Universitat Autònoma de Barcelona

Identification of MIDORI: a novel microprotein that regulates epithelial-mesenchymal plasticity

Emanuela Greco

Universidad Autónoma de Barcelona

Departamento de Bioquímica

Doctoral thesis

Identification of MIDORI: a novel microprotein that regulates epithelial-mesenchymal plasticity

Author:

Emanuela Greco

Thesis director:

María Abad, PhD

Barcelona, 2022

Departamento de Bioquímica
Universidad Autónoma de Barcelona

Doctoral thesis

Identification of MIDORI: a novel microprotein that regulates epithelial-mesenchymal plasticity

Emanuela Greco
Graduate in Biotechnology
Master in Medical biotechnology

The work presented in this thesis has been performed in the Cellular Plasticity & Cancer Group at the Vall d'Hebrón Institute of Oncology (VHIO)

Dra. María Abad Méndez
Director

Dr. Joaquín Arribas Lopéz
Tutor

A Gvanni e I miei Nonni

ACKNOWLEDGMENTS

The work presented in this thesis is the result of six years of learning, joys, disappointments but also a lot of hope of being able to contribute at least with one drop of knowledge to the immense ocean of science. Today, I realise that my thesis and my work are not my final destination, but a starting point to continue my scientific research and try to leave a worthy cultural inheritance to those who will come after me. The thesis stops here, but my research keeps going.
Both this work and all my personal growth to date have been thanks to the people who have made me get here and be what I am today. This job is so yours as mine.

First of all, I would like to thank my supervisor, María. Siempre te agradeceré el voto de confianza que me diste al cogerme. Aunque te hubieran dicho desde el principio que conseguir una beca para mi iba a ser difícil, me has acogido sin pensártelo dos veces (o igual sí), pero nunca me lo has tenido en cuenta. Contigo he aprendido mucho. Desde el punto de vista científico me has ayudado a refinar mi pensamiento crítico, a ser más rigurosa y comunicar de forma más clara mis ideas y pensamientos. Aunque, como ya sabes, mi cerebro funcione al revés que el resto de comunes mortales. En un sentido más personal, me has enseñado a no ahogarme en un vaso de agua, lidiar con todo lo que comporta la vida del investigador y, quizás, lo más importante: a buscarse la vida. Pese a que algunas veces se me haya hecho más difícil buscarme la vida por mi sola, te doy las gracias por haberme enseñado esta virtud y aún más por haberme transmitido tu profunda pasión y dedicación por este trabajo. En fin, gracias por haberme dado la oportunidad de crecer y hacer ciencia en tu laboratorio, me has aportado mucho científicamente y personalmente y espero que, al final de esta aventura, no te hayas arrepentido del voto de confianza que hiciste hace unos años.

En segundo lugar, nada de mi doctorado habría sido lo mismo sin el laboratorio 403 por completo! Aunque los años hayan pasado y algunos cambios ya hayan ocurrido, lo bueno es que determinadas cosas no cambian nunca. Quiero dar las gracias a la F0 del MAL, Elena, Olga e Iñaki por haberme acogido. En todo momento me he sentido parte de la familia y, aunque a vosotros igual os parezca obvio, deseaba agradeceros la acogida me que disteis en aquel entonces. Muchísimas gracias a la Amiga Fritz Postdoc (Elena) por haberme enseñado el castellano gochamanolero que hablo a día de hoy, y estar siempre presente soportándome tanto en las buenas como en las malas, tanto en la ciencia como en la vida real. Gracias por tu amistad, tu resolución y tu sinceridad sin par. A Olga gracias por todo tu apoyo tanto en el labo como en lo personal. Empezamos esta aventura codo con codo compartiendo escritorio y desde entonces hemos vivido muchos momentos y experiencias que me llevo para siempre. Gracias por la paciencia que has tenido en lidiar con una persona un poco “romboidal” como yo y por estar siempre dispuesta a poner de tu parte y ayudar. Al Amigo Fritz (Iñaki), gracias por ser el hombre de la paz y del acuerdo siempre. Durante estos años me has aportado muchísimo, tanto científicamente como personalmente, y aunque no hayamos quedado mucho para ver Anime como hubiera hecho con mi hipotética amiga japonesa, seis años después estoy feliz de tener un amigo vasco-madrileño gracias al cual conozco los tardígrados, sé muchas cosas de pulpos y me mantengo constantemente al día con la geografía. Gracias por compartir conmigo muchas fricadas y por ser mi referente lingüístico, ahora y para siempre.
Aunque la F0 del MAL haya sido muy importante para mí, la F1 no es para menos. A la Maripili 1 (Alba) gracias simplemente por ser tú, con tu falta total de filtros acompañada por una extrema sensibilidad. Gracias por compartir conmigo nuestro proyecto de testis y por estar siempre presente y dispuesta a ayudarme cuando lo necesite, aunque no te lo diga. Aún más, gracias por haberme sacado mil y una veces de mis bucles de negatividad, por apoyarme siempre y mandarme a la mierda cuando es necesario y, como no, por los audios de la mañana. A la Maripili 2 (Marion) gracias por ser un soporte sin igual cada vez que necesito ayuda o simplemente un hombro al cual apoyarme…con

mi copa de vino, no para llorar! Gracias por compartir conmigo la experiencia de ser la "extranjera" del grupo y por entenderme perfectamente en nuestro día-día dentro y fuera del lab cada vez que use dichos, expresiones y palabras que solo nosotras dos entendemos.
Un gracias a parte va a Sandra, mi única estudiante. Gracias por haber sido mi match perfecto en el laboratorio, por ser tan receptiva y por haberme motivado para dar lo mejor de mi como mentora. Han sido unos meses especiales los que hemos pasado juntas y me voy a llevar este recuerdo siempre. Espero que una pequeña parte de mis manías y de MIDORI se queden contigo y te acompañen siempre en tu brillante futuro.
Finalmente, el laboratorio no hubiera sido lo mismo sin nuestros lab managers. Gracias a Marta por habernos hecho de hermana mayor, por la paz y la tranquilidad que inspiras y por todos los consejos de vida que me has dado. Gracias a Lluís por organizarlo todo, por hacer mi vida de laboratorio más fácil y por cuidarnos, siempre intentando quitarnos de encima la mayor carga de trabajo posible.
Aunque el 403 haya sido mi familia, unas gracias especiales van a Mery, Andrea, Annita, Judit y todas las demás iTAGs. Gracias chicas por ser nuestras hermanas directas, por estar siempre presentes, por todos y cada uno de los desahogos que hemos tenido y por todos los Julios: pasados, presentes y futuros.
También gracias a Chiara. Grazie per esserci sempre, per riconoscere la maniera in cui chiudo la porta di Cultivos e per darmi sempre la forza e il coraggio di cui ho bisogno con il tuo sorriso contagioso. Grazie a Fabio, per ricordarmi che nessun ostacolo è insormontabile e che tutto si supera con la giusta dose di sorrisi e abbracci.
Un grazie d'eccezione va a Laura Soucek. Grazie per aver fatto parte del mio Thesis Commeettee Meeting durante questi anni e per essere parte del mio Tribunale, per i consigli scientifici e per rappresentare un modello di vita a seguire.
Un gracias especial va a mi Postdoc preferido (que ya no es Postdoc), Tian Tian. Thank you for helping me and guide me through the tedious but fantastic world of ChIP. You cheered me up and offered your scientific advice when I was a bit lost and demotivated. Thank you very much for understanding that and helping me getting over it.
En fin, hay muchísima gente en el VHIO (pasada y presente) a la cual debería darle las gracias. Faiz, Gemma, Queralt, Laia, Mireia, Sara, Juli, Carmen entre las muchas. Os doy las gracias a todos por haber hecho de VHIO mi segunda casa.

Aunque hayan sido muchas las personas que han contribuido a mi desarrollo científico y personal dentro de VHIO, no hay que olvidar todos aquellos de otras instituciones que han contribuido de la misma forma. En particular, quisiera dar las gracias a Manuel Serrano. Gracias por haber estado siempre involucrado en mi proyecto y por representar por mí un ejemplo a seguir, tanto en la vida como en ciencia.
Un agradecimiento va también a todos los componentes del laboratorio de Manuel Serrano por haber estado siempre dispuestos a echarnos una mano en todo lo posible.

Although my scientific family has been very important to me in these years, I cannot forget to mention my family outside the lab.

Primero de todo, gracias a mi familia de Barcelona. A Simo per esserci da sempre e per essere stata il pilastro fondamentale della mia vita a Barcellona. Grazie per tutti i consigli di vita, per capirmi solo guardandomi in faccia e per incavolarti con me quando sono troppo quadrata ed è necessario strapparmi alla forza dai miei schemi mentali e manie. A SilviaZanini per essere l'amica ritrovata in quel di Barcellona. Grazie per essere sempre tan compresiva e disposta ad ascoltare ogni volta che ne ho bisogno. Sei il collante che ci tiene unite e non potrei immaginare questi ultimi anni a Barcellona senza la tua presenza. A mi Sarita, grazie per essere diventata un'amica speciale negli ultimi anni che abbiamo passato insieme, per essere sempre piena di vita e rendere tutto più leggero con la tua risata. Grazie al mio nipotino Peppe. E`stato un piacere adottarti in

questi anni e ti ringrazio infinitamente per tutte le confidenze e i consigli che mi hai sempre dato. Grazie per essere parte della mia famiglia a Barcellona e per essere stato presenza costante in tutti questi anni. Y por ultimo, gracias a Nachete. Has sido mucho más que un compañero de piso. Has traído en casa la alegría y también un poquito de locura. Gracias por todas las confidencias, el apoyo y la amistad que siempre me has demostrado.

Secondo, ma non meno importante grazie alle Tetraterpene. A Silvia DF90 per essere una presenza costante nella mia vita, per i nostri viaggi d'amore, le maratone nerd durante le quali finisco sempre addormentata e per conoscermi come poche persone a questo mondo. Grazie per essere sempre stata disposta ad ascoltarmi in questi anni e per sostenermi in ogni mia decisione. A Lori Vida, grazie per essere una delle persone che meglio mi comprende a questo mondo. Condividiamo tante manie e tante follie, e sebbene siamo diverse sotto altri punti di vista, ti sento sempre vicina come quando attraversiamo la strada a braccetto (coniglietto impaurito). A Ile. Nonostante la vita non ti abbia trattata bene negli ultimi anni, so che sei sempre lì per me: pronta a farmi stare bene con un consiglio o un sorriso, così come solo tu sai fare.

Un ringraziamento importante va ad Annina e Umberto. Grazie ad Annina per tutto il tempo che abbiamo trascorso insieme, per rimproverarmi tutte le volte che lavoravo troppo e per essere sempre presente con la tua fossetta e il tuo romanaccio innocente anche adesso che viviamo lontane. Grazie a Umberto per riuscire a strapparmi una risata anche nei momenti di disperazione, per tenermi constantemente aggiornata su ogni dettaglio della tua vita e ed essere sempre pronto ad ascoltare religiosamente i miei. Grazie per essere diventato un ottimo amico, per avermi fatto forza durante la stesura della tesi e perché, grazie a te, posso vedere Gomorra senza sottotitoli.

Questa tesi e tutti i traguardi che ho raggiunto non sarebbero stati possibili senza il sostegno delle persone che da sempre fanno parte della mia vita. Grazie a Ste ed Annina per sostenermi e credere in me sempre. Grazie ad Alice e Giuseppe per condividere con me il vostro affetto e vostri consigli. A Rosamaria, grazie per essere amica mia da sempre, per conoscere tutti i lati buoni e negativi del mio carattere e per sostenermi, nel bene e nel male, da tutta la vita. A me cummari Cammen, a Robi, Agata e Damiano grazie per ricordarmi che c'è vita oltre la ricerca scientifica e la tesi e per accogliermi ogni volta che torno come se non me ne fossi mai andata.
Grazie alla Signora Vera per essere stata una constante nella mia vita, per sostenere le mie scelte e per essere una delle ragioni che mi hanno spinto a trovare la mia strada.
Grazie alla mia famiglia: Padre, Madre e Culo ma anche nonni, zie e zii e cugini che hanno creduto in me in questi anni e mi hanno sempre sostenuto affinché realizzassi i miei sogni.

Por ultimo, gracias Eneko. Nuestra amistad especial es tan vieja como este doctorado. Gracias por haber estado siempre presente y apoyarme, sobretodo en todos los momentos malos que ha habido en estos años. Gracias por pensar siempre que merezco la pena y por intentar ayudarme y cuidarme siempre a tu manera toda especial.

RESUMEN

Avances recientes en las técnicas de perfilado de ribosomas, peptidómica y análisis bioinformático han desvelado que muchas regiones genómicas que fueron anotadas como no codificantes en realidad codifican proteínas pequeñas conservadas evolutivamente. Estas proteínas, de menos de 100 aminoácidos de longitud, se han denominado microproteínas, micropéptidos o SEP (de *small-ORF encoded polypeptides*). A día de hoy, solo unos pocos se han caracterizado funcionalmente, pero se han revelado como una nueva clase de reguladores moleculares con funciones cruciales en muchos procesos celulares, como el empalme del ARN, la reparación del ADN y el metabolismo mitocondrial. Estos sorprendentes descubrimientos han abierto un nuevo campo de estudio, con miles de microproteínas con potencial clínico a la espera de ser caracterizadas.

La plasticidad celular es la capacidad de las células de cambiar su identidad y transitar entre distintos estados celulares. Se sabe que es una propiedad fundamental para el desarrollo embrionario y la regeneración de tejidos, pero ya ha quedado claro que la plasticidad celular también es crucial para las células tumorales, que se aprovechan de este proceso para aumentar su capacidad de adaptación.

En esta tesis doctoral nos planteamos como objetivo principal identificar y caracterizar nuevas microproteínas reguladoras de la plasticidad en células tumorales. Usando un algoritmo bioinformático hemos identificado MIDORI, una microproteína conservada a lo largo de la evolución codificada por *ZEB2-AS1*, un gen mal anotado como un ARN largo no codificante (lncRNA). *ZEB2-AS1* se transcribe por la señalización de TGFβ y es necesario para la transición epitelio-mesénquima (EMT). De hecho, hemos observado que la expresión de MIDORI aumenta con TGFβ, así como con estrés genotóxico y ER. Sorprendentemente, nuestros datos demuestran que MIDORI regula negativamente el programa de transcripción mesenquimal en muchos tipos celulares y bloquea la inducción de EMT. Hemos analizado el papel de MIDORI en varios procesos relacionados con la EMT. En células de cáncer de mama, MIDORI induce la formación de uniones adherentes, reduce la secreción de citoquinas pro-inflamatorias, y reduce la migración y la invasión celular *in vitro.* Mas aún, MIDORI es capaz de inhibir la colonización metastásica *in vivo*. A nivel de mecanismo, hemos observado que MIDORI reduce la activación de los efectores de la vía del TGFβ SMAD2, ERK1/2 y NF-κB, y que interactúa con MYBBP1A, un represor transcripcional de NF-κB, lo que sugiere que MIDORI podría estar bloqueando la EMT mediante la regulación de la transcripción.

En línea con el papel de MIDORI en las células cancerosas, nuestros estudios también han revelado que MIDORI mejora la reprogramación de fibroblastos embrionarios de ratón (MEFs) a células madre pluripotentes inducidas (iPSCs), un proceso que requiere una transición mesénquima-epitelial (MET). El medio acondicionado de MEFs que sobreexpresan MIDORI promueve también la reprogramación celular, lo que sugiere que MIDORI mejora la reprogramación de una manera extrínseca. Además, hemos descubierto que MIDORI endógeno se expresa en las primeras etapas de la reprogramación, cuando se produce MET, y su deficiencia reduce drásticamente la eficiencia de la reprogramación. Mas aún, hemos observado que los MEFs deficientes en MIDORI muestran una dinámica transcripcional desregulada de genes relacionados con EMT y un aumento en la expresión de los genes supresores de tumores $p16^{INK4A}$ y $p19^{ARF}$ durante la reprogramación. De hecho, los MEFs deficientes en MIDORI son propensos a entrar en senescencia celular, lo que sugiere que MIDORI regula la reprogramación, al menos en parte, al regular la entrada en senescencia celular.

En resumen, en este trabajo hemos identificado MIDORI, una nueva microproteína inducida por estrés que actúa como efector de un circuito de retroalimentación negativa que revierte la EMT. De acuerdo con lo anterior, MIDORI reduce la metástasis de los tumores de cáncer de mama y mejora la reprogramación celular. Nuestros datos arrojan luz sobre la regulación de la plasticidad epitelio-mesenquima y descubren una potencial diana para medicina regenerativa y la terapia contra el cáncer.

SUMMARY

Recent advances in ribosome profiling, peptidomics and bioinformatics have revealed that many genomic regions previously misannotated as non-protein-coding actually code for evolutionarily conserved small proteins. These proteins, smaller that 100 amino acids in length, have been called microproteins, micropeptides or SEPs (small-ORF encoded polypeptides). To date, only few of them have been functionally characterised but they have been revealed as a new class of molecular regulators with crucial functions in many cellular processes, such as RNA splicing, DNA repair and mitochondrial metabolism. These surprising discoveries have opened a new field of study, with thousands of microproteins with clinical potential waiting to be characterised.

Cellular plasticity is the ability of cells to change their original identity and transit between distinct cell states. It is known to be a fundamental property for embryonic development and tissue regeneration, but it has become clear that cellular plasticity is also crucial for cancer cells, which hijack this process to increase their adaptative capacity.

In this doctoral thesis, we aimed to identify and characterise novel microproteins as regulators of cancer cell plasticity. Using a bioinformatic pipeline, we have identified MIDORI, a microprotein conserved across evolution encoded by *ZEB2-AS1*, a gene misannotated as a long non-coding RNA (lncRNA). *ZEB2-AS1* is known to be upregulated by TGFβ signalling and it is required for the epithelial-to-mesenchymal transition (EMT). Indeed, we have observed that MIDORI protein expression increases upon TGFβ as well as upon genotoxic and ER stress. Surprisingly, we demonstrate that MIDORI downregulates the mesenchymal transcriptional programme in many cell types and blocks the induction of EMT.

Next, we analysed the role of MIDORI in several processes related with EMT. In breast cancer cells, MIDORI promotes the establishment of cell-to-cell junctions, impairs the secretion of proinflammatory cytokines and reduces cell migration and invasion *in vitro*. Moreover, MIDORI overexpression impairs metastatic colonization *in vivo*. Mechanistically, we have shown that MIDORI reduces the activation of the TGFβ effectors SMAD2, ERK1/2 and NF-κB. We have identified MYBBP1A, a NF-κB co-repressor, as a MIDORI interactor, suggesting that MIDORI could be blocking EMT by transcriptional regulation.

In agreement with the role of MIDORI in cancer cells, our studies have also revealed that MIDORI enhances the reprogramming of mouse embryonic fibroblasts (MEFs) to induced pluripotent stem cells (iPSCs), a process known to require a mesenchymal-to-epithelial transition (MET). Of note, the conditioned medium of MIDORI-overexpressing MEFs is enough to promote cellular reprogramming, suggesting the MIDORI enhances reprogramming in a cell extrinsic manner. Importantly, we have found that endogenous MIDORI is transiently expressed at the early stages of reprogramming, when MET takes place, and MIDORI deficiency drastically decreases reprogramming efficiency. Interestingly, we have observed that MIDORI-deficient MEFs display a dysregulated transcriptional dynamics of EMT-related genes and upregulate $p16^{INK4A}$ and $p19^{ARF}$ tumour suppressor genes during the reprogramming process. Indeed, MIDORI-deficient MEFs are prone to enter cellular senescence and MIDORI overexpression decrease the onset of senescence during reprogramming, suggesting that MIDORI regulates *in vitro* reprogramming -at least in part- by regulating the onset of cellular senescence.

Altogether, in this work we have identified MIDORI, a novel microprotein that is expressed upon stress and acts as an effector of a negative feedback loop that shuts down the mesenchymal programme to revert EMT. Consistently, MIDORI reduces metastasis of breast cancer cells and improves cellular reprogramming to iPSCs. Our data shed light in the regulation of the epithelial-mesenchymal plasticity and uncover a potential new target for regenerative medicine and cancer therapy.

ABBREVIATIONS

AKT	V-Akt murine thymoma viral oncogene homologue 1
Alt-ORF	Alternative ORF
ANGPTL7	Angiopoietin like 7
AP	Alkaline phosphatase
ARF	Alternative reading frame
BDFR	Bayesian false discovery rate
BMP	Bone morphogenic protein
C/EBP-α	CCAAT enhancer binding protein alpha
C/EBP-β	CCAAT enhancer binding protein beta
CAF	Cancer associated fibroblast
CCL2	C-C motif chemokine ligand 2
CCL20	C-C motif chemokine ligand 20
CCL5	C-C motif chemokine ligand 5
CDC42	Cell Division Cycle 42
CDH1	Epithelial cadherin
CDH2	Neural cadherin
CDK	Cyclin-dependent kinase
circRNA	Circular RNA
cGAS	Cyclin GMP-AMP synthase
CSC	Cancer stem cell
CSK	Cytoskeletal Buffer
cSSC	Cutaneous skin squamous cell carcinoma
DAPI	4′,6-diamidino-2-phenylindole
dORF	Downstream ORF
DOXO	Doxorubicin
ECM	Extracellular matrix
EGF	Epithelial growth factor
EMP	Epithelial-mesenchymal plasticity
EMT	Epithelial-to-mesenchymal transition
ER	Endoplasmic reticulum
ERK	Extracellular signal–regulated kinase
ESC	Embryonic stem cell
FC	Fold change
FGF	Fibroblast growth factor
GO	Gene ontology
HMGA1	High mobility group AT-hook 1
HNSCC	Head and neck squamous cell carcinoma
HRAS	Harvey rat sarcoma virus GTPase
ICAP1	Integrin cytoplasmic domain-associated Protein 1-alpha
IF	Immunofluorescence
IL-1α	Interleukin 1 alpha
IL-1β	Interleukin 1 beta
IL-6	Interleukin 6
IL-8	Interleukin 8
iPSC	Induced pluripotent stem cell
ITGB1BP1	Integrin beta 1 binding protein 1
ITIH2	Inter-alpha-trypsin inhibitor heavy chain 2
JNK	c-Jun N-terminal kinase
kDA	Kilo Dalton
KO	Knock-out
lncRNA	Long non-coding RNA
MAPK	Mitogen-activated protein kinase
MDM2	Mouse double minute 2 homologue
MET	Mesenchymal-to-epithelial transition

miRNA	Micro RNA
MMTV-PyMT	Mouse mammary tumor virus-polyoma middle tumour-antigen
MS	Mass spectrometry
mTORC	Mammalian target of rapamycin complex
MYBBP1A	MYB Binding Protein 1A
MyoD	Myogenic differentiation 1
NF-κB	Nuclear factor kappa B
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
OIS	Oncogene induced senescence
ORF	Open Reading frame
OSKM	Oct4, Sox2, Klf4, c-Myc
OXPHOS	Oxidative phosphorylation
p16INK4A	Cyclin-Dependent Kinase Inhibitor 2A
p21CIP1	Cyclin Dependent Kinase Inhibitor 1A
PDAC	Pancreatic adenocarcinoma
PDGF	Platelet-derived growth factor
PDGF-AA	Platelet-derived growth factor AA
PI3K	Phosphoinositide 3-kinase
Rb	Retinoblastoma
Ribo-seq	Ribosome-sequencing
RNA POL I	RNA polymerase I
rRNA	Ribosomal RNA
SA-βGal	Senescence-associated β-galactosidase
SASP	Senescence-associated secretory phenotype
SCNT	Somatic cell nuclear transfer
SD	Standard deviation
SEM	Standard error of the mean
SEP	sORF-encoded polypeptide
SG	Stress granules
Shh	Sonic Hedgehog
sORF	Small open reading frame
STAT	Signal transducer and activator of transcription
STING	Stimulator of interferon genes
TAK1	TGFβ activated kinase 1
TGFβ	Transforming-growth-factor beta
TNFα	Tumour necrosis factor alpha
uORF	Upstream ORF
UTR	Untranslated region
ZEB1	E-box–binding homeobox
ZEB2	E-box–binding homeobox 2
αSMA	Alpha smooth muscle actin
μCT	Micro-computed tomography

INDEX

INTRODUCTION

1. Small ORF-encoded Microproteins

1.1. The hidden microproteome

Genome sequencing has been undoubtedly one of most important achievements in biomedicine and represented the beginning of a new era, sowing the seeds for the future basic and applied science. However, it came with an enormous challenge: the extraordinary amount of data gathered by the new sequencing techniques had to be organised and annotated, and functional open reading frames (ORFs) had to be recognised while discarding the ones that are randomly generated within the genome but not translated. An useful approach for this titanic task was developed in the 1990s, when two different authors reported that a cut-off of 100-codons (i.e., 300 nucleotides) was sufficient to significantly reduce the background noise coming from random small ORFs (sORFs) in the genome of *Saccharomyces cerevisiae* (Basrai et al., 1997; Fickett, 1995). This threshold has helped the annotation of coding regions in the genomes as we have known them in the last thirty years. However, this approach had a main drawback: the 100-codons cut-off systematically filtered out many sORFs that could code for functional proteins. Moreover, it indirectly contributed to the establishment of the dogma that eukaryotic genes are monocistronic.

Surprisingly, these assumptions started to change in the last decade, when advances in bioinformatics, ribosome-profiling and peptidomics revealed that many regions assumed to be non-coding, such as lncRNAs, miRNAs, and the untranslated regions (UTRs) of mRNAs contain non-canonical ORFs that actually code for small active proteins (Andrews and Rothnagel, 2014; Brunet et al., 2020; Makarewich and Olson, 2017; Patraquim et al., 2020). These proteins have received different names, such as sORFs-encoded polypeptides (SEPs), micropeptides or microproteins. Based on their genomic location, non-canonical ORFs encoding microproteins can be classified as: 1) small ORFs (sORFs), when they are located in previously assumed non-coding transcripts, such as lncRNAs, miRNAs and circRNAs, and generally are smaller than 100 codons; or 2) alternative ORFs (alt-ORFs), located in canonical protein-coding transcripts, but different from the main annotated ORF. Alt-ORFs may overlap the main ORF (in frame or out of frame) or can be located either in the 5'UTR, named as upstream ORFs (uORFs) or in the 3'UTR, referred to as downstream ORFs (dORFs) (Figure 1) (Merino-Valverde et al., 2020).

Importantly, the identification of microproteins broke the old dogma of the monocistronic nature of eukaryotic genes and it also changed our canonical view of protein translation, since many microproteins start with a codon different from ATG (Cao and Slavoff, 2020). These surprising discoveries opened a new level of complexity, bearing enormous implications from basic research -where many fundamental concepts in biology have been revisited- to the clinical setting, given that the microproteome could be the source of new therapeutic targets and biomarkers.

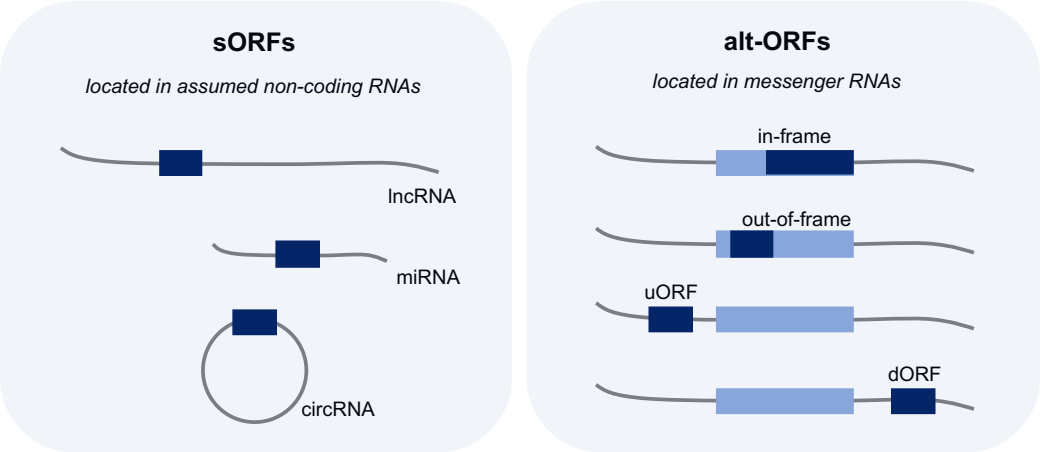


Figure 1. Classification of non-canonical ORFs. sORFs are smaller than 100 codons and are encoded by non-coding RNAs, such as lncRNAs, miRNAs or circRNAs (left panel); alt-ORFs are encoded by canonical protein-coding transcripts (right panel). They can overlap the main-ORF (light blue), in-frame or out-of-frame, or they can be located in the UTRs, named upstream ORFs (uORFs) when they are in the 5′ UTR or downstream ORFs (dORFs) when they are in the 3′ UTR.

1.2. Identification of new microproteins

At present, there are several computational and experimental approaches for the discovery of microproteins. Computational approaches rely on indicators of functionality such as codon conservation across evolution and the presence of domains described in already previously identified proteins. Some bioinformatic algorithms have been developed based on these criteria, such as PhyloCSF, that uses phylogenetic conservation and codon substitution frequencies to predict coding regions (Figure 2). The phylogenetic conservation of codons in regions annotated as noncoding -which are generally non conserved (as lncRNAs)- could mean functionality and, therefore, is a good indicator of coding potential. Once a novel microprotein is predicted, additional features suggestive of meaningful sORF can be analysed, such as the presence of a Kozak sequence, homology with other known proteins or protein domains as well as N-terminal sequences. Although codon conservation is highly suggestive of actual translation, this approach needs to be supported with further experimental evidence.

Experimental methods provide direct experimental evidence of the translation of a new microprotein. One of them is mass-spectrometry (MS), by which proteins of a biological sample are digested into small peptides that are analysed to determine their specific amino acidic spectrum. In common MS experiments, the spectra are aligned to databases containing the spectra of already known proteins, normally using available databases such as UniProt (Figure 2). However, most microproteins are not annotated in normal protein databases, and therefore a customised database that include microproteins is needed during *de novo* identification by MS. One way to generate the database is using transcriptomic data and predict *in silico* all the microproteins that can be potentially translated. Although there has been great technical advance in peptidomics, many microproteins could be overlooked during the identification by MS. In fact, microproteins' small size limits the number of tryptic peptides that are unique and distinguishable for a candidate. Additionally, many times the small size is accompanied

by low abundance compared to other bigger proteins, probably due to the fact that classic proteomic approaches select the top mass-intensities in the analysis resulting in a bias towards highly abundant peptides (Makarewich and Olson, 2017; Peeters and Menschaert, 2020; Slavoff et al., 2013; Zhu et al., 2018).
Another experimental approach used for microprotein identification is ribosome sequencing (Ribo-seq), which identifies actively translated regions of the genome. In this technique, isolated ribosomes bound to transcripts undergo a nuclease treatment prior to sequencing. In this way, regions that are not protected by the ribosomes and, therefore, non-translated are eliminated before the sequencing (Figure 2). The obtained Ribo-seq data are then evaluated using algorithms such as RibORF, that scores ribosome pausing in the ORF based on codon frequency and periodicity (Ruiz-Orera et al., 2018).

The combination of both computational and experimental approaches has allowed the discovery of many novel microproteins. Nevertheless, none of these methods evaluates the biological relevance of the identified microproteins and a functional characterisation needs to be performed case by case. To date, only a small subset of microproteins have been functionally characterised, but they have been demonstrated to play unique biological functions as fine-tune regulators of complex biological processes.

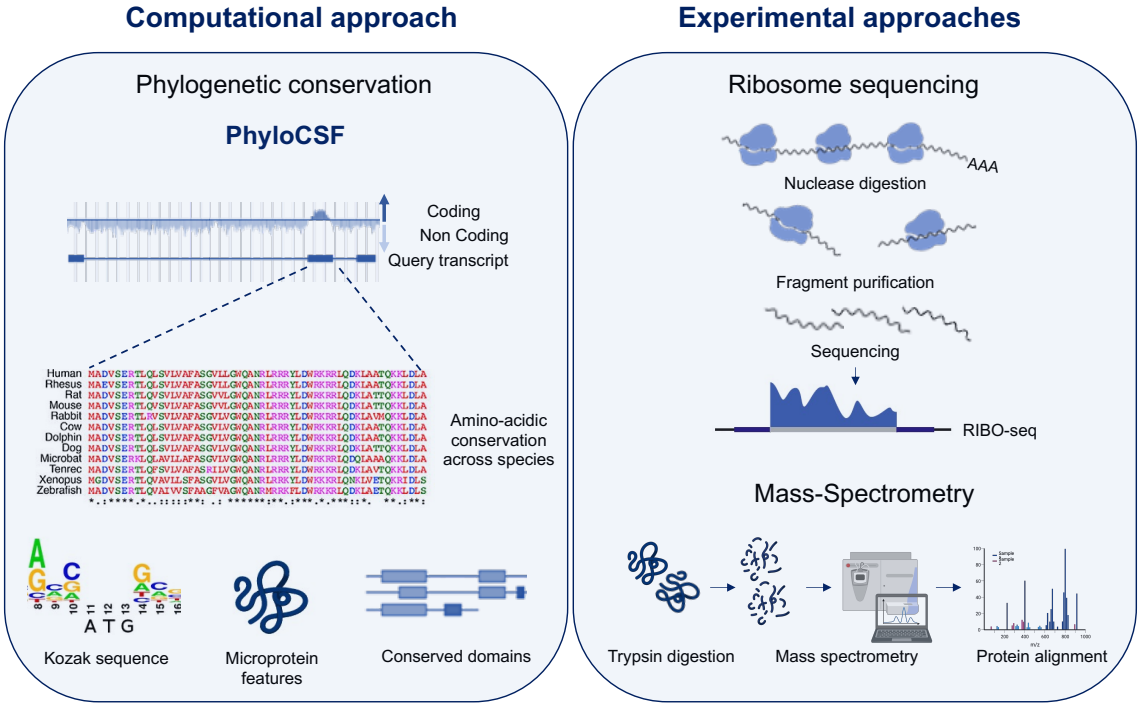


Figure 2. Microprotein discovery strategies. Methods for microprotein discovery can be divided into computational or experimental approaches. While computational approaches evaluate phylogenetic codon conservation and analyse predicted protein characteristics, experimental approaches give direct evidence for protein existence, by direct identification of the microprotein spectrum by MS analysis or by the evaluation of active translation of given transcript using Ribo-Seq.

1.3. Microproteins' biological functions

1.3.1. Microproteins as regulators of cellular processes

Microproteins have been demonstrated to play key functions in several biological processes. As far as we know, they lack enzymatic activity, but their reduced size makes them perfect for allosteric regulation of big macromolecular complexes. This is the case of the regulin family of microproteins composed by Myoregulin, Phospholamban, Sarcolipin, Endoregulin, Another-Regulin and DWORF. All of them have been shown to regulate intracellular calcium dynamics in a tissue specific manner through their interaction with the Calcium-ATPase SERCA and the regulation of its conformational changes (Anderson et al., 2015; Anderson et al., 2016; Nelson et al., 2016). Similarly, the MRI-2 microprotein has been shown to stimulate non-homologous end joining (NHEJ) by interaction with Ku heterodimeric proteins, favouring the conformational stabilisation of the complex and, in turn, the process of DNA-repair (Slavoff et al., 2014). Others microproteins, such as Mitoregulin or SPAR, are structural components of macromolecular complexes. Mitoregulin regulates mitochondrial respiration by stabilising the formation of mitochondrial respiratory super complexes, whereas SPAR regulates amino acid metabolism by the interaction with the lysosomal v-ATPase, in this way regulating mTORC1 activation (Makarewich et al., 2018; Matsumoto et al., 2017; Stein et al., 2018). Other cellular processes regulated by microproteins are "Nonsense mediated decay" and P-bodies formation (D'Lima et al., 2017), mRNA splicing (Huang et al., 2017), fatty acid-oxidation (Makarewich et al., 2018) and transcriptional regulation (Koh et al., 2021) .

Interestingly, microproteins can also act as signals for intracellular or intercellular communication. For example, PIGBOS helps to regulate organelle interactions within the cell. In fact, it regulates the ER-Mitochondrial tethering allowing cells to cope with ER stress and resist to apoptosis (Chu et al., 2019). Moreover, microproteins can be secreted outside the cells as it is the case of Toddler/Elabela and Pegasus. While the first ones are *bona fide* ligands of the apelin receptor and act as mitogens during gastrulation and cardiac development (Chng et al., 2013; Pauli et al., 2014), Pegasus enhances Wingless/Wnt1 protein short-range diffusion and signalling during *Drosophila Melanogaster* wing development (Magny et al., 2021).
There are some processes that have been consistently demonstrated to be regulated by several microproteins, such as muscle function (Anderson et al., 2015; Anderson et al., 2016; Bi et al., 2017; Magny et al., 2013; Makarewich et al., 2018) mitochondrial processes (Makarewich et al., 2018; Stein et al., 2018; Zhang et al., 2020), embryonic development (Chng et al., 2013; Galindo et al., 2007; Magny et al., 2021; Pauli et al., 2014) and immunity (Bhatta et al., 2020; Chu and Saghatelian, 2019; Pueyo et al., 2016).

Summarising, microproteins have been observed in different cellular processes acting as regulators of major protein complexes, as well as intracellular and extracellular messengers. Importantly, the deficiency of some microproteins, such as Toddler/Elabela, has been demonstrated to induce embryonic lethality (Chng et al., 2013; Pauli et al., 2014), demonstrating their physiological relevance.

1.3.2. Microproteins in disease

Just as microproteins have been linked to physiological processes, they can also play a role in pathological contexts. Some microproteins have been found to be regulated in pathological conditions. For instance, Humanin microprotein is downregulated in Alzheimer's disease and it has been linked to neuroprotective effects (Chai et al., 2014; Hashimoto et al., 2001). On the contrary, LINC01013 microprotein is upregulated in cardiac fibrosis and it has been shown to exacerbate the pro-fibrotic phenotype of activated fibroblasts (Quaife et al., 2022).

1.3.2.1. Microproteins in cancer

Cancer is a complex and multistep disease that profoundly affects the biology of the cell. Several microproteins have been described to play a direct role at different stages of tumorigenesis, from tumour initiation till metastasis. In addition, there are some others described in other contexts that, given their molecular function, we hypothesise that can be regulating different aspects of cancer biology as well (Figure 3). We have recently reviewed these evidences in (Merino-Valverde et al., 2020). Importantly, the discovery of novel microproteins in the context of cancer contributes to advance our understanding of the disease and constitutes a novel source of clinical targets.

As in the case of our canonical proteome, microproteins can behave as *bona fide* oncogenes or tumour suppressors. On the one hand, many microproteins display oncogenic activity. For example, CASIMO1 has been shown to activate the MAPK cascade in breast cancer (Polycarpou-Schwarz et al., 2018), whereas in colorectal cancer, CircPPP1R12A-73aa increases cancer cell proliferation through activation of the Hippo-Yap pathway (Zheng et al., 2019) and ASAP promotes cancer progression modulating ATP synthase activity (Ge et al., 2021). Finally, the microprotein APPLE increases the proliferation of hematopoietic cancers by enhancing translation initiation (Sun et al., 2021).

On the other hand, several microproteins are downregulated in cancer and display tumour suppressor activity. The first one identified was HOXB-AS3; this microprotein has been shown to be downregulated in colorectal cancer and its overexpression attenuates cancer cell metabolic reprogramming by altering their glycolytic capacity (Huang et al., 2017). Other examples of tumour suppressor microproteins include PINT-87aa, that is downregulated in glioblastoma and acts as a transcriptional repressor of several oncogenes (Zhang et al., 2018) and YY1BM, a microprotein that is downregulated in Oesophageal Squamous Cell Carcinoma (ESCC) and sensitises cells to apoptosis upon nutrient deprivation (Wu et al., 2020).

Interestingly, several microproteins have been directly related to cancer invasion and metastasis while some others, based on their reported function, might be inferred to regulate the metastatic process as well. Regarding the microproteins with pro-metastatic function, ZFAS1 has been shown to promote cell migration by increasing reactive oxygen species (ROS) production in hepatocellular carcinoma (HCC) (Guo et al., 2019), and STMP1 increases lamellipodia formation and enhances the mitochondrial fission protein machinery (Xie et al., 2022). On the other side, MIAC microprotein inhibits Head and

Neck Squamous Cell Carcinoma (HNSCC) progression and metastasis by inactivating the actin cytoskeleton through interaction with Aquaporin-2 (Li et al., 2020).
Given the pivotal role of TGFβ (Transforming-Growth-Factor β) in metastasis, through the regulation of the epithelial-to-mesenchymal transition (EMT) (David and Massague, 2018), microproteins acting as effectors or targets of the TGFβ pathway may act on metastasis as well. To date, two microproteins have been characterised as directly controlled by TGFβ signalling. The first is CIP2A-BP, whose translation is inhibited by TGFβ. In fact, this microprotein has been described to negatively regulate triple negative breast cancer metastasis by competitive binding to CIP2A protein, which results in the downregulation of key proteins for metastasis, as *SNAIL* or matrix metalloproteinases (Guo et al., 2020). The second is LINC01013 microprotein, whose expression is enhanced by TGFβ signalling. This microprotein enhances fibroblast activation during cardiac fibrosis, thus increasing extracellular matrix (ECM) deposition and the secretion of cytokines and factors that, in turn, exacerbate the fibrotic disease (Quaife et al., 2022). Although this microprotein has not been directly linked to metastasis yet, it might also play a role in a cell extrinsic manner by the secretion of cytokines to the microenvironment.
Altogether, these evidence show the importance of the microproteome in the context of metastasis and, more broadly, in the context of cancer (Figure 3).

Finally, regarding the translation of this new field into the clinic, given that many microproteins can act as oncogenes or tumour suppressors, their expression or inactivation in determined cancer types could represent promising therapeutical approaches (Zhu et al., 2018a). In addition, some microproteins have been reported to be secreted. We have described above the role of secreted microproteins in development (see *Section 1.3.1.*), but others display immunoregulatory functions (Hu et al., 2021) or are involved in cancer growth (Prensner et al., 2021). Thus, secreted microproteins that are detectable in the body-fluids might be used as a source of novel biomarkers in early cancer detection as well as in patients' stratification or prognosis.
Of note, in the same way that canonical proteins accumulate mutations during tumorigenesis, microproteins can also be the target of cancer-driving mutations. In agreement, some studies have revealed that mutations in non-coding regions can act as driver mutations in cancer (Hu et al., 2018; Rheinbay et al., 2020). It is possible that driver mutations in previously assumed non-coding regions can actually affect the function of microproteins yet-to-be-discovered. Moreover, mutated microproteins can be the source of tumour-specific antigens (neoantigens) which might be exploited for cancer immunotherapy (Carbonnelle et al., 2013; Chong et al., 2020; Laumont et al., 2018).
Concluding, microproteins have been revealed as a new class of bioactive molecules that regulate a plethora of cellular processes. The discovery of the microproteome has expanded our vision on the coding capacity of the genome, adding a new layer of biological complexity with a huge impact on basic, translational, and clinical research.

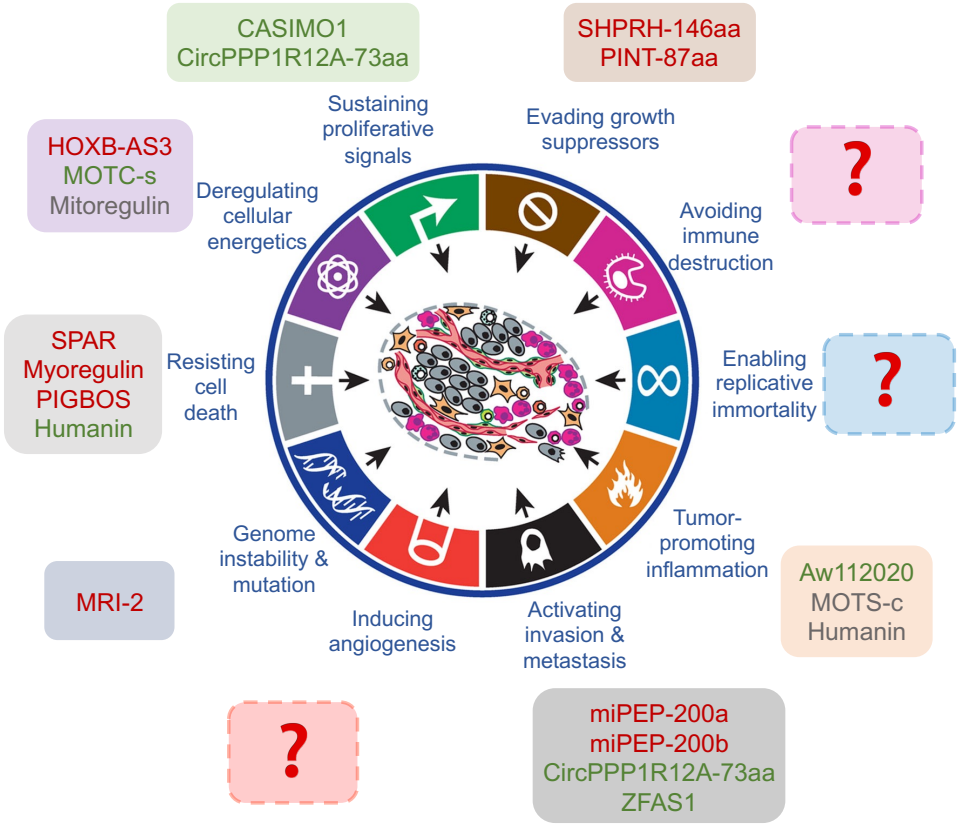


Figure 3. Microproteins as novel regulators of the hallmarks of cancer. A subset of microproteins have been functionally characterised and have been directly related with cancer; some others, based on their function, are likely to be related with cancer. The figure represents the hallmarks of cancer defined by Hanahan and Weinberg and their related microproteins. In green, the microproteins that promotes or activate the hallmark and, in red, the microproteins that function inhibiting or blocking the hallmark. The ones in grey need further investigation to be classified. Of interest, those depicted in red represent tumour suppressor microproteins with potential pharmacological activity, while those in green are pro-oncogenic proteins that could be targeted in the clinic.

2. Cellular Plasticity

Cellular plasticity is a fundamental characteristic of biological systems. It refers to the ability of cells to change their original identity and transit between different cell states.

In the same way procaryotes can acquire different cellular characteristics through horizontal gene transfer, eukaryotes evolved mechanisms enabling plasticity. However, rather than permitting exchange of genetic material, eukaryotic cells maintain their genome integrity and instead rely on processes such as epigenetic regulation or mRNA splicing to diversify cellular functions and acquire distinct phenotypic states (Torborg et al., 2022).

Consistent with this notion, all cells within an organism harbour the genetic information required for acquiring plasticity, what changes, in the end, is the gene expression pattern and the epigenetic landscape that make a specific cell unique and specialised in a determined cellular state.

A paradigmatic example of cellular plasticity is represented by embryonic stem cells (ESCs). These cells are transiently present during embryonic development, and they are

pluripotent, meaning that they have the potential to differentiate into every cell type of the body. The progression of embryonic development is accompanied by an overall decrease in differentiation potential and adult tissues only possess multipotent stem cells, that are able to generate a limited number of terminally differentiated cell types. Finally, terminally differentiated cells have specific functions and under most circumstances, they maintain their cellular identity stable. Nevertheless, fully formed adult tissues retain a certain degree of cellular plasticity. In fact, upon specific environmental cues (i.e., upon tissue injury), adult cells can acquire plasticity transiting between different cellular states, which allows tissue repair (Jopling et al., 2011; Raff, 2003; Rognoni and Watt, 2018). In this scenario, different kinds of conversion are possible: on the one hand, differentiated cells can go back to a more dedifferentiated state, through a process called dedifferentiation (Merrell and Stanger, 2016). On the other hand, cells can be converted into another differentiated cell type through a process named transdifferentiation, including the transition from epithelial phenotypes to more mesenchymal ones and vice versa in the processes known as epithelial-to-mesenchymal transition (EMT) and mesenchymal-to-epithelial transition (MET) (Merrell and Stanger, 2016) (Figure 4).

