CRC-cellen in verschillende matrices in en vergeleken de eiwitexpressie in elk van deze omstandigheden en met cellen gekweekt in een monolaag. Interessant is dat we konden waarnemen dat 3D kweek niet-metastatische KM12C cellen induceert om meer te lijken op KM12SM cellen, wat erop zou wijzen dat ze een meer metastatisch fenotype verwerven wanneer ze in 3D gekweekt worden. Tenslotte hebben we ons gericht op de wisselwerking van KM12C en KM12SM cellen met kanker-geassocieerde fibroblasten, een andere belangrijke speler in het metastatische proces. Onze gegevens toonden aan dat KM12C en KM12SM verschillend interageren met kanker-geassocieerde fibroblasten, en verschillende cellulaire structuren vormen.

In het laatste deel van de resultaten bestudeerden we de rol in CRC-metastase van een van de eiwitten die eerder tot overexpressie bleken te komen in metastatische KM12SM-cellen, het Aryl koolwaterstofreceptor-interagerende eiwit (AIP). Door middel van "gain-of-function" experimenten konden we aantonen dat AIP een drastische verandering teweegbracht in het gedrag van KM12C-cellen, die een agressiever en metastatisch fenotype gingen vertonen. De cellen vertoonden een hoger invasief, adhesief en migratievermogen wanneer AIP werd overgeëxpresseerd. AIP bracht ook veranderingen teweeg in verschillende signaalwegen, waaronder AKT, JNK en SRC en EMT-gerelateerde mediatoren die allemaal veranderd waren in cellen met overexpressie van AIP. Proteomische vergelijking van controles en cellen

die AIP ectopisch tot expressie brengen, toonde een reeks eiwitten aan die verhoogd of verlaagd werden, waarbij 60 eiwitten tussen KM12C en KM12SM gemeenlijk veranderd waren. Western blot analyse van sommige van deze eiwitten bevestigde de proteomische resultaten. Bovendien konden we met behulp van immunofluorescentieanalyses ook veranderingen vaststellen in de niveaus en lokalisatie van E-Cadherine, ZO-1 en Cadherine-17. Dit laatste werd eerder in verband gebracht met metastase en onze resultaten wijzen erop dat AIP-overexpressie induceert dat cadherine-17 wordt getranslokeerd naar de periferie van de cel. Tenslotte bevestigden *in vivo* experimenten dat de waargenomen toename in metastatische capaciteiten niet beperkt was tot *in vitro* studies. KM12C cellen die ectopisch AIP tot expressie brachten waren in staat om de lever te bereiken na intrasplenische injectie, terwijl controle cellen dit niet konden. De overleving van muizen geïnjecteerd met AIP overexpresserende cellen was significant verminderd. Al met al tonen onze resultaten aan dat AIP een centrale rol kan spelen in het proces van CRC metastasering.

Samenvattend tonen de resultaten in dit proefschrift het duidelijke potentieel aan dat proteomische benaderingen kunnen hebben voor het begrijpen van niet alleen CRC metastasen maar ook andere biologische vragen zoals adipeus afgeleide stamceldifferentiatie. Sommige van de hier getoonde resultaten hebben een duidelijk potentieel voor toepassing in de klinische toepassingen en zullen in de toekomst verder worden onderzocht.

ABBREVIATION LIST

AC	Adenocarcinoma
ACS	Adenoma-carcinoma sequence
AD	Adenoma
ADSC	Adipose stem cell
AJCC	American Joint Committee on cancer
APC	Antigen presenting cell
ATCC	American type culture collection
BSA	Bovine serum albumin
CA19-9	Carbohydrate antigen 19-9
CAF	Cancer-associated fibroblast
CEA	Carcinoembryonic antigen
CEB	Cytoplasmic fraction of CRC cells
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CRC	Colorectal Cancer
CSC	Cancer stem cell
DMEM	Dulbecco’s modified Eagle medium
DMSO	Dimethyl sulfoxide
DNA	Desoxiribonucleic acid(s)
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EMT	Epithelial-to -mesenchymal transition
ESI	Electrospray ionization
FAP	Familial Adenomatous polyposis
FBS	Fetal bovine serum
FIT	Faecal immunochemical test
FOBT	Faecal occult blood test
GAM	Goat anti mouse IgG
GAR	Goat anti rabbit IgG
GSEA	Gene set enrichment analysis

Abbreviation list

HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPS	Hamartomatous polyposis syndrome(s)
IBD	Inflammatory bowel disease
IF	Immunofluorescence
IHC	Immunohistochemistry
LC	Liquid chromatography
LF	Label-free
m/z	Charge/mass ratio
MALDI	Matrix assisted laser desorption ionization
MAP	MUTYH associated polyposis
MEB	Membrane fraction of CRC cells
MEM	Minimum essential medium
MET	Mesenchymal-to-epithelial transformation
MeV	Multiexperiment viewer
miRNA	Small micro RNA
MMR	DNA mismatch Repair
MS	Mass Spectrometry
MSI	Microsatellite instability
NEB	Nuclear fraction of CRC cells
NEB-CBP	Chromatin bound fraction of CRC cells
NES	Normalized enrichment score
NK cells	Natural Killer cells
O/N	Overnight
PAGE	Polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PBS-T	Tween 0.1% PBS 1×
PCR	Polymerase chain reaction
PDE	Phosphodiesterase
PEB	Cytoskeletal fraction of CRC cells
PEG	polyethylene glycol
PFA	Paraformaldehyde

PIC	Polyisocianopeptide
RIPA buffer	Radioimmunoprecipitation assay buffer
RNA	ribonucleic acid(s)
ROC	Receiver operating characteristic
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
SILAC	Stable isotope labelling by amino acids in cell culture
TAA	Tumour associated autoantigen
T_c cells	$CD8^+$ cytotoxic T cells
TCDD	2,3,7,8-tetrachlorodibenzo-p-dioxin
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TMA	Tissue microarray
TME	Tumour microenvironment
TMT	Tandem mass tag
TNM	Tumour-Nodule-Metastasis
TOF	Time-of-flight
T_{reg} cell	Regulatory T cells
TSA	Tumour specific antigen
WB	Western blot

INTRODUCTION

The first description of cancer dates back to Ancient Egypt, circa 1600 BCE, and can be found in an old manuscript that describes the appearance and surgery of tumours or ulcers in the breast (1). Such malignancies were already at the time described as “untreatable” (1). Due to its very nature, cancer has burdened humanity since its dawn. Nonetheless, it was not until the 20th and 21st centuries when the overall improvement in lifespan and the establishment of a welfare state led to an increase in cancer occurrence. Additionally, the improvement in both pre- and post-mortem diagnosis also led to a rise in the number of cancer cases detected (2). Since first described, our understanding of the onset, evolution and spread of cancer has drastically expanded. However, despite the advances achieved in cancer research, it remains the second leading cause of death in the United States of America (USA) only preceded by cardiovascular disease (3).

Cancer is a complex disease characterized by a series of molecular mechanisms and footprints (4, 5). These footprints are referred as the hallmarks of cancer and were initially comprised of six alterations (activated invasion/metastasis, angiogenesis induction, cell death resistance, sustained proliferative signalling and growth suppressors evasion) to which two more characteristics (dysregulated cellular energetics and avoidance of immune destruction) and two enabling factors (genome instability and tumour promoting inflammation) were added ten years later (5). Carcinogenesis, also known as tumorigenesis, is the series of events that lead to the formation of a primary tumour. The process varies widely depending on the tissue in which the tumour is formed. Nonetheless it usually follows some common steps. Generally, driven by mutations that induce abnormal and uncontrolled growth, a single cell will start a dysregulated clonal expansion that will lead to the formation of the tumoral mass (6). This clonal expansion does not imply that the cell that started the process bears all the characteristics described

before. Initially, the most important feature is the uncontrolled proliferation, and, in later stages, the expansion of a mutationally unstable clone will lead to the appearance of subpopulations with different traits that altogether give the tumour cells its aggressive behaviour (6).

As the tumour mass becomes larger it will, if possible, hijack the molecular mechanisms of blood vessel formation (angiogenesis) to induce the generation of new circulatory afferents that bring nutrients and oxygen to the poorly irrigated growing malignancy (7). Tumour angiogenesis can be particularly critical for two reasons: firstly, poorly irrigated tumours can evolve into a necrotic or apoptotic state that has been related with poor prognosis and increased aggressiveness (8-12); and secondly, angiogenesis and the formation of new blood vessels serve both as a source of nutrients to fuel the tumour growth and as a connection to the general blood circulation (13). The more irrigated a tumour becomes, the higher the chances for a tumoral cell to escape from its primary niche (13). Although the process of tumour implantation and development of secondary distant tumour masses is highly inefficient, a vast number of tumoral cells make their way to the blood each day (14, 15).

The process by which secondary tumours develop in distant tissues far from where the original cancer was formed is known as metastasis (16). Metastasis is of particular interest as it leads to one of the main causes of cancer associated death which is the spread of the primary tumour with the subsequent systemic damage (17). Although difficult to estimate in detail, some figures indicate that metastasis is responsible for around 70 to 90% of the total amount of cancer associated deaths (18, 19). Taking a closer look to the year cancer statistics for the USA (20), it can be clearly seen how late-stage diagnosis, at a point where distant metastasis have appeared, drastically decreases the probability of survival.

Some common steps of the onset and progression of the metastasis have been extensively described (21). Nonetheless, they are affected by the

tissue of origin and where spread happens, the tumour-host response and the nature of the tumour itself (22, 23). Once the primary tumour becomes large enough and induces the formation of new vessels through stimulating angiogenesis, both isolated cells and larger aggregates will disseminate into the bloodstream (23). Despite being initially debated (24), the consensus hypothesis in the field for the survival and implantation of these aggregates is the “seed and soil” hypothesis (22). The main principle behind this hypothesis is the concept that the primary tumour environment (the “seed”) determines and influences the tissues in which secondary metastasis will be able to settle (the “soil”) and develop a secondary tumour mass (22). This would explain the association between some types of cancer and the most common metastases developed in those malignancies (25, 26). In terms of the tumoral cells, the “seed and soil” hypothesis would translate into the mechanism by which malignant cells would transform into a more invasive phenotype, that upon arrival at a distant tissue would transit back to its original state to start the colonization. Cancer cells can exploit the epithelial to mesenchymal transition (EMT), a biological process crucial in tissue development, to adapt their traits and acquire migrating capacities that allow them to detach from the tumour mass and use the general circulation to migrate. The process would be reverted when the cancer cell (seed) encounters an appropriate tissue (soil) returning to its original proliferative status.

Colorectal cancer

Colorectal cancer (CRC) is the cancer with the third highest incidence rate for both males and females worldwide (20). Annual figures from the global cancer observatory indicate that there were over 1.9 million estimated new cases of CRC for both genders in 2020 (27, 28). In addition, CRC represents the second cause of cancer related death worldwide with nearly a million deaths in 2020 (27, 28). In terms of its global distribution

and epidemiology, CRC could be considered a disease more associated to developed countries. Besides the factors previously described that linked together welfare and increased life expectancy with cancer incidence, most of CRC's risk factors are also correlated with western developed countries (29). Nonetheless, the occidentalising process taking place in Asian and developing countries has also increased their incidence rates. Among the common risk factors for CRC, we can find sedentary behaviour and obesity, diet, smoking and alcohol consumption (29). Although no specific genetic footprint or mutations have been directly associated with the onset of CRC, around 30% of all cases have a family history of the disease, indicating the possible existence of predisposing, yet unknown, germ-line mutations (29).

Regarding CRC survival, if diagnosed in earlier stages of the disease, before metastasis occurs, 5-year survival is very high, 90 to 74% depending on the degree of spreading in the colorectal tissue (20). However, survival percentages plummet down to a 14% in the case of late diagnosis where metastasis has appeared (20). CRC's main metastatic niches are the liver and lungs, delicate tissues in which surgical removal of the metastases can be both difficult and insufficient (26, 30, 31). Precisely for this reason, significant efforts in CRC research are oriented towards developing better screening techniques that allow for a more efficient early detection.

Colorectal cancer aetiology and risk factors

One of the reasons behind the late diagnosis of CRC is the heterogeneity that characterizes this malignancy. Genetic and environmental factors have been described to influence the onset of the disease (32). However, roughly 75% of all diagnosed cases of CRC are sporadic with no previous record of familiar CRC. On the other hand, patients with family history of CRC represent between 15 to 20% of all CRC patients (32). It is worth mentioning though that the increased risk associated with a positive family

history of CRC is not linked to a particular genetic footprint and no genes have been specifically associated with CRC onset (33). Finally, there is a small percentage of hereditary syndromes (Lynch syndrome, adenomatous and hamartomatous polyposis syndromes) that have a clear high lifetime risk of developing CRC and represent around 5% of all CRC patients (33).

Due to the heterogeneity of the disease, the association with particular risk factors remains partially obscure although recent large meta-analysis studies have shed light onto some of the most prominent risk factors for CRC. In an analysis encompassing over 460 research articles Johnson et al. identified a series of risk factors that were significantly associated with CRC (34). The authors found significant associations with body mass index, red meat consumption, smoking cigarettes, inflammatory bowel disease (IBD) and family history of CRC (34). Although not statistically significant, additional factors that showed a positive trend on CRC risk were processed meat and alcohol consumption (34). On the contrary, fruit and vegetable consumption as well as physical activity showed a significant reduction in CRC risk (34). Interestingly, recent epidemiological studies have also shown a clear and significant association between CRC risk and some regions of Portugal and cities in Spain, although the actual cause behind this relation remains unclear (35).

Familial colorectal cancer

A positive familiar history for CRC represents the single most determinant risk factor for CRC, and the risk increases depending on the number of family members affected and the age at which they were diagnosed (32). In such patients either regular or one-off colonoscopy screening is advised (36). The exact frequency and age at which the test should take place will depend on the number of affected family members as well as the age at which they were diagnosed (37, 38). In general, a 5-year colonoscopy screening is recommended after turning 40 years old for

individuals with several family members with history of CRC, while, for patients with a lesser risk, a one-off colonoscopy at 55-year-old is recommended (37, 38).

Hereditary colorectal cancer syndrome

Lynch Syndrome

Lynch syndrome represents the most prevalent type of hereditary colorectal cancer syndrome (39). Although here considered in the context of CRC, Lynch syndrome is actually a predisposition to a broad spectrum of cancers, including endometrium, stomach or brain among others (39, 40). This syndrome is caused by heterozygous germline mutations in MLH1, MSH2, MSH6 or PMS2 that render the product of these genes non-functional (39). All these proteins are involved in DNA mismatch repair (MMR), and thus, the mutations induce a deficiency in MMR, causing microsatellite instability (41). Interestingly, MMR deficiency is not observed and does not play a role in the initiation of sporadic CRC (41). Lynch syndrome is diagnosed via genetic testing in those patients with positive familiar history and colonoscopy since an early age (20 years) is advised for the detection and removal of polyps (40, 42).

Familial adenomatous polyposis

Familial adenomatous polyposis (FAP), the second-most common hereditary CRC syndrome, is an autosomal dominant disease. FAP originates from the mutation of adenomatous polyposis coli (APC), a gene encoding a protein involved in Wnt signalling pathway (32, 43). Mutations in APC lead to the appearance of multiple colorectal adenomas that can later evolve into a colorectal adenocarcinoma if the patients are not diagnosed early enough (43). In recent years the analysis of patients with FAP-like phenotypes without APC mutations led to the categorization of a separate polyposis syndrome known as MUTYH associated polyposis (MAP), the first polyposis syndrome with recessive inheritance (44). Up to

40% patients presenting this phenotype were found to carry MUTYH mutations. Aberrant MUTYH creates somatic transversions in a variety of genes, including APC or KRAS (44).

Hamartomatous polyposis syndromes

The last category of hereditary CRC syndromes are the hamartomatous polyposis syndromes (HPS). Most of these types of syndromes encompass a collection of autosomal-dominant inherited diseases characterized by the presence of hamartomatous polyps, in comparison to the more common adenomatous polyps (45). Among these syndromes we can find the juvenile polyposis syndrome, Peutz-Jeghers syndrome, mixed polyposis syndrome and PTEN hamartoma tumour syndrome (46). HPS represents a rather small percentage (less than 1%) of all CRC diagnosed patients, and compared with adenomatous polyps, the histologic evolution of polyps is drastically different (45). Furthermore, the study of the Peutz-Jeghers syndrome served as a reference for describing an alternative to the adenoma-carcinoma sequence known as the hamartoma-to-carcinoma sequence, broadening the knowledge on sporadic hamartomas formation and progression (45, 47).

Colorectal cancer development and progression

As it was mentioned earlier, there are several factors that can increase the risk of suffering from CRC. Those related with lifestyle, can be modified, and even changed to be protective against CRC development (48). For example, dietary intake of fats and red meat has been associated with high risk of developing CRC while high intake of dietary fibre, calcium or fish have been linked with decreased risk of CRC (49, 50). Besides modifiable risk factors, other unavoidable risks such as type 2 diabetes, IBD, familiar history of CRC and other hereditary conditions as Lynch syndrome, have also been associated with CRC (48, 51). For most cases, initial development of CRC does not produce remarkable symptoms in patients. The tumour usually has a slow growth,

and it is not until it reaches a considerable size that symptoms related with the obstruction of the large intestine appear (48). Such symptoms comprise, but might not be limited to, pain, cramping or bleeding (48). As symptoms are only noticed when the tumour mass starts to grow, which correlates with the increase in its metastatic potential, screening an early diagnosis are crucial to prevent the spread of the disease.

In terms of the progression of the malignancy, adenomatous polyps, sometimes simply referred to as polyps or adenomas, represent the main source from which a tumoral mass may arise (52). This stage is known as the preinvasive neoplasia and its progression depends on both the size and histopathological changes of the tumour (52). Mutations related to cell cycle control and proliferation (APC, KRAS, SMAD4 or the TP53 genes) taking place in the adenoma might also drive the progression of the polyp towards sporadic CRC (53). Recent studies using gene edited organoids have shown the effect that simultaneous mutation of APC, KRAS, SMAD4, PIK3CA and TP53 have on the development of the tumoral mass in mice (54). In this phase, the polyps or the carcinoma mass that may arise from them remain concealed and have not yet invaded other tissues (52).

Mainly three molecular mechanisms have been described to induce a malignant change in the adenoma-carcinoma sequence (ACS) driving the appearance of CRC. Either alone or in combination, chromosomal instability (CIN), CpG island methylator phenotype (CIMP), and microsatellite instability (MSI) can induce changes that lead to the malignancy of benign adenomas (55).

CIN is found in 65 to 70% of all sporadic CRC cases, being the most recurrent alteration in sporadic CRC (55). Initially suggested by Fearon and Vogelstein (56), the tumorigenic process of CRC is initiated by mutations in the newly formed adenoma. Genome-wide sequencing revealed up to 80 mutated genes in colorectal tumours although just few

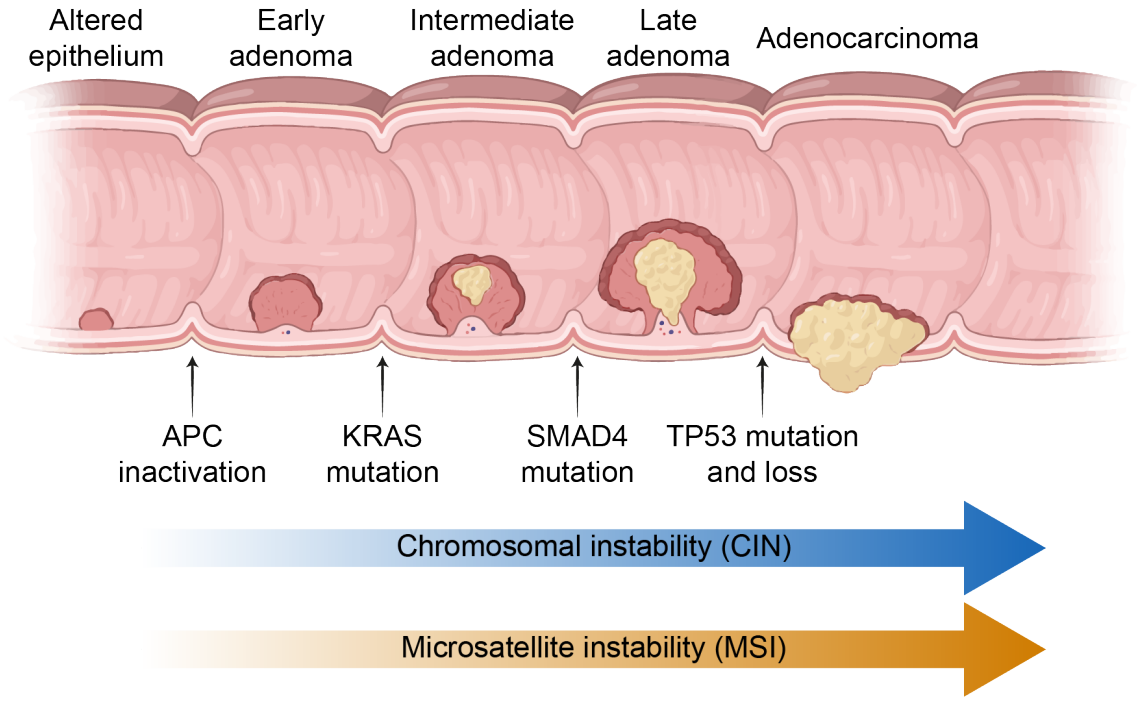


Figure I 1: Adenoma to carcinoma sequence scheme in colorectal cancer.

of them, such as EGFR, MYC or TP53 (57) are considered real initiators of the ACS (58). Remarkably, the order in which such mutations appear is also important (58)(Figure I 1). Taking as an example KRAS and APC, which are often associated with the progression of the ACS, KRAS hyperactivation alone is not able to drive the change from a benign adenoma to a malignant tumour *in vivo* (59). However, in the context of an APC mutated adenoma, KRAS mutation can induce the hyperproliferation of the adenoma cells and thus malignancy (59). Besides point mutations, chromosomal instability can also result in the loss or addition of complete chromosome fragments. Among the most observed losses, the 17p deletion is known to contribute to the metastatic progression of the neoplasia, since the TP53 gene, that encodes for the tumour suppressor protein p53, is encoded in this chromosomal fragment (60, 61).

In CIMP-driven CRC, which is now considered a subtype of CRC (62, 63), an increase in the methylation status of CpG islands leads to the

epigenetic inactivation of hundreds of genes, including tumour suppressor genes (64). Although the exact genes and zones that should be considered to define CIMP-CRC driven cancer are still under debate, a differentiation between CIMP negative and two different CIMP positive phenotypes, CIMP1 and CIMP2, was suggested by Lan et al. (63). Characteristic of CIMP1 is the presence of BRAF mutations and MSI, while CIMP2 often present KRAS mutations but no BRAF or TP53 mutations (63). Other classifications have also been suggested based on different criteria like the methylation status of a series of key genes (62) or the presence or absence of MSI (64).

Colorectal cancer staging

CRC staging is made based on the depth to which the neoplasia has invaded the colorectal tissue (65). Invasion through the muscular mucosa marks the transition between the *in situ* carcinoma (T_{is}) onto Stage I. Further invasion to the *muscularis propria* and across it towards the *serosa* result in Stage IIA (52). Perforation of the *serosa* leads to the transition to Stage IIB (52). Until this point, no metastasis nor invasion of adjacent lymph nodes has occurred and surgical resection of the cancerous lesions is still possible (52). Although the staging of CRC is of paramount importance for diagnosis and crucial in the proper assessment of the treatment required, the ACS is a complex multi-stepped process with a myriad of factors involved (66). As the malignancy progresses, it will invade the surrounding lymph nodes and keep expanding through the colon and rectum walls — Stage III (52). Finally, the cells that shed from the tumoral mass will reach and colonize distant tissues, marking the last stage of CRC in which distant metastases appear — Stage IV (52). These stages correspond to the Tumour-Nodule-Metastasis (TNM) system, which is currently the most widely accepted and internationally advised staging system. It was introduced by the American Joint Committee on Cancer (AJCC) and the most recent version is the 8th iteration of the

manual (67). This system defines different stages for cancer progression based on a series of anatomical and non-anatomical categories that are grouped together and reflect prognosis (68). Anatomical categories would be the primary tumour, nodule invasion or presence of metastases, while histological information would be an example of non-anatomical categories. The status of these different categories can be evaluated by physical or surgical examination or by imaging. The description of each subcategory can be found in Figure I 2 and Table I 1.

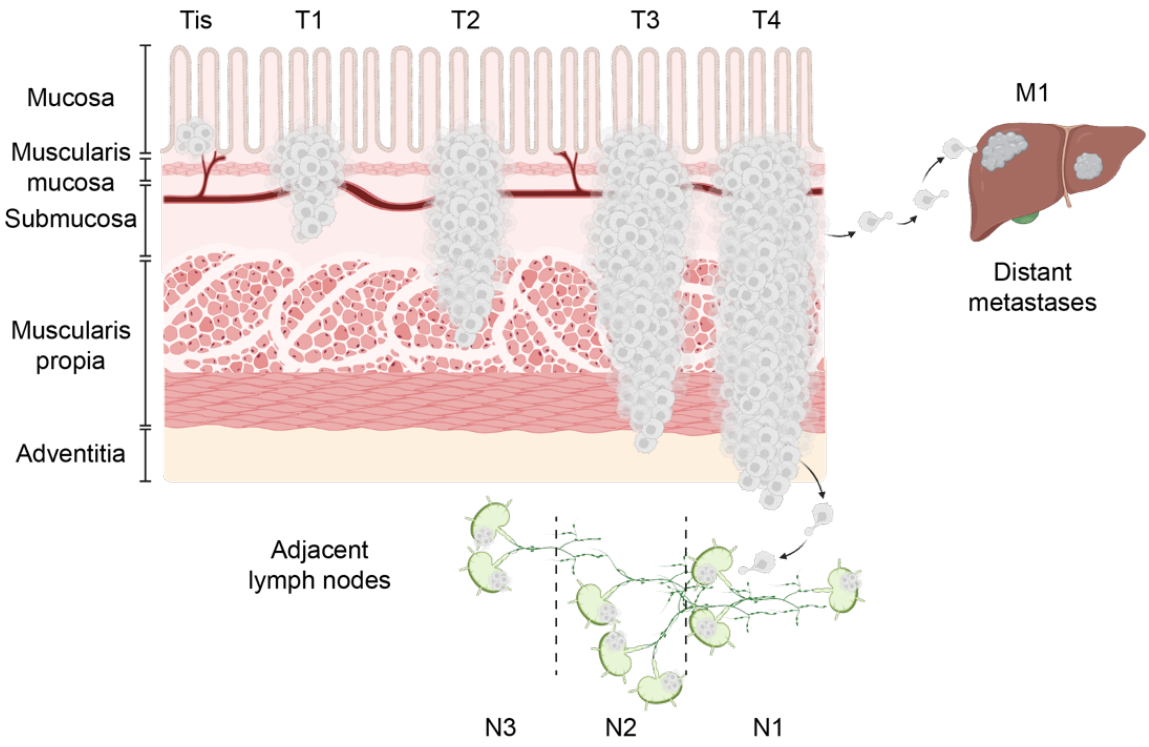


Figure I 2: Scheme of colorectal cancer staging based on the TNM staging.

Table I 1: TNM stages and categories

T category		N category	
Tx	Primary tumour cannot be evaluated	Nx	Regional lymph nodes cannot be evaluated
T0	No evidence of primary Tumour	N0	No metastasis in the regional lymph nodes
Tis	Malignant cells confined in the epithelium	N1	Metastasis in 1-2 regional lymph nodes
T1	Invasion of the lamina propia, muscularis mucosae or submucosa	N2	Metastasis in 3-6 regional lymph nodes
T1a	Invasion of the lamina propia or muscularis mucosae	N3	Metastasis in 7 or more regional lymph nodes
T1b	Invasion of the submucosa	M category	
T2	Invasion of the muscular propia	M0	No distant metastasis
T3	Invasion of the adventitia	M1	Distant metastasis
T4	Invasion of neighbouring structures	G category: histology grade	
T4a	Invasion of the pleura, pericardium, azygos vein, diaphragm or peritoneum	Gx	Unknown grade
		G1	Well differentiated
T4b	Invasion of other structures such as the aortha	G2	Moderately differentiated
		G3	Poorly or undifferentiated

Colorectal cancer diagnosis

As it was mentioned earlier, since the patient prognosis is largely dependent on the stage when CRC is diagnosed, early CRC diagnosis is paramount. However, most of the early symptoms of the malignancy are rather unspecific, like lower abdominal pain or changes in the defecation routine (69). Several meta-analyses have already shown that the diagnostic value of early CRC symptoms is poor and should be accompanied by other sources of information (70-72). Most frequently, early colorectal carcinomas do not present any clear symptoms, and thus, the diagnosis is done either via faecal occult blood test (FOBT), imaging (colonoscopy) or molecular diagnostic based on biomarkers.

Faecal occult blood test and other diagnostic tests

Due to their simplicity and cost-effectiveness, FOBT and faecal immunochemical fest (FIT) are the main non-invasive diagnostic tests used (73). The principle behind FOBT (also known as guaiac-based FOBT) is the detection of the peroxidase activity of haemoglobin's heme group in the stool (74). Similarly, FIT is also based on the detection of haemoglobin although in this case it is detected by specific antibodies against human haemoglobin (74). FIT has several advantages compared to FOBT. The use of specific antibodies against human haemoglobin reduces the ratio of false positive results due to dietary intake of haemoglobin from animal derived blood or of peroxidases present in raw vegetables that would alter the peroxidase-based detection (75). Both tests have rather high specificity for CRC (98-99% for FOBT and 91-98% for FIT) although the sensitivity of FOBT (25-38%) is lower than that of FIT (61-91%) (76). Lately, FIT has been recommended over FOBT due to its higher sensitivity and compliance because of the lack of dietary restrictions and the need for fewer stool samples collection (77). Despite the high specificity and sensitivity values, both tests showed false-positive rates that were close to 50% (78). As a consequence, while these tests

are recommended as a primary source of screening, positive results must be confirmed via colonoscopy.

Endoscopy and image-based diagnostic

All positive FOBT or FIT tests must be confirmed via colonoscopy as this is the most important CRC diagnostic technique. Colonoscopy consists of the visual inspection of the full length of the colon (total colonoscopy) using a flexible tube equipped with a video camera. This technique allows not only for the visualization of the colon to detect the presence of polyps, adenomas or tumours, but also for the collection of tissue samples that can later be subjected to a detailed histological evaluation (79). In addition, small polyps may be completely resected during the process, which is known as polypectomy, making colonoscopy suited not only for diagnosis but also for therapy (79, 80). However, several disadvantages of colonoscopy prevent it from being intensively applied for population screening. Besides its invasiveness other disadvantages of colonoscopy are the high cost and the high level of skill required from the endoscopist, the need for sedation with its corresponding associated risks and the low level of patient compliance (80). A less invasive alternative to total colonoscopy is flexible sigmoidoscopy that is limited to the imaging of the rectum and the distal section of the colon (81). Meta-analysis of the application of sigmoidoscopy report that the rates of colon perforation are reduced by half compared to total colonoscopy, as well as having decreased costs and higher patient compliance (81, 82). Virtual colonoscopy is a last resource imaging alternative for those cases in which endoscopic techniques are not advised. In virtual colonoscopy a 3D representation of the colorectum is generated by merging X-ray images acquired by rotating the imaging plane (83). Less invasive than endoscopic techniques, virtual colonoscopy can have wonderful diagnostic potential but is largely dependent on the skills of the radiologist that interprets the results (83).

Biomarkers and molecular diagnostic of CRC

Molecular diagnostic is based on the detection or characterization of a given molecular footprint (DNA, RNA or protein) that is specifically associated to a disease. These markers can be also helpful in the staging and classification of patients, like the HER2 gene in breast cancer (84). In the context of CRC, the degree of success has been limited, despite the many efforts for finding serum biomarkers that could be used for screening (85). Currently, in the clinic, two biomarkers are mainly used for the detection of CRC: the carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) (86). However, both markers show limited sensitivity and specificity for CRC as their levels also vary in response to other pathologies. Consequently, detection of CA19-9's use is not recommended for diagnosis and CEA is only recommended together with other screening methods to help the prognosis (86). According to the guidelines from the American Society for Clinical Oncology, CEA and CA19-9 should only be used for the evaluation of patient prognosis, detection of recurrence and treatment monitoring (87). Besides CEA and CA19-9, some small microRNAs (miRNAs) have also been postulated as serum diagnostic markers of CRC like miR-19a, miR-210, miR-203 (markers of liver metastasis (88)) or miR-135a and miR-135b. The targets of these microRNA are diverse, miR-19a for example targets the T-cell intracellular antigen 1, a potent tumour suppressor, thus promoting CRC proliferation. miR-210 is a master regulator of hypoxia, with several targets related to cell proliferation and metabolism. miR-203 has also been shown to control cell proliferation through targeting a protein called Hakai while both miR-135 haven shown to target phosphofructokinase-1 to regulate glycolysis, being capable of promoting the progression of pancreatic cancer. Regardless of their target and biological function, none of the miRNA mentioned are currently being used for actual diagnosis.

Genetic biomarkers

As previously mentioned, CRC lacks a clear genetic footprint that could be used for diagnosis (except for hereditary CRC). Nonetheless, over the last years efforts have been made to try to find genetic biomarkers that could serve as diagnostic/prognostic tools (89). These genetic biomarkers could be grouped in those that have proven to be useful for diagnosis/prognosis or those useful for prediction of treatment response (89). In terms of actual diagnostic markers, the genes currently used are related to the same molecular mechanisms involved in triggering a malignant change in the ACS: CIN, CIMP and MSI (90). Within the CIN tumour markers, APC, KRAS, NRAS and BRAF, and in particular the mutational status of KRAS, NRAS and BRAF are important for anti-EGFR therapy (91). APC testing is of special interest as APC germline mutations lead to FAP (32). A test panel based on 5 biomarkers, the so-called Bethesda-markers, has been proposed for the identification of MSI phenotypes which have been associated with better survival and less propensity towards the development of metastatic lesions (85, 90, 92). In addition, MSI status is also important for treatment as studies have already reported a link between this feature and the sensitivity to 5-fluorouracil-based chemotherapy (93). As for CIMP phenotype, no panel of specific methylation markers has been standardized yet. However, hypermethylation of regions like the septin-9 promotor or MLH1 epigenetic silencing have been associated with the early stages of CRC (85).

It is worth noting that, DNA microarrays, widely used for the identification of other cancer types such as breast (94, 95), ovarian (96) or melanomas (97), have not yet been fruitful in CRC. However, one study in 2013 reported the characterization of six subtypes of CRC using gene expression analysis, and the differential response to cetuximab therapy (98).

Proteomic biomarkers

The lack of more sensitive approaches for early detection of CRC opens a whole set of opportunities for proteomic derived biomarker discovery. By using the adequate models or subjects for the different stages and progression of CRC, the identification of proteins that show altered protein expression in these stages represents an outstanding starting point for the discovery of new therapeutic and diagnostic biomarkers. Furthermore, the advances in proteomic analysis techniques in recent years have enabled the identification not only of proteins but also post-translational modifications that might serve as biomarkers of disease (99).

Proteomic approaches focused on CRC biomarkers can be mainly divided into two different categories. On the one hand, differential studies that, by using mass spectrometry, compare protein expression between different conditions, cells or tissues to find proteins that are dysregulated (100-102). On the other hand, humoral response studies that focus on the alterations on the humoral immune response of the organism against tumour derived proteins (103).

Colorectal cancer associated autoantigens

Differential studies can be performed by either working with sera or plasma samples from patients or with tissues or cells. In the first case, protein microarrays can be used to interrogate the sera from control and CRC patients to find proteins that are differentially expressed and could serve as biomarkers (102). Indeed, tumour associated antigens (TAAs) are proteins that can be found in both normal and tumoral cells, but also induce an immune humoral response due to a dysregulation in their expression levels or related to the phenotype they induce (104). Examples of such TAAs include the EGFR (105), EPCAM (104) or p53 (106) and their relevance is related to both their diagnostic/prognostic value and treatment potential (104). In previous works from our research group using commercial and custom-made protein and phage microarrays, a

total of 43 TAAs potential biomarkers for CRC were found (102). The combination of 4 of these TAAs together with an additional peptide from other TAA (identified by phage microarrays) enabled for discriminating between control and CRC patients' sera with great specificity and sensitivity (101, 102).

Besides TAAs, a separate category of tumour antigens known as tumour specific antigens (TSAs) has also been described for the detection of CRC (107). TSAs are neoantigens, i.e., antigens that are specific to the genome of the cancer cell and arise from non-synonymous mutations. They are incorporated to the major histocompatibility complex molecules and recognized by the immune system (107). As their name indicates, TSAs are not expressed in non-tumoral tissues, offering an advantage compared to TAAs as it is the specificity towards the tumour (108). Consequently, TSAs are more useful for cancer immunotherapy than TAAs and have been extensively studied for this purpose (108). Conversely, TAAs are more useful for diagnostic than TSAs. As TSAs arise from mutations, the specific protein that will mutate and the antibody that will be developed against it are difficult to predict and are univocally associated to a phenotype. On the contrary, TAAs are derived from a dysregulation linked to the progression of the tumour and can thus be linked to a particular malignancy.

Autoantibodies and humoral response as biomarkers of CRC

Both TAAs and TSAs will lead to the generation of autoantibodies, i.e., antibodies targeted towards self-molecules. Inside the tumour it is common for proteins to suffer changes in their localisation, structure, abundance or even degradation, all of which can trigger a humoral response from the immune system (109). Similar to the detection of TAAs, microarrays are also frequently used for the identification of autoantibodies. Furthermore, the results derived from TAAs microarrays can be used to prepare complementary autoantibodies microarrays or to

screen for particular antibody expression signatures. For example, in previous research from our group we demonstrated the potential of a panel of autoantibodies for the discrimination between control and CRC patients at early stages (102). Furthermore, the characterization of the exosomes from CRC primary and metastatic surrogates lead to the discovery of a new set of TAAs and autoantibodies with diagnostic capacity to discriminate between control, early and late stages of CRC (110). The main disadvantage of autoantibody detection as CRC diagnostic tool is the elevated cost of protein microarrays. Consequently, autoantibody panels are often derived from the screening of TAAs which are later evaluated via orthogonal techniques, such as enzyme-linked immunosorbent assay (ELISA) or western blot assays.

The tumour microenvironment

The tumour microenvironment (TME) is the collection of immune and stromal cells, blood vessels and extracellular matrix that surround the tumour (111) (Figure I 3). The exact composition of the tumour microenvironment is constantly subjected to changes (112). Furthermore, the tumour and TME establish a dynamic relation where the tumour can modify the TME and, in response, the TME can stimulate cancer resistance and spreading (111). The events taking place in the TME are crucial for the development and progression of the tumour. For instance, the formation of new blood vessels is paramount for the arrival of nutrients to the hypoxic core of the tumour, while anti-tumoral immune response must be modulated for the tumour to continue growing (111, 113, 114). The components of the TME that affect the progression of the tumour could be divided into three categories, immune cells, stromal cells, and the extracellular matrix. The interplay between immune cells and the emerging tumour is a complex process comprised of at least three distinct phases (115). In an initial phase, the so-called ‘elimination phase’, both Natural Killer (NK) cells and $CD8^+$ cytotoxic T (T_C) cells will be able to kill

cancer cells (116, 117), either by recognizing ligands of natural killing receptors (118) or neoantigens presented by antigen-presenting cells (APCs) (119). Eventually, some of the tumoral cells will manage to escape, albeit partially. The surveillance of the immune system enters the 'equilibrium phase'. In this phase, tumoral cells will continue growing and being killed by the immune system, but they will not be completely eliminated (120). Tumoral cells in the equilibrium phase may enter in a dormant estate, without further progressing and waiting for a more favourable microenvironment that enables for the expansion of the

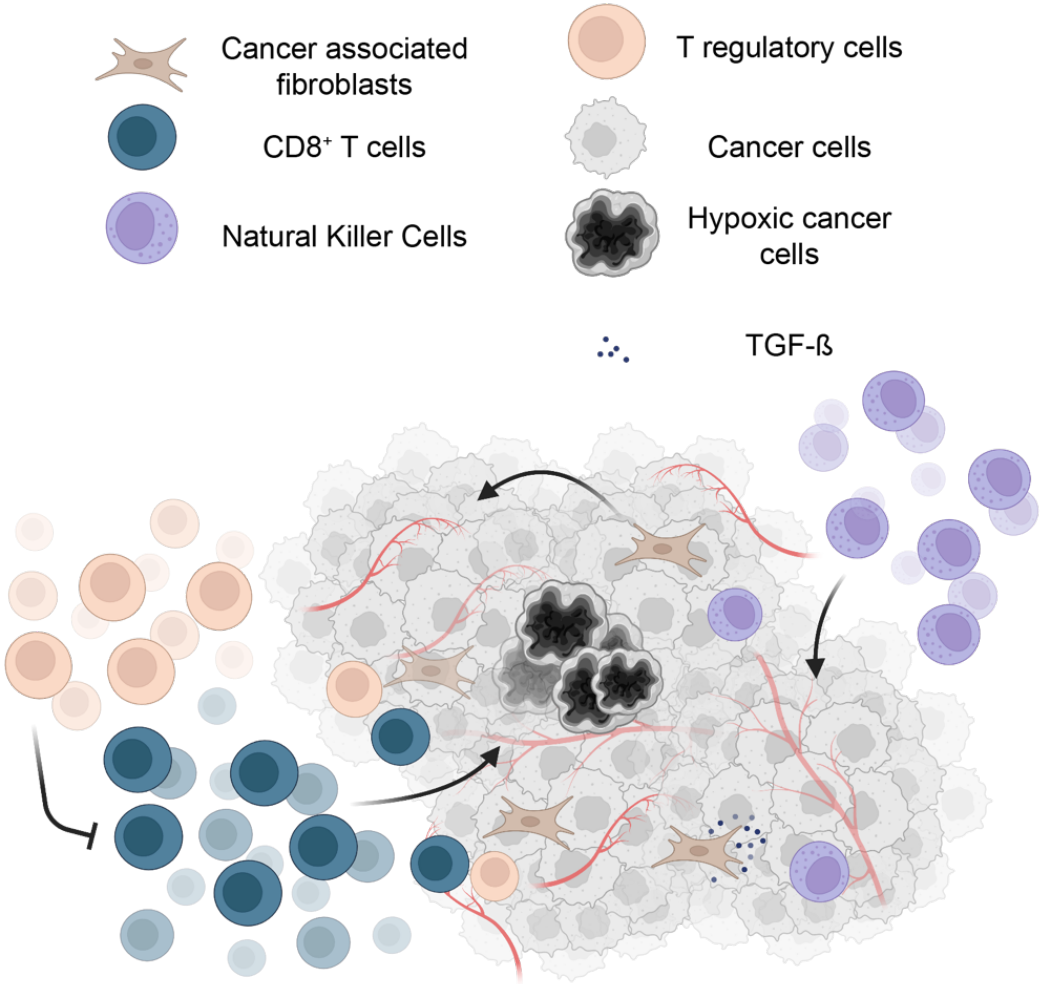


Figure I 3: Tumour microenvironment. The TME is composed of a mixture of cancer cells and other cell types as CAFs and immune cells like NK or T_{reg} cells. The interplay between the heterogenous population present in the TME is a highly complex scenario. As an example of one of the many factors secreted by CAFs and other cells present in the TME TGF-β is shown.

malignancy (121). Angiogenesis has been postulated as one of the mechanisms behind such switch in dormancy (121). Finally, in the 'escape phase' tumoral cells that have gained the ability to avoid immune recognition spread and multiply giving rise to the actual tumour. The mechanisms leading to the escape phase are varied and can be related to both the host and the tumour itself (115). Besides the angiogenic change mentioned previously, other changes in the microenvironment that can drive the escape phase are the secretion of antiapoptotic or immunosuppressive cytokines (as VEGF or TGF-ß) and the recruitment of regulatory T (T_{reg}) cells and other immunomodulatory cells (119).

Stromal cells recruited by the tumour also have an important role in the evolution of the malignancy (111, 122). Tumour cells are able to attract fibroblasts, adipocytes and mesenchymal cells to their surrounding by using different mechanisms, including cytokines and vesicle secretion (123-125). Moreover, recent reports have shown the potential that exosomes from tumoral cells have for providing mesenchymal cells with a tumour-favourable profile (124). Once recruited, a positive feedback loop is established by which the non-cancer-associated cells will transform into cancer-associated stromal cells that will continue secreting cytokines that favour a pro-tumorigenic environment (126, 127). Among the different types of stromal cells found in the TME, cancer associated fibroblasts (CAFs) are of special interest. CAFs will produce most of the extracellular components of the TME, including the extracellular matrix (ECM) and signalling molecules as growth factors (111). In particular, CAFs will secrete TGF-ß, FGF2 and HGF, which stimulate tumour growth and will induce tumoral cells to start the EMT program (111, 127). Furthermore, CAFs will also secrete proteinases that degrade the basement membrane allowing for cancer cells to migrate more easily through the TME (128). Besides this proteinase-based migration model, other reports also suggest that CAFs could stimulate cancer cell migration

in a proteinase independent manner (129). Finally, CAFs are the main secretors and modifiers of the ECM present in the TME.

Although initially overlooked, the role of the ECM in the progression of the tumour has now been shown to be also tremendously relevant (130). For example, NK cells can modify the TME ECM by secreting Fibronectin 1 to alter the architecture that surrounds the tumour and prevent metastasis (131). Both the chemical and mechanical properties of the ECM can affect the tumorigenic capacities of the growing cell mass (132)(Figure I 4). In terms of chemical signalling, the ECM can act as both a reservoir of cytokines and growth factors and a signalling hub by itself by triggering ECM-receptor interactions (111). Overall, it is generally accepted that the high levels of ECM molecules as fibronectin or collagen present in solid tumours are mostly secreted by CAFs and take up to 60% of the total tumour mass (133). Additionally, the components of the TME ECM are also remarkably different to those of the non-malignant tissues (134). A recent proteomic analysis of the ECM of breast cancer xenografts revealed that both stromal and cancer cells can secrete and synthesize collagens, while stromal cells produce the majority of proteoglycans and cancer cells secrete most of the ECM modulating enzymes (135). The exact changes that occur depend on the type of carcinoma. In invasive ductal carcinoma collagen production is shifted towards collagen I and III, while breast carcinomas show an increase in collagen V production (133). Moreover, some studies have already found a role for the ECM in helping to maintain the subset of cancer stem cells (CSCs) inside the tumour (130). More specifically, collagen I, one of the collagens found to be upregulated in the ECM of tumours, can promote the EMT transition in a TGF-ß dependent manner which would be further enhanced in the TME due to the presence of CAFs (136, 137). Induction of EMT would lead part of the differentiated cancer cells towards a stem-like phenotype characteristic of CSC (138). In addition, the increase in ECM deposition will interfere with cell-cell adhesions and polarization, inducing an

increase in the secretion of growth factors and amplifying the feedback loop created by the extra ECM and CAFs (139).

In terms of the signal triggered by ECM molecules, ECM-adhesion through integrins can trigger ERK and PI3K cascades promoting cell growth, survival and migration (140, 141). In addition, matrikines, i.e. components derived from the ECM digestion by cancer and other cells inside the tumour microenvironment, can also regulate angiogenesis and control the angiogenic switch (132). Fragments derived from collagen, elastin or laminin can have opposing effects in the tumour progression

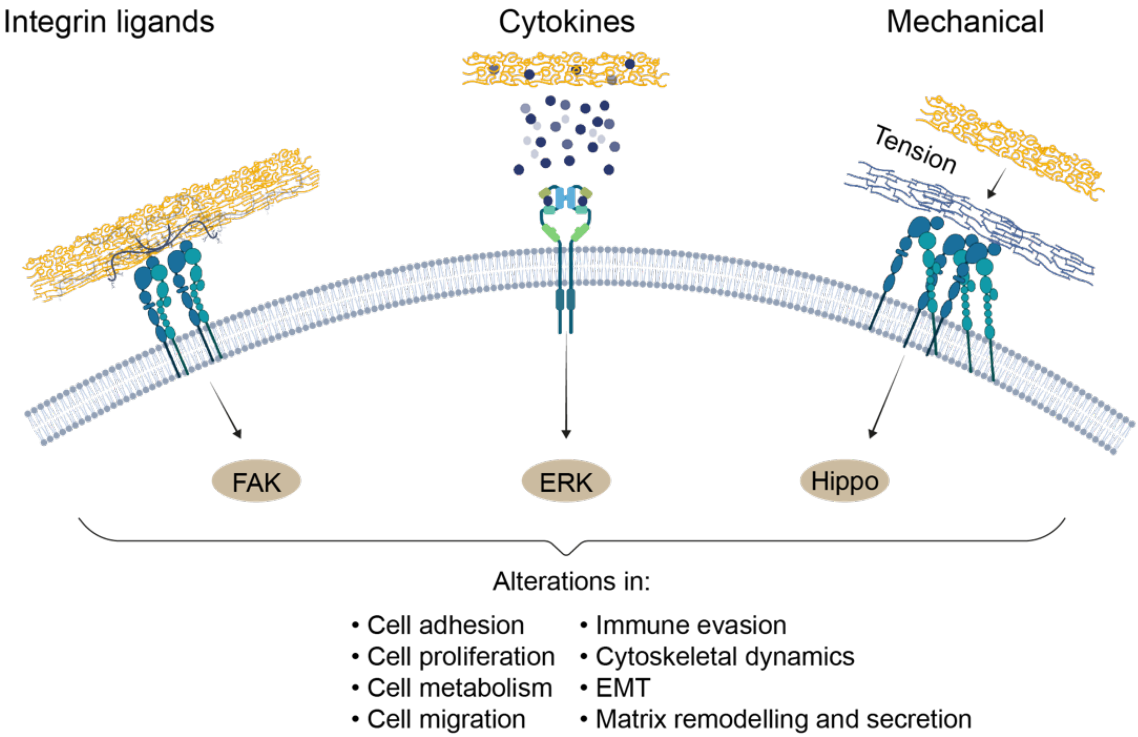


Figure I 4: Intracellular signalling pathways modulated by alterations in the extracellular matrix. From the membrane down in the nucleus, cell-surface receptors act as signalling hubs translating extracellular signals into changes in gene expression. The changes in gene expression control a myriad of processes such as chemosensitivity, EMT programmes, cell differentiation and stemness, proliferation and survival, cell adhesion, migration and invasion, cytoskeletal dynamics. As part of a feedback loop these changes can also regulate the secretion of matrix-remodelling enzymes and matrix proteins from within the tumour.

and the specific effect depends on the cellular receptors being triggered (134). Interestingly, levels of collagen derived matrikines could be used to discriminate between sera from control and CRC patients. This opens the way for a possible use of matrikines as diagnostic/prognostic markers albeit their relatively low levels in sera and other body fluids still hinders such application (134).

At the physical/mechanical level, enhanced ECM secretion makes tumoral tissues significantly stiffer than their non-tumoral counterparts (130). This increased stiffness is mainly caused by collagen and hyaluronan, present in high levels in the TME (142). Collagen shows strain-stiffening, a unique mechanical property characterized by an increase in stiffness when force is applied to the material (143). On the other hand, hyaluronan, negatively charged, generates hydrogel-like networks that can buffer compressive stresses (142).

A central player in sensing and transducing the mechanical changes of the ECM is the Hippo/YAP/TAZ pathway (144). The exact mechanisms that control how the actin cytoskeleton and tensile stress affect and regulate the YAP pathway are unclear. However, it is well established that stiffer ECMs induce the translocation of YAP and TAZ to the nucleus activating their transcriptional activity (145). In the context of tumour progression, this signalling cascade has been shown to induce proliferation (146) and to regulate immune evasion (147) and CSC renewal (148, 149). In addition, the stiffening of the ECM can lead to the secretion of TGF-ß and drive changes in the phenotype of cancer cells towards becoming more aggressive and motile (150). In parallel, the increased ECM stiffness also drives pro-angiogenic signals and VEGF secretion to promote the formation of new blood vessels that irrigate the tumour mass (151). Cancer cells can even physically manipulate the ECM and remodel it by force-mediated realignment of its fibres to generate passages for migration (132).

Epithelial to mesenchymal transition in the context of metastasis

The EMT is a biological process that leads to the conversion of an epithelial cell into a mesenchymal cell (152). In this process the epithelial cells lose their polarization as well as the inter-cellular interactions and links with the surrounding cellular matrix, acquiring properties of mesenchymal cells (Figure I 5). These cells have a very distinct phenotype compared with epithelial cells. Mesenchymal cells show enhanced migratory capacity and invasiveness that is often accompanied by enhanced ECM secretion (153). In the context of cancer progression and metastasis, EMT can induce the transition of tumoral cells from the primary tumour to a more malignant/mesenchymal phenotype with further invasive capacities (152, 154). Recent studies have demonstrated that the transition is more similar to a continuum spectrum rather than just an on and off switch (155-157). According to this model, the initial signals sensed by malignant cells in the primary tumour, that led to a more invasive phenotype, would not be present in the colonized tissue. Consequently, cells would transit back towards an epithelial phenotype allowing for implantation and further development of the metastasis (158).

EMT controlled activation during embryogenesis and wound healing is triggered, mainly, by the activation of SNAI1 and SNAI2 (also known as Slug), respectively (159). During development, fibroblast derived growth factors (FGFs), hepatocyte growth factor (HGF), TGF-ß, BMPs, WNTs or Notch, will activate SNAI1 that will subsequently supress the expression of E-cadherin and other epithelial markers (160). Similar pathways will be activated for wound healing re-epithelialization, although in this case EMT-inhibiting transcription factors will also be activated to keep a tighter control of cell fate and migratory capacities (161). In the tumour itself, EMT markers have been found at the invasive front of carcinomas and circulating tumour cells (159). However, the prognostic value of EMT markers remains still debatable. While numerous reports indicate that

EMT markers could be used as prognostic markers (162-165), some research groups suggest that such correlation is highly dependent on the tissue and cancer subtype (166).

Besides the clear implications of EMT and the reverse mesenchymal-to-epithelial transition (MET) on cell motility and migration, it has also been found that EMT could play a role in immune evasion and the generation of CSCs (154, 159). It has been shown that overexpression of SNAI1 in murine and human melanoma cells induced the activation of T_{reg} cells and prevented dendritic cells activity (167). Furthermore, the

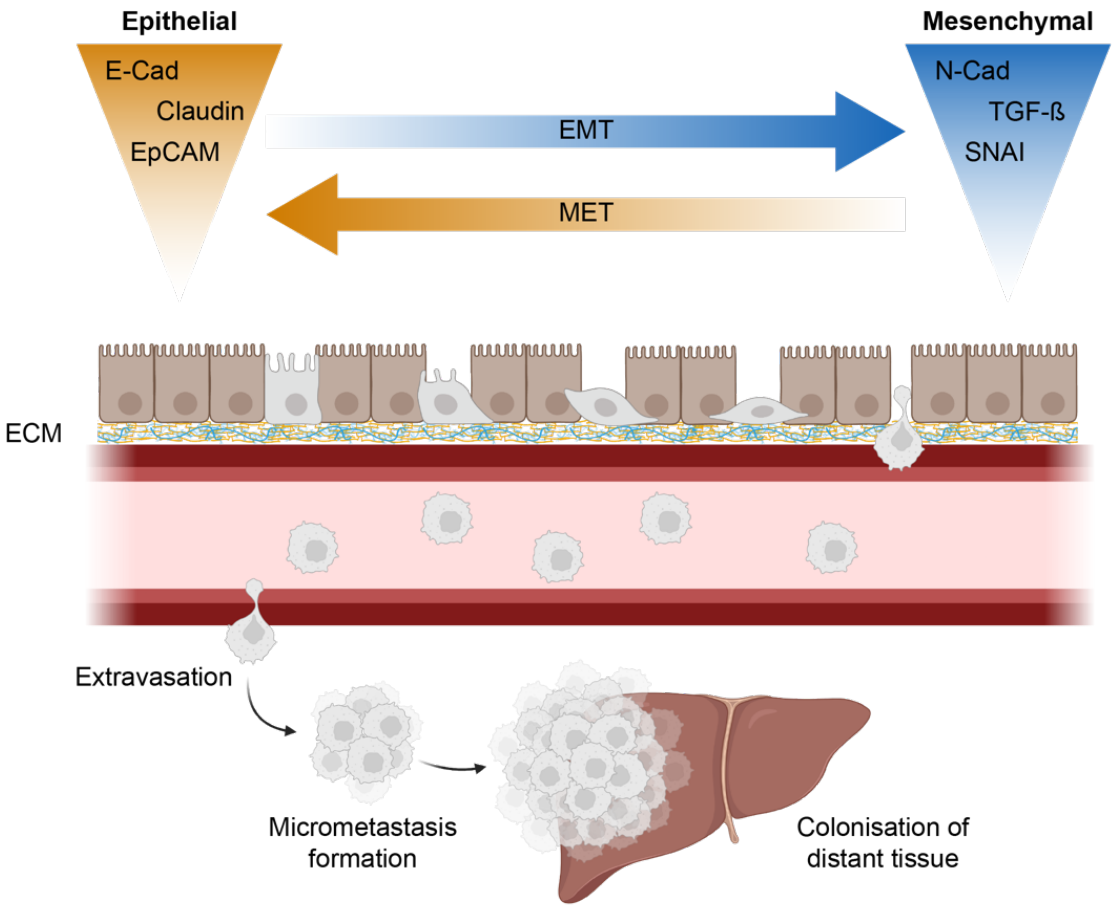


Figure I 5: Epithelial to mesenchymal transition and its role in the migration of tumoral cells and invasion of distant tissues. N-Cad: N-Cadherin, E-Cad: E-Cadherin

silencing of SNAI1 in melanomas that did not respond to immunotherapy significantly reduced tumour growth and metastasis after an increase in tumour-specific and -infiltrating lymphocytes and general systemic immune response (167). In addition, researchers also found that cancer cells undergoing EMT secrete cytokines that result in a more immunosuppressive environment (168). It has also been found that EMT can induce the up-regulation of proteins involved in immune checkpoint such as Programmed Death Ligand-1 and -2 (PD-L1 and PD-L2) (159). On the other hand, it was observed that overexpression of Snail or Twist in tumour-transformed mammary epithelial cells induced an increase in tumorigenic and stem properties of these cells (169). Altogether, the evidence gathered so far point to a clear involvement of the EMT program in the progression and development of tumours, albeit some of the mechanisms or switches might remain partly undefined.

Proteomics and Mass spectrometry

A proteome is defined as the group of proteins produced and/or modified by an organism or living system at a defined point in time and space (170). Proteomics is a collection of methodologies whose ultimate goal is to identify protein footprints that can be associated to particular conditions (171). Among the different approaches we can differentiate between qualitative and quantitative proteomics. In the latter, the aim is to identify those proteins or modifications that would be specific to each of the conditions being studied. As an example, in oncoproteomics the main goal is to identify those proteins associated to a particular malignancy compared with a healthy status. On the other hand, qualitative proteomics focuses on the complete characterization of a set of proteins being expressed. In parallel, proteomics can also be used to analyse the collection of post-translational modifications and protein-protein interactions of a given condition (172).

Proteomic techniques could be divided into two groups depending on whether they are based on mass spectrometry (MS) or not (173). In MS-based proteomics, isolated proteins are examined. This isolation can be done via either mono- or bidimensional polyacrylamide electrophoresis. After the initial isolation step, proteins are digested, and the resulting peptides analysed using MS (170). MS enables the separation and mass identification of all the peptides in the digestion mixture. Identification is done by considering both the enzyme with which the peptides were digested, most frequently trypsin, and the charge/mass (m/z) ratio of the identified peptides. The identified peptides are compared against a database containing the theoretical peptides obtained from the digestion of the known proteins and correspondingly assigned (170). Liquid chromatography (LC) can also be used for the isolation and separation of the digestion peptides (174). In LC-MS/MS, reverse phase chromatography is generally used to separate the different peptides based on their hydrophobicity for a direct subsequent analysis in a mass spectrometer (175). This general approach described corresponds to what is known as bottom-up proteomics (170).

Alternatively, in top-down proteomics the isolated proteins are not digested by trypsin, or other proteases, and thus the intact proteins are introduced into the gas phase of the spectrometer using electrospray ionization (ESI) and fragmented in the actual spectrometer (176). Compared to bottom-up proteomics, top-down simplifies protein identification as there is no need for protein inference (177). Top-down proteomics have other advantages over bottom-up, namely the possible loss of post-translational modifications or the omission of large fragments that can happen in bottom-up proteomics (178). However, top-down proteomics is not without its problems and the technical difficulties associated with intact protein proteome-wide studies have hindered the advance of top-down proteomics and lagged it behind bottom-up proteomics (178).

Paramount to the astounding progress seen in the field of proteomics are the innovations in the field of mass spectrometry. The first step in the process of the MS analysis is the ionization of the sample, which in the case of peptides is achieved either via ESI or matrix-assisted laser desorption/ionization (MALDI) (179). Once the peptides have been converted into ionized analytes the mass analyser is the component in charge of determining the m/z ratios. For proteomics, the most common used mass analysers are quadropoles, ion traps, time-of-flight (TOF) or Fourier-transform ion cyclotron resonance analysers (179). Each of the different type of analysers have marked differences in terms of design and performance with their corresponding advantages and disadvantages (171). Consequently, and to take advantage of the strengths of each type, different analysers are frequently used in tandem rather than just in a stand-alone configuration (171). Some common tandem configurations can be seen in Figure I 6.

Quantitative proteomics

As much as the proteomics field has strived to achieve the complete proteome characterization, knowing the identity of the proteins that are part of such proteome is not the only important information. The levels of expression of one protein can drastically alter its function and consequently quantitative proteomics are fundamental in biomedical research and molecular medicine (180). In quantitative proteomics, the absolute or relative abundance of the proteins present in the proteome of two or more conditions are compared. To detect and measure the abundance of the proteins in each proteome either isotopic labelling or label-free (LF) MS approaches can be used (181). The use of isotopic labelling simplifies the data acquisition process as samples from the different conditions can be combined in the same MS run (181). Isotopic labelling can be done either via chemical linking with a reagent isobarically labelled (Tandem Mass Tag (TMT) or iTraq) (182, 183) or via metabolic

uptake of labelled aminoacids (Stable isotope labelling by amino acids in cell culture (SILAC)) (184).

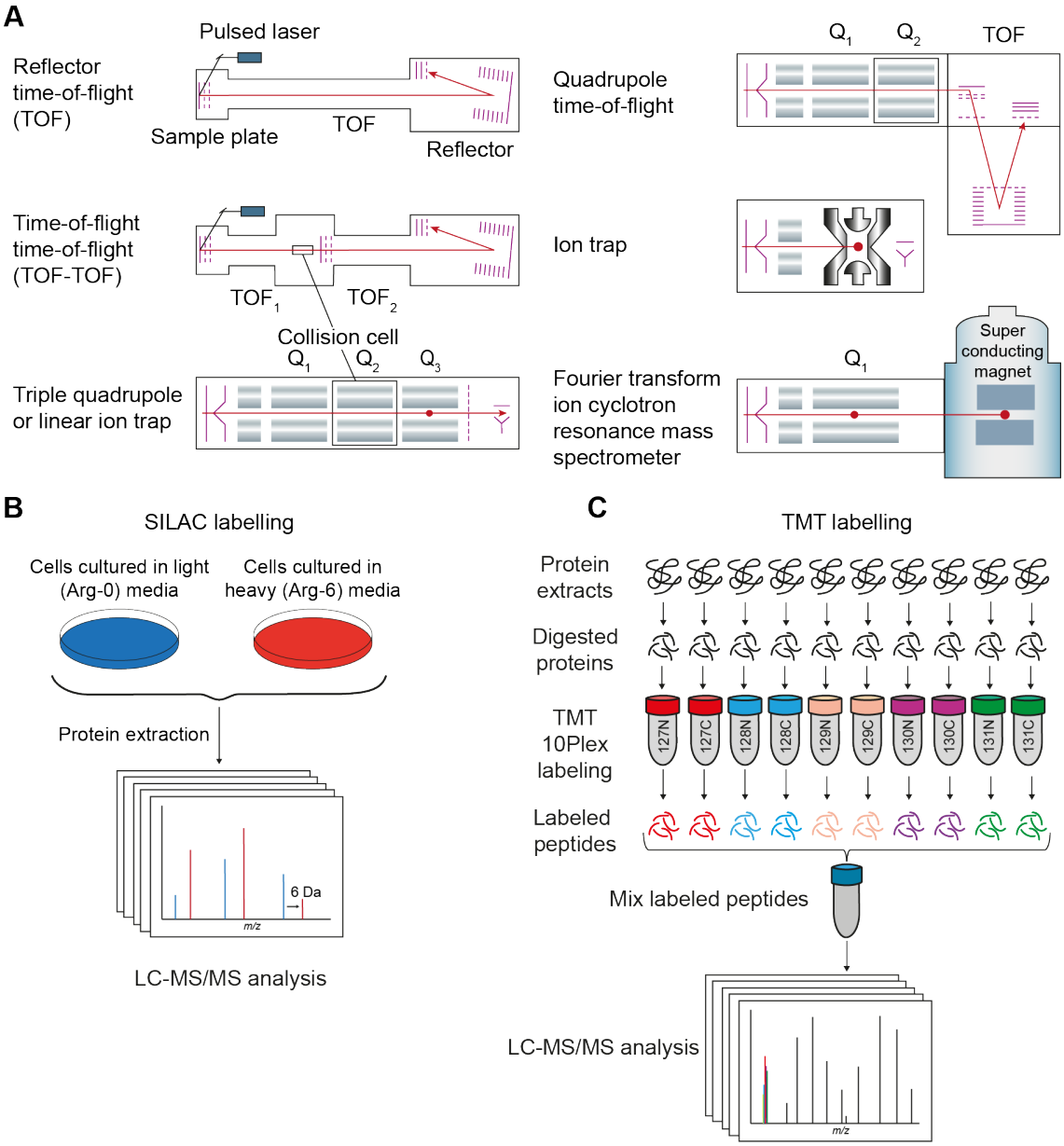


Figure I 6: Mass spectrometer and labelling configurations for proteomics. (**A**) Different mass spectrometer configuration and technologies that can be used for proteomics. (**B**) SILAC isotopic labelling. In SILAC labelling, cells are cultured with media containing aminoacids with a light or heavy isotope, leading to a difference in the mass of the peptides and m/z ratio that can be used to differentiate conditions. (**C**) TMT isobaric labelling. In TMT labelling a whole protein extract is labelled with a mass reporter and the different extracts can then be analysed together.

References

1. Understanding What Cancer Is: Ancient Times to Present.
2. Ahmad AS, Ormiston-Smith N, Sasieni PD. Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960. British Journal of Cancer. 2015;112(5):943-7.
3. Kenneth DK, Jiaquan X, Elizabeth A. Mortality in the United States, 2019. NCHS Data Brief. 2020;395.
4. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell. 2000;100(1):57-70.
5. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;144(5):646-74.
6. Cooper GM. The Development and Causes of Cancer. Sinauer Associates; 2000.
7. Nishida N, Yano H, Nishida T, Kamura T, Kojiro M. Angiogenesis in Cancer. Vascular Health and Risk Management. 2006;2(3):213-9.
8. Bredholt G, Mannelqvist M, Stefansson IM, Birkeland E, Bø TH, Øyan AM, et al. Tumor necrosis is an important hallmark of aggressive endometrial cancer and associates with hypoxia, angiogenesis and inflammation responses. Oncotarget. 2015;6(37):39676-91.
9. Jiao D, Cai Z, Choksi S, Ma D, Choe M, Kwon H-JJ, et al. Necroptosis of tumor cells leads to tumor necrosis and promotes tumor metastasis. Cell Research. 2018;28(8):868-70.
10. Ling Y-h, Chen J-w, Wen S-h, Huang C-y, Li P, Lu L-h, et al. Tumor necrosis as a poor prognostic predictor on postoperative survival of patients with solitary small hepatocellular carcinoma. BMC Cancer. 2020;20(1):1-9.
11. Ford CA, Petrova S, Pound JD, Voss JJLP, Melville L, Paterson M, et al. Oncogenic Properties of Apoptotic Tumor Cells in Aggressive B Cell Lymphoma. Current Biology. 2015;25(5):577-88.
12. Gregory CD, Paterson M. An apoptosis-driven ‘onco-regenerative niche’: roles of tumour-associated macrophages and extracellular vesicles. Philosophical Transactions of the Royal Society B: Biological Sciences. 2018;373(1737).
13. Bielenberg DR, Zetter BR. The Contribution of Angiogenesis to the Process of Metastasis. Cancer journal (Sudbury, Mass). 2015;21(4):267-73.
14. Butler TP, Gullino PM. Quantitation of Cell Shedding into Efferent Blood of Mammary Adenocarcinoma. Cancer Research. 1975;35(3).

15. Fidler IJ. Metastasis: Quantitative Analysis of Distribution and Fate of Tumor Emboli Labeled With 125I-5-Iodo-2′ -deoxyuridine23. JNCI: Journal of the National Cancer Institute. 1970;45(4):773-82.
16. Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y. Molecular principles of metastasis: a hallmark of cancer revisited. Signal Transduction and Targeted Therapy. 2020;5(1):1-17.
17. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. Science. 2011;331(6024):1559-64.
18. Dillekås H, Rogers MS, Straume O. Are 90% of deaths from cancer caused by metastases? Cancer Medicine. 2019;8(12):5574-6.
19. Seyfried TN, Huysentruyt LC. On the origin of cancer metastasis. Critical Reviews in Oncogenesis. 2013;18(1-2):43-73.
20. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA: A Cancer Journal for Clinicians. 2021;71(1):7-33.
21. van Zijl F, Krupitza G, Mikulits W. Initial steps of metastasis: cell invasion and endothelial transmigration. Mutat Res. 2011;728(1-2):23-34.
22. Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nature Reviews Cancer. 2003;3(6):453-8.
23. Talmadge JE, Fidler IJ. AACR Centennial Series: The Biology of Cancer Metastasis: Historical Perspective. Cancer research. 2010;70(14):5649-69.
24. Liu Q, Zhang H, Jiang X, Qian C, Liu Z, Luo D. Factors involved in cancer metastasis: a better understanding to “seed and soil” hypothesis. Molecular Cancer. 2017;16(1):1-19.
25. Gao Y, Bado I, Wang H, Zhang W, Rosen JM, Zhang XHF. Metastasis Organotropism: Redefining the Congenial Soil. Developmental Cell. 2019;49(3):375-91.
26. Obenauf AC, Massagué J. Surviving at a distance: organ specific metastasis. Trends in cancer. 2015;1(1):76-91.
27. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209-49.
28. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: An overview. International journal of cancer. 2021;149(4):778-89.
29. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Przegląd Gastroenterologiczny. 2019;14(2):89-103.

30. Shah SA, Haddad R, Al-Sukhni W, Kim RD, Greig PD, Grant DR, et al. Surgical resection of hepatic and pulmonary metastases from colorectal carcinoma. Journal of the American College of Surgeons. 2006;202(3):468-75.
31. J C, Cr S, Chakedis J, Schmidt CR. Surgical Treatment of Metastatic Colorectal Cancer. Surgical oncology clinics of North America. 2018;27(2):377-99.
32. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, et al. Colorectal cancer. Nature Reviews Disease Primers. 2015;1(1):1-25.
33. Armelao F, De Pretis G. Familial colorectal cancer: A review. World Journal of Gastroenterology. 2014;20(28):9292-8.
34. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, et al. Meta-analyses of Colorectal Cancer Risk Factors. Cancer causes & control : CCC. 2013;24(6):1207-22.
35. Fernandez-Navarro P, Roquette R, Nuñez O, de Sousa-Uva M, Garcia-Perez J, Lopez-Abente G, et al. Atlas of Cancer Mortality in Portugal and Spain (2003–2012): Instituto de Salud Carlos III (Spain); 2021 2021/12//.
36. Cottet V, Pariente A, Nalet B, Lafon J, Milan C, Olschwang S, et al. Colonoscopic Screening of First-Degree Relatives of Patients With Large Adenomas: Increased Risk of Colorectal Tumors. Gastroenterology. 2007;133(4):1086-92.
37. Monahan KJ, Bradshaw N, Dolwani S, Desouza B, Dunlop MG, East JE, et al. Guidelines for the management of hereditary colorectal cancer from the British Society of Gastroenterology (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United Kingdom Cancer Genetics Group (UKCGG). Gut. 2020;69(3):411-44.
38. Burt RW, Barthel JS, Dunn KB, David DS, Drelichman E, Ford JM, et al. Colorectal Cancer Screening. Journal of the National Comprehensive Cancer Network. 2010;8(1):8-61.
39. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895–2015. Nature Reviews Cancer. 2015;15(3):181-94.
40. Idos G, Valle L. Lynch Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW, et al., editors. Seattle: University of Washington, Seattle; 2021.
41. Sekine S, Mori T, Ogawa R, Tanaka M, Yoshida H, Taniguchi H, et al. Mismatch repair deficiency commonly precedes adenoma formation

in Lynch Syndrome-Associated colorectal tumorigenesis. Modern Pathology. 2017;30(8):1144-51.
42. Cohen SA, Pritchard CC, Jarvik GP. Lynch Syndrome: From Screening to Diagnosis to Treatment in the Era of Modern Molecular Oncology. Annual Review of Genomics and Human Genetics. 2019;20:293-307.
43. Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. Orphanet Journal of Rare Diseases. 2009;4(1):4-22.
44. Kantor M, Sobrado J, Patel S, Eiseler S, Ochner C. Hereditary Colorectal Tumors: A Literature Review on MUTYH-Associated Polyposis. Gastroenterology Research and Practice. 2017;2017.
45. Zbuk KM, Eng C. Hamartomatous polyposis syndromes. Nature Clinical Practice Gastroenterology & Hepatology. 2007;4(9):492-502.
46. Jelsig AM, Qvist N, Brusgaard K, Nielsen CB, Hansen TP, Ousager LB. Hamartomatous polyposis syndromes: A review. Orphanet Journal of Rare Diseases. 2014;9(1):1-10.
47. Wang Z-J, Ellis I, Zauber P, Iwama T, Marchese C, Talbot I, et al. Allelic imbalance at the LKB1 (STK11) locus in tumours from patients with Peutz-Jeghers' syndrome provides evidence for a hamartoma–(adenoma)–carcinoma sequence. The Journal of Pathology. 1999;188:9-13.
48. Simon K. Colorectal cancer development and advances in screening. Clinical Interventions in Aging. 2016;11:967-76.
49. Angelo SN, Lourenço GJ, Magro DO, Nascimento H, Oliveira RA, Leal RF, et al. Dietary risk factors for colorectal cancer in Brazil: A case control study. Nutrition Journal. 2016;15(1):1-4.
50. Gonzalez CA, Riboli E. Diet and cancer prevention: Contributions from the European Prospective Investigation into Cancer and Nutrition (EPIC) study. European Journal of Cancer. 2010;46(14):2555-62.
51. Garranzo-Asensio M, Solís-Fernández G, Montero-Calle A, García-Martínez JM, Fiuza MC, Pallares P, et al. Seroreactivity Against Tyrosine Phosphatase PTPRN Links Type 2 Diabetes and Colorectal Cancer and Identifies a Potential Diagnostic and Therapeutic Target. Diabetes. 2022;71(3):497-510.
52. Manne U, Shanmugam C, Katkoori VR, Bumpers HL, Grizzle WE. Development and progression of colorectal neoplasia. Cancer biomarkers : section A of Disease markers. 2010;9(1-6):235-65.
53. Smit WL, Spaan CN, De Boer RJ, Ramesh P, Garcia TM, Meijer BJ, et al. Driver mutations of the adenoma-carcinoma sequence govern the intestinal epithelial global translational capacity. Proceedings of the

National Academy of Sciences of the United States of America. 2020;117(41):25560-70.
54. Matano M, Date S, Shimokawa M, Takano A, Fujii M, Ohta Y, et al. Modeling colorectal cancer using CRISPR-Cas9–mediated engineering of human intestinal organoids. Nature Medicine. 2015;21(3):256-62.
55. Tariq K, Ghias K. Colorectal cancer carcinogenesis: a review of mechanisms. Cancer Biology & Medicine. 2016;13(1):120-35.
56. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61(5):759-67.
57. Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, Parsons DW, et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(42):16224-9.
58. Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology. 2010;138(6):2059-72.
59. Haigis KM, Kendall KR, Wang Y, Cheung A, Haigis MC, Glickman JN, et al. Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. Nature Genetics. 2008;40(5):600-8.
60. Cardoso J, Molenaar L, De Menezes RX, Van Leerdam M, Rosenberg C, Möslein G, et al. Chromosomal Instability in MYH- and APC-Mutant Adenomatous Polyps. Cancer Research. 2006;66(5):2514-9.
61. Orsetti B, Selves J, Bascoul-Mollevi C, Lasorsa L, Gordien K, Bibeau F, et al. Impact of chromosomal instability on colorectal cancer progression and outcome. BMC Cancer. 2014;14(1):121-34.
62. Mojarad EN, Kuppen PJK, Aghdaei HA, Zali MR. The CpG island methylator phenotype (CIMP) in colorectal cancer. Gastroenterology and Hepatology From Bed to Bench. 2013;6(3):120-8.
63. Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, et al. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(47):18654-9.
64. Issa JP. CpG island methylator phenotype in cancer. Nature Reviews Cancer. 2004;4(12):988-93.
65. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more

"personalized" approach to cancer staging. CA: A Cancer Journal for Clinicians. 2017;67(2):93-9.
66. Nguyen HT, Duong HQ. The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy (review). Oncology Letters. 2018;16(1):9-18.
67. Colon and Rectum. In: Greene FL, Compton CC, Fritz AG, Shah JP, Winchester DP, editors. New York, NY: Springer New York; 2006. p. 107-17.
68. Escrig Sos J, Gómez Quiles L, Maiocchi K. The 8th Edition of the AJCC-TNM Classification: New Contributions to the Staging of Esophagogastric Junction Cancer. Cirugía Española (English Edition). 2019;97(8):432-7.
69. Vega P, Valentín F, Cubiella J. Colorectal cancer diagnosis: Pitfalls and opportunities. World Journal of Gastrointestinal Oncology. 2015;7(12):422-33.
70. Jellema P, Van Der Windt DAWM, Bruinvels DJ, Mallen CD, Van Weyenberg SJB, Mulder CJ, et al. Value of symptoms and additional diagnostic tests for colorectal cancer in primary care: systematic review and meta-analysis. The BMJ. 2010;340(7750):21.
71. Astin M, Griffin T, Neal RD, Rose P, Hamilton W. The diagnostic value of symptoms for colorectal cancer in primary care: a systematic review. The British Journal of General Practice. 2011;61(586):e231-e.
72. Ford AC, Veldhuyzen Van Zanten SJO, Rodgers CC, Talley NJ, Vakil NB, Moayyedi P. Diagnostic utility of alarm features for colorectal cancer: systematic review and meta-analysis. Gut. 2008;57(11):1545-53.
73. Ballegooijen Mv, Habbema JDF, Boer R, Zauber AG, Brown ML. A Comparison of the Cost-Effectiveness of Fecal Occult Blood Tests with Different Test Characteristics in the Context of Annual Screening in the Medicare Population. A Comparison of the Cost-Effectiveness of Fecal Occult Blood Tests with Different Test Characteristics in the Context of Annual Screening in the Medicare Population. 2003.
74. Lu M, Luo X, Li N, Chen H, Dai M. Diagnostic Accuracy Of Fecal Occult Blood Tests For Detecting Proximal Versus Distal Colorectal Neoplasia: A Systematic Review And Meta-Analysis. Clinical Epidemiology. 2019;11:943-54.
75. Meklin J, Syrjänen K, Eskelinen M. Colorectal Cancer Screening With Traditional and New-generation Fecal Immunochemical Tests: A Critical Review of Fecal Occult Blood Tests. Anticancer Research. 2020;40(2):575-81.

76. Whitlock EP, Lin JS, Liles E, Beil TL, Fu R. Screening for colorectal cancer: A targeted, updated systematic review for the U.S. Preventive Services Task Force. Annals of Internal Medicine. 2008;149(9):638-58.
77. Akram A, Juang D, Bustamante R, Liu L, Earles A, Ho SB, et al. Replacing the Guaiac Fecal Occult Blood Test With the Fecal Immunochemical Test Increases Proportion of Individuals Screened in a Large Healthcare Setting. Clinical Gastroenterology and Hepatology. 2017;15(8):1265-70.e1.
78. Garcia M, Milà N, Binefa G, Borràs JM, Espinàs JA, Moreno V. False-positive results from colorectal cancer screening in Catalonia (Spain), 2000-2010. Journal of Medical Screening. 2012;19(2):77-82.
79. Atkin WS, Saunders BP. Surveillance guidelines after removal of colorectal adenomatous polyps. Gut. 2002;51(suppl 5):v6-v9.
80. Fyock CJ, Draganov PV. Colonoscopic polypectomy and associated techniques. World Journal of Gastroenterology. 2010;16(29):3630-7.
81. Hawkes EA, Cunningham D. Flexible sigmoidoscopy—valuable in colorectal cancer. Nature Reviews Clinical Oncology. 2010;7(9):488-90.
82. Sharaf RN, Ladabaum U. Comparative Effectiveness and Cost-Effectiveness of Screening Colonoscopy vs. Sigmoidoscopy and Alternative Strategies. Official journal of the American College of Gastroenterology. 2013;108(1).
83. Ganeshan D, Elsayes KM, Vining D. Virtual colonoscopy: Utility, impact and overview. World Journal of Radiology. 2013;5(3):61-7.
84. Tan M, Yu D. Molecular Mechanisms of ErbB2-Mediated Breast Cancer Chemoresistance. Landes Bioscience; 2013.
85. Agrawal S, Bhattacharya A, Manhas J, Sen S. Molecular Diagnostics in Colorectal Cancer. In: Shukla KK, Sharma P, Misra S, editors. Molecular Diagnostics in Cancer Patients. Singapore: Springer Singapore; 2019. p. 143-55.
86. Lakemeyer L, Sander S, Wittau M, Henne-Bruns D, Kornmann M, Lemke J, et al. Diagnostic and Prognostic Value of CEA and CA19-9 in Colorectal Cancer. Diseases. 2021;9(1):21-33.
87. Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, et al. ASCO 2006 Update of Recommendations for the Use of Tumor Markers in Gastrointestinal Cancer. Journal of Clinical Oncology. 2006;24(33):5313-27.
88. Nassar FJ, Msheik ZS, Itani MM, Helou RE, Hadla R, Kreidieh F, et al. Circulating miRNA as Biomarkers for Colorectal Cancer Diagnosis and Liver Metastasis. Diagnostics. 2021;11(2).

89. Coppedè F, Lopomo A, Spisni R, Migliore L. Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. World Journal of Gastroenterology. 2014;20(4):943-56.
90. Tsang AHF, Cheng KH, Wong ASP, Ng SSM, Ma BBY, Chan CML, et al. Current and future molecular diagnostics in colorectal cancer and colorectal adenoma. World Journal of Gastroenterology. 2014;20(14):3847-.
91. De Roock W, De Schutter J, Biesmans B, Tejpar S, Research Centre V, Leuven K, et al. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the effi cacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. Lancet Oncol. 2010;11:753-62.
92. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. Journal of Clinical Oncology. 2005;23(3):609-18.
93. Zhao F, Yuan X, Ren D, Shen G, Wang Z, Zheng F, et al. Predicting the Efficacy of 5-Fluorouracil–Based Adjuvant Chemotherapy in Gastric Cancer by Microsatellite Instability: A Meta-Analysis. 2019;38(1):21-8.
94. Cooper CS. Applications of microarray technology in breast cancer research. Breast Cancer Research. 2001;3(3):158-75.
95. Van't Veer LJ, Dai H, Van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415(6871):530-6.
96. Shridhar V, Lee J, Pandita A, Iturria S, Avula R, Staub J, et al. Genetic Analysis of Early-versus Late-Stage Ovarian Tumors. CANCER RESEARCH. 2001;61:5895-904.
97. De Wit NJW, Burtscher HJ, Weidle UH, Ruiter DJ, Van Muijen GNP. Differentially expressed genes identified in human melanoma cell lines with different metastatic behaviour using high density oligonucleotide arrays. Melanoma research. 2002;12(1):57-69.
98. Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschleger S, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nature medicine. 2013;19(5):619-25.
99. Tomonaga T, Matsushita K, Yamaguchi S, Oh-Ishi M, Kodera Y, Maeda T, et al. Identification of Altered Protein Expression and Post-Translational Modifications in Primary Colorectal Cancer by Using Agarose Two-Dimensional Gel Electrophoresis. Clinical Cancer Research. 2004;10(6):2007-14.

100. Ma H, Chen G, Guo M. Mass spectrometry based translational proteomics for biomarker discovery and application in colorectal cancer. PROTEOMICS – Clinical Applications. 2016;10(4):503-15.
101. Lee PY, Chin S-F, Low TY, Jamal R. Probing the colorectal cancer proteome for biomarkers: Current status and perspectives. Journal of Proteomics. 2018;187:93-105.
102. Babel I, Barderas R, Díaz-Uriarte R, Martínez-Torrecuadrada JL, Sánchez-Carbayo M, Casal JI, et al. Identification of tumor-associated autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays. Molecular and Cellular Proteomics. 2009;8(10):2382-95.
103. Niloofa R, de Zoysa MI, Seneviratne SL. Autoantibodies in the diagnosis, prognosis, and prediction of colorectal cancer. Journal of cancer research and therapeutics. 2021;17(4):819-33.
104. Choi M, Thakur A. Identifying Appropriate Colorectal Cancer-Associated Antigens for the Clinical Trials. Current Colorectal Cancer Reports. 2015;11(1):29-36.
105. Sartore-Bianchi A, Moroni M, Veronese S, Carnaghi C, Bajetta E, Luppi G, et al. Epidermal growth factor receptor gene copy number and clinical outcome of metastatic colorectal cancer treated with panitumumab. Journal of Clinical Oncology. 2007;25(22):3238-45.
106. Bueter M, Gasser M, Schramm N, Lebedeva T, Tocco G, Gerstlauer C, et al. T-cell response to p53 tumor-associated antigen in patients with colorectal carcinoma. International Journal of Oncology. 2006;28(2):431-8.
107. Smith CC, Selitsky SR, Chai S, Armistead PM, Vincent BG, Serody JS. Alternative tumour-specific antigens. Nature Reviews Cancer. 2019;19(8):465-78.
108. Gubin MM, Artyomov MN, Mardis ER, Schreiber RD. Tumor neoantigens: building a framework for personalized cancer immunotherapy. The Journal of Clinical Investigation. 2015;125(9):3413-21.
109. Elkon K, Casali P. Nature and functions of autoantibodies. Nature clinical practice Rheumatology. 2008;4(9):491-8.
110. Montero-Calle A, Aranguren-Abeigon I, Garranzo-Asensio M, Poves C, Fernández-Aceñero MJ, Martínez-Useros J, et al. Multiplexed Biosensing Diagnostic Platforms Detecting Autoantibodies to Tumor-Associated Antigens from Exosomes Released by CRC Cells and Tissue Samples Showed High Diagnostic Ability for Colorectal Cancer. Engineering. 2021;7(10):1393-412.

111. Anderson NM, Simon MC. The tumor microenvironment. Current Biology. 2020;30(16):R921-R5.
112. Cox TR. The matrix in cancer. Nature Reviews Cancer 2021 21:4. 2021;21(4):217-38.
113. Watnick RS. The Role of the Tumor Microenvironment in Regulating Angiogenesis. Cold Spring Harbor Perspectives in Medicine. 2012;2(12).
114. Yuan Y, Jiang YC, Sun CK, Chen QM. Role of the tumor microenvironment in tumor progression and the clinical applications. Oncology Reports. 2016;35(5):2499-515.
115. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: Integrating immunity's roles in cancer suppression and promotion. Science. 2011;331(6024):1565-70.
116. Stojanovic A, Cerwenka A. Natural Killer Cells and Solid Tumors. Journal of Innate Immunity. 2011;3(4):355-64.
117. Weigelin B, Krause M, Friedl P. Cytotoxic T lymphocyte migration and effector function in the tumor microenvironment. Immunology Letters. 2011;138(1):19-21.
118. Marcus A, Gowen BG, Thompson TW, Iannello A, Ardolino M, Deng W, et al. Recognition of tumors by the innate immune system and natural killer cells. Advances in immunology. 2014;122:91-128.
119. Chew V, Toh HC, Abastado JP. Immune microenvironment in tumor progression: Characteristics and challenges for therapy. Journal of Oncology. 2012.
120. Kubick BJ, Fan X, Crouch A, McCarthy R, Roop DR. Tracing the Equilibrium Phase of Cancer Immunoediting in Epidermal Neoplasms via Longitudinal Intravital Imaging. Journal of Investigative Dermatology. 2020;140(4):891-900.e10.
121. Almog N. Molecular mechanisms underlying tumor dormancy. Elsevier Ireland Ltd; 2010. p. 139-46.
122. Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. Breast Cancer Research. 2016;18(1):1-11.
123. Hill BS, Sarnella A, D'Avino G, Zannetti A. Recruitment of stromal cells into tumour microenvironment promote the metastatic spread of breast cancer. Seminars in Cancer Biology. 2020;60:202-13.
124. Lin LY, Du LM, Cao K, Huang Y, Yu PF, Zhang LY, et al. Tumour cell-derived exosomes endow mesenchymal stromal cells with tumour-promotion capabilities. Oncogene. 2016;35(46):6038-42.

125. Arora L, Pal D. Remodeling of Stromal Cells and Immune Landscape in Microenvironment During Tumor Progression. Frontiers in Oncology. 2021;11:401-21.
126. Guo S, Deng CX. Effect of Stromal Cells in Tumor Microenvironment on Metastasis Initiation. International Journal of Biological Sciences. 2018;14(14):2083-93.
127. Denton AE, Roberts EW, Fearon DT, Denton AE, Roberts EW, Fearon DT. Stromal Cells in the Tumor Microenvironment. Advances in Experimental Medicine and Biology. 2018;1060:99-114.
128. Nii T, Makino K, Tabata Y. A cancer invasion model of cancer-associated fibroblasts aggregates combined with TGF-β1 release system. Regenerative Therapy. 2020;14:196-204.
129. Glentis A, Oertle P, Mariani P, Chikina A, El Marjou F, Attieh Y, et al. Cancer-associated fibroblasts induce metalloprotease-independent cancer cell invasion of the basement membrane. Nature Communications. 2017;8(1):1-13.
130. Nallanthighal S, Heiserman JP, Cheon DJ. The Role of the Extracellular Matrix in Cancer Stemness. Frontiers in Cell and Developmental Biology. 2019;7:86-90.
131. Glasner A, Levi A, Enk J, Isaacson B, Viukov S, Orlanski S, et al. NKp46 Receptor-Mediated Interferon-γ Production by Natural Killer Cells Increases Fibronectin 1 to Alter Tumor Architecture and Control Metastasis. Immunity. 2018;48(1):107-19.e4.
132. Winkler J, Abisoye-Ogunniyan A, Metcalf KJ, Werb Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nature Communications 2020 11:1. 2020;11(1):1-19.
133. Henke E, Nandigama R, Ergün S. Extracellular Matrix in the Tumor Microenvironment and Its Impact on Cancer Therapy. Frontiers in Molecular Biosciences. 2020;6:160-84.
134. Brassart-Pasco S, Brézillon S, Brassart B, Ramont L, Oudart JB, Monboisse JC. Tumor Microenvironment: Extracellular Matrix Alterations Influence Tumor Progression. Frontiers in Oncology. 2020;10:397-410.
135. Hebert JD, Myers SA, Naba A, Abbruzzese G, Lamar JM, Carr SA, et al. Proteomic Profiling of the ECM of Xenograft Breast Cancer Metastases in Different Organs Reveals Distinct Metastatic Niches. Cancer Research. 2020;80(7):1475-85.
136. Shintani Y, Maeda M, Chaika N, Johnson KR, Wheelock MJ. Collagen I Promotes Epithelial-to-Mesenchymal Transition in Lung Cancer Cells via Transforming Growth Factor–β Signaling. American Journal of Respiratory Cell and Molecular Biology. 2008;38(1):95-104.

137. Yu Y, Xiao CH, Tan LD, Wang QS, Li XQ, Feng YM. Cancer-associated fibroblasts induce epithelial–mesenchymal transition of breast cancer cells through paracrine TGF-β signalling. British Journal of Cancer 2014 110:3. 2013;110(3):724-32.
138. Singh A, Settleman J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. Oncogene 2010 29:34. 2010;29(34):4741-51.
139. Walker C, Mojares E, Del Río Hernández A. Role of Extracellular Matrix in Development and Cancer Progression. International Journal of Molecular Sciences. 2018;19(10).
140. Mammoto T, Ingber DE. Mechanical control of tissue and organ development. Development (Cambridge, England). 2010;137(9):1407-20.
141. Hastings JF, Skhinas JN, Fey D, Croucher DR, Cox TR. The extracellular matrix as a key regulator of intracellular signalling networks. British Journal of Pharmacology. 2019;176(1):82-92.
142. Gkretsi V, Stylianopoulos T. Cell Adhesion and Matrix Stiffness: Coordinating Cancer Cell Invasion and Metastasis. Frontiers in Oncology. 2018;8:145-52.
143. Motte S, Kaufman LJ. Strain stiffening in collagen I networks. Biopolymers. 2013;99(1):35-46.
144. Zinatizadeh MR, Miri SR, Zarandi PK, Chalbatani GM, Rapôso C, Mirzaei HR, et al. The Hippo Tumor Suppressor Pathway (YAP/TAZ/TEAD/MST/LATS) and EGFR-RAS-RAF-MEK in cancer metastasis. Genes & Diseases. 2021;8(1):48-60.
145. Moroishi T, Hansen CG, Guan KL. The emerging roles of YAP and TAZ in cancer. Nature reviews Cancer. 2015;15(2):73-9.
146. Zanconato F, Forcato M, Battilana G, Azzolin L, Quaranta E, Bodega B, et al. Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. Nature Cell Biology 2015 17:9. 2015;17(9):1218-27.
147. Pan Z, Tian Y, Cao C, Niu G. The Emerging Role of YAP/TAZ in Tumor Immunity. Molecular Cancer Research. 2019;17(9):1777.
148. Basu-Roy U, Bayin NS, Rattanakorn K, Han E, Placantonakis DG, Mansukhani A, et al. Sox2 antagonizes the Hippo pathway to maintain stemness in cancer cells. Nature Communications 2015 6:1. 2015;6(1):1-14.
149. Cordenonsi M, Zanconato F, Azzolin L, Forcato M, Rosato A, Frasson C, et al. The hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. Cell. 2011;147(4):759-72.

150. Li X, Wang J. Mechanical tumor microenvironment and transduction: Cytoskeleton mediates cancer cell invasion and metastasis. International Journal of Biological Sciences. 2020;16(12):2014-28.
151. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. EMBO reports. 2014;15(12):1243-53.
152. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. The Journal of Clinical Investigation. 2009;119(6):1420-8.
153. Kalluri R, Neilson EG. Epithelial-mesenchymal transition and its implications for fibrosis. Journal of Clinical Investigation. 2003;112(12):1776-84.
154. Mittal V. Epithelial Mesenchymal Transition in Tumor Metastasis. Annual Review of Pathology: Mechanisms of Disease. 2018;13:395-412.
155. Panchy N, Azeredo-Tseng C, Luo M, Randall N, Hong T. Integrative Transcriptomic Analysis Reveals a Multiphasic Epithelial–Mesenchymal Spectrum in Cancer and Non-tumorigenic Cells. Frontiers in Oncology. 2020;9:1479-500.
156. Bakir B, Chiarella AM, Pitarresi JR, Rustgi AK, Rustgi AK. EMT, MET, Plasticity, and Tumor Metastasis. Trends in Cell Biology. 2020;30(10).
157. Bhatia S, Wang P, Toh A, Thompson EW. New Insights Into the Role of Phenotypic Plasticity and EMT in Driving Cancer Progression. Frontiers in Molecular Biosciences. 2020;7:71-89.
158. Hamilton G, Rath B, Magbanua MJM, Park JW. Mesenchymal-Epithelial Transition and Circulating Tumor Cells in Small Cell Lung Cancer. Advances in Experimental Medicine and Biology. 2017;994:229-45.
159. Aiello NM, Kang Y. Context-dependent EMT programs in cancer metastasis. Journal of Experimental Medicine. 2019;216(5):1016-26.
160. Barrallo-Gimeno A, Nieto MA. The Snail genes as inducers of cell movement and survival: implications in development and cancer. Development. 2005;132(14):3151-61.
161. Haensel D, Dai X. Epithelial-to-mesenchymal transition in cutaneous wound healing: where we are and where we are heading. Developmental dynamics : an official publication of the American Association of Anatomists. 2018;247(3):473-80.
162. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(35):15449-54.

163. Loboda A, Nebozhyn MV, Watters JW, Buser CA, Shaw PM, Huang PS, et al. EMT is the dominant program in human colon cancer. BMC Medical Genomics. 2011;4:9-19.
164. Byers LA, Diao L, Wang J, Saintigny P, Girard L, Peyton M, et al. An epithelial-mesenchymal transition (EMT) gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clinical cancer research : an official journal of the American Association for Cancer Research. 2013;19(1):279-90.
165. Imani S, Hosseinifard H, Cheng J, Wei C, Fu J. Prognostic Value of EMT-inducing Transcription Factors (EMT-TFs) in Metastatic Breast Cancer: A Systematic Review and Meta-analysis. Scientific Reports 2016 6:1. 2016;6(1):1-10.
166. George JT, Jolly MK, Xu S, Somarelli JA, Levine H. Survival outcomes in cancer patients predicted by a partial EMT gene expression scoring metric. Cancer research. 2017;77(22):6415-28.
167. Kudo-Saito C, Shirako H, Takeuchi T, Kawakami Y. Cancer Metastasis Is Accelerated through Immunosuppression during Snail-Induced EMT of Cancer Cells. Cancer Cell. 2009;15(3):195-206.
168. Yaguchi T, Sumimoto H, Kudo-Saito C, Tsukamoto N, Ueda R, Iwata-Kajihara T, et al. The mechanisms of cancer immunoescape and development of overcoming strategies. International Journal of Hematology. 2011;93(3):294-300.
169. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 2008;133(4):704-15.
170. Dupree EJ, Jayathirtha M, Yorkey H, Mihasan M, Petre BA, Darie CC. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. Proteomes. 2020;8(3):1-26.
171. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003 422:6928. 2003;422(6928):198-207.
172. Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. Proteomics: Technologies and Their Applications. Journal of Chromatographic Science. 2017;55(2):182-96.
173. MacBeath G. Protein microarrays and proteomics. Nature Genetics. 2002;32(4):526-32.
174. Karpievitch YV, Polpitiya AD, Anderson GA, Smith RD, Dabney AR. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. Ann Appl Stat. 2010;4(4):1797-823.

175. Yang F, Shen Y, Camp DG, Smith RD. High pH reversed-phase chromatography with fraction concatenation as an alternative to strong-cation exchange chromatography for two-dimensional proteomic analysis. Expert Review of Proteomics. 2012;9(2):129-34.
176. Chait BT. Mass spectrometry: Bottom-up or top-down? Science. 2006;314(5796):65-6.
177. Toby TK, Fornelli L, Srzentić K, DeHart CJ, Levitsky J, Friedewald J, et al. A comprehensive pipeline for translational top-down proteomics from a single blood draw. Nature Protocols 2018 14:1. 2018;14(1):119-52.
178. Catherman AD, Skinner OS, Kelleher NL. Top Down Proteomics: Facts and Perspectives. Biochemical and biophysical research communications. 2014;445(4):683-93.
179. Han X, Aslanian A, Yates JR. Mass Spectrometry for Proteomics. Current opinion in chemical biology. 2008;12(5):483-90.
180. Cappelletti V, Hauser T, Piazza I, Pepelnjak M, Malinovska L, Fuhrer T, et al. Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. Cell. 2021;184(2):545-59.e22.
181. Pappireddi N, Martin L, Wühr M. A Review on Quantitative Multiplexed Proteomics. Chembiochem : a European journal of chemical biology. 2019;20(10):1210-24.
182. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. Analytical Chemistry. 2003;75(8):1895-904.
183. Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. PROTEOMICS. 2007;7(3):340-50.
184. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Molecular & Cellular Proteomics. 2002;1(5):376-86.

MATERIALS & METHODS

The materials and methods described in the following section have been adapted or partially adapted from the following articles:

Spatial Proteomic Analysis of Isogenic Metastatic Colorectal Cancer Cells Reveals Key Dysregulated Proteins Associated with Lymph Node, Liver, and Lung Metastasis.

Solís-Fernández, G.; Montero-Calle, A.; Martínez-Useros, J.; López-Janeiro, Á.; de los Ríos, V.; Sanz, R.; Dziakova, J.; Milagrosa, E.; Fernández-Aceñero, M.J.; Peláez-García, A.; Casal, J.I.; Hofkens, J.; Rocha, S.; Barderas, R.

Cells 2022, 11, 447. https://doi.org/10.3390/cells11030447

Aryl-hydrocarbon receptor-interacting protein regulates tumorigenic and metastatic properties of colorectal cancer cells driving liver metastasis

Solís-Fernández, G., Montero-Calle, A., Sánchez-Martínez, M. et al.

Br J Cancer (2022). https://doi.org/10.1038/s41416-022-01762-1

Cell lines and cell culture

All the cell lines used in this thesis together with a brief description of the cell line can be found in Table M&M 1.

Table M&M 1: List and brief description of cell lines used.

Cell line	Description	Ref
KM12C	Non-metastatic CRC cells from the KM12 system	(1)
KM12SM	Liver metastasis CRC cells from the KM12 system, isogenic with KM12C	(1)
KM12L4a	Lung and liver metastasis CRC cells from the KM12 system, isogenic with KM12C and KM12SM	(1)
SW480	Established from a primary colon adenocarcinoma	(2)
SW620	Established from a metastatic lymph node, derived from the same tumour from which the SW480 cells were derived.	(2)
HT-29	Human colorectal adenocarcinoma cell line with epithelial morphology	(3)
SW48	Isolated from the large intestine of an 82-year-old, Caucasian, female grade IV Dukes C colorectal cancer patient	(4)
CaCo-2	Epithelial cells isolated from colon tissue derived from a 72-year-old, caucasian, male with colorectal adenocarcinoma	(5)
Lim-1215	Human colorectal carcinoma cell line, derived from an omentum biopsy, with the primary lesion found in the ascending colon	(6)
Colo 320	Established in 1977 from the tumour mass of a 55-year-old woman with a moderately undifferentiated adenocarcinoma of the sigmoid colon	(7)
RKO	Poorly differentiated colon carcinoma cell line	(8)

Adipose Derived Stem Cells	Stem cells derived from adipose tissue and isolated from human lipoaspirate	(9)
Cancer Associated Fibroblasts	Human primary derived fibroblasts immortalized with hTert, labelled with cytoplasmic GFP expression.	(10)

All cells, except CaCo-2, Colo320 and ADSCs, were cultured at 37 °C in a humidified environment at 5% CO_2 in Dulbecco's modified Eagle medium (DMEM; Gibco) containing 10% fetal bovine serum (FBS; Sigma), L-glutamine (2 mM final concentration; Gibco) and either penicillin/streptomycin (100 U/mL final concentration; Gibco) or gentamicin (10 μg/mL final concentration; Gibco). For CaCO-2 minimum essential medium (MEM; Gibco) was used while RPMI medium (GIBCO) was used for Colo 320. For adipose derived stem cells (ADSCs), 5.05 g of alpha-MEM (GIBCO) and 1.1 g of sodium bicarbonate were diluted in distilled, filtered water and FBS was added (10% final concentration) and cells were cultured in this medium.

KM12C, KM12SM and KM12L4a were obtained from the laboratory of Dr. Fidler's lab (MD Anderson Cancer Center, Houston, TX, USA). The remaining CRC cell lines were purchased from the American Type Culture Collection (ATCC). ADSCs were obtained from the Radboud biobank. All cells were cultured for a maximum of 10 passages except for ADSCs that were cultured for less than 7 passages. Cancer associated fibroblasts (CAFs) labelled with GFP were a kind gift from Prof. Olivier de Wever from UGhent.

Protein extraction and quantification

For the protein extracts of 2D cultured cells, cells were grown until 90% confluence and then passaged using PBS-EDTA 4 mM. PBS-EDTA was used to prevent trypsin from digesting any of the proteins being expressed in the cell surface or cell junctions. After passage, cells were harvested

by centrifugation at 1000 *g* for 5 minutes. The supernatant of the cell pellets was removed, and the cell pellets were kept at -20°C until they were used for protein extraction. Protein extracts were obtained by lysing cells with commercial RIPA buffer (Sigma Aldrich) with protease and phosphatase inhibitors (commercial house) at 1× final concentration. After resuspending the cell pellets in RIPA buffer, 25G syringes (BD biosciences) were used to make sure that cells were completely lysed. Protein concentration from each of the protein extracts was determined by tryptophan quantification method (11). Results were confirmed using 10% SDS-PAGE and posterior Coomassie blue staining.

For the spatial analysis of CRC cell lines, a subcellular protein fractionation kit (Subcellular Protein Fractionation Kit for Cultured Cells, Thermofisher Scientific) was used to isolate the proteins from all the different compartments. Following the manufacturer's intructions, cell pellets were resuspended in the appropriate buffer and subsequently lysed. The different subcellular fractions (CEB, cytoplasmic proteins; MEB, membrane proteins; NEB: nuclear soluble proteins; NEB-CBP, chromatin-bound proteins; PEB, cytoskeletal proteins) were isolated by sequential centrifugation and resuspension of the cell pellet in the corresponding buffers. In parallel, for the secretome protein collection, cells were incubated overnight with serum-free DMEM. After overnight incubation the media was collected and proteins in suspension were methanol-chloroform precipitated and re-suspended in RIPA buffer. All the 6 subcellular fractions were quantified the same way as all the other protein extracts using the Trp method (11).

For 3D studies, cells were harvested from the different matrices by dissolving the polymers used, either Matrigel or polyisocianopeptide (PIC), in cold PBS. The cells harvested in cold PBS were washed three times with cold PBS to remove all the remnants of polymer and the cell pellets stored at -20°C until protein was extracted. Protein extraction and

quantification from the cell pellet was performed in the same way as for 2D samples. Likewise, spheroids' protein extract was obtained by from the pelleted spheroids using RIPA buffer for cell lysis in the exact same manner as the rest of the samples.

Isobaric labelling for mass spectrometry analysis

All protein samples analysed via mass spectrometry were labelled using TMT techniques (12). Proteomic analyses of AIP overexpression, 3D differentiation of cancer cells and ADSCs differentiation in 3D were performed using TMT 10-plex labelling kits (ThermoFisher Scientific) while TMT 11-plex labelling kits (ThermoFisher Scientific) were used for the CRC spatial analysis. Regardless of the protein extract used, the labelling procedure as such was the same for all conditions see (13). In brief, 20 μg of each protein extract were diluted to reach a final concentration of 0.2 mg/mL in RIPA buffer (to a final volume of 100 μL) and reduced by adding 10 μL of 100 mM tris(2-carboxyethyl)phosphine (TCEP, Sigma Aldrich). Subsequently, samples were incubated for 45 minutes at 37°C and 600 rpm. An alkylation step was performed by adding 12.2 μL of chloroacetamide 0.4 M (Sigma, final concentration 40 mM) and incubating for 30 minutes at room temperature and 600 rpm. To improve protein digestion and allow for better peptide recovery from the digestion, we used Sera-Mag magnetic beads mix (50% hydrophilic/50% hydrophobic, GE Healthcare). Samples were incubated with 100 μL of the mix of hydrophobic and hydrophilic beads and 230 μL of 100% acetonitrile (PanReac Applichem) for 35 minutes at room temperature and 600 rpm. Then, supernatants were discarded, and magnetic beads were washed twice with ethanol 70% (PanReac AppliChem) and acetonitrile 100%. Proteins bound to magnetic beads were digested for 14 hours at 600 rpm and 37 °C with 1 μg of porcine trypsin (ThermoFisher Scientific) per 20 μg of protein extract, in 100 μL of HEPES buffer (Sigma) 20 mM pH 8.0. To retrieve the maximum number

of peptides from the beads, samples were sonicated for 2 minutes, and the supernatant was collected. Beads were resuspended in 100 μL of HEPES buffer and sonicated again for 2 minutes, and the supernatant was collected and mixed with the previous supernatant. Finally, 20 μL of TMT markers were added to the corresponding tube in two subsequent 30 minutes incubations at room temperature. The content of the 10 or 11 tubes (depending on whether a 10- or 11-plex was used) was pooled together and dried under vacuum. Samples were then reconstituted in 300 μL of TFA (Sigma Aldrich) 0.1% in water and separated based on hydrophobicity using a High pH Reversed-Phase Peptide Fractionation Kit (Pierce Biotechnology). A total of 12 fractions were separated and subsequently dried under vacuum and stored at -80 °C until they were LC—MS/MS analysed in a Q Exactive mass spectrometer (ThermoFisher Scientific).

LC–MS/MS Analysis

LC–MS/MS analyses were made according to established protocols (14, 15). In brief, an Easy-nLC 1000 nano system (ThermoFisher Scientific) was used for peptide separations. For each analysis, samples were loaded into a precolumn Acclaim PepMap 100 (ThermoFisher Scientific) and eluted in a RSLC PepMap C18, 50 cm long, 75 μm inner diameter and 2 μm particle size (Thermo Fisher Scientific). The mobile phase flow rate was 300 nL/min using 0.1% formic acid in water (solvent A, Fisher Chemical) and 0.1% formic acid in acetonitrile (solvent B, Fisher Chemical). The following gradient profile was used: 5 minutes of 3-7% solvent B, 95 minutes of 7 to 25% solvent B, 14 minutes of 25 to 60% solvent B, 1 minute of 60 to 95% solvent B, and, finally, 5 minutes of 95% solvent B. From each sample 4 μL were injected into the Q Exactive mass spectrometer.

Ionization was performed using 1900 V of liquid junction voltage and 270 °C capillary temperature. The full scan method employed a

m/z 300-1800 mass selection, an Orbitrap resolution of 70,000 (at m/z 200), a target automatic gain control value of 3×10^6, and maximum injection times of 100 ms. After the survey scan, the 15 most intense precursor ions were selected for MS/MS fragmentation. Fragmentation was performed with a normalized collision energy of 27 and MS/MS scans were acquired with a dynamic first mass, the AGC target was 1×10^5, with a resolution of 35,000, an intensity threshold of 2×10^4, an isolation window of 1.6 m/z units, and the maximum ion injection time was 100 ms. Charge state screening was enabled to reject unassigned, singly charged, and greater than or equal to seven protonated ions. A dynamic exclusion time of 30 s was used to discriminate against previously selected ions.

MS data analysis

MS raw data were analysed with MaxQuant (Version 1.6.6.0) (16) against UniProt UP000005640_9606.fasta Homo sapiens (human) 2019 database (20,962 protein entries, downloaded: May 2019). For the analysis, Reporter ion MS2 type was used with Trypsin/P as cleavage enzyme allowing for 2 missed tryptic cleavages and a mass tolerance of 20 ppm (Orbitrap). Mass tolerance of precursor and reporter were set to 4.5 ppm and 0.003 Da, respectively. Ser, Thr and Tyr phosphorylation together N-terminal acetylation and methionine oxidation were set as variable modifications, while carbamidomethylation of cysteines was set as the only fixed modification. Overlapping isotopic contributions were taken into account by correcting reporter ion intensities for any bias, following the manufacturer's certificate. Unique and Razor peptides were considered for quantification. Minimal peptide length and maximal peptide mass were fixed to 7 amino acids and 4600 Da, respectively. Precursor intensity fraction with an FDR threshold of 0.01 was used for filtering of identified peptides. Proteins identified as potential contaminants were excluded from the analysis and only those proteins identified with at least one peptide and an ion score above 99% were evaluated. The protein

sequence coverage was estimated as the number of matching amino acids in a specific protein sequence that were found in the peptides sequences having confidence ≥95% divided by the total number of amino acids in the sequence.

All channels from the TMT reagent were computationally normalized using the sum of each channel signals to equal them and correct for the differences resulting from the amount of protein labelled by each TMT reagent. R studio was used to conduct sample loading normalization following established protocols (https://github.com/pwilmart, accessed on 15th may 2020), using “tidyverse”, “psych”, “gridExtra” and “scales” packages (17, 18). Analysis of the normalized data was performed either in Perseus (Version 1.6.6) (19) or Microsoft Excel 2019 and Microsoft Access 2019. In the case of TMT analysis that included replicate samples Perseus was used together with Microsoft Excel, while Microsoft access together with Microsoft Excel was used for the analysis of the rest of TMT datasets.

For the spatial analysis of CRC proteins identified in at least 3 of the 10 samples with one or more peptides and a fold change ≥ 2 for each ratio were selected as potential proteins dysregulated in metastatic cells (upregulated = ratio ≥ 2, or downregulated = ratio ≤ 0.5). For AIP overexpression proteomic analysis and 3D differentiation of cancer cells, the fold change ratio was set at ≥ 1.5 (upregulated = ratio ≥ 1.5, or downregulated = ratio ≤ 0.67). p-value representing the probability that the observed ratio was different than 1 by chance was also calculated. p-value < 0.05 was considered statistically significant. Finally, 3D differentiation of adipose derived stem cells and bicyclic peptide differentiation proteomic characterizations were analysed using Perseus. In this case, the criteria followed for considering up- or downregulated proteins was set at a fold change of 2.

Tissue and plasma samples for Spatial characterization of CRC

This study related to biomarker discovery and validation was approved by the Institutional Ethical Review Boards of the Instituto de Salud Carlos III and Hospital Clínico San Carlos (Madrid) (CEI PI 13_2020-v2). Tissue and plasma samples were obtained from the IdISSC biobank of the Hospital Clínico San Carlos, which belong to the National Biobank Net (ISCIII) cofounded with FEDER funds. Written informed consent was obtained from all patients. Tissue and plasma samples were obtained according to a standardized protocol for sample collection (20-22). Tissue and plasma were stored at –80 °C until use (20-22). Paired OCT -embedded frozen healthy and CRC tissue from 14 CRC patients at stages I to IV was used for western blot analysis (Tables S1 and S2).

Plasma samples from healthy individuals with negative colonoscopy (n = 32), plasma samples from CRC patients at stages I-IV (n = 38), and plasma samples from individuals with premalignant lesions (colorectal adenomas; n = 10) were analysed by ELISA (Table M&M 2).

Table M&M 2: Information of the paired tumoral and non-tumoral tissue samples from CRC patients used for the WB and TMA analysis, and plasma samples from CRC patients used for the ELISA analysis.

	Application	Samples (n)		Age Average ± SD (years)	Age Range (years)	Gender (n)		Stage				
						Male	Female	Premalignant Individuals	I	II	III	IV
Tissue samples	WB analysis	CRC patients	14	72.1 ± 12.3	47–88	5	9	-	2	4	4	4
	IHC and TMA analysis	CRC patients	95	73.0 ± 9.6	94–51	57	38	-	3	30	61	1
Plasma samples	ELISA	CRC patients	48	73.2 ± 10.3	49–88	21	27	10	10	10	11	7
		Healthy individuals	32	57.9 ± 10.4	24–74	16	16	-				

Western blot

For western blot (WB) analysis, equal amounts of protein extracts (μg) from each condition being evaluated were separated on a 10% SDS-PAGE under reducing conditions (23). Samples transfer onto a nitrocellulose membrane (Hybond-C Extra, GE healthcare) was performed at 100 V for 90 minutes using Mini Trans-Blot Module equipment (Bio-rad). After transference, membranes were blocked in blocking buffer for 1 hour at room temperature and subsequently incubated overnight (O/N) at 4 °C with primary antibody diluted in blocking buffer. The dilution of each of the primary antibodies used, together with the blocking buffer used for each antibody and the purpose for which each primary antibody was used can be found in Table S3. After O/N incubation, membranes were washed for 10 minutes three times with 0.1% Tween 20-PBS 1× (PBS-T) to remove unbound primary antibody. Membranes were then incubated for 1 hour at RT with the appropriate secondary antibody (goat anti rabbit, GAR, or goat anti mouse, GAM) labelled with horseradish peroxidase (HRP) in blocking buffer. The washing step was repeated after secondary antibody incubation and signal was developed with SuperSignal West Pico Maximum Sensitivity Substrate (Pierce) or SuperSignal West Femto Maximum Sensitivity Substrate (Pierce) and detected on an Amersham 680 Imager (GE Healthcare).

Immunofluorescence staining

Cells were fixed with 4% paraformaldehyde (Sigma Aldrich) for 20 min at 37 °C and permeabilized with PBS 1×-0.1% Triton X-100 (Sigma Aldrich) for 10 min at room temperature and blocked with PBS-T containing 10% FBS for 1 h at room temperature. Cells were subsequently incubated O/N at 4 °C with primary antibodies and then washed with PBS-T 3 times. Then, immunofluorescence slides were incubated for 1 hour with fluorescently labeled secondary antibody (Table S3). For samples to be

analysed for F-actin, after secondary antibody incubation, slides were incubated at room temperature for 45 minutes with phalloidin labelled with Atto 520 dye (ThermoFisher scientific) diluted 1:400 in blocking buffer. Finally, the cell nucleus was counterstained with DAPI (Agilent) at 10 μg/mL at room temperature for 10 minutes.

For staining of samples encapsulated in either Matrigel of PIC polymer, cells encapsulated were washed with PBS and fixed with 4% PFA for 40 minutes instead of 20 minutes. After fixation, cells were permeabilized in the same way as normal microscopy slide samples. Finally, samples were blocked for 30 minutes at room temperature with BSA 1% in PBS and then stained for 1 hour with Phalloidin Atto-520 diluted 1:20 in blocking buffer (final concentration 10 μM).

***In silico* analysis**

For the *in silico* analysis of CRC spatial characterization, gene set enrichment analysis (GSEA) was performed using fsgea (Version 1.16.0) and msigdbr (Version 7.2.1) R packages (24). In brief, differential protein expression ranked lists obtained for each of the cellular compartments were subjected to pathway enrichment analysis. Pathways belonged to the gene ontology, human reactome and KEGG collections. GSEA was performed using pathways that contained at least one protein from the cross-validated differentially dysregulated proteins. From the 1831 pathways finally used in the present analysis, 1555 belonged to the gene ontology collection, 234 to the human reactome and 42 to the KEGG collection. A Benjamini–Hochberg p-value below 0.05 was required to consider the pathway significantly enriched. String db (version 11.5) for functional protein association networks was used to identify clusters of interaction among the proteins of the dataset.

For AIP prognostic analysis, the GSE17538 database was used. This set contains 244 tumor samples with clinicopathological description from colorectal cancer patients. Data were normalized using Bioconductor's

Affymetrix package and transformed into z-scores. The prognostic value of AIP was assessed by Kaplan-Meier survival curves using the best cut-off method for separating high and low expression populations. The significance of the difference in survival between both populations was estimated by log-rank test. To validate the results with a different cohort of patients, GEPIA2 tool (http://gepia2.cancer-pku.cn/) was employed with colon adenocarcinoma TCGA dataset (270 tumor samples) (25).

Immunohistochemistry, tissue microarrays and ELISA

For immunohistochemistry (IHC) analysis of CRC spatial characterization, tissue microarrays (TMAs) containing 95 core samples from CRC samples (Table S4) and that had been previously constructed were used (26, 27). For staining, 2 μm thick sections were deparaffinized following previously established protocols (26-28) in low pH buffered solution (EnVision™ FLEX Target Retrieval Solution, Dako). Before O/N incubation at 4 °C with the indicated primary antibodies (Table S3), slides were blocked with peroxidase blocking reagent (Dako) to prevent endogenous peroxidase activity. Slides were then incubated with the appropriate anti-Ig HRP-conjugated polymer (EnVision™ FLEX-HRP, Dako). Visualization and immuno-reactivity were conducted according to established protocols using 3,3′-diaminobenzidine as a chromogen (28, 29).

IHC performed for the AIP overexpression studies were performed in the same way, albeit in this case core tumor tissue samples from 50 metastatic and non-metastatic CRC patients and core tumor tissue samples from 94 recurrent or non-recurrent CRC stage II patients composed the two tissue microarrays (TMAs) used in the study.

Finally, for the GLG1 plasma level determination, plasma from patients referred in Table M&M 2 and S2 were diluted 1:5 and GLG1 levels were measured using a commercial GLG1 ELISA kit (Sabbiotech), according to the manufacturer's suggestions.

Polyisocyanopeptide based polymer synthesis

An established protocol (30) was used for the synthesis of Polyisocyanides (PIC). Briefly, isocyanide monomers (azide-appended monomer and non-functional monomer in ratio 1:29) were dissolved in toluene and a catalyst solution ($Ni(ClO_4)2{\bullet}6H_2O$ (0.1 mg/mL) in toluene/ethanol 9:1) was added to achieve a final concentration ratio of monomer to Ni^{2+} equal to 5000:1 or 1000:1 depending on the desired stiffness of the polymer. Monomer concentration was adjusted to 50 mg/mL by adding toluene. The polymerization mixture was stirred at room temperature, and the progress of the reaction was followed by IR-ATR (disappearance of the characteristic isocyanide absorption at 2140 cm^{-1}). Once the isocyanide was consumed (48 h), the polymer was precipitated in diisopropyl ether under vigorous stirring and collected by centrifugation. Then, the polymer was precipitated in two sequential rounds after being resuspended in dichloromethane and, finally, air-dried to obtain an off-white solid. Viscometry was used to measure the molecular weight of the polymer.

PIC polymer bioconjugation

For RGD biofunctionalization of the PIC polymer click chemistry was used. The RGD peptide was bound with a DBCO-terminated spacer composed of PEG repeats. Following previously established protocols the DBCO-PEG_4-RGD moiety was reacted with the azide appended polymers. The reaction was conducted in such a way that, on average, 1% of the monomers that comprise the polymer carry a peptide. Similarly, for the biofunctionalization with bicyclic peptides, the polymer carrying the azide-appended group was resuspended in acetonitrile (2.5 mg/mL), and the appropriate amount of peptide was added. Bicyclic peptides (PEPSCAN) were diluted in DMSO at a final concentration of 6 mg/mL. For both RGD and bicyclic peptides, conjugations were conducted O/N under stirring at room temperature. Finally, conjugated polymers were

collected for centrifugation after precipitation with diisopropyl ether and air-dried in the same way as the un-conjugated polymer was.

Cell encapsulation and 3D culture

KM12C, KM12SM and human adipose-derived stem cells (ADSCs) were all encapsulated following the same procedure that has already been described in (31, 32). After UV sterilization of dry PIC polymers, they were then dissolved in the cell culture medium used to culture the cells that would be encapsulated, at a final polymer concentration of 2 mg/mL. Polymers were dissolved for 24 hours under stirring at 4 °C. Once polymers were dissolved, cells were harvested, counted and subsequently mixed 1:1 with ice-chilled polymer to achieve a final polymer concentration of 1 mg/mL at the desired cell density (200,000 cells/mL). The cell-polymer mixture was transferred onto 48-well plates (Corning) or 8-well chambered coverslips (Sigma) depending on whether the cells were cultured for protein extraction/DNA extraction or imaging and heated to 37 °C. Once seeded in the well plates, cells were cultured under standard cell culture conditions (already described in the cell lines and cell culture section). For encapsulation in Matrigel (Corning), used as control for 3D differentiation, the same procedure was followed although in this case the ratio Matrigel:cell was 7:3. Mixture of Matrigel and cells was also performed on ice. For both PIC and Matrigel after the gel had polymerized because of temperature, additional cell culture medium was added on top of the gels.

Plasmid vectors and cell transfection of KM12 cells

KM12C and KM12SM cells were transfected with a vector to induce the overexpression of AIP. AIP gene was obtained from the DNASU plasmid repository and cloned into the pcDNA3.1(+) vector, obtained from ThermoFisher Scientific. Cloning of the AIP gene in the vector was performed using Gibson assembly (33) (NEB), following standard procedures and the manufacturer's instructions. After plasmid cloning,

Escherichia coli DH5α strain bacterial cells were transformed with the mixture of the Gibson assembly reaction and selected using ampicillin (100 μg/mL) containing agar plates. Individual resistant colonies were harvested and expanded in liquid LB culture and plasmid inside bacteria was purified using a commercial (Neobiotech) plasmid miniprep kit. Once the vector sequence was verified, cells were transfected according to previously described protocols (34-37). Briefly, on a p6 well plate, 250,000 cells were seeded and one day post-seeding cells were transfected using JetPrime transfection reagent (Polyplus) following the manufacturer's instructions. Either 2 μg of pcDNA3.1(+) empty vector or the AIP containing vector were used for cell transfection. Then, 48 hours post-transfection cell media was changed and replaced with DMEM containing 1 mg/mL G418 (Santa Cruz BioTechnologies) for cell selection (3-4 weeks)(38). A concentration of 0.6 mg/mL of G418 was used for routine cell culture after selection.