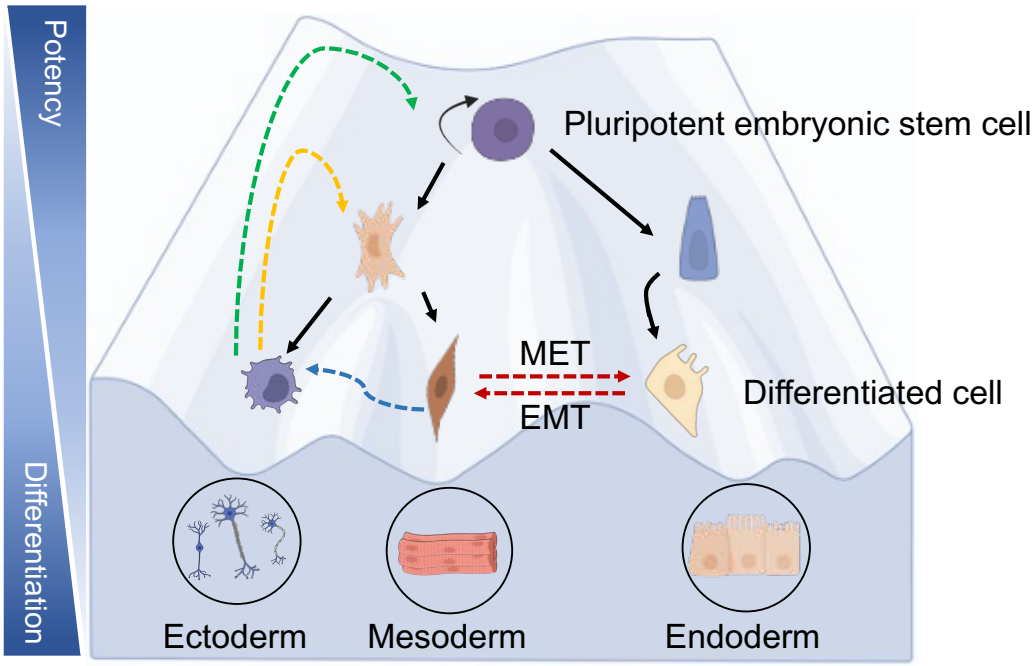


Figure 4. Cellular hierarchies and the conversions between them by cellular plasticity. Adaptation of the Waddington landscape in which the differentiation stages and the possible conversions between them are illustrated. The yellow and green arrows show the cellular dedifferentiation processes, in which cells gain differentiation potential. This process can be partial, generating a cell with limited potency (yellow arrow) or can generate a pluripotent stem cell, which can differentiate into the three germinal layers (in this case the process is known as cellular reprogramming, green arrow). The blue and red arrows represent the transdifferentiation process, by which a terminally differentiated cell loses its cellular identity and is transformed into a different somatic cell type (blue) or loses its epithelial characteristics to gain mesenchymal features and vice versa (red arrows), in the processes of EMT and MET.

2.1. Epithelial-mesenchymal plasticity (EMP)

Epithelial-mesenchymal plasticity (EMP) describes the capacity of cells to interconvert between epithelial and mesenchymal phenotypes. During the EMT, epithelial cells lose their defining characteristics, such as stable cell–cell junctions and apico–basal polarity, and gain the capacity to migrate independently and invade through extracellular matrices (Nieto et al., 2016). Additionally, cells that undergo EMT are characterised by cell cycle arrest and cessation of proliferation (Kerosuo and Bronner-Fraser, 2012), while increasing ribosome biogenesis to sustain their new protein synthesis needs (Prakash et al., 2019).These drastic changes in the cell identity are possible thanks to a tightly regulated transcriptional programme that is activated in different stages of the embryonic development (Nieto et al., 2016; Thiery et al., 2009).

The EMT is a pleiotropic cellular response and a EMT molecular signature has been difficult to establish. Among all the molecular changes, a consensus signature comprises the following features: the loss of the epithelial junctional protein E-CADHERIN (CDH1) and the increased expression of neural CADHERIN (N-CADHERIN; CDH2); increased expression of the intermediate filament protein VIMENTIN (VIM); and the induction of the EMT transcription factors SNAIL, SLUG, ZEB1, ZEB2 and TWIST1 (Figure 5) (Nieto et al., 2016). Of note, beyond these effectors, other non-canonical transcription factors have also been shown to regulate EMT (Stemmler et al., 2019). In fact, many signalling pathways are known to induce or regulate EMT in different cellular contexts. Among them, TGF-β is probably the most characterised inducer but also BMPs, EGF, FGF, PDGF, Wnt, Sonic Hedgehog (Shh), Notch, and integrin signalling are well-known EMT inducers (Al Moustafa et al., 2012; Espinoza and Miele, 2013; Gonzalez and Medici, 2014; Heldin et al., 2012; Katoh and Katoh, 2009; Taipale and Beachy, 2001) .

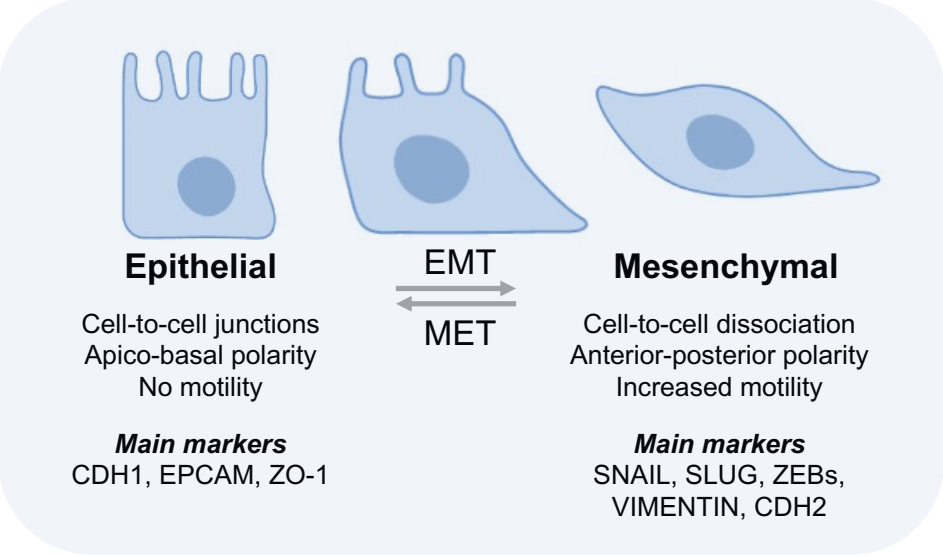


Figure 5. Main characteristics and markers associated with EMT/MET transitions.
Schematic view of the morphological, functional, and canonical gene expression changes associated with the epithelial-to-mesenchymal transition (EMT) and with the mesenchymal-to-epithelial transition (MET).

Of note, other signals from the ECM and hypoxic conditions have been demonstrated to induce EMT as well (Philip et al., 2013).

All these pathways signal through intracellular kinase cascades to induce the transcription factors enumerated above and activate the expression of EMT-related genes. In a very simplified picture, it is possible to distinguish two different activation cascades: a SMAD-dependent (canonical) cascade, in which phosphorylated SMAD2/3 translocate to the nucleus and together with SMAD4 activate the transcription of EMT-related genes (Massague, 2000). And a SMAD-independent (non-canonical) cascade, which relies on the MEK-ERK signalling, PI3K-Akt and TAK1 that in turn can activate NF-kB, JNK and p38-MAPK signalling pathways (Zhang, 2017) (Figure 6).

Importantly, even though we have tended to reduce the processes of EMT-MET to a binary interconversion, nowadays it is clear that several partial EMT states exist (Jolly et al., 2016; Jordan et al., 2011). Thus, the epithelial-mesenchymal plasticity (EMP) brings into the equation the existence of intermediate hybrid epithelial and mesenchymal phenotypes. According to this vision, EMT can be considered as a continuum, whereby cells exhibiting epithelial, intermediate, and mesenchymal phenotypes coexist. Importantly, although full reversibility of the EMT is achievable during embryonic development (Nieto et al., 2016; Thiery et al., 2009) , it is not clear whether this also occurs when the EMT programme is activated in an aberrant way during the adult life, for instance in cancer and fibrotic disease.

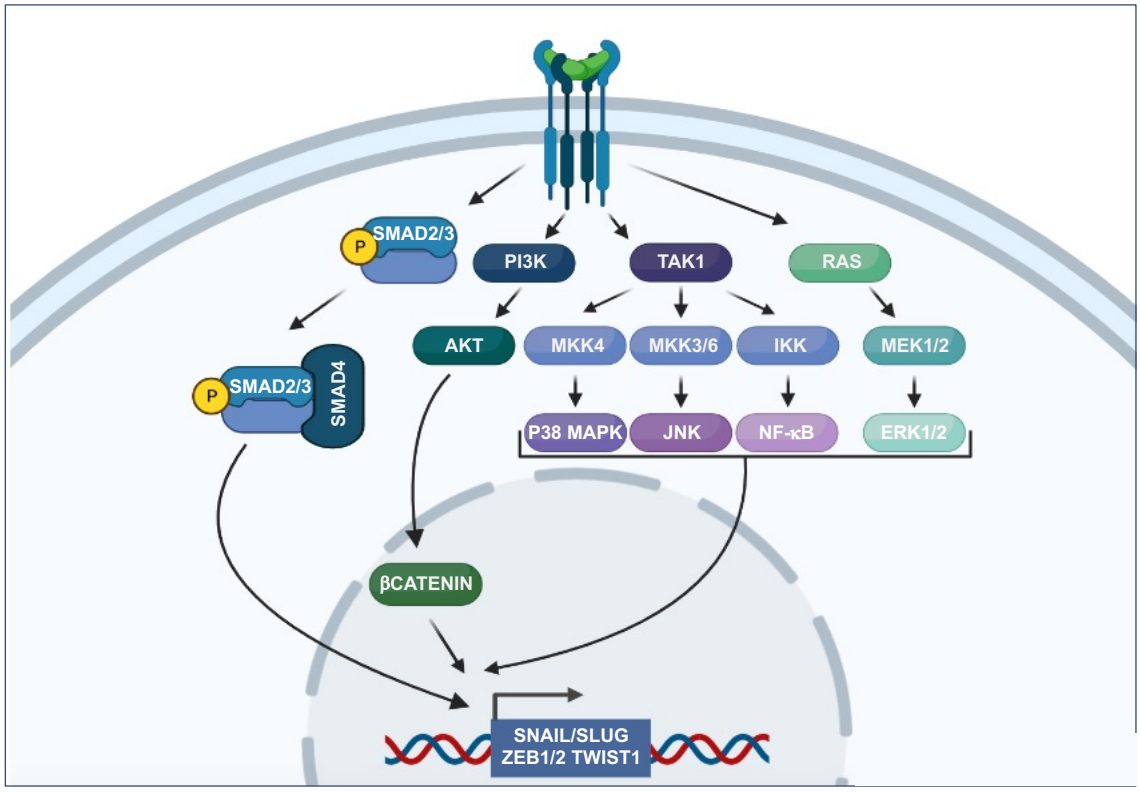


Figure 6. Signalling pathways inducing EMT. When the TGFβ binds to its receptor, this activates various signalling pathways. The canonical pathway is mediated by phosphorylation of SMAD2/3, the formation of a ternary complex with SMAD4, and the shuttle to the nucleus where they control the transcription of EMT-related genes. The non-canonical pathway could be mediated by Ras, which in turn activates ERK1/2, by PI3K-Akt which allows β-catenin translocation to the nucleus and by TAK1 that could activate NF-κB, JNK and p38 MAPK pathways.

2.1.1. EMP and Cancer

As mentioned above, although EMT is essential for development, it can have detrimental consequences in the adult organism, promoting the progression of diseases such as fibrosis and cancer. Since carcinomas are tumours that arise from epithelial tissues, the progression from localised tumour to invasive carcinoma and distal metastasis requires EMT. Importantly, while EMT is needed for metastasis, it must be reversed to colonise a new organ through MET (Hanahan and Weinberg, 2011; Nieto et al., 2016).However, according to the "EMP vision", the conversion is not full and cancer cells can acquire mesenchymal features while continuing to express epithelial traits, resulting in a selective advantage during the metastatic process (Grigore et al., 2016). In fact, such phenotypic plasticity and heterogeneity might provide cancer cells with increased adaptability and resistance, enabling them to respond to the different cues they may encounter while proceeding in the multiple steps of the metastatic cascade (Yang et al., 2020). For example, various epithelial-mesenchymal hybrid phenotypes may provide a survival advantage in distinct microenvironments, such as blood and lymphatic vessels as well as primary and secondary tumour sites (Luond et al., 2021; Pastushenko et al., 2018).
Of note, EMP in cancer has been also associated with the acquisition of additional traits that are not linked to the canonical EMT programme such as stemness features and therapy resistance (Goossens et al., 2017; Puisieux et al., 2014). In line, EMT signature establishment has been linked to the acquisition of stemness property in many tissues and cancer types (Dongre and Weinberg, 2019). Thus, some properties commonly attributed to cancer stem cells (CSCs), such as invasiveness and metastatic potential, may be acquired by the EMP (Gupta et al., 2019).

Finally, EMP has also emerged as a contributor to therapy resistance. Particularly, EMP could contribute to drug resistance by enhancing drug efflux, reducing cell proliferation, evading apoptosis signalling pathways and immunoevasion (Dongre and Weinberg, 2019; Gupta et al., 2019)
Altogether, these observations point to a central and critical role of EMP in cancer. In the next years the challenge will be to apply all the knowledge generated in the last decades to the new model of epithelial-mesenchymal plasticity. Only in this way, we will be able to specifically target the cells able to execute this programme and, thus, overcome metastasis and therapy resistance, which are major challenges in cancer treatment.

2.2. Cellular Reprogramming

2.2.1. History of cellular reprogramming

Perhaps the most striking example of acquired cellular plasticity is the ability of the so-called "Yamanaka factors" (OCT4, SOX2, KLF4, and c-MYC) to induce pluripotency when overexpressed in differentiated cells, in a process named cellular reprogramming (Takahashi and Yamanaka, 2006). Importantly, this amazing discovery that revolutionised forever the stem cell field is the culmination of decades of research willing to reprogramme terminally differentiated cells. In fact, this journey started at the end of the 1950s, when John Gurdon demonstrated that an adult differentiated cell could be

experimentally forced back to an embryonic state by transferring its nucleus into an enucleated oocyte, a technique called somatic cell nuclear transfer (SCNT). That was the first demonstration that cellular identity is determined epigenetically and it could be changed (Gurdon et al., 1958). Another finding that contributed to the discovery of reprogramming was the observation that lineage-associated transcription factors, which help to establish and maintain cellular identity during development and in the adult life, can change cell fate when ectopically expressed in certain heterologous cells. This process, known as transdifferentiation, was formally demonstrated with the interconversion of fibroblast cell lines into myofibers by the ectopic expression of the skeletal muscle factor MyoD (Davis et al., 1987).

With those ideas in mind, Yamanaka and Takahashi designed a screen to identify transcriptional regulators that can reprogramme adult cells into pluripotent cells, starting with a pool of 24 pluripotency-associated candidate genes. The combination of 24 factors, co-expressed by retroviral vectors in embryonic mouse fibroblasts (MEFs), indeed induced the formation of colonies with ESCs characteristics (Takahashi and Yamanaka, 2006). Successive rounds of experiments eliminating individual factors led to the identification of the minimally required core set of four genes comprising KLF4, SOX2, c-MYC, and OCT4, known today as the Yamanaka factors or OSKM. Importantly, these reprogrammed cells, named induced pluripotent stem cells (iPSCs), are undistinguishable from ESCs. In fact, iPSCs can be maintained and propagated in a pluripotent state (they self-renew), and they could also be differentiated into the three germ layers: endoderm, mesoderm and ectoderm (Takahashi and Yamanaka, 2006).

Importantly, one year later cellular reprogramming was also achieved in human somatic cells by using the combination of OCT4, SOX2, KLF4 and LIN28 pluripotency factors (Yu et al., 2007) . Since then, a plethora of different cell types in several species have been successfully reprogrammed to iPSCs (Aoi et al., 2008; Di Stefano et al., 2014; Seki et al., 2010).

Of note, Yamanaka and Takahashi achieved the reprogramming to iPSCs *in vitro,* under very controlled conditions, and this process was thought for a long time to be totally artificial and not translatable to the *in vivo* context. However, in 2013, María Abad in the laboratory of Manuel Serrano generated a transgenic mouse model that allows the inducible expression of the four Yamanaka factors (hereafter abbreviated as i4F), and demonstrated that *in vivo* conditions are also permissive to cellular dedifferentiation and somatic cells within tissues can also acquire pluripotency (Abad et al., 2013). Even more strikingly, they observed that *in vivo* generated iPSCs could also contribute to trophectoderm lineage and that they were able to generate embryo-like structures that express extra-embryonic markers, suggestive of totipotency features. Thus, the four Yamanaka factors induce an even more de-differentiated and plastic phenotype when ectopically expressed *in vivo* (Abad et al., 2013).

2.2.2. Molecular mechanisms governing cellular reprogramming

The reprogramming of somatic cells by the transduction of the OSKM pluripotency factors is normally a poorly efficient process (Masip et al., 2010). More recently, it has been demonstrated that the depletion of some factors, for instance Mbd3, a core member of the Mbd3/NuRD repressor complex, can increase the reprogramming efficiency up to almost one hundred percent (Rais et al., 2013) .

In fact, cellular reprogramming implies the overcoming of the barriers that normally guard cellular identity and the erasure of the epigenetic landscape in the reprogrammed cells. Owing this limitation, a huge effort has been made in the last decades to dissect and understand the molecular mechanisms underlying reprogramming.

Several models have been proposed to explain the molecular bases of reprogramming. Among those, there is the stochastic-deterministic model. According to this model, reprogramming can be divided in two phases. In the early phase, OSKM factors binds many genomic loci to disrupt the somatic cell identity (Koche et al., 2011; Soufi et al., 2012). For example, the fibroblast cell identity gene Thy1 is actively repressed (Stadtfeld et al., 2008) and cells lose their mesenchymal features to undergo a MET (Li et al., 2010). This phase is highly inefficient and stochastic, explaining the low efficiency observed in most of the reprogramming protocols (Masip et al., 2010). In the second phase, late pluripotency genes, such as telomerase and Nanog, and the novo DNA methyltransferases (DNMT) are activated (Brambrink et al., 2008; Papp and Plath, 2013; Polo et al., 2012). At this point, reprogrammed cells become stable and their chromatin status resembles that of ESCs (Papp and Plath, 2013). Once the endogenous pluripotency genes are re-expressed, the exogenous OSKM factors become silenced as part of a general embryonic programme that shuts down retroviral promoters. Additionally, iPSCs undergo several changes that make them more similar to ESCs than their cells of origin. In fact, they show self-renewal capability and undergo a metabolic switch that enables them to manage with their new energetic needs (Folmes et al., 2011; Panopoulos et al., 2012; Teslaa and Teitell, 2015; Vander Heiden et al., 2009).

Although cell intrinsic factors have a direct impact on cellular reprogramming efficiency, the microenvironment is an important determinant that has been demonstrated to affect reprogramming in a non-cell autonomous manner. For example, the extracellular matrix has been shown to play a role in offering the right mechano-transduction stimuli needed to promote reprogramming (Gerardo et al., 2019). In addition, secreted factors have been demonstrated to play a crucial role in cellular reprogramming. That is the case of IL-6, a cytokine that has been shown to improve reprogramming both *in vitro* and *in vivo* (Brady et al., 2013; Chiche et al., 2017; Mosteiro et al., 2016; Mosteiro et al., 2018). Importantly, upon OSKM induction, IL-6 is secreted by cells that activate one of the barriers to reprogramming: cellular senescence. This is an an irreversible, non-proliferative, metabolically active state that plays a dual role in reprogramming: it acts as a barrier in cell intrinsic context, whereas it favours reprogramming in a cell extrinsic manner (see *Section 2.2.2.2-3*) (Chiche et al., 2017; Mosteiro et al., 2016; Mosteiro et al., 2018).

2.2.2.1. Molecular barriers to cellular reprogramming: the EMT

As briefly described above, the EMT and MET programmes are activated during reprogramming into iPSCs. Thus, the epithelial or mesenchymal status of the cell undergoing reprogramming may affect iPSC generation. In fibroblasts, which are mesenchymal cells, the sequential introduction of OSKM showed that an initial EMT-like state is induced and is beneficial for reprogramming (Liu et al., 2013). In line with these data, during B cells reprogramming C/EBP-α transduction increases the efficiency of OSKM-mediated reprogramming by inducing an early EMT (Di Stefano et al., 2014).

Thus, it is possible that an initial EMT might synchronise or prime cells for reprogramming (Pei et al., 2019).

Although some evidences highlight a beneficial role of EMT during the very first step of reprogramming, it is known that it acts as a barrier at later stages (Li et al., 2010). In fact, MET is required during cellular reprogramming in mouse embryonic fibroblasts (MEFs) and human fibroblasts (Hoffding and Hyttel, 2015; Subramanyam et al., 2011) and its blockade results in impaired reprogramming efficiency (Li et al., 2010).
During cellular reprogramming, SOX2, OCT4 and c-MYC suppress TGF-β–Snail signalling, whereas KLF4 induces the epithelial programme, altogether downregulating mesenchymal proteins and upregulating epithelial proteins (Li et al., 2010). In agreement with these data, BMP-driven MET has also been described to improve reprogramming (Samavarchi-Tehrani et al., 2010) whereas Tesk1 and Limk2, which promotes actin polymerisation in fibroblasts, decreases reprogramming efficiency (Sakurai et al., 2014).
Importantly, MET-like changes are also observed during the reprogramming of other somatic cell types. For instance, mouse B cells (Di Stefano et al., 2014), hematopoietic stem cells and progenitors (Gao et al., 2016) and spermatogonial stem cells (An et al., 2017) have all been demonstrated to undergo a MET during cellular reprogramming, while EMT acts as a barrier.

Finally, EMT is a barrier to cellular reprogramming also in a non-cell autonomous manner. In fact, it is known that the presence of TGFβ1, 2 and 3 cytokines is detrimental for reprogramming, whereas the inhibition of TGFβ pathway not only improves reprogramming but also makes SOX2 and c-MYC dispensable during the process (Liu et al., 2013; Maherali and Hochedlinger, 2009).
Taken together, these data clearly point to the acquisition of an epithelial-like status as a prerequisite for cells to reach pluripotency (Figure 7).

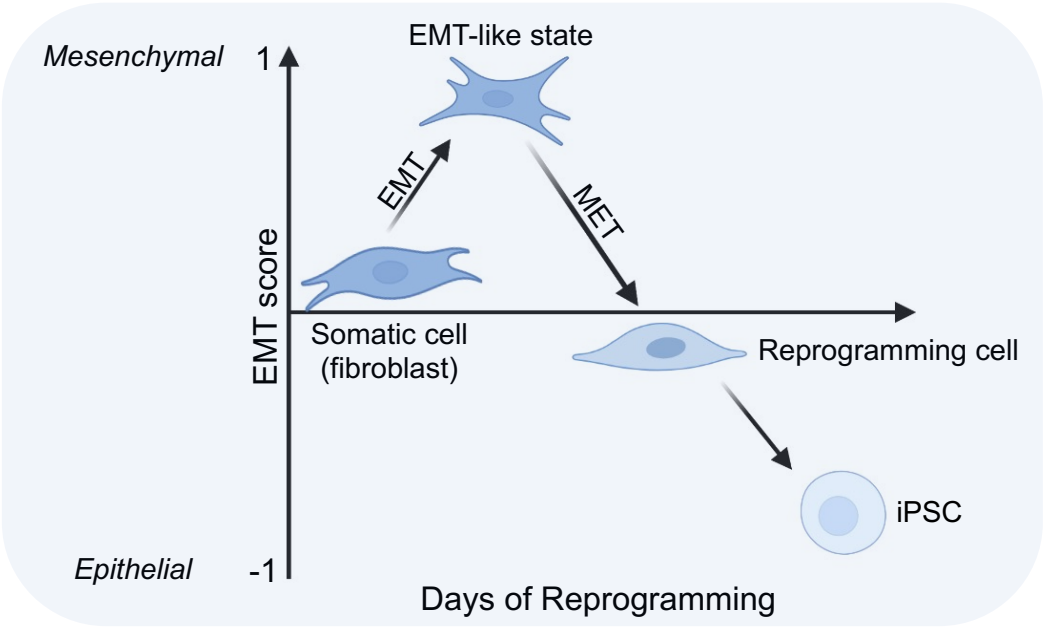


Figure 7. Sequential EMT-MET is required during cellular reprogramming. Schematic outline of reprogramming progression while cells move in the EMT spectrum (adapted from (Pei et al., 2019). The first step induces a mesenchymal phenotype, whereas stemness is acquired during the MET phase.

2.2.2.2. Molecular barriers to cellular reprogramming: cellular senescence

Cellular senescence is a cellular programme characterised by a permanent cell cycle arrest and a prominent secretory phenotype that is induced both *in vitro* and *in vivo* in response to a plethora of stress stimuli (Sapieha and Mallette, 2018). It is possible to distinguish two different types of senescence. On the one hand, replicative senescence refers to the specific programme activated upon the progressive shortening of the telomeres associated with cell division (Blasco et al., 1997; Harley et al., 1990). On the other hand, cells can undergo premature senescence in response to several cues. For instance, oncogene induced senescence (OIS) is a mechanism that blocks oncogenic transformation and tumour growth when oncogenes are aberrantly activated in the cell, but also other cues such as nutrient deprivation, genotoxic stress or irradiation can cause the premature entry of cells into this programme (Kumari and Jat, 2021).
To date, it is well known that essentially all cell types can undergo senescence in response to multiple stresses. For instance, also chemotherapy agents like doxorubicin are known to induce senescence (Piegari et al., 2013).

Regarding the molecular players of cellular senescence, upon damage (i.e. DNA damage) p53 is activated and stabilised, and induces cell cycle arrest via $p21^{CIP1}$, which is a CDK inhibitor (Cazzalini et al., 2010). Another pathway that controls cellular senescence is governed by the *INK4a* locus. Its two products, $p16^{INK4A}$ and $p19^{ARF}$ (or $p14^{ARF}$ in human) (Quelle et al., 1995) can be upregulated by damage and block the cell cycle through different mechanisms. While $p16^{INK4A}$ is a direct inhibitor of CDK4 and 6 (Serrano et al., 1993), $p19^{ARF}$ contributes to the stabilisation of p53 via inactivation of MDM-2 (Chin et al., 1998).

Importantly, cellular senescence cannot be defined by a unique marker that distinguish senescent cells from other non-senescent cells. Therefore, a combination of different biomarkers needs to be studied in order to confidently identify senescent cells. Senescent cells have high activity of lysosomal β-galactosidase due to the augmented lysosomal content (Lee et al., 2006). This activity is known as senescence-associated β-galactosidase (SA-βGal) (Dimri et al., 1995; Kurz et al., 2000; Lee et al., 2006) and it is the property most exploited as a differential marker between senescent and non-senescence cells. *In vitro*, senescence is accompanied by morphological changes; cells become flatter, larger and highly vacuolised. Permanent cell cycle arrest is a key characteristic of senescent cells; thus, they show high expression levels of the tumour suppressor genes $p16^{INK4A}$, $p15^{INK4B}$, $p21^{CIP1}$ and $p27^{KIP1}$, and the activation of p53 (Campisi and d'Adda di Fagagna, 2007; Lin et al., 1998; Munoz-Espin and Serrano, 2014; Serrano et al., 1997). Finally, the characteristic that perhaps differentiates the most senescent form quiescent cells is the acquisition of a prominent pro-inflammatory secretory phenotype. In fact, senescent cells secrete many chemokines, cytokines, proteases, ECM components and growth factors that are collectively known as SASP

(senescence-associated secretory phenotype). We will further discuss the role of the SASP in cellular reprogramming (see *Section 2.2.2.3*).

As mentioned before, cellular senescence acts as a barrier protecting from oncogenic transformation. Importantly, each one of the OSKM factors has been demonstrated to have oncogenic activity (Bass et al., 2009; Rodini et al., 2012; Saiki et al., 2009) and the induction of the OSKM generates genotoxic stress in the reprogramming cells (Gonzalez et al., 2013). Thus, in response to oncogene activation and genotoxic stress, tumour suppressor genes are highly upregulated at early phases of cellular reprogramming. Accordingly, the inactivation of *Ink4a/Arf* locus, Rb, p53 or p21 ^{CIP1} greatly increases the reprogramming efficiency through avoiding cellular senescence response (Banito et al., 2009; Hong et al., 2009; Kareta et al., 2015; Li et al., 2009; Marion et al., 2009; Utikal et al., 2009) .

In summary, the activation of tumour suppressor genes and the induction of cellular senescence act as a barrier for cellular reprogramming in a cell intrinsic manner, inducing cell cycle arrest of cells expressing OSKM.

2.2.2.3. Cellular senescence: a double-edged sword in cellular reprogramming

As explained above, one of the hallmarks of cellular senescence is the SASP. This is a characteristic secretory phenotype associated with senescence that could have autocrine and paracrine effects. Of note, although the panel of secreted molecules could be different depending on the cell type and/or the senescent inducer (Hernandez-Segura et al., 2017), SASP is always composed by the same categories of factors: soluble signalling molecules, secreted proteases (PAI-1 and PAI-2), ECM components and pro-inflammatory signalling molecules such as interleukins, chemokines and growth factors. Among the pro-inflammatory molecules, the key components are interleukin 6 (IL-6), interleukin 8 (IL-8), interleukin 1α and β (IL1-α and IL1-β), monocyte chemoattractant proteins (like CCL2, CCL5 or CCL20) and TGFβ (Coppe et al., 2010). Importantly, these cytokines create an inflammatory microenvironment which recruits cells from the immune system (Di Mitri et al., 2014; Eggert et al., 2016; Kang et al., 2011; Xue et al., 2007) and, moreover, can induce senescence in neighbouring cells acting in a paracrine manner (Acosta et al., 2013; Hubackova et al., 2012; Nelson et al., 2012).

Most of the SASP components are regulated by several signalling pathways. The most important regulator is probably nuclear factor-κB (NF-κB) (Chien et al., 2011). However, other factors such as C/EBP-β (Flanagan et al., 2018; Kuilman et al., 2008), JAK-STAT (Xu et al., 2015), p38/MAPK (Freund et al., 2011), mTOR (Herranz et al., 2015; Laberge et al., 2015) and cGAS-STING pathways (Yang et al., 2017) have been described to control part of the SASP or converge, in the end, to the NF-κB signalling. Remarkably, while NF-κB suppression has a limited impact on the cell cycle arrest establishment in senescent cells, it strongly impairs the SASP (Chien et al., 2011), thus corroborating its role as a master regulator of the SASP phenotype in senescent cells.

While the induction of senescence acts as a barrier for cellular reprogramming in a cell-autonomous manner, the SASP has been demonstrated to promote de-differentiation and reprogramming in a non-cell autonomous manner both *in vitro* and *in vivo*. It has been shown that reprogramming always happens in close proximity to senescent cells. The SASP, and in particular IL-6, were found to induce *in vitro* and *in vivo* reprogramming in a cell extrinsic manner (Mosteiro et al., 2016; Mosteiro et al., 2018). In particular, the OSKM expression induces wide tissue damage that, in turn, activates a senescence response. While senescent cells themselves cannot reprogramme, the induction of the SASP promotes reprogramming in the neighbour cells in a cell extrinsic manner. In agreement with these data, tissue damage-associated senescence triggered by the chemotherapy agent bleomycin is sufficient to induce *in vivo* reprogramming in the lung, an organ normally refractory to *in vivo* reprogramming. Importantly, while the deletion of the *Ink4a/Arf* locus has been described to increase reprogramming *in vitro* (Li et al., 2009), it largely abolishes *in vivo* reprogramming due to the lack of senescence response. In line with these results, the accumulation of senescence during accelerated or physiological aging has been demonstrated to enhance *in vivo* reprogramming (Mosteiro et al., 2016) also in organs normally refractory to reprogramming like skeletal muscle in a cell extrinsic manner (Chiche et al., 2017).

Finally, even though the SASP has been generally reported to have a positive effect on reprogramming, differences in its composition may lead to distinct outcomes in cellular reprogramming. For example, Mahmoudi and collaborators showed that, while some component of the SASP, such as IL-6, are beneficial for reprogramming, others might be detrimental. In particular, they demonstrated that high level of TNFα and IL-1β in the SASP decreases reprogramming efficiency in a non-cell autonomous manner. Interestingly, these two cytokines correlate with an increased profibrotic activity and wound healing (Mahmoudi et al., 2019). In the same line, TGFβ is also a SASP component (Coppe et al., 2010) and similarly to TNFα and IL-1β has a detrimental effect on reprogramming (Liu et al., 2013; Maherali and Hochedlinger, 2009) and displays profibrotic activity (Frangogiannis, 2020).

Concluding, while cellular senescence is a cell intrinsic barrier for reprogramming, it promotes cellular reprogramming in a cell extrinsic manner via the SASP. (Figure 8). Of note, the cell extrinsic component seems to be particularly important in the *in vivo* context. Moreover, the heterogeneity of the SASP could lead to different reprogramming outcomes, either improving or impairing it, depending on its main cytokine composition.

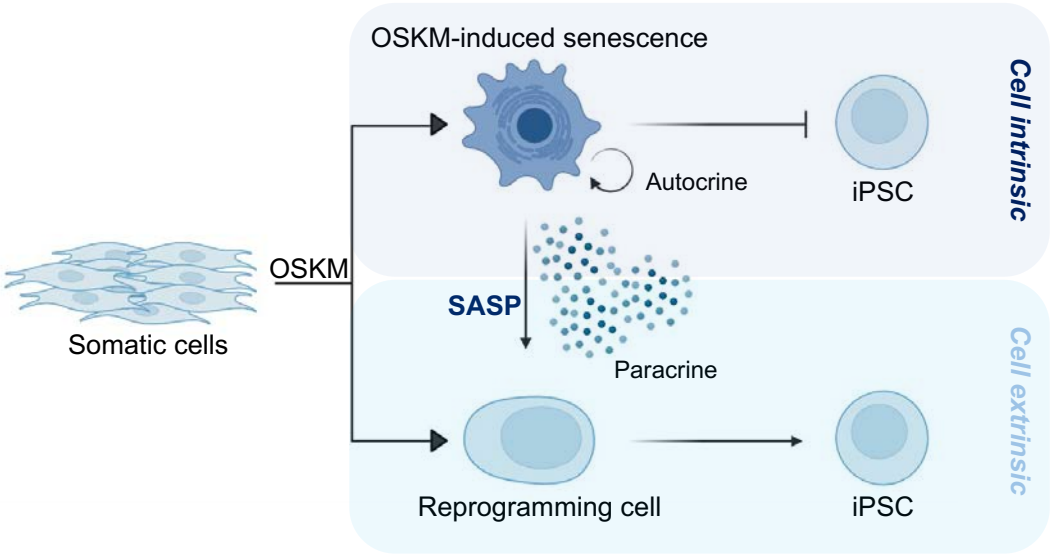


Figure 8. The double role of cellular senescence in reprogramming. Scheme showing the role of cellular senescence as a reprogramming barrier in a cell-autonomous manner and its role as an inducer in a non-cell autonomous manner. While the SASP could act in an autocrine manner reinforcing the senescence phenotype in neighbouring cells, it helps the de-differentiation process in cells undergoing reprogramming.

3. Connections between Damage, Cellular Plasticity and Cancer

3.1. Damage-induced dedifferentiation

As described above, there is a direct link between damage and the induction of cellular dedifferentiation. Importantly, the axis damage-dedifferentiation is not restricted to the artificial context of cellular reprogramming, but it has been shown to be an evolutionary conserved mechanism that operate in several tissues upon injury to promote tissue repair (Yao and Wang, 2020). For example, in *Notophtalmus viridescens* salamander, after limb amputation cells dedifferentiate and form a pool of proliferating cells, known as blastema, that fully regenerates the amputated limb (Iten and Bryant, 1973). In the same line, in Zebrafish, upon heart injury, cardiomyocytes dedifferentiate and start proliferating again to regrow cardiac muscle tissue (Jopling et al., 2010; Kikuchi et al., 2010). In mammals, the process of dedifferentiation-driven regeneration can happen as well, but to a limited extent. In mouse it has been observed that, upon damage, certain secretory progenitor cells of the intestine can revert to stem cells in order to repair the tissue (van Es et al., 2012). Additionally, if the stem cell compartment of the intestine is genetically ablated, terminally differentiated enterocytes can dedifferentiate and repopulate the stem cell compartment (Tetteh et al., 2016). Dedifferentiation also happens in other tissues such as the lung where, similarly to the intestine, epithelial cells can dedifferentiate and reconstitute the stem cell pool upon its ablation *in vivo* (Tata et al., 2013). In line, damage-induced dedifferentiation and subsequently regeneration has

also been observed in the liver, pancreas and nervous system (Kopp et al., 2016; Lin et al., 2017; Painter et al., 2014; Sanges et al., 2013).
Importantly, there are growing evidence showing that senescence can play a key role in the induction of dedifferentiation during regeneration in a cell extrinsic manner. For example, oncogene-induced senescence has been reported to promote the expression of stem cell markers and enhance the regenerative capacity of keratinocyte through the SASP (Ritschka et al., 2017).
Altogether, these evidence reveal a link between damage and the acquisition of dedifferentiation (or plasticity) in order to promote regeneration. Importantly, as described in the case of reprogramming, cellular senescence may have a role by providing an environment that promotes dedifferentiation and favours tissue regeneration.

3.1.1. Damage-induced dedifferentiation in cancer

In the same way damage induces dedifferentiation in order to trigger tissue regeneration, cancer cells hijack this process to acquire a more dedifferentiated phenotype and thus, a selective advantage. In fact, several reports show that cancer cells dedifferentiate upon chemotherapy treatment, which promotes therapy resistance and tumour relapse (Filipponi et al., 2019; Raghavan et al., 2021; Xiong et al., 2019).
Importantly, some evidences show that damage induced dedifferentiation can also lead to tumour initiation, as it is the case of gliomas (Friedmann-Morvinski et al., 2012) or pancreatic adenocarcinoma (Guerra et al., 2007). Consistently, dedifferentiation induced by transient activation of the OSKM reprogramming factors *in vivo* leads to tumour development Abad et al., 2013,(Ohnishi et al., 2014).
Altogether these data suggest that while the acquisition of a certain degree of dedifferentiation could lead to regeneration, if the damage chronically persists the process could aberrantly lead to cancer development.
Of note, although cellular senescence is considered one of the most important cancer protecting barriers (Collado et al., 2005), the SASP has been shown to induce the acquisition of a dedifferentiated phenotype and stem-like properties in cancer cells (Cahu et al., 2012; Milanovic et al., 2018; Parrinello et al., 2005), thus promoting cancer progression and resistance to therapy (Dou and Berger, 2018).
Concluding, we can think about cancer as the result of an aberrant response to damage, where the acquisition of a dedifferentiated phenotype, and therefore plasticity, leads to cancer progression instead of regeneration. In fact, due to its importance in all the phases of neoplastic disease, the acquisition of plasticity has been recently proposed as an emergent hallmark of cancer (Hanahan, 2022).

3.1.2. Parallelism between neoplastic transformation and reprogramming

With the advances in the reprogramming field, it has become evident that cellular reprogramming and cancer can be seen as parallel processes regulated by common molecular mechanisms. Firstly, even though in some tumour types is not clear yet which is the cell of origin, a great number of evidences points to a somatic, differentiated cell that acquires unlimited proliferation as cell of origin (Friedmann-Morvinski and Verma,

2014), like in the case of cellular reprogramming. Second, both processes require the expression of oncogenes; as a matter of fact, the four Yamanaka factors have been demonstrated to be *bona fide* oncogenes (Bass et al., 2009; Iglesias et al., 2017; Rodini et al., 2012; Saiki et al., 2009). Additionally, both transforming cells and reprogramming cells need to overcome the tumour suppressor barriers. Tumour suppressor genes are upregulated at the early phases of cellular reprogramming and represent one of the most important cell intrinsic barrier to the process, by triggering a wide array of cellular responses such as cell-cycle arrest, apoptosis and senescence (Hong et al., 2009; Kareta et al., 2015; Li et al., 2009; Marion et al., 2009). In the same line, in order for oncogenic transformation to occur, the activation of one oncogene is not sufficient to trigger transformation but the inactivation of tumour suppressor genes is needed (DeNicola and Tuveson, 2009; Land et al., 1983).
Finally, the acquisition of unlimited proliferation capacity during both oncogenic transformation and reprogramming requires a metabolic reprogramming. In particular, reprogramming cells switch from an OXPHOS dependent status to a more glycolytic phenotype, which suits better their energetic needs (Liu et al., 2019). In line, cancer cells also shift to a more glycolytic metabolism, a phenomenon that has widely studied as the Warburg effect (DeBerardinis and Chandel, 2020).

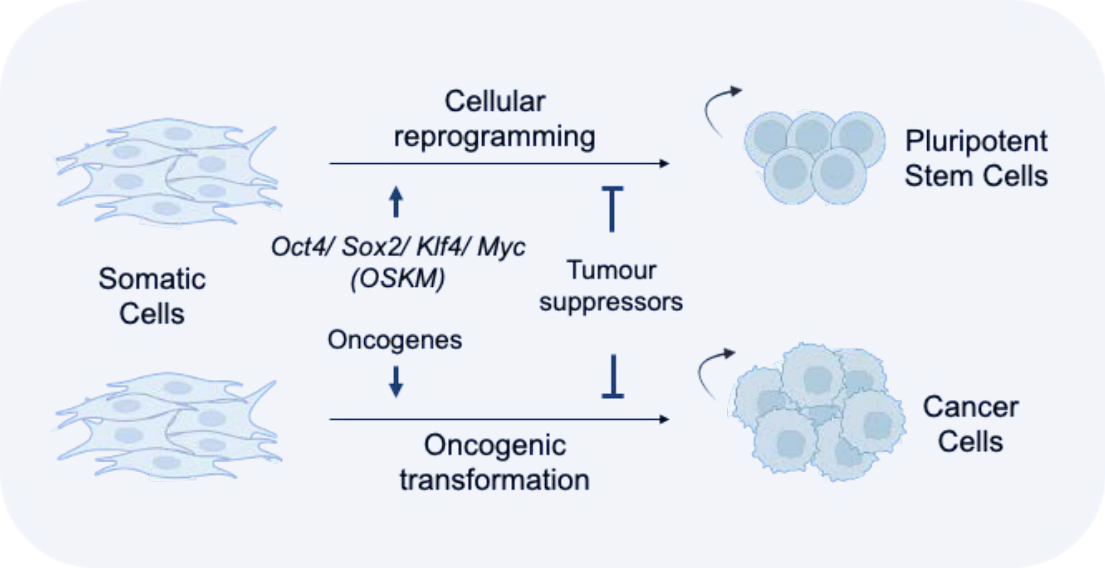


Figure 9. Cell intrinsic mechanisms shared between cellular reprogramming and cancer. Cellular reprogramming as well as neoplastic transformation starts from a population of somatic cells with limited proliferation capacity and end up with an immortal population of cells that self-renew. Additionally, both processes are triggered by oncogenic signals and are blocked by the activation of tumour suppressor genes that protect cellular identity.

3.1.3. EMT-MET: a transition needed in cancer metastasis and cellular reprogramming

Finally, the transition between mesenchymal and epithelial phenotypes is a common cell plasticity feature in both cellular reprogramming and cancer metastasis.
As we discussed above, an EMT has been shown to be beneficial at the beginning of the reprogramming process and is required to start the metastatic cascade (Liu et al.,

2013; Ye and Weinberg, 2015). However, the retention of a mesenchymal phenotype has been demonstrated to be detrimental in both cases. On the one hand, reprogramming cells that do not undergo MET fail to reach the pluripotency state (Li et al., 2010). Indeed, the loss of E-Cadherin in reprogramming cells drastically decreases reprogramming efficiency, further corroborating the importance of MET during the reprogramming process (Li et al., 2010; Redmer et al., 2011).
On the other hand, invasive cancer cells need to undergo a MET in order to colonise distal organs in the metastatic process. As an example, it has been reported that the ablation of the mesenchymal factor PRRX1 in cells that have undergone EMT allows lung colonisation and metastatic outgrowth (Ocana et al., 2012). In agreement, E-Cadherin has been recently shown to be essential for the metastatic process in several model of breast cancer (Padmanaban et al., 2019)
Altogether these data show a requirement for a similar EMT-MET dynamics both in cellular reprogramming and during the cancer metastatic cascade.

In summary, in our view, both cancer and cellular reprogramming are processes that exploit common molecular mechanism for the acquisition of cellular plasticity. Therefore, finding novel regulators of cellular plasticity may open new avenues for new therapeutic interventions both in cancer as well as in regenerative medicine.

HYPOTHESIS & OBJECTIVES

The microproteome is an unexplored field that has opened a new layer of biological complexity and may change current paradigms in basic and clinical research. Our hypothesis is that the microproteome could be a source of novel regulators of cellular plasticity with key functions in cancer and tissue regeneration and, therefore, with clinical potential.

To demonstrate this hyphothesis, we proposed the following specific objectives:

1. Identify novel microproteins related with cancer cell plasticity.
2. Generate gain- and loss-of-function tools to study the most interesting candidate.
3. Characterise the molecular functions of the microprotein, including its subcellular localisation and the regulation of its expression.
4. Study the role of the microprotein in cancer and in cellular plasticity assays, such as reprogramming.
5. Analyse the molecular mechanisms responsible for the microprotein's functions.

MATERIALS & METHODS

1. Mouse experiments

Mice were housed at the Specific Pathogen-Free (SPF) barrier area of the Vall d'Hebron Institute of Oncology (VHIO). All animal procedures were approved by the Animal Care unit and the Ethics Committee for Animal Experimentation (CEEA) of the Vall d'Hebron Research Institute (VHIR), as well as by the Catalan Government, and were performed according to the European legal framework for research animal use and bioethics. Animals were monitored daily and euthanised upon signs of humane endpoints.

1.1. Mouse models

1.1.1. MMTV-PyMT mouse model

The MMTV-PyMT animals were nicely provided by Dr. Laura Soucek (Vall d'Hebron Institute of Oncology). The MMTV-PyVT transgene (MMTV-PyMT) includes the mouse mammary tumour virus (MMTV) long terminal repeat upstream of a cDNA sequence encoding the Polyoma Virus middle T antigen (PyVT). Transgenic mice are viable, but show loss of lactational ability coincident with transgene expression. Adenocarcinomas arise in virgin and breeder females as well as males. Mice were crosses with the MIDORI KO mouse strain.

1.1.2. Reprogrammable mouse (i4F) model

The reprogrammable mouse line known as i4F-B generated by Dr. María Abad (Abad et al., 2013) was used to obtain i4F Mouse Embryonic Fibroblasts (MEFs) and to cross it with MIDORI KO mouse strain. This murine model carries a ubiquitous doxycycline-inducible OSKM transgene, abbreviated as i4F, inserted into the Pparg gene, and the rtTA transcriptional activator under the control of the Rosa26 promoter.