RNA extraction, cDNA synthesis, semi-quantitative PCR, real time quantitative PCR (qPCR)

RNA was extracted from colon cancer cell lines with the RNeasy Mini Kit (Qiagen Inc.) according to the manufacturer's protocol. The extracted RNA was quantified with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc.).

cDNA was synthesized using the NZY First-Strand cDNA Synthesis Kit (NZYtech). Then, cDNA was directly used for semi-quantitative PCR analysis of AIP, EMT markers and GAPDH (as control) mRNA levels using specific primers (Table S5). Alternatively, cDNA was used for qPCR analyses using specific primers of ZO1 (Tjp1 gene), E-Cadherin (CDH1), ZEB1, TGFβ1, Claudin-2, and Snail1 (SNAI1 gene) (Table S5), using 18S as control and for normalization. Semi-quantitative PCR reactions were performed using the Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific). For qPCR, reactions were performed using TB-

Green Premix Ex Taq (Tli RNaseH plus, Takara), and PCR and data collection were performed on a Light Cycler 480 (Roche).

Cell adhesion, invasion, apoptosis detection, proliferation, soft agar colony formation, and wound healing assays

Functional cell-based assays were performed at least in triplicate according to established protocols (39-42). Cell adhesion assays were conducted using 96 well plates that were coated with 0.4 μg/mm^2 of Matrigel in coating buffer (0.1 M $NaHCO_3$ pH 8.8). Coating was done overnight at 4 °C and subsequently unspecific binding was blocked by incubating with adhesion medium (0.5% BSA in serum-free DMEM) at 37 °C for 2 hours. Before seeding, cells were starved for 5 hours without serum and incubated with BCECF-AM (Molecular probes) at 37 °C for 30 minutes to label them. Cells were detached with PBS containing 4 mM EDTA and resuspended in adhesion medium. In five separate wells of the p96 well plate 10^5 cells were seeded and incubated for 30 minutes. Then, to remove non-adherent cells, plates were washed with PBS and bound cells were lysed with 1% SDS in PBS. Fluorescence signal of the lysates was quantified in a Spark multimode microplate reader (Tecan Trading AG).

Matrigel invasion assays were performed by resuspending 8×10^5 cells in invasion medium (0.5% BSA in serum-free DMEM) and transferred onto 8 μm pore-size filters transwells (Costar) that had been previously coated with 35 to 50 μL of Matrigel diluted 1:3 in invasion medium. The bottom compartment of the invasion chambers was filled with standard FBS supplemented cell culture media. After 22 h of incubation at 37 ºC, non-invading cells were removed from the filter's upper surface, and cells that migrated through the filter were fixed with 4% PFA, stained with crystal violet and the invading cells counted under a microscope.

For wound healing, cells were seeded inside IBIDI silicone wound healing inserts (IBIDI, #80209) in 24-well plates at a density of 1.5×10^5 cells per

insert well. The day after, the insert was removed to uncover a cell-free area of 500±100 μm. Fresh cell growth medium was added, and time-lapse imaging of the cells was recorded. For imaging, a confocal microscope (TCS-SP5-AOBS-UV, Leica-Microsystems) imaging position was set using the Mark and Find Leica imaging software tool, and 1500×908 μm (2048×1200 pixels) images were acquired every 90 minutes. Temperature (37 °C) and CO_2 concentration (5%) were kept constant throughout the whole imaging time (48 hours). The cell-free area was measured with the MRI's Wound Healing tool for ImageJ software (NIH)(https://github.com/MontpellierRessourcesImagerie/imagej_macros_and_scripts/wiki/Wound-Healing-Tool) (43), and the calculated areas were visually inspected to verify all the time points were correct.

For cell proliferation assays, the growth medium was changed 24 h after seeding (day 0), and cells were further incubated for three days. Then, medium was removed, and cells were stained with 100 μl of the chromogenic dye 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, Sigma) at a final concentration of 1 mg/mL in DMEM. The cells were further incubated for 1 h at 37 °C and 5% CO_2. Then, medium was carefully aspirated, and cells disrupted with 100 μL of DMSO (Sigma) prior to reading absorbance at 570 nm. All the experiments were done three times in duplicate.

Colony soft agar assay was performed as previously described (42). In brief, 2× concentrated cell culture media were mixed 1:1 with a molten solution of 1% noble agar (Condalab). A total of 1.5 mL per well of the 42 °C mixture of media and agar was poured onto wells of a p6 cell culture treated well plate (Corning) and this bottom layer of agar was allowed to solidify. Subsequently, the top layer of agar containing the cells was prepared. Two different cell densities, 5000 and 25000 cells per well, were used. For both cell densities, a 1:1 mixture of cells diluted in media and 0.6% noble agar diluted in agar was prepared at 42 °C. Finally, 1.5 mL of

the cell-agar mixture were poured on top of the bottom solidified layer of agar. Cells were allowed to grow inside the agar gel for 21 days, refreshing the media inside the wells every 2-3 days by adding 100 μL of culture media per well. Then, colonies were visualised by microscopy by acquiring at least 16 random fields of view.

Confocal microscopy and live cell imaging

All fluorescence confocal microscopy images were acquired on a Leica TCS-SP8-AOBS-UV system (Leica Microsystems). All images were acquired using a 63× water immersion N.A. 1.2 HC PL APO motCORR objective (Leica). A supercontinuum White Light Laser (470–670 nm, pulsed, 80 MHz, NKT Photonics) was used for fluorescence excitation except for DAPI excitation, which was achieved using a UV diode laser (405 nm, pulsed, 40 MHz, PicoQuant). Regardless of the settings used in every particular experiment, they were all kept constant throughout all the samples of the same experiment. For example, all fluorescence images acquired for the spatial analysis of CRC metastasis were acquired in the same conditions of pixel size, z-stack size, excitation laser power and detector sensitivity.

Live cell imaging was performed using a Cytosmart Lux2 inverted bright-field microscope (Cytosmart). The microscope was placed in an incubator (37 °C, humidified) to monitor the morphology of the cells in hydrogels. Cold cell-gel mixtures were prepared in the same way as the described in the cell encapsulation section and 100 μL of the mixture were pipetted onto a 35 mm glass-bottom dish (Cellvis, #1.5) and allowed to gelify at 37 °C for 15 min. Afterward, 3 mL of CO_2-independent α-MEM culture medium (ThermoFisher Scientific) was pipetted on top of the samples. 24-hour monitoring was performed to acquire real-time videos with a 10× air objective acquiring images every 10 minutes.

In vivo animal experiments

The Ethical Committee of the Instituto de Salud Carlos III (Spain) approved the protocols used for experimental work with mouse after approval for the ethical committee OEBA (Proex 285/19). Liver metastasis, *in vivo* homing and subcutaneous experiments were performed in Nude mice according to established protocols (34).

For metastasis experiments, 10^6 AIP- and Mock-stably transfected KM12 cells in 0.1 mL PBS were intrasplenically injected in Swiss nude mice (Charles River; n=6 per group). The spleen was resected the day after, and then mice were daily inspected for signs of disease, such as abdominal distension, locomotive deficit, or tumour presence by palpation. When symptoms were noticeable, mice were euthanized and inspected for metastasis in the liver. For xenografts (n=6 per group), tumours were induced in nude mice by subcutaneous injection of 1×10^6 AIP- and Mock-stably transfected KM12 cells in 0.1 mL PBS. Tumours were measured with an external calliper, and volume was calculated as $(\text{width})^2 \times (\text{length})$. When tumours reached an average size of 1500mm^3, animals were euthanized, and tumours excised for further evaluation.

For homing analysis to liver (n=2 per group), Swiss nude mice were intrasplenically inoculated with 10^6 AIP- and Mock-stably transfected KM12 cells in 0.1 mL PBS. Mice were then euthanized 24 h after inoculation, and RNA was isolated using TriZol Reagent from liver and spleen. RNA was analysed by RT-PCR as described above to amplify human GAPDH and, as loading control, murine β-actin using specific primers (Table S5).

Six-eight weeks Swiss nude mice were used for in vivo experiments. The same proportion of male and females per group were used. To avoid any mistake during identification of the different animals, mice injected with different cells were in different cases; and thus, no randomization and no blinding were used in the study.

Statistical analysis

Statistical analyses were performed with Microsoft Office Excel and Graphpad Prism 8. Data were analysed by one-way analysis of variance followed by Tukey-Kramer multiple comparison test. In both analyses, the minimum acceptable level of significance was p-values < 0.05. For discrete variable data, as presence or absence of metastasis, we used χ^2 test. For continuous variable data with not Gaussian distribution, as overall survival, we performed a U-Mann Whitney test.

For immunohistochemistry analysis, data distribution using the Shapiro-Wilk test and variance homogeneity using the Bartlett test was first evaluated. Since data normality was discarded in all cases, we then assessed whether each indicated group's means were statistically different from each other using the non-parametric U-Mann Whitney test assuming unequal variances. p-values < 0.05 were considered statistically significant. Microsoft Office Excel, GraphPad Prism, and the R program were used for all statistical analyses.

For ELISA, variance homogeneity was evaluated with the Bartlett test. As non-homogeneous variances were observed in the dataset for all groups, the non-parametric Mann–Whitney U test was used to determine whether the mean of the control individuals, the mean of the premalignant individuals, and/or the mean of the CRC groups (grouped or analyzing each CRC stage separately) was statistically different. Mean ± SEM were represented. p-values < 0.05 were considered statistically significant.

ROC curves were used to evaluate GLG1 concentration as marker of CRC patients, premalignant individuals, or control individuals. Alternatively, ROC curves were used to evaluate IHC data. ROC curves were constructed with the R program (Version 3.2.3) using the R package Epi (44). AUC and maximized sensitivity and specificity were calculated.

References

1. Camps J, Morales C, Prat E, Ribas M, Capellá G, Egozcue J, et al. Genetic evolution in colon cancer KM12 cells and metastatic derivates. International Journal of Cancer. 2004;110(6):869-74.
2. Fogh J, Fogh JM, Orfeo T. One Hundred and Twenty-Seven Cultured Human Tumor Cell Lines Producing Tumors in Nude Mice. JNCI: Journal of the National Cancer Institute. 1977;59(1):221-6.
3. Cohen E, Ophir I, Shaul YB. Induced differentiation in HT29, a human colon adenocarcinoma cell line. Journal of cell science. 1999;112 (Pt 16)(16):2657-66.
4. Leibovitz A, Stinson JC, McCombs Iii WB, McCoy CE, Mazur KC, Mabry ND. Classification of Human Colorectal Adenocarcinoma Cell Lines. Cancer Research. 1976;36:4562-9.
5. Jumarie C, Malo C. Caco-2 cells cultured in serum-free medium as a model for the study of enterocytic differentiation in vitro. Journal of cellular physiology. 1991;149(1):24-33.
6. Whitehead RH, Macrae FA, St. John DJB, Ma J. A Colon Cancer Cell Line (LIM1215) Derived From a Patient With Inherited Nonpolyposis Colorectal Cancer. JNCI: Journal of the National Cancer Institute. 1985;74(4):759-65.
7. Quinn LA, Moore GE, Morgan RT, Woods LK. Cell Lines from Human Colon Carcinoma with Unusual Cell Products, Double Minutes, and Homogeneously Staining Regions. Cancer Research. 1979;39:4914-24.
8. Liu Z, Zou D, Yang X, Xue X, Zuo L, Zhou Q, et al. Melatonin inhibits colon cancer RKO cell migration by downregulating Rho-associated protein kinase expression via the p38/MAPK signaling pathway. Molecular Medicine Reports. 2017;16(6):9383-92.
9. Tsuji W, Rubin JP, Marra KG. Adipose-derived stem cells: Implications in tissue regeneration. World Journal of Stem Cells. 2014;6(3):312-21.
10. Stadler S, Nguyen CH, Schachner H, Milovanovic D, Holzner S, Brenner S, et al. Colon cancer cell-derived 12(S)-HETE induces the retraction of cancer-associated fibroblast via MLC2, RHO/ROCK and Ca2+ signalling. Cellular and Molecular Life Sciences. 2017;74(10):1907-21.
11. Wišniewski JR, Gaugaz FZ. Fast and Sensitive Total Protein and Peptide Assays for Proteomic Analysis. Analytical Chemistry. 2015;87(8):4110-6.

12. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. Analytical Chemistry. 2003;75(8):1895-904.
13. Solís-Fernández G, Montero-Calle A, Martínez-Useros J, López-Janeiro Á, de los Ríos V, Sanz R, et al. Spatial Proteomic Analysis of Isogenic Metastatic Colorectal Cancer Cells Reveals Key Dysregulated Proteins Associated with Lymph Node, Liver, and Lung Metastasis. Cells. 2022;11(3).
14. Torres S, Bartolomé RA, Mendes M, Barderas R, Fernandez-Aceñero MJ, Peláez-García A, et al. Proteome Profiling of Cancer-Associated Fibroblasts Identifies Novel Proinflammatory Signatures and Prognostic Markers for Colorectal Cancer. Clinical Cancer Research. 2013;19(21):6006-19.
15. Mendes M, Peláez-García A, López-Lucendo M, Bartolomé RA, Calviño E, Barderas R, et al. Mapping the Spatial Proteome of Metastatic Cells in Colorectal Cancer. Proteomics. 2017;17(19):1700094-.
16. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology 2008 26:12. 2008;26(12):1367-72.
17. Plubell DL, Wilmarth PA, Zhao Y, Fenton AM, Minnier J, Reddy AP, et al. Extended Multiplexing of Tandem Mass Tags (TMT) Labeling Reveals Age and High Fat Diet Specific Proteome Changes in Mouse Epididymal Adipose Tissue. Mol Cell Proteomics. 2017;16(5):873-90.
18. Lopez-Janeiro A, Ruz-Caracuel I, Ramon-Patino JL, De Los Rios V, Villalba Esparza M, Berjon A, et al. Proteomic Analysis of Low-Grade, Early-Stage Endometrial Carcinoma Reveals New Dysregulated Pathways Associated with Cell Death and Cell Signaling. Cancers (Basel). 2021;13(4).
19. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of proteomics data. Nature Methods 2016 13:9. 2016;13(9):731-40.
20. Barderas R, Babel I, Díaz-Uriarte R, Moreno VV, Suárez A, Bonilla F, et al. An optimized predictor panel for colorectal cancer diagnosis based on the combination of tumor-associated antigens obtained from protein and phage microarrays. Journal of proteomics. 2012;75(15):4647-55.
21. Babel I, Barderas R, Díaz-Uriarte R, Martínez-Torrecuadrada JL, Sánchez-Carbayo M, Casal JI, et al. Identification of tumor-associated

autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays. Molecular and Cellular Proteomics. 2009;8(10):2382-95.
22. Barderas R, Villar-Vazquez R, Fernandez-Acenero MJ, Babel I, Pelaez-Garcia A, Torres S, et al. Sporadic colon cancer murine models demonstrate the value of autoantibody detection for preclinical cancer diagnosis. Scientific Reports. 2013;3(1):1-7.
23. Gallagher SR. One-dimensional SDS gel electrophoresis of proteins. Current protocols in immunology. 2006;Chapter 8(1).
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.
25. Hegde P, Qi R, Gaspard R, Abernathy K, Dharap S, Earle-Hughes J, et al. Identification of Tumor Markers in Models of Human Colorectal Cancer Using a 19,200-Element Complementary DNA Microarray1. Cancer Research. 2001;61(21):7792-7.
26. Lee PY, Chin S-F, Low TY, Jamal R. Probing the colorectal cancer proteome for biomarkers: Current status and perspectives. Journal of Proteomics. 2018;187:93-105.
27. Montero-Calle A, Aranguren-Abeigon I, Garranzo-Asensio M, Poves C, Fernández-Aceñero MJ, Martínez-Useros J, et al. Multiplexed Biosensing Diagnostic Platforms Detecting Autoantibodies to Tumor-Associated Antigens from Exosomes Released by CRC Cells and Tissue Samples Showed High Diagnostic Ability for Colorectal Cancer. Engineering. 2021;7(10):1393-412.
28. Garranzo-Asensio M, Segundo-Acosta PS, Martínez-Useros J, Montero-Calle A, Fernández-Aceñero MJ, Häggmark-Månberg A, et al. Identification of prefrontal cortex protein alterations in Alzheimer's disease. Oncotarget. 2018;9(13):10847-67.
29. Garranzo-Asensio M, San Segundo-Acosta P, Povés C, Fernández-Aceñero MJ, Martínez-Useros J, Montero-Calle A, et al. Identification of tumor-associated antigens with diagnostic ability of colorectal cancer by in-depth immunomic and seroproteomic analysis. Journal of Proteomics. 2020;214:103635-.
30. Liu K, Mihaila SM, Rowan A, Oosterwijk E, Kouwer PHJ. Synthetic Extracellular Matrices with Nonlinear Elasticity Regulate Cellular Organization. Biomacromolecules. 2019;20(2):826-34.

31. Grolma JM, Weinand P, Moone DJ. Extracellular matrix plasticity as a driver of cell spreading. Proceedings of the National Academy of Sciences of the United States of America. 2020;117(42):25999-6007.
32. Liu K, Vandaele J, Bernhagen D, Erp Mv, Oosterwijk E, Timmerman P, et al. Rapid stem cell spreading induced by high affinity α5β1 integrin-selective bicyclic RGD peptide in biomimetic hydrogels. bioRxiv. 2022:2022.02.01.478177-2022.02.01.
33. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 2009 6:5. 2009;6(5):343-5.
34. Bartolomé RA, Barderas R, Torres S, Fernandez-Aceñero MJ, Mendes M, García-Foncillas J, et al. Cadherin-17 interacts with α2β1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis. Oncogene. 2013;33(13):1658-69.
35. Peláez-García A, Barderas R, Torres S, Hernández-Varas P, Teixidó J, Bonilla F, et al. FGFR4 Role in Epithelial-Mesenchymal Transition and Its Therapeutic Value in Colorectal Cancer. PLOS ONE. 2013;8(5):e63695-e.
36. Barderas R, Shochat S, Timmerman P, Hollestelle MJ, Martínez-Torrecuadrada JL, Höppener JWM, et al. Designing antibodies for the inhibition of gastrin activity in tumoral cell lines. International journal of cancer. 2008;122(10):2351-9.
37. Borowicz S, Van Scoyk M, Avasarala S, Karuppusamy Rathinam MK, Tauler J, Bikkavilli RK, et al. The Soft Agar Colony Formation Assay. Journal of Visualized Experiments : JoVE. 2014(92):51998-.
38. Peláez-García A, Barderas R, Batlle R, Viñas-Castells R, Bartolomé RA, Torres S, et al. A Proteomic Analysis Reveals That Snail Regulates the Expression of the Nuclear Orphan Receptor Nuclear Receptor Subfamily 2 Group F Member 6 (Nr2f6) and Interleukin 17 (IL-17) to Inhibit Adipocyte Differentiation. Molecular & Cellular Proteomics. 2015;14(2):303-15.
39. Bartolome RA, Barderas R, Torres S, Fernandez-Acenero MJ, Mendes M, Garcia-Foncillas J, et al. Cadherin-17 interacts with alpha2beta1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis. Oncogene. 2014;33(13):1658-69.
40. Pelaez-Garcia A, Barderas R, Torres S, Hernandez-Varas P, Teixido J, Bonilla F, et al. FGFR4 role in epithelial-mesenchymal transition and its therapeutic value in colorectal cancer. PLoS One. 2013;8(5):e63695.

41. Barderas R, Shochat S, Timmerman P, Hollestelle MJ, Martinez-Torrecuadrada JL, Hoppener JW, et al. Designing antibodies for the inhibition of gastrin activity in tumoral cell lines. Int J Cancer. 2008;122(10):2351-9.
42. Borowicz S, Van Scoyk M, Avasarala S, Karuppusamy Rathinam MK, Tauler J, Bikkavilli RK, et al. The soft agar colony formation assay. J Vis Exp. 2014(92):e51998.
43. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671-5.
44. Garranzo-Asensio M, Guzmán-Aránguez A, Povés C, Fernández-Aceñero MJ, Montero-Calle A, Ceron MÁ, et al. The specific seroreactivity to ΔNp73 isoforms shows higher diagnostic ability in colorectal cancer patients than the canonical p73 protein. Scientific Reports. 2019;9(1):1-13.

OBJECTIVES

The main objective of this doctoral thesis has revolved around the study and characterization of colorectal cancer (CRC) metastasis.

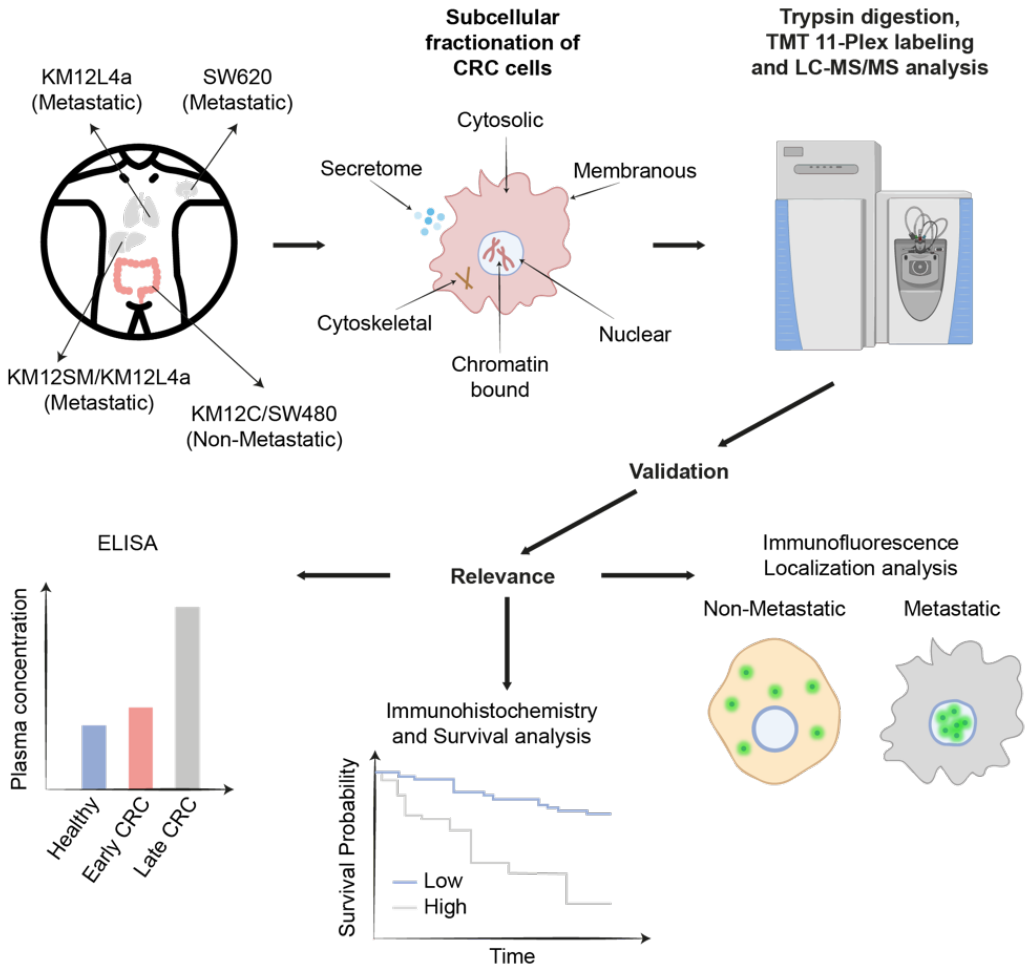
Our first effort was focused on the subcellular spatial proteomic analysis of isogenic cell lines of CRC with different metastatic properties for the discovery of new biomarkers of the disease. Secondly, we focused on expanding the characterisation to isogenic cells cultured in 3D aiming to mimic culture conditions found in living tissues. The initial step of the 3D proteomic analysis involved the study of adipose derived stem cells (ADSCs) as a partially characterized and simpler model. Finally, we focused on the functional analysis of one of the proteins found to be upregulated in liver metastatic KM12SM CRC cells (compared with isogenic non-metastatic KM12C CRC cells) the aryl hydrocarbon receptor-interacting protein (AIP).

The specific objectives from each of the three different sub-objectives are the following:

- Spatial characterization of CRC metastasis.
    - Perform the subcellular spatial proteomics characterization of five isogenic CRC cell lines (KM12C, KM12SM, KM12L4a, SW480 and SW620) with different metastatic properties.
    - Validate the results obtained from the proteomic analysis.
    - Immunofluorescence analysis of selected representative proteins from all of the altered targets to confirm the changes in protein expression and localization.
    - Evaluate the diagnostic and prognostic potential of identified dysregulated proteins (GLG1 in plasma by ELISA and BAIAP2 in tissue by immunohistochemistry).
- Proteomics and PIC hydrogels as tools to understand 3D cell differentiation and its role in cancer progression.

  - Perform the proteomic analysis of ADSCs cultured in PIC-based matrices and 2D to shed light onto proteomic changes previously observed by collaborators.
  - Evaluate the impact of targeting specific integrins ($\alpha_v\beta_3$, $\alpha_5\beta_1$) in the differentiation of ADSCs encapsulated in PIC polymer.
  - Confirm the suitability of our proteomic pipeline for 3D cultured cells
  - Investigate the effect of 3D culture in KM12C and KM12SM cells as a tool to identify new CRC markers.
- AIP as driver of CRC metastasis.
  - Induce the ectopic expression of AIP on KM12C and KM12SM CRC cells and evaluate its effect on the metastatic capacities of both cell lines using *in vitro* functional assays (adhesion, colony formation, invasion, migration and proliferation).
  - Study of the proteins and molecular signalling pathways altered by AIP ectopic expression.
  - Evaluation of the *in vivo* metastatic and tumorigenic potential of cells ectopically expressing AIP.

CHAPTER 1:
SPATIAL CHARACTERIZATION OF CRC METASTASIS



The following chapter is based on the article:

Solís-Fernández G, Montero-Calle A, Martínez-Useros J, López-Janeiro Á, de los Ríos V, Sanz R, et al. Spatial Proteomic Analysis of Isogenic Metastatic Colorectal Cancer Cells Reveals Key Dysregulated Proteins Associated with Lymph Node, Liver, and Lung Metastasis. Cells 2022, 11, 447. https://doi.org/10.3390/cells11030447

Contribution to the article:
I performed the labelling of the protein extracts, analysis of mass spectrometry data, validation via western blot, immunostaining and fluorescence image acquisition, and GLG1 ELISA. Besides the experimental part I wrote and corrected the manuscript.

Introduction

Colorectal cancer, with an estimated overall incidence in the general population of 447 per 100,000, is the second most common cancer in Europe after lung cancer (1, 2). Despite screening programs, CRC is diagnosed at advanced metastatic stages in nearly 20-30% of patients, and relapse occurs in about 40-50% of those patients diagnosed at early cancer stages (3). In the last decade, the use of different schedules of chemotherapies combined with targeted biological therapies has considerably improved the median overall survival for patients with metastatic CRC (1). Nevertheless, the majority of patients with metastatic CRC progress in spite of their initial treatment receiving second and third line treatments, which results in less than 10% 5-year survival probability.

For CRC, besides adjacent lymph node colonization, the most common sites of metastasis are liver (~70%) and lung (~39%); being the other less common metastases sites peritoneum, bone, and nervous system metastases (4). Liver metastases are frequently solitary (~46% of CRC patients), while lung metastases often occur together with liver ones (~68%) (4). Recent progress in metastasis research has expanded our understanding of the molecular and cellular mechanisms behind this process (1, 5); however, it is still far from being fully understood. Metastasis is a multifaceted process comprised of multiple events. The formation of a secondary tumour after colonization at a distant site is preceded by basement membrane invasion and cell migration. Subsequently, cancer cells intravasate into the surrounding vasculature or lymphatic system. Cancer cells that are able to survive in the circulation will be able to reach a secondary tissue by extravasation and colonization (6). Through all these different events cancer cells are exposed to a collection of everchanging microenvironments with unique molecular characteristics (7, 8). Cancer cells must adapt to each of these

situations, altering the expression, localization, and activation of proteins to generate a metastasis.

In this context, isogenic cell lines, that share a common genetic background, but with different metastatic capabilities can be tremendously helpful. Quantitative proteomic analysis of such cell lines has an immense potential to unveil clinically relevant mechanisms underlying the metastatic progression (9-12). The most widely used cell models of CRC metastasis are the KM12 cell system and the SW480/SW620 pair of cell lines. The KM12 cell system includes three isogenic cell lines with different metastatic properties: non- or poorly-metastatic KM12C cells, liver metastatic KM12SM cells, and liver and lung metastatic KM12L4a cells (13, 14). KM12C cells were isolated from a tumour mass developed in a CRC patient in Duke's B_2 stage, equivalent to a T3N0M0 or T4N0M0 stage following the TNM staging described in the introduction section (13). The metastatic potential of these cells was evaluated by spleen injection in nude mice (13). In addition, the KM12L4a cell line was obtained through 3 cycles of isolation and subsequent intrasplenic reinjection of KM12C cells that were able to form liver metastases (13). On the other hand, KM12SM cells were derived from a rare liver metastasis developed after cecum injection of KM12C cells in nude mice (13). Numerous studies support a good correlation between the findings observed using the KM12C and KM12SM cells and patient samples, indicating that these isogenic cell lines recapitulate quite effectively critical issues in CRC liver metastasis (15-19).

Alternatively, the SW480/SW620 pair represents the chromosomal instability subtype of CRC that is clinically commonly observed (20, 21). SW480 cells were established by Dr. Leibovitz in 1976 from a Duke's B_2 primary adenocarcinoma of a 50-year-old male patient (20, 22). SW620 cells were isolated from the same patient as SW480 cells from a lymph

node metastasis that was obtained 6 months after the isolation of SW480 cells, when the cancer recurred presenting widely spread metastasis (20, 22). In terms of TNM staging, SW620 would correspond to either T3N1M0 or T4N1M0, equivalent to a Duke's C stage. SW480 and SW620 possess differences in xenograft metastatic potential *in vivo* (23), migratory propensity, and drug sensitivity (24), which also recapitulates the behaviour observed *in vivo*.

A key defining feature of eukaryotic cells is precisely their high level of compartmentalization, which is crucial for the portioning of biological processes. Such compartmentalization allows for localizing specific proteins to different locations and for creating distinct chemical environments. Hence, control over protein subcellular localization is a central part of cell physiology (25, 26). Indeed, many biological cell processes, such as signalling cascades, involve changes in protein subcellular localization, and protein mislocalization is often linked to disease (27, 28). In the context of cancer, changes in the abundance or subcellular localization of proteins, such as tumour suppressors or oncoproteins, has been frequently reported (27). Mislocalization of these proteins can prevent them from carrying on their function, subsequently altering their ability to suppress tumour cells or, for oncoproteins, increase their potential for inducing cancer development, metastasis or drug resistance. Consequently, the identification of mislocalized proteins can be of special relevance for the discovery of new cancer diagnostic markers or therapeutic targets.

Previous studies have separately compared the KM12 cell model and the SW480/SW620 cell pair at the proteome and secretome levels (12, 13), or focused on the pair KM12C and KM12SM of the KM12 cell system to spatially delineate proteins associated with liver metastasis (29). However, there has been no report applying quantitative proteomics to analyse jointly the different metastatic properties to lymph nodes

(SW480/SW620 cell pair), to the liver, and to the liver and lung (KM12C/KM12SM/KM12L4a cells); to identify proteins associated with CRC metastasis or specific of lymph nodes, liver or lung metastatic niches. Consequently, we compared the multidimensional protein content of five subcellular fractions (cytoplasmatic proteins, membrane proteins, nuclear proteins, chromatin bound proteins, and cytoskeletal proteins) and the secretome of the three KM12 cell lines and the SW480/SW620 pair of cells. Our aim was to identify altered proteins in abundance and localization in the most important metastatic niches of CRC besides lymph nodes (liver and lung). This would allow to gain further insights into metastatic CRC and to try to find novel relevant proteins associated with the disease, which might serve as diagnostic markers and/or therapeutic targets of intervention. Proteins were measured in parallel for the six separate subcellular fractions, outlining metastasis-associated and tropism-associated proteins.

To be able to undertake such characterization we decided to use TMT 11-plex labelling quantitative proteomic analysis. Compared to LF mass spectrometry, isotopic labelling allows for the combination of protein extract coming from different sources, thereby reducing the time required for MS runs. In addition, comparisons between label-free and isotopic labelling MS have shown that although LF can, in some instances, detect a larger number of proteins the difference with isotopic labelling is rather small and at the cost of an increase in reproducibility (30-33). In a direct comparison between TMT and LF for proteomic signalling pathways analysis, it was demonstrated that TMT had a slight advantage over LF regarding the number of altered proteins and pathways identified (30). Researchers demonstrated that compared to both TMT and LF, SILAC outperforms both in terms of technical variability, but LF was still the best alternative for phosphoproteome and proteome coverage (31). Though more stable and reproducible, SILAC labelling is hindered by the number of samples that can be analysed in parallel, which is commonly limited to

three. Consequently, given the nature of our experiment and the large number of samples (30 in total) under evaluation we settled down for chemical linking of isotopic labels. Similarly, at the time that we performed the characterization iTraq labelling was limited to 8 channels, while TMT allows for 11 and even 18 different labels. In addition, TMT was shown to perform better than LF for the evaluation of altered proteins levels (32).

After bioinformatics, alterations in abundance and localization for selected proteins from diverse subcellular localizations were validated by western blot (WB) and immunofluorescence (IF) using CRC cells. Immunohistochemistry (IHC) and WB using CRC patients' tissue samples supported the relevance of the results in the real-life scenario of CRC metastasis. Finally, ELISA confirmed the association and dysregulation of GLG1 in CRC metastasis. GLG1 plasma levels in control and CRC patients had the diagnostic potential to discriminate between control and advanced stages of the disease.

Results

Subcellular fractionation and analysis for differential protein expression and localization in lymph nodes, liver and lung metastatic tropism of colorectal cancer cells

Isogenic KM12 cells (non-metastatic KM12C cells, liver metastatic KM12SM cells and liver and lung metastatic KM12L4a cells) and isogenic SW480 (non-metastatic) and SW620 (lymph node metastasis) CRC cell lines with indicated metastatic tropisms were used to identify metastasis- and tropism-associated proteins (Figure R-1 1). Prior to in-depth proteomics analysis, cells were fractionated into five subcellular fractions (cytoplasm (CEB), membranes (MEB), nuclear proteins (NEB), chromatin-bound proteins (NEB-CBP), and cytoskeletal and insoluble proteins (PEB)). In addition to these five fractions, a sixth fraction corresponding to the conditioned medium of the cells -secretome- was also analysed. The total of six fractions from the five cell lines were separately trypsin-digested and labelled (Figure R-1 1). Each one of the isobaric tags of TMT 11-plex labelling kit was used to label a separate and unique fraction from an independent cell line. According to the labelling scheme used, subcellular compartments were grouped in pairs for the analysis: (i) membrane and cytoplasmic proteins, (ii) nuclear and chromatin-bound proteins, and (iii) cytoskeletal and secreted proteins. Upon labelling, the three independent TMT experiments were separately fractionated into 12 fractions according to peptide hydrophobicity prior to LC–MS/MS analysis using a Q Exactive mass spectrometer.

After normalization (Figure R-1 2), a total of 4031 individual proteins were identified and quantified using MaxQuant, from which, 2305 proteins were observed in at least two or more pairs of subcellular compartments and 1157 proteins in all subcellular fractions (Table S6). Next, we carried out a gene ontology (GO) cellular component classification of the identified and quantified proteins to confirm the correct subcellular fractionation of

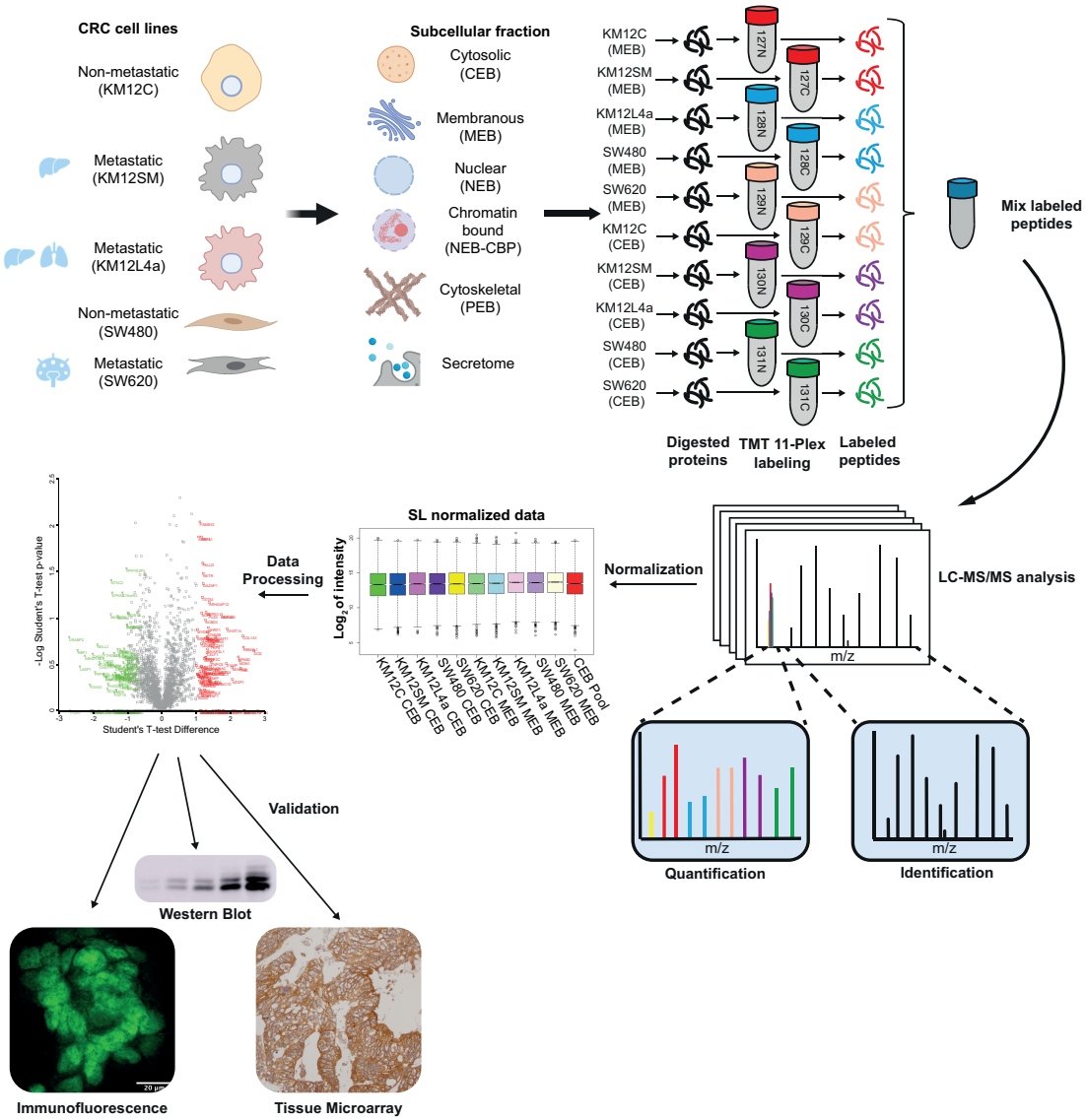


Figure R-1 1: Workflow of the approach for the multidimensional proteomics analysis of the isogenic non-metastatic and metastatic colorectal cancer cell lines. CRC cells with indicated tropisms were subcellularly fractionated prior to TMT 11-plex labeling mixed in a 1:1 proportion and peptides separated using the High pH Reversed-Phase Peptide Fractionation Kit. Then, the fractionated peptides obtained per TMT experiment were analyzed onto a Q Exactive mass spectrometer. MaxQuant and Perseus were used for data analysis and identify proteins dysregulated in CRC metastasis, which were validated by different orthogonal approaches (WB, IF or IHC, among others).

the multidimensional proteomics analysis. In all of the subcellular fractions analysed, proteins were correctly classified within the first two GO cellular component classification hits of each fraction, except for the secretome proteins, where proteins presenting dual localizations are usually found. In the secretome, 434 proteins were classified among extracellular exosome proteins.

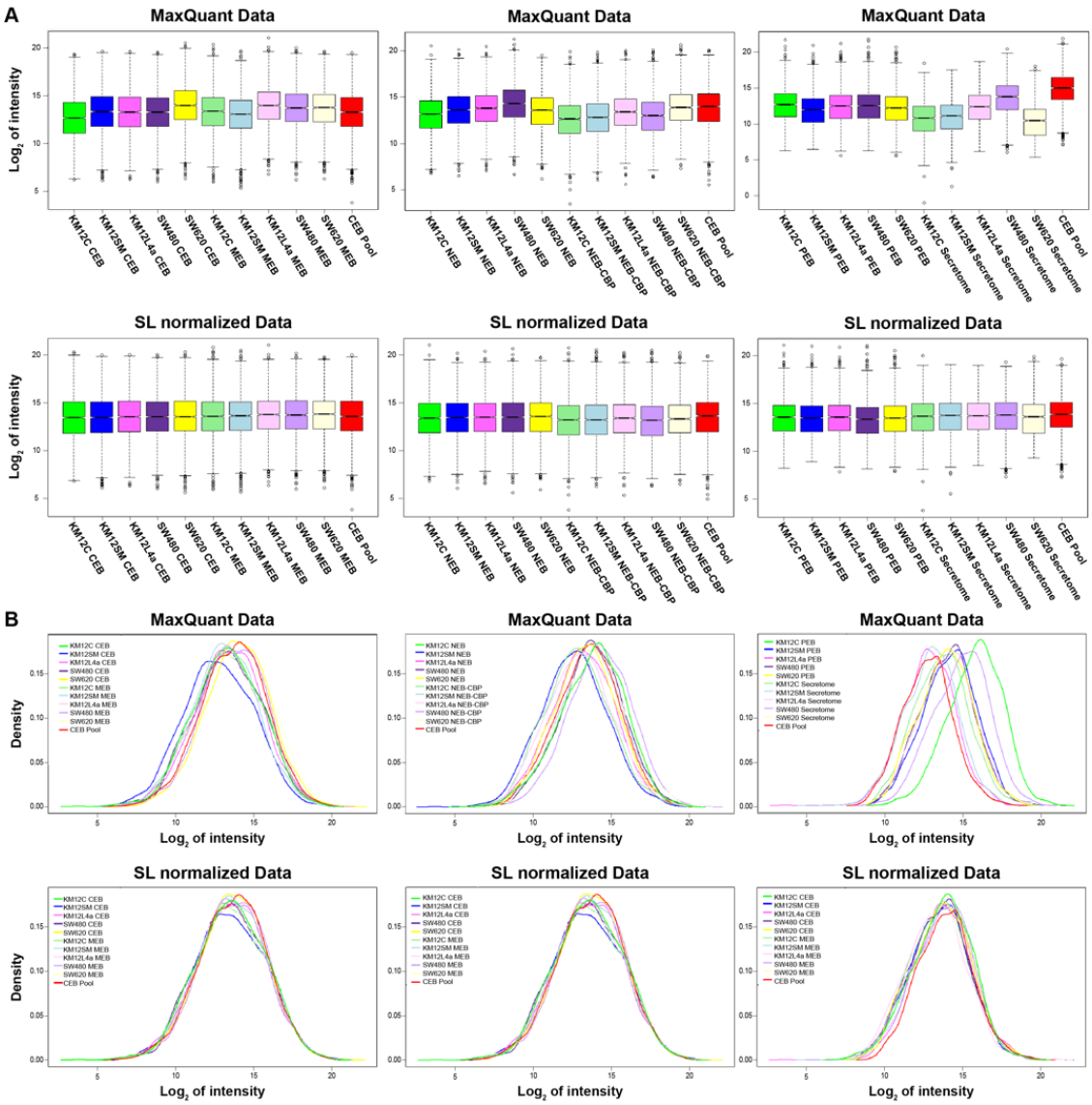


Figure R-1 2: MaxQuant data SL normalization **(A)** Box plots of the log_2 protein intensities mean before (up) and after (down) data normalization of each TMT reporter. **(B)** Histogram of the log_2 protein intensities for each TMT reporter before (up) and after (down) data normalization.

Collectively, these results pointed out to a correct fractionation of the subcellular organelles analysed.

Mapping spatial protein alterations in crc metastatic cells

Then, we focused the analysis on one-to-one (i.e., SW480 vs. SW620) or grouped (non-metastatic vs. metastatic) comparisons for the identification of dysregulated proteins metastasis-associated or associated with a specific tropism using Perseus (Table S7). A fold change ≥ 2 and ≤ 0.5 was used as the cut-off for upregulated and downregulated proteins, respectively. In general, we could observe a higher number of proteins downregulated than upregulated for most of the cell compartments under study. Interestingly, 582 proteins showed an opposite regulation in abundance in different compartments, indicating that these proteins were dysregulated in abundance and in localization.

The total number of dysregulated proteins for all compartments was larger for those one-to-one comparisons in which cells were more different. The total number of dysregulated proteins for the KM12C vs. KM12SM comparison was 1353, for KM12C vs. KM12L4a was 1666 and for KM12SM vs. KM12L4a, it was 1337. Furthermore, the comparison that showed the largest number of dysregulated proteins was the one between SW480 and SW620 cell lines with 2436 dysregulated proteins. These cell lines—SW480 and SW620—were arguably the two most distinct cell lines as they were derived from primary cells from the original tumour mass and a metastasis developed in the lymph nodes, respectively. Regarding comparisons between compartments, the subcellular compartments that showed the highest dysregulation of proteins were the secretome and nuclear and chromatin-bound proteins (Figure R-1 3 and Figure R-1 4).

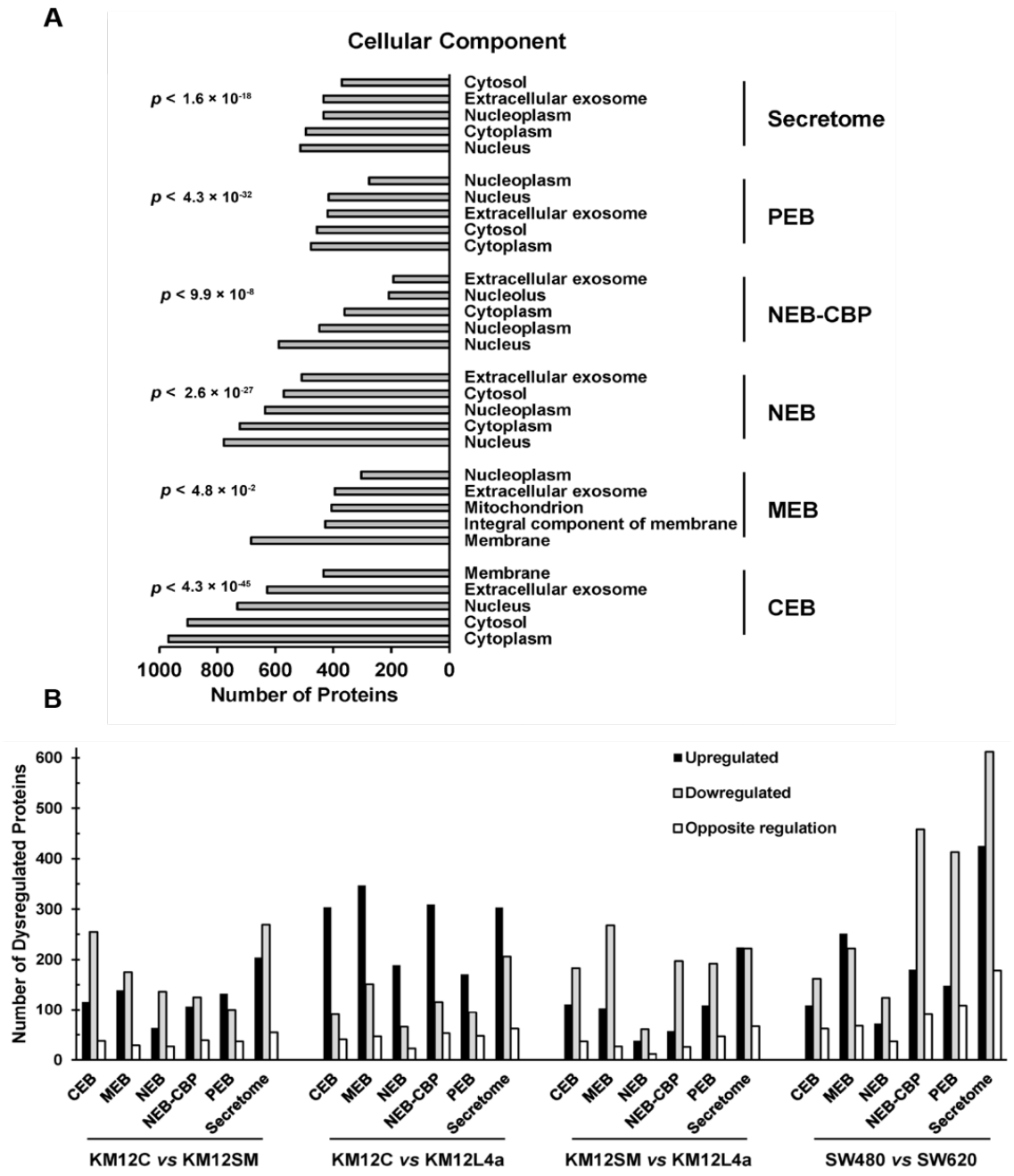


Figure R-1 3: Analysis of the identified and quantified proteins in each subcellular fraction of metastatic and non-metastatic colorectal cancer cells. (**A**) Cellular component GO analysis for the subcellular classification of the proteins in each compartment was made with the DAVID database. The top five subcellular GO classifications in each fraction with the calculated p-value showed a good correlation with the subcellular localization of all identified and quantified proteins. (**B**) The total number of dysregulated proteins (upregulated and downregulated proteins, and those showing an opposite dysregulation in different subcellular fractions) was larger for those comparisons in which cells were more different.

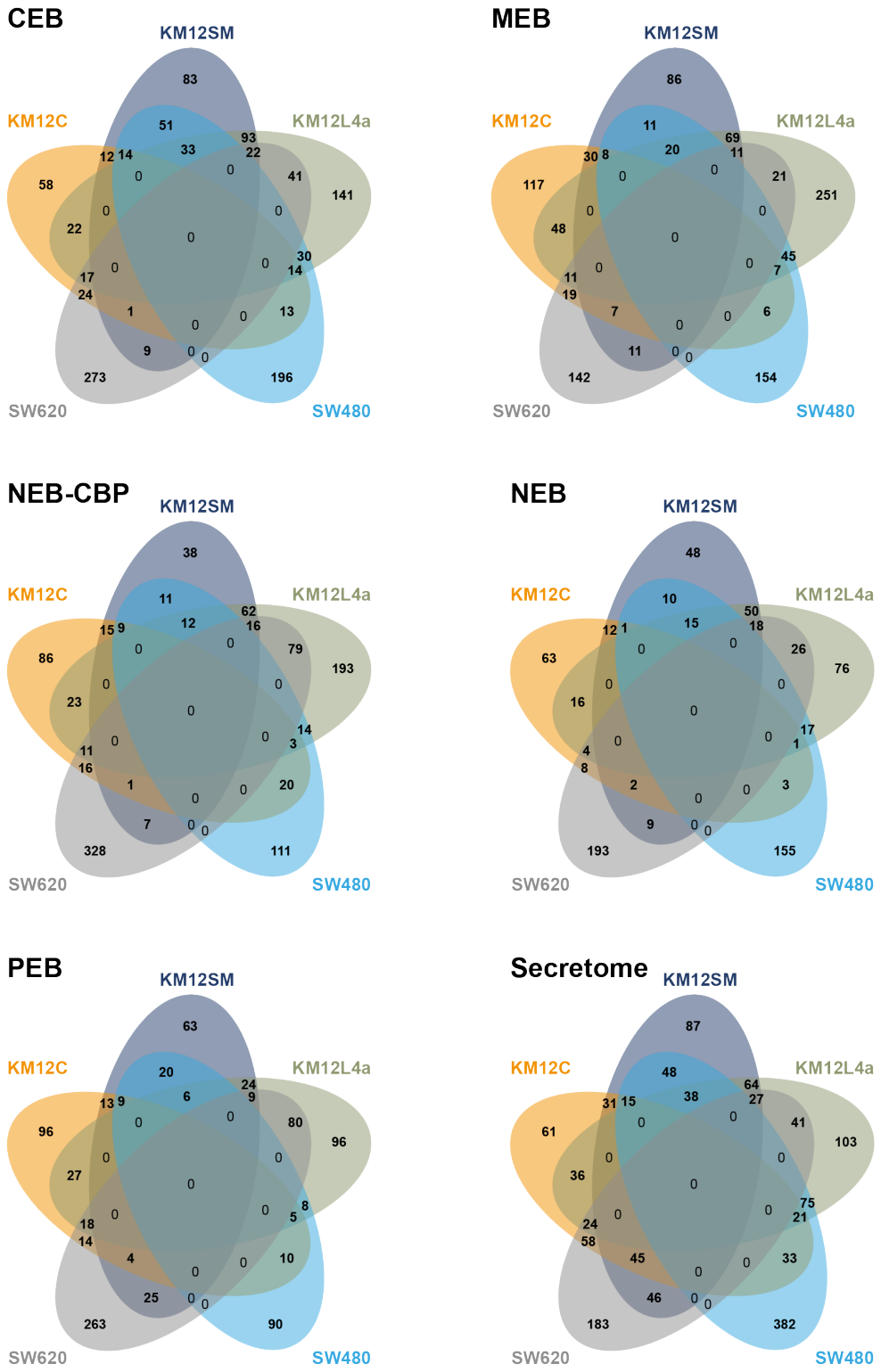


Figure R-1 4: Analysis of the dysregulated proteins in CRC metastatic and non-metastatic cells. Venn diagram of the differentially dysregulated proteins in CEB, MEB, NEB, NEB-CBP, PEB and secretome of non-metastatic (KM12C and SW480) CRC cells and metastatic (KM12SM, KM12L4a, and SW620) CRC cells.

Then, a gene set enrichment analysis (GSEA) was performed to identify the most dysregulated pathways in each subcellular compartment for grouped metastatic vs. non-metastatic CRC cells (Table S8). Taking as reference the pathways of the gene ontology, KEGG pathways and the human reactome, about 1800 altered pathways were observed using the dataset of differentially expressed proteins in each compartment (Figure R-1 5 and Figure R-1 6). For each compartment and each pathway, the normalized enrichment score (NES) was calculated. Among the top significant altered pathways, actin reorganization in CEB, RNA-processing, and response to external stimulus in MEB, vesicle mediated transport in NEB, cellular component disassembly in NEB-CBP, positive regulation of transcription and positive regulation of nucleobase containing compound metabolic process in PEB, and reorganization of extracellular matrix in secretome were observed in the analysis (Figure R-1 5 and Figure R-1 6).

Noticeably, it has been described that the secretome contains, besides exosome and extracellular proteins, other proteins that possess dual localizations in other subcellular compartments (i.e., soluble membrane receptors). Moreover, all the enriched pathways in the secretome were related to encapsulation, extracellular matrix, or integrin cell surface interactions, except reticulum or unfolded protein response that were related to extra-cellular protein constituents.

Collectively, these data highlight not only a correct fractionation of the subcellular fractions (including the secretome) but also a vast dysregulation of proteins and pathways in CRC isogenic cells to acquire metastatic properties.

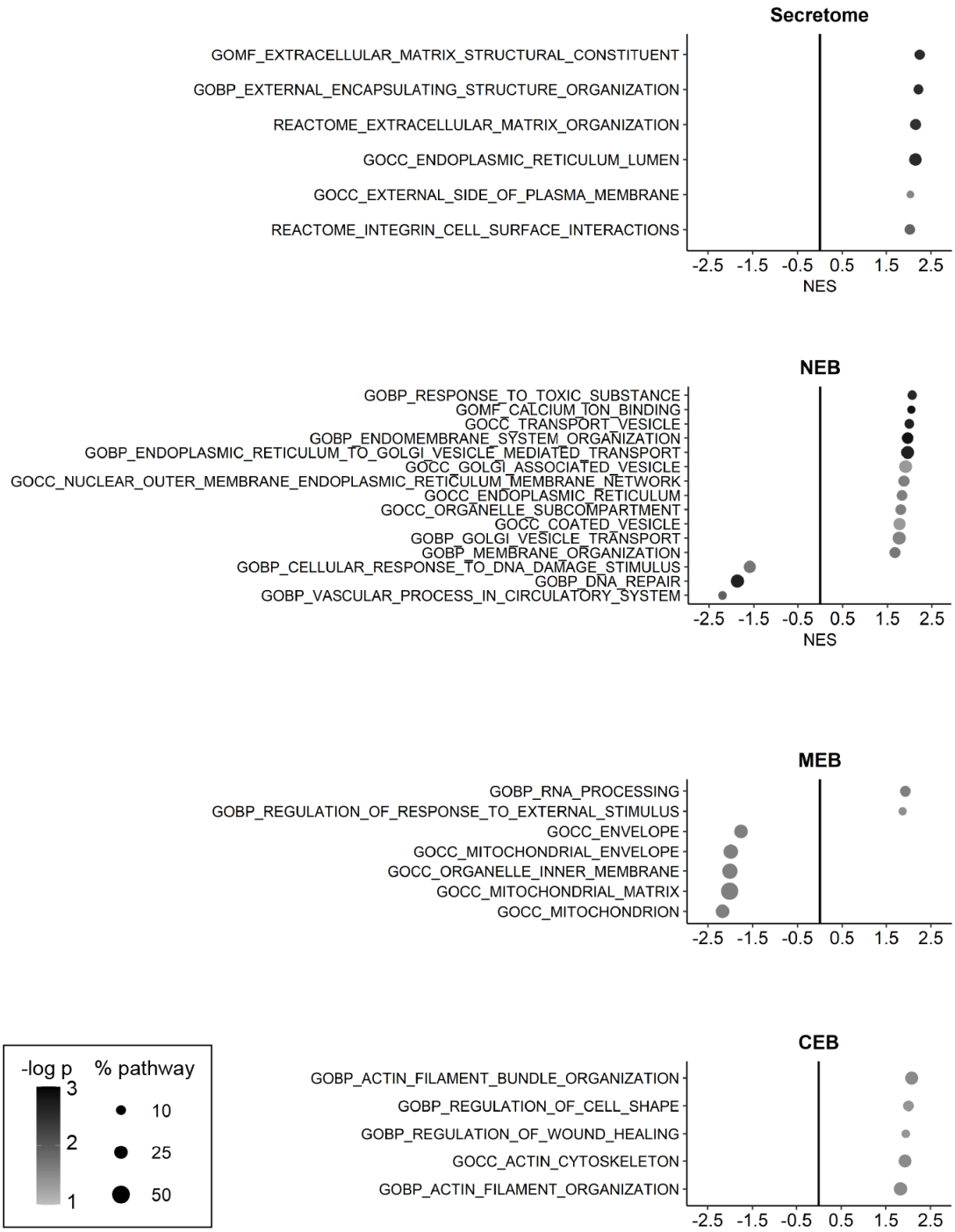


Figure R-1 5: Gene Set Enrichment Analysis allowed the identification of the most dysregulated pathways in Secretome, NEB, MEB and CEB compartments comparing metastatic vs non-metastatic CRC cells. For each compartment and each pathway, the normalized enrichment score (NES) was calculated to identify the most dysregulated pathways according to the biological function.

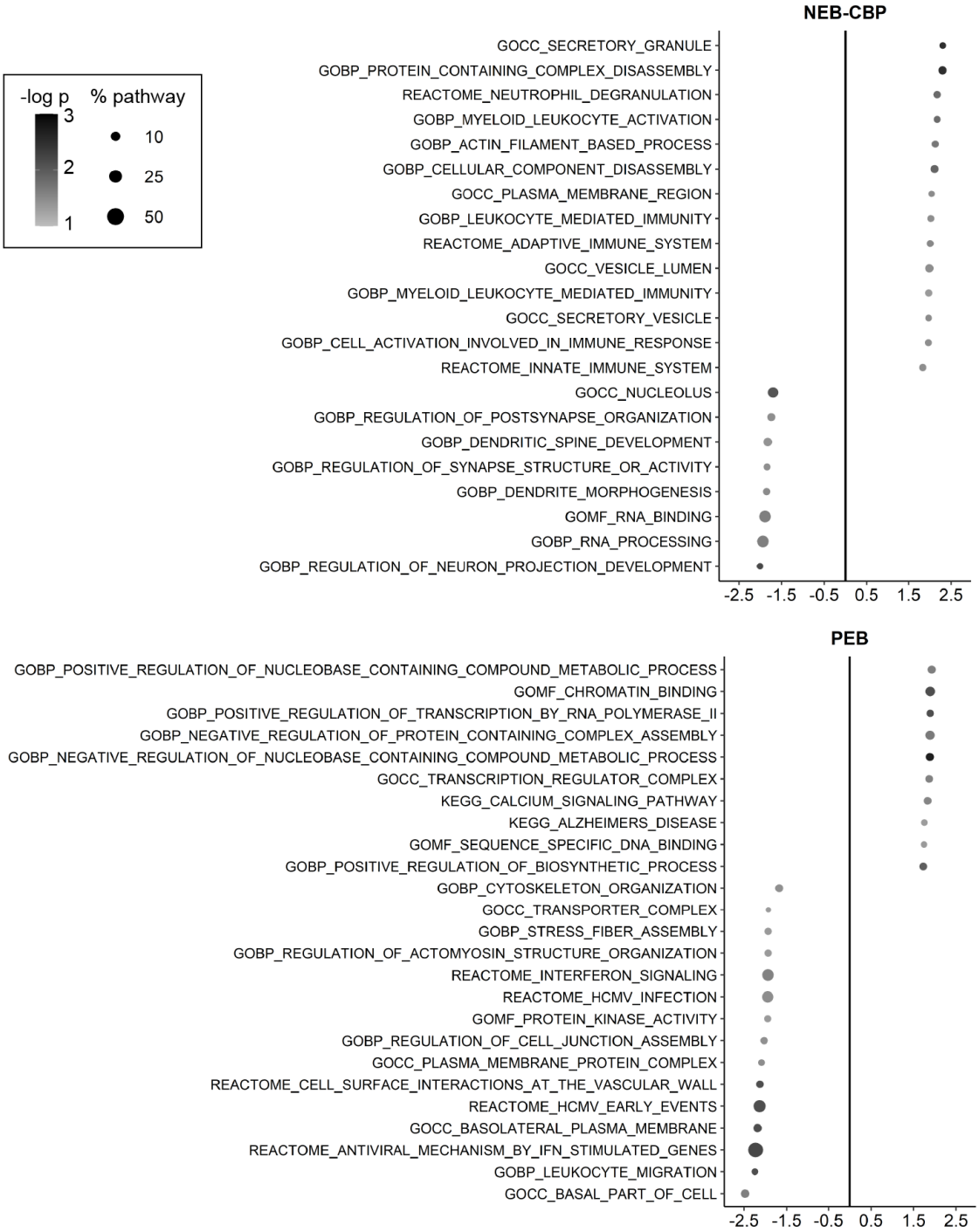


Figure R-1 6: Gene Set Enrichment Analysis allowed the identification of the most dysregulated pathways in NEB-CBP and PEB compartments comparing metastatic vs non-metastatic CRC cells. For each compartment and each pathway, the normalized enrichment score (NES) was calculated to identify the most dysregulated pathways according to the biological function.