1.1.3. MIDORI-deficient mouse model

This animal model was generated in collaboration with Dr. Sagrario Ortega (Head of the Transgenic Mice Unit at CNIO, Madrid). It is a CRISPR-knock-in model, that was developed using a homologous recombination template that includes the mutation of the MIDORI start codon to a stop codon. With this strategy we aim to specifically impair MIDORI translation without affecting the function of the *Zeb2os* lncRNA. Both, a plasmid coding the Cas9 and the gRNA and the homologous recombination template were delivered by microinjection in mouse embryos. The homologous template was designed to include a BclI restriction site, so the modified allele (KI allele) could be detected by enzymatic digestion. Briefly, DNA was extracted from the tail and *Zeb2os* locus was amplified by PCR using Zeb2os_locus amplification pair of primers (Table 2). Once amplified, the PCR product was incubated with BclI restriction enzyme (New England Biolabs) overnight at 37ºC and run in agarose gel to test digestion. Separately, the PCR product was treated with ExoSAP-IT™ PCR Product Cleanup Reagent (Thermo Fisher Scientific) following manufacturer instructions and sequenced by Sanger Sequencing to

confirm the modification. The successfully modified animals were used as founders to generate MIDORI deficient colony (MIDORI KO).

1.1.4. Mice genotyping

DNA was extracted from ear-punch tissue fragments and genotyped by PCR using specific primer pairs (Table 2). For MIDORI-KO colony, specific forward primers were designed to detect wildtype and knock-in alleles (SALgenmismatch1WT_FW and SALgenmismatch1_FW) together with a common reverse primer (SALgenWT_RW). Reprogrammable (i4F) mice were genotyped using OSKM (ORF2 primers) and rtTA (125B, 126B and 127B) pair of primers, while MMTV-PyMT mice were genotyped using a specific set of primers pair (MMTV-PyMT-F and MMTV-PyMT-R).

1.2. *In vivo* experiments

1.2.1. i4F; MIDORI-KO mouse model and induction of *in vivo* reprogramming

The two mouse models described above were crossed to obtain i4F; MIDORI-KO mice. To perform *in vivo* reprogramming experiments, 1 mg/ml or 0,2 mg/ml of doxycycline (Sigma Aldrich) was added to the drinking water supplemented with 7.5% of sucrose either for 1 week or 3 weeks, respectively. Doxycycline was refreshed every two days. Reprogramming experiments were performed with 10-12-weeks old mice of both sexes.

1.2.2. MMTV-PyMT; MIDORI-KO mouse model and the study of breast cancer tumour and metastasis

The two mouse models described above were crossed to obtain MMTV-PyMT; MIDORI-KO mice. Given that the mice show a mixed background, breast tumours start to appear around 12-14 weeks of age in female mice. Tumour size was evaluated weekly by calliper measurements and tumour volume calculated using the following formula: volume = (length x width x width) x 1/6 π. Once the sum of the tumours reached 1200 mm^3, mice were euthanised by CO^2 asphyxiation and breast tumours as well as lungs were processed for further histological analyses.

1.2.3. Folic acid-induced acute kidney injury

Male mice (8-10 weeks old) were intraperitoneally injected with a single dose of vehicle (300 mM NaHCO3) or folic acid (250 mg/Kg). At the end of the experiment, 30 days after the injection, mice were euthanised by CO^2 asphyxiation and the kidneys were collected for molecular biology and histological analyses.

1.2.4. *In vivo* wound healing

Mice of both sexes (4 weeks old) were used in the experiments before the hair growing started to be asynchronized. Mice were anesthetised by Isoflurane inhalation (5% for induction and 2% for the maintenance during the surgery) and shaved in the back. A 6-mm diameter biopsy punch was used to create a circular wound in the shaved area. The closure of the wound was monitored every day and measured by calliper. Wounded area was calculated using the following formula: area = (length/2 x width/2) x π.

1.2.5. Model of lung colonisation by tail vein injection

Immune-compromised NMRI nu/nu mice (6-8 weeks old females by Janvier) were injected with 100.000 MDA-MB-231cells constitutively expressing MIDORI or the empty vector as a control. 100 µl of cells in PBS were injected into the dorsal tail vein. The injection was followed by IVIS imaging. The progression of lung colonisation was monitored every week by IVIS imaging until week 10 post injection. When the majority of control mice showed lung tumours (confirmed by at least three repetitive detections of lung signal by IVIS), we ended the experiment. Mice were euthanised by CO^2 asphyxiation and the lungs were harvested and used for further histological analyses. In the case of the survival analysis, mice were euthanised when reaching the human end point.

1.2.6. Model of breast cancer with surgical resection and metastatic spread

Immune-compromised NMRI nu/nu mice (6-8 weeks old females by Janvier) were injected with a suspension 1:1 of Matrigel mixed with 1.5 million of MDA-MB-231cells constitutively expressing MIDORI or the empty vector as a control in PBS. 50 µl of the obtained cell suspension was injected between the fourth and the fifth right mammary fat pad. Before surgery, mice were anesthetised by Isoflurane inhalation (5% for induction and 2% for the maintenance during the surgery). Air flow was set at 0.8 L/min. Tumour size was evaluated weekly by calliper measurements and tumour volume calculated using the following formula: volume = (length x width x width) x 1/6 π. When tumours reached around 400-500 mm^3, they were surgically resected. The evolution of metastases was followed weekly by IVIS imaging. At the endpoint of the experiment, when all control mice developed metastases (confirmed by at least three repetitive detections of lung signal by IVIS), mice were euthanised by CO^2 asphyxiation and *ex vivo* IVIS scan was performed. Organs that were positive for luciferase activity were collected and used for further histological analyses.

1.2.7. Orthotopic model of cutaneous skin squamous-cell carcinoma

Immune-compromised NOD/SCID gamma mice (6-8 weeks old females by Janvier) were injected with a cell suspension containing 1.5 million of cSSC11 cutaneous skin squamous cell carcinoma cells (Bernat-Peguera et al., 2021; Bernat-Peguera et al., 2019) expressing a vector for the doxycycline-inducible expression of MIDORI or the empty vector as a control in PBS. 100 µl of the cell suspension was injected in the right

and left flank of each mouse. Before the injection, mice were anesthetised by Isoflurane inhalation (5% for induction and 2% for the maintenance during the surgery). Air flow was set at 0.8 L/min. Doxycycline was administered in the drinking water at 1 mg/ml supplemented with 7.5% of sucrose every two days throughout all the experiment. Tumour size was evaluated weekly by calliper measurements and tumour volume calculated using the following formula: volume = (length x width x width) x 1/6 π. The evolution of the tumours and eventual metastases was followed weekly by IVIS imaging. At the endpoint of the experiment, when all the mice developed tumours around 1200 mm^3, mice were euthanised by CO^2 asphyxiation and tumours were collected and used for further histological and molecular biology analyses.

1.2.8. IVIS and μCT imaging

IVIS studies were performed with a Xenogen IVIS Spectrum (Perkin Elmer) and micro-computed tomography (μCT) studies were performed with a Quantum GX microCT Imaging System (Perkin Elmer). Acquisition and analysis were carried out by the Preclinical Imaging Platform staff at VHIR.
For IVIS imaging, mice were injected intraperitoneally with a D-luciferin solution (150 mg/Kg in PBS) 5-10 minutes prior to acquisition. For both IVIS and μCT, mice were anesthetised with isoflurane (5% for induction and 2% during acquisition). Air flow was set at 0.8 L/min. IVIS data were analysed with Living Image® software (Perkin Elmer). Study analysis consists in a light radiance quantification. Signals from the light sources were detected and characterised. Working units were photons/sec/cm^2/sr, which allow the comparison between images obtained by different acquisition parameters.
μCT reconstructions were performed with the Quantum GX microCT software and ADIME software was used for the analysis (http://amide.sourceforge.net).

1.2.9. Lung tumours quantification

At the endpoint of the experiments, lungs were perfused with 4% buffered formaldehyde and washed with PBS. Lungs were then fixed for 24 hours in buffered 4% formaldehyde, transferred to 70% ethanol and embedded in paraffin. The whole lung was sectioned and serial sections in four different planes were stained with H&E. H&E staining was performed on 3 μm paraffin sections in a Robust carousel tissue stainer (Slee Medical) according to common method. Metastatic foci were counted and their area measured with ImageJ.

1.2.10. Immunohistochemistry

Immunohistochemistry was performed on paraffin-embedded mouse tissues. For 21 staining, paraffin blocks were sliced into 3 μm sections, deparaffinised with xylene (Fisher Scientific, Waltham, MA, United States) and rehydrated with decreasing concentrations of ethanol in water. Heat-mediated antigen retrieval was done in citrate buffer containing 0,05% Tween at pH 6.0 and endogenous peroxidase was blocked. Slides were first incubated with p21 primary antibody (Table 1) overnight at 4ºC and then

with the secondary antibody anti-goat (Abcam Ab6697) (1:200 in goat serum) conjugated with horseradish peroxidase for 1h at room temperature. Immunohistochemical reaction was developed using 3,30-diaminobenzidine tetrahydrochloride (DAB) as a chromogen and nuclei were counterstained with haematoxylin. Lastly, the slides were dehydrated, cleared, and mounted using DPX as mounting medium for microscopic evaluation.
For Ki67 and MIDORI staining, paraffine sections underwent antigenic exposure process into the Discovery Ultra (Ventana) system with CC1 buffer for 64 minutes at 95ºC. Anti-MIDORI antibody was incubated for 1 hour at RT (Table 1). Next, slides were incubated with the secondary antibody Discovery UltraMap anti-Rabbit HRP (Ventana).

1.2.11. Sirius red staining

To evaluate collagen deposition, kidney tissues were fixed 24 hours in buffered 4% formaldehyde, transferred to 70% ethanol and embedded in paraffin. Paraffin blocks were sliced into 3 µm sections, deparaffinised with xylene, rehydrated through graded series of ethanol: Once hydrated, sections were stained with Sirius red (saturated aqueous solution of picric acid containing 0.2% Direct Red 80) (Sigma-Aldrich) and counterstained with 0,2% Fast Green (Sigma-Aldrich) in 1% acetic acid.

1.2.12. Sirius red, H&E and immunohistochemistry quantification

To quantify the collagen deposition in the kidney, the percentage of dysplastic area in the pancreas and the percentage of KI67 and p21 positive cells in tissues, slides were scanned using a Nano Zoomer pathology scanner. The areas of collagen deposition and the number of positive cells in the immunohistochemistry were measured using Qu Path software.

2. General cell culture and treatments

2.1. Cell culture conditions

Human Embryonic Kidney (HEK) 293T, primary Mouse Embryonic Fibroblasts (MEFs), IMR90 human fibroblasts and MDA-MB-231 triple negative breast cancer cells were cultured in DMEM with GlutaMAX supplemented with 10% of Foetal Bovine Serum (FBS) and 1% of Penicillin-Streptomycin (P/S) (Gibco). NMuMG breast epithelial cells were cultured with the same condition but supplementing the medium with human Insulin (Sigma 11061-68-0) diluted 1:1000. BxPC-3 human pancreatic cancer cells were cultured in RPMI supplemented with 10% of Foetal Bovine Serum (FBS) and 1% of Penicillin-Streptomycin (P/S) (Gibco). Mouse PDAC and Cancer associated fibroblasts (CAFs) were kindly provided by Dr. Laura Soucek (Vall d'Hebron Institute of Oncology) and were cultured in DMEM-F12 with GlutaMAX supplemented with 10% of Foetal Bovine Serum (FBS) and 1% of Penicillin-Streptomycin (P/S) (Gibco). MMTV-PyMT; WT and MIDORI KO were cultured in cultured in DMEM-F12 with GlutaMAX supplemented with 10% of Foetal Bovine Serum (FBS), 1% of Penicillin-Streptomycin (P/S) and ITS

Liquid Media Supplement (100×) (Gibco). Finally, during reprogramming experiments, reprogramming cells were cultured in DMEM GlutaMax supplemented with 15% KO serum replacement (Gibco), 50 mM β-mercaptoethanol, 1% NEAA (Invitrogen), 1% P/S and 1000 U/ml LIF (ESGRO, Chemicon).
Cells were maintained in a humidified incubator at 37ºC with 5% CO2. Cultures were routinely tested for mycoplasma and were always negative.

2.2. Isolation of primary Mouse Embryonic Fibroblasts (MEFs)

Mouse embryos were extracted at 13.5 d.p.c. and suspended in PBS. The head and foetal liver were removed and then, embryos were chopped and incubated in 0.2% Trypsin for 20 minutes at 37ºC in a cell culture incubator, pipetted up and down 15 times, and incubated for 20 more minutes at 37ºC. Trypsin was inactivated adding DMEM GlutaMAX supplemented with 10% FBS and 1% P/S and cells were cultured in normal conditions or were frozen after one passage.

2.3. Isolation of primary MMTV-PyMT breast cancer cell lines

Tumours were extracted from MMTV-PYMT; WT and MIDORI KO mice when they reached 1200 mm^3 and susopended in PBS. Tumours were chopped and incubated with the digestion medium (DMEM-F12 with GlutaMAX supplemented with 2,5% of Foetal Bovine Serum (FBS), 1% of Penicillin-Streptomycin (P/S), Collagenase II (5 mg/ml) and Dispase (1.25 mg/ml) (Gibco) for 2 hours at 37ºC in a cell culture shaker. Then, the digestion was inactivated by adding DMEM-F12 with GlutaMAX supplemented with 10% of Foetal Bovine Serum (FBS) and 1% of Penicillin-Streptomycin (P/S) to the reaction. Cells were spun down for 5 minutes at 300 g and finally resuspended and cultured in DMEM-F12 with GlutaMAX supplemented with 10% of Foetal Bovine Serum (FBS), 1% of Penicillin-Streptomycin (P/S) and ITS Liquid Media Supplement (100×) (Gibco).

2.4. Treatments

To induce the expression of Tet-ON inducible systems, cells were treated when indicated with Doxycycline (Sigma Aldrich) at 1 µg/ml. For EMT induction and NF-κB Western blot, cells were treated with human recombinant TGFβ1 and human recombinant TNFα (Peprotech) at a final concentration of 5 ng/ml in the cell culture medium. Doxorubicin and Tunicamycin (Selleckchem) were used at a final concentration of 1µM while control cells were treated with the vehicle, DMSO, as a control. For senescence experiments, Palbociclib was added to the culture medium at a final concentration of 2µM while the vehicle, DMSO, served as control.

3. Cloning procedures

MIDORI mouse (long and short) and human ORFs were synthesised (IDT technologies) fused with a flexible linker (3xGGGGS) and an HA tag epitope at the C-terminal part of the microprotein and flanked by EcoRI and XhoI enzyme restrictions sites at the two ends (Complete Sequences can be found in Table 4) After enzymatic digestion, constructs were ligated into the pMSCV retroviral vector and pENTR1A vector (Addgene). For the lentiviral vectors, the MIDORI-HA tag construct was obtained by recombining the donor pENTR1A-MIDORI-HA vector with the lentiviral inducible system pINDUCER20 (Invitrogen) using the Gateway Cloning Technology LR clonase (Invitrogen), following manufacturer's instructions.

4. Viral infections

HEK293T cells were used as packaging cells and transfected with the lentiviral and packaging plasmids indicated in Table 3 using Polyethylenimine (PEI) (Polyscience Europe GmbH) reagent following the manufacturer instructions. Viral supernatants were collected twice a day on two consecutive days starting 36 hours after transfection, filtered through a 0.45 µm syringe filter, supplemented with 8 µg/ml of polybrene and used to infect the previously plated cells of interest. Successfully infected cells were established by geneticin (Gibco) selection at 400-600µg/ml (pINDUCER-20) or puromycin (VWR) selection 1-2 µg/ml (pMSCV).

5. Protein analysis by Western blotting

Cells were homogenized in 2% SDS lysis buffer (50 mM Tris-HCl, pH 8.0, 1mM EDTA, 2% SDS) or medium-salt lysis buffer (150 mM NaCl, 50 mM Tris pH 8, 1% NP40) supplemented with protease (Roche) and phosphatase (Sigma Aldrich) inhibitors cocktails. Protein concentration was determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher). A total protein amount of 50 µg were loaded in 12% Bis-Tris acrylamide gels and transferred to nitrocellulose membranes. Primary antibodies were incubated overnight at 4°C. Secondary HRP-conjugated antibodies were incubated the following day for 1h at room temperature, and ECL Prime Western Blotting Detection Reagent (Fisher Scientific) or SuperSignal™ West Dura Extended Duration Substrate were used as a chemiluminescent reagent for protein detection. Antibodies and dilutions are listed in Table 1.

5.1. Subcellular Fractionation

Cells were homogenized in Buffer A (HEPES pH 7.8 (10mM), $MgCl_2$ (1.5 mM), KCl (10 mM) and DTT). The homogenate was incubated on ice for ten minutes and then, 10% Triton-X was added to favour cellular disruption. Samples were centrifuged at 11000 rpm for one minute at 4ºC and supernatant (Cytoplasmic extract, CE) was separated from the pellet (Nuclear extract, NE). CE supernatant was washed with Buffer B (HEPES pH 7.8 (0.3 mM), $MgCl_2$ (1.4 mM) and KCl (30 mM)), followed by centrifugation at 15000 g

for 15 minutes at 4ºC. The collected supernatant was used as final cytoplasmic extract and the pellet was washed with Buffer B and disrupted using Buffer C (HEPES pH 7.8 (20 mM), $MgCl_2$ (1.5 mM), NaCl (0.42 mM), EDTA (0.2 mM), glycerol (25%) and DTT). After thirty minutes of ice-incubation and centrifugation at 15000 g for fifteen minutes at 4ºC, supernatant was used as final nuclear extract. Finally, the remaining pellet was disrupted in SDS lysis buffer (Tris-HCl pH 8.0 (50 mM), EDTA (1 mM), SDS 2%) and sonicated to obtain the chromatin bound extract.

6. mRNA analysis by quantitative real-time PCR (qRT-PCR)

Total RNA was extracted with Trizol Reagent (Invitrogen), chloroform and isopropanol following the manufacturer protocol. 1 µg of total RNA was retrotranscribed into cDNA using the iScriptTM Reverse Transcription Supermix Kit (BioRad). Gene expression was analysed by qRT-PCR using PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) in the 7900HT Fast Real-Time PCR System (Applied Biosystems). Gene-specific primers are listed in Table 2.

7. Immunofluorescence

Cells were seeded in fibronectin-coated (Sigma Aldrich) or collagen I-coated (Corning) coverslips. When cells reached the desired confluency, cells were fixed in 4% paraformaldehyde for fifteen minutes and permeabilized with 0.2% Triton X-100 in PBS for ten minutes at room temperature. Blocking step was made in 3% Bovine Serum Albumin (BSA) for one hour. Cells were incubated overnight at 4°C with the primary antibody diluted in the blocking buffer. The day after, secondary antibodies were incubated for one hour at room temperature in the dark. Antibodies and dilutions are listed in Table 1. Finally, coverslips were mounted on the slides using Prolong Mounting Medium with DAPI (Invitrogen), and images were taken in Nikon C2 Plus Confocal Microscope. For quantification, at least 200 cells per staining were counted using ImageJ software.

In the case of the immunofluorescence with cytoskeletal buffer (CSK) extraction, cells were incubated with the CSK buffer (25mM HEPES pH7.4, 50mM NaCl, 3mM $MgCl_2$, 300mM sucrose and 0.5% Triton X-100) 10 minutes at room temperature prior to the fixation step.

8. Cell proliferation analysis

Cells were seeded into 24-well cell culture plates at density $5x10^3$ cells/well. Cells were trypsinised and counted using a Neubauer chamber every day. When using MIDORI-inducible cell lines, doxycycline was added every two days.

9. Cell migration assay

Cells were seeded in a 6-well cell culture plates and cultured until confluence was reached. A pipette tip was used to scratch the surface of the cell monolayer, forming a wound. An Olympus CellR microscope equipped with a Hamamatsu C9100 camera was used to follow the closure of the wound up to the indicated timepoints.
The wound closure was measured by ImageJ software.

10. Cell invasion assay

$2x10^4$ cells were seeded in the upper compartment of a Boyden chamber in a Corning cell culture insert coated with Matrigel (Corning). In the bottom compartment, culture medium supplemented with 10% FBS, conditioned medium from MDA-MB-231 cells or $2x10^5$ mCAFs were seeded as attractant for the invading cells. Cells were then incubated for 24 hours at 37°C in a cell culture incubator in order to invade through the Boyden chamber membrane. Then, culture inserts were fixed in methanol for 5 minutes and stained using Crystal Violet during 20 minutes at room-temperature. Pictures were taken using an Olympus CellR microscope. Cells were counted using ImageJ.

11. Matrix degradation assay

Gelatine from pig skin Oregon green 488 conjugate (Molecular probe life technologies) was used to coat glass coverslips. In brief, coverslips were coated with the fluorescent gelatine solution in the dark for 20 minutes. Coverslips were washed three times in PBS before cross-linking with a 0.5% solution of glutaraldehyde for 40 minutes followed by three washes in PBS. Before plating cells, coverslips were incubated with Collagen I (Corning) during 4 hours at 37°C in a cell culture incubator. After washing with normal culture medium, MDA-MB-231 cells were plated at a density of $2x10^4$ on top of the coverslips for 24 before fixation in 4% PFA. Gelatine degradation was measured by quantifying the mean area of nonfluorescent pixels per field using a manual threshold in ImageJ (National Institutes of Health). 10–15 random fields were imaged per condition, and each independent experiment was performed at least three times and averaged.

12. *In vitro* transcription/translation

LncRNA *Zeb2os* CDS was cloned in a pcDNA-3.1 (+) vector under the control of the T7 promoter and incubated for one hour with rabbit reticulocyte-coupled transcription/translation system (Promega) in the presence of [35S] methionine. After incubation, translated product was resolved by 15% SDS-polyacrylamide gel electrophoresis (PAGE) and detected by autoradiography.

13. Ribosome profiling analysis

We retrieved a public ribosome profiling dataset from mouse brain (ArrayExpress accession number E-MTAB-7247) (Wang et al., 2020) and we adapted a computational approach to identify translated sORFs (Ruiz-Orera et al., 2018). In brief, read adapters were trimmed and reads mapping to annotated ribosomal and transfer RNAs were filtered out. Resulting reads were mapped to the assembled mouse genome (mm10). Next, mapped reads from experimental replicates were merged and we used the riboORF algorithm (Ji et al., 2015) to predict translated sORFs with significant read uniformity and frame periodicity (score ≥0.7), as this feature is indicative of active ORF translation.

14. Cellular reprogramming

MEFs were transduced with Tet-O-FUW-OSKM inducible lentiviral vector together with rtTA vector. FUGW-GFP vector was used as transduction control. Transduction efficiency was assessed by Flow Cytometry using the FUGW-GFP infected cells. To study the effect of MIDORI on cellular reprogramming, MEFs were infected with the pINDUCER20 doxycycline-inducible lentiviral vector expressing either long mMIDORI or short mMIDORI together with the Tet-O-FUW-OSKM inducible lentiviral vector. MEFs transduced with the pINDUCER20 empty and the Tet-O-FUW-OSKM served as control. After infection, different MEFs cell lines derived from 3 embryos were plated in a 6-well plate at 10^5 cells/well. Once attached, OSKM cassette was induced by treating with doxycycline at 1 µg/ml in the cellular reprogramming medium, which was changed every other day for 12 days, until iPSC colonies were evident.
For reprogramming experiments using i4F MEFs, different MEFs cell lines derived from 3 embryos were directly seeded at $5 \cdot 10^5$ cells/well and induced with doxycycline at 1µg/ml in the cellular reprogramming medium. The medium was changed every other day for 12 days, until iPSC colonies were well defined. To determine cellular reprogramming efficiency in both cases, iPSC colonies were quantified manually after staining for alkaline phosphatase (AP) activity (Promega) following manufacturer instructions. In the case of the cellular reprogramming time course, cells were harvested at different timepoints and subsequently processed for protein and RNA extraction.

15. Generation of MEFs-conditioned medium (CM)

To produce conditioned medium (CM), 3.5×10^6 WT, MIDORI KO, MIDORI overexpressing and MIDORI KO overexpressing MEFs were seeded in 15-cm-diameter plates and cultured with reprogramming medium. After 3 days, the conditioned medium was collected and centrifugated for 10 minutes at 500 g. After the centrifugation, the conditioned medium was supplemented with doxycycline at 1 µg/ml and used to culture i4F MEFs in a reprogramming experiment, as previously described.

16. MEFs-secretome preparation

The CM, produced as described above, was spun down at 200 × g for 5min, the supernatants were collected, and filtered through a Millex-GP 0.22 µm pore syringe driven filter (Millipore, Ireland). Then secretomes were first concentrated using a 10000 MWCO Millipore Amicon Ultra (Millipore) at 4000 g at room temperature until a final volume of 250 µl, and then using a 10000 MWCO Microcon (Amicon) at 14000 g until a final volume of 50 µl. Protein concentration was determined with a Pierce BCA protein assay kit (Thermo Scientific).

17. Secretome analysis by mass spectrometry

All samples were in-solution digested previous to HPLC-MS analysis. Fifteen micrograms of each secretome preparation were first dissolved in 15 µl of 50% 2,2,2-Trifluoroethanol and reduced with tris (2-carboxyethyl) phosphine hydrochloride to a final concentration of 5 mm for 1 h at 60 °C and 700 rpm, and alkylated in 10 mm of Iodoacetic acid at 25 °C for 20min at 700 rpm in the dark. Before trypsin digestion, samples were diluted with 50 mm AB (ammonium bicarbonate) to a final concentration of 10% 2,2,2-Trifluoroethanol, and then proteins were digested in a ratio of 1:20 (w/w) with trypsin for 5 h at 37 °C. The reaction was stopped with formic acid (FA) to give a final concentration of 0.4% FA in the digested solution. After digestion samples were cleared at 10000 rpm for 10 minutes, dried, and re-dissolved in 30% acetonitrile, 0.1% FA to a final concentration of 1 µg/µl before liquid chromatography (LC)-MS analysis. Samples were analysed using an linear ion trap Velos-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Instrument control was performed using Xcalibur software package, version 2.1.0 (Thermo Fisher Scientific, Bremen, Germany). Peptide mixtures were fractionated by on-line nanoflow liquid chromatography using an EASY-nLC system (Proxeon Biosystems, Thermo Fisher Scientific) with a two-linear-column system. Digests were loaded onto a trapping guard column (EASY-column, 2 cm long, ID 100 µm and packed with Reprosil C18, 5 µm particle size from Proxeon, Thermo Fisher Scientific) at a maximum pressure of 160 Bar. Then, samples were eluted from the analytical column (EASY-column, 10 cm long, ID 75 µm and packed with Reprosil, 3 µm particle size from Proxeon, Thermo Fisher Scientific). Separation was achieved by using a mobile phase from 0.1% FA (Buffer A) and 100% acetonitrile with 0.1% FA (Buffer B) and applying a linear gradient from 5 to 35% of buffer B for 60 min at a flow rate of 300 nL/min. Ions were generated applying a voltage of 1.9 kV to a stainless-steel nano-bore emitter (Proxeon, Thermo Fisher Scientific), connected to the end of the analytical column.

The LTQ Orbitrap Velos mass spectrometer was operated in data-dependent mode. A scan cycle was initiated with a full-scan MS spectrum (from m/z 300 to 1600) acquired in the Orbitrap with a resolution of 30,000. The 20 most abundant ions were selected for collision-induced dissociation fragmentation in the linear ion trap when their intensity exceeded a minimum threshold of 1000 counts, excluding singly charged ions. Accumulation of ions for both MS and MS/MS scans was performed in the linear ion trap, and the AGC target values were set to 1×10^6 ions for survey MS and 5000 ions for MS/MS experiments. The maximum ion accumulation time was 500 and 200 ms in the

MS and MS/MS modes, respectively. The normalised collision energy was set to 35%, and one microscan was acquired per spectrum. Ions subjected to MS/MS with a relative mass window of 10 ppm were excluded from further sequencing for 20 s. For all precursor masses a window of 20 ppm and isolation width of 2 Da was defined. Orbitrap measurements were performed enabling the lock mass option (m/z 445.120024) for survey scans to improve mass accuracy. XCalibur 2.05 software (Thermo Fisher Scientific) was used to generate RAW files of each MS run. The .RAW files were processed using Proteome Discoverer 1.2 (Thermo Fisher Scientific). All cancer cell line secretomes MS/MS were searched against the human database Swiss-Prot 2010.11 (20332 sequences). In all cases the search engine used was MASCOT (Matrix Science, London, U.K.; version 2.2.04). The data was searched with fragment ion mass tolerance of 0.8 Da and a parent ion tolerance of 10 ppm. Oxidation of methionine and carbamidomethylation of cysteins were specified in Mascot as dynamic and static modification respectively, and one missed cleavage was allowed for tryptic cleavage. For the SILAC data, 13C6-l-Lysine and 13C615N4-l-Arginine modifications were specified in Mascot as dynamic or static modification depending on the experiment. The files generated from MASCOT (.DAT files) were then uploaded into Scaffold (version 3 00 07; Proteome software, Inc., Portland, OR) resulting in a nonredundant list of identified proteins per sample. Peptide identifications were accepted if they could be established at a PeptideProphet probability greater than 95% (18). Protein identifications were accepted if they could be established at greater than 95% probability and contained at least two identified spectra. Using these filters we always achieve a protein false discovery rate (FDR) under 1.0%, as estimated by a search against a decoy database. Protein isoforms and members of a protein family would be identified separately only if peptides that enable differentiation of isoforms had been identified based on generated MS/MS data. Otherwise, Scaffold would group all isoforms under the same gene name. Different proteins that contained similar peptides and which were not distinguishable based on MS/MS data alone were grouped to satisfy the principles of parsimony.

Relative label-free protein quantification analysis was performed on the different samples analysed using spectral counting. The “Number of assigned spectra” function of Scaffold software, which provides the total number of spectra that matched to a protein identified in each sample was used. Scaffold files containing all of the spectral counts for each sample and its replicates of a given experiment was generated and then exported to R software for normalisation and statistical analysis. The data from a MS/MS experiment was assembled in a matrix of spectral counts where the different conditions are represented by the columns, and the identified proteins are represented in the rows of that matrix. The need for normalisation was assessed by comparing the total spectral counts (SpC) in technical replicates of each sample. As the quantity of substance for each sample, in each experiment, was the same, any substantial deviation was corrected by normalizing to the median total sample counts. An exploratory data analysis by means of principal components analysis (PCA) and hierarchical clustering of the samples on the SpC matrix was performed to find potential outliers and patterns in the data. Dealing with counts precludes the use of statistical tests and procedures based on the normal distribution, and restricts the appropriate methods to those in the general frame of the Generalized Linear Models (GLM) (Agresti A. (2010) Analysis of Ordinal Categorical Data, John Wiley & Sons Inc) with discrete distributions. As no substantial biological variability is expected from cell line data, according to our experience, a Poisson

regression was used for significance testing throughout this work. The GLM model based on the Poisson distribution was used as a significance test throughout our work. Finally, we used the Benjamini and Hochberg multitest correction to adjust the p values with control on the FDR (Benjamini Y., et al. 1995).

18. Senescence-associated β-galactosidase staining

Cells seeded in 6-well plates were washed with PBS and fixed using the fixation solution (0.2% glutaraldehyde and 2% formaldehyde in PBS) for 15 minutes at room temperature. After washing with PBS, cells were incubated overnight at 37°C with the staining solution containing 1 mg/ml X-Gal, 40 mM citric acid/phosphatase buffer at pH 6.0, 2 mM $MgCl_2$, 150 mM NaCl, 5 mM $K_3Fe(CN)_6$ and 5 mM $K_4Fe(CN)_6$. Afterwards, cells were washed in PBS and pictures were taken using bright-field microscopy. 3 representative fields per well were counted and averaged to quantify the percentage of SA-β-Gal+ cells.

19. MIDORI interactome analysis by mass spectrometry

MDA-MB-231 cells overexpressing MIDORI-HA or the Empty vector for 24 hours were lysed in a buffer containing 50 mM Tris-HCl pH 7.5–8, 150 mM NaCl, 1%Triton X-100 and protease and phosphatase inhibitors and homogenised for 30 min in a rotor wheel. 5 mg of lysates were immunoprecipitated with 5 µg of monoclonal HA-tag antibody (Sigma) overnight at 4°C. Immunocomplexes were collected using PureProteome™ Protein A Magnetic Beads (MERCK) and eluted by competition incubating with synthetic HA peptide (Sigma) 5 hours at room temperature. Eluate was digested with trypsin and analysed by liquid chromatography-mass spectrometry on an Orbitrap Fusion Lumos™ Tribrid (Thermo Scientific). A twin database search was performed with two separated softwares, Thermo Proteome Discoverer v2.3.0.480 (PD) and MaxQuant v1.6.6.0 (MQ). The search engine nodes used were Sequest HT for PD and Andromeda for MQ. The databases used in the search was SwissProt Human (released 2019_05) including contaminants and the user proteins.
The search was run against targeted and decoy databases to determine the false discovery rate (FDR). Search parameters included trypsin enzyme specificity, allowing for two missed cleavage sites, oxidation in methionine and acetylation in protein N-terminal as dynamic modifications and Carbamidomethylation in cystein as static modification. Peptide mass tolerance was 10 ppm and the MS/MS tolerance was 0.6 Da. Peptides with a q-value lower than 0.01 and a FDR < 1% were considered as positive identifications with a high confidence level. For the quantitative analysis, contaminant identifications were removed and unique peptide spectrum matches of protein groups identified with Sequest HT and Andromeda were analysed with SAINTexpress-spc v3.11 (SAINTe) (Teo et al., 2014). SAINTe compares the prey control spectral counts with the prey test spectral counts for all available technical replicates. For each available bait and for each available replicate, as prey count the maximum count result between PD and MQ was taken. Once obtained this combined dataset, SAINTe algorithm was run. High confidence interactors were defined as those with Bayesian false discovery rate BFDR ≤ 0.02 and fold change FC ≥ 3.

19.1. Validation of MIDORI's interactors

For the validation of interactome analysis, MDA-MB-231 overexpressing MIDORI-HA or the Empty vector for 24 hours were used in a MIDORI pull-down, as described above, and the resulting IP output was incubated with the interactor antibody.
In alternative, U2OS cells or HEK293T cells were co-transfected with MIDORI-HA and ITGB1BP1-FLAG plasmids or in the case of MYBBP1A-HA it was transfected alone. The immunoprecipitation was performed as explained before. SDSPAGE
was used to visualise by anti-HA MYBBP1A or MIDORI-HA and anti-Flag proteins. When MYBBP1A was immunoprecipitated using anti-HA antibody, MIDORI was detected using the polyclonal custom-antibody (Proteogenix).

20. *In silico* analyses

20.1. Codon conservation

ZEB2-AS1 and *Zeb2os* coding potential was assessed using PhyloCSF (Lin et al., 2011), a comparative genomics algorithm that analyzes a multispecies nucleotide sequence alignment and score it according to phylogenetic codon conservation.

20.2. *LncRNA* structure prediction

Zeb2os secondary RNA structure was predicted using MFold web-server software (Zuker, 2003) using the cDNA sequence.

20.3. Protein features prediction

The different characteristics of the protein were predicted using publicly available prediction software in different web-servers. Using the amino acid sequence of human and the two mouse MIDORI, the three-dimensional protein-structure prediction was obtained i-Tasser software (Yang and Zhang, 2015). The subcellular protein localisation prediction was determined using DeepLoc 1.0 software (Almagro Armenteros et al., 2017). Finally, the prediction of a nuclear localisation signal (NLS) was carried out using the cNLS Mapper software (Kosugi et al., 2009).

20.4. Generation of Kaplan-Meier Plots

Kaplan-Meier (KM) plots were generated for survival analysis using KM plotter database (http://kmplot.com/analysis), a website database based on resources from TCGA database. The final prognostic KM plots were presented with a hazard ratio (HR), 95% confidence interval (CI) and log-rank P value. P value < .05 was considered statistically significant.

21. Statistical analysis

Data are presented as mean ± standard deviation (SD) or standard error of the mean (SEM) when working with mice. Differences between groups were analysed using two-way ANOVA or Student's T-test, or Fisher's Exact as specified. Differences were considered statistically significant based on p-value (*p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001). All statistical tests were two-sided and performed using GraphPad Prism (GraphPad Software Inc., San Diego, CA, United States).

22. Ethical statement

The animal studies in this work comply with the European, Spanish and Catalan Regulations for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes (Directive 2010/63; Spanish BOE RD 53/2013; Catalan DOGC 214/1997). All studies were carried out in the "Lab Animal Service Campus Vall d'Hebron (LAS-CVH)", registered and accredited at the Department de Medi Ambient i Habitatge by Generalitat of Catalonia government with register number B9900062. The experiments performed for this manuscript were linked to a project approved by Vall d'Hebron Ethics Committee, and the Commission of Animal Experimentation of Generalitat of Catalonia government.

23. Table 1. Antibodies

Target	Reference	Dilution WB	Dilution IF	Dilution IHC
HA-Tag	Ab9110, Abcam	1:2500		
HA-Tag	H6908, Sigma		1:150	
GAPDH	AM4300, Thermo Fisher Scientific	1:10000		
MIDORI	Proteogenix	1:500	1:100	1:500
SNAIL	3879, Cell Signaling	1:1000		
p53	SC-126, Santa Cruz	1:1000		
p21	Hugo291 Ab107799, Abcam			No dilution
β–TUBULIN	SC-32293, Santa Cruz	1:1000		
β–CATENIN	GTX633010, GeneTex		1:100	
E-CADHERIN	610182, BD	1:1000	1:100	
58K	NB600-412SS, NovusBio		1:100	
LAMP-2	NBP2-22217SS, NovusBio		1:100	
H3-HISTONE	Ab18521, Abcam	1:10000		
LAMIN-B1	PA5-19468, Thermo Fisher Scientific	1:1000		
Phalloidin-TRITC	P1951, Sigma-Aldrich		20 µg/ml	
pERK1/2	9101, Cell Signaling	1:1000		
Total ERK1/2	4695, Cell Signaling	1:1000		
pSMAD2	18338, Cell Signaling	1:1000		
pNF-kB	3033, Cell Signaling	1:1000		
Total NF-kB	8242, Cell Signaling	1:1000		
Acetyl-NF-kB	3045, Cell Signaling	1:500		
Flag-Tag	F1804, Sigma	1:2000		
HMGA1	Provided by Josep Villanueva (VHIO)	1:1000		
pS6	4858, Cell Signaling	1:1000		
Total S6	14467, Cell Signaling	1:1000		

24. Table 2. Primers

Human primers for RT-qPCR analysis

Primer Name	Forward	Reverse
GAPDH	GGACTCATGACCACAGTCCATGCC	TCAGGGATGACCTTGCCCACAG
hmMIDORI	CAGCTCCCGGAGCAAACTG	TCCGGCACATCATACGGATA
ZEB2-AS1	ATGAAGAAGCCGCGAAGTGT	CACACCCTAATACACATGCCCT

SNAIL	ACCACTATGCCGCGCTCTT	GGTCGTAGGGCTGCTGGAA
SLUG	TCGGACCCACACATTACCTT	ATGAGCCCTCAGATTTGACCT
ZEB1	GCACCTGAAGAGGACCAGAG	TGCATCTGGTGTTCCATTTT
ZEB2	CGCTTGACATCACTGAAGGA	CTTGCCACACTCTGTGCATT
CDH1	GTTTTCCACCAAAGTCACGC	TGTGAGCAATTCTGCTTGGA
CDH2	CTGCACAGATGTGGACAGGA	CCACAAACATCAGCACAAGG
TWIST1	GGACAAGCTGAGCAAGATTCA	CGGAGAAGGCGTAGCTGAG
VIMENTIN	GACAATGCGTCTCTGGCACGTCTT	TCCTCCGCCTCCTGCAGGTTCTT
FIBRONECTIN	ATCCCTGGTGCTGATGGAC	ACCTTGTTTGCCAGGTTCAC
COLLAGEN I	ATGTCTAGGGTCTAGACATGTTCA	CCTTGCCGTTGTCGCAGACG
MMP-2	CTCATCGCAGATGCCTGGAA	TTCAGGTAATAGGCACCCTTGAAGA
MMP-9	ACGCACGACGTCTTCCAGTA	CCACCTGGTTCAACTCACTCC
pre45S	CCGTCCGTCCGTCGTCCTCCTCGC	TGTACCGGCCGTGCGTACTTAGAC
18S	CGGCTACCACATCCAAGG	TACAGGGCCTCGAAAGAGTC
28S	CTAAATACCGGCACGAGACC	TTCACGCCCTCTTGAACTCT
5.8S	ACTCTTAGCGGTGGATCACTC	AAGCGACGCTCAGACAGG
IL1α	GAATGACGCCCTCAATCAAAGT	TCATCTTGGGCAGTCACATACA
IL1β	CTGAAAGCTCTCCACCTCCA	CCAAGGCCACAGGTATTTTG
IL6	GGTACATCCTCGACGGCATCT	GTGCCTCTTTGCTGCTTTCAC
TNFα	CTCTTCTGCCTGCTGCACTTTG	ATGGGCTACAGGCTTGTCACTC
CCL2	CCCCAGTCACCTGCTGTTAT	TGGAATCCTGAACCCACTTC
CCL5	GAGTATTTCTACACCAGTGGCAAG	TCCCGAACCCATTTCTTCTCT
CCL20	CCAAGAGTTTGCTCCTGGCT	TGCTTGCTGCTTCTGATTCG
CXCL10	TGGCATTCAAGGAGTACCTCTC	GGACAAAATTGGCTTGCAGGA
SERPINE1	GATTCATCATCAATGACTGGGTG	GCCGTTGAAGTAGAGGGCAT
αSMA	CTGAAGAGCATCCCACCCT	ACATGGCTGGGACATTGAAA

Mouse primers for genotyping

Primer Name	**Forward**	**Reverse**
MIDORI WT allele	AGTTTTGGCCAGAAATGGTG	AATTATCCCTCCCCACAAAG
MIDORI KO allele	TAGTTTTGGCCAGAATGATC	AATTATCCCTCCCCACAAAG
Zeb2os locus amp.	CATGAAGAAGCCGCGAAGTG	GCGTTTGCGGAGACTTCAAG
ORF2 (OSKM)	GGATGGAGTGGGACAGAGAA	GTGCCGATCCGTTCACTAAT
125B (RTTA)	AAAGTCGCTCTGAGTTGTTAT	
126B (RTTA)	GCGAAGAGTTTGTCCTCAACC	
127B (RTTA)	GGAGCGGGAGAAATGGATATG	
MMTV-PyMT	GGAAGCAAGTACTTCACAAGGG	GGAAAGTCACTAGGAGCAGGG

Mouse primers for RT-qPCR analysis

Primer Name	**Forward**	**Reverse**

Gapdh	TGTGTCCGTCGTGGATCTGA	TTGCTGTTGAAGTCGCAGGAG
hmMIDORI	TAGTTTTGGCCAGAATGATC	AATTATCCCTCCCCACAAAG
Zeb2os	AAGAGTGTCTGGAGGCAGG	TGGAGAAGGGAGGGAGGG
Zeb2os isoform1	TCTGCTTAGAAAACAGCCTTCTC	CATTTCTGGCCAAAACAAATTCG
Zeb2os isoform2	CGCTGTGCCTCGAGGATTAG	AGACAATAGGTGCAGGGTGTAG
Snail	AAGATGCACATCCGAAGCCA	CTCTTGGTGCTTGTGGAGCA
Slug	AAGGCTTTCTCCAGACCCTG	GTGCCCTCAGGTTTGATCTG
Zeb1	GGCCTTCAAGTACAAACACCA	AGGAGCCAGAATGGGAAAAC
Zeb2	AGGCATATGGTGACGCACAA	CTTGAACTTGCGGTTACCTGC
Cdh1	CCGGGACTCCAGTCATAGG	CAGCTCTGGGTTGGATTCAG
Cdh2	GAAGGTGGAGGAGAAGAAGACC	TCGTCTAGCCGTCTGATTCC
Twist1	GGACAAGCTGAGCAAGATTCA	CGGAGAAGGCGTAGCTGAG
Vimentin	CCAACCTTTTCTTCCCTGAA	TGAGTGGGTGTCAACCAGAG
Fibronectin	ATCCCTGGTGCTGATGGAC	ACCTTGTTTGCCAGGTTCAC
Collagen I	ATGGATTCCCGTTCGAGTACG	TCAGCTGGATAGCGACATCG
Mmp-2	AACCCAGATGTGGCCAACTA	TTCAGGGTCCAGGTCAGGT
Mmp-9	TAGCTACCTCGAGGGCTTCC	GCTGTGGTTCAGTTGTGGTG
αSma	CTGAAGAGCATCCCACCCT	ACATGGCTGGGACATTGAAA
Ink4a	CGTACCCCGATTCAGGTGAT	TTGAGCAGAAGAGCTGCTACGT
Arf	GCCGCACCGGAATCCT	TTGAGCAGAAGAGCTGCTACGT
Il1α	AAGTCTCCAGGGCAGAGAGG	CTGATTCAGAGAGAGATGGTCAA
Il1β	AAAAGCCTCGTGCTGTCG	AGGCCACAGGTATTTTGTCG
Il6	GTTCTCTGGGAAATCGTGGA	GGTACTCCAGAAGACCAGAGGA
Tnfα	GCCTCTTCTCATTCCTGCTT	CTCCTCCACTTGGTGGTTTG
Ccl2	TCTCTCTTCCTCCACCACCA	TCATTGGGATCATCTTGCTG
Ccl5	TGCCCACGTCAAGGAGTATTTC	AACCCACTTCTTCTCTGGGTTG
Tgfβ	ACCATGCCAACTTCTGTCTGGGA	ATGTTGGACAACTGCTCCACCTTG

25. Table 3. Plasmids

Retroviral packaging vectors

pCL-Ampho
pCL-Eco

Retroviral vectors

pMSCV-hMIDORI_Cterm
pMSCV-mMIDORI_short
pMSCV-mMIDORI_long
pMSCV-empty

Lentiviral packaging vectors

pLP1
pLP2
pLP- VSVG

Lentiviral vectors

pFUW-TetO-OSKM
pFUGW
pFUW-RTTA
pINDUCER20-empty
pINDUCER20-hMIDORI
pINDUCER20-mMIDORI_short
pINDUCER20-mMIDORI_long

26. Table 4. Overexpression constructs

Construct Name	***Sequence***
hMIDORI_HA	taagcaCTCGAGCGCCACC**ATG**GTGAGAAGAAAAAGCATGAAGAAGCC GCGAAGTGTGGGGGAGAAAAAGGTGGAAGCGAAGAAACAGCTCCCG GAGCAAACTGTACAAAAACCTCGCCAAGAGTGTCGGGAGGCAGGACC GTTATTCCTGCAGAGCAGGAGAGAGACGAGAGACCCTGAAACACGCG CCACCTATCTTTGTGGGGAGGGAGGCGGTGGTGGCAGTGGTGGCGG AGGAAGCGGTGGGGGAGGCAGCTATCCGTATGATGTGCCGGATTATG CG**TGA**GAATTctgctt
Long mMIDORI_HA	taagcaGAATTCCTCGAGCGCCACC**ATG**CAAACTGTAACACAAGGGCCA TTAACCCTTTCTCTGCCGCCCAGCACGCCCAGTCCCTACACCCTGCAC CTATTGTCTCCGGTGCCACAAGCTTTGCGGAAAACCTGGAAATTTTGG CCAGAAATGGTGAGAAGAAAAAGCATGAAGAAGCCGCGAAGTGTGGG GGAGAAAAAGGTGGAGGCGAAGAAACAGCTCCCGGAGCAAACTGAA CAAAACCTCGCCAAGAGTGTCTGGAGGCAGGACCGTTATTCCTGCCG AGCAGGGGAGAGAAGAGAGACCCTGAAACACGCGCCACCTATCTTTG TGGGGAGGGATAATGGCGGTGGTGGCAGTGGTGGCGGAGGAAGCG GTGGGGGAGGCAGCTATCCGTATGATGTGCCGGATTATGCG**TGA**GAA TTctgctt
Short mMIDORI_HA	taagcaGAATTCCTCGAGCGCCACC**ATG**GTGAGAAGAAAAAGCATGAAG AAGCCGCGAAGTGTGGGGGAGAAAAAGGTGGAGGCGAAGAAACAGC TCCCGGAGCAAACTGAACAAAACCTCGCCAAGAGTGTCTGGAGGCAG GACCGTTATTCCTGCCGAGCAGGGGAGAGAAGAGAGACCCTGAAACA CGCGCCACCTATCTTTGTGGGGAGGGATAATGGCGGTGGTGGCAGTG GTGGCGGAGGAAGCGGTGGGGGAGGCAGCTATCCGTATGATGTGCC GGATTATGCG**TGA**GAATTctgctt

RESULTS

1. Identification of MIDORI microprotein

1.1. *ZEB2-AS1* lncRNA encodes for a 64-amino acid microprotein

To identify novel microproteins relevant for cancer cell plasticity we followed a computational approach. Among the "assumed" non-coding regions that could encode for microproteins we focused in lncRNAs, given that most of them are annotated and in many cases their expression and regulation is characterised. We generated a list of lncRNAs deregulated in cancer and in cellular differentiation-dedifferentiation processes and analysed their coding potential (Figure 10A). For that we used PhyloCSF, a comparative genomics algorithm that analyses codon substitution frequencies across 35 vertebrate species to predict coding sequences based in phylogenetic conservation (Lin et al., 2011). Following this approach, we identified a previously unrecognised sORF in the lncRNA *ZEB2-AS1,* that we have named MIDORI (Figure 10A).

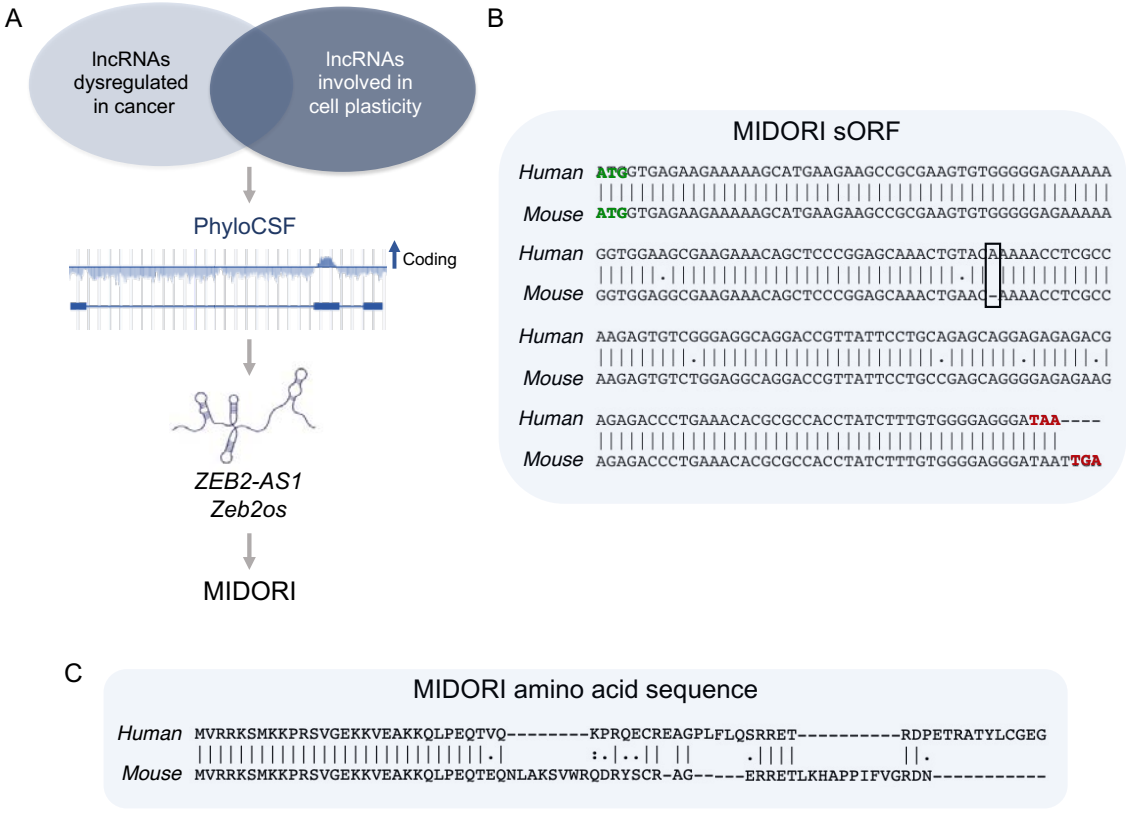


Figure 10. Identification of MIDORI sORF in *ZEB2-AS1*. (A) Strategy for the discovery of novel microproteins involved in cancer cell plasticity. (B) Alignment of human and mouse MIDORI sORF nucleotide sequence. In primates occurred a single nucleotide deletion that is highlighted with the black box. Start codons are marked in green while stop codons are in red.

Phylogenetic analyses showed that MIDORI is conserved across placental mammals. However, in primates occurred a single nucleotide deletion that led to a frameshift and changed the amino acid sequence of the C-terminal part of MIDORI (Figure 10B-C). We analysed in detail the human and mouse locus. In humans, MIDORI sORF span exons 3 and 4 of the *ZEB2-AS1* transcript and potentially produces a 64-amino acid

microprotein that is totally conserved through primates (Figure 11A). In mouse, the orthologue locus (*Zeb2os*) produces two main splicing isoforms that differ in the first exon (Figure 11B). PhyloCSF analysis revealed that the murine locus could encode two MIDORI protein isoforms: one of 65 amino acids (short MIDORI) that starts in a methionine in exon 2, and other of 107 amino acids (long MIDORI) coming from an upstream methionine only present in exon 1 of Isoform 1 (long MIDORI) (Figure 11B). Since the two starting methionines are in frame, both microproteins share most of the sequence, and long MIDORI contains short MIDORI (Figure 11B).

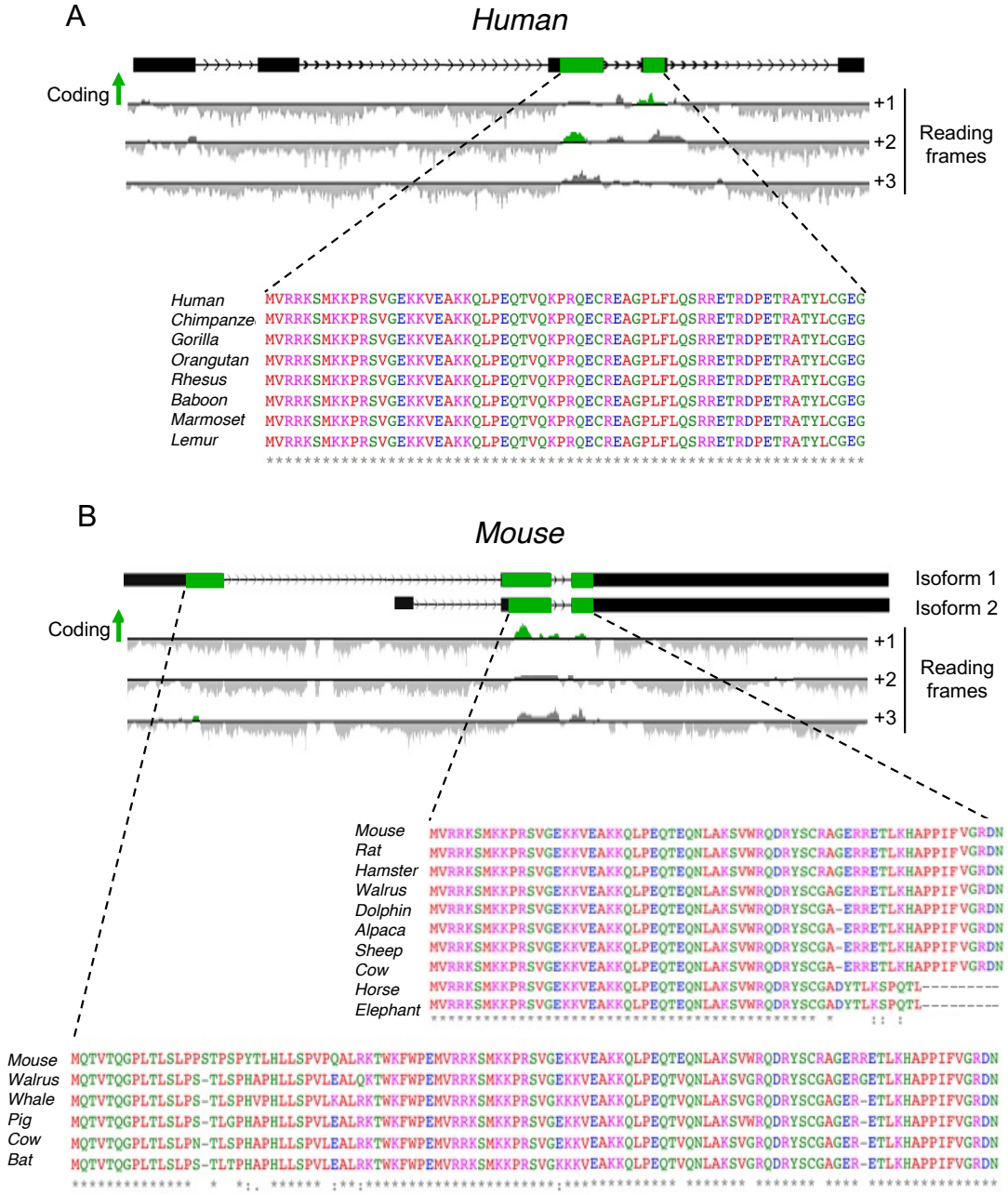


Figure 11. *ZEB2-AS1* and *Zeb2os* are evolutionarily conserved at the aminoacidic level. (A) (Top) Analysis of codon conservation in human *ZEB2-AS1* locus by PhyloCSF software, which shows a positive deflection in the first and second reading frame corresponding to MIDORI sORF (highlighted in green). (Bottom) Human MIDORI sORF amino acidic conservation across primates. (B) (Top) Analysis of codon conservation in the murine *Zeb2os* locus by PhyloCSF software, which shows a positive deflection (green) in the first reading frame corresponding to

short mouse MIDORI. An additional peak in the third reading frame corresponds to the upstream methionine of long mouse MIDORI localised in the first exon of *Zeb2os* Isoform 1 (long and short MIDORI sORF are highlighted in green). (Bottom) Short mouse MIDORI and long mouse MIDORI sORFs amino acidic conservation across vertebrates.

ZEB2-AS1 is described as a lncRNA antisense to *ZEB2* gene that is involved in the TGFβ pathway. Its expression is directly controlled by SNAIL, and it is known to allow ZEB2 translation by regulating its splicing (Beltran et al., 2008). In fact, the expression of this lncRNA is required for the progression through the Epithelial-to-Mesenchymal Transition (EMT) (Beltran et al., 2008) (Figure 12A). In agreement, high expression of *ZEB2-AS1* has been correlated with tumour metastasis and worse prognosis in bladder cancer (Zhuang et al., 2015) as well as in hepatocellular carcinoma (Lan et al., 2016). We have corroborated these data by extending the study to other cancer types (Figure 12B). *ZEB2-AS1* lncRNA high expression significantly correlates with reduced overall survival (OS) of patients with breast cancer (Log rank test $p=1.3e^{-06}$, HR=2.24), pancreatic adenocarcinoma (PDAC) (Log rank test p=0.0023, HR=2.24) and ovarian cancer (Log rank test $p=1.5e^{-05}$, HR=1.81).

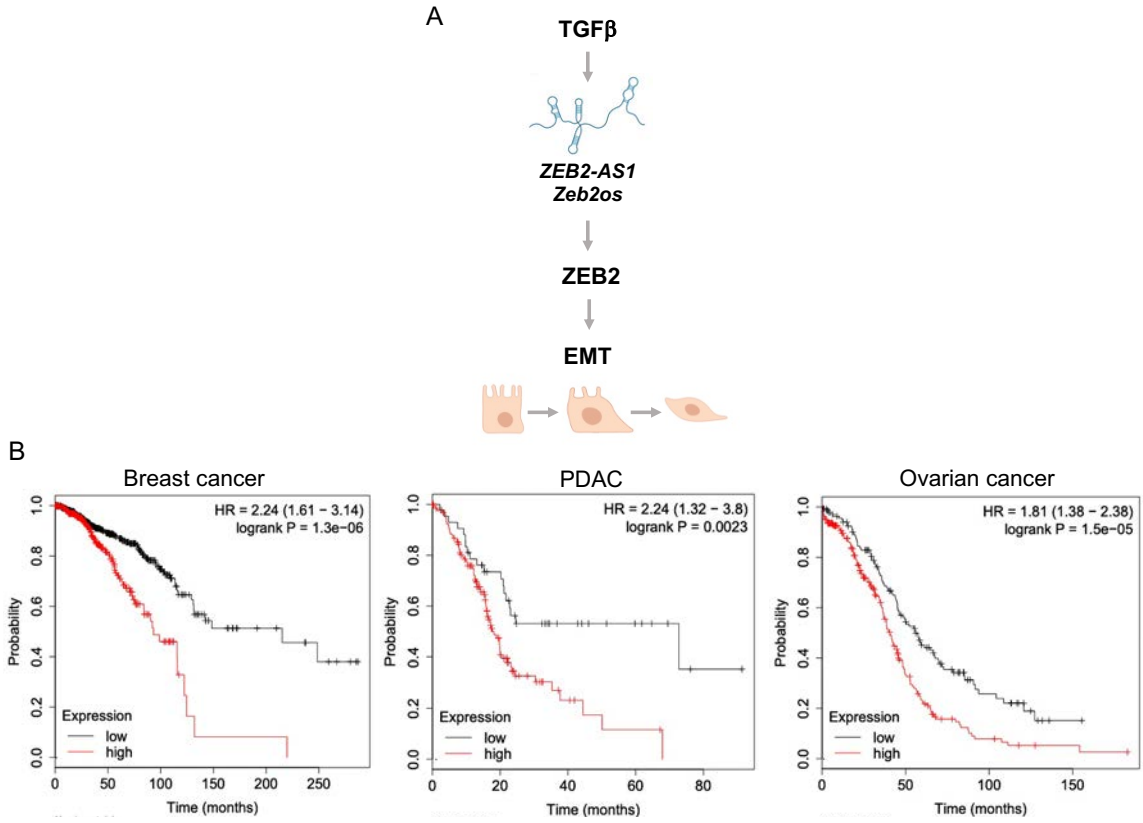


Figure 12. Role of *ZEB2-AS1* in cancer. (A) Scheme summarising the role of *ZEB2-AS1* in the TGFβ pathway and EMT. (B) Kaplan-Meier curves showing the correlation between *ZEB2-AS1* expression and patient overall survival in the indicated tumour types.

Of note, *ZEB2-AS1* has not only been described to be dysregulated in cancer, but it also plays a role in cellular reprogramming and in the maintenance of pluripotency (Bernardes de Jesus et al., 2018). In particular, *Zeb2os* (*ZEB2-AS1* mouse orthologue) has been

showed to be upregulated in mouse aged fibroblasts and acts as a barrier for cellular reprogramming to iPSCs (Figure 13A). Consistently, knock-down of Zeb2os increases cellular reprogramming and prevents mouse embryonic stem cells (mESCs) differentiation (Bernardes de Jesus et al., 2018). Importantly, we have confirmed that *ZEB2-AS1* is also upregulated in IMR90 human fibroblasts upon replicative senescence (Figure 13B).

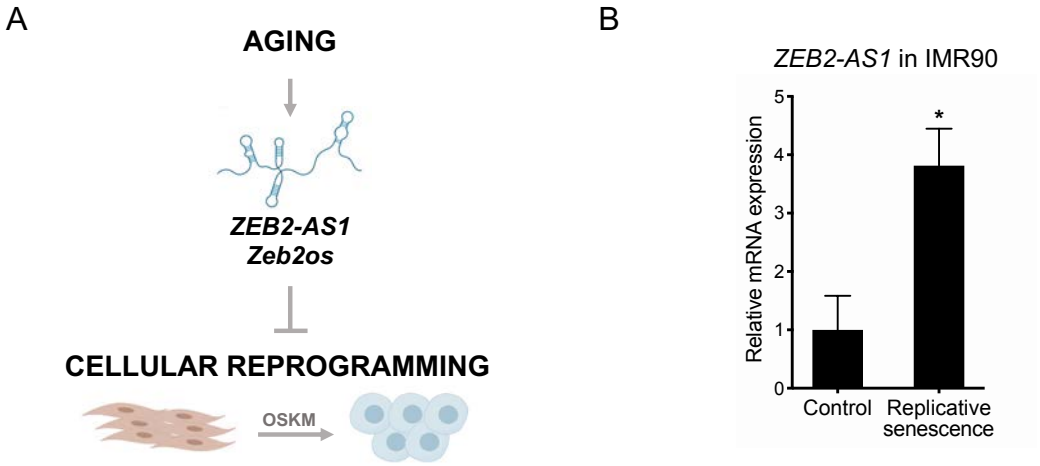


Figure 13. *ZEB2-AS1* is upregulated during aging and impairs cellular reprogramming. (A) Scheme summarising the regulation of *ZEB2-AS1* in aging and its role in cellular reprogramming. (B) mRNA expression of *ZEB2-AS1* in low passage human IMR90 fibroblasts and upon replicative senescence. Expression is normalized to *GAPDH* and relativized to the control. N=3 of three technical replicates. Error bar represents the mean ± SD. **P* < 0.05 using Student's t-test.

Altogether, this evidence points to *ZEB2-AS1* at the perfect overlay between cancer and cellular plasticity and encouraged us to pursue the functional characterization of MIDORI, its encoded microprotein.

First of all, we wanted to validate the translation of MIDORI sORF into a protein. We analysed publicly available data of ribosome profiling experiments performed in mouse hippocampus (Wang et al., 2020) using RiboORF, a software that identifies regions of active translation based on read distribution features such as codon periodicity and uniformity (Ruiz-Orera et al., 2018). Importantly, RiboORF analysis showed that *Zeb2os* is translated with a codon periodicity that corresponds to the sORFs predicted by the phylogenetic analysis (Figure 14A). Of note, ribosomes were found to bind from the first methionine of *Zeb2os* Isoform 1, demonstrating that long MIDORI is in fact translated, although we cannot distinguish if short MIDORI is also translated from the second methionine in exon 2. To further validate the coding potential of *Zeb2os*, we performed *in vitro* translation using the full length *Zeb2os* isoform 1 in presence of 35^{s}-methionine; we obtained a peptide product of ~12 kDa that fits long MIDORI's molecular weight (12.5 kDa) and a secondary band that matches the size of short MIDORI (7.5 kDa) (Figure 14B). Altogether, we conclude that ZEB2AS1 is a transcript misannotated as a lncRNA that actually codes for MIDORI microprotein.

Both analyses -RiboORF and *in vitro* translation- strongly support the coding potential of *Zeb2os* at the experimental level. However, with these approaches we were not able to decipher which one is the main protein product of the murine *Zeb2os* locus. Since

Zeb2os is known to be upregulated by TGFβ (Beltran et al., 2008), we measured the expression of the two transcripts by using isoform-specific primer pairs (Figure 14C). Although the two transcripts seem to have some level of basal expression, we saw that only *Zeb2os* transcript 1 is significantly upregulated after TGFβ treatment. Taking into account that normally ribosomes start translating the first methionine of a transcript, this result suggests that long MIDORI isoform might be the prevalent isoform in mouse.

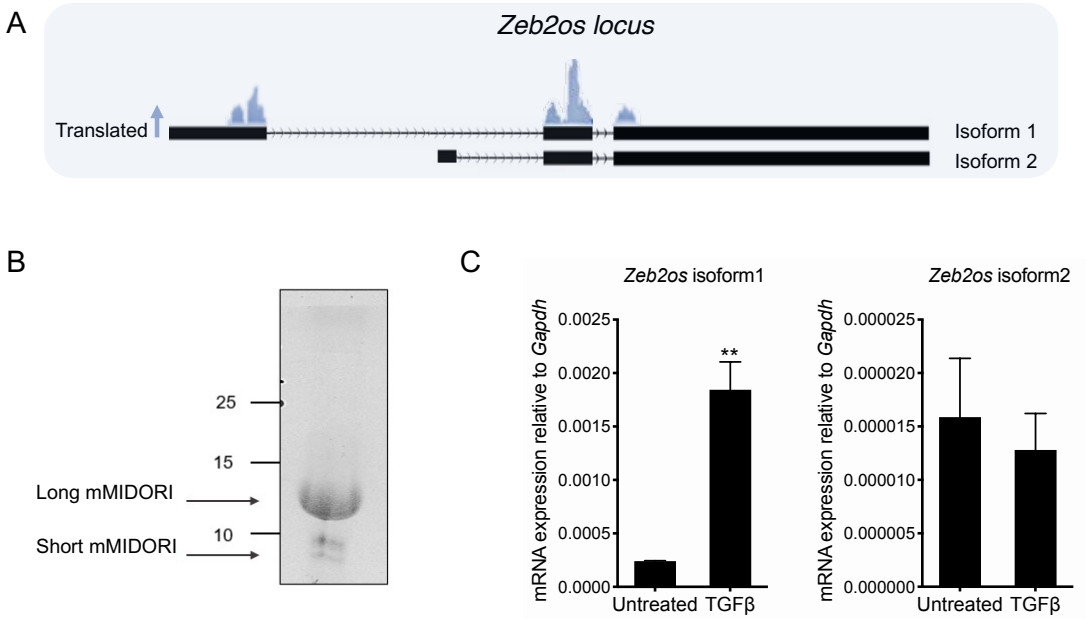


Figure 14. Experimental evidences of *Zeb2os* translation. (A) Representation of Ribo-Seq data performed in mouse hippocampus and analysed with RibORF (B) *Zeb2os in vitro* translation. (C) mRNA expression of Zeb2os isoforms in mouse embryonic fibroblasts (MEFs) in basal conditions and upon 24 hours of TGFβ treatment. Values are represented as mean ± SD of three technical replicates normalised to *Gapdh*. $^{**}P < 0.01$ using Student's t-test.

Then, we performed different *in silico* analyses to predict the main features of MIDORI. On the one hand, we predicted the 3D structure of the microprotein by i-TASSER software (Yang and Zhang, 2015) (Figure 15A), and observed that some short helices are predicted with low confidence score (-4.11, where the lowest score is -5 and the highest +2), but MIDORI is essentially predicted to be an unstructured protein. On the other hand, using the DeepLoc 1.0 software (Almagro Armenteros et al., 2017), we observed that both human MIDORI (hMIDORI) and the two mouse MIDORI (mMIDORI) isoforms are predicted to be located in the nucleus (Figure 15B). In agreement with this prediction, a nuclear localisation signal (NLS) was found in MIDORI's sequence (Figure 15C) using the cNLS Mapper software (Kosugi et al., 2009), further suggesting a nuclear localisation of MIDORI microprotein.

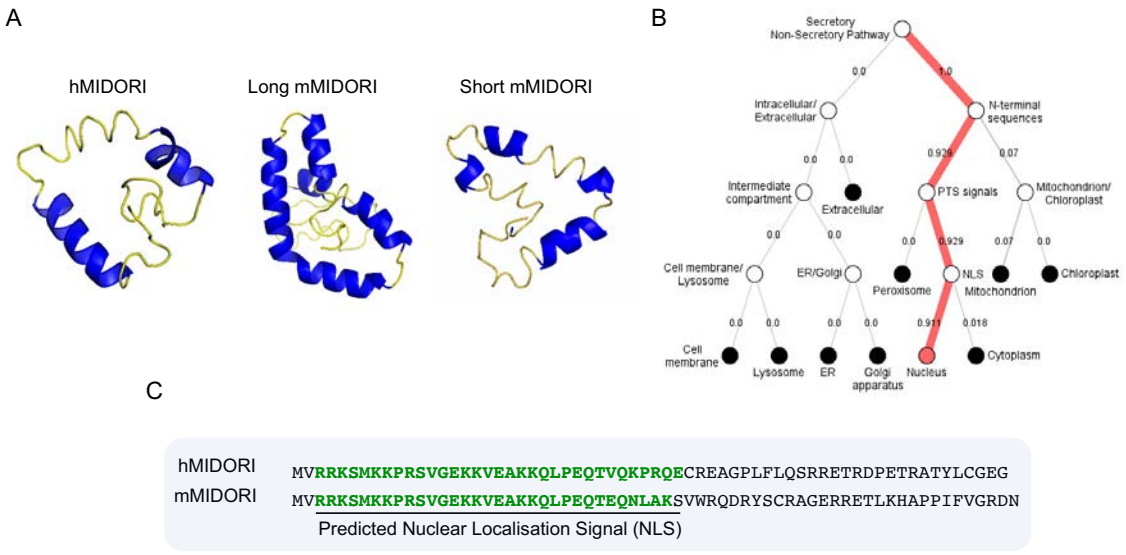


Figure 15. *In silico* characterisation of MIDORI microprotein. (A) Predicted tertiary structure of hMIDORI, long mMIDORI and short mMIDORI by i-TASSER. (B) Predicted nuclear localisation of hMIDORI (as an example of the three isoforms) using DeepLoc 1.0. Numbers in the branches indicate the confidence score of the protein being at that compartment (0.911 in the case of both human and mouse MIDORI) (C) Nuclear localisation signal (NLS, shown in green) prediction by cNLS Mapper software in MIDORI's sequence.

1.2. Generation of the tools to study MIDORI microprotein

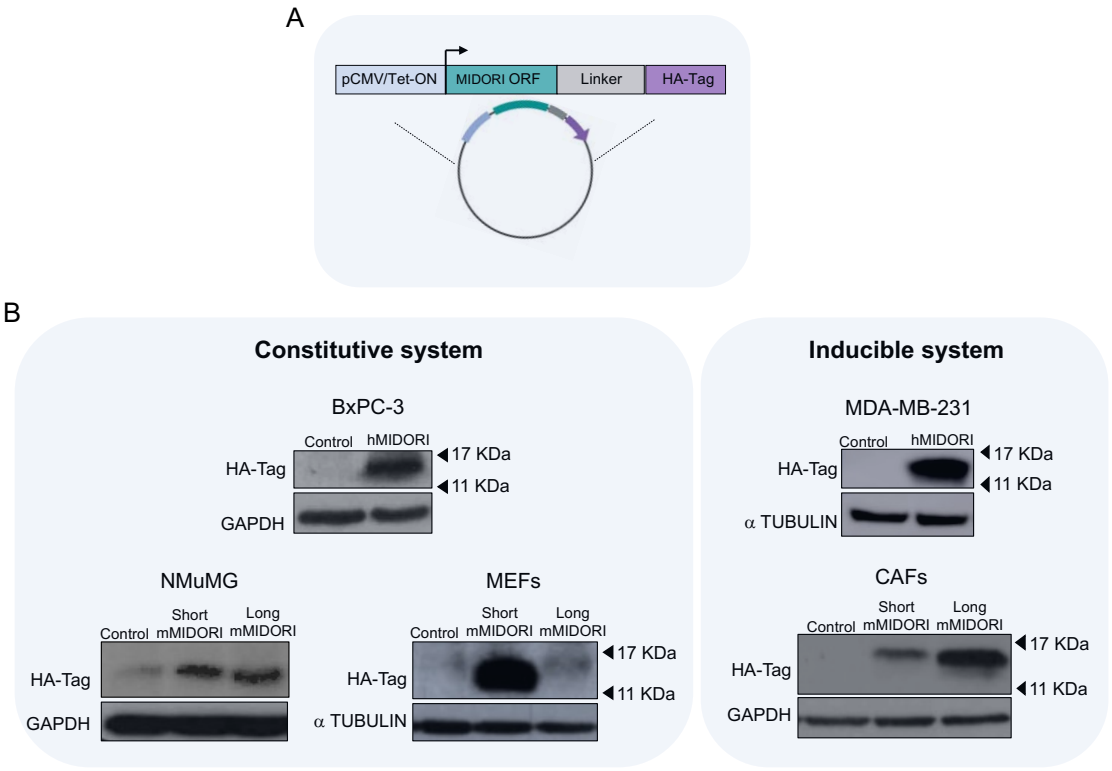
1.2.1. Generation of MIDORI overexpression vectors

To functionally characterise MIDORI microprotein we used gain-of-function and loss-of-function approaches. To study MIDORI gain-of-function, we generated MIDORI-overexpressing vectors by cloning the sORF of the human and murine MIDORI (short and long isoforms) tagged with an HA into the pMSCV constitutive retroviral vector and into the pINDUCER20 doxycycline-inducible lentiviral vector. To minimise the possible effect of the tag over MIDORI, we separated the microprotein and the HA by a linker of low interacting amino acids (3xGGGGS) (Figure 16A).

We transduced human and murine breast cell lines (MDA-MB-231, NMuMG), a human PDAC cell line (BxPC-3), murine CAFs and primary mouse embryonic fibroblasts (MEFs) with the different expression vectors: the human cell lines with hMIDORI, the murine cells with the two mMIDORI isorforms. Importantly, we detected hMIDORI and the two mMIDORI isoforms by immunoblotting against the HA-tag in all the analysed conditions (Figure 16B), showing that MIDORI microproteins can be expressed and are stable in cellular conditions. Additionally, we performed immunofluorescence (IF) to detect the exogenous MIDORI using an antibody against the HA-Tag (Figure 16C-D). Both hMIDORI and mMIDORI isoforms are detected by IF in the cytoplasm and in the nucleus. To further study MIDORI's localisation, we carried out subcellular fractionation in hMIDORI-overexpressing MDA-MD-231 cells and analysed the different fractions by Western blot. Exogenous hMIDORI was detected in cytoplasmatic, nuclear and

chromatin fractions suggesting that, at least when overexpressed, MIDORI seems to be an ubiquitous microprotein in the cell (Figure 16E).

Although all the experiments with MIDORI have been performed with the two murine isoforms, given that our results suggest that in mouse long MIDORI is the predominant isoform (Figure 14) and for the sake of simplicity, in this work from now we show the results obtained with human MIDORI (in human cells) and with mouse long MIDORI (in murine cells). However, after the end of each section we show that we can corroborate the main observed phenotypes also with short MIDORI (see below).



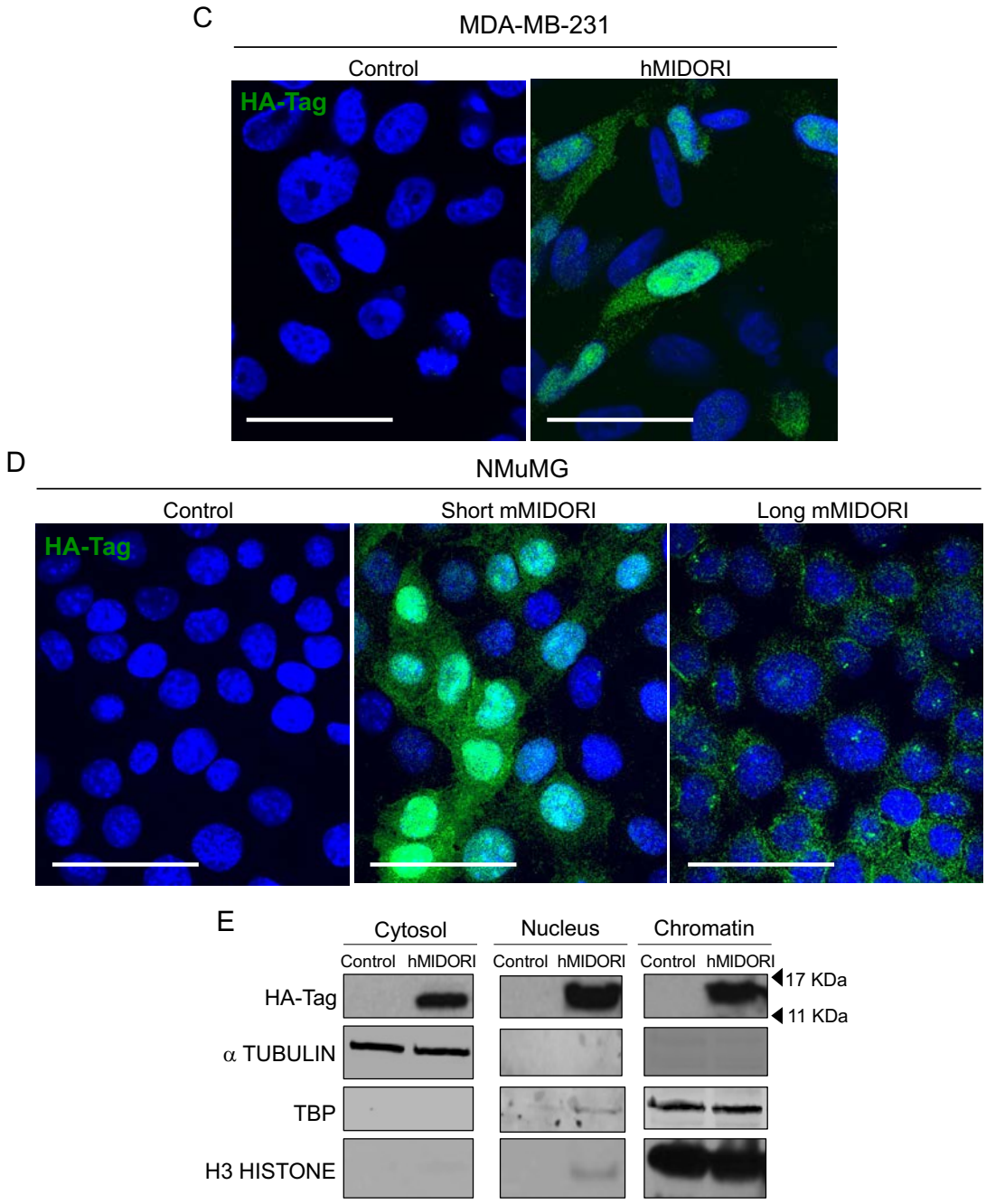


Figure 16. Generation and validation of MIDORI gain-of-function expression vectors. (A) Design of the constitutive and the doxycycline-inducible MIDORI-overexpressing constructs. (B) Detection of human and mouse MIDORI overexpression in the indicated cells lines after retroviral transduction and selection (in the case of constitutive systems) or after 48 hours of doxycycline induction (1 μg/ml) (in the case of inducible systems). Control cells were transduced with the empty backbone. Protein expression was detected by immunoblot using an anti-HA-tag antibody. (C-D) Representative immunofluorescence images of exogenous MIDORI detected by using an anti-HA-Tag antibody (green) in control and hMIDORI-overexpressing MDA-MB-231 breast cancer cells (C) and mMIDORI-overexpressing NMuMG breast epithelial-like cells (D). Nuclei are counterstained with DAPI. Scale bars=25μm. (E) WB of subcellular fractionation of cytosol, nucleus, and chromatin fractions in MDA-MB-231 cells. MIDORI overexpression was detected by anti-HA-Tag antibody while α-Tubulin, TBP and H3 histone are markers for cytoplasmatic, nuclear and chromatin fractions, respectively.

1.2.2. Generation of a MIDORI-deficient mouse strain

To assess MIDORI loss-of-function, we generated a MIDORI-deficient mouse strain using CRISPR-Cas9 technology. Since the secondary and tertiary structure of lncRNAs is known to be important for their function (Johnsson et al., 2014; Zampetaki et al., 2018), we sought to disrupt MIDORI translation without affecting the folding and function of *Zeb2os* lncRNA. To do so, instead of disrupting the locus with indels, we used a homologous recombination template in which we substituted short MIDORI start codon by a stop codon and mutated the PAM sequence to avoid recognition by the gRNA-Cas9 complex once the genomic region has been modified. We decided to mutate short MIDORI start codon because this way we can disrupt both MIDORI isoforms. On the one hand, if the expressed *Zeb2os* isoform is the 1, only a truncated form of the long MIDORI lacking the majority of the microprotein sequence would be translated. On the other hand, if the expressed *Zeb2os* isoform is the 2, the microprotein translation would be completely abolished. In addition, while mutating the PAM, we introduced a GTC>CTA mutation to generate a restriction site in order to be able to detect the recombination by enzymatic digestion. Finally, since MIDORI has an additional methionine 16 nucleotides downstream the start codon we also mutated it (ATG>TTG) to avoid downstream translation (Figure 17).

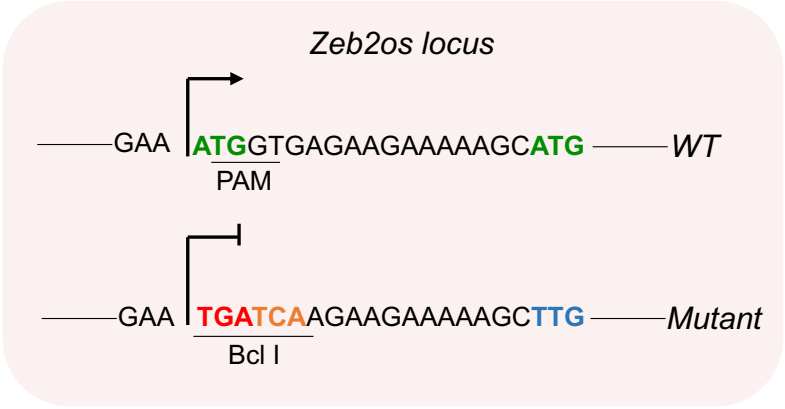


Figure 17. Strategy for the generation of a MIDORI-deficient mouse strain using CRISPR-Cas9 technology. Image showing the MIDORI WT allele and the MIDORI-KO allele. We substituted the start codon (green) by a stop codon (red), mutated the PAM sequence, introduced GTC>CTA mutation to include a restriction site (in yellow) and mutated the second downstream MIDORI start codon (in blue).

We analysed the wild type and the mutated *Zeb2os* transcript sequences using Mfold lncRNA structure predictor (Zuker, 2003), and we observed a change in the ATG loop but not in the overall lncRNA structure and the mutations did not significantly affect the stability of the entire molecule (Fig 18).

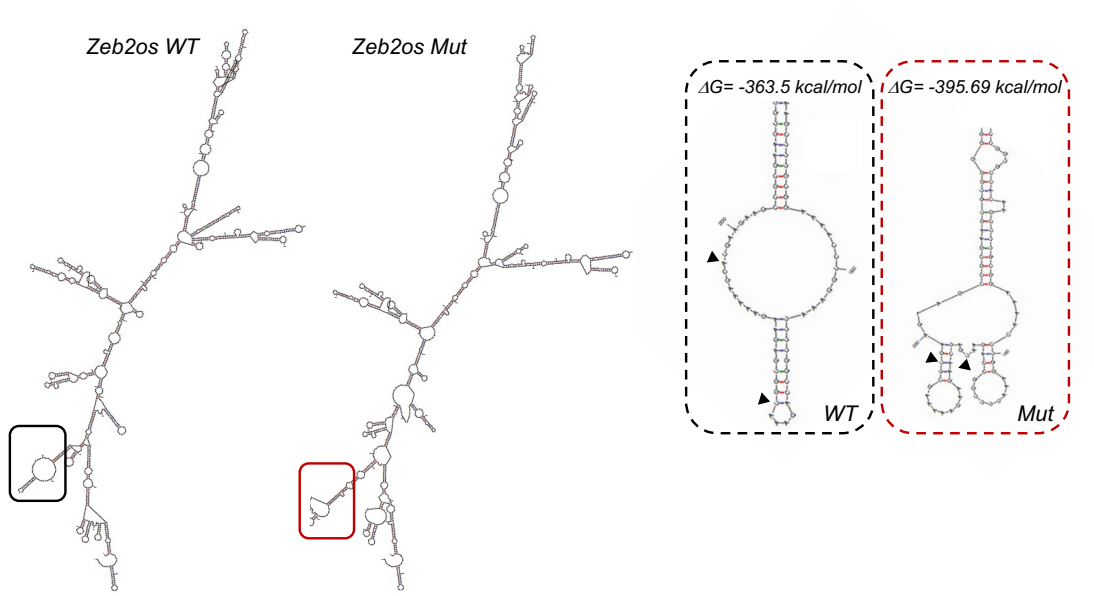


Figure 18. *Zeb2os* engineering did not significantly affect RNA stability or structure. Left, structure of wild type and mutated *Zeb2os* predicted by Mfold software. Right, detail of the modification in the RNA structure with the minimum free energy of both WT and mutated molecules. The arrows indicate the modified ATG codons.

In collaboration with Sagrario Ortega (Head of the Transgenic Mice Unit at the CNIO), we delivered the Cas9, gRNA and the homologous recombination template in mice embryos by microinjection and we successfully obtained the desired modification in some mice that we used as founders to establish a colony of MIDORI deficient mice (Fig 19A-B). Once the colony was established, we isolated WT and KO mouse embryonic fibroblasts (MEFs) and we analysed *Zeb2os* mRNA expression in basal conditions and upon TGFβ (Figure 19C). We observed that the lncRNA is similarly expressed in both WT and KO MEFs in basal conditions and, more importantly, it is upregulated following TGFβ treatment. Therefore, we have disrupted the translation of MIDORI microprotein without affecting the regulation of *Zeb2os* lncRNA.

We performed a histological analysis of many organs in WT and MIDORI-KO mice, and we did not observe any clear difference or alteration in any of the analysed tissues (data not shown). Additionally, MIDORI-KO mice show normal behaviour, are fertile and the offspring analysis showed that there is not any allelic skewing, suggesting that MIDORI deficiency does not induce embryonic lethality. Therefore, we concluded that, at least in homeostatic conditions, MIDORI deficiency does not have any profound effect in adult mice.

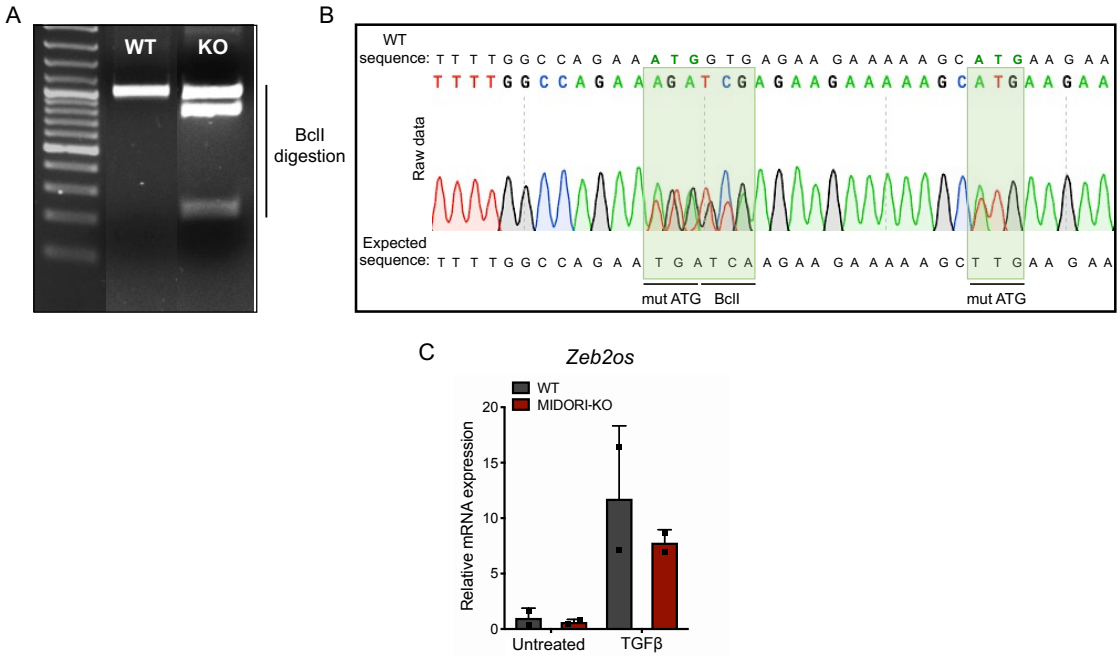


Figure 19. *Zeb2os* engineering does not affect the regulation of the *Zeb2os* lncRNA. (A) Genotyping of the engineered mice to detect the desired modification by BclI restriction enzyme digestion. (B) Result of the sequencing of *Zeb2os* locus in one of the founders carrying the modification in heterozygosis. The changes introduced in the locus are highlighted in green: start codon substitution, mutation of the PAM sequence, introduction of a restriction site for genotyping and mutation of the downstream ATG. (C) mRNA expression of *Zeb2os* in MEFs in basal conditions and after 24 hours of TGFβ treatment. Values are represented as mean ± SD of two biological replicates normalised to *Gapdh* and relativised to the untreated WT control.

1.2.3. Generation of an anti-MIDORI antibody

Finally, in order to detect the expression of the endogenous MIDORI microprotein we generated a custom rabbit polyclonal antibody. We commissioned the generation of the antibody to the company Proteogenix, which synthesised hMIDORI microprotein and used it to immunise two rabbits. The purified antibody has been used throughout this thesis.

1.3. Analysis of the regulation of MIDORI expression and its subcellular localisation

1.3.1. *ZEB2-AS1* and MIDORI expression in tissues

Many lncRNAs are expressed in a tissue-specific manner (Cabili et al., 2011). To study the expression in human tissues, we analysed public data retrieved from the GTEX database, and observed that *ZEB2-AS1* is expressed at low levels in all the analysed tissues (Fig 20A). We studied the *Zeb2os* expression in a set of mouse organs by qRT-PCR. Similarly to what we observed in human, *Zeb2os* is expressed at low or very low levels in the analysed organs (Figure 20B). Thus, we did not observe a clear pattern of specific tissue expression in basal conditions.