Protein clusters, pathways and network analysis of differentially expressed proteins in CRC metastatic cells

To characterize at the protein level the different metastatic properties of the cells, we carried out a global *in silico* analysis of the differentially expressed proteins to visualize protein-protein interactions and the most significantly altered protein clusters and macromolecular complexes related to proteins associated with metastasis in one, two, or three of the metastatic cell lines. Among the observed vast dysregulation of proteins

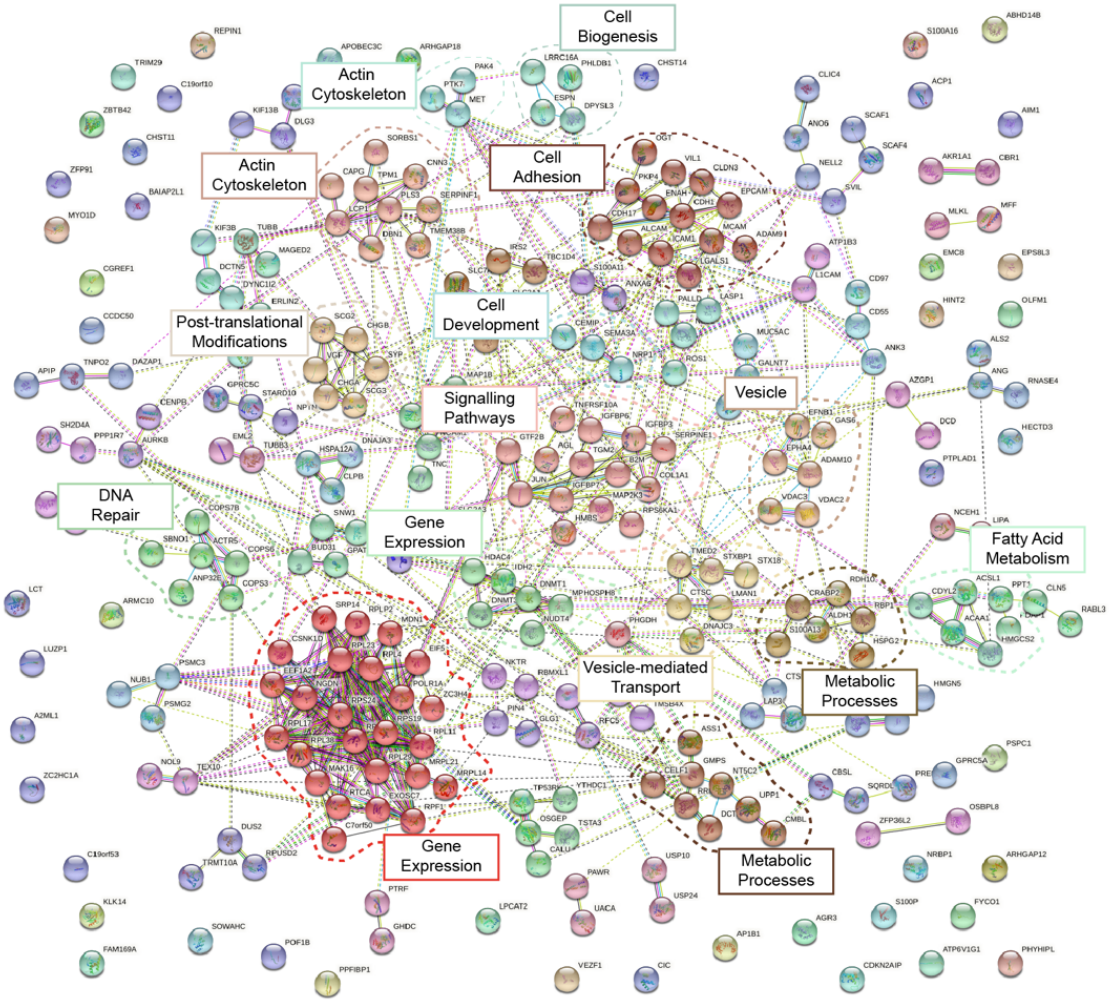


Figure R-1 7: Bioinformatics analysis of the dysregulated proteins in colorectal cancer metastasis. String revealed 15 different clusters of interaction among the dysregulated proteins associated with CRC related to DNA repair, gene expression, metabolism, signaling, cell development, cell adhesion, actin cytoskeleton, and transport or vesicle-mediated transport.

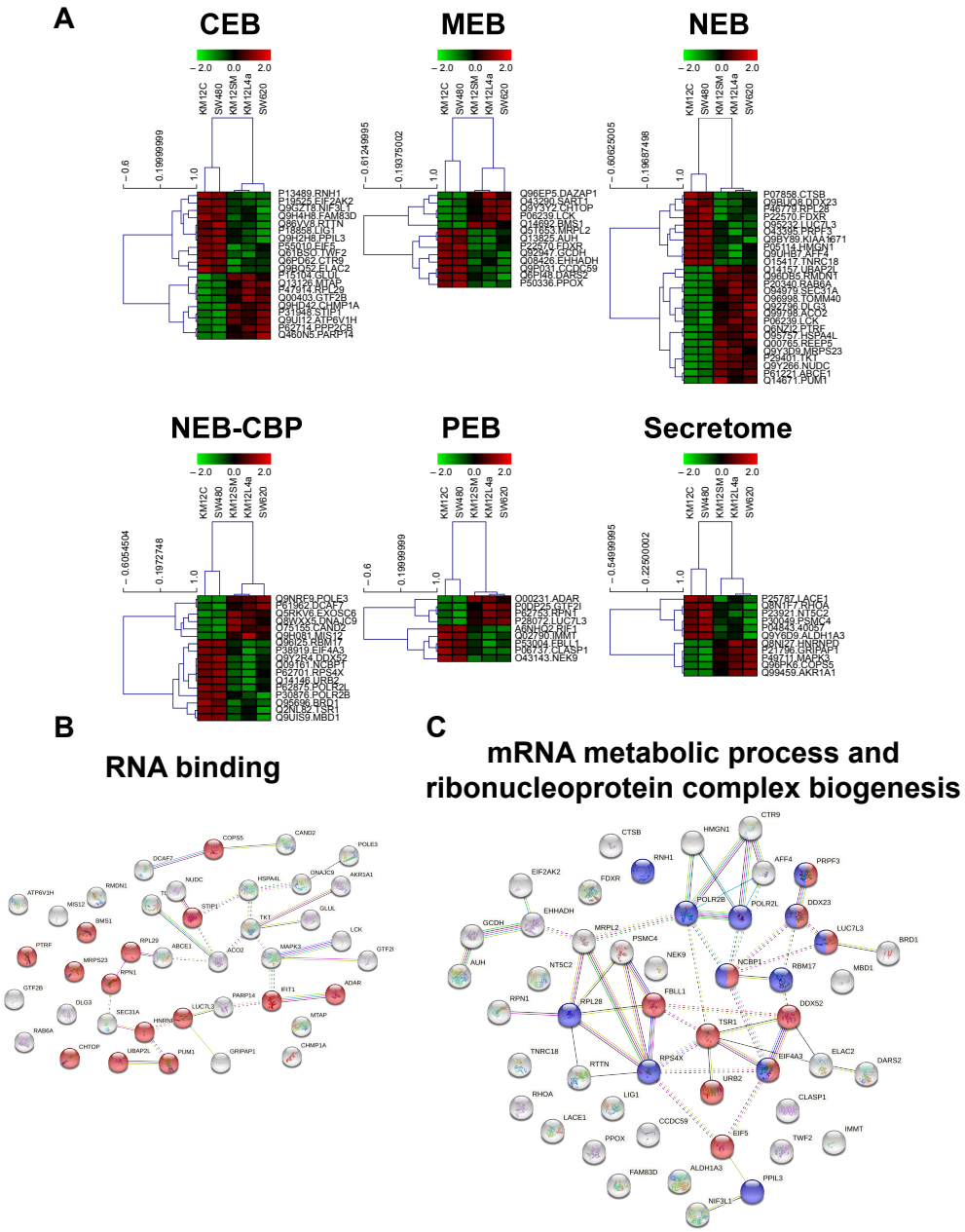


Figure R-1 8: (**A**) Hierarchical clustering of dysregulated—upregulated and downregulated—proteins in the different compartments showed the significant discrimination between non-metastatic (KM12C and SW480) CRC cells and metastatic (KM12SM, KM12L4a, and SW620) CRC cells ($p < 0.05$). (**B**) RNA binding in red was the molecular function more enriched in upregulated proteins in all metastatic localizations in comparison to non-metastatic cells. (**C**) mRNA metabolic process (in red) and ribonucleoprotein complex biogenesis (in blue) were the molecular functions more enriched among the downregulated proteins found to be downregulated in all metastatic localizations in comparison to non-metastatic cells.

associated with metastasis (lymph nodes, liver, and liver and lung metastasis), released proteins together with nuclear proteins were amongst the most dysregulated proteins as assessed by String (Figure R-1 7). Which contributed to the dysregulation of different clusters of proteins related to metabolic processes, fatty acid metabolism, vesicle-mediated transport, DNA repair, cell development, signalling pathways, post-translational modifications, gene expression, and cell adhesion ($p < 0.05$). Among them, gene expression, with two different clusters containing 6 and 25 proteins ($p < 0.05$), was the most overrepresented dysregulated process.

Then, we surveyed the proteomics dataset for the identification of dysregulated proteins associated with metastasis in comparison to non-metastatic cells (Table R-1 1 and S8 and Figure R-1 8). Proteins identified and quantified in each compartment were subjected to supervised clustering analysis (p value < 0.05; Kendall's Tau algorithm) using MultiExperiment Viewer (MeV), according to their differential expression in the subcellular fractions and the metastatic properties of the cells (Figure R-1 8A). There, the interactome of those proteins showing metastasis associated dysregulation in abundance in the different subcellular localizations was visualized. Interestingly, among the proteins observed to be upregulated, we found that 15 out of the 45 commonly upregulated metastasis-associated proteins were proteins involved in RNA binding (FDR = 0.0064; Figure R-1 8B). Regarding downregulated proteins, we observed that the most overrepresented biological processes involved mRNA metabolic process and ribonucleoprotein complex biogenesis (FDR = 0.00011; Figure R-1 8C).

Table R-1 1: Dysregulated proteins associated with CRC metastasis (fold change ≥2 or ≤0.5) in the indicated subcellular compartment as observed in the quantitative proteomics analysis. * Fold change: ratio between KM12SM and KM12C expression

Protein	Accession Number	Association	Compartment	Fold Change *	Up- or Downregulated
GAS6	Q14393	Metastatic cells	Secretome	37.24	Up
DCD	P81605	Metastatic cells	Secretome	35.46	Up
c-MET	P08581	Metastatic cells	Secretome	23.31	Up
LMAN1	P49257	Metastatic cells	Secretome	19.17	Up
CEMIP	Q8WUJ3	Metastatic cells	Secretome	12.39	Up
HMGN5	P82970	Metastatic cells	PEB	11.34	Up
MUC5AC	P98088	Metastatic cells	MEB	9.85	Up
TNFRSF10A	O00220	Metastatic cells	CEB	9.37	Up
VIL1	P09327	Metastatic cells	NEB	7.71	Up
GLG1	Q92896	Metastatic cells	Secretome	7.54	Up
LCP1	P13796	Metastatic cells	CEB	6.30	Up

AGR3	Q8TD06	Metastatic cells	Secretome	5.51	Up
PLS3	P13797	Metastatic cells	Secretome	5.33	Up
CDKN2AIP	Q9NXV6	Metastatic cells	CEB	5.04	Up
SNX9	Q9Y5X1	Liver and Lung metastasis	Secretome	4.97	Up
PHYHIPL	Q96FC7	Metastatic cells	MEB	4.07	Up
ARHGAP18	Q8N392	Metastatic cells	NEB	4.01	Up
S100A16	Q96FQ6	Metastatic cells	NEB	3.98	Up
FBP1	P09467	Metastatic cells	CEB	3.96	Up
S100P	P25815	Metastatic cells	Secretome	3.30	Up
PTRF	Q6NZI2	Metastatic cells	NBP	3.02	Up
CLDN3	O15551	Liver and Liver and Lung metastasis	PEB	2.57	Up
AHR	P35869	Liver and Lung metastasis	NBP	2.52	Up
SCRIB	Q14160	Liver metastasis	MEB	2.13	Up
MAP2K3	P46734	Metastatic cells	NBP	2.10	Up

KIF13B	Q9NQT8	Metastatic cells	PEB	2.06	Up
BAIAP2	Q9UQB8	Lymph nodes metastasis	MEB	2.04	Up
LGALS1	P09382	Metastatic cells	CEB	0.20	Down
PHLDB1	Q86UU1	Metastatic cells	MEB	0.19	Down
SLC2A1	P11166	Metastatic cells	MEB	0.18	Down
FYCO1	Q9BQS8	Metastatic cells	MEB	0.16	Down
L1CAM	P32004	Metastatic cells	MEB	0.16	Down
ASS1	P00966	Metastatic cells	Secretome	0.15	Down
ICAM1	P05362	Metastatic cells	MEB	0.15	Down
MCAM	P43121	Metastatic cells	MEB	0.13	Down
TGM2	P21980	Metastatic cells	Secretome	0.05	Down
CRABP2	P29373	Metastatic cells	Secretome	0.02	Down
RBP1	P09455	Metastatic cells	Secretome	0.02	Down

Validation of dysregulation of proteins in abundance and/or localization

Validation of the MS dataset and of proteins associated with metastasis was performed by meta-analysis, WB and IF. First, we observed by meta-analysis proteins previously associated with CRC metastasis, such as GAS6, MET, or MUC5AC (Table R-1 1), which highlighted the utility of the spatial proteomics approach for the identification of CRC biomarkers (34, 35). Next, we focused the validation on selected dysregulated proteins from different subcellular compartments (Table R-1 1). These proteins were upregulated in metastatic cell lines in two or more subcellular compartments or showed the highest upregulation in CRC metastatic cell lines in comparison to isogenic non-metastatic cells. According to these criteria, TNFRSF10A and CDKN2AIP from the

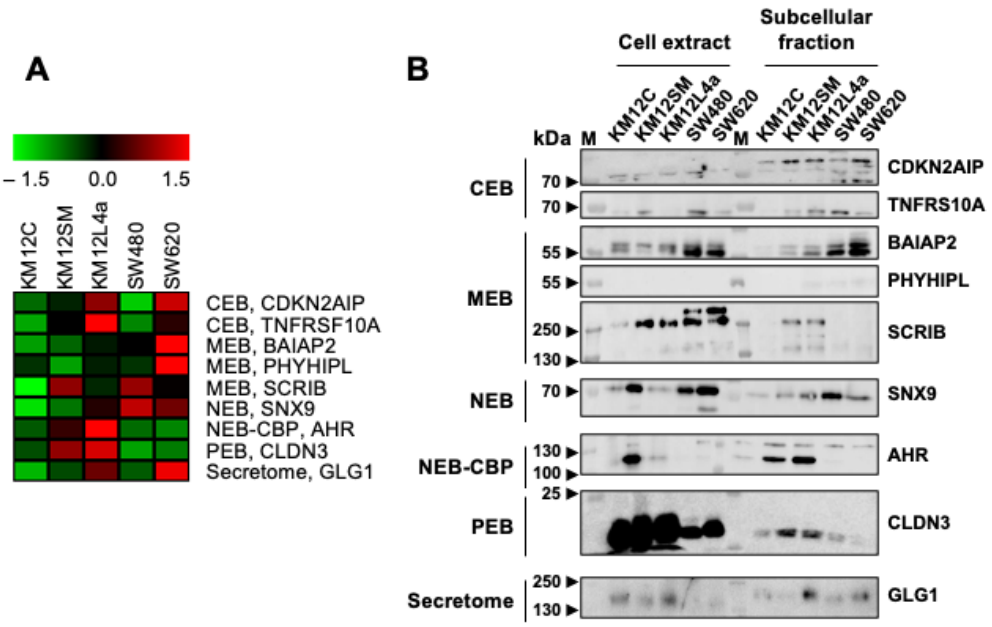


Figure R-1 9: Hierarchical clustering and validation of indicated dysregulated proteins. (**A**) Unsupervised cluster analysis of the indicated altered proteins in different subcellular localizations. Green, downregulation, and red, overexpression. Color scale is related to the fold-change observed for each protein in each subcellular fraction. (**B**) A total of 10 µg of the indicated total extracts or subcellular fractions of the five isogenic CRC cells of the study were subjected to WB analysis with specific antibodies.

cytoplasm, BAIAP2, PHYHIPL, and SCRIB from membrane, SNX9 from nucleus, AHR from chromatin-bound proteins, CLDN3 from the cytoskeletal fraction and GLG1 from the secretome were selected for validation (Table R-1 1 and Figure R-1 9). By WB analysis, BAIAP2, SCRIB, SNX9, TNFRSF10A, PHYHIPL, CDKN2AIP, CLDN3, AHR, and GLG1 showed a concordant protein dysregulation as observed by MS (Figure R-1 9). A clear increment in the protein expression levels of BAIAP2, CDKN2AIP, AHR, CLDN3, and SCRIB could be observed in parallel to the metastatic properties of the cells. In addition, the differences detected at the subcellular level either by proteomics or by WB were in general not observed at the whole extract level. This further supports the potential of the subcellular fractioning for elucidation of dysregulation of proteins in CRC metastasis of previously overlooked markers or highly expressed proteins.

In parallel, we performed IF staining of selected markers in the five CRC cells of the study under native conditions to further investigate for alterations in their subcellular localization and demonstrate their relevance in CRC (Figure R-1 10 and Figure R-1 11). Besides the further confirmation that AHR, BAIAP2, CLDN3, PHYHIPL, and SCRIB are highly expressed in highly metastatic CRC cells in comparison to non-metastatic isogenic cells, changes in the localization of these proteins could also be observed between non-metastatic and metastatic cells. In this sense, we could observe differences in the localization of AHR, BAIAP2, PHYHIPL, and SCRIB and among cells. AHR shifted from a dot-like distribution, which could be associated with a vesicular distribution in the cytoplasm of KM12C cells, to a more marked nuclear localization in the metastatic KM12SM and KM12L4a cells. In the SW480 and SW620 cells, AHR was mainly located to the nucleus of the cells; although based on the fluorescence intensity, higher levels of nuclear AHR in the SW620 cells were observed in comparison to SW480. Moreover, in concordance with proteomics data, in the non-metastatic KM12C and SW480 cells, BAIAP2

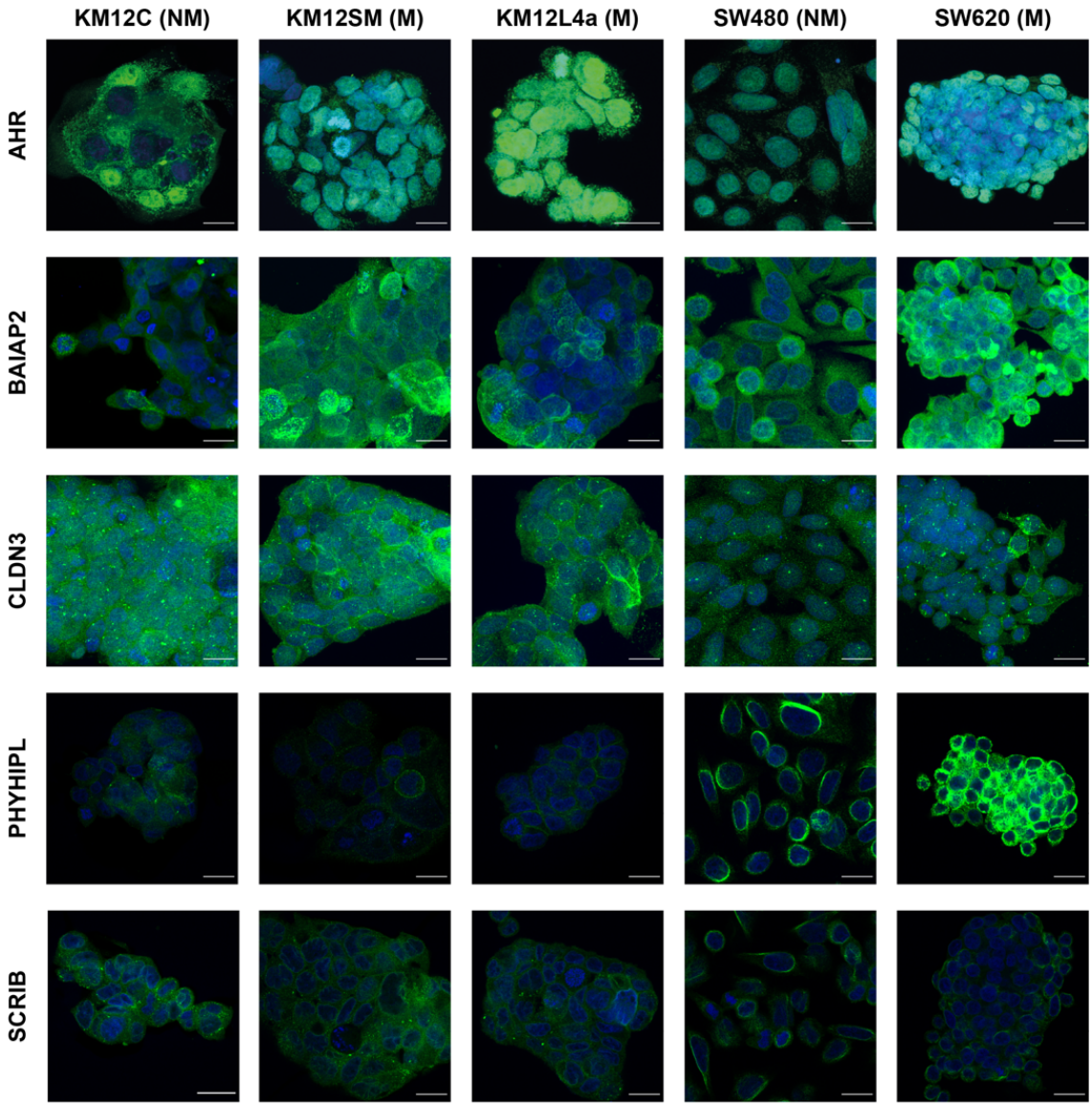


Figure R-1 10: Confocal microscopy analysis of differential protein localization among KM12C, KM12SM, KM12L4a, SW480, and SW620 colorectal cancer cells. Cells were cultured on glass coverslips pretreated with Matrigel for 24 h and subjected to confocal microscopy analysis using antibodies for AHR, BAIAP2, CLDN3, PHYHIPL, and SCRIB (green). Cells were counterstained with the nuclear probe DAPI (blue). Representative images show a differential staining distribution in the different cellular compartments for these proteins between the metastatic (KM12SM, KM12L4a, and SW620 cells) and non-metastatic (KM12C and SW480) cells. M, metastatic cells. NM, non-metastatic cells. Scale Bar 20 μm. PHYHIPL and SCRIB are single stacks, whereas the other are max projections. BAIAP2 KM12L4a cells settings were corrected to avoid saturation.

was mainly localized in the cytoplasm, whereas BAIAP2 was mainly localized in the cell membrane in metastatic KM12SM, KM12L4a, and SW620 cells. PHYHIPL was localized to the cell edges and contact areas between cells in the metastatic KM12SM and KM12L4a cells, in concordance with its increase in the membrane fraction in comparison to KM12C cells previously observed by proteomics. Furthermore, we could see a change in the distribution of SCRIB from being mainly located in the periphery of the nucleus to being more evenly distributed throughout the whole cell when comparing SW480 and SW620 cells. Finally, for CLDN3, KM12L4a cells showed a more peripheral localization of the protein, while KM12C had a more even, unspecific, distribution. However, in KM12SM cells the change towards the cell edges was not as noticeable as for KM12L4a but could still be detected.

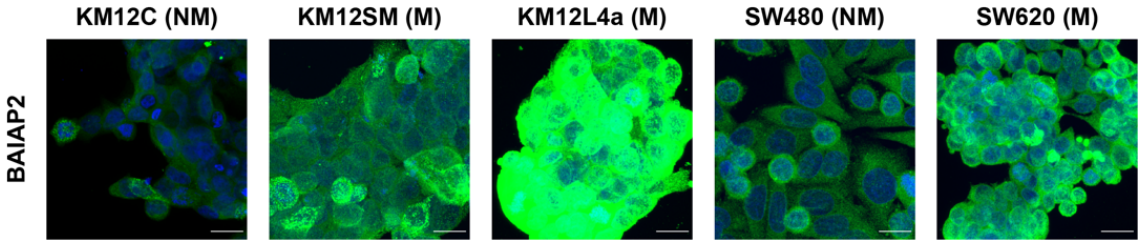


Figure R-1 11: Confocal microscopy analysis of BAIAP2 under same acquisition conditions, acquired in parallel to the acquisition of the images shown Figure R-1 10 in the panel corresponding to BAIAP2.

Collectively, isogenic CRC metastatic cells were observed to present a vast number of changes in the abundance and in the spatial distribution of proteins that might be relevant for CRC metastasis. Such changes would have been impossible to detect without subcellular fractionation.

Relevance of Dysregulated Proteins in Colorectal Cancer

Next, the selected proteins were further analysed by WB and IHC using actual CRC samples to determine the relevance of dysregulated proteins in the disease. By WB we studied the protein content of selected proteins

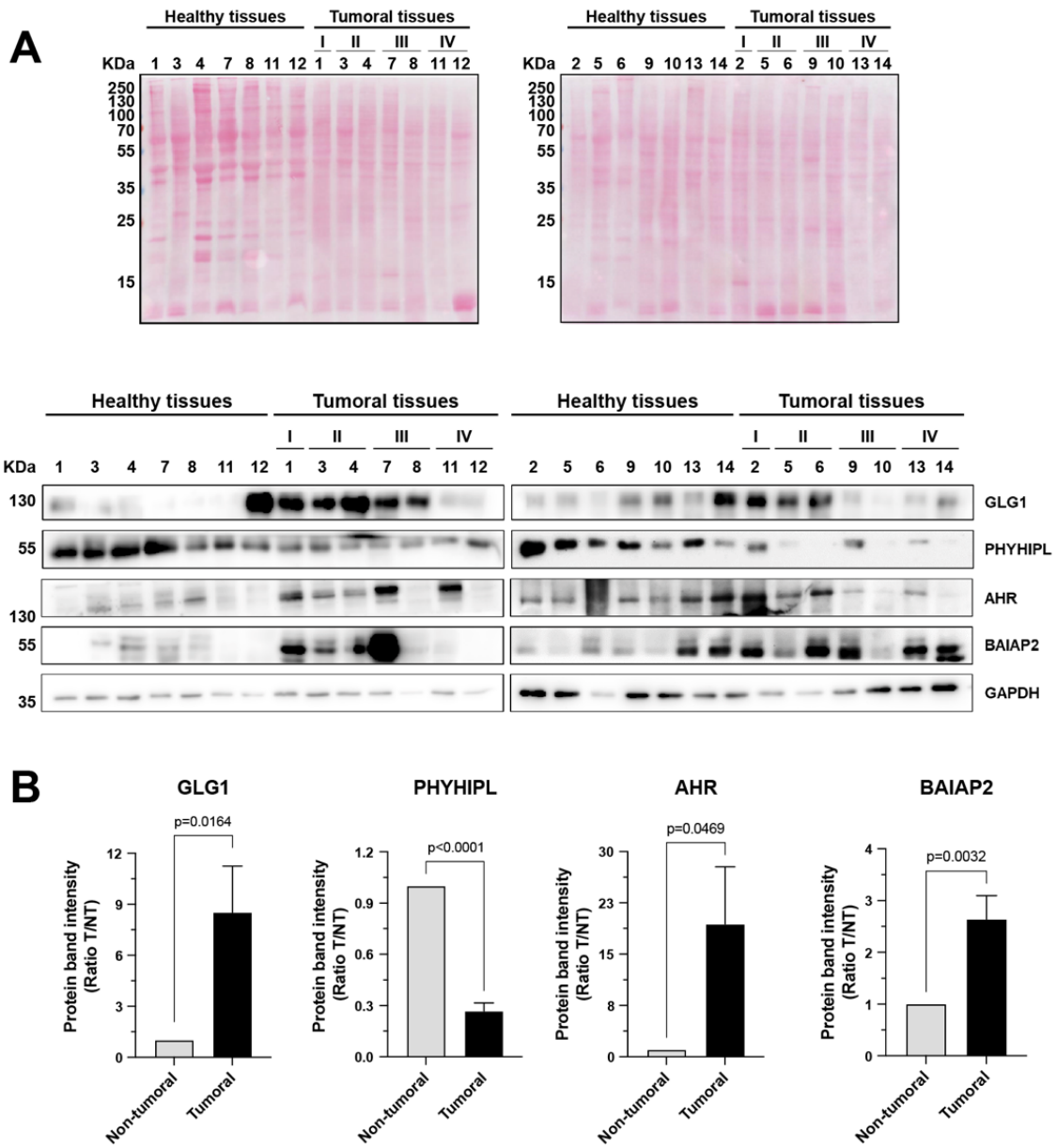


Figure R-1 12: Analysis of the dysregulated proteins in CRC patients tissue samples. (**A**) Ponceau red staining and WB analysis of 14 paired tumoral and non-tumoral tissue samples from CRC patients against indicated dysregulated proteins in CRC. GLG1, AHR, and BAIAP2 were upregulated in tumoral tissue samples, with GLG1 and AHR mainly overexpressed at earlier stages of the disease. On the contrary, PHYHIPL was downregulated in tumoral tissues in comparison with paired non-tumoral tissue samples from patients. GAPDH was used as the control in the assay. Protein bands were quantified with ImageJ and normalized according to the total protein lane content of the Ponceau red staining. (**B**) GLG1, AHR, and BAIAP2, and PHYHIPL were found significantly upregulated and downregulated, respectively, in tumoral tissue samples in comparison with non-tumoral tissue samples from CRC patients.

in paired tumoral and non-tumoral CRC patients' tissue samples (n = 14) at stages I-IV (Figure R-1 12). We found a statistically significant up-regulation of GLG1, AHR, and BAIAP2 protein levels in tumoral tissue samples in comparison with normal tissues, with higher levels of GLG1 and AHR at early stages (stage I-II) of the disease. On the contrary, PHYHIPL was statistically found downregulated in tumoral tissues in comparison with non-tumoral tissue samples (Figure R-1 12). Noticeably, the absence of metastatic tumoral samples avoided us to confirm their dysregulation in the different sites of metastasis.

Finally, we analysed the physiological relevance of the dysregulation of BAIAP2 by tissue microarray analysis (TMA) of patient derived samples (Figure R-1 13). From the markers previously validated and analysed, we

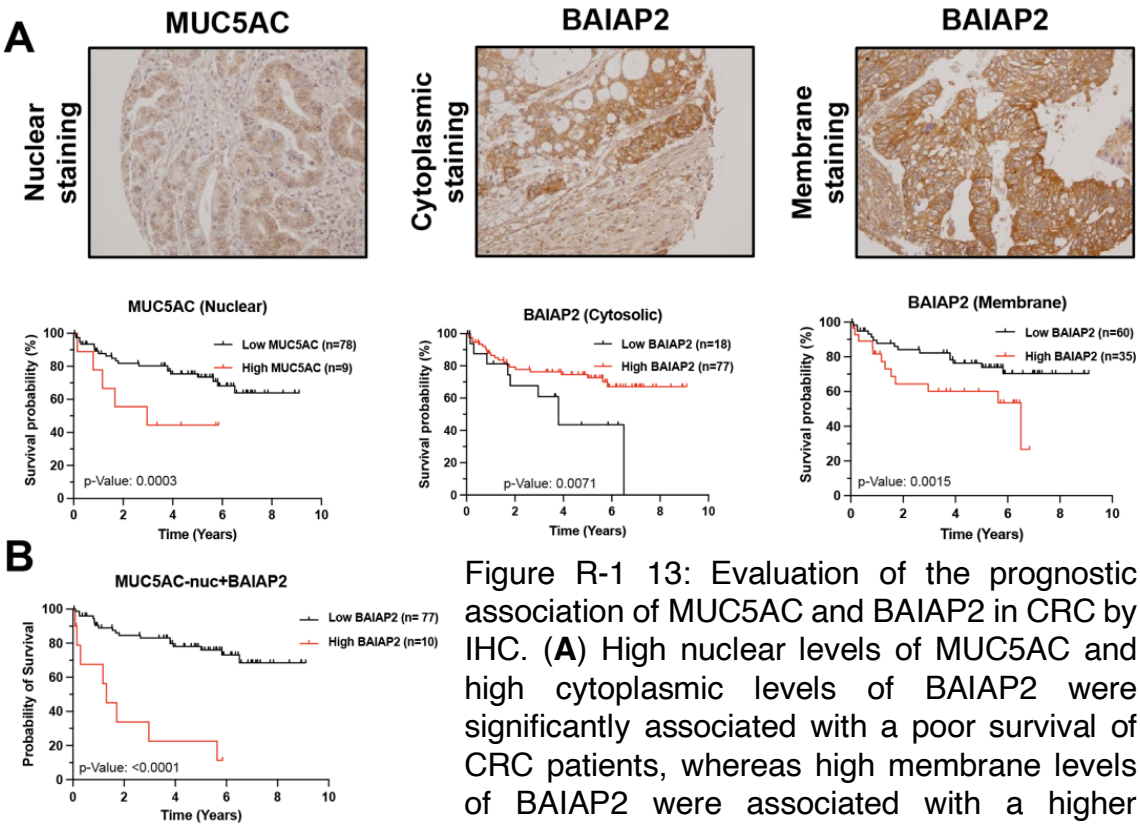


Figure R-1 13: Evaluation of the prognostic association of MUC5AC and BAIAP2 in CRC by IHC. (**A**) High nuclear levels of MUC5AC and high cytoplasmic levels of BAIAP2 were significantly associated with a poor survival of CRC patients, whereas high membrane levels of BAIAP2 were associated with a higher survival of patients. (**B**) The combination of nuclear MUC5AC and membrane BAIAP2 protein staining showed the most significant differences in survival between CRC patients and healthy individuals

focused on BAIAP2 as the other markers for which we attempted TMA validation (PHYHIPL, GLG1, SCRIB and BAIAP2) we did not find a significant association between protein levels and survival (data not shown). MUC5AC was used as control because of its known marker character in CRC (36-38), and its identification and quantification in our experimental setup. MUC5AC has been previously associated with different types of cancer both as a good and bad prognostic marker (36-38). Here, it was observed that nuclear levels of MUC5AC had a significant association with poor patient survival (Figure R-1 13A). TMA analysis of BAIAP2 revealed an opposite effect for BAIAP2 depending on its localization. Patients with higher levels of BAIAP2 in the cytoplasm showed significantly better survival than those with lower levels of cytoplasmic BAIAP2. Strikingly, when looking at the membrane levels of BAIAP2, an opposite trend was observed—increased expression of BAIAP2 in the membrane led to a significant decrease in patient survival, suggesting a role for BAIAP2 in signalling when localized in the membrane (Figure R-1 13A). Finally, we analysed the effect of the combination of nuclear MUC5AC and membrane BAIAP2 protein levels on survival (Figure R-1 13B). High expression of BAIAP2 in the membrane and high expression of nuclear MUC5AC showed the most significant differences in survival between patient groups.

Collectively, our results suggest that metastatic cells present changes in the spatial distribution of altered proteins mimicking actual changes in CRC tumoral samples, which were indeed associated with the prognosis of patients.

GLG1 analysis as blood-based candidate biomarkers for colorectal cancer diagnosis

Finally, we hypothesized that dysregulated proteins in the secretome could serve as plasma biomarkers of CRC. Thus, the levels of GLG1 in the plasma from 48 CRC patients at different stages and patients with

premalignant lesions and 32 healthy individuals as controls were tested by ELISA (Figure R-1 14). GLG1 plasma levels significantly discriminated

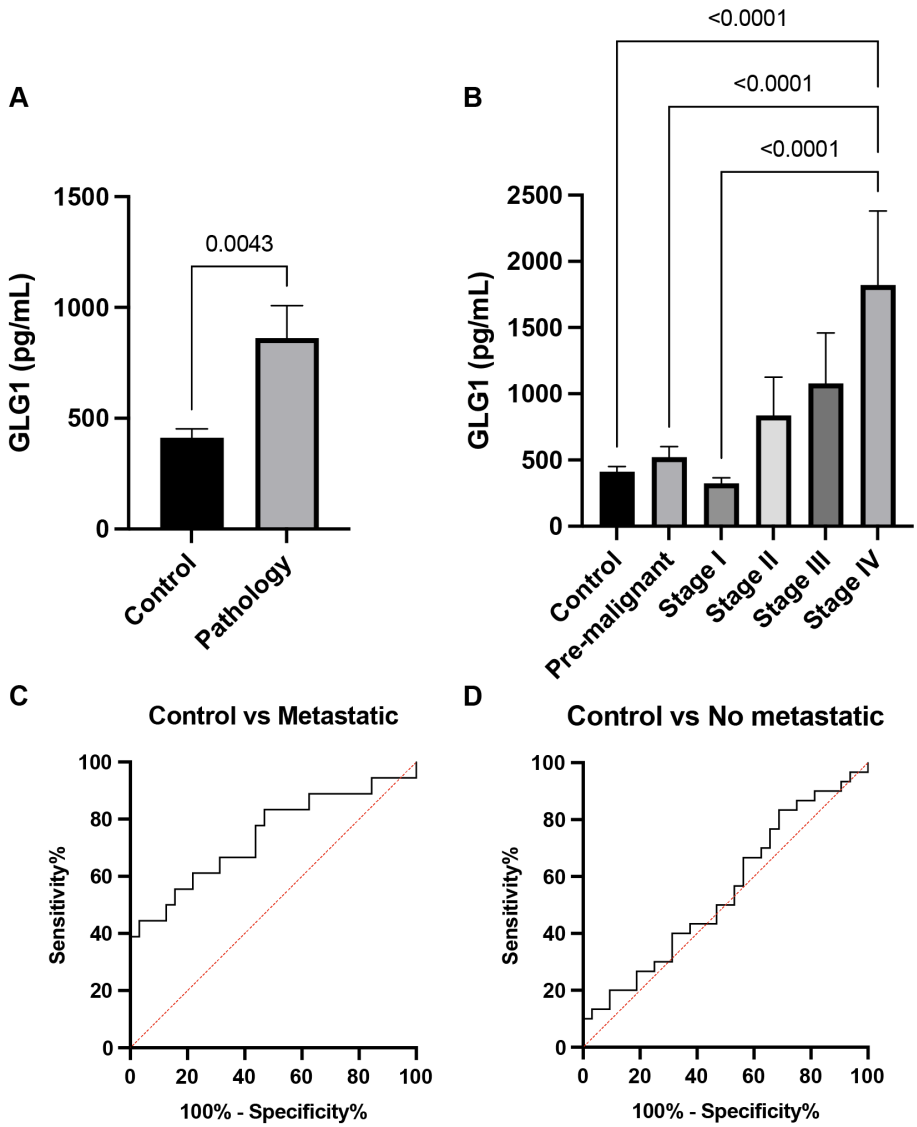


Figure R-1 14: Evaluation of the plasma biomarker potential of GLG1 in colorectal cancer. (**A**) Quantification of GLG1 using commercially available ELISAs in plasma from healthy individuals as controls, and premalignant individuals and CRC patients as the pathology group. (**B**) GLG1 levels in plasma from healthy individuals as controls, premalignant individuals and stage I, II, III, and IV CRC patients. (**C**, **D**) Determination of the GLG1 value as discriminating plasma biomarker was carried out through ROC curves calculating AUC, sensitivity and specificity values. (**C**) ROC curve analysis discriminating controls and CRC patients (stages III and IV). (**D**) ROC curve analysis discriminating controls and premalignant individuals and early-stage CRC patients (stages I and III).

patients from controls samples (mean ± SEM = 863.65 ± 146.21 pg/mL for the pathological group -premalignant individuals and CRC patients- versus 412.26 ± 39.82 pg/mL for controls; p-value = 0.0043) (Figure R-1 14A). Importantly, GLG1 plasma levels significantly increased from premalignant individuals to stage IV CRC patients, where the highest difference in GLG1 plasma levels were observed (1822.74 ± 556.89 pg/mL for stage IV CRC patients vs. 412.26 ± 39.82 pg/mL for controls, p-value = 0.00017) (Figure R-1 14B).

Finally, we surveyed the usefulness of the plasma measurement of GLG1 to discriminate CRC and premalignant individuals from controls calculating sensitivity, specificity, and AUC by means of ROC curves. The highest AUC value for GLG1 was observed for discriminating CRC stage IV from controls (p-value = 0.0009) with 90.63% and sensitivity and specificity of 85.71% and 78.13%, respectively. Moreover, GLG1 was observed useful for discriminating advance CRC (stages III and IV) from controls (Figure R-1 14C), with AUC of 74.13% and sensitivity and specificity of 66.67% and 68.75%, respectively. In contrast, GLG1 was not useful and non-significant for the discrimination of CRC (stages I and II) and premalignant individuals from controls (p-value = 0.9251) with AUC, sensitivity and specificity of 50.78%, 54.79%, and 50%, respectively (Figure R-1 14D).

Collectively, these results not only confirmed the predictive value as a biomarker in plasma of GLG1 for CRC patients at advanced stages, but they also agreed with its higher expression in metastatic cells (as a model of advanced CRC stage) as observed by proteomics in comparison to non-metastatic cells.

Discussion

The compartmentalization of eukaryotic cells and their distribution allows biological processes to occur synchronously requiring the specialization of multiple cellular functions. The localization of proteins to specific subcellular niches is usually a requirement to fulfil their functions. In addition, the capacity to dynamically transit between compartments is also essential for multiple cellular processes related to signalling, growth, proliferation, motility or programmed cell death. Accordingly, mislocalization of proteins has been implicated in various different pathological states, including cancer (39, 40). In this context, the exploration of the cell proteome related to metastasis in subcellular organelles or subcompartments is not only a practical approach but also a mandatory study. Proper interpretation of proteomic data requires information about compartmentalization of protein machinery to get further insights into cancer metastatic processes. Therefore, determining the subcellular location of proteins and how they change in metastasis would be essential for understanding the protein's altered biochemical functions associated with this process, comprised of invasion, and cell migration of tumoral cells from the primary tumour; intravasation and survival of tumoral cells in the circulation; extravasation, colonization, and proliferation at secondary tumour sites (6). Moreover, the identification of specific proteins associated with metastasis or to a specific organ of colonization is a requirement to identify specific proteins as prognostic markers. Altered proteins can serve as predictive biomarkers for specific metastasis and new therapeutic targets of intervention to reduce the mortality associated with metastasis.

The combination of traditional biochemical fractionation coupled to mass spectrometry-based identification has been the next step in the characterization of the proteome subcellular organization (40). Here, we used the most common isogenic cells in CRC as models of metastasis to

lymph nodes (SW480/SW620) and the KM12 cell system (KM12C, KM12SM and KM12L4a), which mimics the metastasis to liver (KM12SM) and the metastasis to liver and lung (KM12L4a). In previous proteomics studies using KM12 isogenic cell lines, many identified proteins were described as key molecules in CRC, as VEGFA, ERBB2, EGFR, MMP7, FGFR4, cadherin-17 (CDH17), or IL13Rα2 (13, 14, 29, 41). In addition, in SW480 and SW620 isogenic cell lines, several proteins were also found as interesting proteins in CRC, as ITGB3, CacyBP, TFF3, or GDF15 (42-44). These results highlight the necessity to continue characterizing these cell lines to increase our knowledge of CRC metastasis by quantitative proteomics analyses. In this sense, we utilized SILAC in a previous report to identify dysregulated proteins in abundance and localization between the poorly metastatic KM12C cells and the liver metastatic KM12SM CRC cells (29). Here, our purpose of the study consisted of providing the widest quantitative analysis of multidimensional proteome alterations in CRC metastatic cells by analysing the dysregulated proteome associated with lymph nodes, liver and lung tropisms, and their spatial distributions in the cytoplasm, membrane, nucleus, chromatin, and cytoskeletal fractions, and the secretome. This kind of high-throughput approach could provide an important way to elucidate the protein functions and to identify new roles for specific subcellular compartments. Here, this simultaneous proteomic study of different human cellular compartments provided us with novel and significant insights into protein function during differentiation and transformation of poorly metastatic cells into highly metastatic cells. Importantly, in contrast to many studies on organelle proteomics, we here provide not only a detailed list of the protein content of these organelles and the dysregulated proteins in CRC metastasis, but we also provide data validating the results of selected proteins by WB, IF, IHC, and ELISA. We also show for GLG1 a remarkable association of its protein levels in plasma to late CRC stages that could be used as a predictive marker of CRC aggressiveness.

The analysis of the pathways differing between metastatic and non-metastatic CRC cells revealed that several cellular canonical pathways were over-represented among the different subcellular organelles of study. For example, the most significantly represented functions in cytoplasm (CEB) proteins were those related to actin organization and regulation of wound healing. In the membrane (MEB) fraction, RNA processing was the most enriched pathway, whereas in the nucleus the endomembrane system organization and vesicle transport appeared as enriched and DNA repair, cellular response to DNA damage, and vascular processes as diminished pathways in metastasis. Regarding cytoskeletal (PEB) proteins, the most enriched pathways were those related to metabolism, chromatin binding and regulation of transcription, whereas for the secretome upregulated proteins were associated with extracellular matrix constituents and organization and encapsulating structure organization. All of these processes are hallmarks of cancer and have been largely associated with metastasis (5). These results and those related to the cellular component analysis pointed out to a correct subcellular fractionation of the cells.

In this work, we focused our attention on interesting proteins not previously associated with CRC or CRC metastasis and whose information in databases was scarce, such as BAIAP2, GLG1, PHYHIPL, TNFRSF10A, and CDKN2AIP, in contrast to other proteins, such as SCRIB, SNX9, CLDN3, or AHR, which have been previously associated with CRC and/or CRC metastasis (45-48).

Among them, BAIAP2 has been described as active mainly in neurons, diffused in the cytoplasm and localized in the membrane upon association to BAI1. Consequently, it has been proposed that it may have a potential role in lamellipodia and filopodia formation in motile cells and as a cell adhesion molecule inducing growth cone guidance in the process of neuronal growth (49). GLG1 has been postulated as an important

regulator of membrane trafficking and described to be instrumental for metastatic colonization in bone of BM2 myeloid cells and M1a cells (derivative from the SUM159 triple negative breast cancer cell line) by binding E-selectin (50). PHYHIPL, a paralog of the phytanoyl-CoA hydroxylase-interacting protein, is altered in glioblastoma multiforme (51), where its function remains unknown, and has been proposed as a therapeutic target gene (51). TNFRSF10A, the receptor for the cytotoxic ligand TNFSF10/TRAIL, can regulate apoptosis mediated by TRAIL; and inactivating mutations have been demonstrated to play a role in metastasis of breast cancer (52). CDKN2AIP was initially identified as the binding partner or ARF and several studies showed that CDKN2AIP amplification can enhance angiogenesis (53). In parallel, its expression was closely associated with higher expression of several markers involved in angiogenesis and metastasis in breast, skin, prostate, liver and kidney cancer (53).

One of the goals of the study was to generate validated data on the here identified dysregulated proteins involved in CRC metastasis to lymph nodes, liver, and lung. To this end, we validated the results by WB, IF, IHC, and ELISA on indicated proteins from the different compartments, with and without known association to CRC metastasis. Expression levels of the selected proteins analysed by WB were in agreement with the proteomics data. In addition, IF analysis allowed us to confirm that some of these proteins changed its localization and abundance between poorly metastatic and highly metastatic CRC cells. For example, AHR went from the cytoplasm to the nucleus, PHYHIPL from the cytoplasm to the membrane, or CLDN3 mainly from the cytoplasm to the membrane, especially in the KM12 cell system. Using actual CRC samples from patients, either by WB or IHC, it was observed that the dysregulation of GLG1, AHR, or BAIAP2 was associated with CRC (paired normal/tumoral samples) and to prognosis of CRC patients. In particular, the altered localization of BAIAP2 observed by IHC, from the cytoplasm to the

membrane, was associated with a worst overall survival of CRC patients. These results validated, at least partially, the proteomics dataset pointing out to a vast dysregulation of proteins in CRC metastasis involved in processes highly related to cancer as cell biogenesis, cell adhesion, cell development, actin cytoskeleton, gene expression, signalling pathways, vesicle-mediated transport, and metabolic processes (5, 54).

Finally, we demonstrated by ELISA the good performance of GLG1 as a plasma marker predictive of late stages of CRC. Our results highlight the usefulness of the multidimensional proteomics approach to identify dysregulated proteins in abundance and localization as novel markers in CRC. Therefore, these results support the initial premise of the study, encouraging us to perform subsequent functional analyses focused on the mechanism of action of selected dysregulated proteins to determine their relevance in the formation, progression and metastasis of CRC.

Altogether, these results demonstrate that combining or doing a whole cell analysis often dilutes protein changes altered in abundance in specific localizations that might only be observed when looking at specific subcellular fractions. Our results prove the usefulness of studying protein changes within subcellular fractions to identify dysregulated proteins associated with CRC metastasis. In addition, it was possible to identify proteins whose expression in one subcellular compartment was decreased, while increased in another. Indeed, we found that about 10% of the dysregulated proteins showed opposite protein abundance dysregulation in different subcellular organelles, which would suggest regulation or activation of specific pathways involved in CRC metastasis.

In summary, we provided a comprehensive proteomic analysis of CRC metastasis using isogenic CRC cell lines with different metastatic tropisms, identifying several proteins and pathways upregulated in CRC metastasis. The information gained from this study generated a large amount of data useful for determining proteins potentially involved in colon

epithelial cell differentiation, transformation, and metastatic processes. Our study goes one step beyond conventional studies by providing subcellular localization of proteins associated with metastasis of CRC encompassing lymph nodes, liver and lung metastasis, completing a previous study focused only on CRC liver metastasis (29). A further multidimensional study using actual CRC samples from patients should eventually allow for better classification and identification of dysregulated pathways. Such approach would ensure achieving better diagnostics and an increased ability to provide patients with the best treatments for personalized medicine. Finally, our findings provide validated novel dysregulated proteins not-previously associated with CRC and CRC metastasis. BAIAP2, GLG1, PHYHIPL, TNFRSF10A and CDKN2AIP, should be further explored in subsequent studies to determine their usefulness as advanced CRC stage biomarkers. They should also be the focus of functional experiments to determine their roles in CRC formation, progression, and metastasis, to potentially identify new therapeutic targets of the disease.

References

1. Rodriguez-Salas N, Dominguez G, Barderas R, Mendiola M, Garcia-Albeniz X, Maurel J, et al. Clinical relevance of colorectal cancer molecular subtypes. Crit Rev Oncol Hematol. 2017;109:9-19.
2. Holleczek B, Rossi S, Domenic A, Innos K, Minicozzi P, Francisci S, et al. On-going improvement and persistent differences in the survival for patients with colon and rectum cancer across Europe 1999-2007 - Results from the EUROCARE-5 study. Eur J Cancer. 2015;51(15):2158-68.
3. van der Geest LG, Lam-Boer J, Koopman M, Verhoef C, Elferink MA, de Wilt JH. Nationwide trends in incidence, treatment and survival of colorectal cancer patients with synchronous metastases. Clin Exp Metastasis. 2015;32(5):457-65.
4. Riihimaki M, Hemminki A, Sundquist J, Hemminki K. Patterns of metastasis in colon and rectal cancer. Sci Rep. 2016;6:29765.
5. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;144(5):646-74.
6. Hapach LA, Mosier JA, Wang W, Reinhart-King CA. Engineered models to parse apart the metastatic cascade. NPJ Precis Oncol. 2019;3:20.
7. Torres S, Bartolomé RA, Mendes M, Barderas R, Fernandez-Aceñero MJ, Peláez-García A, et al. Proteome Profiling of Cancer-Associated Fibroblasts Identifies Novel Proinflammatory Signatures and Prognostic Markers for Colorectal Cancer. Clinical Cancer Research. 2013;19(21):6006-19.
8. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19(11):1423-37.
9. Cai Z, Chiu JF, He QY. Application of proteomics in the study of tumor metastasis. Genomics Proteomics Bioinformatics. 2004;2(3):152-66.
10. Aleckovic M, Wei Y, LeRoy G, Sidoli S, Liu DD, Garcia BA, et al. Identification of Nidogen 1 as a lung metastasis protein through secretome analysis. Genes Dev. 2017;31(14):1439-55.
11. Song P, Bao H, Yu Y, Xue Y, Yun D, Zhang Y, et al. Comprehensive profiling of metastasis-related proteins in paired hepatocellular carcinoma cells with different metastasis potentials. Proteomics Clin Appl. 2009;3(7):841-52.

12. Sun B, Zhang S, Zhang D, Li Y, Zhao X, Luo Y, et al. Identification of metastasis-related proteins and their clinical relevance to triple-negative human breast cancer. Clin Cancer Res. 2008;14(21):7050-9.
13. Morikawa K, Walker SM, Jessup JM, Fidler IJ. In vivo selection of highly metastatic cells from surgical specimens of different primary human colon carcinomas implanted into nude mice. Cancer Research. 1988;48(7):1943-8.
14. Morikawa K, Walker SM, Nakajima M, Pathak S, Jessup JM, Fidler IJ. Influence of organ environment on the growth, selection, and metastasis of human colon carcinoma cells in nude mice. Cancer Research. 1988;48(23):6863-71.
15. Li A, Varney ML, Singh RK. Constitutive expression of growth regulated oncogene (gro) in human colon carcinoma cells with different metastatic potential and its role in regulating their metastatic phenotype. Clin Exp Metastasis. 2004;21(7):571-9.
16. Kuniyasu H, Ohmori H, Sasaki T, Sasahira T, Yoshida K, Kitadai Y, et al. Production of interleukin 15 by human colon cancer cells is associated with induction of mucosal hyperplasia, angiogenesis, and metastasis. Clin Cancer Res. 2003;9(13):4802-10.
17. Calon A, Espinet E, Palomo-Ponce S, Tauriello DV, Iglesias M, Cespedes MV, et al. Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. Cancer Cell. 2012;22(5):571-84.
18. Barderas R, Bartolome RA, Fernandez-Acenero MJ, Torres S, Casal JI. High expression of IL-13 receptor alpha2 in colorectal cancer is associated with invasion, liver metastasis, and poor prognosis. Cancer Res. 2012;72(11):2780-90.
19. Hegde P, Qi R, Gaspard R, Abernathy K, Dharap S, Earle-Hughes J, et al. Identification of Tumor Markers in Models of Human Colorectal Cancer Using a 19,200-Element Complementary DNA Microarray1. Cancer Research. 2001;61(21):7792-7.
20. Capes-Davis A. Cell Line Detective Work: Basic Principles and Molecular Applications. Advances in Molecular Pathology. 2018;1(1):229-38.
21. Ribas M, Masramon L, Aiza G, Capellà G, Miró R, Peinado MA. The structural nature of chromosomal instability in colon cancer cells. The FASEB Journal. 2003;17(2):289-91.
22. Leibovitz A, Stinson JC, McCombs WB, McCoy CE, Mazur KC, Mabry ND. Classification of Human Colorectal Adenocarcinoma Cell Lines. Cancer Research. 1976;36(12):4562-9.

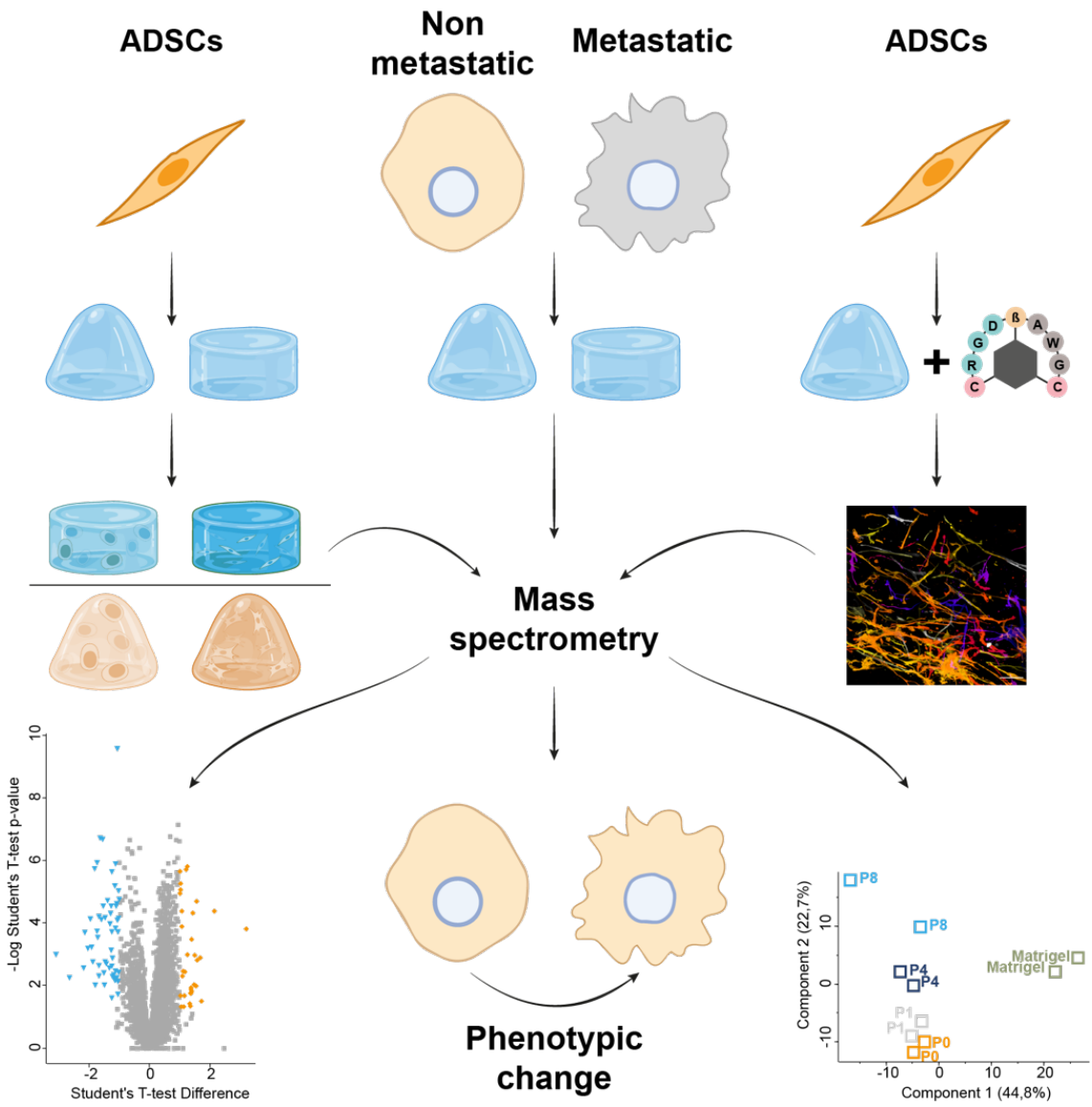
23. Hewitt RE, McMarlin A, Kleiner D, Wersto R, Martin P, Tsokos M, et al. Validation of a model of colon cancer progression. The Journal of pathology. 2000;192(4):446-54.
24. Fior R, Póvoa V, Mendes RV, Carvalho T, Gomes A, Figueiredo N, et al. Single-cell functional and chemosensitive profiling of combinatorial colorectal therapy in zebrafish xenografts. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(39):E8234-E43.
25. Lundberg E, Borner GHH. Spatial proteomics: a powerful discovery tool for cell biology. Nat Rev Mol Cell Biol. 2019;20(5):285-302.
26. Bauer NC, Doetsch PW, Corbett AH. Mechanisms Regulating Protein Localization. Traffic. 2015;16(10):1039-61.
27. Wang X, Li S. Protein mislocalization: mechanisms, functions and clinical applications in cancer. Biochim Biophys Acta. 2014;1846(1):13-25.
28. Banworth MJ, Li G. Consequences of Rab GTPase dysfunction in genetic or acquired human diseases. Small GTPases. 2018;9(1-2):158-81.
29. Mendes M, Peláez-García A, López-Lucendo M, Bartolomé RA, Calviño E, Barderas R, et al. Mapping the Spatial Proteome of Metastatic Cells in Colorectal Cancer. Proteomics. 2017;17(19):1700094-.
30. Wang Z, Karkossa I, Großkopf H, Rolle-Kampczyk U, Hackermüller J, von Bergen M, et al. Comparison of quantitation methods in proteomics to define relevant toxicological information on AhR activation of HepG2 cells by BaP. Toxicology. 2021;448:152652.
31. Stepath M, Zülch B, Maghnouj A, Schork K, Turewicz M, Eisenacher M, et al. Systematic Comparison of Label-Free, SILAC, and TMT Techniques to Study Early Adaption toward Inhibition of EGFR Signaling in the Colorectal Cancer Cell Line DiFi. Journal of Proteome Research. 2020;19(2):926-37.
32. O'Connell JD, Paulo JA, O'Brien JJ, Gygi SP. Proteome-Wide Evaluation of Two Common Protein Quantification Methods. Journal of Proteome Research. 2018;17(5):1934-42.
33. Collier TS, Sarkar P, Franck WL, Rao BM, Dean RA, Muddiman DC. Direct Comparison of Stable Isotope Labeling by Amino Acids in Cell Culture and Spectral Counting for Quantitative Proteomics. Analytical Chemistry. 2010;82(20):8696-702.
34. Muñoz-San Martín C, Pérez-Ginés V, Torrente-Rodríguez RM, Gamella M, Solís-Fernández G, Montero-Calle A, et al. Electrochemical

immunosensing of Growth arrest-specific 6 in human plasma and tumor cell secretomes. Electrochem Sci Adv. 2021:e2100096.
35. Betge J, Schneider NI, Harbaum L, Pollheimer MJ, Lindtner RA, Kornprat P, et al. MUC1, MUC2, MUC5AC, and MUC6 in colorectal cancer: expression profiles and clinical significance. Virchows Arch. 2016;469(3):255-65.
36. Lakshmanan I, Rachagani S, Hauke R, Krishn SR, Paknikar S, Seshacharyulu P, et al. MUC5AC interactions with integrin beta4 enhances the migration of lung cancer cells through FAK signaling. Oncogene. 2016;35(31):4112-21.
37. Zhang CT, He KC, Pan F, Li Y, Wu J. Prognostic value of Muc5AC in gastric cancer: A meta-analysis. World J Gastroenterol. 2015;21(36):10453-60.
38. Pothuraju R, Rachagani S, Krishn SR, Chaudhary S, Nimmakayala RK, Siddiqui JA, et al. Molecular implications of MUC5AC-CD44 axis in colorectal cancer progression and chemoresistance. Mol Cancer. 2020;19(1):37.
39. Neel DS, Allegakoen DV, Olivas V, Mayekar MK, Hemmati G, Chatterjee N, et al. Differential Subcellular Localization Regulates Oncogenic Signaling by ROS1 Kinase Fusion Proteins. Cancer Res. 2019;79(3):546-56.
40. Christopher JA, Stadler C, Martin CE, Morgenstern M, Pan Y, Betsinger CN, et al. Subcellular proteomics. Nat Rev Methods Primers. 2021;1.
41. Peláez-García A, Barderas R, Torres S, Hernández-Varas P, Teixidó J, Bonilla F, et al. FGFR4 Role in Epithelial-Mesenchymal Transition and Its Therapeutic Value in Colorectal Cancer. PLOS ONE. 2013;8(5):e63695-e.
42. Xue H, Lu B, Zhang J, Wu M, Huang Q, Wu Q, et al. Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. J Proteome Res. 2010;9(1):545-55.
43. Ghosh D, Yu H, Tan XF, Lim TK, Zubaidah RM, Tan HT, et al. Identification of key players for colorectal cancer metastasis by iTRAQ quantitative proteomics profiling of isogenic SW480 and SW620 cell lines. J Proteome Res. 2011;10(10):4373-87.
44. Lei Y, Huang K, Gao C, Lau QC, Pan H, Xie K, et al. Proteomics identification of ITGB3 as a key regulator in reactive oxygen species-induced migration and invasion of colorectal cancer cells. Mol Cell Proteomics. 2011;10(10):M110 005397.