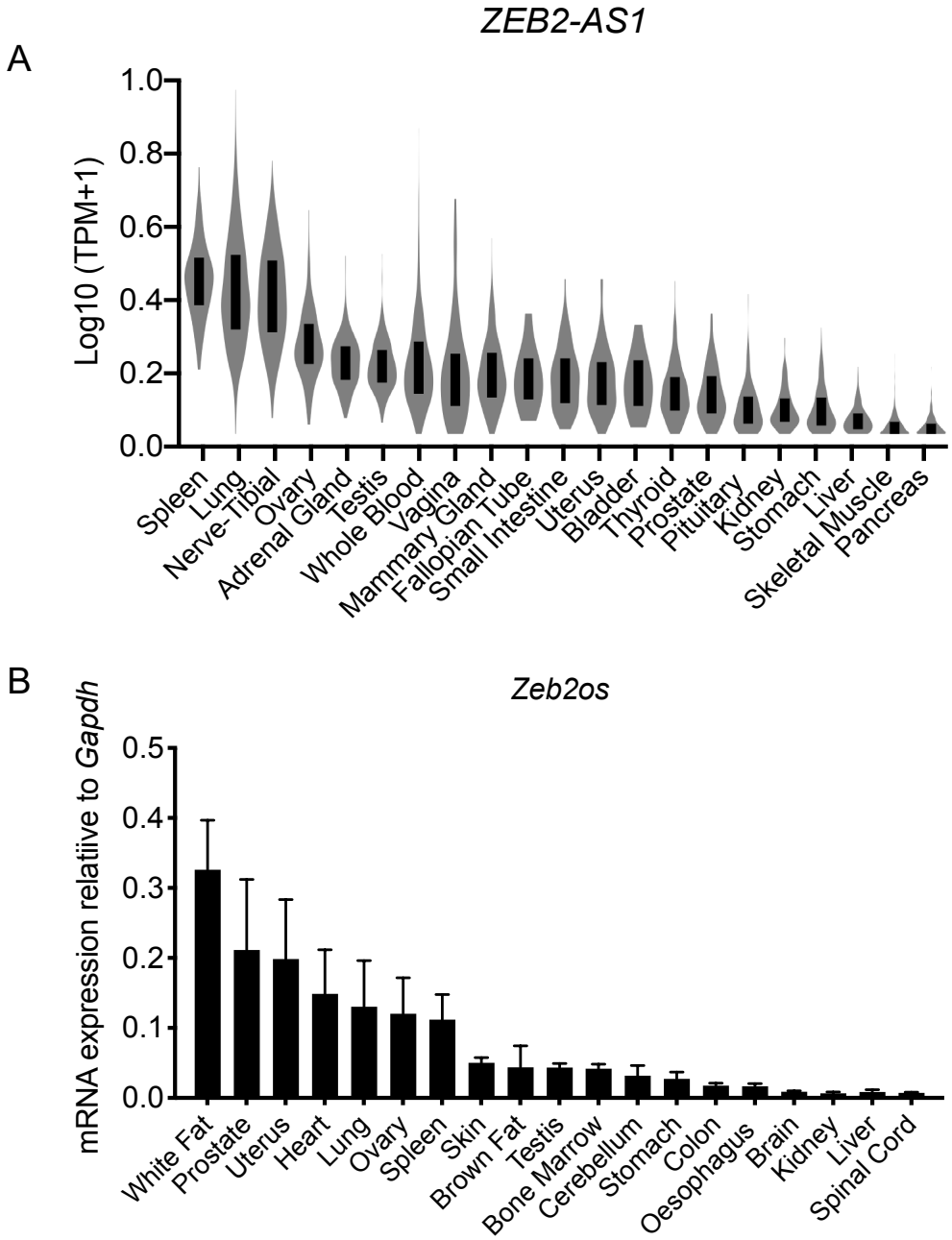


Figure 20. *ZEB2-AS1* and *Zeb2os* are expressed at low levels in several human and mouse tissues in basal conditions. (A) Violin plot of *ZEB2-AS1* expression in human tissues. Data extracted from Gtex database. (B) mRNA expression levels of *Zeb2os* in 8-12 weeks old mouse tissues analysed by qRT-PCR. Bars represent mean expression ± SEM of n=6 mice normalised to *Gapdh.*

We then analysed the expression of endogenous MIDORI microprotein by Western Blot using our custom-made polyclonal antibody in the same battery of mouse organs analysed by RT-qPCR. In this experiment, MIDORI KO organs were used as a negative control for the antibody. We only obtained clear results in kidney, where we observed a band that could correspond to MIDORI according to the molecular weight, whereas no signal was observed in MIDORI KO kidneys (Figure 21A). Of note, we obtained similar results using our polyclonal antibody in immunohistochemistry (IHC) (Figure 21B),

supporting the observation that endogenous MIDORI is expressed in kidney. Interestingly, MIDORI is expressed in the tubular epithelium at the interface between the kidney cortex and the medulla, showing a dotted pattern that is mostly cytoplasmatic.
Finally, we checked the level of *Zeb2os* expression and observed that WT and MIDORI KO kidneys show similar expression level of the lncRNA (Figure 21C), corroborating that our knock-out strategy does not affect *Zeb2os* expression but it does affect MIDORI translation.

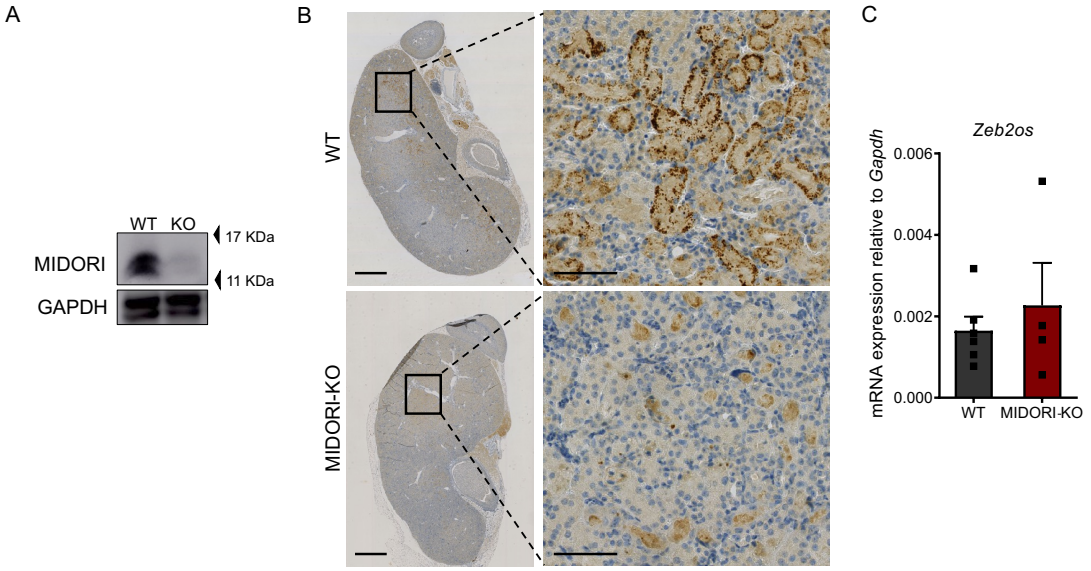


Figure 21. MIDORI is expressed in mouse kidney. (A) Western blot analysis showing MIDORI expression in WT and MIDORI KO mouse kidney. Protein expression was detected by immunoblot using an anti-MIDORI polyclonal custom-made antibody. (B) Representative immunohistochemistry images of MIDORI in WT and MIDORI KO kidneys derived from 10-week-old mice. MIDORI was stained using an anti-MIDORI polyclonal custom-made antibody. Two different magnifications of the kidney are shown: left pictures, scale bar=2mm; right pictures, scale bar=100μm. (C) mRNA expression levels of *Zeb2os* in kidneys isolated from 8-10 weeks old WT and MIDORI KO mice analysed by qRT-PCR. Each dot represents a single animal, bars represent mean expression ± SEM of n=4 mice normalised to *Gapdh.*

1.3.2. MIDORI is upregulated upon damage at the mRNA and protein level

It is known that the expression of some lncRNAs is regulated by different cues (i.e. under stress conditions) rather than in a tissue-specific manner (Valadkhan and Valencia-Hipolito, 2016). This is the case of *ZEB2-AS1*, whose expression is very low in basal conditions but it is known that certain stimuli such as TGFβ or the process of ageing upregulate the lncRNA both in human and in mouse (Beltran et al., 2008; Bernardes de Jesus et al., 2018) (Figure 12A and 13). Therefore, we sought to investigate the regulation of *ZEB2-AS1* and MIDORI protein expression by different stimuli. We treated MDA-MB-231 cells with the chemotherapeutic agent Doxorubicin (DOXO) to induce genotoxic stress and we observed an upregulation of the lncRNA, even if less pronounced than after TGFβ treatment (Figure 22A). Using our anti-MIDORI custom-made antibody, we then tested MIDORI protein levels in those samples and observed that MIDORI protein expression increases upon TGFβ and DOXO treatment (Figure

22B). As we do not have a human MIDORI KO cell line, we used a lysate from MIDORI-HA overexpressing cells as a positive control for the antibody. Of note, we tested other types of stress and found that also Tunicamycin, a drug known to induce endoplasmic reticulum (ER) stress (Guha et al., 2017), induces an increase in MIDORI protein levels in human cells (Figure 22C). We conclude that MIDORI is upregulated by genotoxic and ER stress.

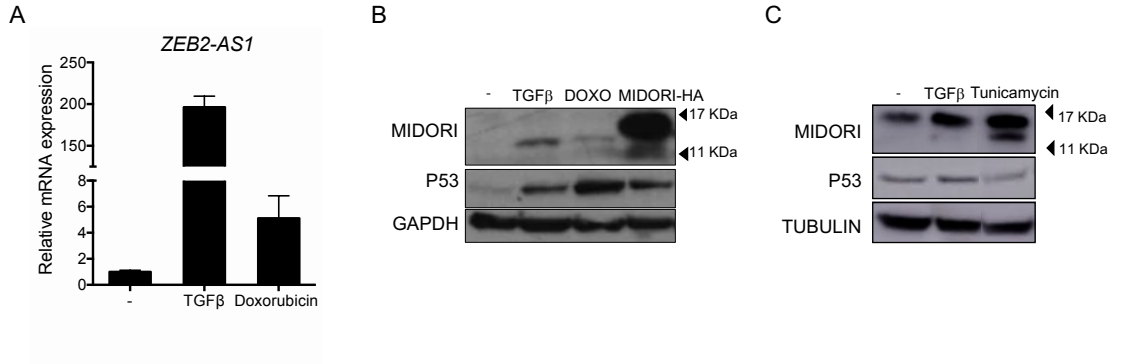


Figure 22. *ZEB2-AS1* and MIDORI are upregulated upon TGFβ, genotoxic and ER stress. (A) mRNA expression levels of *ZEB2-AS1* in MDA-MB-231 breast cancer cells after 48 hours of TGFβ treatment and 24 hours of DOXO treatment (1μM), analysed by qRT-PCR. Bars represent mean expression ± SD of three technical replicates normalised to *GAPDH* and relativised to no treated cells. (B) Western blot analysis showing MIDORI endogenous expression in MDA-MB-231 breast cancer cells after 48 hours of TGFβ treatment and 24 hours of DOXO treatment. Protein expression was detected by immunoblot using an anti-MIDORI polyclonal custom-made antibody. (C) Western blot analysis showing MIDORI endogenous expression in MDA-MB-231 breast cancer cells after 48 hours of TGFβ treatment and 24 hours of Tunicamycin treatment (1μM). Protein expression was detected by immunoblot using an anti-MIDORI polyclonal custom-made antibody.

1.3.3. MIDORI subcellular localisation

We performed IF experiments using our anti-MDORI custom-made antibody in MDA-MB-231 cells in basal conditions and upon TGFβ or DOXO treatment (Figure 23A). We observed that endogenous MIDORI is predominantly detected in the cytoplasm, where it shows a punctuated perinuclear pattern that in many cells acquired a spheroidal shape (Figure 23B). Consistent with previous results, treatment with TGFβ or with DOXO increases MIDORI protein expression. We performed co-staining experiments to determine the precise subcellular localisation but we did not observe colocalisation either with the Golgi apparatus or the lysosomes (Figure 24). We plan to perform stainings to test colocalization with ER or mitochondria.

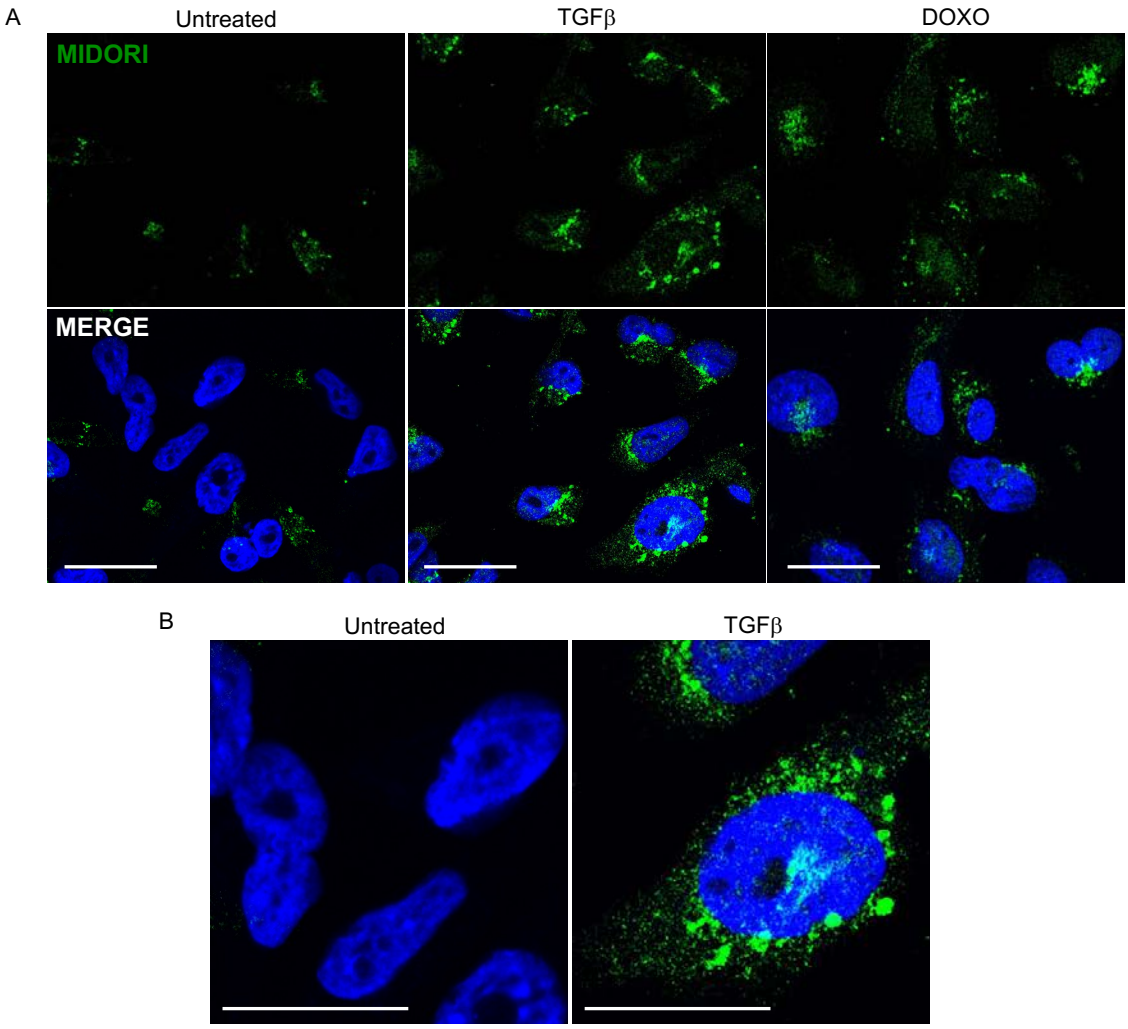


Figure 23. Endogenous MIDORI detection by IF. Representative immunofluorescence images of MIDORI (green) in untreated MDA-MB-231 cells and upon TGFβ or DOXO, scale bar=25μm. MIDORI was stained using an α-MIDORI polyclonal custom-made antibody. Nuclei were counterstained with DAPI. (B) Digital zoom showing untreated and TGFβ treated MDA-MB-231 cells stained using an α-MIDORI polyclonal custom-made antibody, scale bar=56μm. Nuclei were counterstained with DAPI.

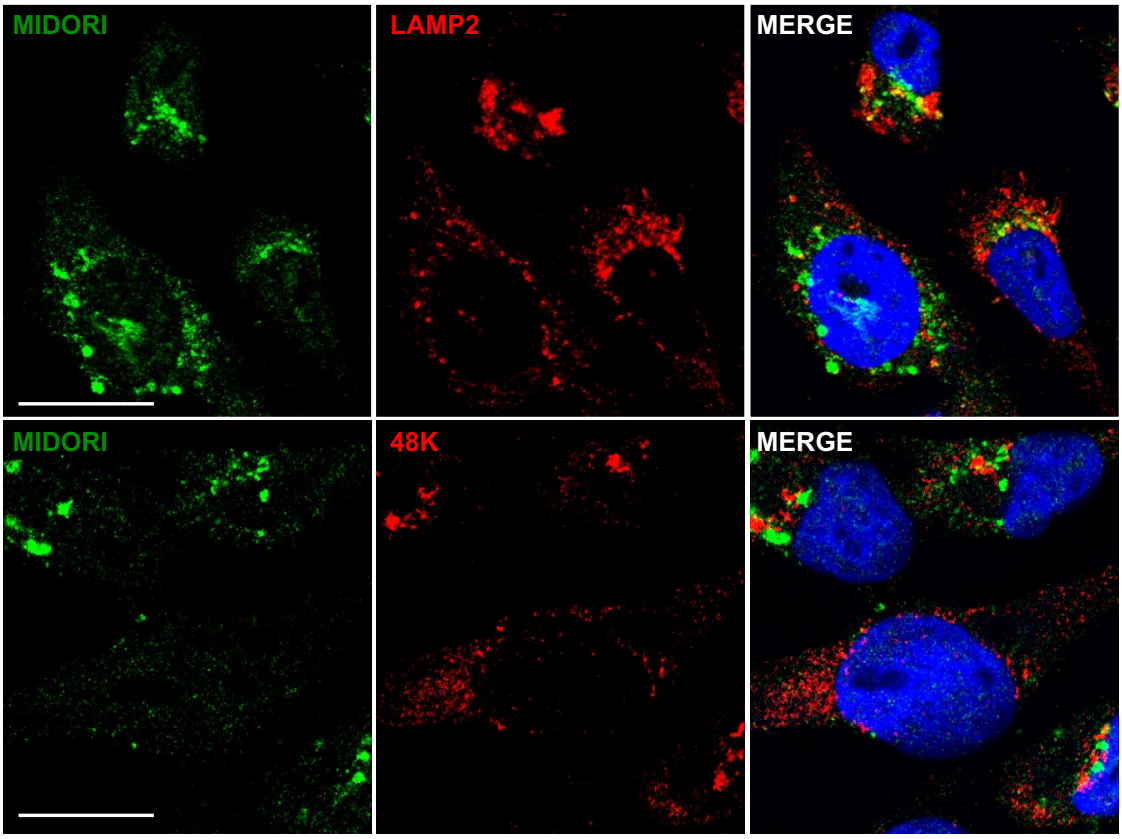


Figure 24. Endogenous MIDORI does not seem to colocalise with lysosomes or Golgi apparatus. Representative immunofluorescence images performed in TGFβ treated MDA-MB-231 cells, scale bar=40μm. MIDORI was stained using an α-MIDORI polyclonal custom-made antibody, LAMP2 lysosomal marker and 48K Golgi marker are shown in red. Nuclei were counterstained with DAPI.

Next, we performed subcellular fractionation followed by WB analysis with the anti-MIDORI antibody and observed that upon treatment with TGFβ or DOXO MIDORI is detected in the nucleus and not in the cytoplasm (Figure 25A), in line with the localisation prediction by DeepLoc 1.0 (Figure 15B). Given the apparent controversy with previous results, to validate the nuclear localisation we performed IF experiments washing out the cytoplasmatic signal by adding a pre-cytoplasmatic extraction step, avoiding this way diluting the nucleus signal with the noise coming from the cytoplasm. Using this approach, we did detect endogenous MIDORI in the nucleus (Figure 25B). Concluding this part, our results suggest that endogenous MIDORI localise both, in the nucleus and in the cytoplasm.

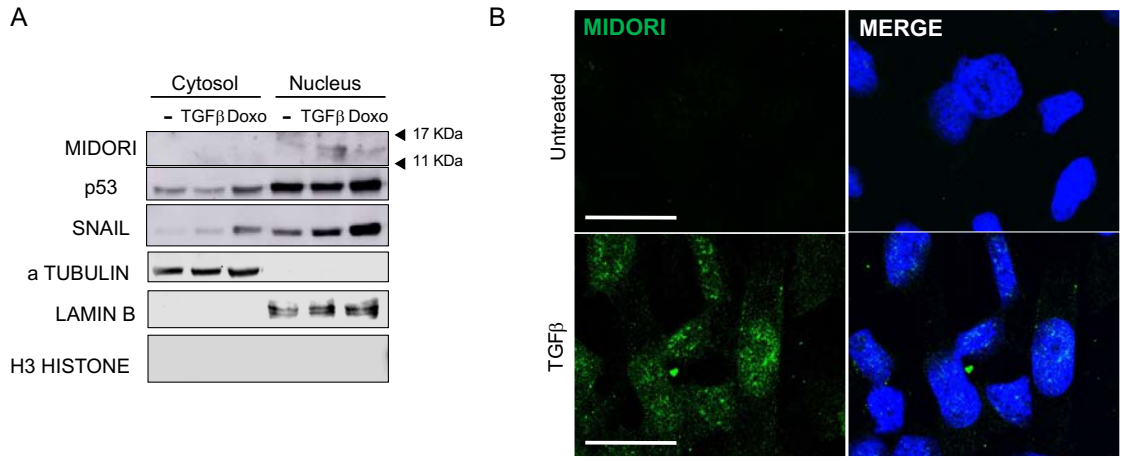


Figure 25. Endogenous MIDORI is detected in the nucleus. (A) Subcellular fractionation of cytosol, nucleus, and chromatin fractions in MDA-MB-231 cells. α-Tubulin, Lamin B and H3 histone are markers for cytoplasmatic, nuclear and chromatin fractions, respectively. MIDORI was detected using an α-MIDORI polyclonal custom-made antibody. (B) Representative immunofluorescence images after cytoplasmatic extraction of MIDORI (green) in control MDA-MB-231 cells and upon TGFβ, scale bar=25 μm. MIDORI was stained using an α-MIDORI polyclonal custom-made antibody. Nuclei were counterstained with DAPI.

2. MIDORI is a negative regulator of the EMT

2.1. MIDORI overexpression downregulates the mesenchymal programme

Since *ZEB2-AS1* is known to promote the EMT, we started the functional characterisation of MIDORI by analysing the impact of MIDORI overexpression on the EMT transcriptional programme in primary cells and established cell lines of mouse and human origin. To our surprise, we observed that MIDORI induces a downregulation of EMT-related genes, including EMT transcription factors (such as *SNAIL*, *SLUG* and *ZEB* proteins), mesenchymal markers (as *VIMENTIN* and *αSMA* in fibroblasts), ECM proteins (such as *COLLAGEN I* and *FIBRONECTIN)* and ECM remodelling factors as *MMP-9* (Figure 26). Of note, in MDA-MB-231 and NMuMG breast cell lines the downregulation of EMT markers is accompanied by an upregulation of the epithelial marker *E-CADHERIN*. Altogether these data suggest that MIDORI, contrary to the activator role of *ZEB2-AS1*, might act as a negative regulator of the EMT programme.

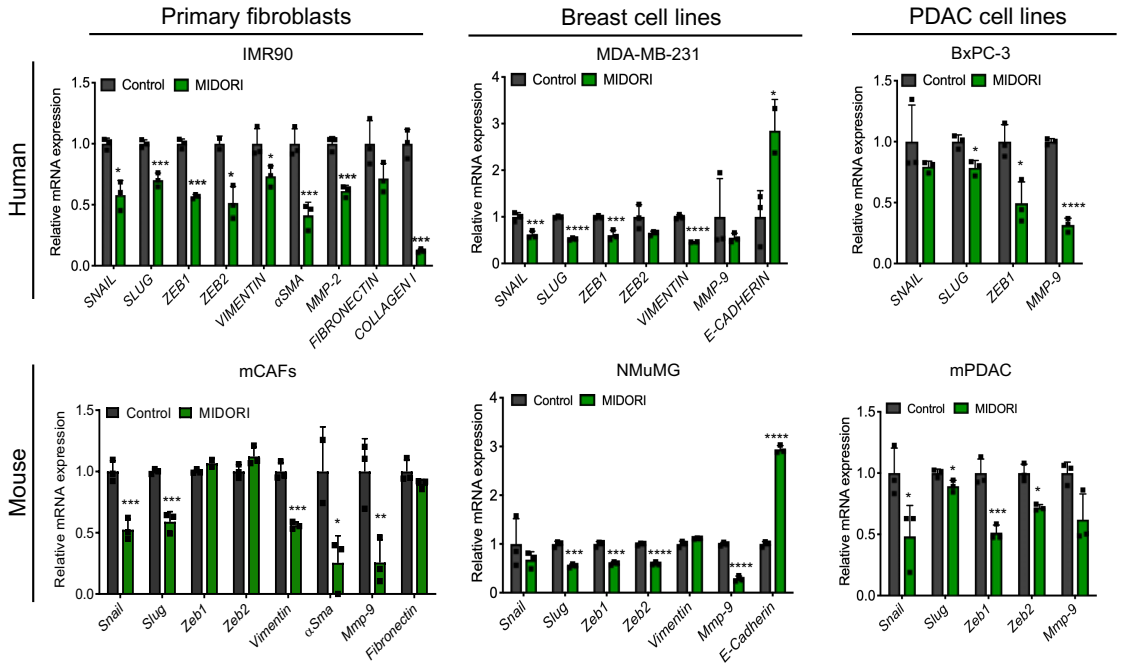


Figure 26. MIDORI downregulates EMT transcriptional programme in primary and cancer cell lines. mRNA expression analysis of the indicated genes in the indicated cell lines by RT-qPCR. mRNA expression was evaluated 4 days post induction with doxycycline in the case of cell lines transduced with the inducible construct (IMR90, MDA-MB-231 and NMuMG) or 4 days post transduction in the case of cell lines transduced with the constitutive construct (mCAFs, BxPC-3 and mPDAC). Values are relativised to *GAPDH* and normalised to cells expressing the empty vector as a control. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0005$ using Student's t-test for statistics. Bars represent mean expression ± SD of three technical replicates.

To further characterise MIDORI-induced phenotype we performed a set of *in vitro* assays in our MIDORI-overexpressing cancer cell lines. Firstly, we checked the proliferation rate in human breast and pancreatic cancer cells and we did not observe significant differences in the proliferation rate upon MIDORI overexpression (Figure 27), suggesting that MIDORI does not have a role in cell proliferation.

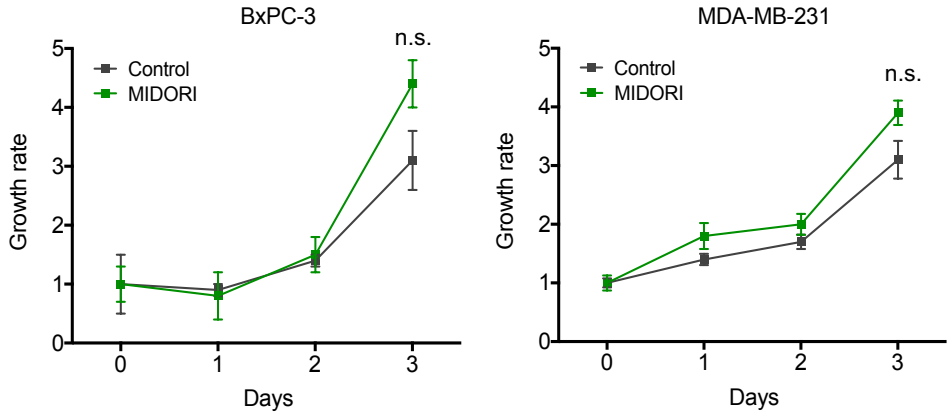


Figure 27. MIDORI does not affect cell proliferation *in vitro*. (A) Proliferation curve of the indicated cell lines. Proliferation curve analysis was started 4 days post-induction with doxycycline and it was refreshed every 48 hours during the experiment. Error bars represent mean expression ± SD of three technical replicates.

Then, we studied the migration and invasion capability of MIDORI-overexpressing cells by performing scratch and Matrigel-coated Boyden chamber assays, respectively. We observed that MIDORI-overexpressing cells are less migratory (Figure 28A) and significantly less invasive (Figure 28B-C) compared to control cells.

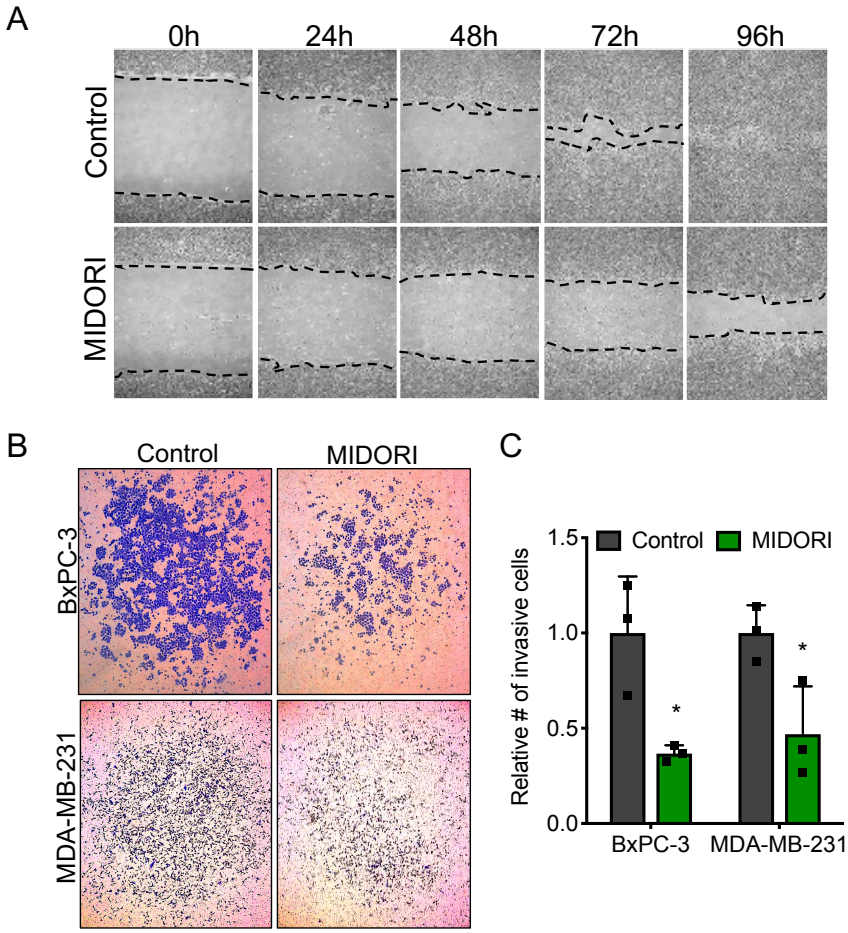


Figure 28. MIDORI reduces migration and invasion *in vitro*. (A) Cell migration assay of MIDORI-overexpressing BxPC-3 pancreatic cancer cells and control. Pictures were taken at 0 hours and different time points post scratch. Cell migration was assessed by recover of the scratch. Migration assay was started 4 days post-induction with doxycycline and it was refreshed every 48 hours during the assay. (B) Invasion of the indicated cell lines transduced with inducible MIDORI or empty vector across matrigel-covered transwells. Representative pictures of invading cells 24 hours after seeding. Invasion assay was performed 4 days doxycycline post-induction and it was refreshed every 48 hours during the assay. (C) Relative number of invading cells. Bars represent mean expression ± SD of three technical replicates relativised to invading cells in the empty vector condition. *$P < 0.05$, using Student's t-test for statistics.

Moreover, MIDORI overexpression is sufficient to significantly reduce the matrix degradation capacity of MDA-MB-231 cells, triple negative breast cancer cells known to be highly invasive (Kang et al., 2003; Minn et al., 2005) (Figure 29A-B).

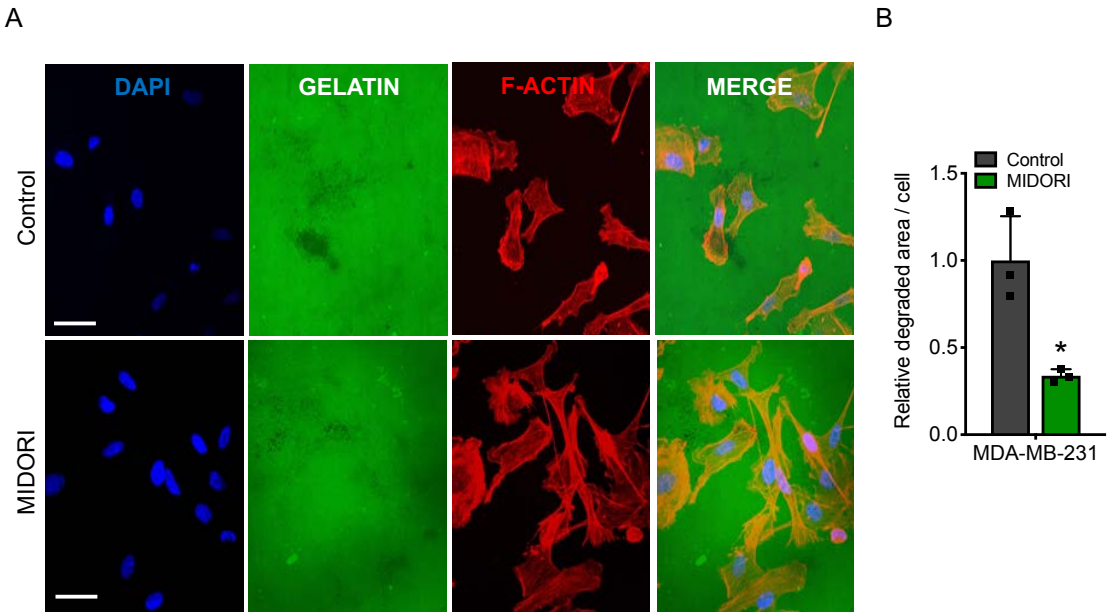


Figure 29. MIDORI reduces MDA-MB-231 matrix degradation capacity. (A) Representative pictures of matrix degradation assay performed with MIDORI-overexpressing MDA-MB-231 cells. After 4 days of doxycycline induction, cells were seeded on top of Collagen I and fluorescent gelatine coated coverslips and let degrade the matrix during 24 hours. Nuclei were counterstained with DAPI and cells were visualised by phalloidin staining that specifically binds to filamentous actin, scale bar=25μm. (B) Quantification of the degraded area per cell. Bars represent mean expression ± SD of three biological replicates relativised to the degraded area per cell in the empty vector condition. *$P < 0.05$, using Student's t-test for statistics.

In addition, MIDORI-overexpressing MDA-MB-231 cells display a different morphology compared to controls (Figure 30A). While control cells are spindle-shaped and they do not form clusters (regardless of the confluency), MIDORI-overexpressing cells look bigger, less refringent and they often form clusters. Therefore, we studied if MIDORI-overexpressing MDA-MB-231 were able to form cell-to-cell junctions. Since we saw that MIDORI induces an upregulation of *E-CADHERIN*, we firstly checked if we could see E-CADHERIN at the cell membrane by immunofluorescence, but we did not detect a membranous pattern (data not shown). We then analysed another component of adherent junctions, β-CATENIN (Figure 30B-C). In this case, we could see that while in control cells β-CATENIN signal is absent or diffuse in the cytoplasm, MIDORI-overexpressing MDA-MB-231 cells assemble β-CATENIN at cell membrane, suggesting that MIDORI overexpression can, at least partially, revert the mesenchymal phenotype of MDA-MB-231 cells. Altogether these data suggest that MIDORI impairs the mesenchymal programme at the transcriptional and functional level.

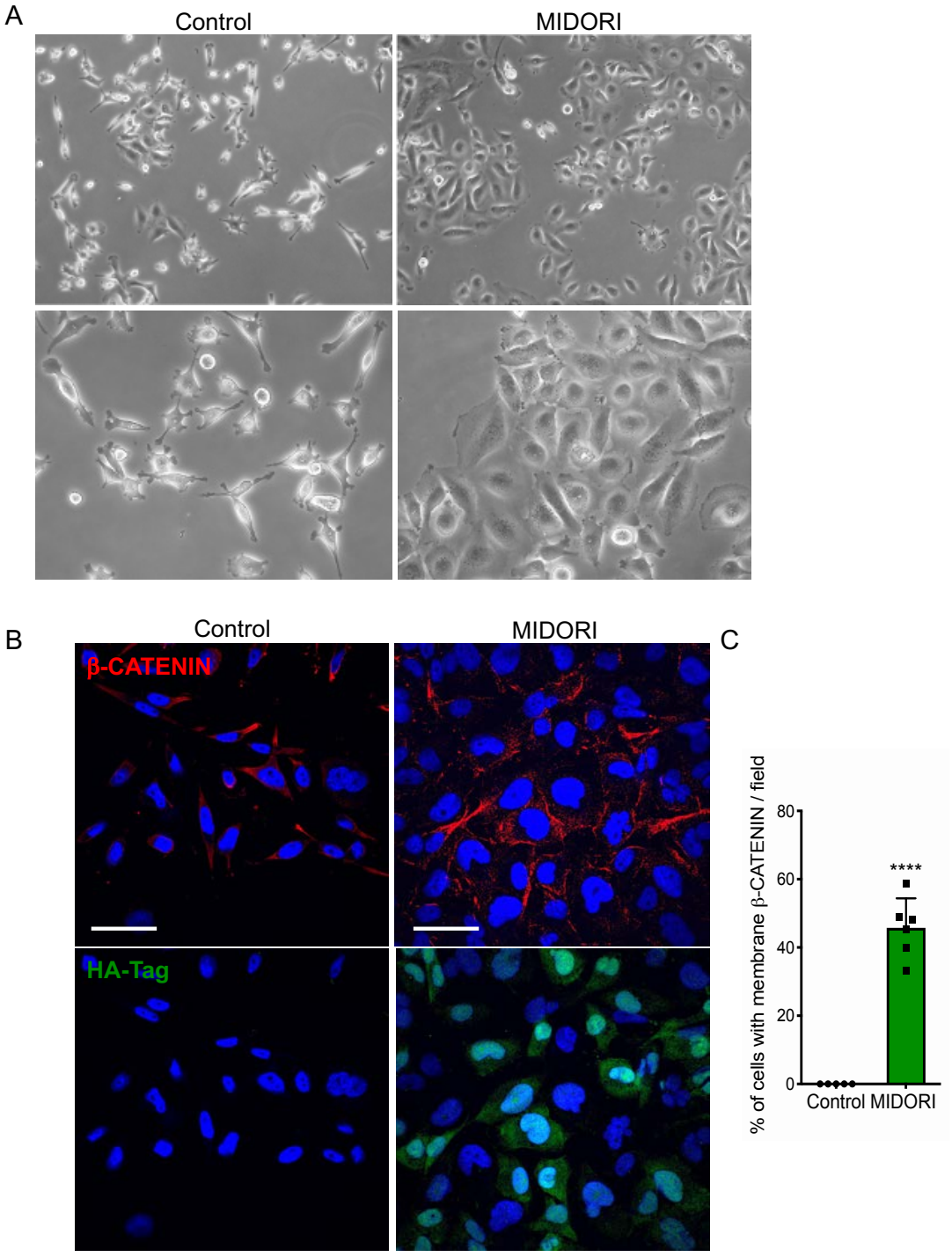


Figure 30. MIDORI overexpression induces morphological changes in MDA-MB-231 and β-CATENIN assembly on the cell membrane. (A) Representative bright-field microscopy images of MIDORI-overexpressing cells and control cells expressing the empty vector. Pictures were taken with 10X (upper) or 20X (bottom) magnification. (B) Representative immunofluorescence images of β-CATENIN (red) and HA-Tag (green) in control and MIDORI-overexpressing MDA-MB-231 cells 4 days after doxycycline induction, scale bar=25μm. Nuclei were counterstained with DAPI. (C) Percentage of cells with membranous β-CATENIN pattern per field. Bars represent mean ± SD in 5 different fields. At least 200 cells were counted per condition. ****$P < 0.0005$, using Student's t-test for statistics.

2.2. MIDORI impairs TGFβ-induced EMT *in vitro*

Since MIDORI overexpression decreases the mesenchymal properties in fibroblasts as well as in cancer cells lines with different degrees of mesenchymal traits, next we sought to determine if MIDORI overexpression is also capable of inhibiting the mesenchymal programme when cells are forced to undergo EMT. To do so, we treated MDA-MD-231 and NMuMG cells with TGFβ. Importantly, we saw that while control cells treated with TGFβ upregulate mesenchymal markers, MIDORI overexpression prevents or significantly reduces the transcriptional upregulation of EMT markers (Figure 31).

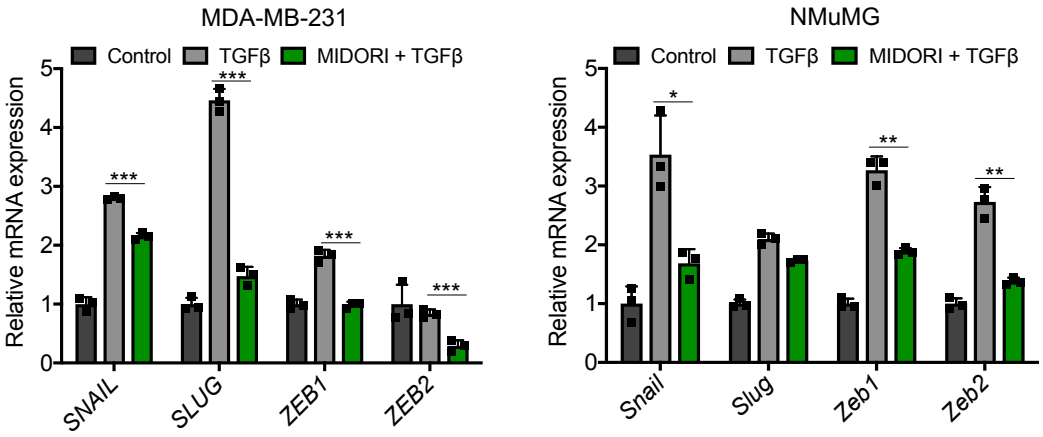


Figure 31. MIDORI overexpression impairs the upregulation of EMT markers upon TGFβ treatment. mRNA expression analysis of the indicated genes in the indicated cell lines by RT-qPCR. mRNA expression was evaluated 4 days post doxycycline induction (with the last two days (MDA-MB-231) or one day (NMuMG) adding or not TGFβ to the culture medium). Values are relativised to *GAPDH* and normalised to cells expressing the empty vector as a control. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 using Student's t-test for statistics. Bars represent mean expression ± SD of three technical replicates.

Next, we analysed the presence of adherent junctions in NMuMG as a readout of their mesenchymal state, given that this mouse epithelial cell line is known to undergo a robust EMT upon TGFβ treatment (Vincent et al., 2009). Remarkably, while in control cells the membranous β-CATENIN and E-CADHERIN staining is totally lost after 24 hours of TGFβ treatment, both proteins are mostly retained at cell membrane of MIDORI-overexpressing cells (Figure 32A-B-C), suggesting that MIDORI prevents them to undergo an EMT.

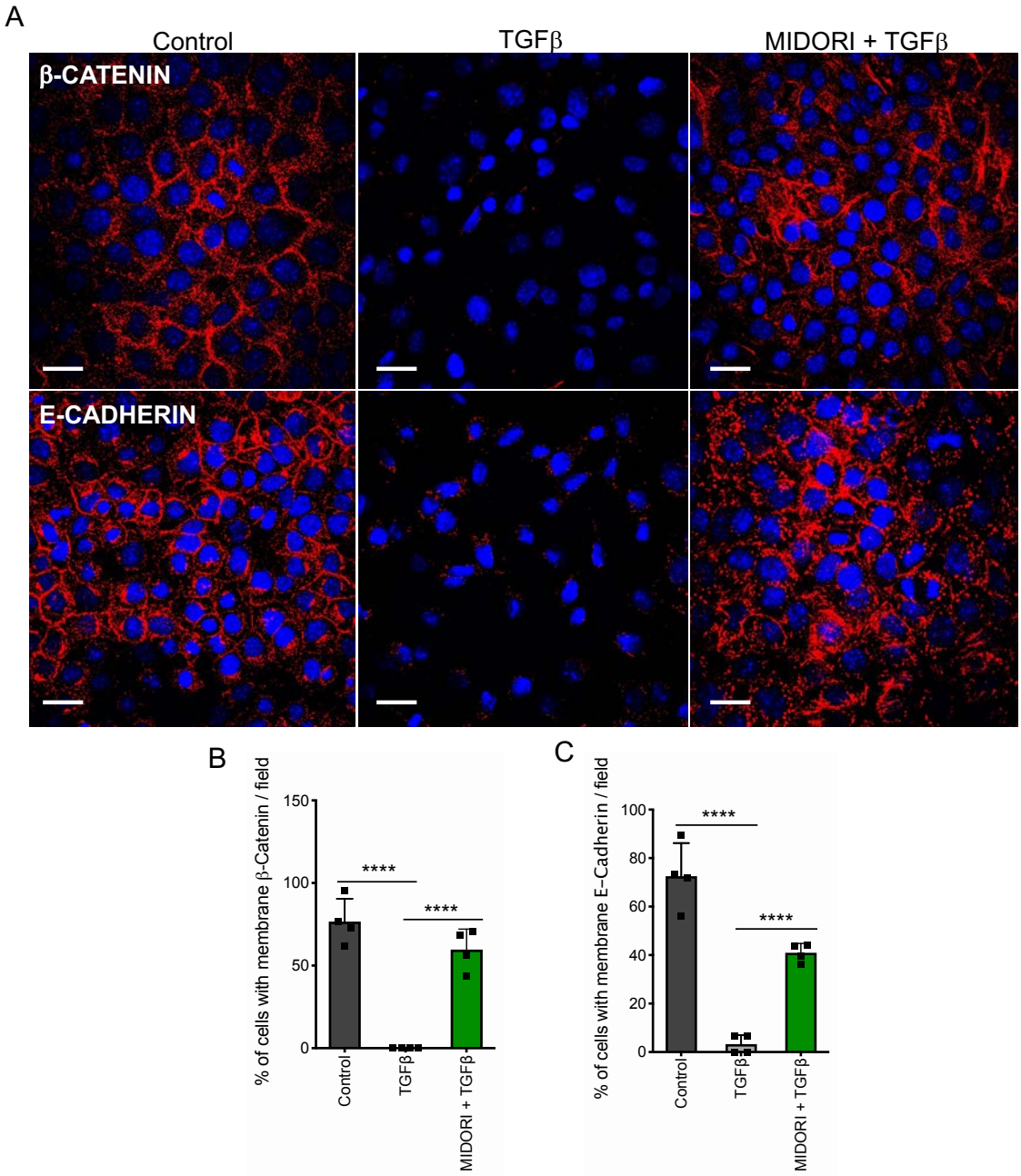


Figure 32. MIDORI-overexpressing NMuMG cells retain adherent junctions after TGFβ treatment. (A) Representative immunofluorescence images of β-CATENIN and E-CAHERIN stainings in control and MIDORI-overexpressing NMuMG cells (4 days post doxycycline induction) after 24 hours of TGFβ treatment, scale bar=25μm. Nuclei were counterstained with DAPI. (B) Percentage of cells presenting a membranous β-CATENIN pattern per field. Bars represent mean expression ± SD in four different fields. At least 300 cells were counted per condition. ****$P < 0.0005$, using Student's t-test for statistics. (C) Percentage of cells presenting a membranous E-CADHERIN pattern per field. Bars represent mean expression ± SD in four different fields. At least 300 cells were counted per condition. ****$P < 0.0005$, using Student's t-test for statistics.

In addition, by performing Western blot in these cells we observed that MIDORI overexpression induces an increase of E-CADHERIN protein expression, in basal conditions and upon TGFβ (Figure 33).

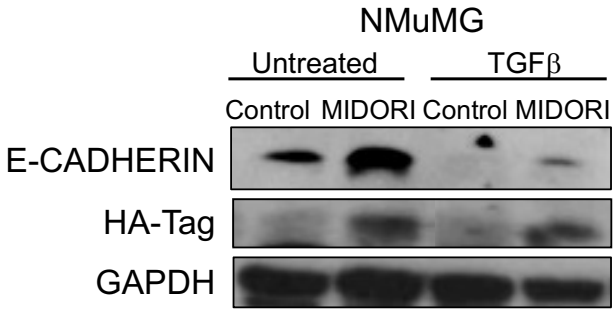


Figure 33. MIDORI overexpression increases the expression of E-CADHERIN in NMuMG cells undergone EMT. (A) Western blot analysis showing E-CADHERIN expression in MIDORI-overexpressing NMuMG cells or control (4 days post doxycycline induction) after 24 hours of TGFβ treatment.