45. Shen H, Huang C, Wu J, Li J, Hu T, Wang Z, et al. SCRIB Promotes Proliferation and Metastasis by Targeting Hippo/YAP Signalling in Colorectal Cancer. Front Cell Dev Biol. 2021;9:656359.
46. Tanigawa K, Maekawa M, Kiyoi T, Nakayama J, Kitazawa R, Kitazawa S, et al. SNX9 determines the surface levels of integrin beta1 in vascular endothelial cells: Implication in poor prognosis of human colorectal cancers overexpressing SNX9. J Cell Physiol. 2019;234(10):17280-94.
47. Cherradi S, Martineau P, Gongora C, Del Rio M. Claudin gene expression profiles and clinical value in colorectal tumors classified according to their molecular subtype. Cancer Manag Res. 2019;11:1337-48.
48. Xie G, Raufman JP. Role of the Aryl Hydrocarbon Receptor in Colon Neoplasia. Cancers (Basel). 2015;7(3):1436-46.
49. Oda K, Shiratsuchi T, Nishimori H, Inazawa J, Yoshikawa H, Taketani Y, et al. Identification of BAIAP2 (BAI-associated protein 2), a novel human homologue of hamster IRSp53, whose SH3 domain interacts with the cytoplasmic domain of BAI1. Cytogenet Cell Genet. 1999;84(1-2):75-82.
50. Esposito M, Mondal N, Greco TM, Wei Y, Spadazzi C, Lin SC, et al. Bone vascular niche E-selectin induces mesenchymal-epithelial transition and Wnt activation in cancer cells to promote bone metastasis. Nat Cell Biol. 2019;21(5):627-39.
51. Fu H, Ge B, Chen D, Wu Y, Luo Q, Li X, et al. Phytanoyl-CoA 2-Hydroxylase-Interacting Protein-Like Gene Is a Therapeutic Target Gene for Glioblastoma Multiforme. Med Sci Monit. 2019;25:2583-90.
52. Shin MS, Kim HS, Lee SH, Park WS, Kim SY, Park JY, et al. Mutations of tumor necrosis factor-related apoptosis-inducing ligand receptor 1 (TRAIL-R1) and receptor 2 (TRAIL-R2) genes in metastatic breast cancers. Cancer Res. 2001;61(13):4942-6.
53. Kalra RS, Chaudhary A, Yoon AR, Bhargava P, Omar A, Garg S, et al. CARF enrichment promotes epithelial-mesenchymal transition via Wnt/beta-catenin signaling: its clinical relevance and potential as a therapeutic target. Oncogenesis. 2018;7(5):39.
54. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell. 2000;100(1):57-70.

CHAPTER 2:

PROTEOMICS AND PIC HYDROGELS AS TOOLS TO UNDERSTAND 3D CELL DIFFERENTIATION AND ITS ROLE IN CANCER PROGRESSION



Introduction

Tissue microenvironment and extracellular matrix architecture have emerged in recent years as paramount factors involved in cell differentiation and behaviour (1-5). The importance of ECM composition and stiffness in controlling cell fate has been now vastly proven (4-9). What was once considered a mere physical support has been shown to be able to trigger drastic changes in cells. Integrins and cadherins are two of the best-studied classes of adhesion receptors. Integrins mediate the adhesion between the cell and ECM, whereas cadherins mediate homotypic adhesion between cells (10).

At the chemical level the most prevalent signals are those related to integrins, the main ECM receptor. Integrins are transmembrane receptors formed by the dimerization of two different integrin monomers (11). An alpha and a beta subunit dimerize to form a functional integrin receptor. There are 18 α-subunits and 8 β-subunits that combine to form, at least, 24 different heterodimers, each of them with different ligand specificities (11). Integrins are not only able to bind to components of the ECM but also to sense the mechanical properties of the matrix (12). Equally important to chemical binding, ECM stiffness is crucial for guiding stem cell commitment to specific phenotypes (3, 8, 9, 13). Already in 2002, Flanagan et al. demonstrated that the elastic properties of the substrate can alter the branching and neurite formation of neurons (14). Growing neurons on softer substrates, more similar to brain tissue, led to the development of a higher number of branches (14). Furthermore, Engler et al. could demonstrate that after the initial weeks in culture, matrix elasticity was the main determinant for mesenchymal stem cell differentiation (15). Nonetheless, besides the undisputed importance of the ECM and its properties not only for cell differentiation but also in the context of cancer onset and progression, many studies still overlook the role the ECM has.

In this sense, in a recent proteomic study that we carried out comparing paired tumoral and non-tumoral tissue samples from patients, we still found a large number of proteins altered that we had not identified in our spatial analysis of CRC metastasis (Figure R-2 1). Indeed, around 50% of all the proteins (212) we found to be upregulated in adenoma and adenocarcinoma (AD/AC) tissues were not present in any of the three different metastases (liver, liver and lung and lymph nodes) that we had studied (Figure R-2 1). Thus, we hypothesized that this difference could be due to the oversimplification that 2D culture represents compared to the physiological situation.

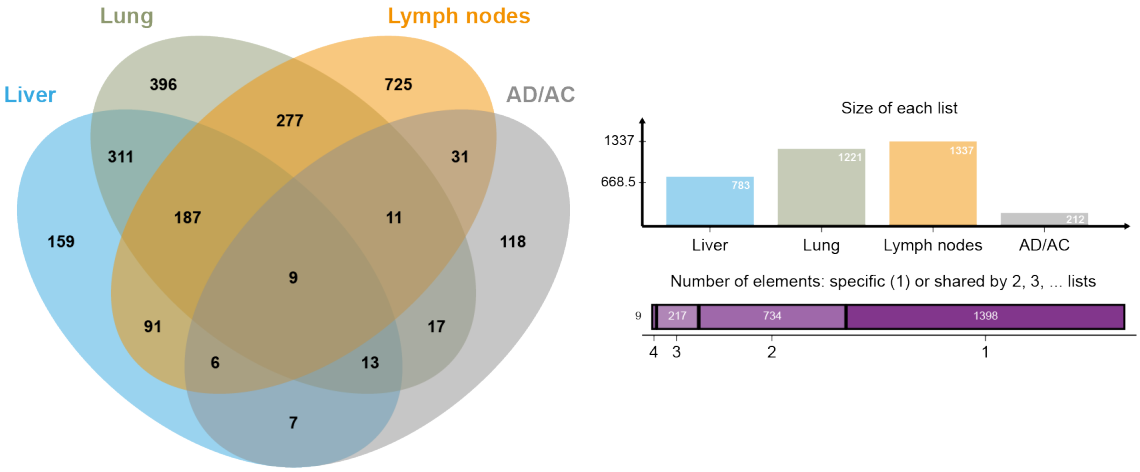


Figure R-2 1: Venn diagram displaying data of the proteins overexpressed in the different metastatic models we have analysed (KM12C, KM12SM, KM12L4a, SW480 and SW620 cells) and samples from adenoma and adenocarcinoma samples.

One of the main hindrances of studying 3D differentiation and ECM lies in finding the adequate model. ECM alternatives for cell culture can be divided onto two categories, cell-derived matrices, such as collagen or Matrigel, and synthetic matrices. Cell-derived matrices are difficult to modify chemically to introduce ligands for specific receptors (16, 17). In addition, Matrigel in particular can be especially problematic and introduce additional experimental uncertainties due to differences between brands and production batches (18). On the other hand,

synthetic matrices are interesting alternatives as they can be easily modified and show little variance between batches and manufacturers (19). However, they fail to fully replicate the mechanical properties of natural polymers. Among the synthetic hydrogels available, polyisocyanopeptide (PIC) based hydrogels represent an outstanding alternative due to their unique biomimetic properties (20). PIC hydrogels are highly tuneable and show a mechanical property known as strain stiffening, characteristic of natural matrices as collagen or Matrigel (20, 21). Strain stiffening is a non-linear mechanical behaviour characterised by an increase in the stiffness of the matrix as a response to an increment in the strain applied to it (21). In other words, Matrigel and other cell-derived matrices become stiffer as the force applied to them increases. This prevents large deformations that could rupture the tissue and also helps in long distance transmission of cell-cell communication (21, 22).

As a demonstration of the usability and potential of PIC based hydrogels, Liu et al. demonstrated that the secretome of adipose derived stem cells (ADSCs) could be modified by the mechanical properties of PIC polymers and the presence of RGD (23). RGD is a tripeptide formed arginine (R), glutamine (G) and aspartic acid (D) and is one of the principal ligands recognized by integrins (24, 25). RGD was proposed as an integrin binding peptide back in 1984 by Pierschbacher and Ruoslathi (26). They generated a series of fragments from a varying size derived from fibronectin sequence and evaluated which of those peptides could induce cell adhesion to sepharose beads (26). RGD was the smallest peptide able to induce cell binding and has since then been extensively used in synthetic matrices to stimulate cell adhesion (26). Liu et al. found that ADSCs showed a clear morphological change in gels containing RGD, being more pronounced in softer gels (23). Furthermore, proliferation and cell viability were also enhanced in gels containing RGD compared with uncoated PIC polymer (23). Luminex analysis of the secretome of the

cells cultured in the different matrices revealed changes in the levels of several cytokines, including VEGF or eotaxin whose levels were remarkably higher in the secretome of RGD cells (23).

In this context, we decided to use this platform as a first step to test whether our proteomic analysis pipelines would be suitable to also investigate 3D cultured cells. Imaging of 3D encapsulated KM12C and KM12SM cells had not shown striking differences in cell morphology compared to 2D cultures (Figure R-2 2). Consequently, if the proteomic analysis of 3D cultured CRC cells did not show any significant differences in protein expression, we would not be able to know if it was due to a problem in our analysis or culture methods or an actual result. Protein extraction from cells growing in 3D is not as straightforward as it is for cells grown in 2D. Specially for cells encapsulated in Matrigel, any remnants of Matrigel that would be left behind would contaminate and bias the proteomic analysis. For ADSCs, brightfield and fluorescence imaging together with Luminex secretome analysis had already demonstrated that the ECM was able to induce drastic changes in cell morphology and

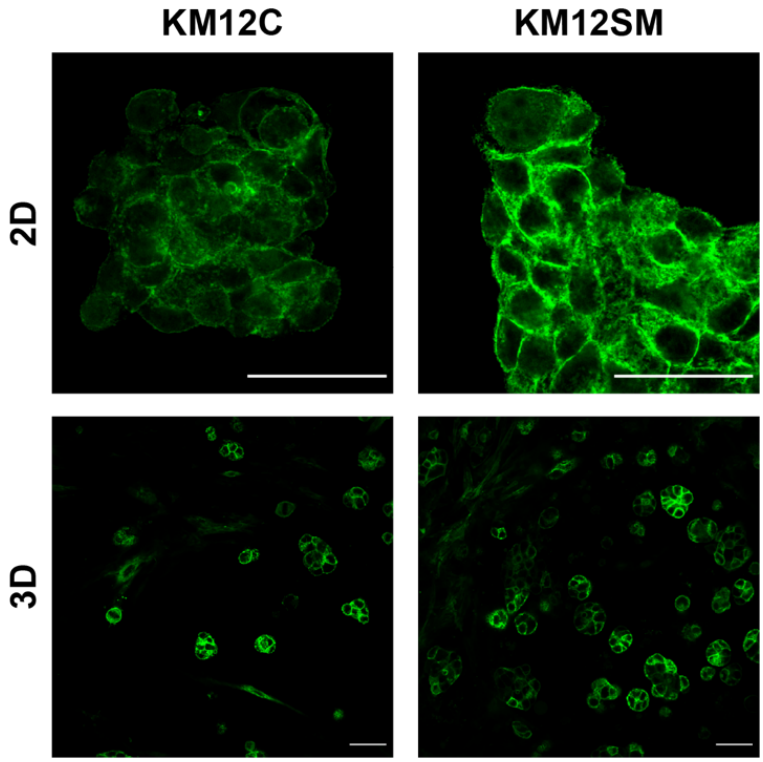


Figure R-2 2: Fluorescence images of phalloidin stained KM12C and KM12SM cells cultured in 2D and encapsulated in Matrigel (3D). No clear differences in cell morphology between 2D and encapsulated cells could be observed. Scale bar: 50 μm

secretion. Thus, it seemed reasonable to assume that if we did not detect any difference in protein expression using our proteomic analysis pipeline it was likely that the problem was within the analysis itself.

Once we tested our proteomic pipeline and applied it to a simpler model, as is the case of ADSCs, we decided to continue. Subsequently, we performed the proteomic characterization of the KM12 model in different 3D conditions, to try to find new proteomic markers that had been overlooked so far.

Results

3D culture alters protein expression in human adipose derived stem cells

To evaluate the importance of 3D culture in the CRC metastatic model we have studied so far, we sought to answer the question of whether we would be able to apply proteomics to a 3D differentiation model. From the work of the group of Prof. Dr. Kouwer that we discussed in the introduction, we knew that the different 3D matrices used had an effect in the secretome, morphology and viability of ADSCs. The authors found that all 3D conditions had an increased production of IL-10, compared to bidimensional cultured cells, and that this increase was more marked in inert gels (PIC-N3 gels). Furthermore, the secretome of these cells, that contained IL-10, induced an increase in wound closure speed of fibroblasts. Such increment was prevented by using an IL-10 neutralizing antibody. However, IL-10 supplementation of the culture media used in the wound healing assay did not induce a significant change in wound closure speed, indicating a synergistic effect between IL-10 and some other mediator(s) present in the conditioned media.

Then, we decided to evaluate whether the morphological changes of the cells would also translate into changes detectable using proteomic approaches. Subsequently, we grew ADSCs in the same conditions they were grown for the analysis of the morphology and protein secretion and after 7 days in culture collected them by removing the ECM. Whole protein extracts from the different conditions were then labelled using a TMT 10-plex kit. Upon labelling, samples were fractionated based on peptide hydrophobicity into 12 fractions prior to LC-MS/MS analysis onto a Q Exactive mass spectrometer.

Mass spectrometry data were analysed and a total of 3673 proteins were identified (Table S9). Principal component analysis (PCA) of the different

conditions showed the clear co-localization of replicates. Furthermore, RGD samples were grouped closer together similar to how N3 samples were, and all 3D conditions were clearly separated from 2D cultured cells (Figure R-2 3).

Bioinformatic analysis of proteins upregulated in 2D and 3D pooled conditions (soft or stiff and N3 or RGD PIC) (Figure R-2 4) revealed the upregulation in 2D of molecular pathways related to extracellular matrix organization and the formation of elastic fibres. In parallel, we found that the pathways upregulated in 3D cultured cells belonged to Vpr and HIV induced signalling cascades, sumoylation and collagen chain trimerization. Looking in detail at the proteins upregulated in either N3 or RGD coated PIC we found that 9 proteins were commonly upregulated in both conditions and only two more were specific of RGD grown cells (from a total of 13 proteins upregulated in RGD compared with 2D) (Figure R-2 5A). Surprisingly, when we evaluated the protein expression of N3 cells

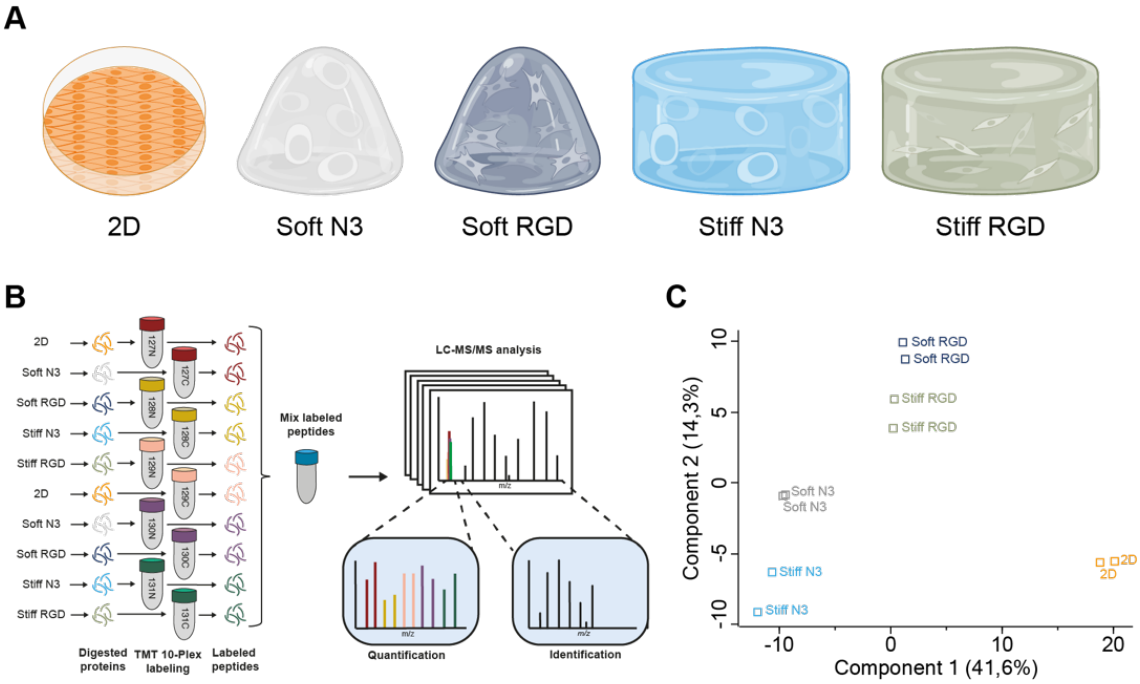


Figure R-2 3: ADSCs quantitative proteomic analysis. (**A**) ADSCs culture conditions analysed in the proteomic study. (**B**) Labelling scheme of the ADSC protein extracts for the analysis. (**C**) Principal component analysis of the protein extracts of ADSC, values represent the different duplicates.

we found 48 proteins that were upregulated specifically compared to 2D grown ADSCs (Figure R-2 5B). Reactome analysis of the proteins upregulated in N3 cultured cells indicated that most of those proteins belonged to pathways involved in cellular responses to stress and stimuli. In addition, some of the proteins upregulated were part of sumoylation pathways, and IL-12 response pathways. The latter would be a possible explanation of the increase observed by Liu et al. in the secretion of IL-10 (23). IL-12 has been shown to induce the secretion of IL-10 by immune cells as a feedback mechanism to control proinflammatory responses triggered by IL-12 (27).

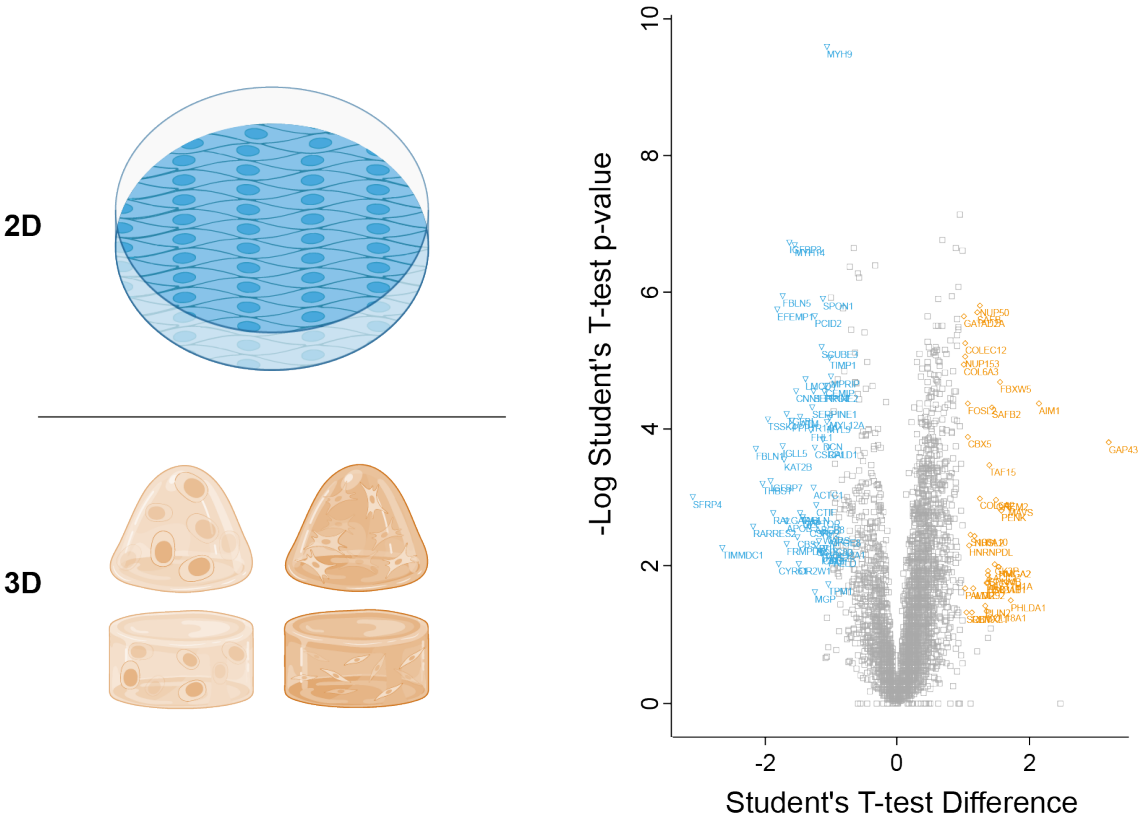


Figure R-2 4: Vulcano plot of the proteins significantly overexpressed in 2D (blue) and 3D (orange) conditions. The 3D conditions grouped for the analysis are depicted in the left bottom panel.

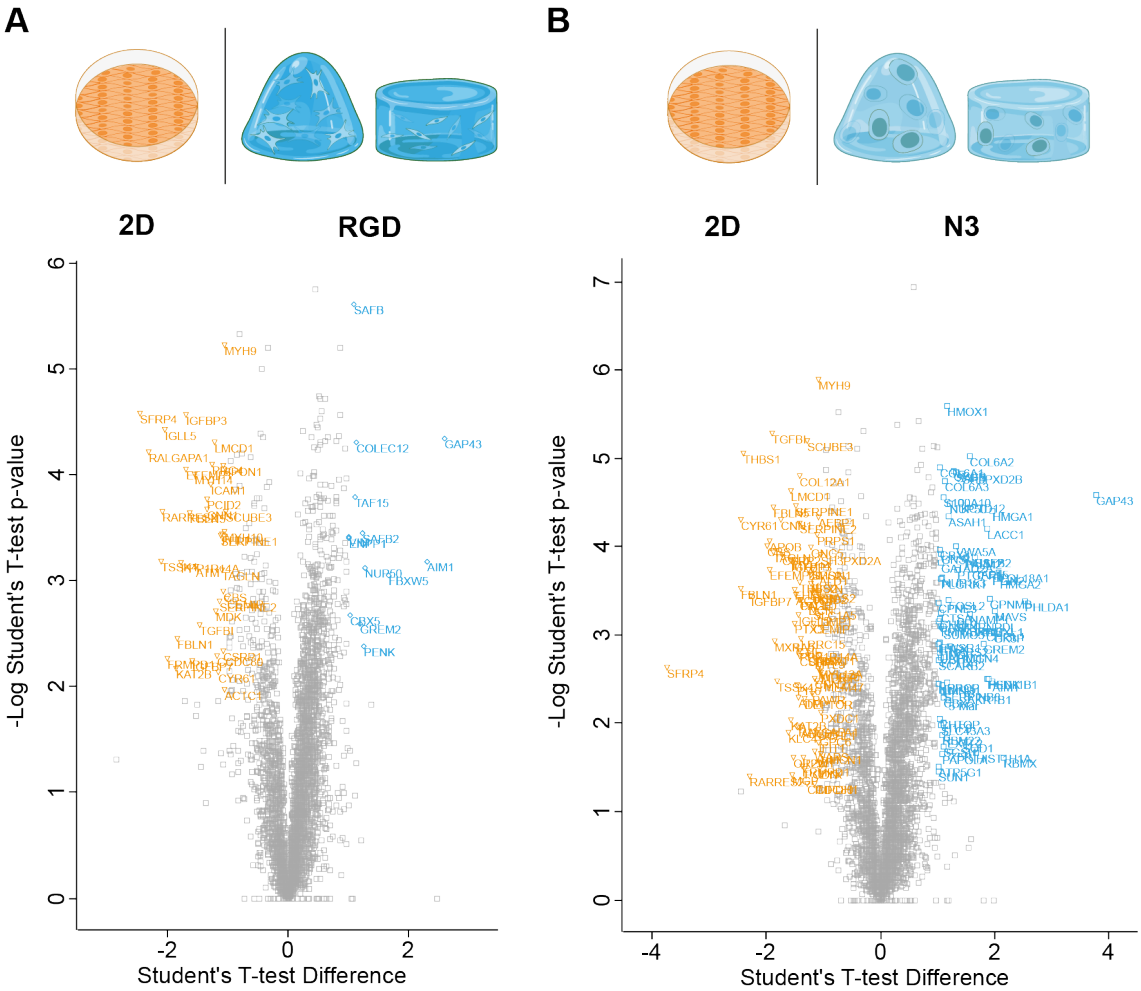


Figure R-2 5: Comparison of RGD and undecorated PIC with 2D cultured cells. (**A**) Vulcano plot of proteins overexpressed by cells encapsulated in RGD decorated PIC (blue) compared with 2D cultured cells (orange). (**B**) Vulcano plot of proteins overexpressed by cells encapsulated in undecorated PIC (blue) compared with 2D cultured cells (orange).

To consider the effect that matrix stiffness would have on protein expression we compared the proteome of all the soft gel conditions with the stiff ones (Figure R-2 6). Surprisingly, we could only find 2 proteins overexpressed in stiffer gels and no proteins overexpressed in soft gels. To separate the potential influence that the presence or absence of RGD might had, we decided to divide the analysis by equivalent conditions, without considering together N3 and RGD gels (Figure R-2 7). In Figure R-2 7A it can be clearly observed that in the presence of RGD there are

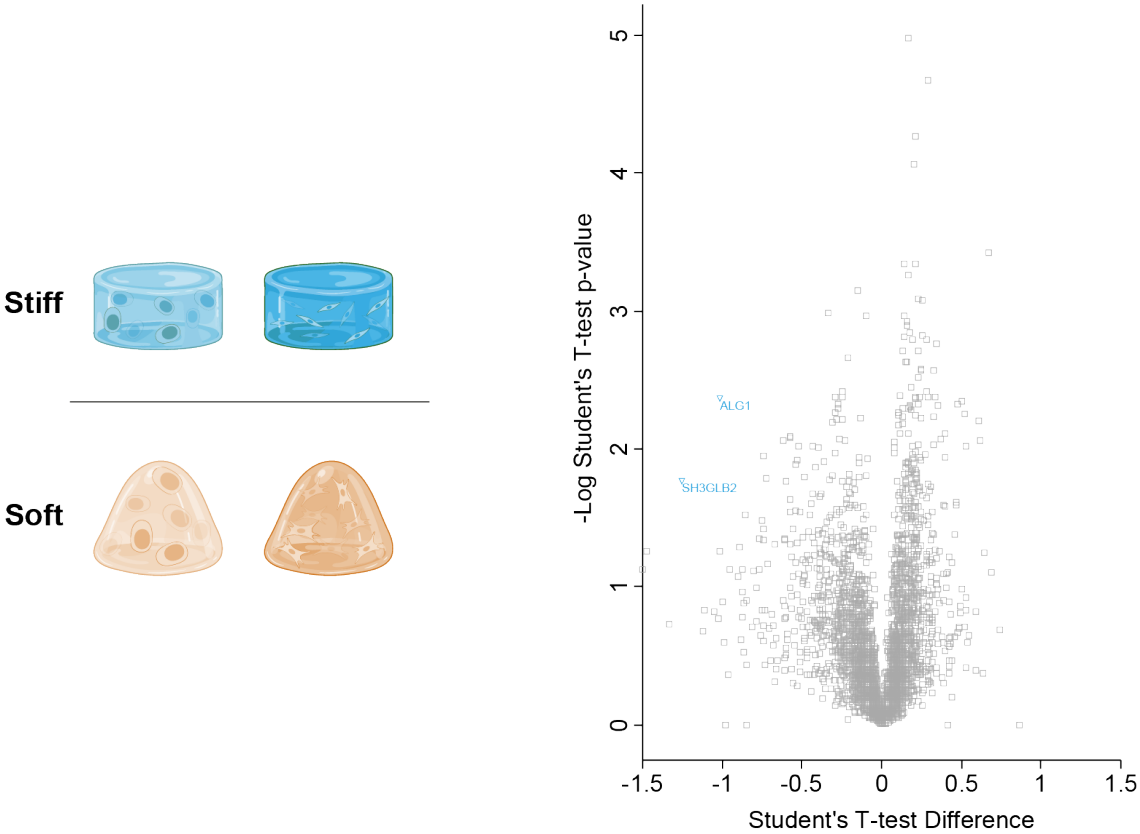


Figure R-2 6: Vulcano plot of proteins overexpressed in soft (orange) compared with stiff (blue) PIC polymer (with and without RGD). No proteins were significantly overexpressed in soft gels.

almost no differences between cells cultured in soft or stiff gels. On the other hand, when there is no RGD present, a clear overexpression pattern can be seen in cells cultured in stiffer gels (Figure R-2 7B). Interestingly, there were no proteins upregulated in cells grown in soft gels, while in stiff gels we could find 29 proteins overexpressed. Reactome and string analysis of these 29 proteins revealed that they belong to the regulation and induction of cell senescence.

Finally, from the previous comparison between soft and stiff gels, it became clear the tremendous impact that the presence of RGD could have. Thus, we decided to compare the expression pattern of matrices with and without RGD (Figure R-2 8). We could only find 4 proteins overexpressed in RGD grown cells while 16 proteins were overexpressed in N3 PIC cultured cells. The 4 proteins from RGD conditions did not

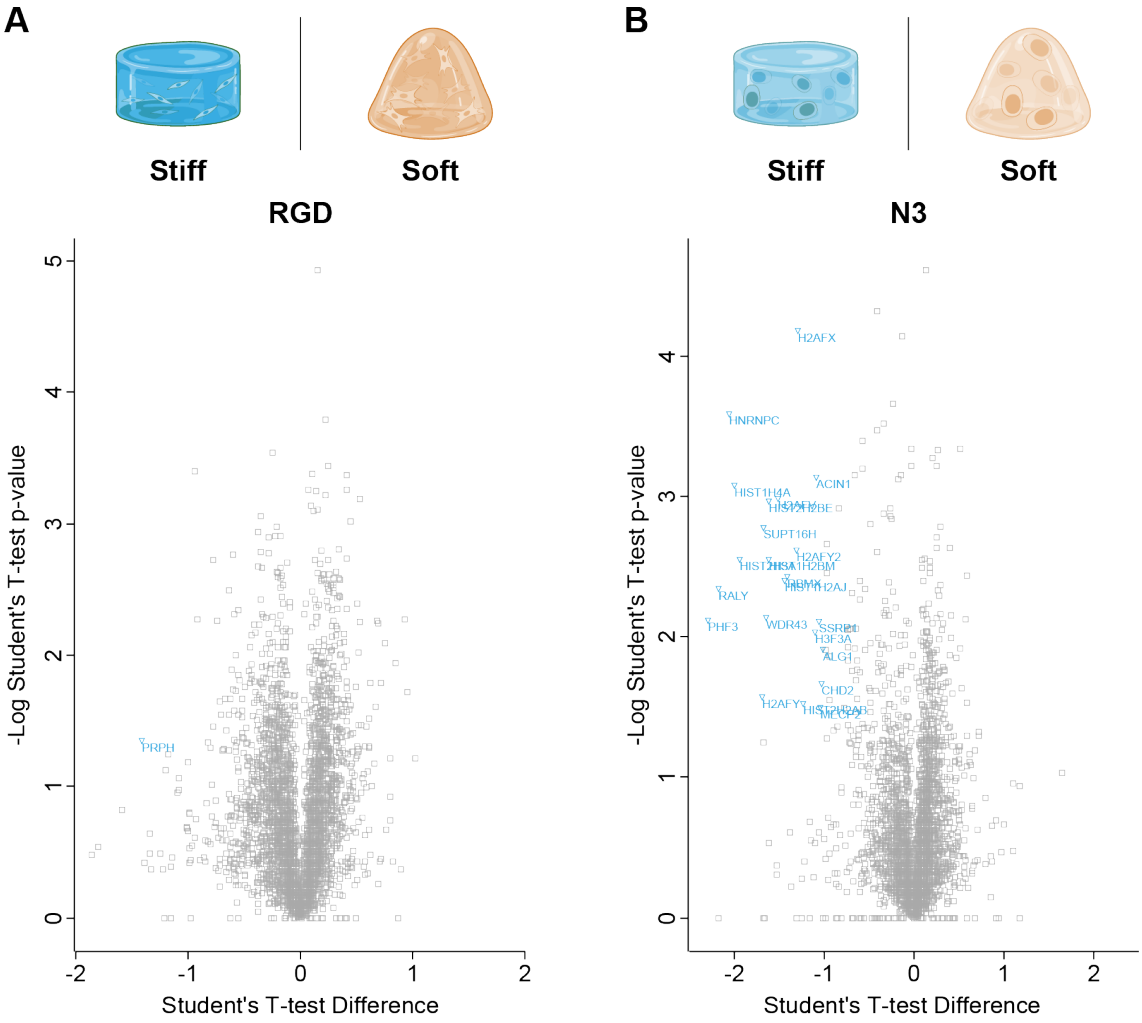


Figure R-2 7: Comparison of protein expression between soft and stiff PIC polymers. (**A**) Vulcano plot of proteins overexpressed in soft (orange) and stiff (blue) RGD decorated PIC polymer. (**B**) Vulcano plot of proteins overexpressed in soft (orange) and stiff (blue) undecorated PIC polymer. No proteins were significantly overexpressed in any of the soft gel conditions.

belong to any overrepresented signalling pathway. However, N3 overexpressed proteins were part of cell senescence pathways, which would agree with what we found for cells grown in stiff gels. Lastly, we evaluated the individual conditions with and without RGD to separate the effect of matrix stiffness from the presence of RGD (Figure R-2 9). In this case we observed more proteins to be upregulated in both soft (22) and stiff gels (29) when there was no RGD. On the other hand, only a small number of proteins were upregulated in RGD decorated soft (8) and stiff

(4) gels. In terms of the pathways upregulated, only in the case of stiff gels without RGD we could find proteins belonging to cell senescence pathways to be upregulated. This would be in agreement with the previous comparisons in which we had already observed that stiffer gels led to the activation of cell senescence, a process that seems to be prevented by the presence of RGD in the ECM.

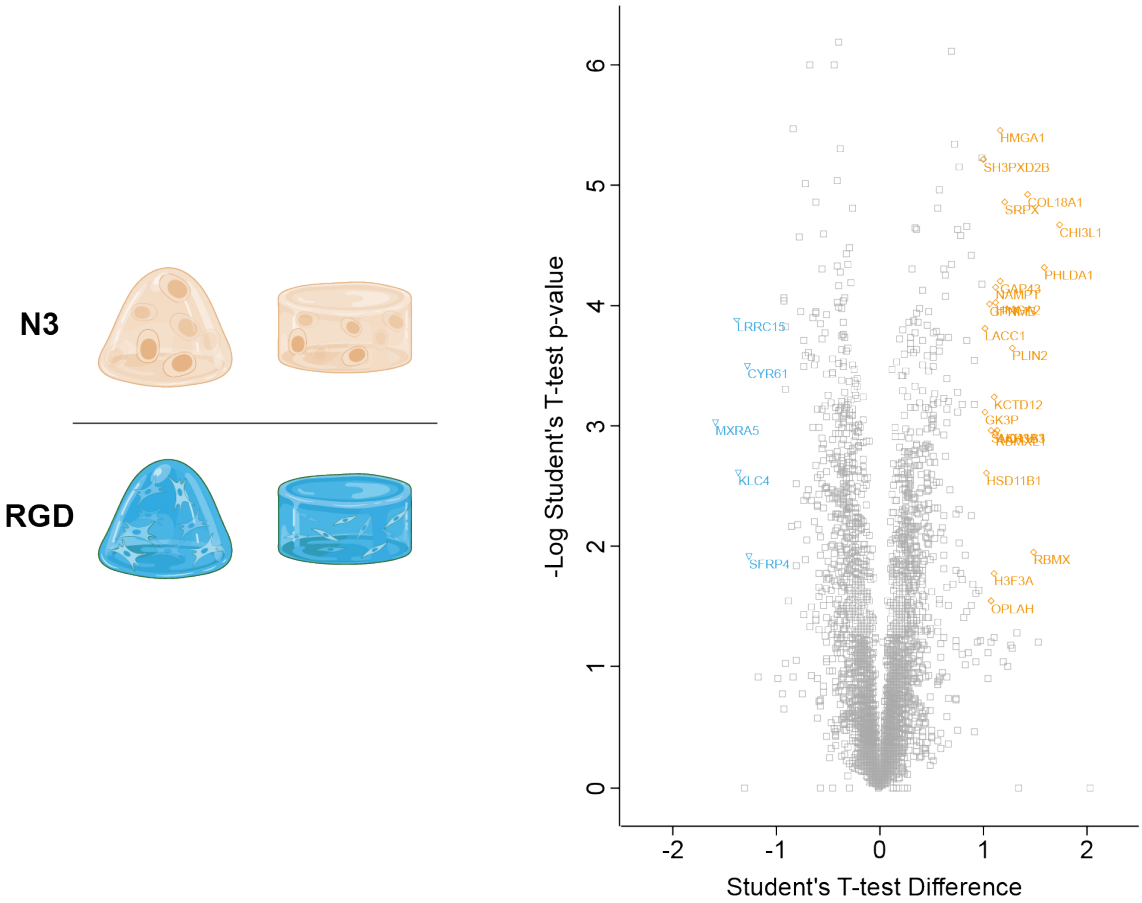


Figure R-2 8: Vulcano plot of proteins overexpressed in cells encapsulated in RGD decorated PIC (blue) compared with undecorated PIC polymer (orange) (stiff and soft polymers).

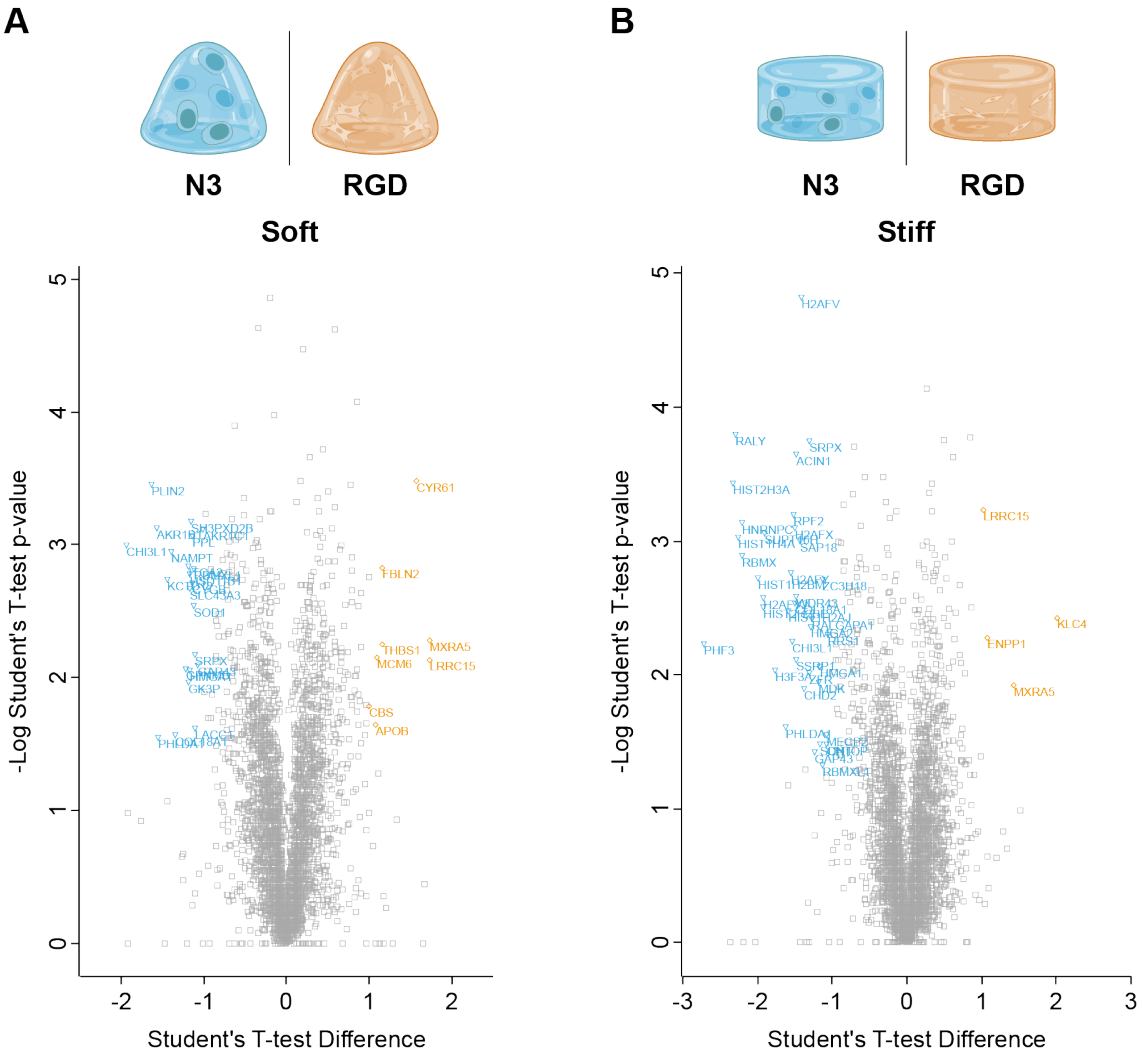


Figure R-2 9: Comparison of protein expression between RGD and undecorated polymers. (**A**) Vulcano plot of proteins overexpressed in cells encapsulated in RGD (orange) and N3 (blue) soft PIC polymer. (**B**) Vulcano plot of proteins overexpressed in stiff RGD decorated PIC (orange) and stiff undecorated PIC (blue).

Bicyclic peptides induce drastic changes in human adipose derived stem cells' morphology

After the characterization of ADSCs differentiation in different 3D matrices we observed that RGD presence could outperform the effect of matrix stiffness as the differences between soft and stiff gels were lost when RGD was present. Consequently, we considered the possibility of evaluating other integrin ligands that might be recognized by other integrin dimers. From the 24 integrin heterodimers described so far, eight of them (αvβ1, αvβ3, αvβ5, αvβ6, αvβ8, α5β1, α8β1, and αIIbβ3) bind the RGD tripeptide. Cyclic peptides that show more specificity than RGD and that induce a stronger integrin binding response had already been designed using random design (28, 29). These peptides, mostly targeting integrins $\alpha_v\beta_3$ and $\alpha_5\beta_1$, were able to promote cell adhesion and spreading and showed high affinity and selectivity (28, 29). We combined the properties of fully synthetic PIC hydrogels with cyclic peptides to determine the effect of integrin signalling in ADSCs differentiation in a controlled environment. Bicyclic peptides (Figure R-2 2) were linked to PIC hydrogels and ADSCs

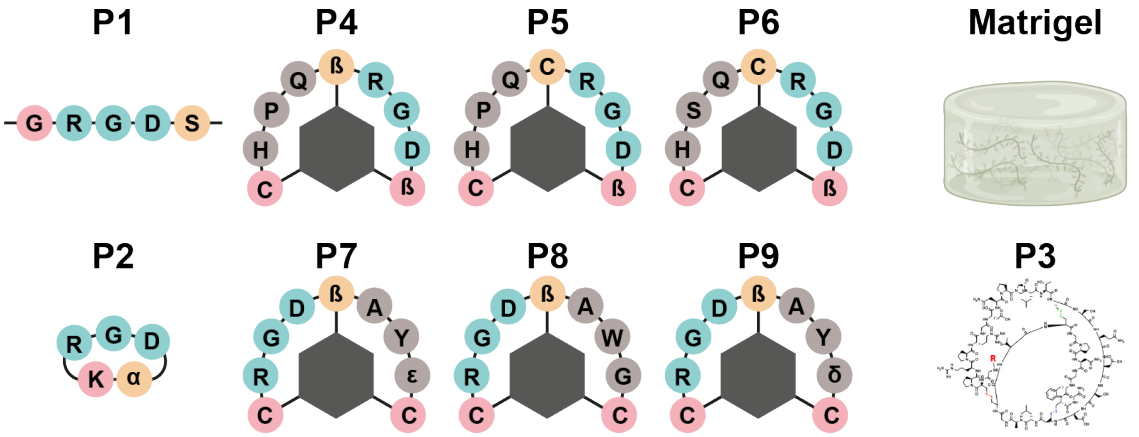


Figure R-2 10: Schematic representation of the different bicyclic peptides and matrices whose effect on cell differentiation we have studied. The structure of P3 is shown differently due to its complexity (sequence of P3: K(BCN)-linker-GCS-SRPRPRGDNPPLTCS-SSQDSDCS-SLAGCS-SVCS-SGPNGFCSSG(K(BCN)-linker-knottin-RGD). Single letter code represents aminoacids following the standard one letter aminoacid code. Special characters stand for α, D-phenylalanine, β, D-cysteine, δ, D-alanine, and ε, D-leucine.

cultured in the modified hydrogels in parallel to Matrigel that was used as control.

First, we investigated the behaviour of ADSCs in undecorated PIC (P0) and PIC decorated with different peptides: linear RGD (P1), cyclic RGD (P2), and 7 bicyclic peptides (P3-P9), using Matrigel as control. To do so, we followed cell morphology (both migration and spreading) during the first 24 hours after encapsulating cells (Figure R-2 11). We limited the imaging time to 24 hours to prevent the possible interference due to the deposition of ECM by the ADSCs. As it can be observed in Figure R-2 11, only the cells grown in P8 decorated PIC showed a clear spreading after 24 hours. Interestingly, although P7 and P9 peptides should bind with

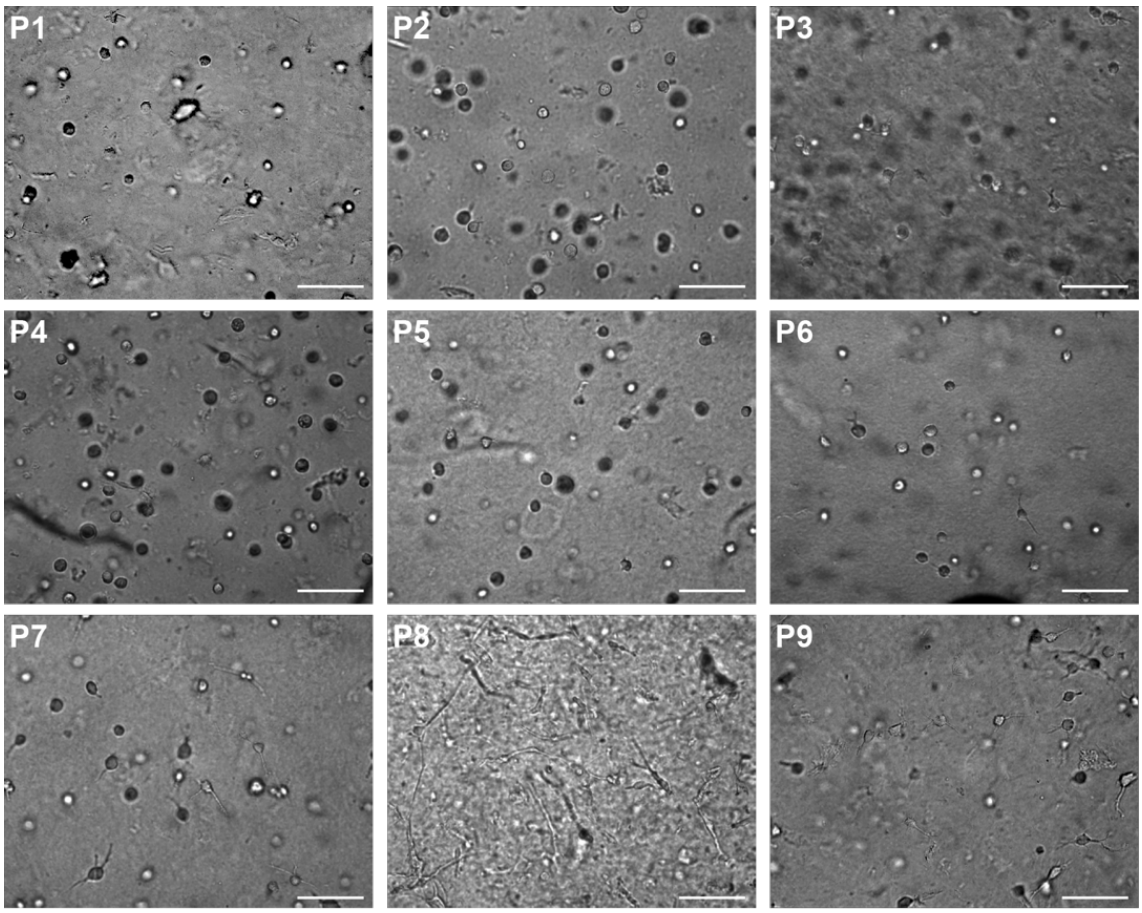


Figure R-2 11: Brightfield images of encapsulated ADSCs in PIC polymers (P1-P9). All images were acquired under the same conditions 24 hours after encapsulation of the cells. Scale bar: 100 μm

similar affinity to the $\alpha_5\beta_1$ integrin that P8 also binds to we could not observe similar spreading in the gels decorated with these peptides. Alternatively, those cyclic peptides aimed towards $\alpha_v\beta_3$ did not elicit any clear effect on cell morphology. P3, a longer peptide structure that strongly binds to both $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins did not induce changes in cell morphology. Based on these results, we decided to perform an in-depth proteomic and imaging analysis of cells cultured in five selected matrices: P0 (same PIC-N3 as in the previous results section, as a negative control), P1 (standard linear PIC-RGD as in the previous results section), P4 (bicyclic peptide with high affinity towards $\alpha_v\beta_3$ integrin), P8 (bicyclic peptide with high affinity towards $\alpha_5\beta_1$ integrin) and Matrigel.

Fluorescence imaging of cells cultured in the 5 selected matrices acquired at 7 and 24 hours confirmed the striking differences we had observed in brightfield images. Furthermore, already at 7 hours we could see a clear spreading of ADSCs in P8 decorated polymers, while nothing similar could be detectable in any of the other matrices (Figure R-2 12). Protein extracts from ADSCs cultured for 24 hours in the different matrices were labelled in duplicate using a TMT 10-plex kit (Figure R-2 13A) following standardized protocols (isobaric labelling in materials and methods). Labelled samples were separated into 6 fractions according to peptide hydrophobicity prior to LC–MS/MS analysis using a Q Exactive mass spectrometer. Mass spectrometry raw data were analysed using Maxquant and normalised, and 2406 proteins were identified (Table S10). Principal component analysis of the different samples and duplicates showed a clear correlation between samples duplicates, except for P8 samples for which single experiments were not grouped so close (Figure R-2 13B). The number of upregulated proteins when comparing all the different conditions is shown in Table R-2 1. From the information in Table R-2 1 it can be clearly seen the trend we already observed with the principal component analysis about the similarities between P0

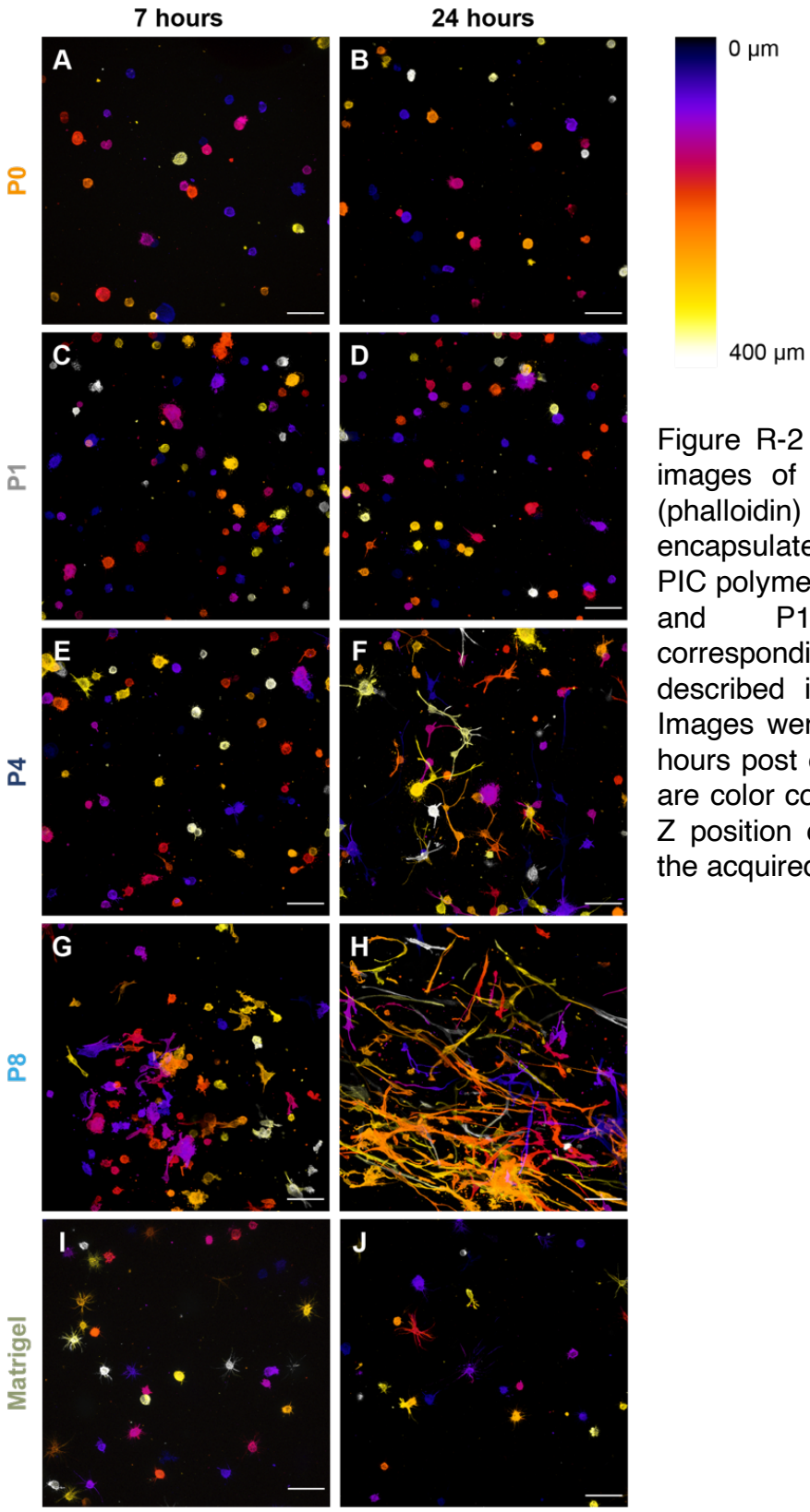


Figure R-2 12: Fluorescence images of filamentous actin (phalloidin) staining of ADSCs encapsulated in the indicated PIC polymers. P0, P1, P4, P8 and P10 are the corresponding PIC polymer as described in Figure R2 10. Images were taken 7 and 24 hours post encapsulation and are color coded based on the Z position of the cells within the acquired stacks.

and P1 conditions and P4 being slightly different to them. Matrigel appears as the condition with the largest number of up- and down-regulated proteins when compared to any of the other matrices being P8 the second most different.

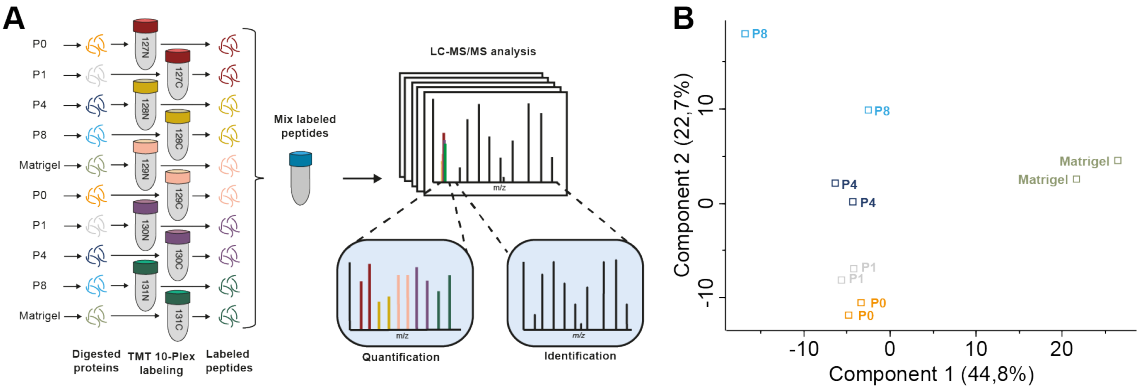


Figure R-2 13: Proteomic analysis of bicyclic peptide induced ADSCs differentiation. (**A**) Labelling scheme of the protein extracts derived from ADSCs encapsulated in each of the conditions described. (**B**) Principal component analysis of the different samples from the proteomic study.

Table R-2 1: Number *of proteins overexpressed in each matrix compared with the other matrices. OE indicates the direction of the overexpression, for example, there are 84 proteins overexpressed in Matrigel when compared with P0.*

OE ⇩	Matrigel	P0	P1	P4	P8
Matrigel	0	84	91	97	36
P0	100	0	2	32	40
P1	53	1	0	23	29
P4	53	13	3	0	3
P8	53	30	28	18	0

Venn diagrams of the proteins overexpressed in each condition can be seen in Figure R-2 14. Proteins overexpressed by cells grown in Matrigel are mostly found in all comparisons, in other words, regardless of which other matrix Matrigel is compared to the proteins upregulated are very similar (Figure R-2 14A). On the other hand, for all the other matrices the overexpressed proteins were mostly specific to each of the comparisons (Figure R-2 14B-E). From the proteins overexpressed in Matrigel two sets

could be considered, one for those proteins upregulated compared to P0, P1 and P4, and a second set for those upregulated compared to all bicyclic peptides. Reactome and string analysis of the 46 proteins that correspond to that first set showed enrichment in proteins related to extracellular matrix recognition, adhesion and remodelling. Proteins from the second set did not belong to any pathway significantly overrepresented. Remarkably, the proteomic comparison between Matrigel and P8 showed that proteins overexpressed in Matrigel belonged to pathways related to formation of components from the ECM and remodelling and organization of the ECM. On the other hand, proteins overexpressed in P8 compared with Matrigel were part of Rho signalling, smooth muscle contraction or axon guidance.