At the functional level, MIDORI significantly decreases the invasion capability of both MDA-MB-231 and NMuMG cell lines when stimulated with TGFβ (Figure 34A-B), confirming that, at least *in vitro*, MIDORI overexpression retains epithelial identity even when cells are forced to acquire a mesenchymal phenotype.

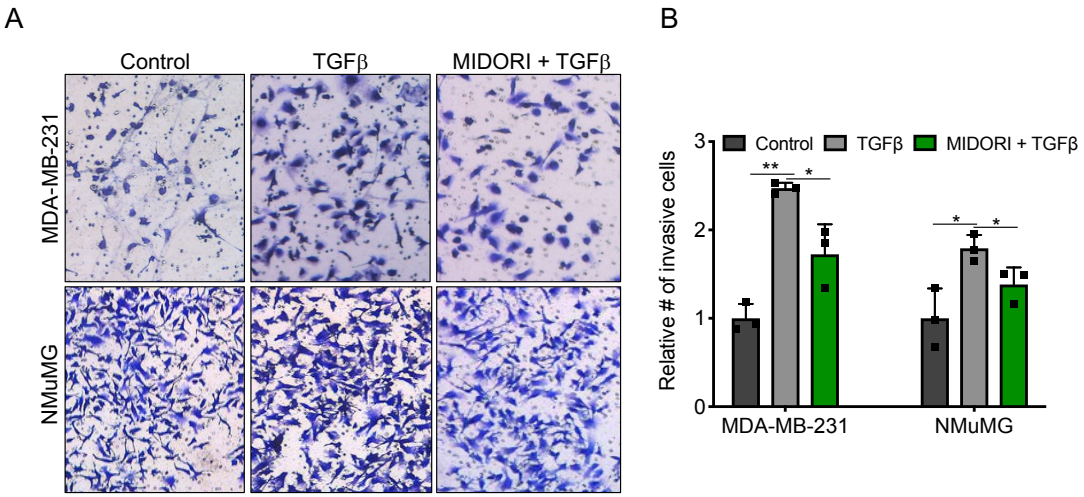


Figure 34. MIDORI overexpression decreases cell invasion in cells undergone EMT. (A) Invasion of the indicated cell lines transduced with MIDORI or empty vector across matrigel-covered transwells. Invasion assay was performed 4 days post doxycycline induction, and 2 days (MDA-MB-231) or 1 day (NMuMG) after adding or not TGFβ to the culture medium. Representative pictures of invading cells 24 hours after seeding. (C) Relative number of invading cells. Bars represent mean expression ± SD of three technical replicates relativised to invading cells in the empty vector condition. *$P < 0.05$, **$P < 0.01$ using Student's t-test for statistics.

2.3. MIDORI impairs the proinflammatory secretome

Another feature of mesenchymal cells is their secretory phenotype, which include the secretion of ECM remodelling factors, cytokines and chemokines. These factors have a paracrine effect on the neighbour cells, that in the context of cancer can make them more invasive (Scheel et al., 2011), escape from the immune system (Kudo-Saito et al., 2009) or even help them to prepare the metastatic niche (Pein et al., 2020). By qRT-PCR analyses we observed that MIDORI overexpression decreases the expression of proinflammatory cytokines and chemokines in basal conditions and when treating the cells with TGFβ (Figure 35A). To corroborate the role of MIDORI inhibiting the pro-inflammatory secretome, we tested the effect of MIDORI overexpression on senescent cells, a paradigmatic example of secretory and pro-inflammatory cells. We treated the MDA-MB-231 cells with the CDK4/6 inhibitor Palbociclib, a known senescence inducer (Dickson, 2014), and observed that senescent cells upregulate several cytokines as part of the SASP. Remarkably, MIDORI overexpression impairs the upregulation of the SASP, with the exception of *IL-6* that in fact is increased in MIDORI-overexpressing cells (Figure 35B). We then looked at the activation of NF-κB and mTOR pathway (measuring activation of S6), both known to control the expression of the SASP (Salminen et al., 2012) (Herranz et al., 2015). While we did not observe any change in the phosphorylation status of S6, we did observe a decreased phosphorylation of NF-κB, meaning less activation, in senescent cells overexpressing MIDORI (Figure 35C). Thus, MIDORI impairs the expression of proinflammatory cytokines and chemokines possibly by impairing the activation of NF-κB. Importantly, NF-κB is known to be activated as an alternative branch of TGFβ pathway (Fernandez-Gonzalez, 2021; Freudlsperger et al., 2013).

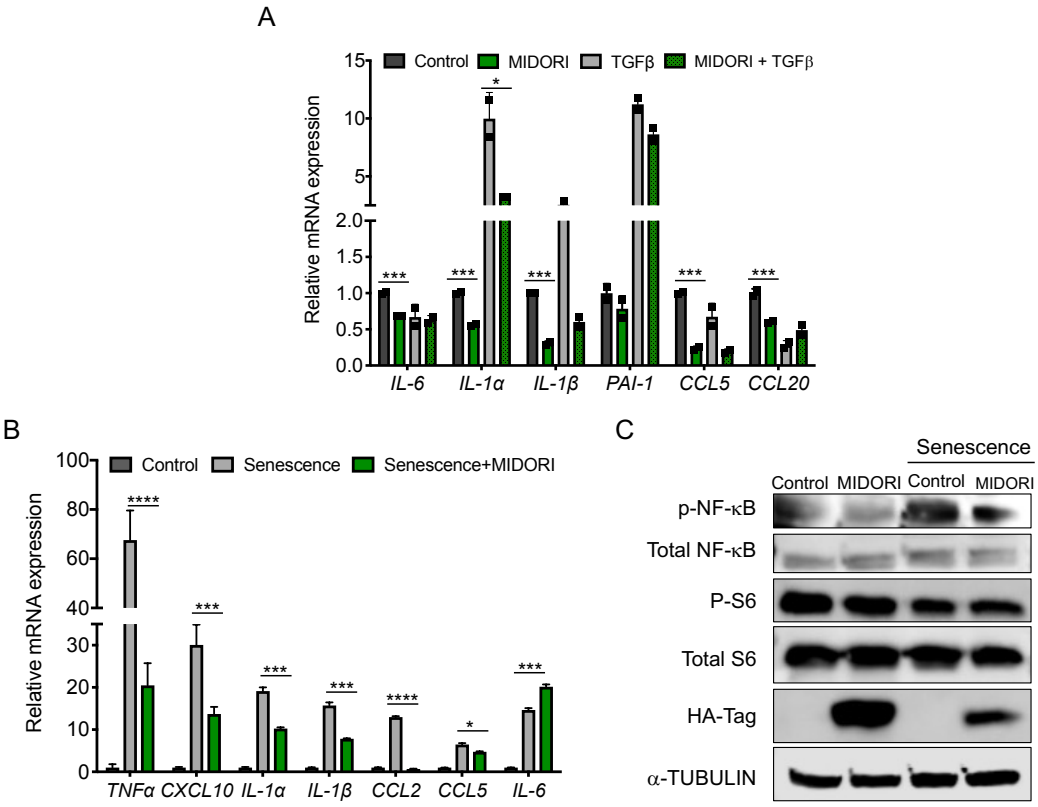


Figure 35. MIDORI overexpression decreases the expression of proinflammatory cytokines and chemokines. (A) mRNA expression analysis of the indicated genes in MDA-MB-231 cell line treated with TGFβ during 48 hours after 4 days of doxycycline induction by RT-qPCR. Values are relativised to *GAPDH* and normalised to cells expressing the empty vector as a control. **P* < 0.05, ****P* < 0.001 using Student's t-test for statistics. Bars represent mean expression ± SD of three technical replicates. (B) mRNA expression analysis by RT-qPCR of the indicated genes in MDA-MB-231 cells overexpressing control vector or MIDORI, and after inducing senescence with Palbociclib. Cells were treated 1 week with Palbociclib to induce senescence and, afterwards, MIDORI was induced during 4 days by adding doxycycline to the culture medium. Values are relativised to GAPDH and normalised to cells expressing the empty vector as a control. **P* < 0.05, ****P* < 0.001 using Student's t-test for statistics. Bars represent mean expression ± SD of three technical replicates. (C) Western blot analysis against the indicated proteins in MDA-MB-231 cells after the induction of senescence by Palbociclib. Cells were treated as in (B).

We then wanted to test if MIDORI overexpression has an impact in a paracrine way. To test this hypothesis, we took the conditioned medium (CM) from control and MIDORI-overexpressing MDA-MB-231 cells treated or not with TGFβ and we used it as attractant for MDA-MB-231 in Boyden chamber assay (Figure 36A). As expected, the CM of TGFβ–treated cells induced the invasion of recipient cells, but MIDORI overexpression did not have a significant effect (Figure 36B-C) suggesting that, at least in MDA-MB-231 cells, MIDORI does not seem to impact cell invasion in a cell extrinsic manner. However, when we performed the invasion assay seeding at the bottom of the chamber competent secretory cells, as cancer associated fibroblasts, over mouse pancreatic cancer cells (Figure 36D), we saw that the CM of MIDORI-overexpressing mCAFs decreases the invasion of mPDAC cells (Figure 36E-F). Thus, MIDORI decreases the pro-invasive activity of CAFs, probably by modulating their secretome.

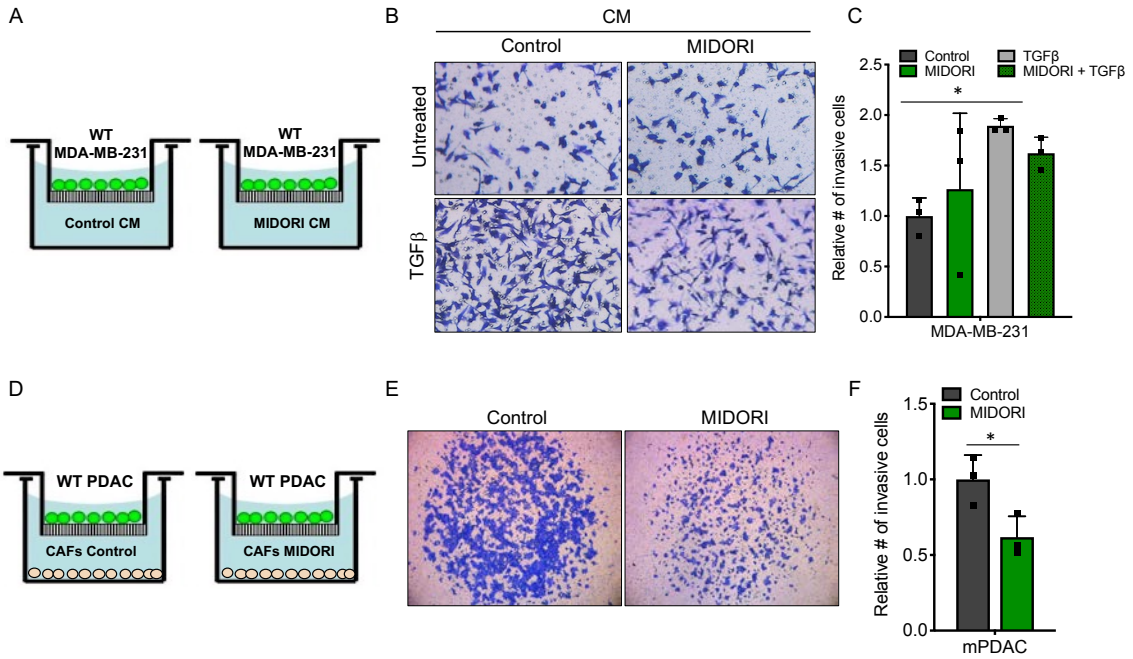


Figure 36. MIDORI-overexpressing mCAFs CM decreases cancer cell invasion. (A) Scheme of invasion assay performed with MDA-MB-231 overexpressing MIDORI or the empty vector CM. The CM was generated incubating the culture medium during 3 days after inducing MIDORI expression 4 days by doxycycline induction, while WT MDA-MB-231 cells were seeded in the top of the matrigel-covered transwell. (B) Invasion of the WT MDA-MB-231 cells exposed to CM of MDA-MB-231 cells transduced with MIDORI or empty vector across matrigel-covered transwells.

Representative pictures of invading cells 24 hours after seeding. (C) Relative number of invading cells. Bars represent mean expression ± SD of three technical replicates relativised to invading cells in the empty vector condition. *$P < 0.05$, using Student's t-test for statistics. (D) Scheme of invasion assay performed with mCAFs overexpressing MIDORI or the empty vector as a control. CAFs constitutively expressing MIDORI construct or the empty vector as a control were seeded on the bottom compartment of the Boyden chamber, while WT PDAC cells were seeded in the top of the matrigel-covered transwell. (E) Invasion of the WT mPDAC cells exposed to CM of mCAFs transduced with MIDORI or empty vector across matrigel-covered transwells. Representative pictures of invading cells 24 hours after seeding. (F) Relative number of invading cells. Bars represent mean expression ± SD of three technical replicates relativised to invading cells in the empty vector condition. *$P < 0.05$, using Student's t-test for statistics.

2.4. MIDORI is a negative regulator of TGFβ pathway

Our previous results showed that MIDORI impairs the activation of NFκB, which is known to be induced by TGFβ (Fernandez-Gonzalez, 2021; Freudlsperger et al., 2013) . To expand this analysis to other TGFβ signalling effectors, we performed Western blot in both MDA-MB-231 and NMuMG cells overexpressing MIDORI and treated or not with TGFβ. We observed that MIDORI reduces the expression of SNAIL in basal conditions and upon TGFβ treatment, confirming that the decrease observed at the mRNA level (Figure 26 and 31) is also accompanied by a decrease in the protein levels. Of note, we also observed that MIDORI overexpression decreases the phosphorylation of SMAD2 and ERK1/2 (Figure 37A). Together with the result showing the effect on NF-κB (Figure 35C), these data indicate that MIDORI might actually act as a negative regulator of TGFβ pathway upstream of *ZEB2-AS1* (Figure 37B).

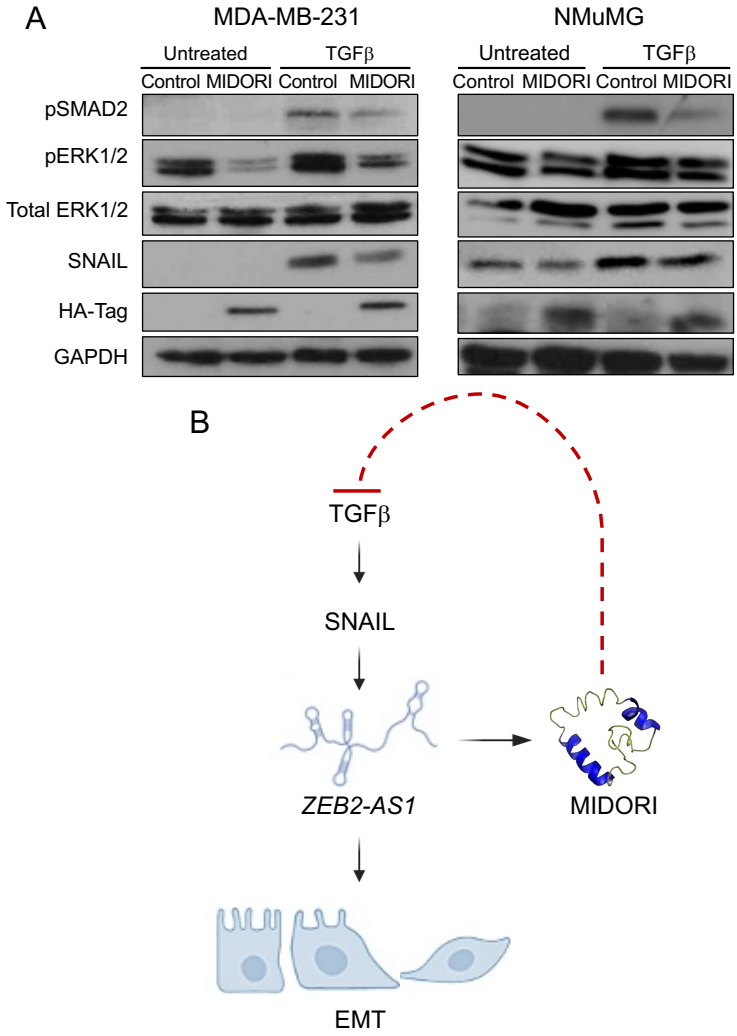


Figure 37. MIDORI overexpression decreases the activation of TGFβ pathway effectors. (A) Western blot analysis showing the indicated proteins in MDA-MB-231 and NMuMG cells overexpressing MIDORI or the empty vector as a control (4 days post doxycycline induction) after 48 and 24 hours of TGFβ treatment, respectively. (B) Scheme showing how MIDORI could act upstream of *ZEB2-AS1* as a negative regulator of TGFβ pathway.

2.4. Short MIDORI phenocopies long MIDORI

As mentioned in section 1.2.1., although we simplified the reading of this thesis showing first the results obtained with long MIDORI, we have performed all the gain-of-function experiments with the two murine MIDORI isoforms. Here we show the results of the most informative experiments performed with short MIDORI.

As in the case of the long isoform, we observed that short MIDORI downregulates the expression of EMT transcription factors, ECM proteins and remodelling factors both, in primary and cancer cell lines (Figure 38A). The transcriptional downregulation of the EMT programme is also accompanied by a decreased invasion capability (Figure 38B-C), suggesting that short MIDORI also act as a negative regulator of mesenchymal features.

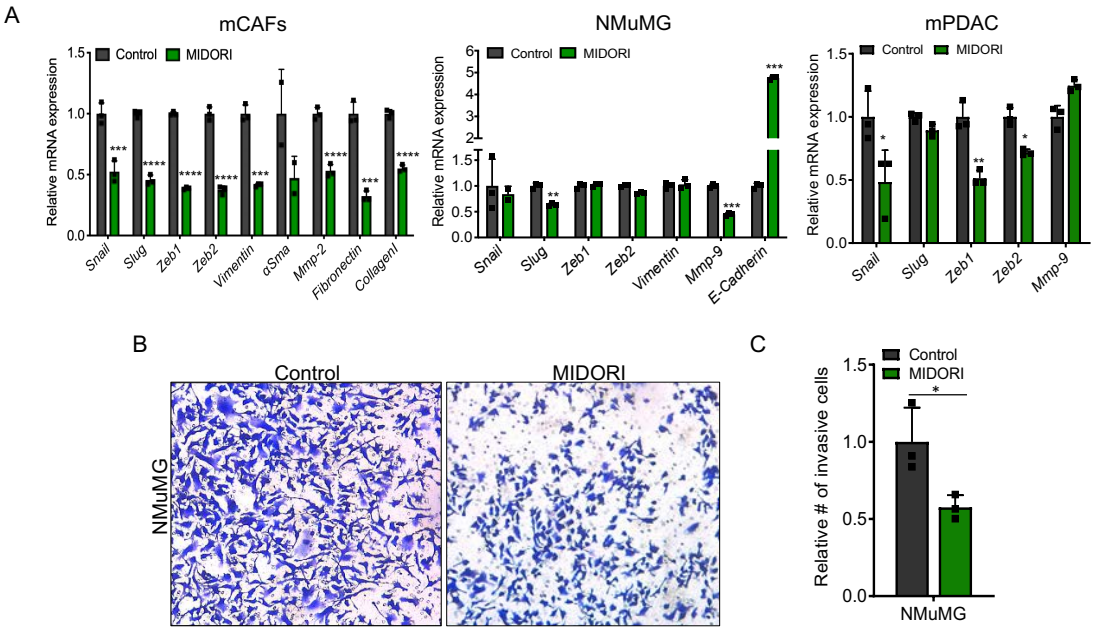


Figure 38. Short MIDORI overexpression downregulates EMT markers and decreases cell invasion. (A) mRNA expression analysis of the indicated genes in the indicated cell lines by RT-qPCR. mRNA expression was evaluated 4 days post induction with doxycycline in the case of cell lines transduced with the inducible construct (NMuMG) or 4 days post transduction in the case of cell lines transduced with the constitutive construct (mCAFs, and mPDAC). Values are relativised to *Gapdh* and normalised to cells expressing the empty vector as a control. ***$P < 0.001$, ****$P < 0.0005$ using Student's t-test for statistics. Bars represent mean expression ± SD of three technical replicates. (B) Invasion of the NMuMG transduced with short MIDORI or empty vector across matrigel-covered transwells after 4 days of doxycycline induction. Representative pictures of invading cells 24 hours after seeding. (D) Relative number of invading cells. Bars represent mean expression ± SD of three technical replicates relativised to invading cells in the empty vector condition. *$P < 0.05$, using Student's t-test for statistics.

We next wanted to address the effect of short MIDORI when forcing cells to undergo EMT. We observed that short MIDORI overexpression prevents *Zeb1* and *Zeb2* upregulation upon TGFβ treatment (Figure 29A), recapitulating in part long MIDORI phenotype. We then looked at TGFβ pathway effectors and observed a decreased phosphorylation of SMAD2 upon TGFβ, but no significant difference in the case of ERK1/2 (Figure 39B). Although the decrease in the EMT markers and TGFβ signalling effectors activation is not fully recapitulated, we did see a significant decrease in the invasion capability of short MIDORI-overexpressing cells upon TGFβ treatment (Figure 39C-D), suggesting that short MIDORI overexpression can functionally mimic long MIDORI.

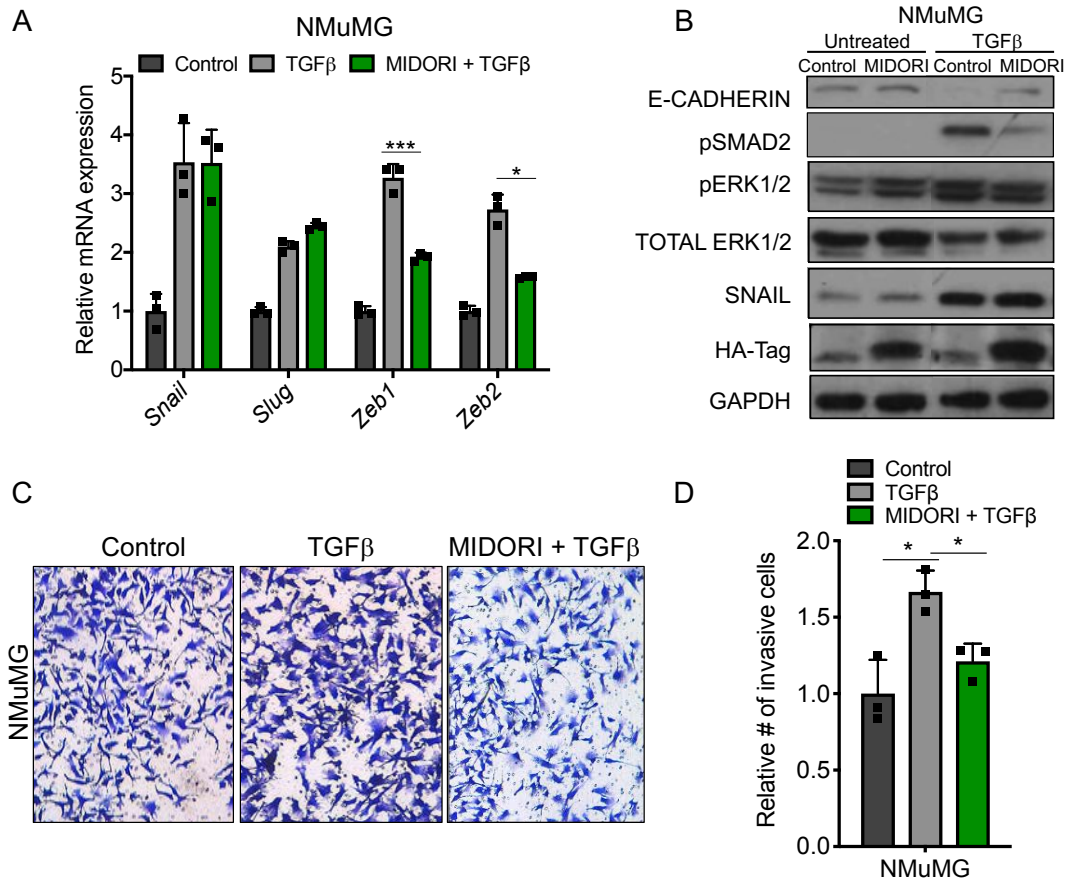


Figure 39. Short MIDORI overexpression decreases the activation of TGFβ pathway effectors and decreases cell invasion in cells undergone EMT. (A) mRNA expression analysis of the indicated genes in the indicated cell line by RT-qPCR. mRNA expression was evaluated 4 days post doxycycline induction (with the last day adding or not TGFβ to the culture medium). Values are relativised to *Gapdh* and normalised to cells expressing the empty vector as a control. *$P < 0.05$, ***$P < 0.001$ using Student's t-test for statistics. Bars represent mean expression ± SD of three technical replicates. (B) Western blot analysis showing the indicated proteins NMuMG cells after 4 days of doxycycline induction and 24 hours of TGFβ treatment. (C) Invasion of the indicated cell lines expressing short MIDORI or empty vector (4 days of doxycycline induction) across matrigel-covered transwells. Representative pictures of invading cells 24 hours after seeding. (D) Relative number of invading cells. Bars represent mean expression ± SD of three technical replicates relativised to invading cells in the empty vector condition. *$P < 0.05$, using Student's t-test for statistics.

Of note, short MIDORI overexpression in NMuMG cells also induced an increase in E-CADHERIN expression (Figure 39B), and prevented the loss of adherent junctions upon treatment with TGFβ (Figure 40 A-C), further suggesting a similar role of both isoforms.

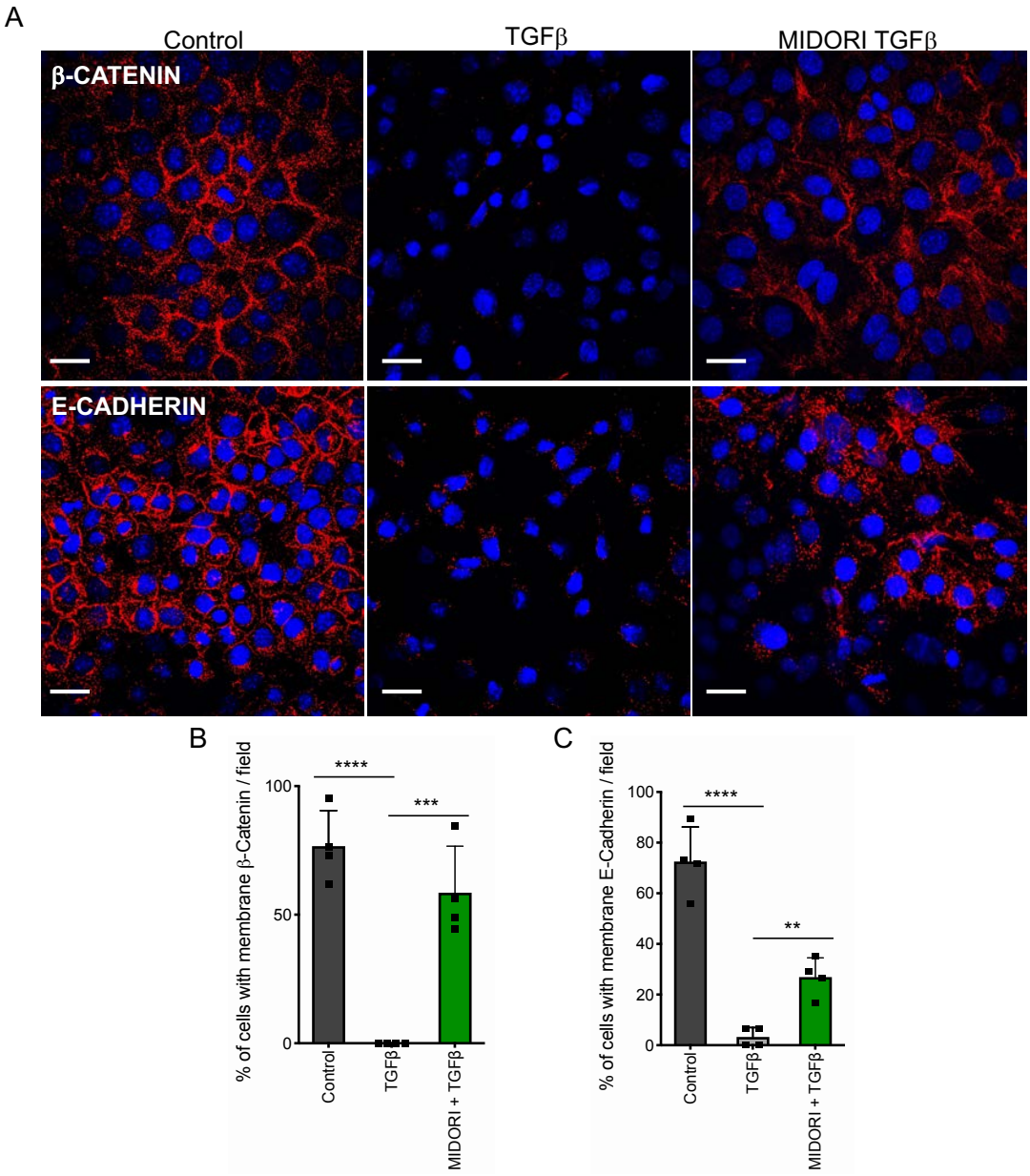


Figure 40. Short MIDORI-overexpressing NMuMG retain β-CATENIN and E-CADHERIN at cell membrane after TGFβ treatment. (A) Representative immunofluorescence images of β-CATENIN and E-CAHERIN (red) in control and short MIDORI-overexpressing NMuMG cells after 4 days of doxycycline induction and 24 hours of TGFβ treatment, scale bar=25μm. Nuclei were counterstained with DAPI. (B) Percentage of cells with β-CATENIN staining per field. Bars represent mean expression ± SD of the cell counting of four different pictures. At least 300 cells were counted per condition. ***$P < 0.001$, ****$P < 0.0005$, using Student's t-test for statistics. (C) Percentage of cells with E-CADHERIN staining per field. Bars represent mean expression ± SD of the cell counting of four different pictures. At least 300 cells were counted per condition. **$P < 0.01$, ****$P < 0.0005$, using Student's t-test for statistics.

Finally, we also tested if short MIDORI has a decreased pro-invasive cell-extrinsic capacity. As observed for long MIDORI, we saw that the CM of short MIDORI-overexpressing mCAFs decreases the invasion capability of WT mPDAC cells (Figure 41A-B).

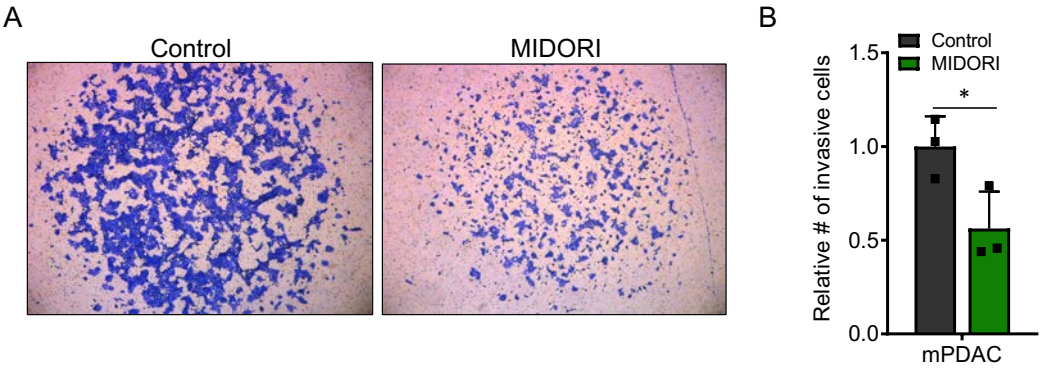


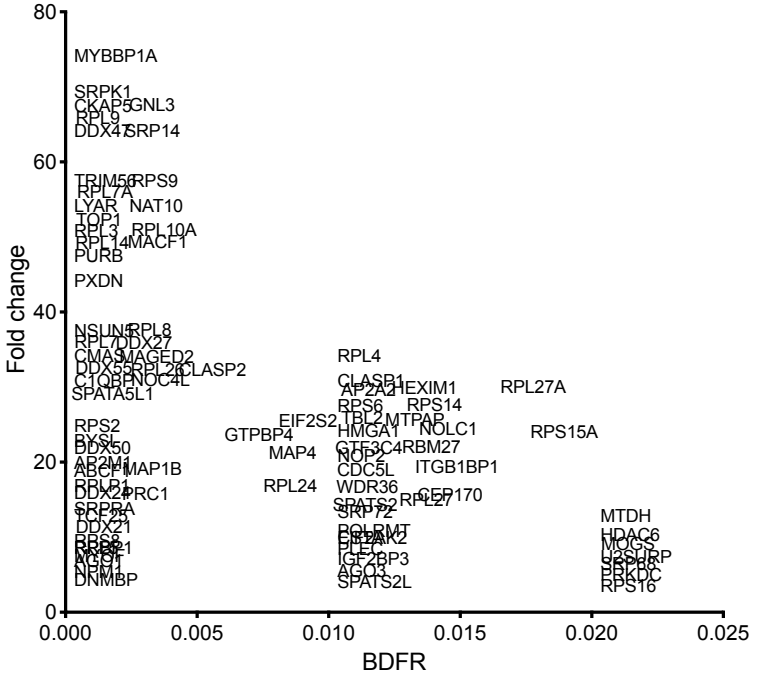
Figure 41. Short MIDORI-overexpressing mCAFs CM decreases cancer cell invasion. (A) Invasion of the WT mPDAC cells exposed to CM of mCAFs constitutively expressing short MIDORI or empty vector across matrigel-covered transwells. Representative pictures of invading cells 24 hours after seeding. (B) Relative number of invading cells. Bars represent mean expression ± SD of three technical replicates relativised to invading cells in the empty vector condition. *$P < 0.05$, using Student's t-test for statistics.

Concluding, even if the downregulation of EMT markers is not totally recapitulated by the overexpression of short MIDORI, all the functional analyses let us conclude that the two microproteins phenotypically behave in the same way regarding the regulation of mesenchymal properties.

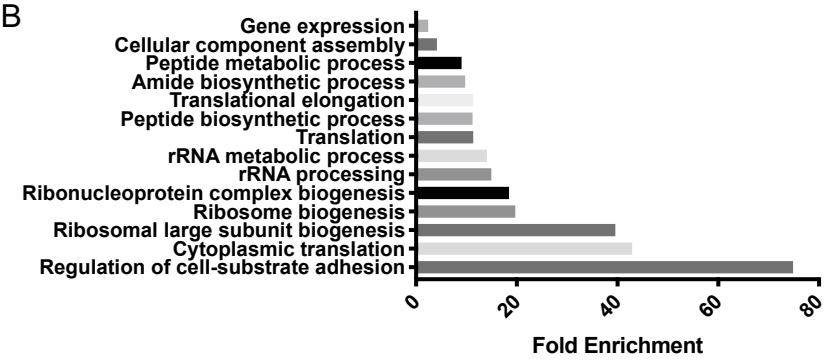
2.5. Analysis of MIDORI interactome

In order to understand the molecular mechanisms behind MIDORI's function, we studied MIDORI's interactome. We overexpressed MIDORI in MDA-MB-231 cells and performed MIDORI immunoprecipitation (using an anti-HA antibody) followed by mass-spectrometry. This analysis revealed several interactor candidates. We ranked the candidates according to their fold change detection in the MIDORI-overexpressing condition (FC) and the Bayesian false discovery rate (BDFR) obtained from the spectra (Fig 42A). There were 89 of candidates with a FC >3 and an BDFR < 0.02. Surprisingly, we observed that the vast majority of MIDORI's interactor candidates are proteins whose function is related with RNA translation and ribosome biogenesis (Figure 42A). We performed Gene Ontology analysis (GO analysis) of MIDORI's interactome and, in fact, we found an enrichment in GO terms related with rRNA processing, ribosome biogenesis and translation, as well as with the ribosomal or nucleolar subcellular components (Figure 42C-D). Interestingly, the most enriched term was "Regulation of cell-substrate adhesion", which is consistent with the role of MIDORI inhibiting migration, invasion and matrix degradation.

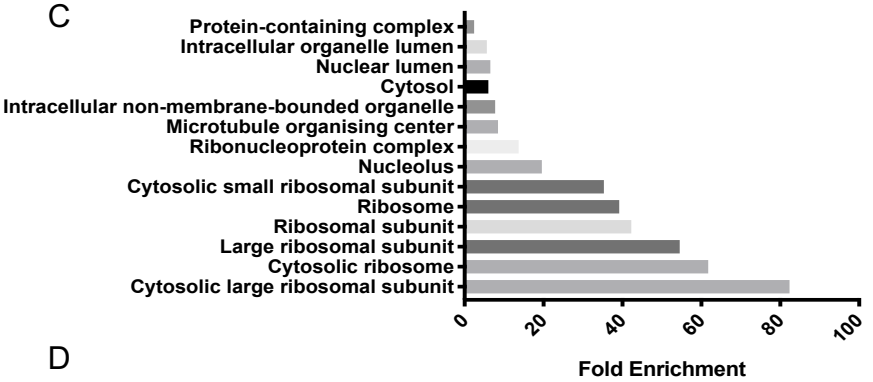
A



B



C



D

Cellular compartment	Interactor candidates
Ribosome	RPL3, RPS15A, RPS14,RPS9, RLP24, RPS6, RPL7, RPL7A, RPL8, RPL4, RPL27, RPL27A, RPL10A, RPL26, RPL14, RPL9
Ribonucleoprotein complex	SRP14, HEXIM1, NAT10, CDC5L, NOLC1, NOC4L, WRD36, AP2A2, CEP170, HMGA1, MAP4, EIF2S2
Nucleolus	RPL3, TOP1, RPS14, RPS9, RPS6, RPL7, GTPBP4, RPL7A, NAT10, RPL4, MYBBP1A, DDX47, CKAP5, NSUN5, DDX27, DDX55, LYAR, NOP2, NOLC1, RPL27, RPL26, GNL3, NOC4L, WDR36, MAGED2
Nucleoplasm	TOP1, RPS15A, TRIM56, CMAS, RPS14, RPS9, RPS6, GTPBP4, HEXIM1, NAT10, HMGA1, DDX47, PURB, MYBBPYA, NSUN5, DDX55, LIAR, CDC5L, NOP2, NOLC1, GTF3C4, RPL27, RPL26, GNL3, MTPAP, NOC4L, SRPK1, RBM27
Focal adhesion	CLASP1, CLASP2, MACF1, MAP4, DNMBP, CKAP5, ITGB1BP1

Figure 42. MIDORI's interactome candidates. (A) Graphical representation of the interaction candidates identified by mass-spectrometry, represented according to the Fold change (FC>3) (Y axis) and its BDFR score (BDFR<0.02) (X axis). (B) GO molecular function analysis of MIDORI's interactors using PantherGO tool. (C) GO cellular component analysis of MIDORI's interactors using PantherGO tool. (D) Table showing the subcellular localisation of MIDORI's interactor candidates in different compartment.

We validated some interactor candidates that we found particularly interesting by co-immunoprecipitation experiments. The first candidate we validated is ITGB1BP1 (Figure 43A). This protein, also known as ICAP1, is an INTEGRINβ1 binding protein (Chang et al., 1997) and it is known to inhibit the Rho GTPase CDC42 by blocking its binding to GTP (Degani et al., 2002). This way ITGB1BP1 impairs the formation of focal adhesions and consequently, negatively regulate cell spreading through the ECM and cell migration (Bouvard et al., 2003; Degani et al., 2002) (Figure 43B). Thus, it is possible that MIDORI interacts with ITGB1BP1 enhancing its function, which in turn impairs the cell interaction with the matrix, cell migration and invasion.

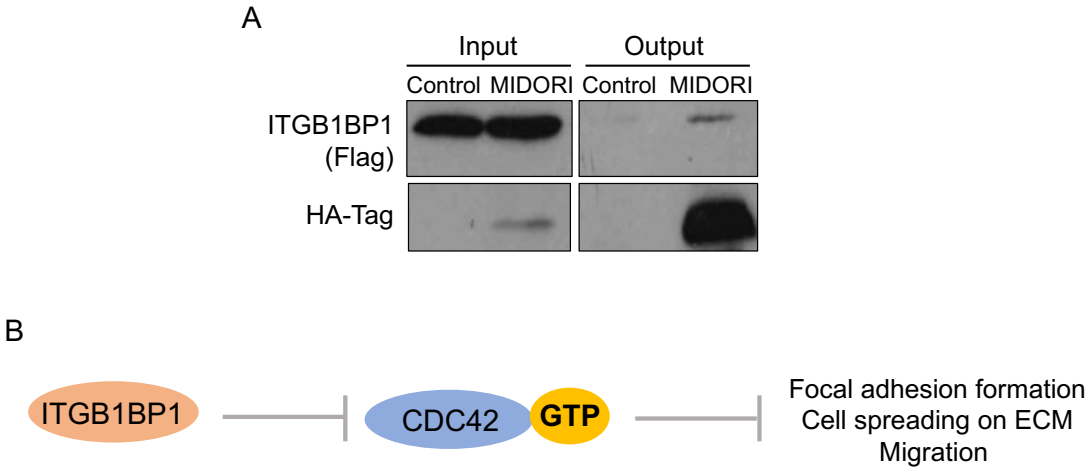


Figure 43. MIDORI interacts with ITGB1BP1. (A) Validation of the interaction of MIDORI with ITGB1BP1 by co-immunoprecipitation using an anti-HA antibody in 293T cells transiently transduced with MIDORI-HA or the empty vector as control and Flag-ITGB1BP1, followed by Western blot. (B) Scheme of ITGB1BP1 mechanism of action.

The second validated candidate is HMGA1 (Figure 44A). This is a chromatin-associated protein involved in the regulation of gene transcription associated to many cellular processes (Cleynen and Van de Ven, 2008), as well as in ribosome biogenesis (Hall et al., 2006). Of note, HMGA1 can also be secreted, for example upon TGFβ signalling (Zhong et al., 2017), and it has been demonstrated to increase cell invasion and breast cancer metastasis (Mendez et al., 2018).

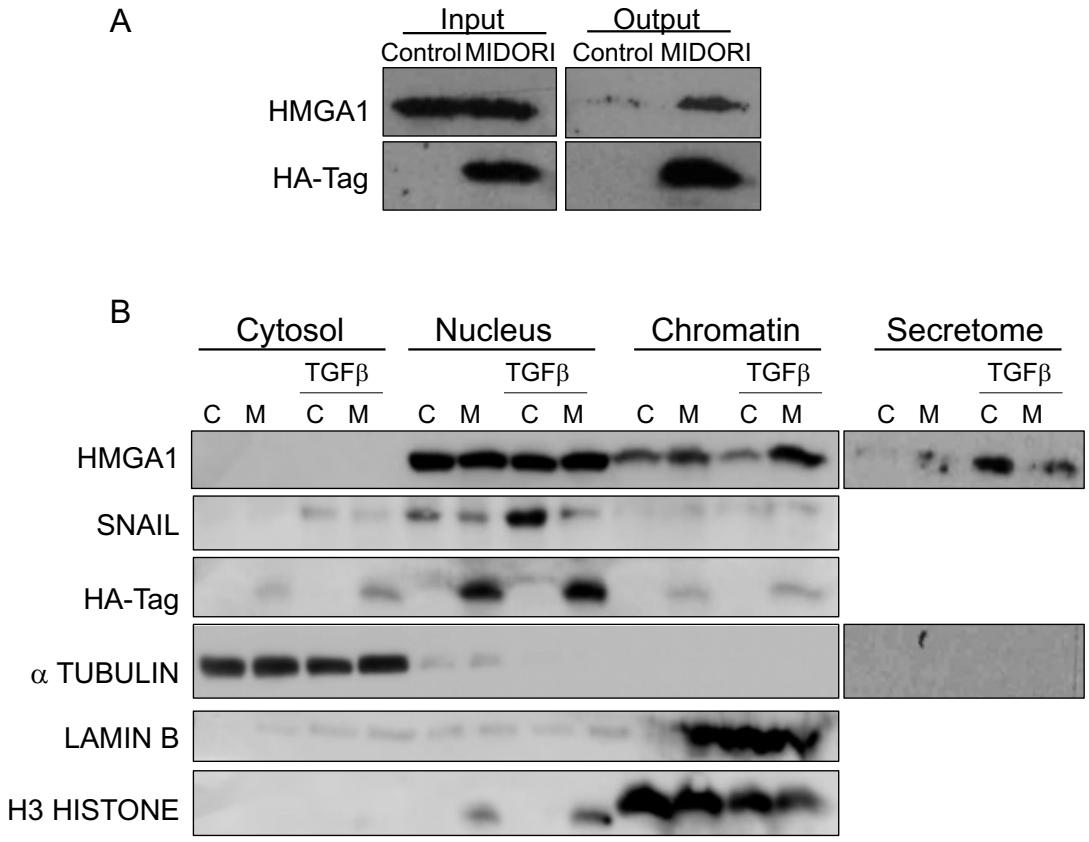


Figure 44. MIDORI interacts with HMGA1 and decreases its secretion upon TGFβ treatment. (A) Validation of the interaction of MIDORI with HMGA1 by co-immunoprecipitation using an anti-HA antibody in MDA-MB-231 cells overexpressing exogenous MIDORI-HA (one day of doxycycline induction), followed by Western blot. (B) Western blot analysis performed in the subcellular fractionations cytosol, nucleus, chromatin and in the concentrated conditioned medium of MDA-MB-231 cells overexpressing MIDORI or the empty vector after 4 days of doxycycline induction, treated or not with TFGβ during 48 hours. α-Tubulin, Lamin B and H3 histone are markers for cytoplasmatic, nuclear and chromatin fractions, respectively. The absence of α-Tubulin in the secretome was used as a quality control of the secreted fraction.

Given the role of the HMGA1 in breast cancer invasion and metastasis, we sought to investigate if MIDORI overexpression alters its expression, localisation or its secretion. Of note, we found that HMGA1 is enriched in the chromatin fraction upon MIDORI overexpression, but it decreases in the secreted fraction upon TGFβ treatment (Figure 44B). Importantly, these results are in line with the role MIDORI reducing breast cancer invasion.

The last candidate that we validated is MYBBP1A, the top one in our list according to its FC and BDFR (Figure 45A). MYBBP1A is a corepressor normally located in the nucleolus. However, upon several stresses, such as genotoxic stress or ribosomal stress, it shuttles to the nucleoplasm where it acts inhibiting a plethora of transcription factors (Felipe-Abrio and Carnero, 2020). In fact, it is reported to be a negative regulator of MYB protein, it inhibits RNA POL I blocking ribosomal RNA (rRNA) transcription and ribosome biogenesis, it favours the acetylation and stabilisation of p53 and, finally, it inhibits the transcriptional activity of NF-κB (Felipe-Abrio and Carnero, 2020).

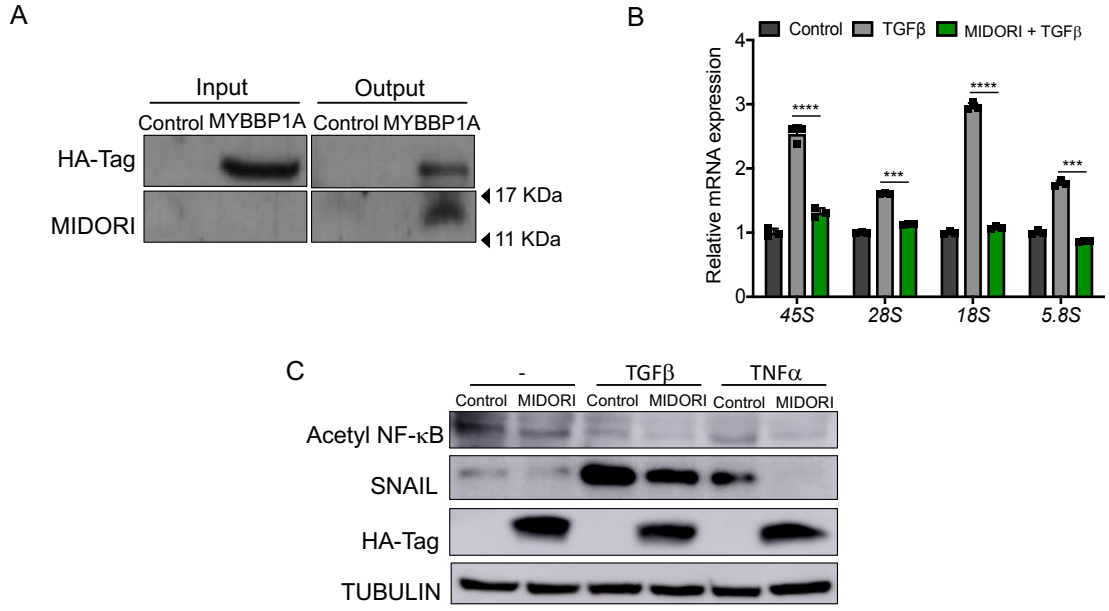


Figure 45. MIDORI interacts with MYBBP1A and decreases rRNA transcription upon TGFβ treatment. (A) Validation of the interaction of MIDORI with MYBBP1A by co-immunoprecipitation using an anti-HA antibody in 293T cells transiently transduced with MYBBP1A-HA or the empty vector as control, followed by Western blot. MIDORI was detected using our custom-made antibody. (B) mRNA expression analysis of the indicated genes in MDA-MB-231 cells overexpressing MIDORI or the empty vector as a control after 4 days of doxycycline induction, with or without 48 hours of TGFβ treatment by RT-qPCR. Values are relativised to *GAPDH* and normalised to cells expressing the empty vector as a control. ***$P < 0.001$, ****$P < 0.0005$ using Student's t-test for statistics. Bars represent mean expression ± SD of three technical replicates. (C) Western blot analysis showing the indicated proteins in MDA-MB-231 breast cancer inducible cells after 4 days of doxycycline induction and 48 hours of TGFβ treatment or 24 hours of TNFα treatment.

Given rRNA transcription is known to increase during EMT (Prakash et al., 2019), we tested if MIDORI might have an effect in this process. Similarly to what is reported for MYBBP1A, we saw a significant decrease in rRNAs transcription when overexpressing MIDORI upon TGFβ treatment (Figure 45B). Finally, it is known that MYBBP1A reduces NF-κB transcriptional activity by competing with p300, reducing in turn NF-κB acetylation (Owen et al., 2007). Remarkably, we observed that MIDORI overexpression induces a decrease in the acetylation of NF-κB (Figure 45C). Taken together, these data could suggest that MIDORI may be acting as a negative regulator of NF-κB through MYBBP1A. As a last note, we observed that MIDORI overexpression reduces the expression of SNAIL upon TGFβ but also upon TNFα (Figure 45C), suggesting that MIDORI might be acting also as a negative regulator of TNFα-induced EMT.

In conclusion, we validated three interactors that could help explaining the functions of MIDORI at the molecular level. However, further analyses are needed to define the molecular mechanisms behind the role of MIDORI as a negative regulator of the EMT programme.

3. Analysis of the role of MIDORI in EMT-related processes

Given that our *in vitro* results strongly supported an inhibitory role of MIDORI in EMT, we sought to investigate the function of MIDORI in processes where the EMT is known to play a role such as cancer metastasis, wound healing, fibrosis and cellular reprogramming.

3.1. Role of MIDORI in cancer

3.1.1. MIDORI overexpression impairs breast cancer metastasis *in vivo*

As we demonstrated that MIDORI overexpression is sufficient to prevent cells from undergoing EMT, we sought to investigate if it could also prevent metastasis, a process known to require partial EMT (see Introduction). To study MIDORI's role in metastasis we used two different experimental settings using human MDA-MB-231 breast cancer cell line, overexpressing or not human MIDORI in a constitutive manner. On the one hand, we performed a metastatic colonisation assay by injecting intravenously MDA-MB-231 cells in mice tail vein. We followed up metastases' appearance in the lungs by the detection of a luciferase reporter and we finally compared the number of metastatic foci at the experimental endpoint (Figure 46A). On the other hand, we used an orthotopic model that resembles better the metastatic process. We orthotopically injected control or MIDORI-overexpressing MDA-MB-231 cells in the mammary fat pad, we resected the primary tumours when they reached a volume of 400-500 mm^3 and we followed up the appearance of metastases by the luciferase reporter (Figure 46B).

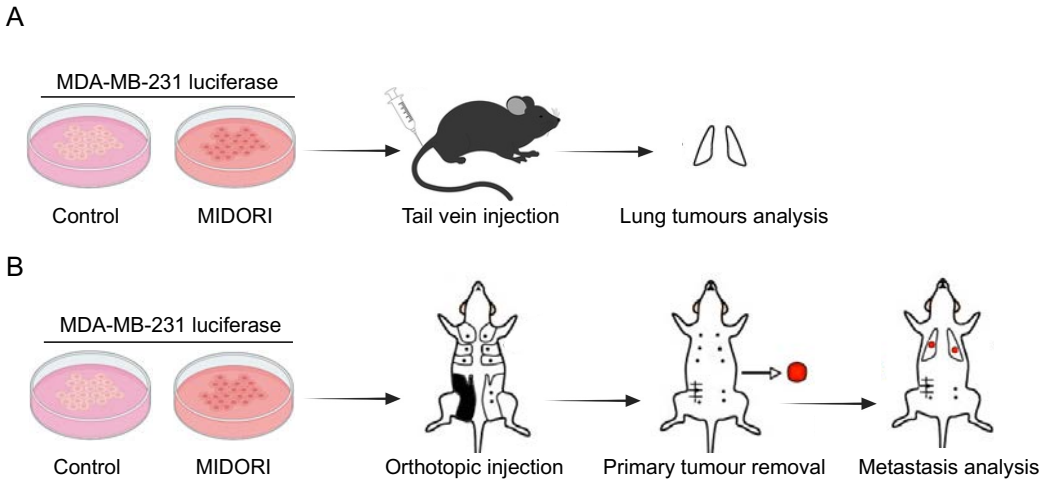


Figure 46. *In vivo* experimental settings for studying MIDORI's role in metastasis. (A) Schematic representation of the lung colonisation experiment. (B) Schematic representation of the orthotopic tumours and metastasis experiment.

In the first experimental setting we observed that mice bearing MIDORI-overexpressing cells show less lung colonisation over time compared to controls (Figure 47A-B). To better analyse lung lesions, we also performed micro-computed tomography (μCT) of the thoracic cavity of these mice 9 weeks after the injection. Lung metastasis burden was measured and the percentage of healthy tissue volume per mouse was calculated (Figure 47C-D). Performing this analysis, we confirmed that mice bearing MIDORI-

overexpressing cells display healthier lungs. Consistently, when we analysed the lungs at the endpoint of the experiment, we observed a reduction in the number of metastatic foci in mice inoculated with MIDORI-overexpressing cells (Figure 47E-F).

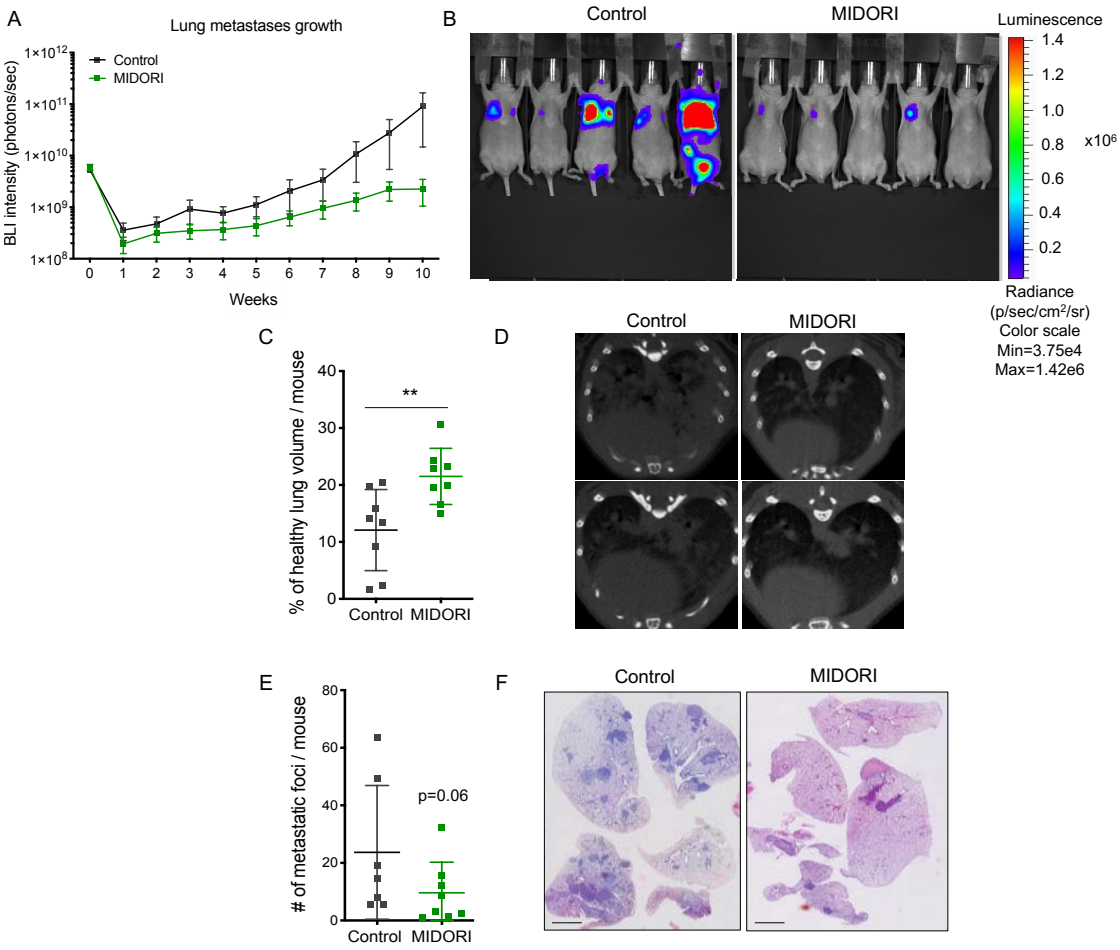


Figure 47. MIDORI-overexpressing MDA-MB-231 cells display an impaired capacity of lung colonisation. (A) Luciferase activity measured weekly by IVIS imaging as bioluminescence (BLI) intensity in mice inoculated with MIDORI-overexpressing MDA-MB-231 cells compared to control mice. Error bars represent mean ± SEM of n=8 mice per group. (B) Luminescence images representative of 5 mice inoculated with MDA-MB-231 cells expressing the empty vector or MIDORI in a model of lung colonisation upon tail vein cells injection. (C) Percentage of healthy lung volume per mouse measured from µCT images. Each single dot represents a single animal of the group. Error bars represent ± SEM of n=8 mice per group. $^{**}P < 0.01$ using Student's t-test for statistics. (D) Representative µCT images of two mice bearing MDA-MB-231 cells expressing the empty vector as a control or MIDORI nine weeks after tail vein injection. (E) Number of metastatic foci per mouse in mice inoculated with MDA-MB-231 cells overexpressing MIDORI or the empty vector as a control. Each dot represents an animal. Error bars represent mean ± SEM of n=7 mice per group (Control) and n=8 mice per group (MIDORI). (F) Representative H&E pictures of lungs from mice bearing MIDORI-overexpressing MDA-MB-231 or control, scale bar=2mm.

We repeated the tail vein injection experiment performing a survival study. Mice bearing MIDORI-overexpressing cells show an increased metastasis-free survival and lung metastasis-free survival compared to the controls, demonstrating that MIDORI inhibits the metastatic colonization process *in vivo* (Figure 38A-B).

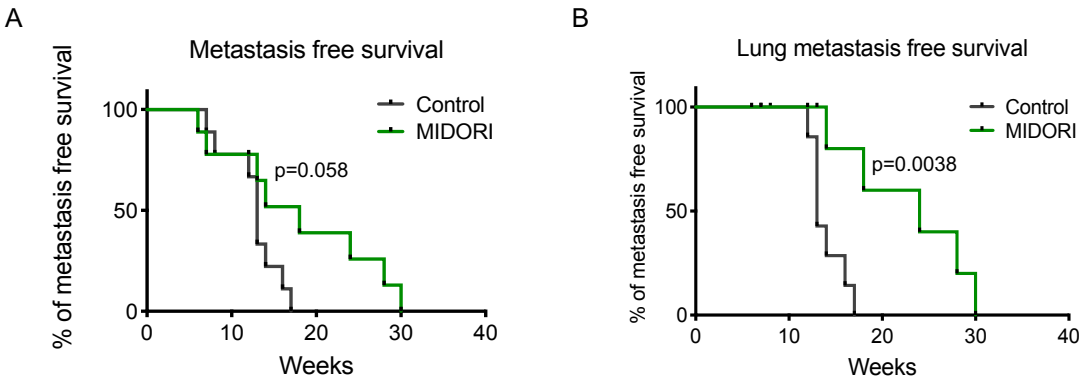


Figure 48. MIDORI overexpression increases metastasis-free survival. (A) Kaplan-Meier graph showing the metastasis free survival of mice inoculated with MIDORI-overexpressing MDA-MB-231 or control cells. P=0.058 using Mantel-Cox test for statistics (n=10 mice per group). (D) Kaplan-Meier graph showing the lung metastasis free survival of mice bearing MIDORI-overexpressing MDA-MB-231 or control. P=0.0038 using Mantel-Cox test for statistics (n=10 mice per group).

To support these data, as mentioned above, we also performed orthotopic injection of MDA-MB-231 cells in the mammary fat pad. Importantly, MIDORI overexpression does not affect the growth of primary tumours (Figure 49A), in agreement with our data that shows that MIDORI does not affect cell proliferation *in vitro* (Figure 27). We resected the primary tumours, we let metastases appear and we analysed the number of organs with metastasis at the end of the experiment, when mice reached the human endpoint (see *Materials and Methods*). We did not observe a clear effect in the distribution pattern of metastases in the different organs (Figure 49B), but we saw a significant decrease in the number of colonised organs (Figure 49C) and a decrease metastasis signal per organ in mice inoculated with MIDORI-overexpressing cells (Figure 49D-E).
Therefore, we conclude that MIDORI acts as a negative regulator the EMT programme also *in vivo* impairing the metastatic process.

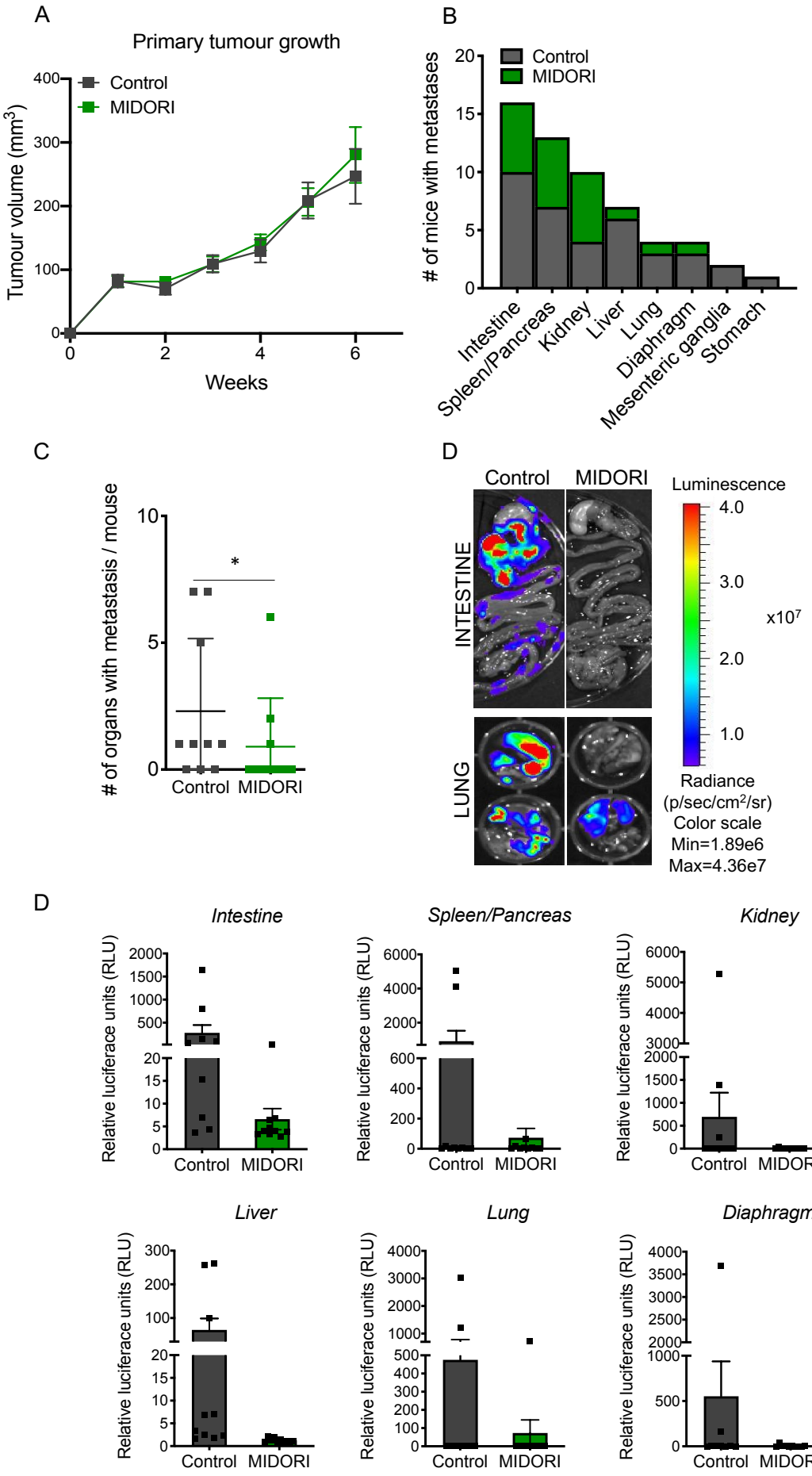

A
Primary tumour growth
Control
MIDORI
Tumour volume (mm3)
Weeks
B
Control
MIDORI
# of mice with metastases
Intestine
Spleen/Pancreas
Kidney
Liver
Lung
Diaphragm
Mesenteric ganglia
Stomach
C
# of organs with metastasis / mouse
Control
MIDORI
D
Control
MIDORI
INTESTINE
LUNG
Luminescence
x10^7
Radiance (p/sec/cm2/sr)
Color scale
Min=1.89e6
Max=4.36e7
D
Intestine
Spleen/Pancreas
Kidney
Liver
Lung
Diaphragm
Relative luciferace units (RLU)
Control
MIDORI

Figure 49. MIDORI significantly impairs metastasis formation after primary tumour resection. (A) Primary tumour growth curve of mice bearing MDA-MB-231 cells overexpressing MIDORI or the empty vector as a control. Error bars represent mean ± SEM of n=10 mice per group. (B) Metastasis distribution pattern in the indicated organs in mice bearing MIDORI-overexpressing MDA-MB-231 or control cells (n=10 mice per group). (C) Number of organs with metastasis per mouse in mice bearing MIDORI-overexpressing MDA-MB-231 or control cells. Each dot represents an animal. Error bars represent mean ± SEM of n=10 mice per group. **P* < 0.05 using Fisher's exact test for statistics. (D) Representative luminescence images of organs with metastases isolated from mice bearing MDA-MB-231 cells expressing the empty vector as a control or MIDORI in a model of orthotopic cell injection and metastatic colonisation. (E) Quantification of the luminescence signal in the indicated organs by IVIS at the end of the experiment. Each dot represent an animal. Error bars represent mean ± SEM of n=10 mice per group.

3.1.2. MIDORI downregulates the mesenchymal programme and impairs the growth of cSCC tumors *in vivo*

To assess the effect of MIDORI overexpression in other cancer types, we selected a patient-derived cutaneous skin squamous cell carcinoma (cSSC) cell line, cSSC11, that grows in spheres and shows robust mesenchymal properties (Figure 50A) (Bernat-Peguera et al., 2021; Bernat-Peguera et al., 2019). Indeed, the overexpression of MIDORI in these cells (Figure 50B) downregulates the expression of EMT-related genes (Figure 50C), consistently with our previous observations.

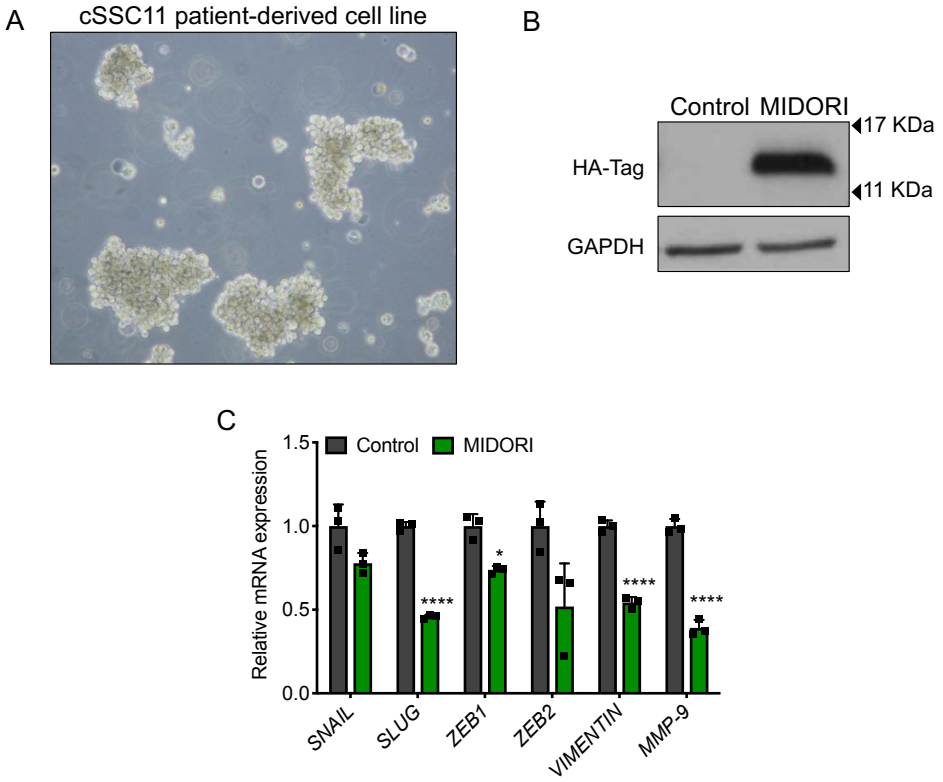


Figure 50. MIDORI reduces the expression of EMT markers in cSSC11 patient-derived cell line. (A) Representative bright-field microscopy picture of cSSC11 patient-derived cell line taken using a 10X magnification. (B) Analysis of exogenous MIDORI expression in the cSSC11 cell line after 48 hours of doxycycline induction (1 µg/ml). Control cells were transduced with the empty backbone. Protein expression was detected by immunoblot using an anti-HA-tag antibody. (C) mRNA expression analysis of the indicated genes in cSSC11 cells 4 days post doxycycline

induction by RT-qPCR. Values are relativised to *GAPDH* and normalised to cells expressing the empty vector as a control. *P < 0.05, ***P < 0.001, ****P < 0.0005 using Student's t-test for statistics. Bars represent mean expression ± SD of three technical replicates.

We generated PDXs by injecting cSSC11 cells subcutaneously in immunocompromised mice and, to our surprise, we observed that MIDORI overexpression significantly impairs primary tumour growth in this model (Figure 51A-B).

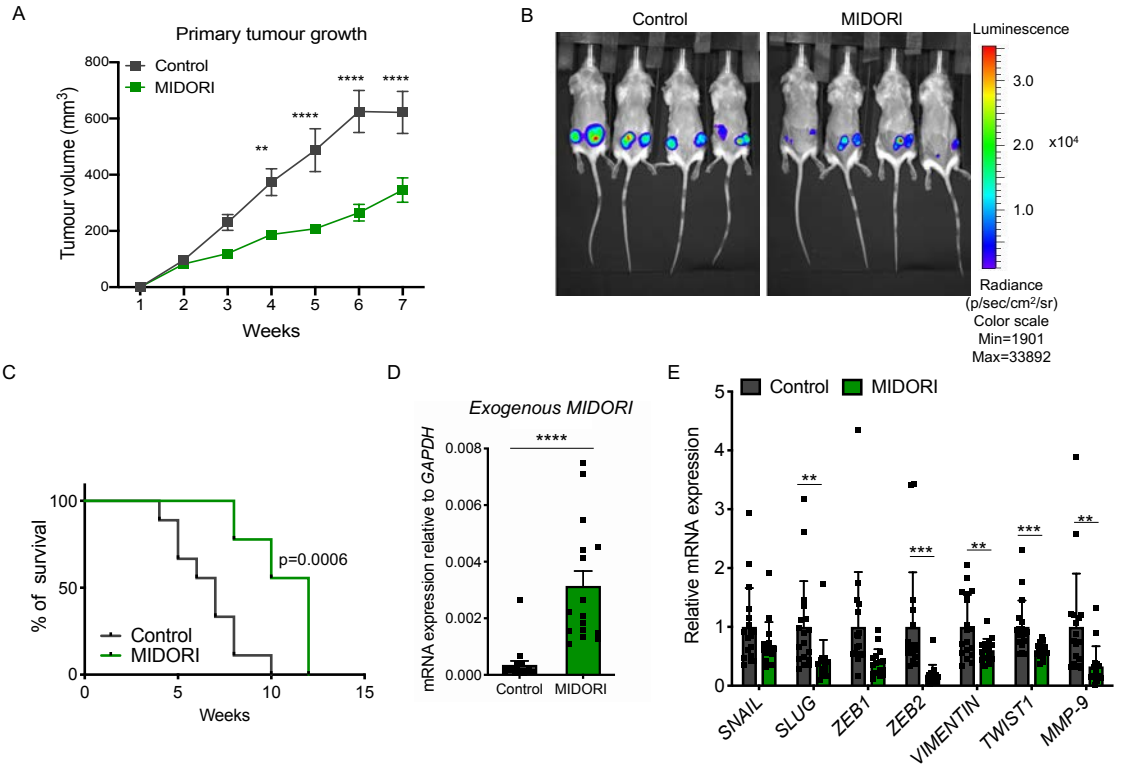


Figure 51. MIDORI impairs cSSC11 tumour growth and downregulates EMT markers *in vivo*. (A) Primary tumour growth curve of mice bearing MIDORI-overexpressing cSSC11 or control. Error bars represent mean ± SEM of n=18 tumours per group. **P < 0.01, ****P < 0.0005 using Student's t-test for statistics. (B) Luminescence images representative of 4 mice bearing cSSC11 cells expressing the empty vector as a control or MIDORI. (C) Kaplan-Meier graph showing the tumour free survival of mice bearing MIDORI-overexpressing cSSC11 or control. P=0.0006 using Mantel-Cox test for statistics (n=9 mice per group). (D) Detection of MIDORI overexpression by RT-qPCR in cSSC11 tumours. Each dot represents a single tumour. Error bars represent mean ± SEM of n=18 tumours per group. ***P < 0.0005 using Student's t-test for statistics. (E) mRNA expression analysis of the indicated genes in cSSC11 primary tumours by RT-qPCR. Values are relativised to *GAPDH* and normalised to tumours expressing the empty vector as a control. **P < 0.01, ***P < 0.001, using Student's t-test for statistics. Bars represent mean expression ± SEM of eighteen biological replicates.

Since we had to euthanise the mice due to the primary tumors at different timepoints (Figure 51C), we could not reach any conclusion regarding metastasis using this experimental model. However, we checked the expression of EMT markers in the primary tumours. Similarly to what observed *in vitro*, we also saw a significant decrease in the expression of EMT markers in MIDORI-overexpressing cSSC11 tumours (Figure 51E). Thus, we conclude that MIDORI overexpression can reduce the mesenchymal programme *in vivo* also in a patient-derived cSCC model.

3.1.3. Analysis of the effect of MIDORI deficiency in breast cancer

To study MIDORI loss-of-function in cancer we crossed our MIDORI-KO mouse with the MMTV-PyMT mouse model (Figure 52A). This transgenic mouse strain expresses the polyomavirus middle T-antigen (PyMT) in the mammary gland epithelium under the control of the *MMTV* promoter. It spontaneously develops breast cancer and progresses from adenocarcinoma until the stage of metastatic breast cancer (Guy et al., 1992). We analysed the onset of breast tumours and the overall survival of these mice, but we did not observe any difference of MIDORI-KO versus MIDORI WT mice (Figure 52B). In fact, we found no metastasis in the lungs of these mice, which are the organs most affected by metastasis in this model (data not shown). Since it is known that metastasis appearance is really compromised when the MMTV-PyMT alleles are carried in a mixed background (Attalla et al., 2021), we are currently backcrossing our animal to the FVB genetic background before analysing the effect of MIDORI deficiency in metastasis.
In the meanwhile, we isolated breast cancer cell lines from tumours that arose in MIDORI WT and MIDORI-KO mice (Figure 52C) and we used them to perform migration assays (Figure 53A).

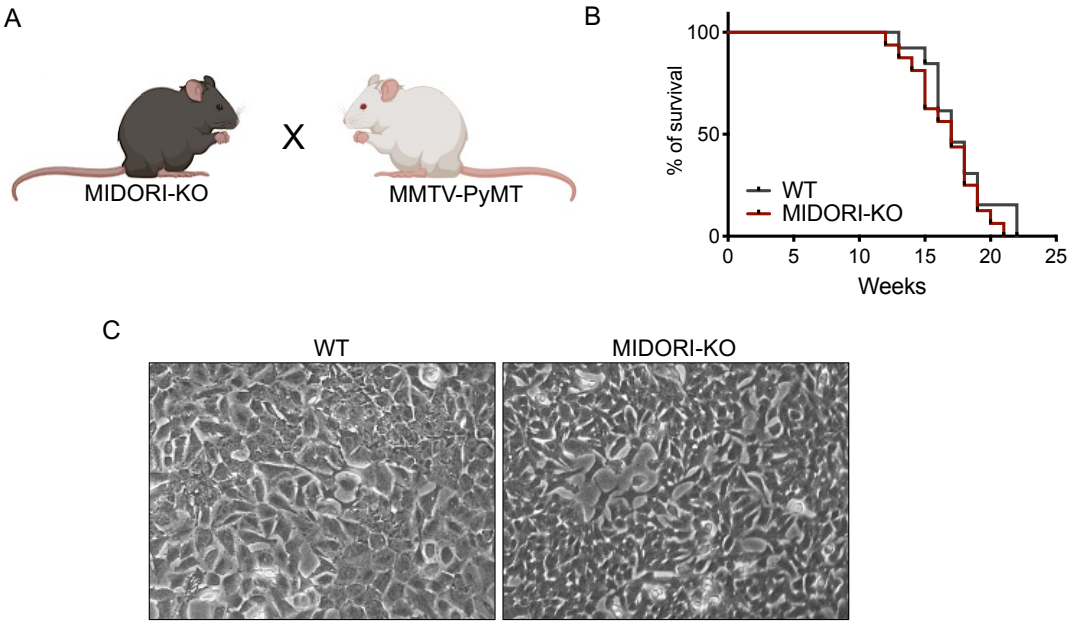


Figure 52. Study of MIDORI loss-of-function *in vivo* using the MMTV-PyMT; MIDORI-KO mouse strain. (A) Scheme of the crossing between MIDORI-KO and MMTV-PyMT mouse strains. (B) Kaplan-Meier graph showing the survival of MMTV-PyMT; MIDORI-KO mice or WT (n=13 mice per group). (C) Representative bright-field microscopy picture of the breast cancer cell lines isolated from MIDORI WT and MIDORI KO; MMTV-PyMT mouse models using 10X magnification.

We saw that MIDORI-KO breast cancer cells migrate faster compared to the WT counterpart (Figure 53B), suggesting that MIDORI deficiency could enhance the migration capability of cancer cells.

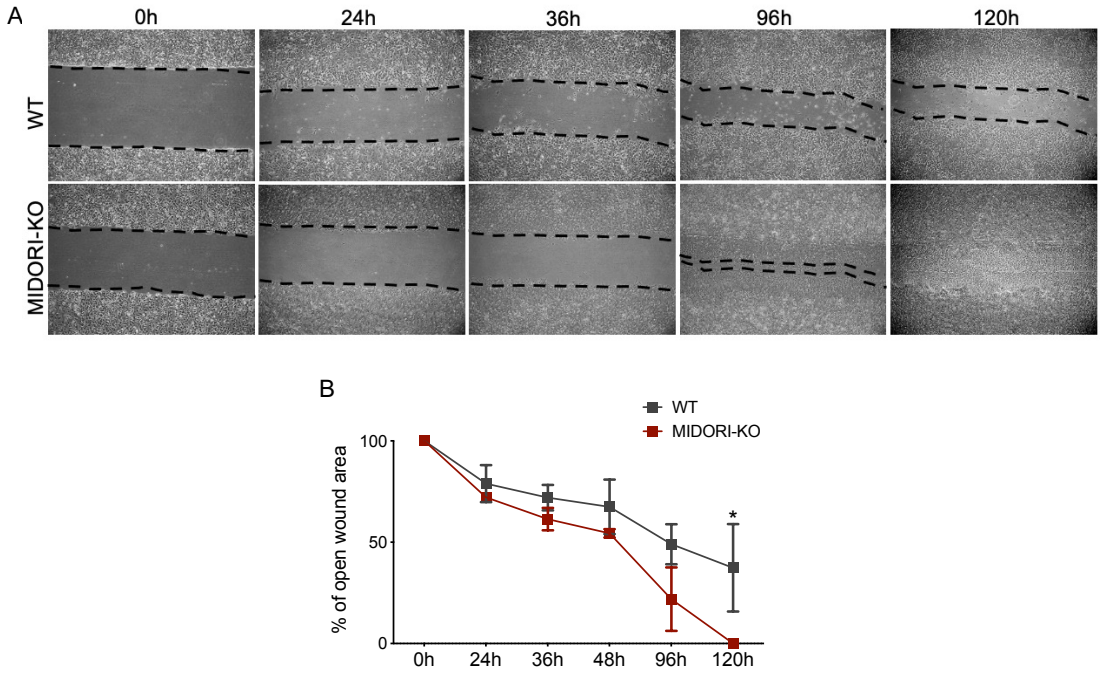


Figure 53. MIDORI deficiency increases breast cancer cell migration. (A) Cell migration assay of WT and MIDORI-KO MMTV-PyMT breast cancer cells. Pictures were taken at 0 hours and different time points post-scratch. Cell migration was assessed by the measurement of the scratch closure. (B) Quantification of the wound healing assay as percentage of open wound area at the indicated timepoints. Error bars represent mean expression ± SD of two biological replicates relativised to the area at 0 hours post scratch. *$P < 0.05$, using two-way ANOVA test for statistics.

3.2. Role of MIDORI in *in vivo* wound healing

We next wanted to assess the role of MIDORI in other processes related with the activation EMT, as wound healing (Nieto et al., 2016). We carried out wound healing experiments *in vivo* by performing cutaneous wounds in the back skin of WT and MIDORI-KO mice and scored wound closure (Figure 54A). We saw that MIDORI-KO mice heal significantly faster compared to their WT counterpart (Figure 54B), additionally pointing to an increased EMT activation when MIDORI expression is lost.

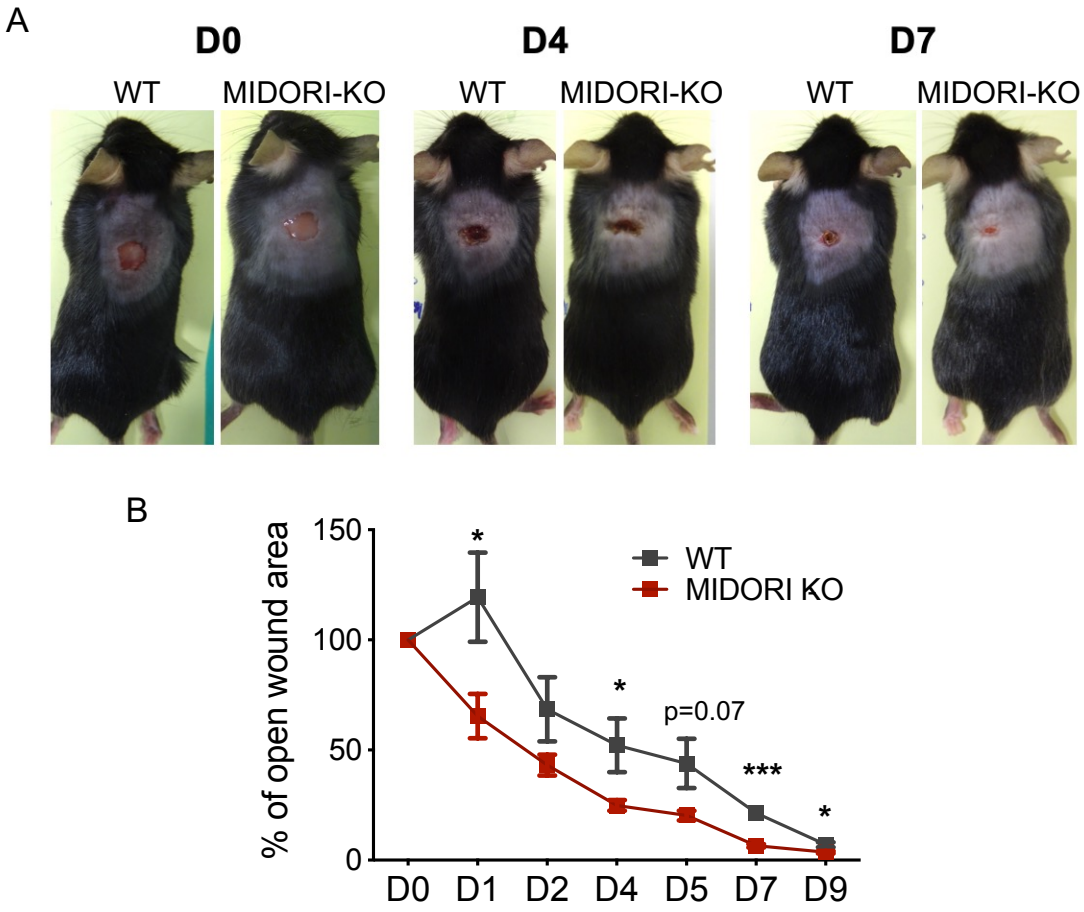


Figure 54. MIDORI loss-of-function increases *in vivo* wound healing. (A) Representative pictures of WT and MIDORI-KO mice at the indicated timepoints of *in vivo* wound healing. (B) Quantification of the wound healing assay as percentage of open wound area at the indicated timepoints. Error bars represent mean expression ± SEM of n=5 mice per group relativised to the area at 0 hours post wound. *$P < 0.05$, ***$P < 0.001$ using multiple non-parametric T-Test for statistics.

3.3. Role of MIDORI in renal fibrosis

It is known that EMT is also activated in the process of fibrosis (Grande et al., 2015; Nieto et al., 2016). Of note, we saw that MIDORI acts as a negative regulator of *Collagen I* expression (Figure 26 and 38A), a major component of the fibrotic scar (Wells, 2022). Altogether, these observations made us wonder whether MIDORI deficiency has an effect in fibrosis, and we focused in kidney fibrosis, given that endogenous MIDORI seems to be translated in kidney (Figure 21). To investigate that, we induced acute kidney injury by administering folic acid intraperitoneally to our mice (Figure 55A). We then analysed the kidneys thirty days after the injection, when kidney fibrosis has already occurred. First, we looked at the expression of EMT markers (Figure 55B) and proinflammatory cytokines (Figure 55C) that are known to be upregulated upon renal fibrosis (Grande et al., 2015).

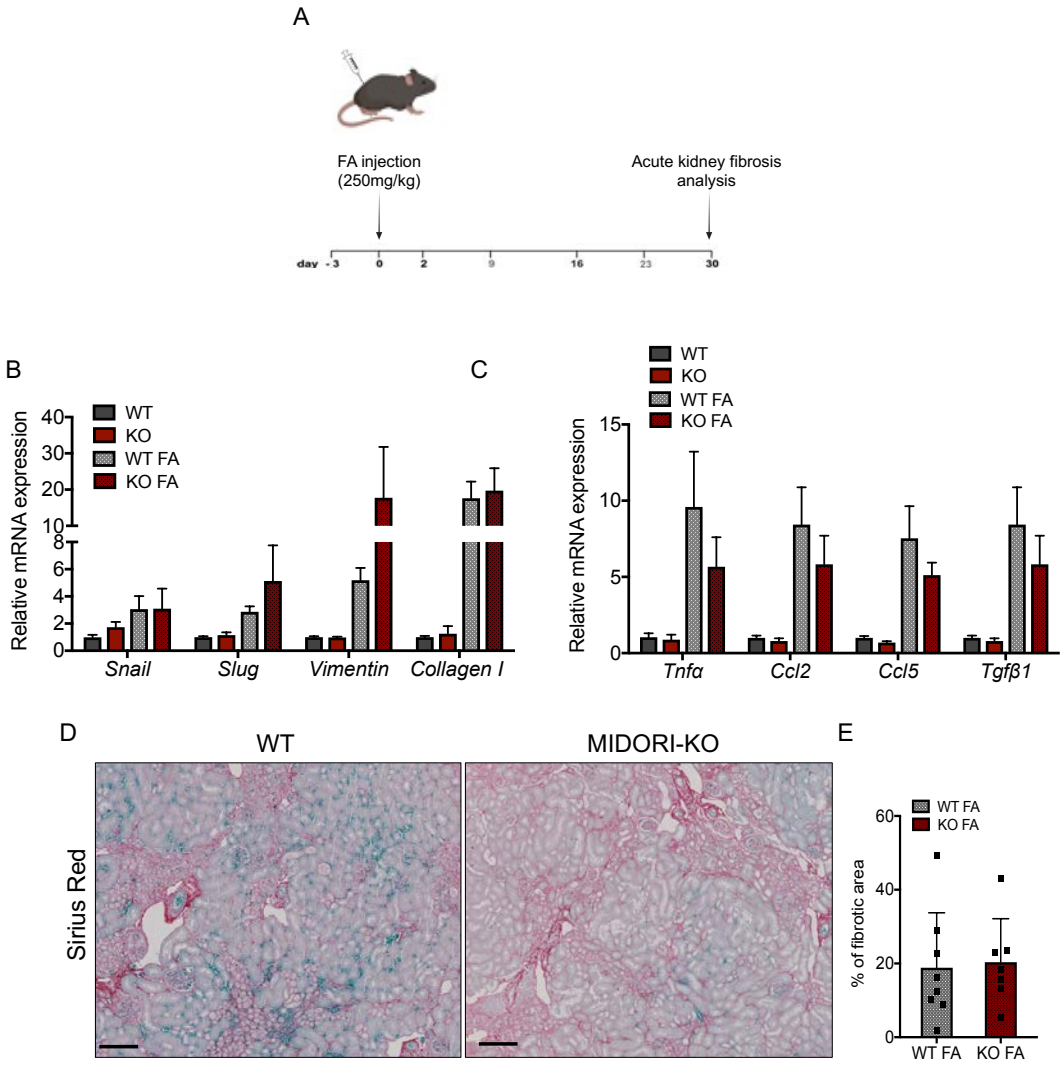


Figure 55. MIDORI loss-of-function does not affect acute kidney fibrosis. (A) Scheme of acute kidney fibrosis induction in WT and MIDORI KO mice. (B-C) mRNA expression analysis of the indicated genes by RT-qPCR. Values are relativised to *Gapdh* and normalised to WT untreated mice as control. Bars represent mean expression ± SEM of n=8. (D) Representative pictures of Sirius Red staining in WT and MIDORI KO kidneys. Scale bar = 200μm. (E) Quantification of the percentage of fibrotic area per kidney. Each dot represents a single animal. Bars represent mean expression ± SEM of n=8 control mice and n=7 MIDORI-KO mice.

Although fibrotic MIDORI-KO kidneys showed an increase in the expression of *Slug* and *Vimentin*, we did not observe significant changes in gene expression between MIDORI-KO and its WT counterparts. We then performed Sirius Red staining to measure the fibrotic area in the kidneys, and we did not observe any difference between WT and MIDORI-KO mice (Figure 55D y E).

Concluding, even if MIDORI deficiency seems to have a positive effect on wound healing, so far we have not observed any effect on acute kidney fibrosis. However, since the induction of fibrosis with this protocol is very fast, in future experiments we will examine earlier timepoints, before the fibrosis is already established, when perhaps is more difficult to see significant differences.

3.4. Functional characterisation of MIDORI in cellular reprogramming

It is well established that MET is a critical step during reprogramming to iPSCs (Li et al., 2010; Samavarchi-Tehrani et al., 2010). However, it has been reported that having a more mesenchymal phenotype at the early stage of the process could facilitate the cell fate transition, and a sequential EMT-MET enhances significantly reprogramming efficiency (Di Stefano et al., 2014; Liu et al., 2013) (Figure 56A).
Given the role of MIDORI inhibiting EMT we wanted to study the possible role of MIDORI as a regulator of cellular reprogramming.

3.4.1. Endogenous MIDORI is transiently expressed at the early stages of cellular reprogramming

Firstly, we wanted to assess whether endogenous MIDORI is naturally expressed during cellular reprogramming of MEFs to iPSCs by Oct4, Sox2, Klf4 and c-Myc (hereafter OSKM). For that, we transduced MEFs with a doxycycline-inducible lentiviral vector encoding OSKM, and treated the infected cells with doxycycline until iPSC colonies were apparent and fully formed.

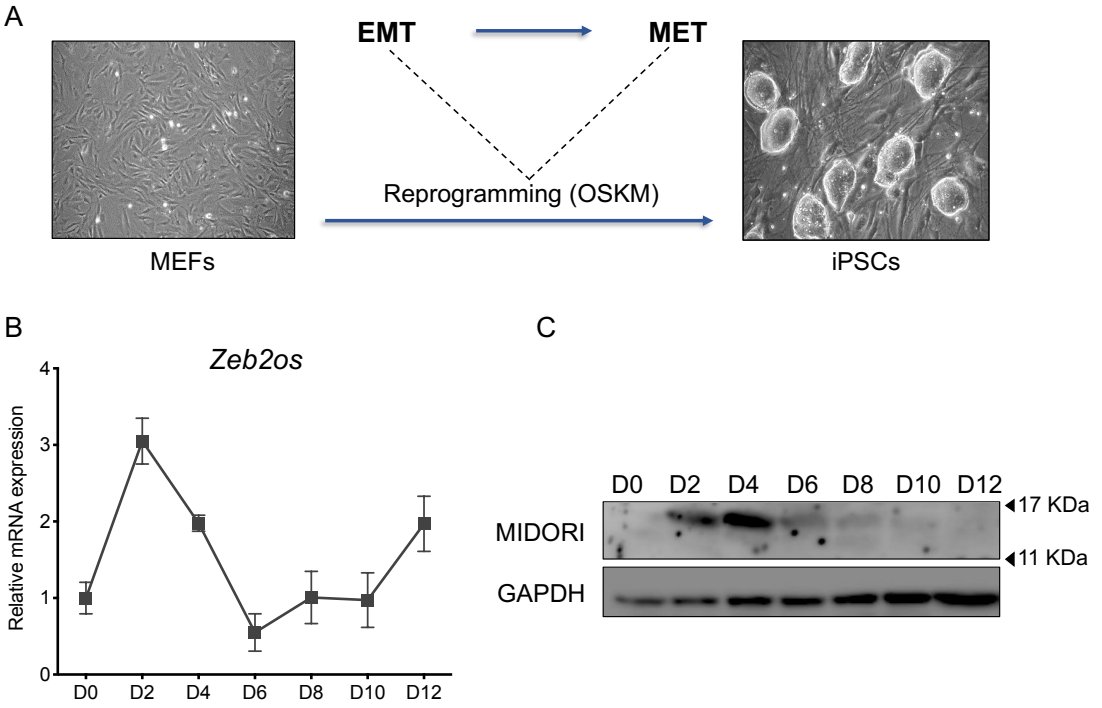


Figure 56. MIDORI is upregulated in the first stages of cellular reprogramming. (A) Scheme showing EMT and MET sequential activation during cellular reprogramming. (B) mRNA expression of *Zeb2os* in WT MEFs during the cellular reprogramming process, measured by qRT-PCR. Values are represented as mean ± SD of three technical replicates normalised to *Gapdh* and relativised to day 0. (C) Western blot showing endogenous MIDORI protein expression at indicated timepoints during cellular reprogramming.