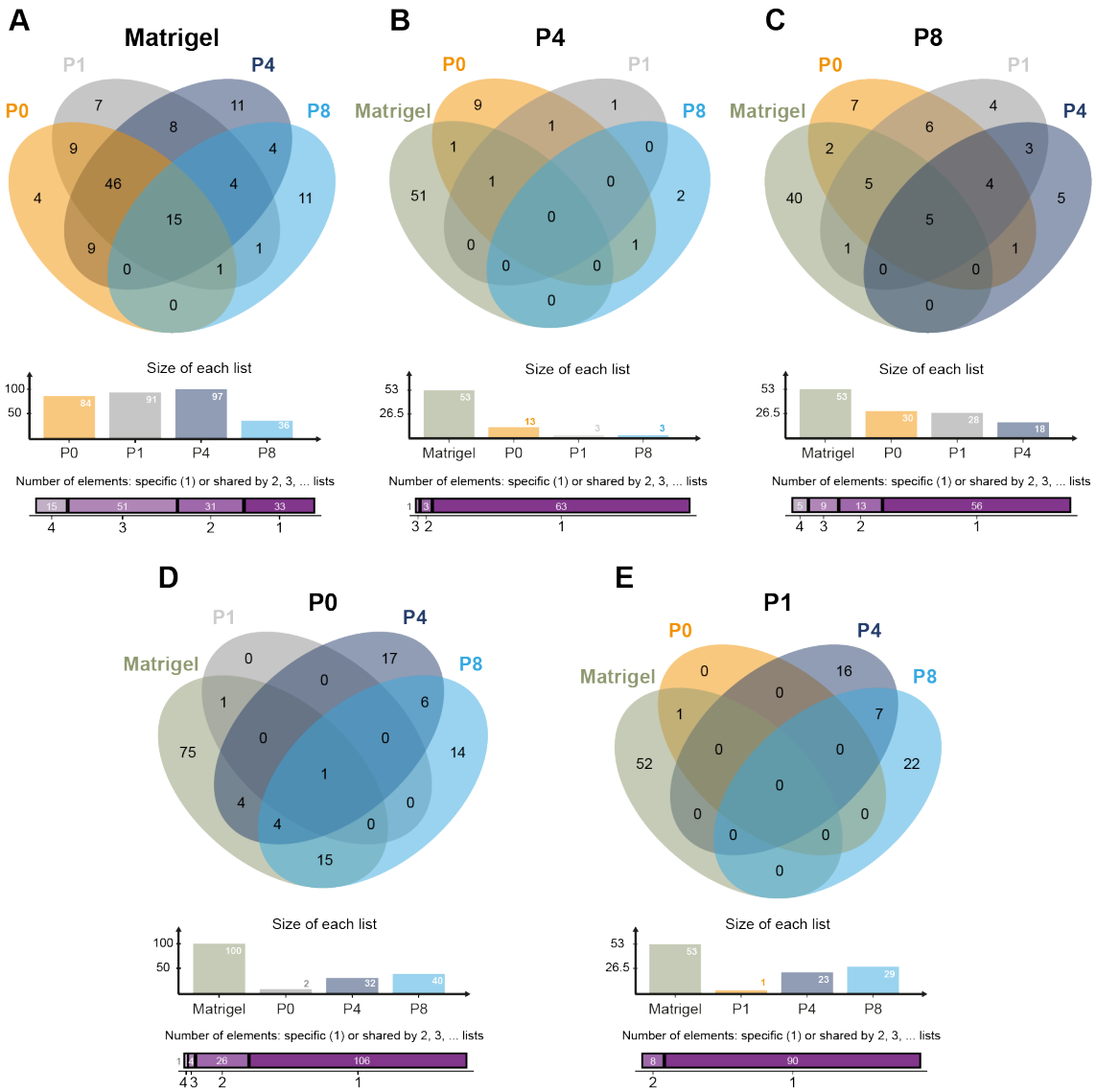


Figure R-2 14: Venn diagrams of the proteins overexpressed in each of the evaluated matrices compared with the remaining four matrices. Proteins overexpressed in Matrigel (**A**), P4 (**B**), P8 (**C**), P0 (**D**) and P1 (**E**).

3D culture can drive non-metastatic CRC cells to become more metastatic

The analysis of the ADSCs differentiation in 3D demonstrated that our proteomic analysis pipeline was suited for investigating protein expression patterns of cells encapsulated in different conditions and growing in 3D. Consequently, we decided to take a step forward and apply the same methodology this time to the KM12 CRC model. As we mentioned already in the introduction, the proteomic spatial analysis of the CRC models described in chapter 1, did not yet fully cover the proteins we found to be upregulated in adenocarcinoma samples derived from patients (Figure R-2 1).

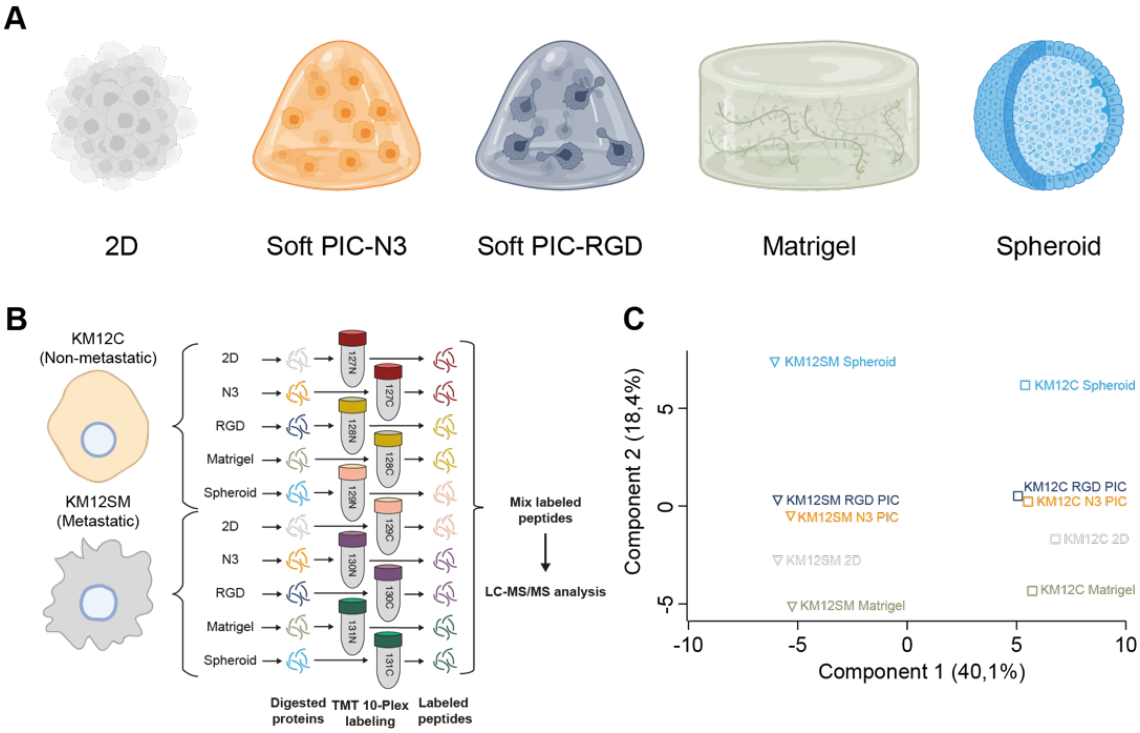


Figure R-2 15: Proteomic analysis of 3D culture of CRC metastasis model. (**A**) Schematic representation of the different matrices and culture conditions that were considered for the study. (**B**) TMT labelling scheme of the protein extracts from the different conditions. (**C**) Principal component analysis of all the labelled samples, squares represent non-metastatic cells (KM12C) and triangles represent metastatic (KM12SM) cells.

To try to identify markers that we might have overlooked by culturing cells in traditional 2D conditions, we decided to grow both KM12C and KM12SM in different matrices (Figure R-2 15A). The conditions we chose were undecorated soft PIC, RGD decorated soft PIC and Matrigel. In addition to these 3 matrices, we used 2D cultured cells as a negative control and spheroids from the two cell lines. For all conditions cells were grown for a week before protein extraction and cell morphology was followed during the differentiation. We did not observe major differences among the different matrices, except for the cells that were forming the spheroids that due to the physical constraints and adhesive properties of the agar were different to the other cells. After culturing cells for 7 days, protein extracts were TMT 10-plex labelled (Figure R-2 15B). Samples were separated onto 6 fractions based on hydrophobicity prior to LC—MS/MS analysis using a Q Exactive mass spectrometer. Raw files were analysed with MaxQuant and subsequently normalised prior to protein expression analysis with Perseus. A total of 1357 proteins were identified and quantified (Table S11).

Principal component analysis of the different samples showed that KM12C and KM12SM samples were easily distinguishable and separated in two different regions of the plot (Figure R-2 15C). The differences in component 1 of the analysis defined two clouds of points, one for the KM12C and another one for the KM12SM. Most of the conditions were similar in component 1 content and showed differences in the contribution of component 2. A small exception to this would be KM12C cells cultured in 2D that showed a displacement towards the right side of the graph, hinting that KM12C cells cultured in the different matrices and 3D conditions are more similar to the KM12SM than the 2D cultured cells.

As a measure of similarity between KM12C and KM12SM cells that would allow to evaluate whether they become more similar depending on the culture conditions we evaluated the number of upregulated proteins in

each condition comparing non-metastatic against metastatic cells (Table R-2 2) different when they are grown in 2D, and show more similarities when they are grown in any of the evaluated 3D conditions. We decided to compare the protein expression pattern of all KM12C 3D cultured cells with the KM12SM 2D cultured to try to discern whether the similarities that we had observed meant that KM12C cells were becoming more metastatic or not (Figure R-2 16). From the Venn diagram in Figure R-2 16, it can be seen that there are 10 proteins that are upregulated in KM12SM cells in 2D that also become upregulated in KM12C cells when they are grown in any of the evaluated 3D cultures. Overall, one third of the proteins overexpressed by KM12SM cells grown in 2D were also upregulated in one or more of the KM12C 3D cultured cells. Although no signalling pathways were overrepresented among the altered proteins, some of them were highly interesting as the EGFR or β-catenin (CTNNB1), both associated with tumoral progression (30-33). On the other hand, 13 of the 53 proteins downregulated in KM12SM cells were also common to one or more conditions of KM12C cells although as for upregulated proteins, no signalling pathway was significantly overrepresented.

Table R-2 2: Upregulated proteins in the different matrices compared with 2D cultured cells. Third row corresponds to proteins altered (up- and downregulated) in KM12C cells compared with KM12SM in each of the culture conditions.

	2D	N3	RGD	Matrigel	Spheroid
KM12C	-	21	43	18	23
KM12SM	-	13	15	12	75
KM12C vs KM12SM	152	115	131	111	116

Interestingly, we also observed that the number of altered proteins in KM12SM cells when cultured in 3D was smaller than for KM12C when comparing individual conditions with 2D (Figure R-2 17). However, the altered proteins in KM12SM cells were mostly specific to each matrix, leading to a total higher number of altered proteins. A total of 135 proteins were altered in KM12C cells grown in 3D compared to 168 in KM12SM. In terms of the differences between the matrices selected, we could see that KM12C cells growing in 3D shared some common upregulated proteins in all conditions although downregulated proteins were more specific of each matrix (Figure R-2 16C-D). Similarly, altered proteins in KM12SM cells were mostly specific of each individual matrix, with only a handful of proteins being upregulated in 2 (8) or 3 (5) matrices at the same time and no protein being altered in all of them, neither up nor downregulated (Figure R-2 17).

Finally, we incorporated to the comparison those proteins up- and downregulated in the adenoma and adenocarcinoma analysis to evaluate whether this new approach had been more successful for the identification of altered proteins than our previous 2D studies (Figure R-2 18). Unfortunately, few proteins were common between adenoma and adenocarcinoma samples and 3D cultured samples, even for both KM12C and KM12SM.

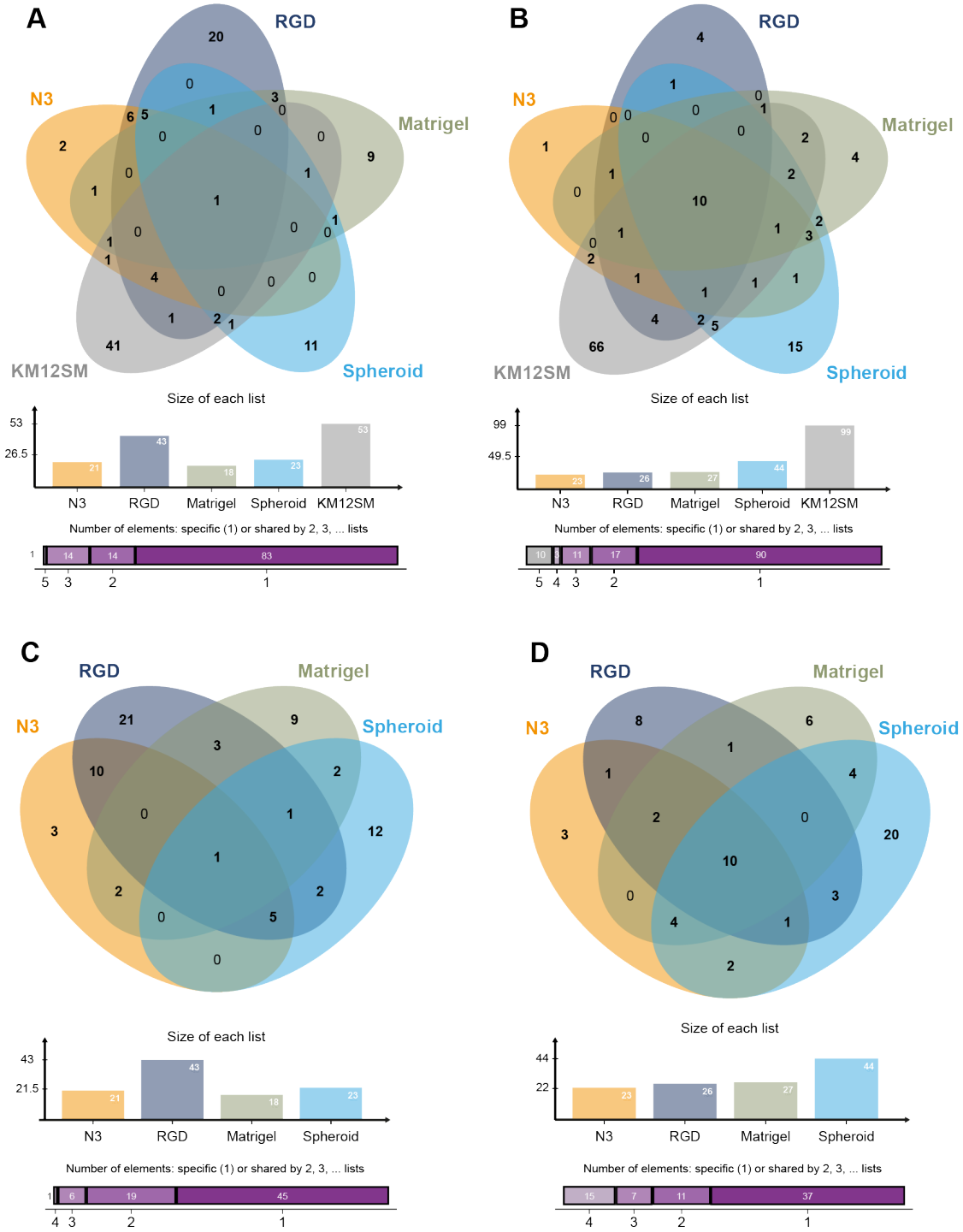


Figure R-2 16: Venn diagrams of proteins up- and downregulated in KM12C cells in the corresponding culture conditions. Down- (**A**) and upregulated (**B**) proteins in KM12C cells grown in 3D. Proteins altered in KM12SM cells cultured in 2D compared with KM12C also cultured in 2D were also included. A more simplified version of (**A**) and (**B**) without KM12SM altered proteins can be seen in (**C**) and (**D**), respectively.

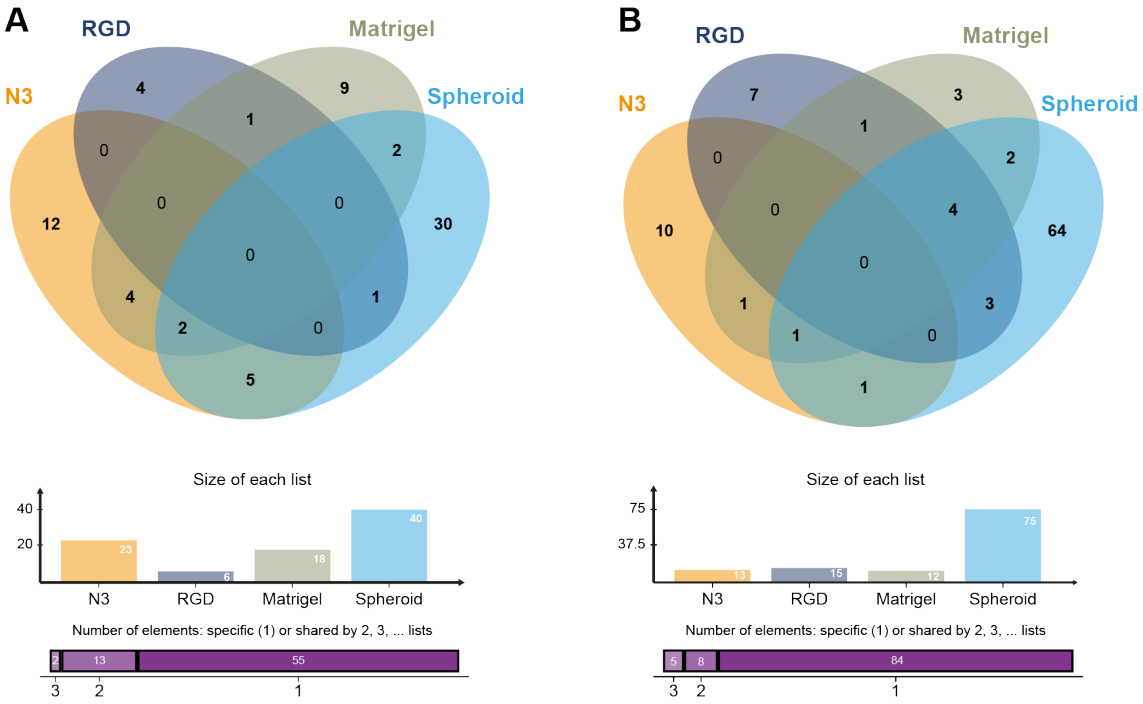


Figure R-2 17: Venn diagrams of proteins down- (**A**) and upregulated (**B**) in KM12SM cells compared with 2D cultures.

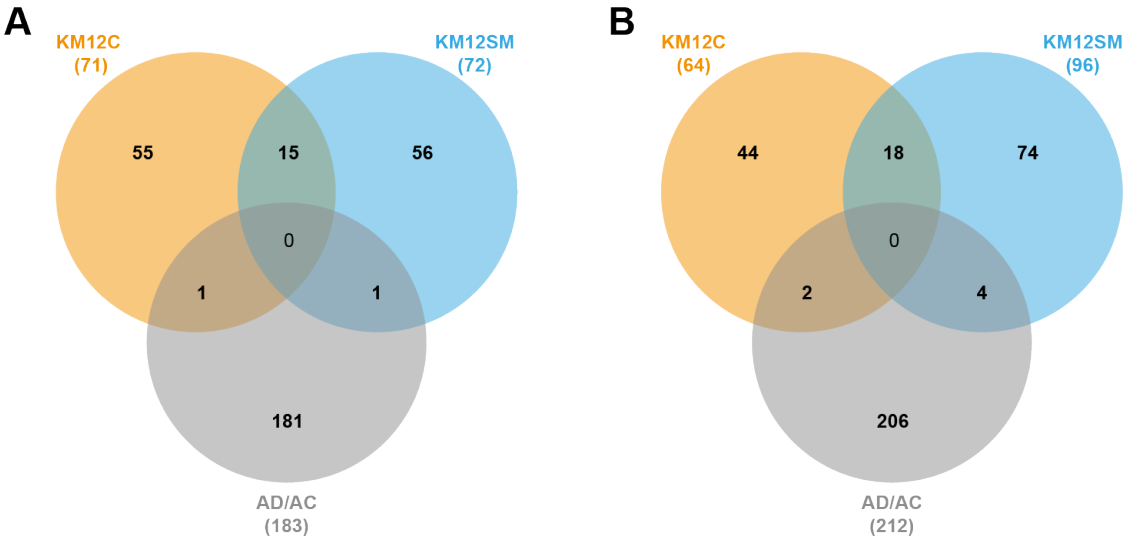


Figure R-2 18: Venn diagrams of proteins down- (**A**) and upregulated (**B**) in 3D cultured KM12C and KM12SM cells and adenoma/adenocarcinoma tissues.

Metastatic and non-metastatic colorectal cancer cells interact differently with cancer associated fibroblasts

So far, in our aim to characterize the role of the extracellular matrix in cancer progression, we had only focused on the matrix itself. Nevertheless, as it was discussed in the introduction, the tumour microenvironment (TME) is not only composed of ECM. Furthermore, cells present in the TME can have a tremendous impact in the evolution of the tumour. Cancer associated fibroblasts (CAF) in particular, have been shown to have a central role in the interplay between tumoral cells and the TME, secreting ECM and signalling mediators like TGF-β. Subsequently, we sought to better understand the interaction between our CRC metastasis model and CAFs. Proteomic characterization of co-cultures from CRC cells together with CAFs, although possible (34, 35), would require setting up a more complex analysis pipeline. Consequently, we decided to start with a simpler approach and rely on fluorescence microscopy.

To study the interaction between KM12 cells and CAFs, we encapsulated GFP-labelled CAFs together with KM12 cells in different matrices. Cells were co-cultured for 5 days and then stained with phalloidin to observe cytoskeletal organisation of both CAF and KM12 cells. First, we encapsulated cells in Matrigel to use it as a positive control (Figure R-2 19). We could not detect major differences in cell morphology of CAF nor KM12 cells when encapsulated in Matrigel. CAFs were clearly spread, similar to what we had observed for ADSCs in the different decorated PIC hydrogels, while KM12 cells were mostly aggregated in clump-like structures. However, we could see a clear change in the distribution of CAFs and KM12 cells. In KM12C-CAF co-culture, CAFs were interwoven between KM12C aggregates, that had a relatively small size (Figure R-2 19A-C). On the other hand, KM12SM aggregates were larger than those

of KM12C cells and CAFs were segregated from these structures (Figure R-2 19D-F).

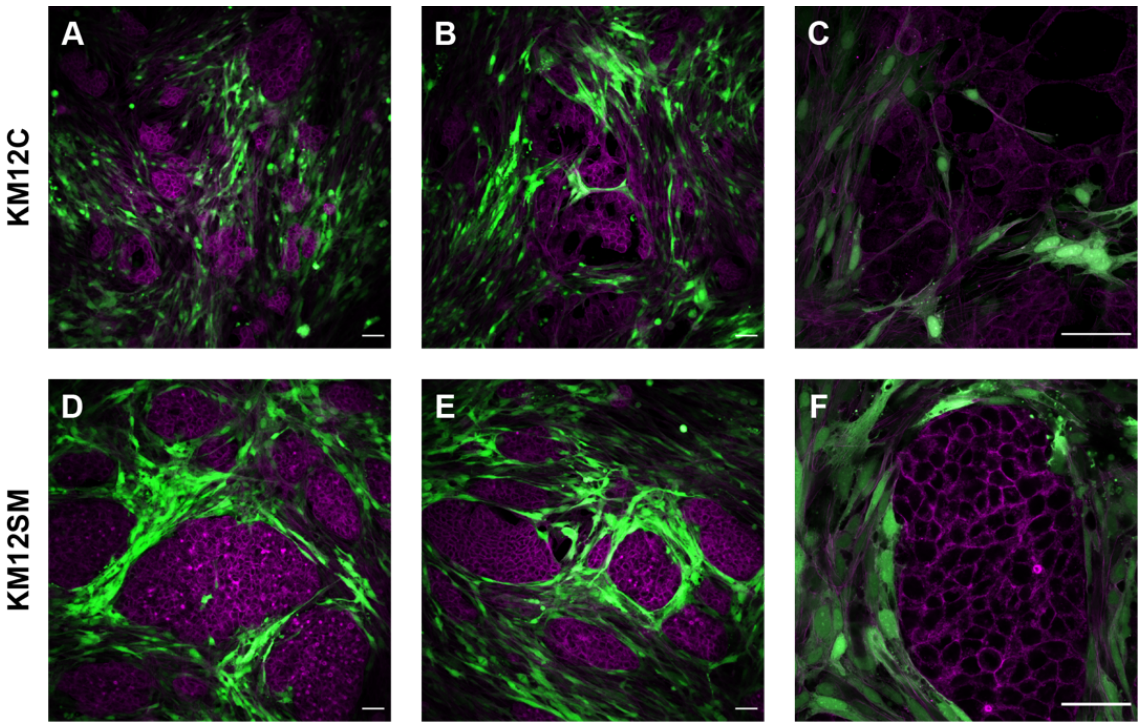


Figure R-2 19: Fluorescence images of KM12C (**A** to **C**) and KM12SM (**D** to **F**) cells cultured together with CAFs in Matrigel. CAFs were GFP labelled and shown in green. To observe KM12C and KM12SM cells, they were stained with phalloidin to label filamentous actin (magenta). Scale bar: 50 μm.

Then, to better assess the mechanical forces applied by CAFs in 3D and how were they affected by the presence of CRC cells with different metastatic potential we encapsulated CAFs and KM12 in fluorescently labelled PIC polymer with or without RGD. Strikingly, CAFs were more sensitive to the presence of RGD compared to ADSCs (Figure R-2 20). CAFs encapsulated in undecorated PIC appeared as apoptotic cells with abundant accumulation of apoptotic blebs (Figure R-2 20A-B). This effect was independent of the cell type with which CAFs were co-cultured and neither KM12C (Figure R-2 20A) nor KM12SM (Figure R-2 20B) were able to rescue this phenotype. Conversely, CAFs encapsulated in PIC decorated with RGD (Figure R-2 20C-D) had a similar morphology to that observed in Matrigel or in 2D. Besides the change observed in CAF

morphology in RGD decorated PIC, we could also detect how CAFs exerted mechanical force in the PIC polymer (Figure R-2 20C-D). Regions of the matrix being pulled by the CAFs appeared as brighter, more intense, regions among the structure of the PIC due to the local increase in polymer concentration.

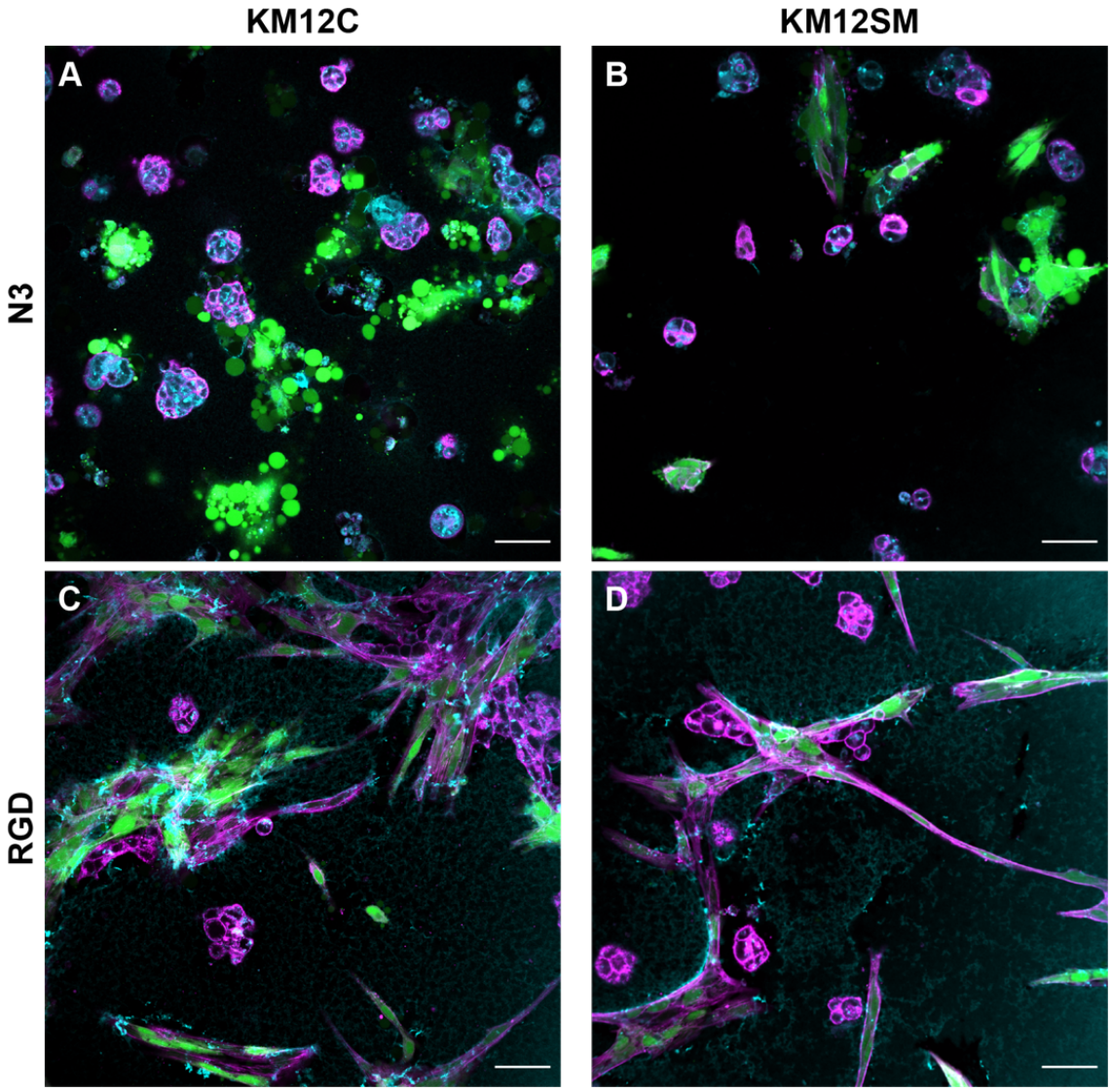


Figure R-2 20: Fluorescence images of KM12C (**A** and **C**) and KM12SM cells (**B** and **D**) encapsulated in PIC polymer with CAFs. PIC polymer was labelled with Atto 655 (cyan). (**A**) and (**B**) correspond to cells encapsulated in undecorated PIC while (**C**) and (**D**) correspond to cells encapsulated in RGD decorated PIC. CAFs were labelled with GFP (green) while phalloidin (magenta) was used to stain the actin cytoskeleton. Scale bar: 50 μm.

Discussion

In this work we have conducted the proteomic characterisation of 3D cultured cells to, on the one hand, test the suitability of our current proteomic analysis pipeline for the analysis of this model and, on the other hand, evaluate the role of the matrix in CRC metastasis and diagnosis.

In the first part of our study, we focused on the characterisation of adipose derived stem cells cultured in a monolayer or encapsulated in PIC polymers with no chemical ligands and PIC polymers decorated with RGD. Interestingly, the differences between cells encapsulated in RGD decorated PIC and cells cultured in 2D were not as marked as those of cells cultured in undecorated PIC. A possible explanation for this observation would be that ADSCs cultured in 2D display adhesion molecules that signal downstream the same mediators that integrins activated by RGD would be channelled through. Indeed, RGD is only one of the many ligands that integrins can have and cell-cell adhesion molecules like PECAM-1 (CD31) or VCAM-1 (CD106) can also be recognised by integrins (11). Among the molecules overexpressed by cells grown in 2D compared with ADSCs cultured in RGD-PIC we could find ICAM-1 (CD54), that can also be recognised by integrins ($\alpha_L\beta_2$). Activation of integrin signalling in ADSCs via ICAM-1 heterotypic interaction in 2D cultured cells and via RGD in 3D encapsulated cells could lead to a similar phenotype, explaining the relative absence of differences between these two conditions (36). Remarkably, Hutton et al showed that ADSCs cultured under high confluency were able to self-organize into vascular structures formed by a mixture of PECAM-1(+), α-smooth muscle actin (+) and double negative ADSCs (36). Furthermore, RGD activation has been shown to induce an unresponsiveness to TNF-α mediated increase in ICAM-1 in immune cells, which would also explain why ICAM-1 levels are reduced in RGD-PIC encapsulated cells (37).

In terms of the differences between the matrix conditions evaluated for ADSCs differentiation, it was striking to find that the presence of RGD can clearly outplay the role of matrix stiffness, as seen by the similarities between soft and stiff matrices when RGD was present. Matrix stiffness has been shown to induce an increase in latent TGF-β activation, a process that has been associated with cellular senescence (38). TGF-β secreted onto the ECM is often coupled to latent TGF-β binding protein (LTBP) and kept in an inactive form, to be activated only upon protease digestion of LTBP. However, in the context of increased matrix stiffness due to aberrant ECM deposition, the mechanical pulling fibroblasts could exert in the matrix could induce the release of this TGF-β leading to the induction of senescence in some cell types (38-40). The results of our analysis of ADSCs differentiation do not only give an initial explanation to the differences observed in the secretome of ADSCs cultured in different matrix, but also prove the suitability and potential of PIC for 3D cell culture.

Secondly, based on the changes induced by the presence of RGD in the matrix, we decided to evaluate the effect that other integrin ligands would have. In particular, we investigated bicyclic peptides developed to bind integrin $\alpha_V\beta_3$ and $\alpha_5\beta_1$. From these, the ligand that had the most remarkable effect on ADSCs differentiation was P8, a ligand specifically binding to integrin $\alpha_5\beta_1$. In principle, the three peptides (P7-P9) designed to bind $\alpha_5\beta_1$ integrin had similar binding affinity (28, 29, 41), nonetheless, these rate constants had not been evaluated with a constrained peptide, as it would be the case of the ligand bound to the PIC skeleton. In addition, chemical binding to PIC could also affect the affinity of the peptides further explaining the differences observed. As for the differences that we have detected between triggering $\alpha_V\beta_3$ integrin or $\alpha_5\beta_1$, the proteomic analysis of the ADSCs showed that both alpha subunits had similar expression levels, while there was clearly more abundance of β_1 subunit than of β_3. In terms of the proteins overexpressed, we could clearly differentiate

between three conditions, cells encapsulated in Matrigel, cells encapsulated in P8 and cells encapsulated in the other matrices. When we compared the proteins overexpressed in Matrigel with those overexpressed in P8 we found riveting differences. While cells encapsulated in Matrigel had higher levels of proteins involved in matrix recognition and remodelling, cells encapsulated in P8 had increased expression of proteins involved in actin cytoskeleton reorienting and migration. Altogether it seems that P8 matrix is more suited for ADSC differentiation. Cells can recognise the matrix as soon as they are encapsulated and then start migrating without the need of remodelling the matrix. On the other hand, in Matrigel, cells recognise the matrix but need to adapt it and modify it so that it becomes more suited for them. This would also agree with what was observed via microscopy, where cells encapsulated in P8 matrix spread already 7 hours after seeding while only some minor spreading could be observed in cells encapsulated in Matrigel after 24 hours.

The analysis of ADSCs differentiation confirmed that the data obtained from 3D cultured cells was in agreement with what had been previously reported in literature. Consequently, we decided to tackle the more complex question of how 3D culture and the matrix would affect CRC metastasis. Our data suggest that non-metastatic KM12C cells become more aggressive and potentially more metastatic when they are cultured in 3D. Although no particular signalling pathways were enriched among the proteins upregulated in KM12C 3D cultured cells that were common to KM12SM overexpressed cells some of them had previously been associated with tumour progression. Contrastingly, downregulated proteins did not follow the same trend and we could not detect any pattern when comparing KM12C 3D cultured cells and KM12SM cells. In terms of the proteins expressed in the different conditions, we found that the vast majority of the proteins altered in KM12SM cells were compartment specific. On the other hand, proteins upregulated in KM12C cells were

more recurrent and we found that nearly 50% percent of all the upregulated proteins were altered in at least two different conditions. This was not the case for downregulated proteins, where there was not so much overlap between conditions. Although the proteins we found to be altered are promising candidates for future studies, we were not able to increase the coverage of proteins altered in patient-derived tumoral tissues. One plausible reason behind this would be the fact that tumoral samples are derived from tissues and thus contain a mixture of different cells. In addition, patient variability can also affect protein expression, as it can be seen by the rather small number of proteins we found to be altered (212) compared with, for example, the lymph node metastasis model (1337).

Finally, our study of the interplay between KM12C and KM12SM cells with CAFs, although limited in the extent to which it was performed, revealed some interesting results. As it can be seen in the images in Figure 19, it is clear that CAFs distribute differently when they are co-cultured with KM12C or KM12SM cells. Although at present we do not have any clear explanation for this, it is reasonable to think that the metastatic status of cells can condition how they interact with the cells surrounding them. KM12C cells represent an earlier stage of the disease, when the initial tumoral mass is forming. It is possible that in this initial stage cancer cells are more flexible to the interaction with CAFs. CAFs have been shown to be able to stimulate the generation of vessels (42, 43), much needed by the growing tumoral mass. On the other hand, in the metastatic phase of the disease, cancer cells need to settle in the new tissue and have to grow rapidly and attach to be able to colonize it. In this context, the presence of CAFs would not be as helpful as in the previous stage, explaining the distribution we have observed. In the future, it would be interesting to also include in our model metastasis associated fibroblasts or subject it to other signals or alterations that might perturb the interaction between CRC cells and CAFs.

In the work here presented we were able to on the one hand, successfully apply our proteomic analysis pipeline for the study of 3D encapsulated cells, and on the other hand demonstrate the usefulness of PIC-based hydrogels for cell culture. The changes we detected in protein expression in the different matrices we evaluated for ADSCs differentiation are in agreement with the data reported in literature and partially explain previous observations from our collaborators. Moreover, the results from ADSCs differentiation in PIC gels decorated with bicyclic peptides clearly demonstrate that the platform has tremendous potential as a novel culture system. Not only did cells differentiate morphologically faster in P8 decorated PIC than in Matrigel but also, at the protein level, we could demonstrate that P8 decorated PIC was more suited for direct ADSCs spreading than Matrigel. Finally, our characterization of 3D encapsulated CRC cells revealed some interesting changes in non-metastatic KM12C cells and open up the way for future research on the role of the matrix in the metastatic process. In addition, the changes in the distribution of CAFs and KM12C and KM12SM cells are also very interesting and will be further researched on.

Altogether, the evidence shown here demonstrate the suitability of both our proteomic analysis pipeline and PIC-based hydrogels to continue the study of the KM12 CRC metastasis model in 3D and expand our knowledge on how the cells interact with the ECM.

References

1. Fu J, Wang YK, Yang MT, Desai RA, Yu X, Liu Z, et al. Mechanical regulation of cell function with geometrically modulated elastomeric substrates. Nature methods. 2010;7(9):733-.
2. Kim SH, Turnbull J, Guimond S. Extracellular matrix and cell signalling: the dynamic cooperation of integrin, proteoglycan and growth factor receptor. Journal of Endocrinology. 2011;209(2):139-51.
3. Sun M, Chi G, Xu J, Tan Y, Xu J, Lv S, et al. Extracellular matrix stiffness controls osteogenic differentiation of mesenchymal stem cells mediated by integrin α5. Stem Cell Research and Therapy. 2018;9(1):1-13.
4. Ozdil B, Güler G, Acikgoz E, Kocaturk DC, Aktug H. The effect of extracellular matrix on the differentiation of mouse embryonic stem cells. Journal of Cellular Biochemistry. 2020;121(1):269-83.
5. Novoseletskaya E, Grigorieva O, Nimiritsky P, Basalova N, Eremichev R, Milovskaya I, et al. Mesenchymal Stromal Cell-Produced Components of Extracellular Matrix Potentiate Multipotent Stem Cell Response to Differentiation Stimuli. Frontiers in Cell and Developmental Biology. 2020;8:983-.
6. Gattazzo F, Urciuolo A, Bonaldo P. Extracellular matrix: A dynamic microenvironment for stem cell niche. Biochimica et Biophysica Acta. 2014;1840(8):2506-.
7. Watt FM, Huck WTS. Role of the extracellular matrix in regulating stem cell fate. Nature Reviews Molecular Cell Biology 2013 14:8. 2013;14(8):467-73.
8. Nallanthighal S, Heiserman JP, Cheon DJ. The Role of the Extracellular Matrix in Cancer Stemness. Frontiers in Cell and Developmental Biology. 2019;7:86-90.
9. Guilak F, Cohen DM, Estes BT, Gimble JM, Liedtke W, Chen CS. Control of stem cell fate by physical interactions with the extracellular matrix. Cell stem cell. 2009;5(1):17-.
10. Kim HN, Ruan Y, Ogana H, Kim YM. Cadherins, Selectins, and Integrins in CAM-DR in Leukemia. Frontiers in Oncology. 2020;10:2554-.
11. Humphries JD, Byron A, Humphries MJ. Integrin Ligands. Journal of cell science. 2006;119(Pt 19):3901-.
12. Kechagia JZ, Ivaska J, Roca-Cusachs P. Integrins as biomechanical sensors of the microenvironment. Nature Reviews Molecular Cell Biology 2019 20:8. 2019;20(8):457-73.

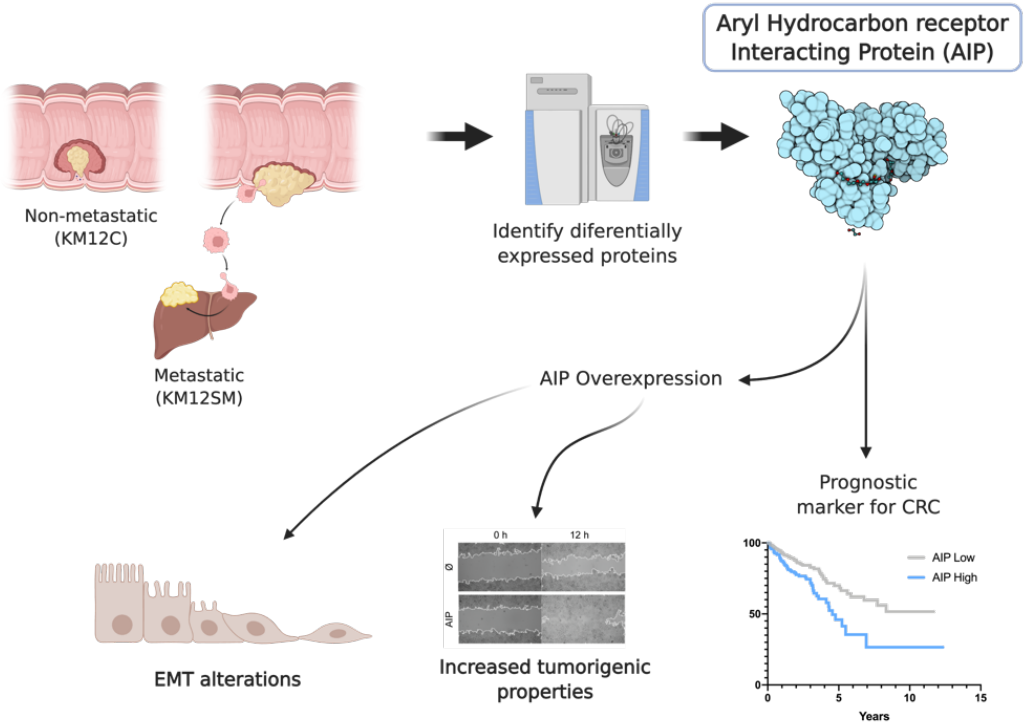
13. Smith LR, Cho S, Discher DE. Stem Cell Differentiation is Regulated by Extracellular Matrix Mechanics. Physiology. 2018;33(1):16-.
14. Flanagan LA, Ju YE, Marg B, Osterfield M, Janmey PA. Neurite branching on deformable substrates. Neuroreport. 2002;13(18):2411-.
15. Engler AJ, Sen S, Sweeney HL, Discher DE. Matrix Elasticity Directs Stem Cell Lineage Specification. Cell. 2006;126(4):677-89.
16. Guvendiren M, Burdick JA. Engineering synthetic hydrogel microenvironments to instruct stem cells. Current Opinion in Biotechnology: NIH Public Access; 2013. p. 841-6.
17. Geckil H, Xu F, Zhang X, Moon S, Demirci U. Engineering hydrogels as extracellular matrix mimics. Nanomedicine: NIH Public Access; 2010. p. 469-84.
18. Hughes CS, Postovit LM, Lajoie GA. Matrigel: A complex protein mixture required for optimal growth of cell culture. PROTEOMICS. 2010;10(9):1886-90.
19. Gyles DA, Castro LD, Silva JOC, Ribeiro-Costa RM. A review of the designs and prominent biomedical advances of natural and synthetic hydrogel formulations. European Polymer Journal: Elsevier Ltd; 2017. p. 373-92.
20. Das RK, Gocheva V, Hammink R, Zouani OF, Rowan AE. Stress-stiffening-mediated stem-cell commitment switch in soft responsive hydrogels. Nature Materials. 2016;15(3):318-25.
21. Storm C, Pastore JJ, MacKintosh FC, Lubensky TC, Janmey PA. Nonlinear elasticity in biological gels. Nature 2005 435:7039. 2005;435(7039):191-4.
22. Rudnicki MS, Cirka HA, Aghvami M, Sander EA, Wen Q, Billiar KL. Nonlinear Strain Stiffening Is Not Sufficient to Explain How Far Cells Can Feel on Fibrous Protein Gels. Biophysical Journal. 2013;105(1):11-20.
23. Liu K, Veenendaal T, Wiendels M, Ruiz-Zapata AM, Van Laar J, Kyranas R, et al. Synthetic Extracellular Matrices as a Toolbox to Tune Stem Cell Secretome. ACS Applied Materials and Interfaces. 2020;12(51):56723-30.
24. Ruoslahti E, Pierschbacher MD. New Perspectives in Cell Adhesion: RGD and Integrins. Science. 1987;238(4826):491-7.
25. Ruoslahti E. RGD and other recognition sequences for integrins. http://dxdoiorgkuleuvene-bronnenbe/101146/annurevcellbio121697. 2003;12:697-715.

26. Pierschbacher MD, Ruoslahti E. Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. Nature. 1984;309(5963):30-3.
27. Meyaard L, Hovenkamp E, Otto SA, Miedema F. IL-12-induced IL-10 production by human T cells as a negative feedback for IL-12-induced immune responses. The Journal of Immunology. 1996;156(8).
28. Cipriani F, Bernhagen D, Garcia-Arévalo C, De Torre IG, Timmerman P, Rodriguez-Cabello JC. Bicyclic RGD peptides with high integrin αvβ3 and α5β1 affinity promote cell adhesion on elastin-like recombinamers. Biomedical Materials. 2019;14(3):035009-.
29. Bernhagen D, Jungbluth V, Gisbert Quilis N, Dostalek J, White PB, Jalink K, et al. High-Affinity α5β1-Integrin-Selective Bicyclic RGD Peptides Identified via Screening of Designed Random Libraries. ACS Combinatorial Science. 2019;21(8):598-607.
30. Arnold A, Tronser M, Sers C, Ahadova A, Endris V, Mamlouk S, et al. The majority of β-catenin mutations in colorectal cancer is homozygous. BMC Cancer. 2020;20(1).
31. Markman B, Javier Ramos F, Capdevila J, Tabernero J. EGFR and KRAS in colorectal cancer. Adv Clin Chem. 2010;51:71-119.
32. Sigismund S, Avanzato D, Lanzetti L. Emerging functions of the EGFR in cancer. Mol Oncol. 2018;12(1):3-20.
33. Yu Jj, Zhou Dd, Yang Xx, Cui B, Tan Fw, Wang J, et al. TRIB3-EGFR interaction promotes lung cancer progression and defines a therapeutic target. Nature Communications 2020 11:1. 2020;11(1):1-16.
34. Chignell JF, Park S, Lacerda CMR, De Long SK, Reardon KF. Label-Free Proteomics of a Defined, Binary Co-culture Reveals Diversity of Competitive Responses Between Members of a Model Soil Microbial System. Microbial Ecology. 2018;75(3):701-19.
35. Liu R, Wang Y, Li B, Wang H, Guan F, Tan Z, et al. Screening differentially expressed proteins from co-cultured hematopoietic cells and bone marrow-derived stromal cells by quantitative proteomics (SILAC) method. Clinical Proteomics. 2019;16(1):32.
36. Hutton DL, Logsdon EA, Moore EM, Gabhann FM, Gimble JM, Grayson WL. Vascular Morphogenesis of Adipose-Derived Stem Cells is Mediated by Heterotypic Cell-Cell Interactions. Tissue Engineering Part A. 2012;18(15-16):1729-.
37. Ogawara KI, Kułdo JM, Oosterhuis K, Kroesen BJ, Rots MG, Trautwein C, et al. Functional inhibition of NF-κB signal transduction in αvβ3 integrin expressing endothelial cells by using RGD-PEG-modified

adenovirus with a mutant IκB gene. Arthritis Research and Therapy. 2006;8(1):1-10.
38. Blokland KEC, Pouwels SD, Schuliga M, Knight DA, Burgess JK. Regulation of cellular senescence by extracellular matrix during chronic fibrotic diseases. Clinical Science (London, England : 1979). 2020;134(20):2681-.
39. Tominaga K, Suzuki HI. TGF-β Signaling in Cellular Senescence and Aging-Related Pathology. International Journal of Molecular Sciences. 2019;20(20).
40. Hassona Y, Cirillo N, Lim KP, Herman A, Mellone M, Thomas GJ, et al. Progression of genotype-specific oral cancer leads to senescence of cancer-associated fibroblasts and is mediated by oxidative stress and TGF-β. Carcinogenesis. 2013;34(6):1286-95.
41. Bernhagen D, Jungbluth V, Quilis NG, Dostalek J, White PB, Jalink K, et al. Bicyclic RGD Peptides with Exquisite Selectivity for the Integrin α v β 3 Receptor Using a "random Design" Approach. ACS Combinatorial Science. 2019;21(3):198-206.
42. Tang D, Gao J, Wang S, Ye N, Chong Y, Huang Y, et al. Cancer-associated fibroblasts promote angiogenesis in gastric cancer through galectin-1 expression. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016;37(2):1889-99.
43. Sewell-Loftin MK, Bayer SVH, Crist E, Hughes T, Joison SM, Longmore GD, et al. Cancer-associated fibroblasts support vascular growth through mechanical force. Scientific Reports. 2017;7(1).

CHAPTER 3:

AIP AS DRIVER OF CRC METASTASIS



The following chapter is adapted from the publication:

Solís-Fernández G, Montero-Calle A, Sánchez-Martínez M, Peláez-García A, Fernández-Aceñero MJ, Pallarés P, et al. Aryl-hydrocarbon receptor-interacting protein regulates tumorigenic and metastatic properties of colorectal cancer cells driving liver metastasis. British Journal of Cancer. 2022;126(11):1604-15.

https://doi.org/10.1038/s41416-022-01762-1

Contribution to the article:

I performed all the functional assays (proliferation, migration, invasion and colony agar), western blot and PCR and qPCR, immunofluorescence and image acquisition, survival data analysis, TMT labelling and mass spectrometry data analysis and writing and correction of the manuscript.

Introduction

Metastasis is the final step of malignant transformation and the main responsible for morbidity and mortality in cancer. Indeed, more than 90% of the mortality associated with cancer is due to metastasis (1). Metastasis cannot be understood as a unique process but as a collection of different events with unique molecular characteristics, where cancer cells need to interact with different microenvironments, affecting and being affected by cell and extracellular matrix of such tissues (2, 3). There, cancer cells must adapt to each situation, altering the expression, localization and activation of proteins to generate a metastasis.

In the first chapter of this thesis, we focused on the identification of metastasis-associated proteins and altered pathways taking advantage of the possibilities of both quantitative proteomics and isogenic cell lines studies (4). Here, our aim was to discover new targets of intervention and diagnosis that would help improving patients' survival. Quantitative proteomic analysis of cell lines with the same genetic background, but differing in their metastatic capabilities, has an immense potential to unveil clinically relevant underlying mechanisms (5-8). In previous works from our research group, we quantitatively studied by in-depth proteomics the secretome and spatial proteome of both KM12C and KM12SM cells to shed light onto the biology of CRC liver metastasis (9, 10). Compared to the characterization described in chapter 1, the previous work on KM12C and KM12SM cells was done using SILAC labelling and limited to those two cell lines. As we mentioned in chapter 1, this cell system derived from a CRC patient classified as Duke's B (actual T3N0M0 or T4N0M0 of TNM classification) allows for the study of late metastatic events to liver in colorectal cancer, including liver colonization and survival (11, 12). Numerous studies support a good correlation between the findings observed in the KM12 cell system and patient samples, indicating that these isogenic cell lines recapitulate quite effectively critical issues in CRC

liver metastasis (13-17). In the initial analysis of KM12C and KM12SM, many identified proteins had been previously described as key molecules in CRC, as VEGFA, ERBB2, EGFR, MMP7, FGFR4, Cadherin-17 (CDH17) or IL13Rα2. These results encouraged us to continue characterizing these cell lines and gain further insights into proteins dysregulated between KM12C and KM12SM cells (9, 10). In this sense and considering that the function of about 20% of all human proteins remains unknown (18, 19), we focused this work on the analysis of one of these barely known proteins in CRC: aryl hydrocarbon receptor-interacting protein (AIP).

The exact function of AIP, also known as ARA9 or XAP2, remains rather obscure. AIP was first described in 1997 as a novel interactor of the aryl hydrocarbon receptor (AHR) discovered in yeast two-hybrid interaction experiments by two independent groups (20, 21). In the work of Carver et. al, AIP, named here as ARA9, was found to interact with AHR in a ligand dependent manner, showing a 11-fold increase in affinity for the AHR when it was bound to one of its native ligands, the aryl compound ß-naphtoflavone (21). In addition, they demonstrated that AIP was able not only to bind to AHR but also to AHR coupled with the 90-kDa heat shock protein (Hsp90), a chaperone involved in AHR signalling (21, 22). In parallel, another research group, also discovered the interaction between AIP and the AHR and showed that indeed AIP was able to bind to unliganded AHR and Hsp90 (20). Remarkably, in the work of Ma et al. it was found that treatment with 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) induces the dissociation of the Hsp90-AHR-AIP complex (20). However, Ma et al. also discovered that the presence of the AHR nuclear translocator (ARNT) was required for TCDD to disrupt the Hsp90-AHR-AIP complex (20). Moreover, they also demonstrated that AIP is constrained to the cell cytosol, while AHR is translocated to the nucleus upon binding to TCDD, which would suggest a transition from the Hsp90-AHR-AIP complex towards and AIP-ARNT complex upon TCDD

activation of AHR (20). Altogether, this evidence supported the role of AIP/ARA9 in AHR signalling although its exact function in the pathway remained yet unclear.

AHR signalling is crucial for the processing of xenobiotics, that should be removed from the cell to prevent its toxic effect (23). As such, a myriad of ligands for AHR have been found, although most of them are aromatic hydrocarbons both exogenous and endogenous (23). In addition to the tertiary structure described before, another co-chaperone, p23, was also found to interact with AHR to form what is considered the non-activated form of the receptor (24). Two different signalling pathways have been described to be triggered by the activation of the AHR upon ligand binding (23). In the genomic pathway, the one that is best characterized, AHR dissociates from the quaternary complex in the cytosol, and it translocates to the nucleus where it binds to ARNT, this active heterodimer is the responsible for the activation of AHR targets (23). The AHR-ARNT functional dimer will bind to a series of co-activators and repressors to recognize the consensus xenobiotic response element and induce the expression of proteins like the cytochromes CYP1A1 or P450 (25). In addition to the role in xenobiotic metabolism, studies on knock-out mice have also found a role for AHR in development of the liver (26) and vascular system (27, 28).

In the context of cancer and tumoral progression, AHR mediates the toxic and pro-tumorigenic effect of TCDD and its levels are often increased in tumours (29). However, the role of AHR in cancer as pro- or antitumorigenic remains unclear. Moreover, the different ligands can have different effects on the activity of AHR in tumour progression (30). Additionally, the different specificity of human and murine AHR for AHR ligands makes the elucidation of the actual role of AHR in tumour progression more complicated(29). A rather large body of evidence supports that AHR acts as a pro-tumorigenic factor in hepatocarcinoma

via TCDD sustained activation (31, 32). Nonetheless, one of the central questions around the implication of AHR in tumorigenesis is whether endogenous ligands would be able to achieve the same effect exogenous TCDD has (29). In a proinflammatory environment, the activation of AHR via the endogenous ligand kynurenine led to the generation of T_{reg} cells, which would link AHR to a more general pro-tumorigenic activity (33). Moreover, in breast cancer MCF-7 cells, AHR in conjunction with inflammatory signalling induced a synergistic expression of interleukin-6, further supporting the involvement of AHR in immune regulation (34).

Altogether, it is thus not surprising that the role of AIP in tumour progression has also been disputed. One of the first descriptions of AIP's involvement in tumour development was found in pituitary adenomas (35). In 2006, the germline mutation analysis of a population of pituitary adenoma predisposed subjects in Finland revealed a strong association between germline loss-of-function mutations of AIP and pituitary adenoma predisposition (35). The authors, however, were unable to find the mechanism by which AIP was exerting its alleged tumour-suppressive activity (35). Since then, several articles have also found a correlation between AIP mutations in pituitary adenomas (36-38). Up to 15-30% of family isolated pituitary adenomas and 20% of the sporadic ones carry inactivating AIP mutations (36-38). Strikingly, somatic mutation screening of CRC, breast and prostate tumour samples did not reveal a similar association between AIP mutations and cancer development (39). So far, no gain-of-function mutations for AIP have been described. Consequently, the evidence from pituitary adenomas would indicate that AIP has an anti-tumorigenic effect, which would, to some extent, contravene our previous findings that AIP is overexpressed in highly metastatic cell lines (10).

Here, we observed a clear association of AIP expression with increased liver metastasis and a worse prognosis of CRC patients. At the molecular

level, gain-of-function experiments in the KM12 cell system of CRC liver metastasis showed EMT dysregulation and a strong increase in cell adhesion, invasion, colony formation, migration, *in vivo* liver homing and liver metastasis. Proteomics analysis pointed out to an AIP-associated expression of transcription factors, EGFR and CDH17.

RESULTS

AIP overexpression in colorectal cancer patients correlates with lower overall survival and liver metastasis

The overexpression of AIP was previously observed by multidimensional proteomics in highly metastatic KM12SM compared to isogenic poorly-metastatic KM12C CRC cells (Figure R-3 1A) (10). Here, we further assessed the differential expression of AIP by semi-quantitative PCR and WB analysis (Figure R-3 1B). In addition, we also investigated whether AIP dysregulation could be associated to other CRC cells. AIP mRNA expression was observed in 5 out of the 8 tested cell lines by semi-quantitative PCR analyses (Figure R-3 1C). qPCR analysis confirmed that the highest AIP expression mRNA levels were observed in the metastatic SW620 and LIM1215 colon cancer cell lines, together with Colo320 cells (Figure R-3 1D). Cell lines with lower metastatic capacity, like SW480 cells from the SW480/SW620 isogenic pair, RKO or HT29 (Figure R-3 1D) (40), showed the lowest mRNA expression levels of AIP. SW480/SW620 results were in concordance with KM12 cells results, where metastatic cells showed higher AIP expression than poorly or non-metastatic cells from the isogenic pair.

Then, to investigate the clinical relevance in human CRC, we analysed AIP mRNA expression in tumour tissue samples using two public cohorts containing 508 CRC patients (Figure R-3 2A). Kaplan-Meier analysis showed a strong significant association between high AIP expression and lower overall survival in the GSE17538 cohort ($p = 0.0035$). These results were validated with the COAD TCGA dataset containing 270 colon adenocarcinoma samples ($p = 0.0038$). Moreover, AIP protein expression was analysed using tissue microarrays containing 144 core tissue samples from CRC patients followed for more than 5 years and retrospectively selected (Figure R-3 2B) (41). AIP high expression

significantly correlated with CRC relapse and lower survival (p=0.0309) (Figure R-3 2B).

These results demonstrated an association between AIP expression and poor prognosis of CRC patients, besides its association with liver metastasis observed by proteomics.

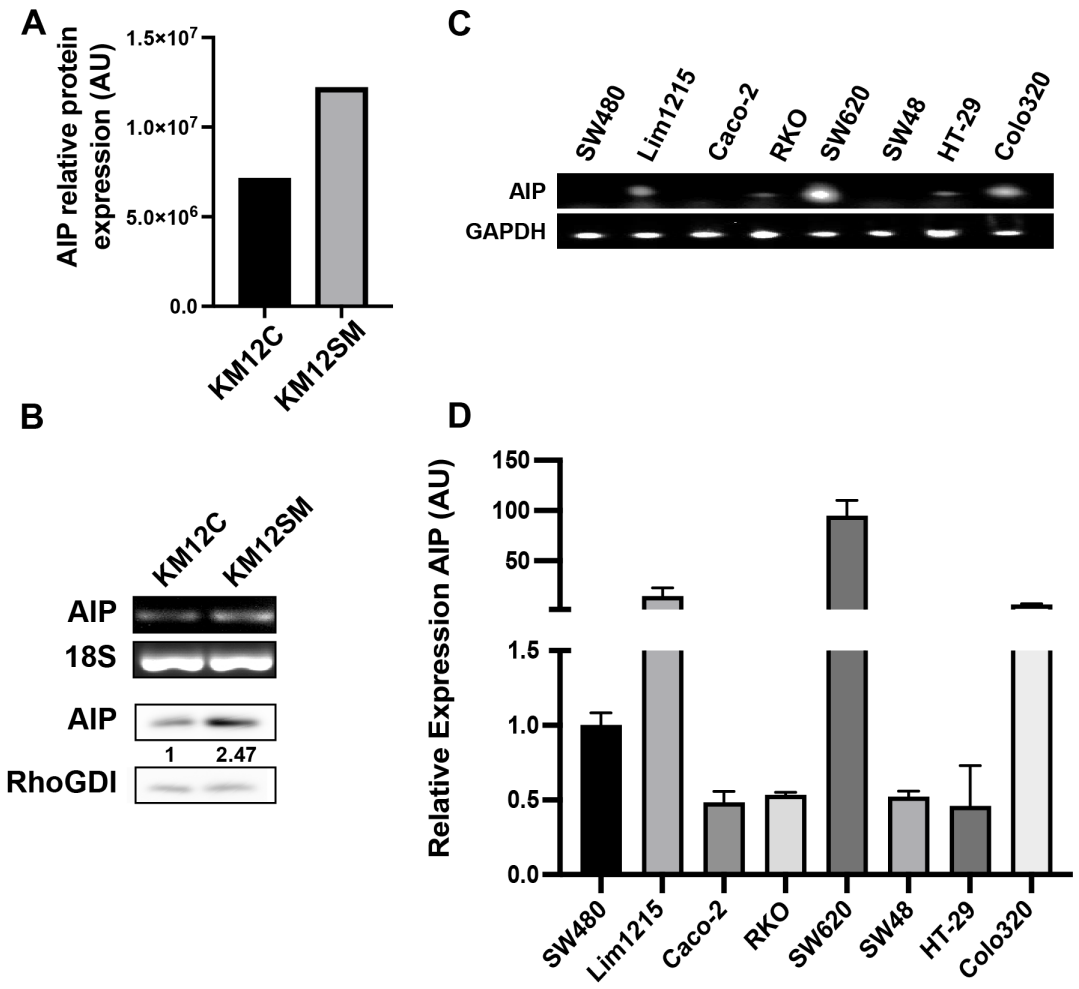


Figure R-3 1: Analysis of AIP expression in CRC cell lines. (**A**) AIP protein expression levels depicted as bar graph were higher in KM12SM liver metastatic cells than in the poorly metastatic KM12C colon cancer cells by spatial proteomics. (**B**) AIP mRNA and protein expression levels were assessed by semi-quantitative PCR and WB analyses using 18S and RhoGDI as controls, respectively. AU, arbitrary units. (**C**) AIP expression was assessed by semi-quantitative PCR in the eight indicated colon cancer cell lines using GAPDH as control. (**D**) qPCR analysis of AIP expression levels in the same CRC cell lines using 18S for normalization purposes.

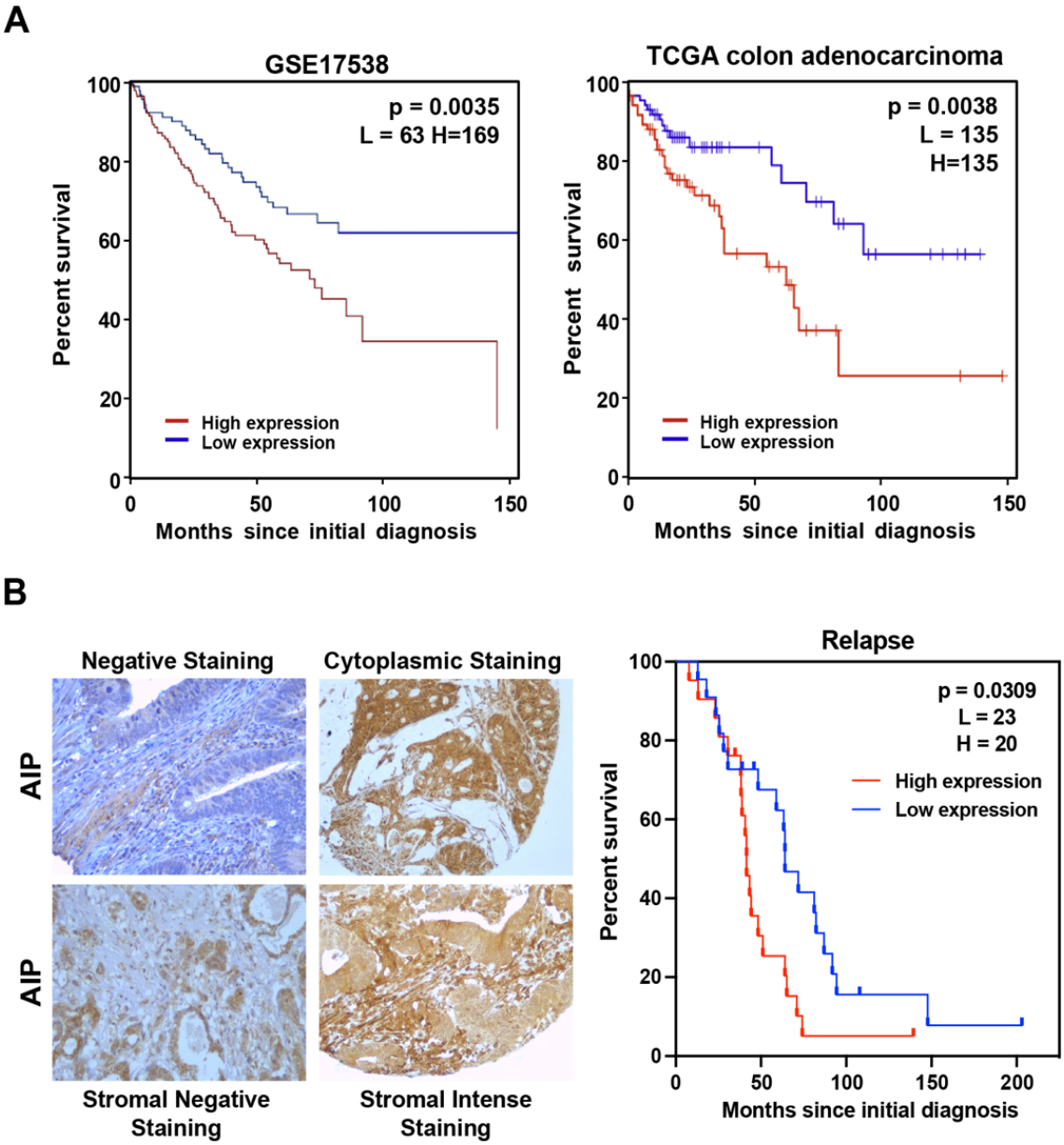


Figure R-3 2: (**A**) Kaplan–Meier analyses of overall survival of patients with colon cancer with the log-rank test according to the expression of AIP. Significant association of AIP expression gene with lower overall survival was found by comparing differences between high- versus low-expression groups with the log-rank test. The publicly available GSE17538 cohort containing colorectal cancer samples with clinicopathological data was used for the prognostic study. The prognostic value of AIP was independently assessed with a dataset containing 270 tumour samples from the cancer genome atlas (TCGA). (**B**) Immunohistochemical analysis of AIP expression in tissue microarrays showing representative images of weak, moderate or intense staining of different colon carcinomas. Counterstaining was made with hematoxylin. Pictures were taken at x100 or x200 magnification. Significant association of AIP tumoral stromal overexpression with poor survival was found with the log-rank test.

AIP overexpression promotes adhesion, colony formation, migration and invasion of colorectal cancer cells

To address the role of AIP overexpression in tumorigenesis and metastasis, we studied the effect of stably overexpressing AIP in the KM12 cell model of CRC liver metastasis in comparison to Mock-stably transfected control cells. A significant increase in AIP expression was observed in KM12 cells stably transfected with AIP by WB (Figure R-3 3A), RT-PCR (Figure R-3 3B), and immunofluorescence (Figure R-3 3C). AIP stable transfection effect was more pronounced on KM12C, which showed less AIP protein expression by proteomics, than on KM12SM cells.

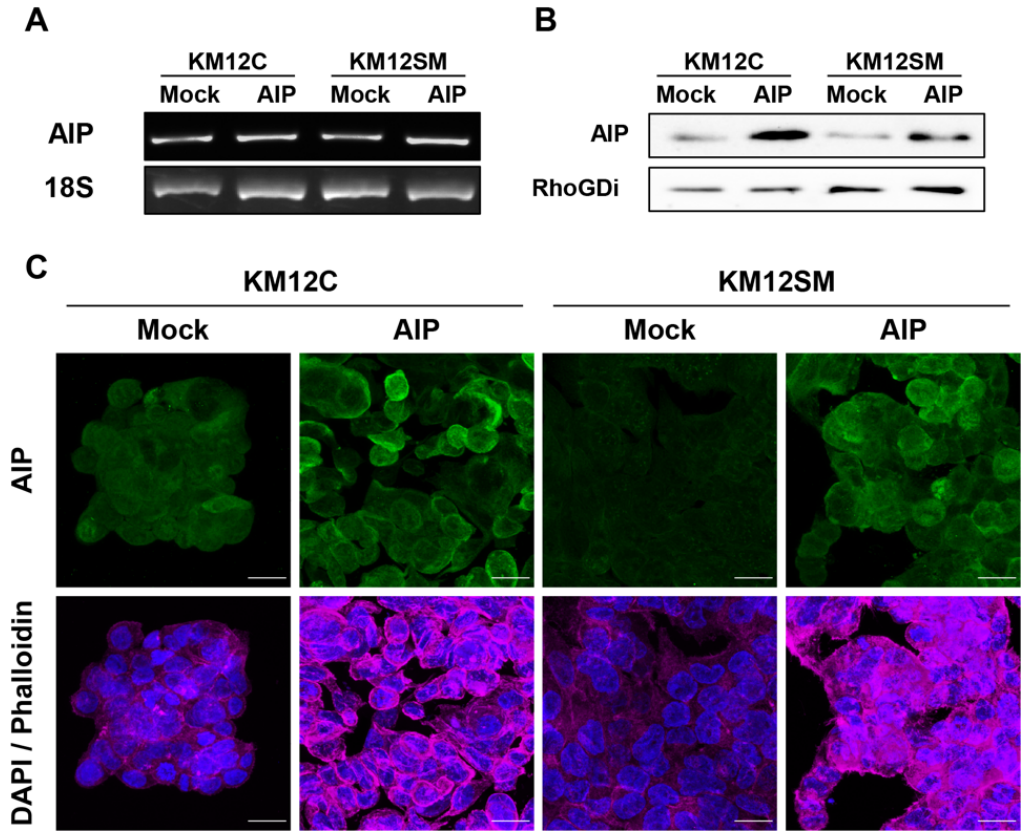


Figure R-3 3: (**A**) Semi-quantitative PCR analysis of AIP expression. 18S was used as total RNA content control. (**B**) WB analysis of AIP in KM12C and KM12SM cells stably expressing AIP and Mock confirmed the ectopic expression of AIP in both cell lines (**C**) Analysis of the expression of AIP in AIP- and Mock-stably transfected cell lines by Confocal Microscopy. Representative images show AIP (green), F-actin (magenta) and cell nucleus (blue). Scale bar: 20 μm.

Next, the tumorigenic (proliferation and colony formation) and metastatic (adhesion, migration and invasion) properties of the AIP- and Mock-stably transfected KM12 cells were investigated. AIP-overexpressing KM12 cells showed non-significant changes in proliferation (Figure R-3 4A). Regarding colony-forming ability, AIP-stably transfected KM12C and KM12SM cells showed a 2.5-fold increase and about 3.5-fold, respectively, respect to Mock cells. Similar colony formation ability for AIP-stably transfected KM12C cells was observed in comparison to Mock KM12SM cells (Figure R-3 4B).

Then, the adhesive properties of the cell lines were analysed using Matrigel assays. AIP induced a striking 5-fold higher adhesion capacity in KM12C cells. In KM12SM cells, a 155% increase in their adhesion capacity was observed (Figure R-3 4C). Remarkably, AIP ectopic expression induced KM12C and KM12SM cells to have a similar adhesion capacity. AIP-stably transfected cells also showed increased migration and invasive capacities. A significant increase in the invasive properties induced by AIP, more pronounced on KM12C cells, was found on KM12 cells (Figure R-3 4D). Changes in migration, assessed via wound healing assays in 24-well plates coated with Matrigel, showed that KM12 Mock controls were unable to close the wound (Figure R-3 5). On the other hand, AIP-overexpressing cells showed a steep increase in migration speed, more accentuated in KM12C cells.

Collectively, these results demonstrate that AIP significantly augments the tumorigenic and metastatic properties of KM12 CRC cells. The alterations were more evident for the poorly metastatic KM12C cells, which upon AIP-stable transfection increased their tumoral and metastatic properties nearby to KM12SM Mock cells or surpassed them regarding invasive and migration capacities.

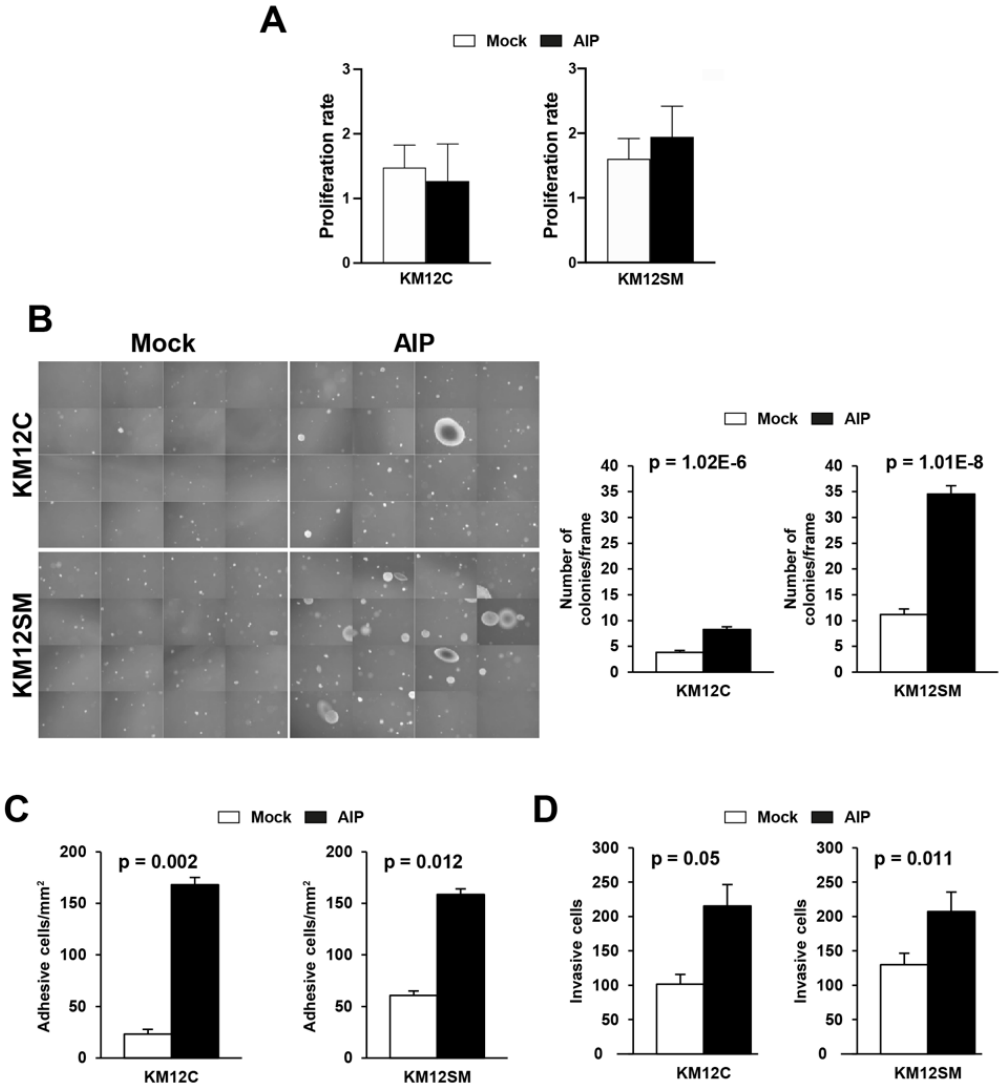


Figure R-3 4: AIP-ectopic expression in colorectal cancer cells increases tumorigenic and metastatic properties of colorectal cancer cells. (**A**) Proliferation was determined by MTT assays after 96 h of culture. A non-significant slightly decreased optical density in AIP-overexpressing KM12C and KM12SM cells was observed in comparison to Mock control cells. (**B**) AIP-ectopic expression induces the formation of soft agar colonies in KM12C and KM12SM cells in comparison to Mock control cells. Single cell suspensions of AIP- and Mock-stably transfected KM12C and KM12SM cells were seeded in soft agar and allowed to form colonies for 21 days in 6-well plates. Then, colonies were visualized by microscopy by taking photographs of 16 random fields. The average colony number per frame was represented by bar graphs. (**C**) Cell adhesion to Matrigel of AIP- or Mock-stably transfected, after starving cells for 5 h in medium alone. (**D**) KM12C and KM12SM ectopically expressing AIP showed approximately 2-fold higher invasion than Mock-stably transfected cells.

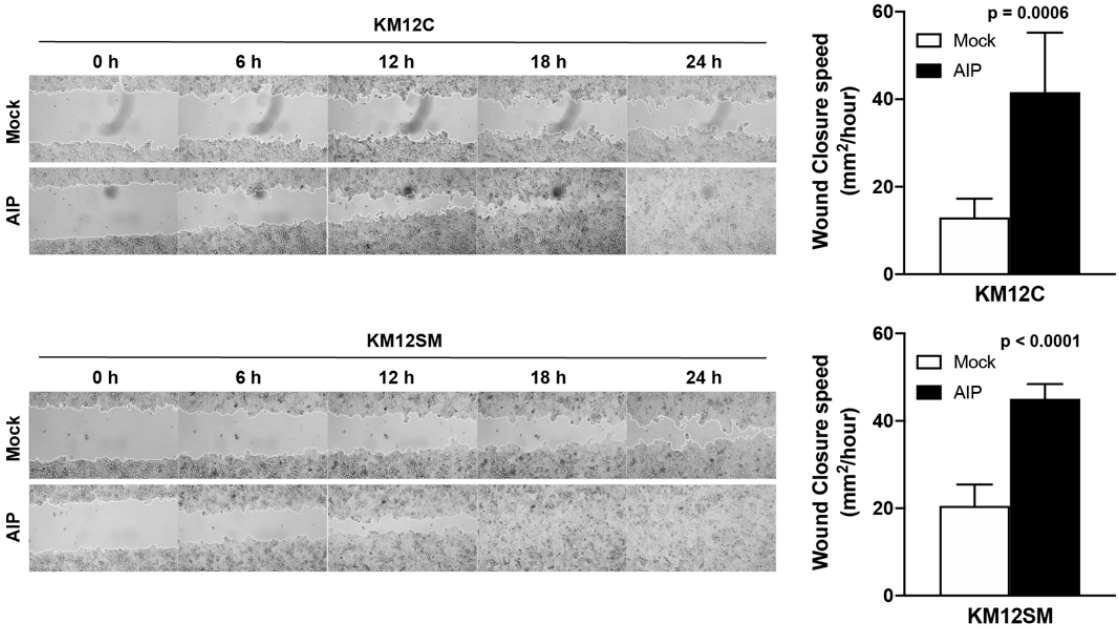


Figure R-3 5: Migration analysis of AIP-stably expressing and mock control cells. Cells were grown until confluence and their migratory capabilities were analysed in a wound-healing assay every 90 minutes. KM12C and KM12SM cells migratory capabilities were significantly enhanced by AIP ectopic expression. Representative images of the wound-healing assay are shown. Migration speed ($\mu m^2/h$) of AIP and Mock control cells was calculated as the distance covered every 24 h. Data for all the experiments represent the mean±SD of 3 independent experiments. p values of all the experiments are shown.

AIP overexpression modifies the expression of inducers of the mesenchymal phenotype

Since cell adhesion, migration and invasive capacity of epithelial cells correlate with the epithelial-to-mesenchymal transition (EMT), we investigated for alterations in EMT inducers. We studied changes in the mRNA expression levels of Snail1 (SNA1), ZEB1, TGFβ1, Claudin-2, ZO-1, and E-Cadherin (CDH1) by semi-quantitative PCR and qPCR analyses. In KM12C and KM12SM cells, AIP overexpression caused a significant alteration in the EMT inducers TGFβ1, Snail1 and ZEB1 accompanied by a large decrease in the epithelial marker CDH1 (Figure R-3 6A). Semi-quantitative PCR results for Tjp1 (ZO-1), CDH1, Snail1, ZEB1, TGFβ1 and Cldn2 were further assessed by qPCR analysis (Figure R-3 6B). Collectively, AIP overexpression induced in KM12 cells a large decrease in ZEB1, Snail1 and CDH1, and an increase in TGFβ1. Furthermore, AIP ectopic expression altered the expression of the mRNAs of the tight junctions' proteins Tjp1 (ZO-1) and Claudin.

At protein level, Snail1, ZO-1 and E-cadherin confirmed mRNA results (Figure R-3 6C). We observed a considerable decrease of E-cadherin together with a noticeable increase in N-cadherin supporting their opposite dysregulation (42), indicative of a reduction of the mesenchymal phenotype. Observed changes in E-Cadherin and adherens junctions (ZO-1) upon AIP ectopic expression were also confirmed by IF (Figure R-3 7). Differences in E-cadherin expression were more evident in the cell membrane, suggesting that AIP facilitated N-cadherin expression and the suppression of functional E-cadherin on the cell surface. Collectively, these data confirm that AIP induces a significant alteration on EMT effectors.

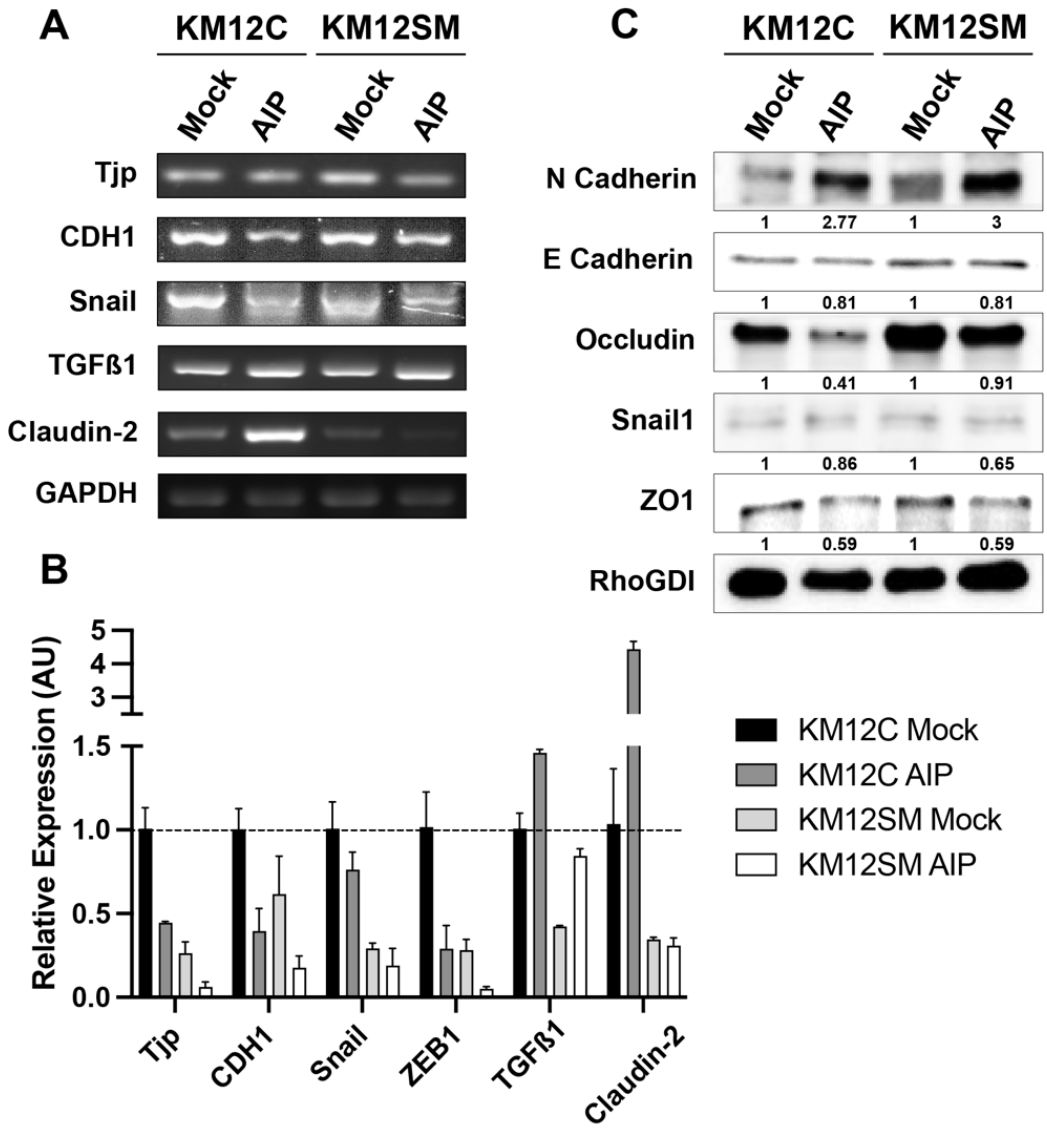


Figure R-3 6: Alterations in EMT inducers after AIP-ectopic expression in colorectal cancer cells. (**A**) cDNA synthesized from total RNA from AIP- and Mock-stably transfected cells was subjected to semi-quantitative RT-PCR using specific primers (Table S5) for the EMT inducers Tjp (ZO-1), CDH1, Snail, TGFß1 and Cluadin-2, using GAPDH for normalization. (**B**) cDNA synthesized from total RNA from AIP- and Mock-stably transfected cells was subjected to qPCR analysis using specific primers (Table S5) for the EMT inducers Tjp1 (ZO-1), CDH1, Snail, ZEB1, TGFβ1 and Claudin-2 using 18S for normalization. Data are shown as mean±SD. (**C**) KM12C and KM12SM Mock and AIP cells were lysed and subjected to WB analysis using specific antibodies against the indicated proteins and EMT markers. The abundance of each protein was quantified by densitometry. RhoGDI was used as loading control.

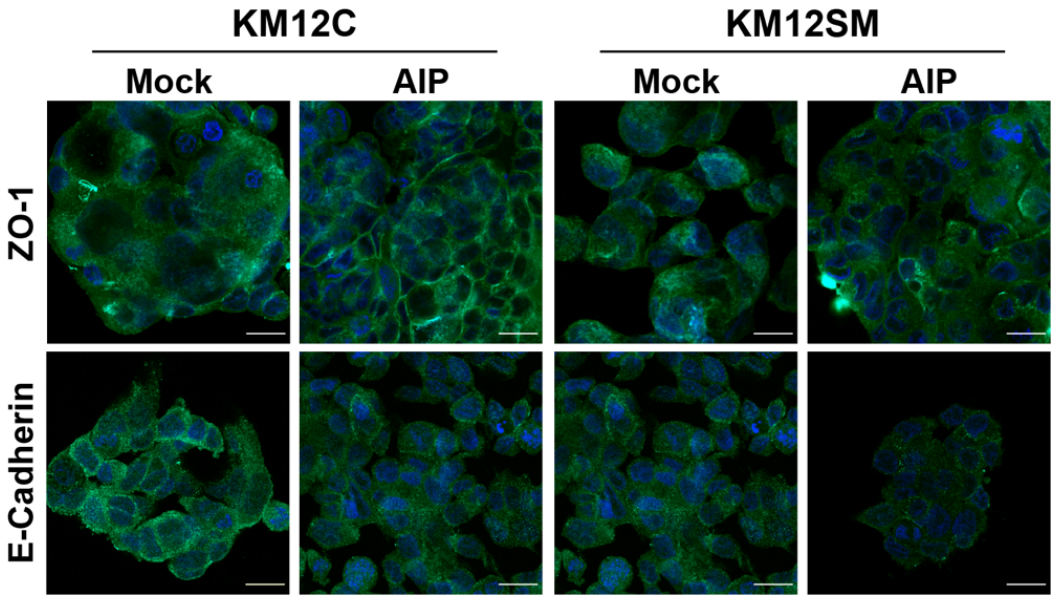


Figure R-3 7: Alterations in E-cadherin and ZO-1, EMT markers, after AIP-ectopic expression in colorectal cancer cells. Immunofluorescence analysis of ZO-1, E-Cadherin in Mock and AIP-stably transfected KM12C and KM12SM cells. DAPI was used for counterstaining of the nucleus in blue. Scale bar: 20 μm.

Signalling analysis in AIP-stably transfected colorectal cancer cells

Then, the effect of AIP ectopic expression on signalling pathways associated with effects on tumorigenic and metastatic properties was analysed. We observed a significant activation of p SRC, p-JNK, and p-AKT in AIP-overexpressed KM12 cells (Figure R-3 8). Surprisingly, although the proliferation levels of AIP and mock controls were similar, we detected a noticeable reduction of p-ERK1/2. Likewise, we also observed a decrease in p-FAK despite the increase in adhesion we had observed for AIP overexpressing cells. These changes suggest an effect mediated by AIP through AKT on EMT and cell survival, and SRC and JNK on cell migration, adhesion, and invasion, which may play a role in advanced CRC facilitating liver metastatic colonization.

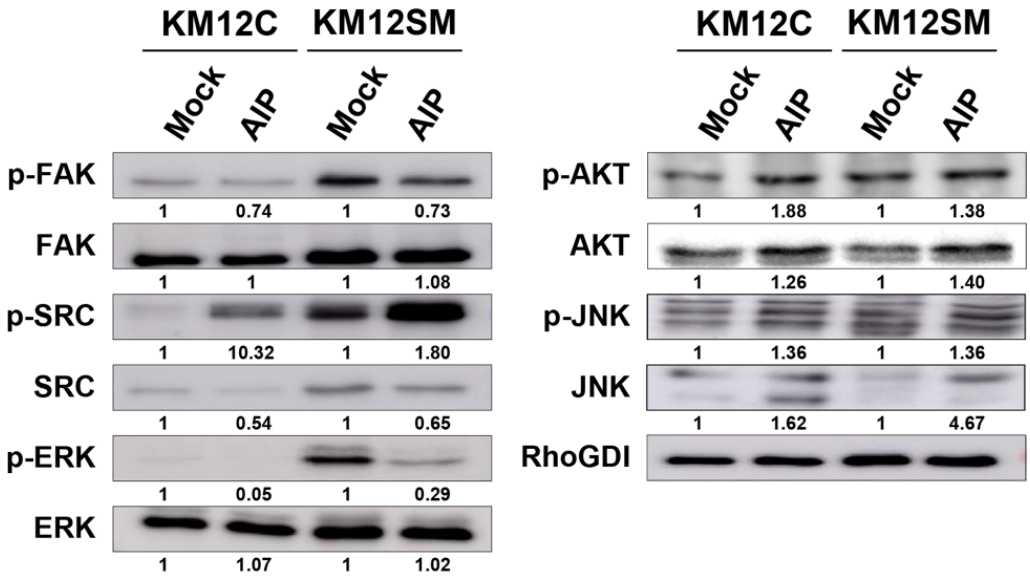


Figure R-3 8: Analysis of alterations in EMT inducers by WB after AIP-ectopic expression in colorectal cancer cells. KM12C and KM12SM Mock and AIP cells were lysed and subjected to WB analysis using specific antibodies against the indicated proteins and EMT markers. The abundance of each protein was quantified by densitometry. RhoGDI was used as loading control.

Identification of proteins affected by AIP overexpression on KM12 colorectal cancer cells by proteomics

We then carried out a proteomic approach to identify AIP-modulated proteins and characterize their interaction network. AIP-stably transfected KM12 lysed cells were analysed by quantitative TMT proteomics analysis. As control, we included in the assay Mock and parental cells. After data normalization a total of 3124 proteins were identified and quantified with at least one peptide (Table S12). Among them, 569 proteins identified and quantified with two or more peptides showed upregulation or downregulation because of AIP-overexpression with a fold change ≥ 1.5 or ≤ 0.67 (Figure R-3 9A-B and Table S13). To identify proteins highly specifically modulated by AIP, we focused on those proteins commonly dysregulated in KM12C and KM12SM cells. In total, we found 60 proteins up- or down-regulated (Figure R-3 9B and Table R-3 1). As expected, AIP was among the overexpressed proteins.

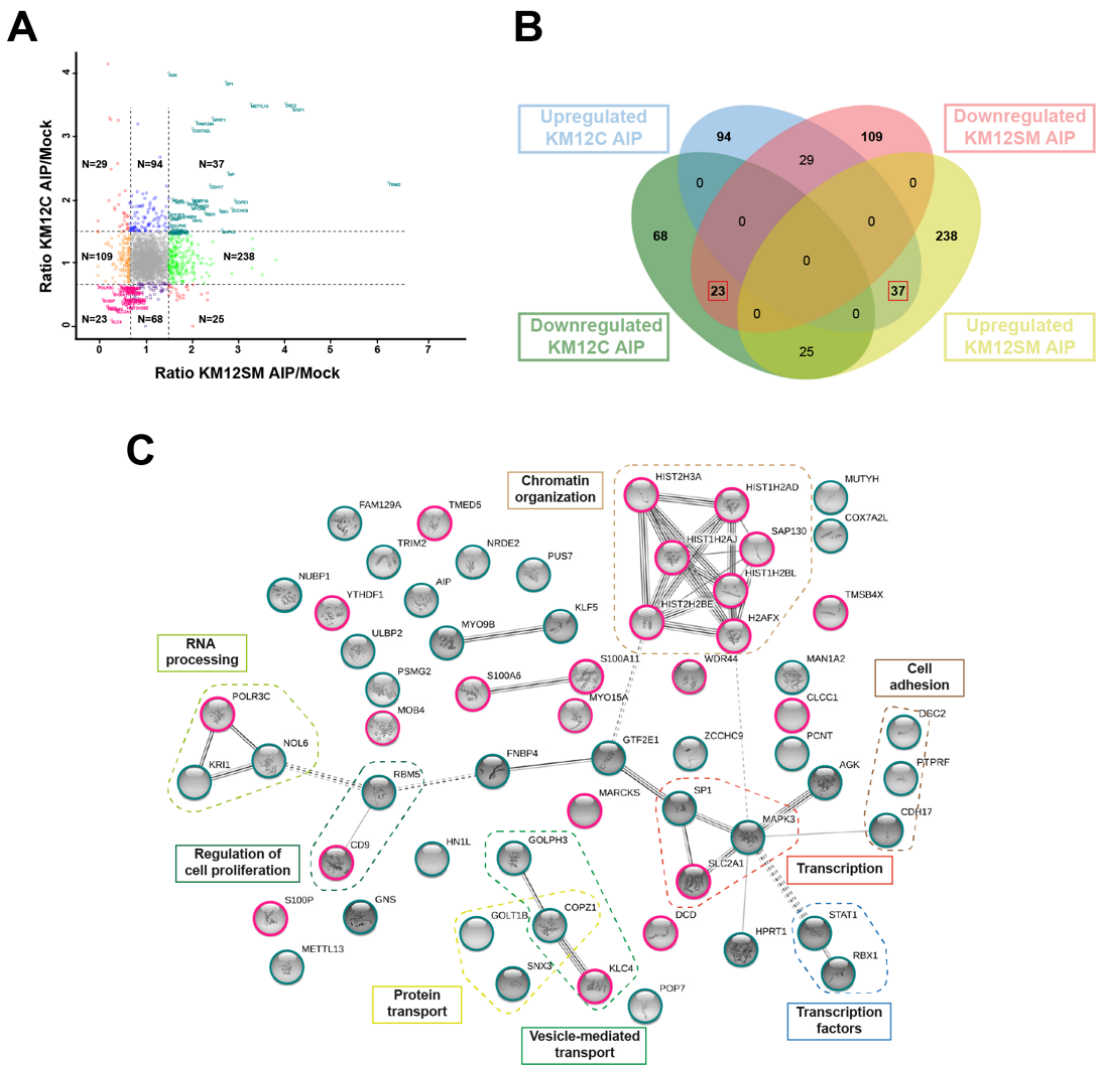


Figure R-3 9: Mass spectrometry analysis of protein alterations modulated by AIP overexpression in CRC cells. (**A**) Correlation scatterplot of proteins altered in AIP-overexpressing KM12C and KM12SM cells compared to Mock controls. Coloured dots represent differentially expressed proteins upregulated (green) or downregulated (orange) in AIP-overexpressing KM12SM; and upregulated (blue) and downregulated (orange) in AIP-overexpressing KM12C with 1.5-fold expression difference (dashed lines). Proteins altered in both AIP-overexpressing KM12C and KM12SM cells are represented in bluish green (upregulated) and pink (downregulated) dots, whereas proteins showing opposite alteration cells are represented in red. (**B**) Venn Diagram of the proteins dysregulated in AIP-overexpressed KM12C and KM12SM cells. Red squares, the 60 proteins commonly dysregulated in both AIP-overexpressed cell lines. (**C**) Protein interactome map of the proteins significantly dysregulated by AIP in both KM12C and KM12SM cells. Upregulated proteins by AIP overexpression are circled in bluish green, and downregulated proteins in pink

Table R-3 1: Proteins upregulated or downregulated in both AIP-overexpressed KM12C and KM12SM cells

Protein IDs	Protein names	Gene names	AIP-transfected cells
P06703	Protein S100-A6	S100A6	Down
P11166	Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1	Down
P16104	Histone H2AX	H2AFX	Down
P20671	Histone H2A type 1-D	HIST1H2AD	Down
P21926	CD9 antigen	CD9	Down
P25815	Protein S100-P	S100P	Down
P29966	Myristoylated alanine-rich C-kinase substrate	MARCKS	Down
P31949	Protein S100-A11; Protein S100-A11, N-terminally processed	S100A11	Down
P62328	Thymosin beta-4; Hematopoietic system regulatory peptide	TMSB4X	Down
P81605	Dermcidin; Survival-promoting peptide; DCD-1	DCD	Down
Q16778	Histone H2B type 2-E; Histone H2B type 1-B; Histone H2B type 1-O; Histone H2B type 1-J	HIST2H2BE	Down
Q5JSH3	WD repeat-containing protein 44	WDR44	Down
Q71DI3	Histone H3.2	HIST2H3A	Down
Q96S66	Chloride channel CLIC-like protein 1	CLCC1	Down
Q99878	Histone H2A type 1-J	HIST1H2AJ	Down
Q99880	Histone H2B type 1-L	HIST1H2BL	Down
Q9BUI4	DNA-directed RNA polymerase III subunit RPC3	POLR3C	Down
Q9BYJ9	YTH domain-containing family protein 1	YTHDF1	Down
Q9H0E3	Histone deacetylase complex subunit SAP130	SAP130	Down
Q9NSK0	Kinesin light chain 4	KLC4	Down
Q9UKN7	Unconventional myosin-XV	MYO15A	Down
Q9Y3A3	MOB-like protein phocein	MOB4	Down
Q9Y3A6	Transmembrane emp24 domain-containing protein 5	TMED5	Down
O00170	**AH receptor-interacting protein**	**AIP**	**Up**

O14548	Cytochrome c oxidase subunit 7A-related protein, mitochondrial	COX7A2L	Up
O60476	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IB	MAN1A2	Up
O60493	Sorting nexin-3	SNX3	Up
O75817	Ribonuclease P protein subunit p20	POP7	Up
O95613	Pericentrin	PCNT	Up
P00492	Hypoxanthine-guanine phosphoribosyltransferase	HPRT1	Up
P08047	Transcription factor Sp1	SP1	Up
P10586	Receptor-type tyrosine-protein phosphatase F	PTPRF	Up
P15586	N-acetylglucosamine-6-sulfatase	GNS	Up
P27361	Mitogen-activated protein kinase 3	MAPK3	Up
P29083	General transcription factor IIE subunit 1	GTF2E1	Up
P42224	Signal transducer and activator of transcription 1-alpha/beta	STAT1	Up
P52756	RNA-binding protein 5	RBM5	Up
P53384	Cytosolic Fe-S cluster assembly factor NUBP1	NUBP1	Up
P61923	Coatomer subunit zeta-1	COPZ1	Up
P62877	E3 ubiquitin-protein ligase RBX1; E3 ubiquitin-protein ligase RBX1, N-terminally processed	RBX1	Up
Q02487	Desmocollin-2	DSC2	Up
Q12864	Cadherin-17	CDH17	Up
Q13459	Unconventional myosin-IXb	MYO9B	Up
Q13887	Krueppel-like factor 5; Krueppel-like factor 6; Krueppel-like factor 7; Krueppel-like factor 2; Krueppel-like factor 1; Krueppel-like factor 4	KLF5	Up
Q53H12	Acylglycerol kinase, mitochondrial	AGK	Up
Q8N3X1	Formin-binding protein 4	FNBP4	Up
Q8N567	Zinc finger CCHC domain-containing protein 9	ZCCHC9	Up
Q8N6R0	Methyltransferase-like protein 13	METTL13	Up
Q8N9T8	Protein KRI1 homolog	KRI1	Up
Q969U7	Proteasome assembly chaperone 2	PSMG2	Up
Q96PZ0	Pseudouridylate synthase 7 homolog	PUS7	Up
Q9BZM5	NKG2D ligand 2	ULBP2	Up
Q9BZQ8	Protein Niban	FAM129A	Up

Q9C040	Tripartite motif-containing protein 2	TRIM2	Up
Q9H4A6	Golgi phosphoprotein 3	GOLPH3	Up
Q9H6R4	Nucleolar protein 6	NOL6	Up
Q9H7Z3	Protein NRDE2 homolog	NRDE2	Up
Q9H910	Hematological and neurological expressed 1-like protein	HN1L	Up
Q9UIF7	A/G-specific adenine DNA glycosylase	MUTYH	Up
Q9Y3E0	Vesicle transport protein GOT1B	GOLT1B	Up

Using STRING and data mining (43), proteins were classified into eight clusters of interaction, including proteins related to cell adhesion, cell cycle and regulation of proliferation, transport, transcription, chromatin organization, transcription factors and RNA processing (Figure R-3 9C). Using Reactome (44), altered processes due to AIP ectopic expression were observed. Among them, AIP was observed to induce changes in DNA repair (p-value = $4.08 \cdot 10^{-4}$), cell cycle (p-value = $1.34 \cdot 10^{-2}$), gene expression (p-value = $4.57 \cdot 10^{-2}$) and metabolism of proteins (mainly protein deubiquitination (p-value = $2.26 \cdot 10^{-2}$) and asparagine N-glycosylation (p-value = $2.81 \cdot 10^{-2}$)). Moreover, the dysregulation of all the cell adhesion proteins upregulated by AIP -CDH17, DSC2 and PTPRF- has been described to increase metastasis, *in vivo* homing and proliferation, while also contributing to adherens junctions redistribution and cytoskeletal rearrangement (45-48). Furthermore, AIP was found to upregulate the transcription factors SP1 and STAT1, which play a major role in cancer and metastatic progression (49-51).

WB and IF analyses of selected targets confirmed the dysregulation of indicated proteins (Figure R-3 10A-B). The overexpression observed by proteomics of EGFR, p38, GOLPH3, MUTYH, STAT1 and CDH17 due to the ectopic expression of AIP in KM12C and KM12SM cells was confirmed, whereas SP1 overexpression could be only validated in KM12SM cells. In addition, AIP was able to activate p-p38, in contrast to p-STAT1, whose expression decreased in parallel with AIP expression. Moreover, the expression of AIP receptor -Aryl-hydrocarbon receptor (AHR)- was also analysed by WB. It was observed that the ectopic expression of AIP also induced its expression in KM12C and KM12SM cells. Remarkably, IF analysis showed that the ectopic expression of AIP induced changes in the abundance and the localization of CDH17, with a recruitment of CDH17 to the plasmatic membrane (Figure R-3 10B).

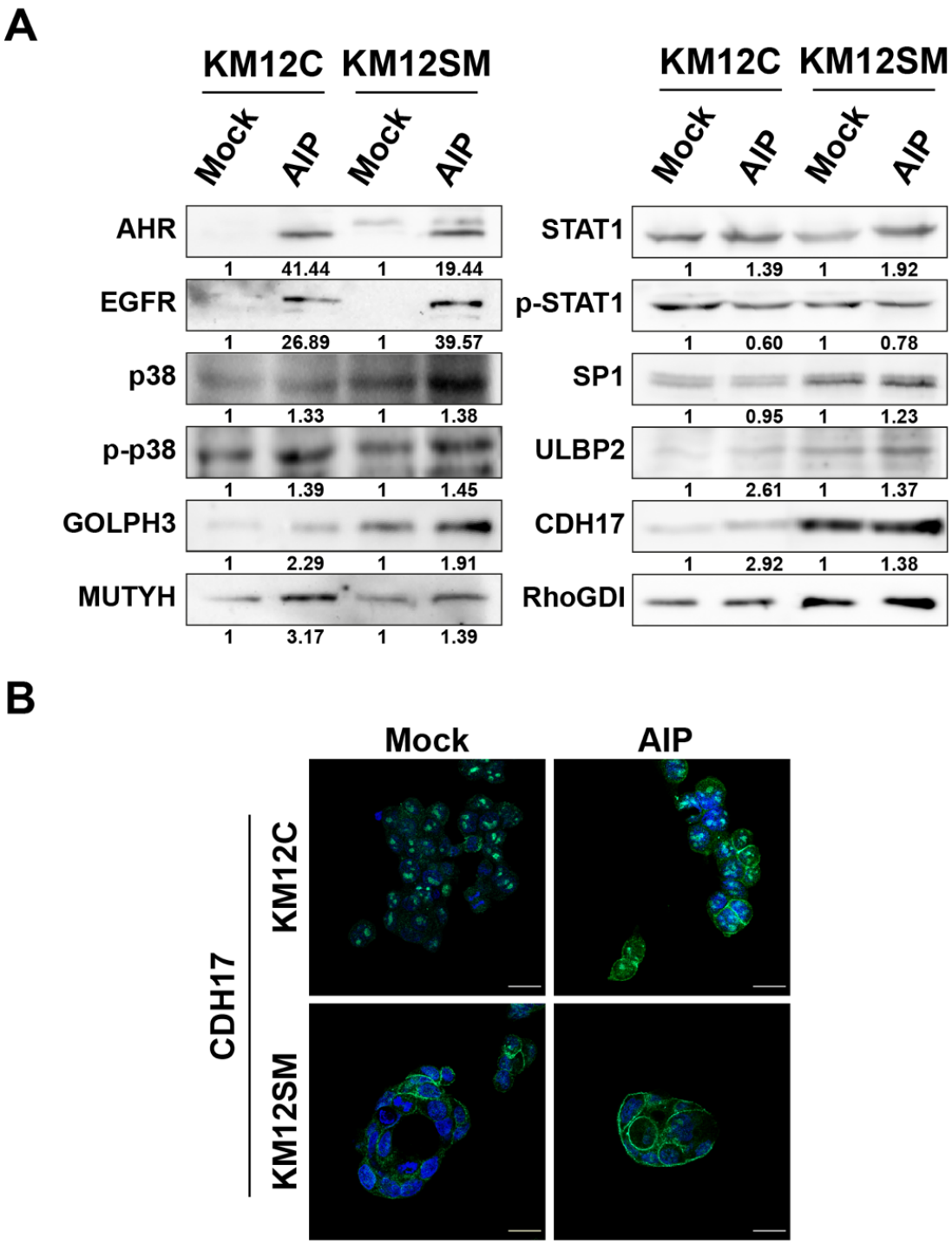


Figure R-3 10: (**A**) Verification of the dysregulation of AIP-associated proteins identified and quantified by proteomics by WB analysis for verification of protein alterations using optimized dilutions of the antibodies. RhoGDI was used as loading control in the same gels. (**B**) Immunofluorescence analysis of CDH17 in Mock and AIP-stably transfected KM12C and KM12SM cells. DAPI was used for counterstaining of the nucleus in blue. Scale bar: 20 μm.

AIP induces *in vivo* tumour growth, liver metastasis and decreases mice survival

Finally, we investigated the *in vivo* effects of AIP ectopic expression. First, we examined its effects on the capacity of KM12 cells inoculated in the spleen of nude mice for liver homing. As a surrogate marker for homing, human GAPDH was highly detected in the livers of mice inoculated with AIP-stably transfected KM12 cells in comparison to Mock cells (Figure R-3 11A). More importantly, AIP ectopic expression induced in KM12C cells the ability to colonize liver.

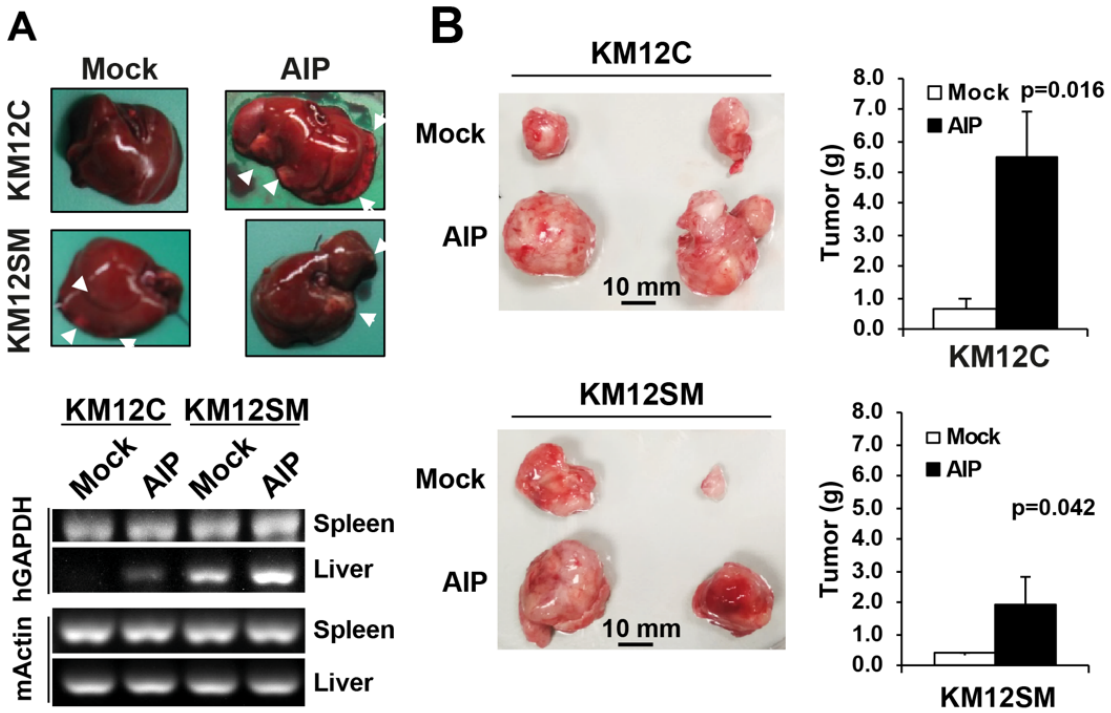


Figure R-3 11: AIP ectopic expression induces tumor growth and liver metastasis in KM12 cells. (**A**) Nude mice intrasplenically inoculated with indicated KM12 cells were sacrificed 24 hours after inoculation for analysis of *in vivo* homing. RNA was isolated from the liver and subjected to RT-PCR to amplify human GAPDH (hGAPDH). Representative experiments out of 3 are shown. Murine β-actin (mβ-actin) was amplified as control. (**B**) KM12 transfectants were inoculated subcutaneously in nude mice (n=6 per group). Tumor size was measured every day for 4 weeks and mean±SEM of the endpoint represented by bar graphs. Representative tumor images from KM12 cell stably transfectants are also depicted.

Then, the role of AIP in CRC tumour growth, metastasis and survival was investigated by subcutaneous and intrasplenic cell inoculations. AIP-stably transfected cells developed significantly higher measurable tumours after subcutaneous inoculation than Mock KM12 cells (Figure R-3 11B). In addition, mice inoculated intrasplenically with AIP-stably

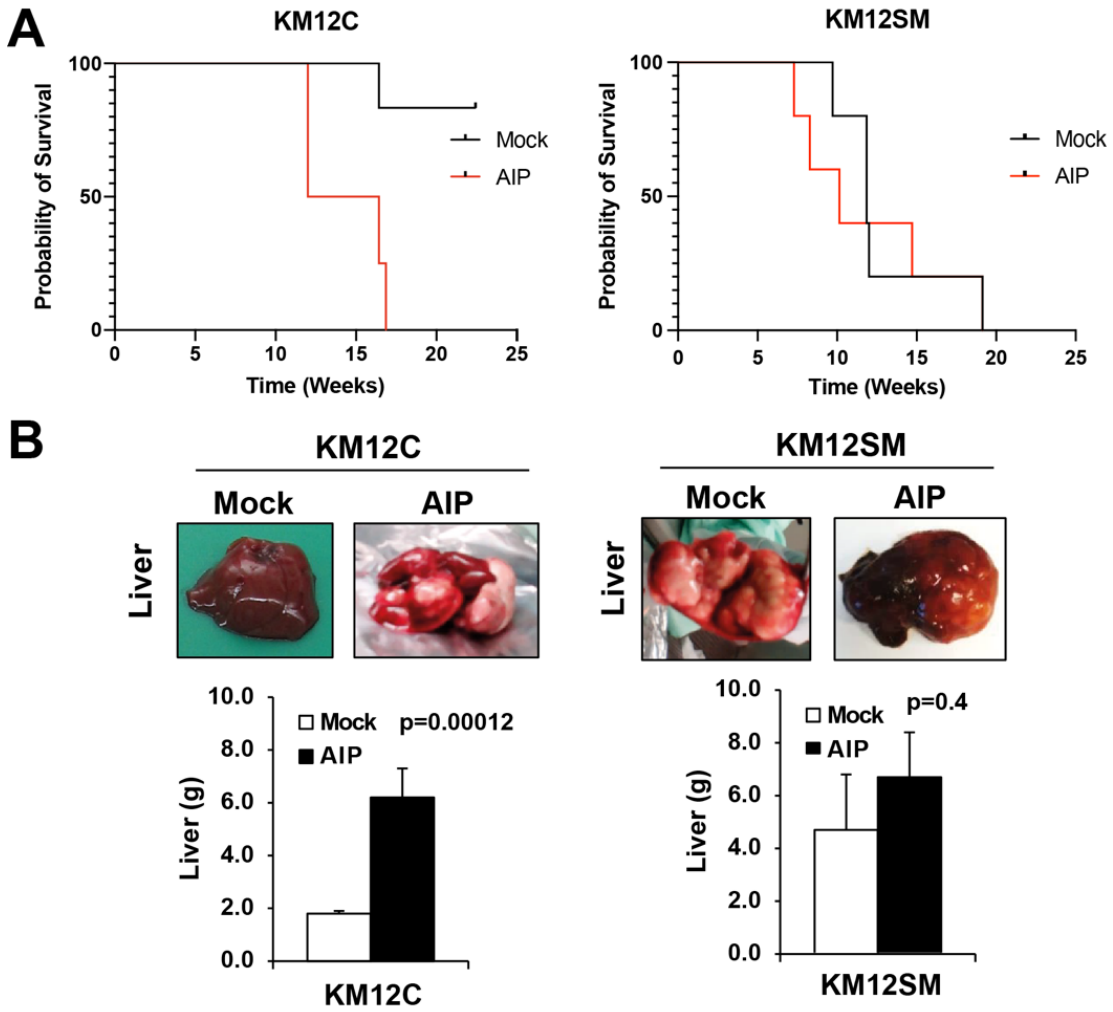


Figure R-3 12: AIP ectopic expression induces lower survival associated to liver metastasis growth in KM12 cells. (**A**) Kaplan-Meier survival curve of nude mice inoculated intrasplenic with the indicated KM12 cell transfectants. Survival of mice inoculated with AIP-stably transfected cells significantly decreased ($p < 0.01$) when compared with those inoculated with Mock control cells. (**B**) Mice were examined for macroscopic metastases in liver. Representative images of macroscopic metastases in the liver are shown. Livers were weighted at endpoint and mean±SEM represented.

transfected KM12C cells showed shorter survival than those inoculated with Mock control cells (Figure R-3 12A). Importantly, Mock KM12C control cells did not develop any metastasis to liver as opposed to AIP-stably transfected KM12C cells. This reduced survival was associated with the AIP-induced higher capacity for liver colonization, as visually inspected and depicted by the mice's liver weight at endpoint (Figure R-3 12B). Regarding KM12SM cells, although no significant differences were observed in survival curves because AIP-stably transfected and Mock KM12SM cells develop liver metastasis (Figure R-3 12A), liver metastasis were considerably higher in AIP-stably transfected cells (Figure R-3 12B). Finally, to confirm that the observed differences were due to the ectopic expression of AIP in the injected cells, we analysed AIP protein abundance on tumour sections by immunohistochemistry (Figure R-3 13). We could observe a more intense AIP staining in the tumours of mice injected with AIP-stably expressing cells. In addition, AIP-stably expressing cells showed a clear increase in the marker for cellular proliferation KI67 (52) compared to Mock conditions (Figure R-3 13), indicating that tumours overexpressing AIP were more proliferative than their Mock counterparts.

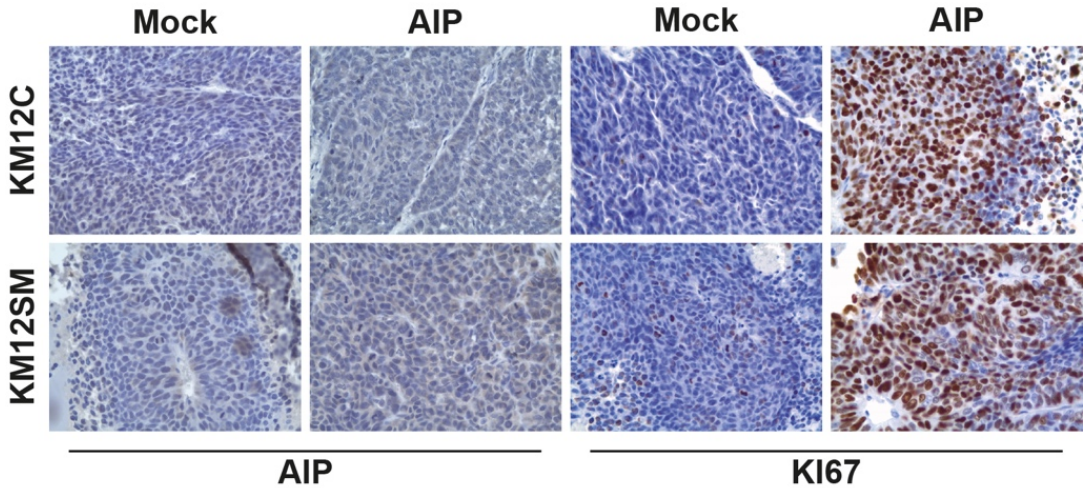


Figure R-3 13: Analysis of tumoral samples for AIP and Ki67 expression. Representative IHC images of tumours developed in the mice injected with either KM12C or KM12SM Mock or AIP-stably expressing cells. More intense AIP staining was observed in AIP-stably expressing cells.

Discussion

We here found that AIP is highly overexpressed in CRC metastatic cell lines and liver metastasis, and acts as a novel key player promoting CRC metastasis. AIP affected a plethora of transcription factors, cell adhesion molecules and signalling cascades leading to SRC, JNK and PI3K/AKT pathways' activation. Associated with these changes, AIP altered adhesion, colony formation, migration, and invasion capacity of cells. Remarkably, these effects together with AIP-associated dysregulation of multiple EMT factors induced non-metastatic KM12C cells to become metastatic to liver.

Germline AIP mutations have been strongly associated with familial isolated pituitary adenoma (53). However, AIP mutation's analysis in colorectal, breast and prostate cancer showed that the presence of somatic mutations is not a common finding (39). In contrast, AIP overexpression in tumours has been strongly associated with a poor outcome in gastric (54), pancreatic (55), and colorectal cancer patients (here presented data). Notably, AIP-induced alterations in the EMT process have been previously reported. Gene expression analysis of AIP germline-mutated pituitary adenomas showed that the EMT pathway was altered with 16 upregulated and 31 downregulated genes (56). Here, we observed that AIP induced alterations in EMT, as depicted by the downregulation of Snail, Slug, and E-cadherin and an increase of N-cadherin and TGFβ1 accompanied by CDH17 overexpression. Altogether, these observations suggest that AIP-mediated EMT dysregulation is a common event on AIP-mutated or AIP-overexpressing tumours.

Previous clinical data supported the role of AIP as a tumour suppressor since AIP germline mutated gene was associated with familial isolated pituitary adenoma (57). Germline mutational analysis of these adenomas detected 50 different pathogenic mutations leading to AIP disruption.

Moreover, wild-type AIP overexpression in human fibroblast and pituitary cell lines reduced cell proliferation *in vitro*, whereas the mutant AIP loses this ability compared to the wild-type AIP (37). However, AIP overexpression in gastrointestinal -colorectal, gastric and pancreatic cancers was associated with a worse prognosis, and in CRC cells produced a significant increase in overall metastatic capacities. In addition, this work showed the first association of AIP overexpression with liver homing and liver metastasis in CRC. Therefore, our results indicate that beyond its role as a tumour suppressor in pituitary adenomas, AIP acts as an oncogene in CRC.

Twenty interaction partners have been described for AIP, of which fourteen were confirmed to interact directly with AIP. These includes viral proteins (HBV X and EBNA-3), chaperones (Hsp90 and Hsc70), PDEs (PDE4A5 and PDE2A3), nuclear (AHR, PPARα and TRβ1) and transmembrane (RET) receptors, G proteins (Gα13 and Gαq), survivin and a mitochondrial import receptor (TOMM20) (58). Besides the initial interest drawn upon AIP and AHR in the 2000s, the interactome or the cellular pathways in which AIP is involved remain obscure. Here, we have shed some light on these processes by quantitative proteomics and orthogonal techniques. In this sense, AIP ectopic expression induced a vast protein dysregulation in KM12 cells, associated to chromatin organization, DNA repair, cell cycle, or signal transduction, among others. Notably, adhesion proteins -as CDH17- and transcription factors -as SP1 and STAT1- known to play central roles in cancer and metastasis were also dysregulated by AIP (45, 49-51). In addition to the vast number of proteins dysregulated by AIP, we found AIP in the cytoplasm and nucleus by IF and IHC from AIP-overexpressing cells and CRC tumoral tissue in Nude mice, respectively. Thus, it is plausible to think that beyond its interaction with AHR and the translocation of this receptor to the nucleus where it functions as a transcription factor (59, 60), AIP might act independently to AHR as a (direct or indirect) transcription factor itself

when overexpressed. Two of the most interesting proteins upregulated by AIP were the atypical cadherin CDH17 and EGFR. CDH17 facilitates cell clustering via homotypic cadherin interactions and cell adhesion through integrin activation to initiate micrometastasis formation (45). Therefore, AIP-mediated increase of CDH17 should facilitate liver adhesion and *in vivo* homing of CRC cells ectopically expressing AIP. On the other hand, previous reports have already established the role of EGFR in tumour onset and progression (61-63). Mutation or overexpression of EGFR leads to altered EGFR signalling, which in turn induces abnormal trafficking and contributes to increased signalling and tumour development (61).

In summary, this work demonstrates the value of AIP, previously identified from a spatial proteomic analysis of metastatic cells (10), and multidimensional proteomics for identifying relevant proteins in metastasis with actual value in CRC patients. Our conclusions were based on the following observations: (i) AIP high expression was associated with liver metastasis and poor overall survival, (ii) AIP ectopic overexpression increased cell adhesion, migration, invasion and colony formation in KM12 cells, particularly in non-metastatic KM12C cells, (iii) AIP overexpression induced the dysregulation of proteins related to the EMT transition or involved in cancer and metastasis, suggesting a transcription factor role for AIP, (iv) AIP overexpression also induced an increase in the phosphorylation levels of JNK, SRC and AKT, (v) AIP overexpression increased *in vivo* tumour growth and decreased mice survival, and (vi) AIP overexpression induced KM12C cells to acquire ability for liver colonization. Together, these data confirm that AIP has a key role in CRC and liver metastasis, possessing oncogene features.

References

1. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. Science. 2011;331(6024):1559-64.
2. Torres S, Bartolomé RA, Mendes M, Barderas R, Fernandez-Aceñero MJ, Peláez-García A, et al. Proteome Profiling of Cancer-Associated Fibroblasts Identifies Novel Proinflammatory Signatures and Prognostic Markers for Colorectal Cancer. Clinical Cancer Research. 2013;19(21):6006-19.
3. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19(11):1423-37.
4. Solís-Fernández G, Montero-Calle A, Martínez-Useros J, López-Janeiro Á, de los Ríos V, Sanz R, et al. Spatial Proteomic Analysis of Isogenic Metastatic Colorectal Cancer Cells Reveals Key Dysregulated Proteins Associated with Lymph Node, Liver, and Lung Metastasis. Cells. 2022;11(3).
5. Cai Z, Chiu JF, He QY. Application of proteomics in the study of tumor metastasis. Genomics Proteomics Bioinformatics. 2004;2(3):152-66.
6. Aleckovic M, Wei Y, LeRoy G, Sidoli S, Liu DD, Garcia BA, et al. Identification of Nidogen 1 as a lung metastasis protein through secretome analysis. Genes Dev. 2017;31(14):1439-55.
7. Song P, Bao H, Yu Y, Xue Y, Yun D, Zhang Y, et al. Comprehensive profiling of metastasis-related proteins in paired hepatocellular carcinoma cells with different metastasis potentials. Proteomics Clin Appl. 2009;3(7):841-52.
8. Sun B, Zhang S, Zhang D, Li Y, Zhao X, Luo Y, et al. Identification of metastasis-related proteins and their clinical relevance to triple-negative human breast cancer. Clin Cancer Res. 2008;14(21):7050-9.
9. Barderas R, Mendes M, Torres S, Bartolome RA, Lopez-Lucendo M, Villar-Vazquez R, et al. In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration, and invasion. Mol Cell Proteomics. 2013;12(6):1602-20.
10. Mendes M, Peláez-García A, López-Lucendo M, Bartolomé RA, Calviño E, Barderas R, et al. Mapping the Spatial Proteome of Metastatic Cells in Colorectal Cancer. Proteomics. 2017;17(19):1700094-.
11. Morikawa K, Walker SM, Jessup JM, Fidler IJ. In vivo selection of highly metastatic cells from surgical specimens of different primary human colon carcinomas implanted into nude mice. Cancer Research. 1988;48(7):1943-8.

12. Morikawa K, Walker SM, Nakajima M, Pathak S, Jessup JM, Fidler IJ. Influence of organ environment on the growth, selection, and metastasis of human colon carcinoma cells in nude mice. Cancer Research. 1988;48(23):6863-71.
13. Li A, Varney ML, Singh RK. Constitutive expression of growth regulated oncogene (gro) in human colon carcinoma cells with different metastatic potential and its role in regulating their metastatic phenotype. Clin Exp Metastasis. 2004;21(7):571-9.
14. Kuniyasu H, Ohmori H, Sasaki T, Sasahira T, Yoshida K, Kitadai Y, et al. Production of interleukin 15 by human colon cancer cells is associated with induction of mucosal hyperplasia, angiogenesis, and metastasis. Clin Cancer Res. 2003;9(13):4802-10.
15. Calon A, Espinet E, Palomo-Ponce S, Tauriello DV, Iglesias M, Cespedes MV, et al. Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. Cancer Cell. 2012;22(5):571-84.
16. Barderas R, Bartolome RA, Fernandez-Acenero MJ, Torres S, Casal JI. High expression of IL-13 receptor alpha2 in colorectal cancer is associated with invasion, liver metastasis, and poor prognosis. Cancer Res. 2012;72(11):2780-90.
17. Hegde P, Qi R, Gaspard R, Abernathy K, Dharap S, Earle-Hughes J, et al. Identification of Tumor Markers in Models of Human Colorectal Cancer Using a 19,200-Element Complementary DNA Microarray1. Cancer Research. 2001;61(21):7792-7.
18. Ellens KW, Christian N, Singh C, Satagopam VP, May P, Linster CL. Confronting the catalytic dark matter encoded by sequenced genomes. Nucleic Acids Res. 2017;45(20):11495-514.
19. Perdigão N, Heinrich J, Stolte C, Sabir KS, Buckley MJ, Tabor B, et al. Unexpected features of the dark proteome. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(52):15898-903.
20. Ma Q, Whitlock JP. A Novel Cytoplasmic Protein That Interacts with the Ah Receptor, Contains Tetratricopeptide Repeat Motifs, and Augments the Transcriptional Response to 2,3,7,8-Tetrachlorodibenzo-p-dioxin *. Journal of Biological Chemistry. 1997;272(14):8878-84.
21. Carver LA, Bradfield CA. Ligand-dependent Interaction of the Aryl Hydrocarbon Receptor with a Novel Immunophilin Homolog in Vivo*. Journal of Biological Chemistry. 1997;272(17):11452-6.

22. Tsuji N, Fukuda K, Nagata Y, Okada H, Haga A, Hatakeyama S, et al. The activation mechanism of the aryl hydrocarbon receptor (AhR) by molecular chaperone HSP90. FEBS Open Bio. 2014;4:796-.
23. Larigot L, Juricek L, Dairou J, Coumoul X. AhR signaling pathways and regulatory functions. Biochimie Open. 2018;7:1-.
24. Petrulis JR, Perdew GH. The role of chaperone proteins in the aryl hydrocarbon receptor core complex. Chemico-Biological Interactions. 2002;141(1-2):25-40.
25. Mulero-Navarro S, Fernandez-Salguero PM. New trends in Aryl hydrocarbon receptor biology. Frontiers in Cell and Developmental Biology. 2016;4(MAY):45-.
26. Schmidt JV, Su GHT, Reddy JK, Simon MC, Bradfield CA. Characterization of a murine Ahr null allele: Involvement of the Ah receptor in hepatic growth and development. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(13):6731-6.
27. Lahvis GP, Lindell SL, Thomas RS, McCuskey RS, Murphy C, Glover E, et al. Portosystemic shunting and persistent fetal vascular structures in aryl hydrocarbon receptor-deficient mice. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(19):10442-7.
28. Lahvis GP, Pyzalski RW, Glover E, Pitot HC, McElwee MK, Bradfield CA. The Aryl Hydrocarbon Receptor Is Required for Developmental Closure of the Ductus Venosus in the Neonatal Mouse. Molecular Pharmacology. 2005;67(3):714-20.
29. Murray IA, Patterson AD, Perdew GH. Aryl hydrocarbon receptor ligands in cancer: friend and foe. Nature Reviews Cancer 2014 14:12. 2014;14(12):801-14.
30. Paris A, Tardif N, Galibert MD, Corre S. AhR and Cancer: From Gene Profiling to Targeted Therapy. International Journal of Molecular Sciences. 2021;22(2):1-22.
31. Knerr S, Schrenk D. Carcinogenicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin in experimental models. Molecular Nutrition and Food Research. 2006;50(10):897-907.
32. Bock KW, Köhle C. Ah receptor- and TCDD-mediated liver tumor promotion: clonal selection and expansion of cells evading growth arrest and apoptosis. Biochemical Pharmacology. 2005;69(10):1403-8.
33. Mezrich JD, Fechner JH, Zhang X, Johnson BP, Burlingham WJ, Bradfield CA. An interaction between kynurenine and the aryl

hydrocarbon receptor can generate regulatory T cells. Journal of immunology (Baltimore, Md : 1950). 2010;185(6):3190-.
34. Hollingshead BD, Beischlag TV, DiNatale BC, Ramadoss P, Perdew GH. Inflammatory Signaling and Aryl Hydrocarbon Receptor Mediate Synergistic Induction of Interleukin 6 in MCF-7 Cells. Cancer research. 2008;68(10):3609-.
35. Vierimaa O, Georgitsi M, Lehtonen R, Vahteristo P, Kokko A, Raitila A, et al. Pituitary adenoma predisposition caused by germline mutations in the AIP gene. Science. 2006;312(5777):1228-30.
36. Lloyd C, Grossman A. The AIP (aryl hydrocarbon receptor-interacting protein) gene and its relation to the pathogenesis of pituitary adenomas. Endocrine. 2014;46(3):387-96.
37. Leontiou CA, Gueorguiev M, Van Der Spuy J, Quinton R, Lolli F, Hassan S, et al. The Role of the Aryl Hydrocarbon Receptor-Interacting Protein Gene in Familial and Sporadic Pituitary Adenomas. The Journal of Clinical Endocrinology & Metabolism. 2008;93(6):2390-401.
38. Cazabat L, Guillaud-Bataille M, Bertherat J, Raffin-Sanson ML. Mutations of the Gene for the Aryl Hydrocarbon Receptor-Interacting Protein in Pituitary Adenomas. Hormone Research in Paediatrics. 2009;71(3):132-41.
39. Georgitsi M, Karhu A, Winqvist R, Visakorpi T, Waltering K, Vahteristo P, et al. Mutation analysis of aryl hydrocarbon receptor interacting protein (AIP) gene in colorectal, breast, and prostate cancers. British Journal of Cancer 2007 96:2. 2007;96(2):352-6.
40. Kubens BS, Zanker KS. Differences in the migration capacity of primary human colon carcinoma cells (SW480) and their lymph node metastatic derivatives (SW620). Cancer Lett. 1998;131(1):55-64.
41. Garranzo-Asensio M, San Segundo-Acosta P, Povés C, Fernández-Aceñero MJ, Martínez-Useros J, Montero-Calle A, et al. Identification of tumor-associated antigens with diagnostic ability of colorectal cancer by in-depth immunomic and seroproteomic analysis. Journal of Proteomics. 2020;214:103635-.
42. Araki K, Shimura T, Suzuki H, Tsutsumi S, Wada W, Yajima T, et al. E/N-cadherin switch mediates cancer progression via TGF-beta-induced epithelial-to-mesenchymal transition in extrahepatic cholangiocarcinoma. Br J Cancer. 2011;105(12):1885-93.
43. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-D13.

44. Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. Mol Cell Proteomics. 2020;19(12):2115-25.
45. Bartolomé RA, Barderas R, Torres S, Fernandez-Aceñero MJ, Mendes M, García-Foncillas J, et al. Cadherin-17 interacts with α2β1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis. Oncogene. 2013;33(13):1658-69.
46. Tian X, Yang C, Yang L, Sun Q, Liu N. PTPRF as a novel tumor suppressor through deactivation of ERK1/2 signaling in gastric adenocarcinoma. Onco Targets Ther. 2018;11:7795-803.
47. Fang WK, Liao LD, Li LY, Xie YM, Xu XE, Zhao WJ, et al. Down-regulated desmocollin-2 promotes cell aggressiveness through redistributing adherens junctions and activating beta-catenin signalling in oesophageal squamous cell carcinoma. J Pathol. 2013;231(2):257-70.
48. Bujko M, Kober P, Mikula M, Ligaj M, Ostrowski J, Siedlecki JA. Expression changes of cell-cell adhesion-related genes in colorectal tumors. Oncol Lett. 2015;9(6):2463-70.
49. Bajpai R, Nagaraju GP. Specificity protein 1: Its role in colorectal cancer progression and metastasis. Crit Rev Oncol Hematol. 2017;113:1-7.
50. Wang H, Li K, Mei Y, Huang X, Li Z, Yang Q, et al. Sp1 Suppresses miR-3178 to Promote the Metastasis Invasion Cascade via Upregulation of TRIOBP. Mol Ther Nucleic Acids. 2018;12:1-11.
51. Khodarev NN, Roach P, Pitroda SP, Golden DW, Bhayani M, Shao MY, et al. STAT1 pathway mediates amplification of metastatic potential and resistance to therapy. PLoS One. 2009;4(6):e5821.
52. Basak O, van de Born M, Korving J, Beumer J, van der Elst S, van Es JH, et al. Mapping early fate determination in Lgr5+ crypt stem cells using a novel Ki67-RFP allele. EMBO J. 2014;33(18):2057-68.
53. Coopmans EC, Muhammad A, Daly AF, de Herder WW, van Kemenade FJ, Beckers A, et al. The role of AIP variants in pituitary adenomas and concomitant thyroid carcinomas in the Netherlands: a nationwide pathology registry (PALGA) study. Endocrine. 2020;68(3):640-9.
54. Diaz Del Arco C, Estrada Munoz L, Barderas Manchado R, Pelaez Garcia A, Ortega Medina L, Molina Roldan E, et al. Prognostic Role of Aryl Hydrocarbon Receptor Interacting Protein (AIP) Immunohistochemical Expression in Patients with Resected Gastric Carcinomas. Pathol Oncol Res. 2020;26(4):2641-50.

55. Fernández-Aceñero MJ, Barderas R, Peláez-García A, Martínez-Useros J, Díez-Valladares L, Ortega-Medina L, et al. Aryl hydrocarbon receptor interacting protein (AIP) significantly influences prognosis of pancreatic carcinoma. Annals of Diagnostic Pathology. 2021;151742.
56. Barry S, Carlsen E, Marques P, Stiles CE, Gadaleta E, Berney DM, et al. Tumor microenvironment defines the invasive phenotype of AIP-mutation-positive pituitary tumors. Oncogene. 2019;38(27):5381-95.
57. Vasilev V, Daly AF, Trivellin G, Stratakis CA, Zacharieva S, Beckers A. HEREDITARY ENDOCRINE TUMOURS: CURRENT STATE-OF-THE-ART AND RESEARCH OPPORTUNITIES: The roles of AIP and GPR101 in familial isolated pituitary adenomas (FIPA). Endocr Relat Cancer. 2020;27(8):T77-T86.
58. Trivellin G, Korbonits M. AIP and its interacting partners. J Endocrinol. 2011;210(2):137-55.
59. Rothhammer V, Quintana FJ. The aryl hydrocarbon receptor: an environmental sensor integrating immune responses in health and disease. Nat Rev Immunol. 2019;19(3):184-97.
60. Meyer BK, Pray-Grant MG, Vanden Heuvel JP, Perdew GH. Hepatitis B virus X-associated protein 2 is a subunit of the unliganded aryl hydrocarbon receptor core complex and exhibits transcriptional enhancer activity. Mol Cell Biol. 1998;18(2):978-88.
61. Sigismund S, Avanzato D, Lanzetti L. Emerging functions of the EGFR in cancer. Mol Oncol. 2018;12(1):3-20.
62. Markman B, Javier Ramos F, Capdevila J, Tabernero J. EGFR and KRAS in colorectal cancer. Adv Clin Chem. 2010;51:71-119.
63. Sigismund S, Argenzio E, Tosoni D, Cavallaro E, Polo S, Di Fiore PP. Clathrin-mediated internalization is essential for sustained EGFR signaling but dispensable for degradation. Dev Cell. 2008;15(2):209-19.

General Discussion

In the work presented in this thesis we have focused on increasing the knowledge in colorectal cancer proteome, using isogenic cell lines as model and validating the data *in vitro* and *in vivo* using cells, animal models and colorectal cancer tissue. The aim behind this approach was dual, on the one hand to deepen the understanding of CRC metastasis' onset and progression and, on the other hand, find markers that could be used for early diagnosis of the disease. The three results chapters represent the different steps of this process. In a first step we performed the multidimensional proteomics characterisation of five isogenic cell lines with different metastatic properties. Then, we attempted to improve the model by incorporating the effect of 3D differentiation of cancer cells. Finally, we functionally evaluated the role of one of the proteins we found to be altered, Aryl-Hydrocarbon receptor-interacting protein (AIP), in the metastatic process by overexpressing it in the KM12 cell model of CRC metastasis. Additionally, as part of our effort to better understand cell differentiation in 3D we performed the proteomic analysis of ADSCs grown in different PIC-based synthetic matrices to both test the suitability of PIC for cell culturing, compared with Matrigel, and of our proteomic analysis pipeline for 3D cultured cells.

The steps described in this work could be considered as a general approach that could be further on systematically applied to similar biological questions. Proteomics has proven its usefulness for the identification of targets that are reliable in a clinical scenario. Furthermore, AIP, whose role on the metastatic progression of CRC was demonstrated here, has also been shown to have a prognostic value in the diagnosis of gastric and prostate cancer (1, 2). In addition, the proteomic analysis of PIC cultured cells contributes to confirm its tremendous value as a tool for the investigation of cell behaviour in 3D. Our data, clearly show that PIC stiffness and ligands linked to it can heavily affect cell behaviour, opening the way for almost endless research possibilities due to the high tuneability that PIC offers.

The spatial analysis of CRC isogenic cell lines revealed a series of compelling markers whose role in the progression of the disease is yet unclear. From all the proteins we found to be upregulated BAIAP2 and GLG1 were particularly interesting. In the case of BAIAP2, as we already discussed, we found that the localization of BAIAP2 drastically changes its correlation with patient survival. When localized at the cytosol, its most frequent location, BAIAP2 levels were correlated with good patient survival. Conversely, a shift from the cytosol towards the membrane was associated with poor patient survival, a change we also observed in the metastatic CRC cell lines (KM12SM, KM12L4a and SW620) in comparison to isogenic non-metastatic CRC cell lines. BAIAP2 is involved in the reorganization of the actin cytoskeleton as part of the signalling of Cdc42/Rho. Although we did not detect BAIAP2 in the 3D proteomic analysis of KM12 cells, we still would like to analyse BAIAP2 localization using immunofluorescence in 3D. Especially considering that KM12C cells acquired a more metastatic phenotype in terms of protein expression when cultured in 3D. Among the proteins commonly upregulated between KM12SM cultured in 2D and 3D cultured KM12C we found CYRIB, that is also involved in Rac1 and Cdc42 signalling. Alternatively, in the future we would also like to follow BAIAP2 changes in localization in metastatic and non-metastatic cells upon the activation of pro- and anti-tumorigenic stimuli.

Likewise, we would also like to expand on the potential diagnostic ability of GLG1. In the work here presented we demonstrate the ability of GLG1 to discriminate between CRC patients at advance stages and control patients with an area under the curve of 74.13%. We also found a significant difference in GLG1 plasmatic levels of control and pathological subjects. The discrimination capacity of GLG1 plasmatic levels is promising although it could be considered limited because of the number of samples analysed, or taking into account that early CRC could not be discriminated from controls. However, GLG1 could be

considered as prognostic blood-based marker, being its usefulness associated to the detection of advance CRC stages or for recurrence. Diagnostic panels based on plasmatic levels of markers are generally composed of several markers that altogether provide the best combination of sensitivity and specificity. There are still a large number of proteins upregulated in metastatic cells that could also have diagnostic or prognostic potential in combination with GLG1. For example, GAS6, that was more than 35 times upregulated in the secretome of metastatic cells. In collaboration with the group of Prof. Dr. Susana Campuzano, we were able to develop an electrochemical immunosensor for GAS6 that replicated the results we obtained via proteomic and western blot analysis (3). Furthermore, this sensor was able to detect significant differences in serum levels of GAS6 in pancreatic ductal adenocarcinoma patients and healthy controls (3). Combination of GLG1 and GAS6 for CRC might also improve the detection efficiency and sensitivity for the screening of CRC patients. Research into some of the proteins also found to be altered in the adenoma/adenocarcinoma proteomic analysis should also provide with more markers that could be further evaluated and considered for the incorporation into a wide serum screening panel.

The proteomic analysis of ADSCs cultured in PIC further support the previous results from our collaborator (4) and provide with a potential mechanism for the changes in protein secretion that had been observed before. In addition, the drastic change induced by bicyclic peptides grafting of PIC in ADSC differentiation open the way for a simplified understanding of the interaction of cells with its surrounding matrix. For the moment we have only tested ligands that were specific for integrins $\alpha_v\beta_3$ and $\alpha_5\beta_1$, but there are still many more integrins and many other adhesion molecules that could be tried. In the work presented here we observed a stronger effect of $\alpha_5\beta1$-binding peptides. This integrin is activated by the recognition of fibronectin, a common component of the

ECM, and triggers the activation of angiogenesis. Consequently, it is reasonable to assume that the effect observed is due to the activation of similar pathways to those triggered by fibronectin. Interestingly, though RGD should also bind both $\alpha_v\beta_3$ and $\alpha_5\beta_1$ we could not observe the same effect for cyclic RGD. This would imply that there is an additional effect likely due to the three-dimensional organization of the bicyclic peptide compared with the monocyclic RGD. In collaboration with the group of Prof. Kouwer, we are currently trying to apply the same principle to EMT and E-cadherin and N-cadherin. These two cadherins show opposite regulation and are key in the transition from epithelial to mesenchymal phenotype. When we overexpressed AIP in KM12C and KM12SM cells we could observe a decrease in E-Cadherin and an increase in N-cadherin levels, which we associated with the increase of a partial EMT phenotype. PIC grafting would allow to coat the matrix with ligands specific to each of these cadherins. In addition, it would still be possible to use PIC with different mechanical properties. These properties could be adjusted to mimic the progression observed in tumours, that become stiffer, compared to healthy tissues, as ECM deposition by CAFs increases.

In terms of the outlook of the bicyclic peptide platform we are going to repeat the encapsulation and acquire data at later time points (48 and 72 hours) to observe the evolution of cells encapsulated in Matrigel. Our hypothesis for the protein profiles we observed was that P8-PIC represented a better suited matrix for ADSCs while Matrigel still had to be modified. Consequently, we would expect that if cells encapsulated in Matrigel become morphologically similar to those encapsulated in P8-PIC after a longer period of time, the protein expression profile should also be similar. Furthermore, if the morphology and protein expression are similar, it might be possible by incorporating other ligands besides P8 to PIC gels to obtain a matrix that not only mimics Matrigel while being synthetic but even outperforms it for 3D cell culture.

Finally, in the last chapter of the results we demonstrate the value of proteomic derived markers by showing the potential of AIP, which was identified as upregulated in a previous work using SILAC quantitative proteomics analysis (5). Here, using another quantitative proteomic technique (TMT) we could not observe the dysregulation of AIP. This could be associated to the use of different quantitative proteomics techniques (TMT 10-plex versus SILAC), the simultaneous comparison of subcellular fractions in the same quantitative proteomic experiment (TMT), and/or to the equipment used. Importantly, we were able to demonstrate that AIP overexpression could induce KM12C cells to become metastatic to liver cells as proven by the different functional *in vitro* and *in vivo* studies done. Although AIP depletion was not explored in CRC cells as a reciprocal phenotype to overexpression -as a limitation of the study-, we report here a new role for AIP in adhesion, invasion, migration, colony formation and liver metastasis in CRC as depicted from AIP overexpression. Given the prognostic capacity of AIP in gastric and pancreatic cancer it would also be interesting to explore the effect of AIP knock in or down in gastric or pancreatic cancer cell lines. Additionally, we also plan to continue the research on AIP by inducing AIP overexpression in other isogenic cell lines (namely SW480 and SW620 cell lines) to test whether the effect observed is independent of the CRC cell type. Moreover, it would be also interesting to explore the potential of AIP as a druggable protein target either by using specific inhibitors or antibodies to try to improve the survival of CRC in advanced stages.

Among the proteins that AIP induced alterations on were the EGFR and CDH17. Both molecules have been associated with the progression and onset of CRC. In the case of CDH17 the change was especially significant because we did not only observe a change in the levels of the proteins but also on its localization. Much like in the case of BAIAP2, we could see a clear shift in the localization of CDH17 from what appeared

to be a nucleolar like distribution towards the membrane. Exploiting the highly tuneable capacities of PIC, it would be tremendously interesting to decorate PIC polymers with CDH17 ligands and encapsulate inside them KM12C and KM12SM cells overexpressing AIP. CDH17 has been shown to bind to $\alpha_2\beta_1$ although the exact binding mechanisms that CDH17 could use are still under debate (6, 7)(REF). In fact, based on our results, we could consider AIP as a metastatic switch, with the potential to be used to induce a metastatic change when desired. Consequently, it could be used to study, for example, the interaction of non-metastatic cells with CAFs and how does it change when the cells become metastatic.

Altogether, the work presented in this thesis demonstrates the vast potential that oncoproteomics has, and how it can benefit and exploit the advantages of the use of synthetic matrices for expanding our current knowledge on cell differentiation and behaviour.

References

1. Fernández-Aceñero MJ, Barderas R, Peláez-García A, Martínez-Useros J, Díez-Valladares L, Ortega-Medina L, et al. Aryl hydrocarbon receptor interacting protein (AIP) significantly influences prognosis of pancreatic carcinoma. Annals of Diagnostic Pathology. 2021;151742.
2. Diaz Del Arco C, Estrada Munoz L, Barderas Manchado R, Pelaez Garcia A, Ortega Medina L, Molina Roldan E, et al. Prognostic Role of Aryl Hydrocarbon Receptor Interacting Protein (AIP) Immunohistochemical Expression in Patients with Resected Gastric Carcinomas. Pathol Oncol Res. 2020;26(4):2641-50.
3. Muñoz-San Martín C, Pérez-Ginés V, Torrente-Rodríguez RM, Gamella M, Solís-Fernández G, Montero-Calle A, et al. Electrochemical immunosensing of Growth arrest-specific 6 in human plasma and tumor cell secretomes. Electrochem Sci Adv. 2021:e2100096.
4. Liu K, Veenendaal T, Wiendels M, Ruiz-Zapata AM, Van Laar J, Kyranas R, et al. Synthetic Extracellular Matrices as a Toolbox to Tune Stem Cell Secretome. ACS Applied Materials and Interfaces. 2020;12(51):56723-30.
5. Mendes M, Peláez-García A, López-Lucendo M, Bartolomé RA, Calviño E, Barderas R, et al. Mapping the Spatial Proteome of Metastatic Cells in Colorectal Cancer. Proteomics. 2017;17(19):1700094-.
6. Bartolomé RA, Barderas R, Torres S, Fernandez-Aceñero MJ, Mendes M, García-Foncillas J, et al. Cadherin-17 interacts with α2β1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis. Oncogene. 2013;33(13):1658-69.
7. Gray ME, Sotomayor M. Crystal structure of the nonclassical cadherin-17 N-terminus and implications for its adhesive binding mechanism. Acta Crystallogr F Struct Biol Commun. 2021;77(Pt 3):85-94.

Conclusions

Chapter 1

- We have demonstrated the potential that proteomic analysis of CRC isogenic cell lines has for the determination of biomarkers and provided a comprehensive analysis of the proteins altered in different metastatic tropisms.
- The subcellular proteomic analysis has demonstrated that the alterations found can be easily overlooked when analysing at a whole cell level.
- We have found novel dysregulated proteins not previously associated with CRC. Of special interest are GLG1 and BAIAP2. GLG1 plasma levels were significantly different between control and CRC patients while BAIAP2's change in localization in metastatic cells was associated with poor patient survival.

Chapter 2

- Once the usefulness of PIC derived hydrogels for cell culture was proven, quantitative proteomic analysis demonstrated that PIC hydrogels induce cellular responses similar to those observed in natural tissues.
- Thanks to its high tuneability, PIC grafting with bicyclic peptides was shown to induce faster differentiation of adipose stem cells, opening the way for an almost endless array of possibilities for 3D cell culturing.
- Three-dimensional culture of isogenic KM12C and KM12SM cells has demonstrated that culture dimensionality can also have an impact in the behaviour of metastatic and non-metastatic cells.
- The differences observed between KM12C and KM12SM cells indicate that non-metastatic cells can acquire a more metastatic phenotype when they are cultured in 3D. As with ADSCs, different

ligands can be used to decorate PIC polymers to evaluate in detail the interaction of cancer cells with different elements of the ECM.

Chapter 3

- Aryl-hydrocarbon receptor-interacting protein (AIP) has been shown to have a tremendous impact in KM12C and KM12SM CRC cells phenotype.
- AIP was able to induce drastic changes in the metastatic capacities of KM12C CRC cells as proven by *in vitro* functional assays and confirmed *in vivo* through homing and survival assays.
- AIP induced alterations in a myriad of signalling mediators including JNK, SRC or AKT, in proteins involved in EMT and proteins previously associated to metastasis and progression of colorectal cancer as EGFR and CDH17.
- AIP has a key role in CRC and liver metastasis.

Protein list

Due to the large number of proteins mentioned in this thesis we have considered useful to list them here including a brief description of their function together with their Uniprot identifier (with the link towards the Uniprot web entry).

ACTA2: α-smooth muscle actin (ASMA). Actins are highly conserved proteins that are involved in various types of cell motility and are ubiquitously expressed in all eukaryotic cell. Uniprot: P62736

AHR: Ligand-activated transcription factor that enables cells to adapt to changing conditions by sensing compounds from the environment, diet, microbiome and cellular metabolism, and which plays important roles in development, immunity and cancer. Uniprot: P35869

AIP: AH receptor-interacting protein. May play a positive role in AHR-mediated (aromatic hydrocarbon receptor) signaling, possibly by influencing its receptivity for ligand and/or its nuclear targeting. Uniprot: O00170

AKT: AKT1 is one of 3 closely related serine/threonine-protein kinases (AKT1, AKT2 and AKT3) called the AKT kinase, and which regulate many processes including metabolism, proliferation, cell survival, growth and angiogenesis. Uniprot: P31749

APC: Adenomatous polyposis coli protein. Tumor suppressor. Promotes rapid degradation of CTNNB1 and participates in Wnt signaling as a negative regulator. APC activity is correlated with its phosphorylation state. Uniprot: P25054

ARF: Capable of inducing cell cycle arrest in G1 and G2 phases. Acts as a tumor suppressor. Binds to MDM2 and blocks its nucleocytoplasmic shuttling by sequestering it in the nucleolus. Uniprot: Q8N726

ARNT: Required for activity of the Ah (dioxin) receptor. This protein is required for the ligand-binding subunit to translocate from the cytosol to the nucleus after ligand binding. Uniprot: P27540

BAI1: Adhesion G protein-coupled receptor B1. Phosphatidylserine receptor which enhances the engulfment of apoptotic cells. Also mediates the binding and engulfment of Gram-negative bacteria. Uniprot: O14514

Protein list

BAIAP2: Brain-specific angiogenesis inhibitor 1-associated protein 2. Adapter protein that links membrane-bound small G-proteins to cytoplasmic effector proteins. Necessary for CDC42-mediated reorganization of the actin cytoskeleton and for RAC1-mediated membrane ruffling. Uniprot: Q9UQB8

BIRC5/Survivin: Multitasking protein that has dual roles in promoting cell proliferation and preventing apoptosis. Uniprot: O15392

BMP: Bone morphogenetic protein 1. Metalloprotease that plays key roles in regulating the formation of the extracellular matrix (ECM) via processing of various precursor proteins into mature functional enzymes or structural proteins. Uniprot: P13497

BRAF: Serine/threonine-protein kinase B-raf. Protein kinase involved in the transduction of mitogenic signals from the cell membrane to the nucleus (Probable). Phosphorylates MAP2K1, and thereby activates the MAP kinase signal transduction pathway. Uniprot: P15056

CacyBP: Calcyclin-binding protein. May be involved in calcium-dependent ubiquitination and subsequent proteasomal degradation of target proteins. Uniprot: Q9HB71

CacyBP: May be involved in calcium-dependent ubiquitination and subsequent proteasomal degradation of target proteins. Probably serves as a molecular bridge in ubiquitin E3 complexes. Uniprot: Q9HB71

CD8: T-cell surface glycoprotein CD8. Integral membrane glycoprotein that plays an essential role in the immune response and serves multiple functions in responses against both external and internal offenses. In T-cells, functions primarily as a coreceptor for MHC class I molecule:peptide complex. Uniprot: P01732

CDH17: Cadherin-17. Cadherins are calcium-dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic

manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types. Uniprot: Q12864

CDKN2AIP: CDKN2A-interacting protein. Regulates DNA damage response in a dose-dependent manner through a number of signaling pathways involved in cell proliferation, apoptosis and senescence. Uniprot: Q9NXV6

Claudin-2: Plays a major role in tight junction-specific obliteration of the intercellular space, through calcium-independent cell-adhesion activity. Uniprot: P57739

CLDN3: Claudin-3. Plays a major role in tight junction-specific obliteration of the intercellular space, through calcium-independent cell-adhesion activity. Uniprot: O15551

Collagen I: Type I collagen is a member of group I collagen (fibrillar forming collagen). Uniprot: P02452

Collagen III: Collagen type III occurs in most soft connective tissues along with type I collagen. Involved in regulation of cortical development. Uniprot: P02461

Collagen V: Type V collagen is a member of group I collagen (fibrillar forming collagen). It is a minor connective tissue component of nearly ubiquitous distribution. Type V collagen binds to DNA, heparan sulfate, thrombospondin, heparin, and insulin. Uniprot: P20908

CTNNB1: Catenin beta-1. Key downstream component of the canonical Wnt signaling pathway. Uniprot: P35222

CYP1A1: A cytochrome P450 monooxygenase involved in the metabolism of various endogenous substrates, including fatty acids, steroid hormones and vitamins. Uniprot: P04798

DSC2: Desmocollin-2. Component of intercellular desmosome junctions. Involved in the interaction of plaque proteins and intermediate filaments mediating cell-cell adhesion. Uniprot: Q02487

E-Cadherin: Cadherin-1. Cadherins are calcium-dependent cell adhesion proteins. CDH1 is involved in mechanisms regulating cell-cell adhesions, mobility and proliferation of epithelial cells. Uniprot: P12830

E-selectin: Cell-surface glycoprotein having a role in immunoadhesion. Mediates in the adhesion of blood neutrophils in cytokine-activated endothelium through interaction with SELPLG/PSGL1. May have a role in capillary morphogenesis. Uniprot: P16581

EBNA3: Plays an essential role for activation and immortalization of human B-cells. Uniprot: P12977

EGFR: Epidermal growth factor receptor. Receptor tyrosine kinase binding ligands of the EGF family and activating several signaling cascades to convert extracellular cues into appropriate cellular responses. Uniprot: P00533

EPCAM: Epithelial cell adhesion molecule. May act as a physical homophilic interaction molecule between intestinal epithelial cells (IECs) and intraepithelial lymphocytes (IELs) at the mucosal epithelium for providing immunological barrier as a first line of defense against mucosal infection. Uniprot: P16422

ERK1/MAPK3: Mitogen-activated protein kinase 3. Serine/threonine kinase which acts as an essential component of the MAP kinase signal transduction pathway. MAPK1/ERK2 and MAPK3/ERK1 are the 2 MAPKs which play an important role in the MAPK/ERK cascade. They participate also in a signaling cascade initiated by activated KIT and KITLG/SCF. Uniprot: P27361

ERK2/MAPK1: Mitogen-activated protein kinase 1. Serine/threonine kinase which acts as an essential component of the MAP kinase signal

transduction pathway. MAPK1/ERK2 and MAPK3/ERK1 are the 2 MAPKs which play an important role in the MAPK/ERK cascade. They participate also in a signaling cascade initiated by activated KIT and KITLG/SCF. Uniprot: P28482

FAK: Focal Adhesion Kinase 1. Non-receptor protein-tyrosine kinase that plays an essential role in regulating cell migration, adhesion, spreading, reorganization of the actin cytoskeleton, formation and disassembly of focal adhesions and cell protrusions, cell cycle progression, cell proliferation and apoptosis. Uniprot: Q05397

FGF-2: Fibroblast growth factor 2. Acts as a ligand for FGFR1, FGFR2, FGFR3 and FGFR4. Also acts as an integrin ligand which is required for FGF2 signaling. Uniprot: P09038

FGF: Fibroblast growth factor 1. Plays an important role in the regulation of cell survival, cell division, angiogenesis, cell differentiation and cell migration. Functions as potent mitogen in vitro. Acts as a ligand for FGFR1 and integrins. Uniprot: P05230

FGFR4: Fibroblast growth factor receptor 4. Tyrosine-protein kinase that acts as cell-surface receptor for fibroblast growth factors and plays a role in the regulation of cell proliferation, differentiation and migration, and in regulation of lipid metabolism, bile acid biosynthesis, glucose uptake, vitamin D metabolism and phosphate homeostasis. Uniprot: P22455

Fibronectin 1: Fibronectins bind cell surfaces and various compounds including collagen, fibrin, heparin, DNA, and actin. Fibronectins are involved in cell adhesion, cell motility, opsonization, wound healing, and maintenance of cell shape. Uniprot: P02751

GAPDH: Glyceraldehyde-3-phosphate dehydrogenase. Glyceraldehyde-3-phosphate dehydrogenase is a key enzyme in glycolysis that catalyzes the first step of the pathway by converting D-glyceraldehyde 3-phosphate (G3P) into 3-phospho-D-glyceroyl phosphate. Uniprot: P04406

GDF15: Growth/differentiation factor 15. Regulates food intake, energy expenditure and body weight in response to metabolic and toxin-induced stresses. Uniprot: Q99988

GDF15: Regulates food intake, energy expenditure and body weight in response to metabolic and toxin-induced stresses. Uniprot: Q99988

GLG1: Golgi apparatus protein 1. Binds fibroblast growth factor and E-selectin (cell-adhesion lectin on endothelial cells mediating the binding of neutrophils). Q92896

GNA13: Guanine nucleotide-binding proteins (G proteins) are involved as modulators or transducers in various transmembrane signaling systems. Uniprot: Q14344

GNA14: Guanine nucleotide-binding proteins (G proteins) are involved as modulators or transducers in various transmembrane signaling systems. Uniprot: O95837

HBV-X:

HER2/ERBB2: Receptor tyrosine-protein kinase erbB-2. Protein tyrosine kinase that is part of several cell surface receptor complexes, but that apparently needs a coreceptor for ligand binding. Uniprot: P04626

HGF: Hepatocyte growth factor. Potent mitogen for mature parenchymal hepatocyte cells, seems to be a hepatotrophic factor, and acts as a growth factor for a broad spectrum of tissues and cell types. Activating ligand for the receptor tyrosine kinase MET by binding to it and promoting its dimerization. Uniprot: P14210

Hippo: Serine/threonine-protein kinase 4. Stress-activated, pro-apoptotic kinase which, following caspase-cleavage, enters the nucleus and induces chromatin condensation followed by internucleosomal DNA fragmentation. Key component of the Hippo signaling pathway which

plays a pivotal role in organ size control and tumor suppression by restricting proliferation and promoting apoptosis. Uniprot: Q13043

Hsc70: Molecular chaperone implicated in a wide variety of cellular processes, including protection of the proteome from stress, folding and transport of newly synthesized polypeptides, activation of proteolysis of misfolded proteins and the formation and dissociation of protein complexes. Uniprot: P11142

Hsp90:

hTERT: Telomerase reverse transcriptase. Telomerase is a ribonucleoprotein enzyme essential for the replication of chromosome termini in most eukaryotes. Active in progenitor and cancer cells. Uniprot: O14746

ICAM-1: Intercellular adhesion molecule 1. ICAM proteins are ligands for the leukocyte adhesion protein LFA-1 (integrin alpha-L/beta-2). During leukocyte trans-endothelial migration, ICAM1 engagement promotes the assembly of endothelial apical cups through ARHGEF26/SGEF and RHOG activation. Uniprot: P05362

IL13Rα2: Interleukin-13 receptor subunit alpha-2. Binds as a monomer with high affinity to interleukin-13 (IL13), but not to interleukin-4 (IL4). Uniprot: Q14627

ITGA2B: Integrin alpha-IIb. Integrin alpha-IIb/beta-3 is a receptor for fibronectin, fibrinogen, plasminogen, prothrombin, thrombospondin and vitronectin. It recognizes the sequence R-G-D in a wide array of ligands. Uniprot: P08514

ITGA3: Integrin alpha-3. Integrin alpha-3/beta-1 is a receptor for fibronectin, laminin, collagen, epiligrin, thrombospondin and CSPG4. Uniprot: P26006

ITGA5: Integrin alpha-5. Integrin alpha-5/beta-1 (ITGA5:ITGB1) is a receptor for fibronectin and fibrinogen. It recognizes the sequence R-G-D in its ligands. Uniprot: P08648

ITGAL: Integrin alpha-L. Integrin ITGAL/ITGB2 is a receptor for ICAM1, ICAM2, ICAM3 and ICAM4. Integrin ITGAL/ITGB2 is a receptor for F11R. Uniprot: P20701

ITGAV: Integrin alpha-V. The alpha-V (ITGAV) integrins are receptors for vitronectin, cytotactin, fibronectin, fibrinogen, laminin, matrix metalloproteinase-2, osteopontin, osteomodulin, prothrombin, thrombospondin and vWF. Uniprot: P06756

ITGB1: Integrin beta-1. Integrins alpha-1/beta-1, alpha-2/beta-1, alpha-10/beta-1 and alpha-11/beta-1 are receptors for collagen. Integrins alpha-1/beta-1 and alpha-2/beta-2 recognize the proline-hydroxylated sequence G-F-P-G-E-R in collagen. Uniprot: P05556

ITGB2: Integrin beta-2. Integrin ITGAL/ITGB2 is a receptor for ICAM1, ICAM2, ICAM3 and ICAM4. Integrin ITGAL/ITGB2 is also a receptor for the secreted form of ubiquitin-like protein ISG15; the interaction is mediated by ITGAL. Uniprot: P05107

ITGB3: Integrin beta-3. Integrin alpha-V/beta-3 (ITGAV:ITGB3) is a receptor for cytotactin, fibronectin, laminin, matrix metalloproteinase-2, osteopontin, osteomodulin, prothrombin, thrombospondin, vitronectin and von Willebrand factor. Uniprot: P05106

ITGB5: Integrin beta-5. Integrin alpha-V/beta-5 (ITGAV:ITGB5) is a receptor for fibronectin. It recognizes the sequence R-G-D in its ligand. Uniprot: P18084

JNK/MAPK8: Mitogen-activated protein kinase 8. Serine/threonine-protein kinase involved in various processes such as cell proliferation, differentiation, migration, transformation and programmed cell death. Uniprot: P45983

KRAS: GTPase KRas. Ras proteins bind GDP/GTP and possess intrinsic GTPase activity. Plays a role in promoting oncogenic events by inducing transcriptional silencing of tumor suppressor genes (TSGs) in colorectal cancer (CRC) cells in a ZNF304-dependent manner. Uniprot: P01116

LTBP: Latent-transforming growth factor beta-binding protein 1. Key regulator of transforming growth factor beta (TGFB1, TGFB2 and TGFB3) that controls TGF-beta activation by maintaining it in a latent state during storage in extracellular space. Uniprot: Q14766

MLH1: DNA mismatch repair protein Mlh1. Heterodimerizes with PMS2 to form MutL alpha, a component of the post-replicative DNA mismatch repair system (MMR). Uniprot: P40692

MMP7: Matrilysin. Degrades casein, gelatins of types I, III, IV, and V, and fibronectin. Activates procollagenase. Uniprot: P09237

MSH2: DNA mismatch repair protein Msh2. Component of the post-replicative DNA mismatch repair system (MMR). Forms two different heterodimers: MutS alpha (MSH2-MSH6 heterodimer) and MutS beta (MSH2-MSH3 heterodimer) which binds to DNA mismatches thereby initiating DNA repair. Uniprot: P43246

MSH6: DNA mismatch repair protein Msh6. Component of the post-replicative DNA mismatch repair system (MMR). Heterodimerizes with MSH2 to form MutS alpha, which binds to DNA mismatches thereby initiating DNA repair. Uniprot: P52701

MUTYH: Adenine DNA glycosylase. Involved in oxidative DNA damage repair. Initiates repair of A*oxoG to C*G by removing the inappropriately paired adenine base from the DNA backbone. Possesses both adenine and 2-OH-A DNA glycosylase activities. Uniprot: Q9UIF7

MYC: Myc proto-oncogene protein. Transcription factor that binds DNA in a non-specific manner, yet also specifically recognizes the core sequence

5'-CAC[GA]TG-3'. Activates the transcription of growth-related genes. Uniprot: P01106

N-Cadherin: Cadherin-2. Calcium-dependent cell adhesion protein; preferentially mediates homotypic cell-cell adhesion by dimerization with a CDH2 chain from another cell. Cadherins may thus contribute to the sorting of heterogeneous cell types. Uniprot: P19022

Notch: Neurogenic locus notch homolog protein 1. Functions as a receptor for membrane-bound ligands Jagged-1 (JAG1), Jagged-2 (JAG2) and Delta-1 (DLL1) to regulate cell-fate determination. Uniprot: P46531

NRAS: GTPase NRas. Ras proteins bind GDP/GTP and possess intrinsic GTPase activity. Uniprot: P01111.

P23

P38/MAPK11: Mitogen-activated protein kinase 11. Serine/threonine kinase which acts as an essential component of the MAP kinase signal transduction pathway. Q15759

PDE2A3: cGMP-activated cyclic nucleotide phosphodiesterase with a dual-specificity for the second messengers cAMP and cGMP, which are key regulators of many important physiological processes. Uniprot: O00408

PDE4A5: Hydrolyzes the second messenger 3',5'-cyclic AMP (cAMP), which is a key regulator of many important physiological processes. Uniprot: P27815

PDL1: Programmed cell death 1 ligand 1. Plays a critical role in induction and maintenance of immune tolerance to self. s a ligand for the inhibitory receptor PDCD1/PD-1, modulates the activation threshold of T-cells and limits T-cell effector response. Uniprot: Q9NZQ7

PDL2: Programmed cell death 1 ligand 2. Involved in the costimulatory signal, essential for T-cell proliferation and IFNG production in a PDCD1-independent manner. Interaction with PDCD1 inhibits T-cell proliferation by blocking cell cycle progression and cytokine production. Uniprot: Q9BQ51

PECAM-1: Platelet endothelial cell adhesion molecule. Cell adhesion molecule which is required for leukocyte transendothelial migration (TEM) under most inflammatory conditions. Uniprot: P16284

PHYHIPL: Phytanoyl-CoA hydroxylase-interacting protein-like. May play a role in the development of the central system. Uniprot: Q96FC7

PI3K: Phosphatidylinositol 3-kinase. phosphorylates phosphatidylinositol (PI) and its phosphorylated derivatives at position 3 of the inositol ring to produce 3-phosphoinositides. Uniprot: P42336

PIK3CA: Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform. Phosphoinositide-3-kinase (PI3K) phosphorylates phosphatidylinositol (PI) and its phosphorylated derivatives at position 3 of the inositol ring to produce 3-phosphoinositides. Uniprot: P42336

PMS2: Mismatch repair endonuclease PMS2. Component of the post-replicative DNA mismatch repair system (MMR). Heterodimerizes with MLH1 to form MutL alpha. Uniprot: P54278

PPARA: Ligand-activated transcription factor. Key regulator of lipid metabolism. Activated by the endogenous ligand 1-palmitoyl-2-oleoyl-sn-glycerol-3-phosphocholine. Uniprot: Q07869

PTEN: Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN. Tumor suppressor. Acts as a dual-specificity protein phosphatase, dephosphorylating tyrosine-, serine- and threonine-phosphorylated proteins. Uniprot: P60484

Protein list

PTPRF: Receptor-type tyrosine-protein phosphatase F. Possible cell adhesion receptor. It possesses an intrinsic protein tyrosine phosphatase activity (PTPase) and dephosphorylates EPHA2 regulating its activity. Uniprot: P10586

RET: Receptor tyrosine-protein kinase involved in numerous cellular mechanisms including cell proliferation, neuronal navigation, cell migration, and cell differentiation upon binding with glial cell derived neurotrophic factor family ligands. Uniprot: P07949

Rho: Transforming protein RhoA. Small GTPase which cycles between an active GTP-bound and an inactive GDP-bound state. Mainly associated with cytoskeleton organization, in active state binds to a variety of effector proteins to regulate cellular responses such as cytoskeletal dynamics, cell migration and cell cycle. Uniprot: P61586

RhoGDI: Rho GDP-dissociation inhibitor 1. Controls Rho proteins homeostasis. Regulates the GDP/GTP exchange reaction of the Rho proteins by inhibiting the dissociation of GDP from them, and the subsequent binding of GTP to them. Uniprot: P52565

SCRIB: Protein scribble homolog. Scaffold protein involved in different aspects of polarized cell differentiation regulating epithelial and neuronal morphogenesis and T-cell polarization. Uniprot: Q14160

SEPTIN9: Septin-9. Filament-forming cytoskeletal GTPase (By similarity). May play a role in cytokinesis (Potential). May play a role in the internalization of 2 intracellular microbial pathogens, Listeria monocytogenes and Shigella flexneri. Uniprot: Q9UHD8

SMAD4: Mothers against decapentaplegic homolog 4. Promotes binding of the SMAD2/SMAD4/FAST-1 complex to DNA and provides an activation function required for SMAD1 or SMAD2 to stimulate transcription. Uniprot: Q13485

SNAI1: Zinc finger protein SNAI1. Involved in induction of the epithelial to mesenchymal transition (EMT), formation and maintenance of embryonic mesoderm, growth arrest, survival and cell migration. Uniprot: O95863

SNAI2: Zinc finger protein SNAI2, also known as Slug. Transcriptional repressor that modulates both activator-dependent and basal transcription. Involved in the generation and migration of neural crest cells. Uniprot: O43623

SNX9: Sorting nexin-9. Involved in endocytosis and intracellular vesicle trafficking, both during interphase and at the end of mitosis. Required for efficient progress through mitosis and cytokinesis. Uniprot: Q9Y5X1

SP1: Transcription factor Sp1. Transcription factor that can activate or repress transcription in response to physiological and pathological stimuli. Binds with high affinity to GC-rich motifs and regulates the expression of a large number of genes involved in a variety of processes such as cell growth, apoptosis, differentiation and immune responses. Uniprot: P08047

SRC: Proto-oncogene tyrosine-protein kinase Src. on-receptor protein tyrosine kinase which is activated following engagement of many different classes of cellular receptors including immune response receptors, integrins and other adhesion receptors, receptor protein tyrosine kinases, G protein-coupled receptors as well as cytokine receptors. Uniprot: P12931

STAT1: Signal transducer and activator of transcription 1-alpha/beta. Signal transducer and transcription activator that mediates cellular responses to interferons (IFNs), cytokine KITLG/SCF and other cytokines and other growth factors. Uniprot: P42224

TAZ: Tafazzin. Acyltransferase required to remodel newly synthesized phospholipid cardiolipin (1',3'-bis-[1,2-diacyl-sn-glycero-3-phospho]-glycerol or CL), a key component of the mitochondrial inner membrane,

with tissue specific acyl chains necessary for adequate mitochondrial function. Uniprot: Q16635

TFF3: Trefoil factor 3. Involved in the maintenance and repair of the intestinal mucosa. Promotes the mobility of epithelial cells in healing processes (motogen). Uniprot: Q07654

TGF-ß: Transforming growth factor beta-1 proprotein. Transforming growth factor beta-1 proprotein: Precursor of the Latency-associated peptide (LAP) and Transforming growth factor beta-1 (TGF-beta-1) chains, which constitute the regulatory and active subunit of TGF-beta-1, respectively. Uniprot: P01137

TNFA: Tumor necrosis factor. Cytokine that binds to TNFRSF1A/TNFR1 and TNFRSF1B/TNFBR. It is mainly secreted by macrophages and can induce cell death of certain tumor cell lines. Uniprot: P01375

TNFRSF10A: Tumor necrosis factor receptor superfamily member 10A. Receptor for the cytotoxic ligand TNFSF10/TRAIL. Uniprot: O00220

TOMM20: Central component of the receptor complex responsible for the recognition and translocation of cytosolically synthesized mitochondrial preproteins. Uniprot: Q15388

TP53: Cellular tumor antigen p53. Acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and cell type. Uniprot: P04637

TRB1: Adapter protein involved in protein degradation by interacting with COP1 ubiquitin ligase. Uniprot: Q96RU8

Twist: Twist-related protein 1. Acts as a transcriptional regulator. Inhibits myogenesis by sequestrating E proteins, inhibiting trans-activation by MEF2, and inhibiting DNA-binding by MYOD1 through physical interaction. This interaction probably involves the basic domains of both proteins. Uniprot: Q15672

VCAM-1: Vascular cell adhesion protein 1. Important in cell-cell recognition. Appears to function in leukocyte-endothelial cell adhesion. Interacts with integrin alpha-4/beta-1 (ITGA4/ITGB1) on leukocytes, and mediates both adhesion and signal transduction. Uniprot: P19320

VEGF: Vascular endothelial growth factor A. Growth factor active in angiogenesis, vasculogenesis and endothelial cell growth. Induces endothelial cell proliferation, promotes cell migration, inhibits apoptosis and induces permeabilization of blood vessels. Uniprot: P15692

Wnt: Protein Wnt. Ligand for members of the frizzled family of seven transmembrane receptors. Uniprot: A0A384N611

YAP: Transcriptional coactivator YAP1. Transcriptional regulator which can act both as a coactivator and a corepressor and is the critical downstream regulatory target in the Hippo signaling pathway that plays a pivotal role in organ size control and tumor suppression by restricting proliferation and promoting apoptosis. Uniprot: P46937

ZEB1: Zinc finger E-box-binding homeobox 1. Acts as a transcriptional repressor. Inhibits interleukin-2 (IL-2) gene expression. Enhances or represses the promoter activity of the ATP1A1 gene depending on the quantity of cDNA and on the cell type. Uniprot: P37275

ZO-1: Tight junction protein ZO-1. TJP1, TJP2, and TJP3 are closely related scaffolding proteins that link tight junction (TJ) transmembrane proteins such as claudins, junctional adhesion molecules, and occludin to the actin cytoskeleton. Uniprot: Q07157

Publication list & CV

List of articles published during the completion of my PhD thesis, listed in chronological order.

Publications derived from the work presented in this thesis:

- *Aryl-hydrocarbon receptor-interacting protein regulates tumorigenic and metastatic properties of colorectal cancer cells driving liver metastasis.* G. Solís-Fernández et al. **British Journal of Cancer** (2022) **(*)**.
- *Spatial Proteomic Analysis of Isogenic Metastatic Colorectal Cancer Cells Reveals Key Dysregulated Proteins Associated with Lymph Node, Liver, and Lung Metastasis.* G. Solís-Fernández et al. **Cells** (2022) **(*)**.
- *Prognostic Role of Aryl Hydrocarbon Receptor Interacting Protein (AIP) Immunohistochemical Expression in Patients with Resected Gastric Carcinomas.* C. Díaz del Arco et al. **Pathologoy and Oncology Research** (2020).

Other publications not included in this thesis:

- *In-depth proteomics characterization of ΔNp73 effectors identifies key proteins with diagnostic potential implicated in lymphangiogenesis, vasculogenesis and metastasis in colorectal cancer.* M Garranzo-Asensio et al. **Molecular oncology** (2022)
- *Simultaneous electrochemical immunosensing of relevant cytokines to diagnose and track cancer and autoimmune diseases.* B Arévalo et al. **Bioelectrochemistry** (2022).
- *Binary MoS2 nanostructures as nanocarriers for amplification in multiplexed electrochemical immunosensing: simultaneous determination of B cell activation factor and proliferation-induced signal immunity-related cytokines.* B Arévalo et al. **Microchimica Acta** (2022).
- *Seroreactivity Against Tyrosine Phosphatase PTPRN Links Type 2 Diabetes and Colorectal Cancer and Identifies a Potential Diagnostic and Therapeutic Target.* M. Garranzo-Asensio et al. **Diabetes** (2022) **(*)**.
- *Phage-Derived and Aberrant HaloTag Peptides Immobilized on Magnetic Microbeads for Amperometric Biosensing of Serum Autoantibodies and Alzheimer's Disease Diagnosis.* A. Valverde et al. **Analysis & Sensing** (2021).
- *Multiplexed magnetic beads-assisted amperometric bioplatforms for global detection of methylations in nucleic acids.* E. Povedano et al. Analytica Chimica Acta (2021).
- *Multiomics Profiling of Alzheimer's Disease Serum for the Identification of Autoantibody Biomarkers.* P. San Segundo-Acosta. **Journal of Proteome Research** (2021).
- *Multiplexed Biosensing Diagnostic Platforms Detecting Autoantibodies to Tumor-Associated Antigens from Exosomes Released by CRC Cells and Tissue Samples*

Showed High Diagnostic Ability for Colorectal Cancer. A. Montero-Calle et al. **Engineering** (2021).

- *Quantification of FRET-induced angular displacement by monitoring sensitized acceptor anisotropy using a dim fluorescent donor.* D. Laskaratou et al. **Nature Communications** (2021). **Second author**
- *Electrochemical immunosensing of Growth arrest-specific 6 in human plasma and tumor cell secretomes.* C. Muñoz-San Martín. **Electrochemical Science Advances** (2021).
- *Protein Microarrays for Ocular Diseases.* G. Solís-Fernández et al. **Methods in Molecular biology** (2021) **(*)**.
- *Phage Microarrays for Screening of Humoral Immune Responses.* A. Montero-Calle et al. **Protein Microarrays for Disease Analysis** (2021).
- *Magnetic microbeads-based amperometric immunoplatform for the rapid and sensitive detection of N6-methyladenosine to assist in metastatic cancer cells discrimination.* E. Povedano et al. **Biosensors & Bioelectronics** (2021).
- *Identification of tumor-associated antigens with diagnostic ability of colorectal cancer by in-depth immunomic and seroproteomic análisis.* M. Garranzo-Asensio et al. **Journal of Proteomics** (2020).
- *Multiplexed monitoring of a novel autoantibody diagnostic signature of colorectal cancer using HaloTag technology-based electrochemical immunosensing platform*. M. Garranzo-Asensio et al. **Theranostics** (2020).
- *Raster Image Correlation Spectroscopy Performance Evaluation.* M. Longfils et al. **Biophysical Journal** (2019).
- *Single-Step Synthesis of Dual Phase Bright Blue-Green Emitting Lead Halide Perovskite Nanocrystal Thin Films.* H. Bhatia et al. **Chemistry of Materials** (2019).
- *Protein Microarrays: Valuable Tools for Ocular Diseases Research.* M. Garranzo-Asensio et al. **Current Medicinal Chemistry** (2019).
- *C(sp^3)–H Bond Activation by Perovskite Solar Photocatalyst Cell.* H. Huang et al. **ACS Energy Letters** (2018).
- *Effects of Ole e 1 on Human Bronchial Epithelial Cells Cultured at the Air-Liquid Interface.* E. Batanero et al. **Journal of Investigational Allergy and Clinical Immunology** (2018).
- *Efficient and selective photocatalytic oxidation of benzylic alcohols with hybrid organic–inorganic perovskite materials.* H. Huang et al. **ACS Energy Letters** (2018).
- *Electricity Production Plan for Belgium post-2025.* S. Burssens et al. **Transdisciplinary Insights** (2017) **(*)**.

(*) First author publications

Oral and poster contributions at conferences during the PhD:

- 19th International Microscopy Congress, poster presentation. Held in Sydney from the 9th until the 14th of September 2018.
- Joint meeting of the Belgian Society of Developmental Biology and the Royal Belgian society of Microscopy, poster presentation. Held the 21st of September 2018 in Antwerp.
- Dutch Biophysics Meeting 2018, poster presentation. Held from the 2nd to the 3rd of October 2018 in Veldhoven.
- EACR Tracking Cancer, poster presentation. Held in Barcelona from the 2nd to the 4th of February 2019.
- Royal Belgian Society of Microscopy, poster presentation. Held in Louvain La Neuve the 9th of September 2019.
- EMBO Seeing is Believing conference, poster presentation. Held in Heidelberg from the 9th to the 12th of October 2019.
- EACR tumour microenvironment conference, poster presentation. Held in Lisbon from the 2nd to the 4th of February 2020.
- EACR virtual congress, poster presentation. Held online from the 18th to the 19th of June.
- Human Proteome Organization online conference, poster presentation. Held online the 22nd of October 2020.
- Spanish Society of Proteomics online meeting 2020, oral contribution. Held online the 22nd of November 2020.
- Spanish Society of Biochemistry and Molecular Biology 2021 online meeting, poster contribution. Held online from the 19th to the 22nd of July online.
- EMBO Seeing is Believing online conference, poster presentation. Held online from the 5th to the 8th of October 2021.
- Human Proteome Organization online conference, poster presentation. Held online the 22nd of October 2021.
- Joint French, Portuguese and Spanish proteomics societies, Oral contribution and poster presentation. Held in Vilamoura from the 11th to the 13th of May 2022.

Supervised master and bachelor thesis:

- Master theses:
    - Quinten Coucke. Master in Biochemie en biotechnologie, KU Leuven. Academic year 2017-2018. Title: Studying the multimerization of Integrase during HIV infection with FLIM-based FRET phasor analysis.
    - Elfriede Herwegh. Master in Biochemie en biotechnologie, KU Leuven. Academic year 2018-2019.Title: The role of cancer associated fibroblasts in tumor invasion.

  - Charlotte Cresens. Master in Biochemie en biotechnologie, KU Leuven. Academic year 2019-2020. Title: Imaging focal adhesions in cancer metastasis via single molecule localization microscopy.
  - Stéphanie Geerts. Master in Biochemie en biotechnologie, KU Leuven. Academic year 2020-2021. Title: The influence of the ECM on Cancer Associated Fibroblasts and cancer cell interplay.
- Bachelor theses:
  - Gudrun Vermeulen. Bachelor in Biochemie en Biotechnologie, KU Leuven. Academic year 2017-2018. Title: Performance of novel fluorescent protein pairs for Förster resonance energy transfer fluorescence lifetime imaging microscopy (FRET- FLIM).
  - Kristof Flebus: Bachelor in chemistry (option Life Sciences) UHasselt. Academic year 2018-2019. Title: Fluorescent proteins for super-resolution microscopy using STED.
  - Lorien Bogaerts: Bachelor in chemistry (option Biochemistry) Uhasselt. Academic year 2019-2020. Title: Influence of p53 hotspot mutations on DNA binding in glioblastoma cell lines.

Other scientific outreach and communication activities:

- 2x Volunteer in the Dag Van de Wetenschap (Science Day) organized by the Flemish government.
- Volunteer in the open day from the Chemistry Department of the KU Leuven.
- Followed the following courses in communication and outreach:
  - Science communication and outreach (3 ECTS)
  - Science Figured Out (1.5 ECTS) on how to present scientific research effectively, lead to the publication of a scientific outreach 2-minute pitch of my thesis
  - Academic English writing (1.5 ECTS)
  - Presentation and seminar skills (1 ECTS)