At the mRNA level, we saw that *Zeb2os* lncRNA expression is transiently upregulated during the reprogramming process, reaching its maximum expression on day 2 and decreasing afterwards (Figure 56B). This result is in line with the activation of EMT during the first stages of cellular reprogramming and its shutdown at later stages. At the protein level, we observed by using our custom-made antibody that endogenous MIDORI expression is transiently detected at the early stages of reprogramming, reaching its maximum on day 4 (Figure 56C). Indeed, the fact that the expression of the microprotein is induced during the reprogramming process suggests that it might be required for cellular reprogramming.

3.4.2. Effect of MIDORI overexpression in cellular reprogramming

To test MIDORI's role in cellular reprogramming, we reprogrammed MEFs by transducing them with a doxycycline-inducible lentiviral vector encoding the OSKM factors together with an inducible-vector encoding murine MIDORI or the empty control. We found that iPSCs colonies appear faster in MIDORI-overexpressing MEFs (Figure 57A) and they produce more iPSCs colonies at the end of the process (Figure 57B-C), indicating that MIDORI overexpression accelerates cellular reprogramming and significantly improves reprogramming efficiency. This is consistent with the role of MIDORI as a negative regulator of EMT.

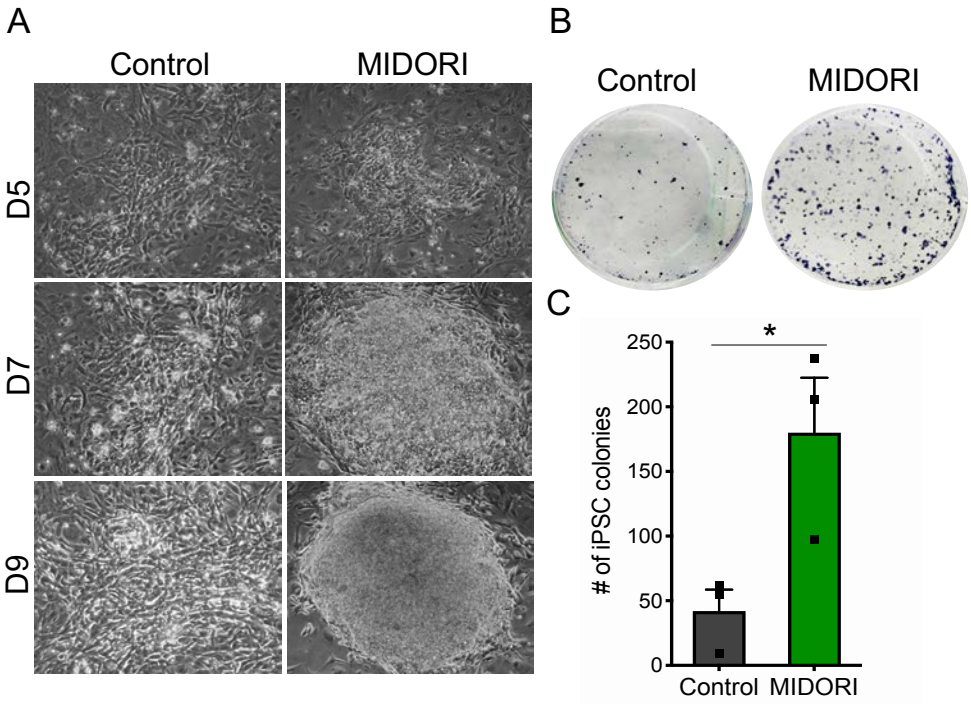


Figure 57. MIDORI overexpression increases cellular reprogramming efficiency. (A) Representative bright-field microscopy images of the formation of iPSCs colonies derived from MEFs transduced with a doxycycline-inducible lentiviral vector encoding the OSKM factors together with the murine MIDORI-inducible vector or the empty as control. Pictures were taken on days 5, 7 and 9 post doxycycline induction to induce reprogramming with 20X magnification. (B) Representative iPSCs colonies positive for alkaline phosphatase (AP) staining at day 12 of reprogramming. (C) AP staining quantification performed at day 12 of reprogramming. Each dot

represents a single biological replicate. Bars represent the mean ± SD of three biological replicates. *P<0.05 using Student's T-Test for statistics.

Given the role of MIDORI modulating the secretome, and the fact that reprogramming is regulated by cytokines and secreted factors we wanted to test if the improved reprogramming efficiency is due to a cell-autonomous or non-cell-autonomous effect. We collected the CM from MEFs constitutively expressing MIDORI or the empty vector as a control and we used them to reprogramme MEFs isolated from reprogrammable mice (from now on i4F mice), which ubiquitously express OSKM in a doxycycline-inducible manner (Abad et al., 2013). Strikingly, the CM derived from MIDORI-overexpressing cells increases very significantly the efficiency of cellular reprogramming to iPSCs (Figure 58A-C).
These results indicate that MIDORI might be increasing cellular reprogramming in a cell extrinsic manner, probably by modulating the cell secretome.

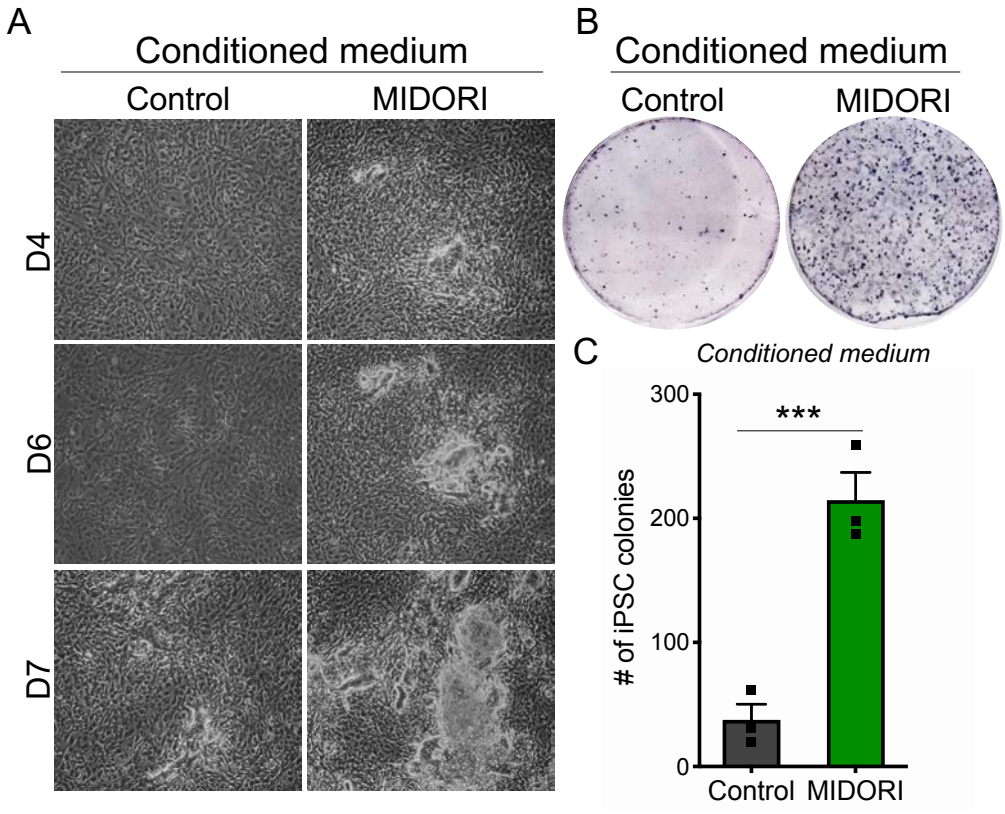


Figure 58. The conditioned medium of MIDORI-overexpressing cells increases cellular reprogramming efficiency. (A) Representative bright-field microscopy images of the formation of iPSCs colonies derived from i4F MEFs reprogrammed using MIDORI-overexpressing MEFs CM or from control cells. Pictures were taken on days 4, 6 and 7 post doxycycline induction in i4F MEFs with 10X magnification. (B) Representative iPSCs colonies positive for alkaline phosphatase (AP) staining at day 12 of reprogramming. (C) AP staining quantification performed at day 12 of reprogramming. Each dot represents a single biological replicate. Bars represent the mean ± SD of three biological replicates. ***P<0.001 using Student's T-Test for statistics.

3.4.3. MIDORI deficiency dramatically impairs cellular reprogramming

Next, we wanted to address whether MIDORI is required for cellular reprogramming by using a loss-of-function approach. We reprogrammed WT and MIDORI-KO MEFs by transducing them with a doxycycline-inducible lentiviral vector encoding the OSKM factors. We found that MIDORI deficiency delays the appearance of iPSCs colonies (Figure 59A) and dramatically reduces the reprogramming efficiency (Figure 59B-C), suggesting that MIDORI is required for cellular reprogramming.

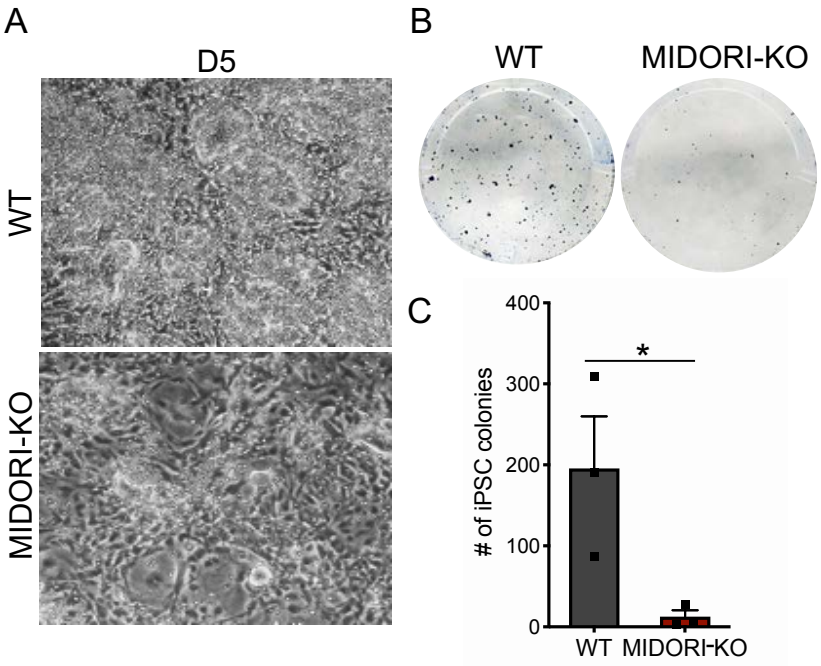


Figure 59. MIDORI deficiency significantly decreases reprogramming efficiency. (A) Representative bright-field microscopy images of the formation of iPSCs colonies derived from control or MIDORI-KO MEFs lentivirally transduced with OSKM. Pictures were taken on day 5 with 20X magnification. (B) Representative alkaline phosphatase (AP) staining of iPSCs colonies derived from WT and MIDORI-KO MEFs at day 12 of reprogramming. (C) Quantification of AP staining performed at day 12 of reprogramming. Each dot represents a single biological replicate. Bars represent the mean ± SD of three biological replicates. *P<0.05 using Student's T-Test for statistics.

We then studied the effect of MIDORI deficiency in reprogramming in a cell extrinsic context. We collected the CM from WT and MIDORI-KO MEFs and used it to reprogramme i4F MEFs. We observed that cells receiving MIDORI-KO MEFs CM show a decreased reprogramming efficiency (Figure 60A-C), although this decrease is not as pronounced as when reprogramming MIDORI-KO MEFs.
Thus, MIDORI deficiency impairs cellular reprogramming, but this effect does not seem to rely only on the secretome.

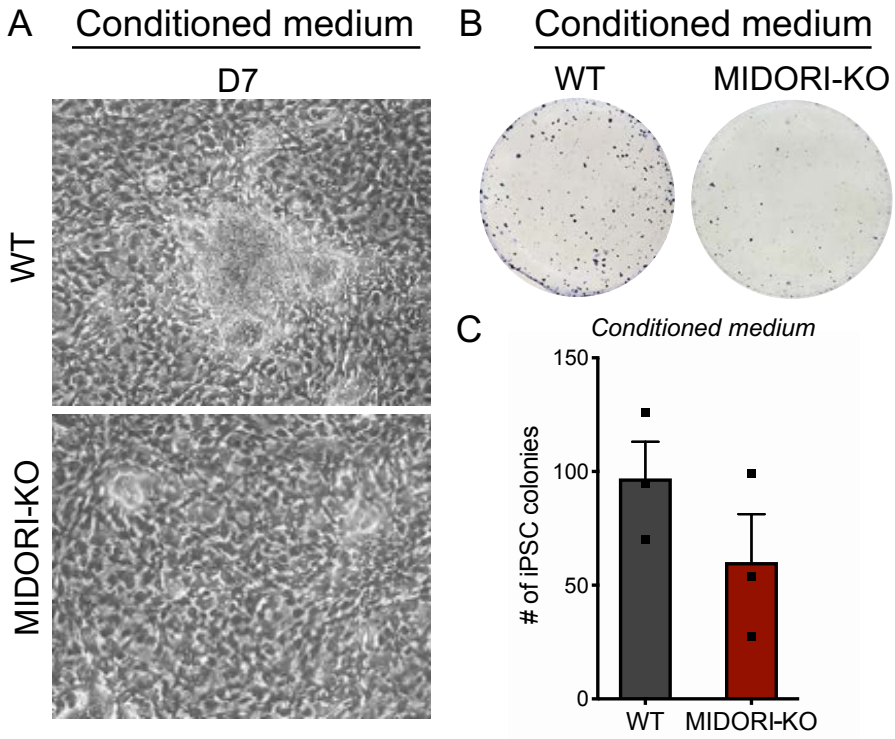


Figure 60. MIDORI-KO cells conditioned medium decreases cellular reprogramming efficiency. (A) Representative bright-field microscopy images of the formation of iPSCs colonies derived from i4F MEFs reprogrammed using CM from MIDORI-KO or WT MEFs. Pictures were taken on day 7 of reprogramming with 20X magnification. (B) Representative alkaline phosphatase (AP) staining of iPSCs colonies generated using the conditioned medium of MIDORI-KO or WT MEFs. The staining was performed at day 12. (C) AP staining quantification performed at day 12 of reprogramming. Each dot represents a single experiment performed with three biological replicates. Bars represent the mean ± SD of three experiments performed with three biological replicates.

Of note, the expression of exogenous MIDORI in MIDORI-KO MEFs rescues the phenotype both, in a cell intrinsic manner (Figure 61A-B) and in a cell extrinsic manner (Figure 61C-D), suggesting that MIDORI overexpression is sufficient to overcome the barrier that prevents MIDORI-KO MEFs to reprogramme, and it restores the secretome which enhances reprogramming in a paracrine manner.

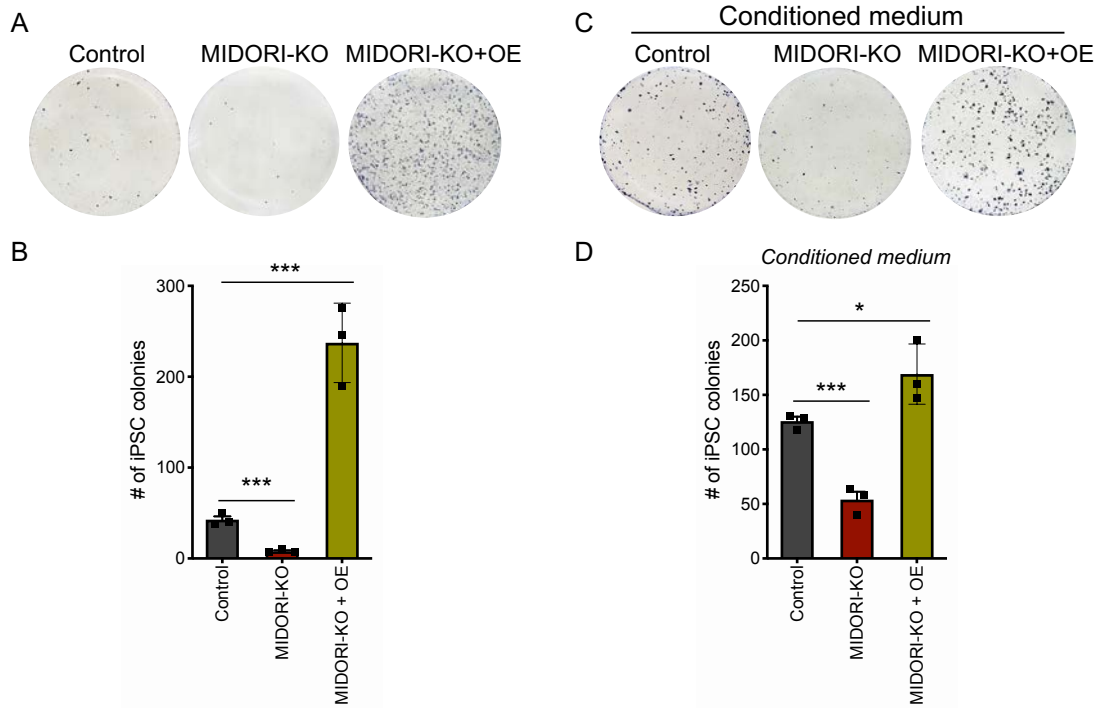


Figure 61. MIDORI exogenous expression in MIDORI-KO MEFs rescues cellular reprogramming in cell intrinsic and cell extrinsic manner. (A) Alkaline phosphatase (AP) staining of iPSCs colonies derived from WT and MIDORI-KO MEFs transduced with a doxycycline-inducible lentiviral vector encoding the OSKM factors together with the murine MIDORI-inducible vector or the empty as control at day 12 of reprogramming. (B) AP staining quantification at day 12 of reprogramming. Each dot represents a single biological replicate. Bars represent the mean ± SD of three biological replicates. ****P*<0.001 using Student's T-Test for statistics. (C) Representative iPSCs colonies positive for alkaline phosphatase (AP) staining in i4F MEFs receiving CM from control, MIDORI-KO and MIDORI-KO overexpressing MIDORI (OE) MEFs at day 12 of reprogramming. (D) AP staining quantification at day 12 of reprogramming. Each dot represents a single biological replicate. Bars represent the mean ± SD of three biological replicates. .**P*<0.05, ****P*<0.001 using Student's T-Test for statistics.

3.4.4. Study of MIDORI's molecular mechanisms in cellular reprogramming

3.4.4.1. MIDORI deficiency dysregulates the expression dynamics of EMT factors and the *Ink4a/Arf* locus during cellular reprogramming

To find out the molecular mechanisms behind the observed phenotype, we analysed the expression of several genes known to act as barriers for cellular reprogramming. We performed a time course reprogramming experiment with WT and MIDORI-KO MEFs in which cells were collected at different time points and processed for transcriptomic analysis. As expected, in WT MEFs the expression of *Snail* and *Slug* was upregulated at the early phase of reprogramming -to undergo an early EMT- and subsequently downregulated to proceed with the reprogramming process (Liu et al., 2013). However, MIDORI-KO MEFs show a significantly higher expression *of Snail* and *Slug* at the intermediate stage (days 6 to 8) of reprogramming compared to WT MEFs (Figure 62A).

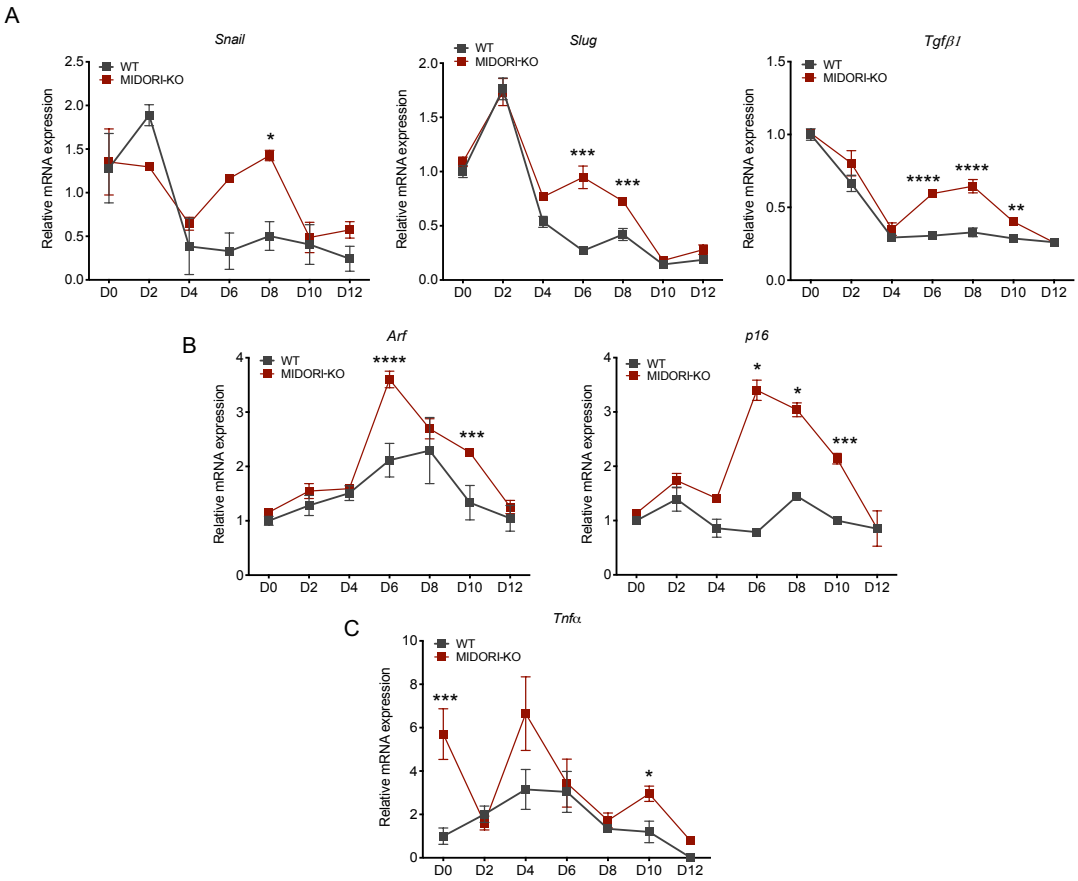


Figure 62. MIDORI deficiency dysregulates EMT dynamics and *Ink4a/Arf* locus expression during cellular reprogramming. (A) *Snail, Slug, and Tgfβ1,* (B) *Arf* and *p16* and (C) *Tnfα* mRNA levels in WT and MIDORI-KO MEFs measured by qRT-PCR at the indicated time points during a time-course reprogramming experiment. Values are represented as mean ± SD of three technical replicates, normalised to *Gapdh* and relativised to day 0. Statistical significance was assessed using Two-way ANOVA. *p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001.

Moreover, we also found that *Tgfβ1*, whose expression is detrimental for cellular reprogramming (Liu et al., 2013), is significantly upregulated at the same days in MIDORI-KO MEFs (Figure 62A). These results indicate that, in the absence of MIDORI, MEFs are not able to downregulate mesenchymal genes properly, suggesting an active role of MIDORI shutting down the expression of EMT-related factors to induce a MET.

On the other hand, it is well established that the activation of the tumour suppressor locus *Ink4a/Arf* represents a cell intrinsic barrier to cellular reprogramming by inducing cellular senescence (Li et al., 2009). We also assessed the expression dynamics of *p19Arf* and *p16Ink4a* during cellular reprogramming. Of note, MIDORI-KO MEFs transiently upregulate these genes at the intermediate phases of the reprogramming process (Figure 62B), pointing to cellular senescence as a potential additional barrier for reprogramming in MIDORI-KO MEFs. Finally, we observed that MIDORI-KO MEFs, already in basal conditions, express high levels of *Tnfα* (Figure 62C), and it has been recently described that high levels of this cytokine are detrimental for reprogramming (Mahmoudi et al., 2019).

Altogether, these results suggest that MIDORI is a key regulator of cellular reprogramming, acting in a cell-autonomous manner by inducing a MET and regulating the expression of tumour suppressor genes (such as *p16* and *p19Arf*), but also acts in a cell extrinsic manner, possibly in part by regulating the expression of *Tnfα* and *Tgfβ1.*

3.4.4.2. MIDORI-deficient MEFs are prone to enter in cellular senescence

Given the significant upregulation of the *Ink4a/Arf* locus detected during reprogramming in MIDORI-KO MEFs (Figure 62B), we tested the effect of MIDORI loss-of-function in cellular senescence. We assessed the induction of senescent cells by staining for senescence-associated-β-galactosidase (SA-β-Gal) activity in WT, MIDORI-overexpressing and MIDORI-KO MEFs in basal conditions and after treating them with Palbociclib, a CDK4/6 inhibitor known to induce cellular senescence (Dickson, 2014).
Of note, we found that MIDORI-KO MEFs cultures show a significantly higher number of senescent cells compared to WT and MIDORI-overexpressing MEFs, in basal conditions and after treatment with Palbociclib (Figure 63A-B). MIDORI overexpression does not seem to have any effect in basal conditions (Figure 63A-B), but we observed a decrease in the number of SA-β-Gal positive cells upon Palbociclib treatment (although not statistically significant).
These data suggest a possible role of MIDORI as a negative regulator of cellular senescence.

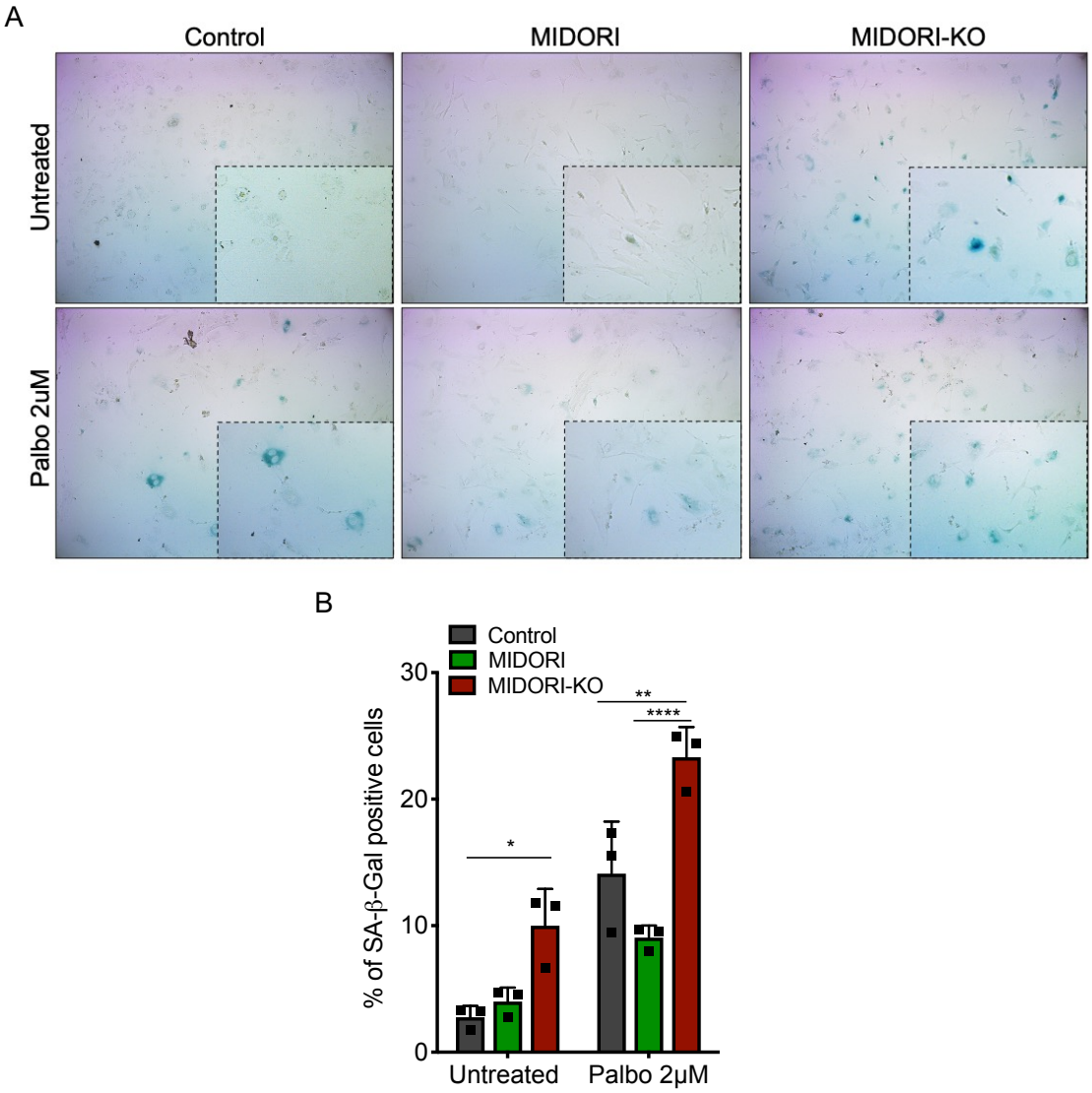


Figure 63. MIDORI deficiency increases senescence in basal conditions and upon Palbociclib treatment. (A) Representative images of SA-β-Gal staining under bright-field microscopy with 10X magnification in the no treated (DMSO, top) and the treated (Palbociclib at 2μm, bottom) condition of WT control cells expressing the empty vector (left), MIDORI-overexpressing (center) and MIDORI-KO (right) MEFs after 7 days. (B) Percentage of the SA-β-Gal positive cells. Each point represents a single biological replicate (n=3). *P<0.05, ** P<0.01 and **** P<0.0001 using two-way ANOVA test for statistics.

3.4.4.3. MIDORI prevents the onset of cellular senescence during reprogramming

We next wanted to address if the increased cellular senescence in MIDORI-KO cells is actually the cause for the reduced reprogramming efficiency. To assess that, we reprogrammed WT, MIDORI-overexpressing, MIDORI-KO and MIDORI KO MEFs overexpressing exogenous MIDORI (MIDORI-KO+OE) and performed SA-β-Gal staining at different timepoints of reprogramming (Figure 64A).

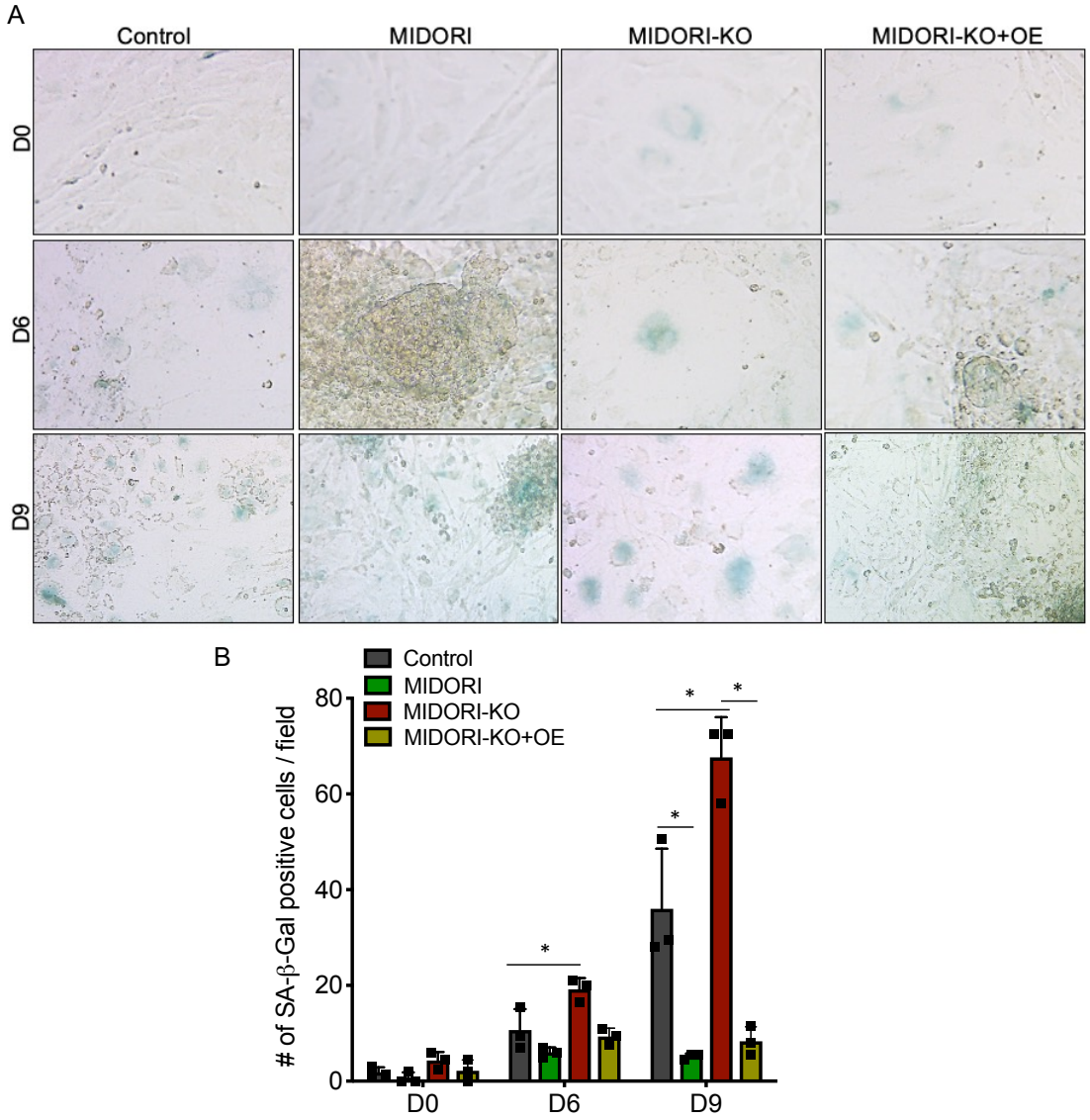


Figure 64. MIDORI deficiency increases senescence during cellular reprogramming while MIDORI overexpression prevents it. (A) Representative images of SA-β-Gal staining of MEFs of the indicated genotypes at day 0, 6 and 9 of reprogramming. WT and MIDORI-KO MEFs were transduced with a doxycycline-inducible lentiviral vector encoding the OSKM factors together with the murine MIDORI-inducible vector or the empty as control. Pictures were taken under bright-field microscopy with 20X magnification and digital zooming. (B) Number of the SA-β-Gal positive cells per field. Each point represents a single biological replicate (n=3). *P<0.05, using Student's T-Test for statistics.

We saw that already at day six of reprogramming MIDORI-KO MEFs show a significant increase in the number of SA-β-Gal positive cells compared to WT and the MIDORI-overexpressing counterparts' (Figure 64A-B). At day nine, when miPSCs colonies were present in all the other conditions, MIDORI-KO MEFs did not form iPSCs colonies and the increase in senescent cells was even more pronounced. In agreement, MIDORI overexpression prevents the onset of senescence in WT MEFs and, remarkably, also in MIDORI-KO MEFs. It is true that at day 9 the cultures (especially MIDORI-overexpressing ones) are mixed populations comprising MEFs, transiently-reprogrammed cells and iPSCs, so results are more difficult to interpret. Anyway, there

is a clear correlation between the induction of cellular senescence and the efficiency of reprogramming.

Concluding, we show that MIDORI regulates the onset of cellular senescence during reprogramming. We suggest that MIDORI prevents the upregulation of *p19Arf* and *p16Ink4a*, thus overcoming cellular senescence and substantially increasing reprogramming efficiency.

Altogether, we hypothesise that MIDORI is transiently induced after the expression of OSKM to prevent the onset of senescence and downregulate the expression of mesenchymal genes, inducing a MET and promoting the transition to iPSCs.

3.4.4.4. Analysis of MIDORI-induced secretome

Finally, to identify the secreted factors that promotes reprogramming in MIDORI-overexpressing MEFs, we concentrated the CM from control and MIDORI-overexpressing MEFs and we analysed them by mass spectrometry.

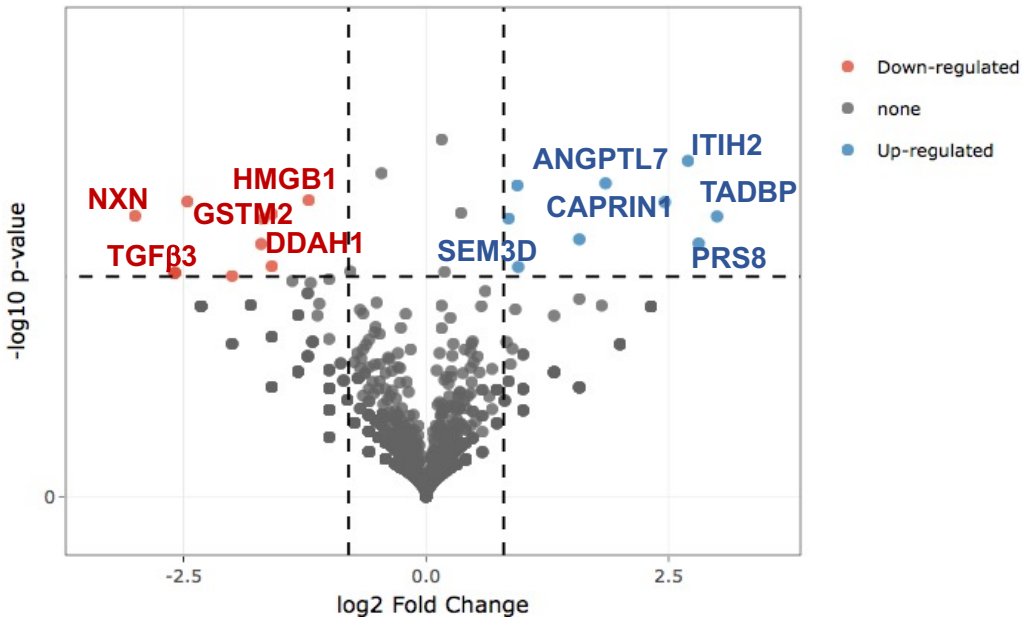


Figure 65. Mass spectrometry analysis of MIDORI-overexpressing MEFs secretome. Volcano plot showing proteins downregulated (red) or upregulated (blue) in MIDORI-overexpressing MEFs secretome compared to the control. The analysis shows proteins downregulated or upregulated that are common for long MIDORI and short MIDORI.

Importantly, we found that TGFβ3, a member of the TGFβ family, is downregulated in the secretome of MIDORI-overexpressing cells (Figure 65). Given the deleterious role of TGFβ cytokines in cellular reprogramming (Liu et al., 2013), MIDORI might be increasing reprogramming, at least in part, by decreasing the secretion of this cytokine. Regarding the upregulated proteins, although none of them have been directly related to cellular reprogramming, ANGPTL7 and ITIH2 look like interesting candidates. Both of them play a role in the formation and organisation of the ECM, being ITIH2 already known as an

anti-metastatic protein (Weidle et al., 2018), and ANGPTL7 is a growth factor important for the expansion of hematopoietic stem cells (Xiao et al., 2015). We are planning to test if blocking TGFβ3 by antibodies or the supplementation with recombinant ANGPTL7 and ITIH2 could mimic MIDORI cell extrinsic phenotype in cellular reprogramming. In addition, we are currently performing an extensive analysis of cytokines and chemokines by Elisa to find other secreted MIDORI effectors.

3.4.5. Short MIDORI recapitulates long MIDORI's phenotype in cellular reprogramming

As we did in previous sections, we performed the most important experiments overexpressing short MIDORI to address its role in cellular reprogramming. As with long MIDORI, we saw that the overexpression of short MIDORI in MEFs increases cellular reprogramming (Figure 66A-B) and, by using the CM, we saw that it does it in a non-cell-autonomous manner (Figure 66C-D). These results confirm that short MIDORI behaves as its longer isoform in cellular reprogramming.

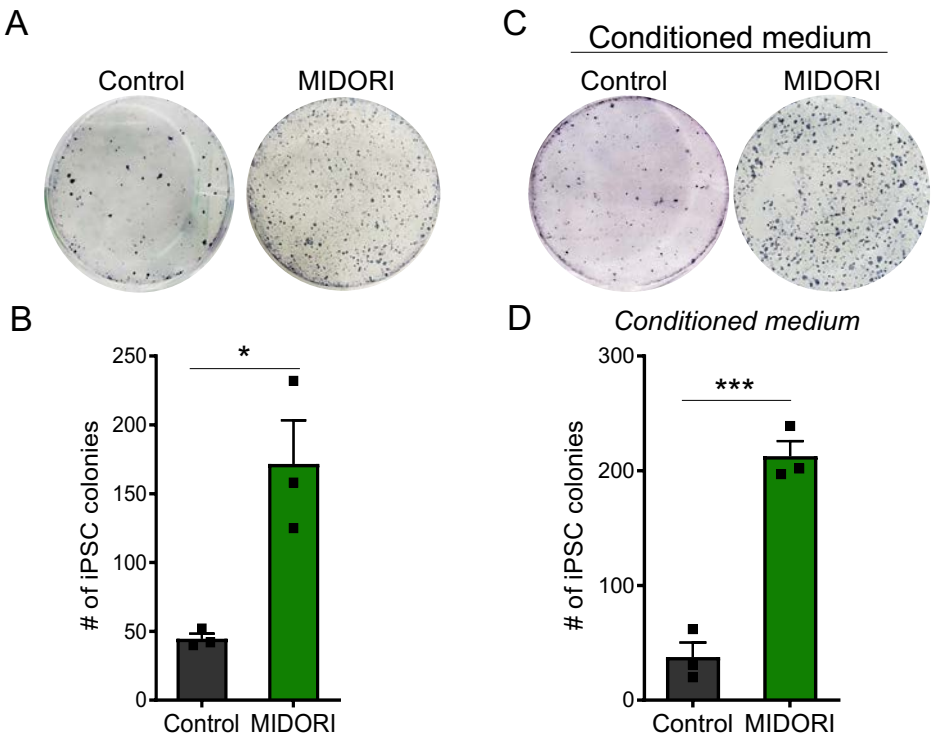


Figure 66. Short MIDORI increases cellular reprogramming efficiency in a non-cell autonomous manner. (A) Representative iPSCs colonies positive for AP staining at day 12 of reprogramming derived from control MEFs transduced with a lentiviral vector expressing the OSKM factors together with an empty vector or short MIDORI construct. (B) AP staining quantification at day 12 of reprogramming. Each dot represents a single biological replicate. Bars represent the mean ± SD of three biological replicates. *P<0.05 using Student's T-Test for statistics. (C) Representative iPSCs colonies positive for AP staining derived from i4F MEFs reprogrammed using short MIDORI-overexpressing MEFs CM or from control cells (expressing the empty vector) at day 12 of reprogramming. (D) AP staining quantification at day 12 of reprogramming. Each dot represents a single biological replicate. Bars represent the mean ± SD of three biological replicates. ***P<0.001 using Student's T-Test for statistics.

Of note, the overexpression of exogenous short MIDORI in MIDORI-KO cells also rescue the reprogramming efficiency, both in cell intrinsic and in a cell extrinsic manner (Figure 67A-D).

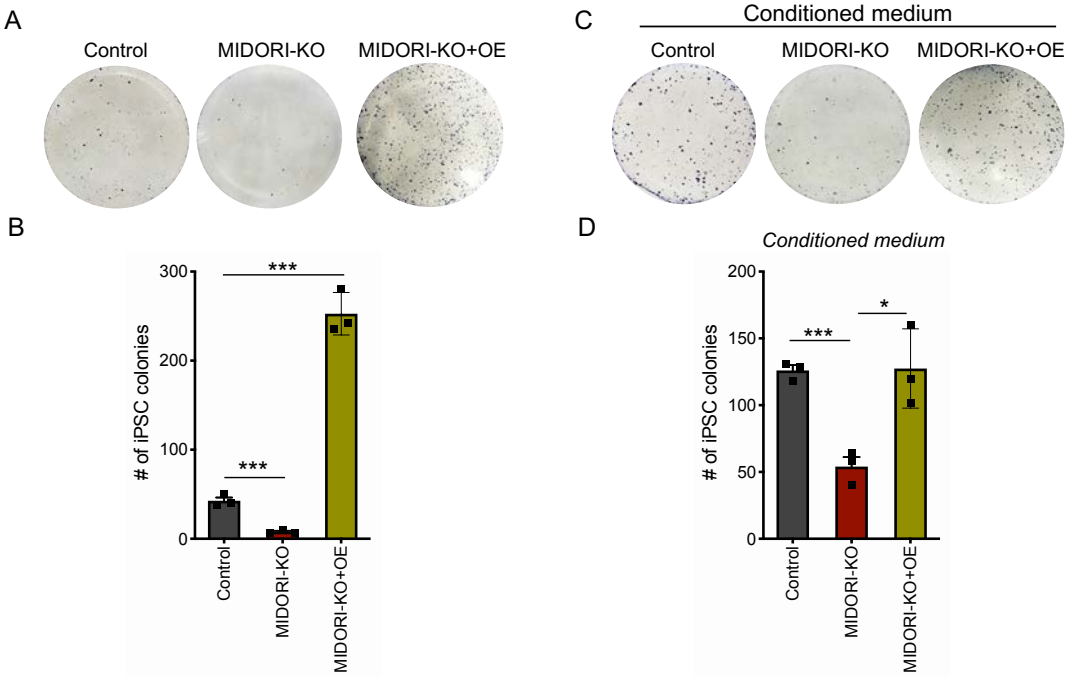


Figure 67. Short MIDORI overexpression in MIDORI-KO MEFs rescues the efficiency in cellular reprogramming. (A) Representative iPSCs colonies positive for AP staining in control, WT and MIDORI-KO MEFs transduced with a lentiviral vector expressing the OSKM factors together with an empty vector or short MIDORI construct at day 12 of reprogramming. (B) AP staining quantification at day 12 of reprogramming. Each dot represents a single biological replicate. Bars represent the mean ± SD of three biological replicates. ****P*<0.001 using Student's T-Test for statistics. (C) Representative iPSCs colonies positive for AP staining in i4F MEFs receiving CM from control, MIDORI-KO and MIDORI-KO MEFs overexpressing short MIDORI (OE) at day 12 of reprogramming. (D) AP staining quantification at day 12 of reprogramming. Each dot represents a single biological replicate. Bars represent the mean ± SD of three biological replicates. .**P*<0.05, ****P*<0.001 using Student's T-Test for statistics.

We then studied the induction of cellular senescence, in basal conditions, upon Palbociclib treatment (Figure 68A-B) as well as during cellular reprogramming (Figure 68C). As observed in the case of long MIDORI, we saw that cells overexpressing short MIDORI show a decreased SA-β-Gal staining upon Palbociclib treatment and are protected from undergoing senescence during cellular reprogramming.

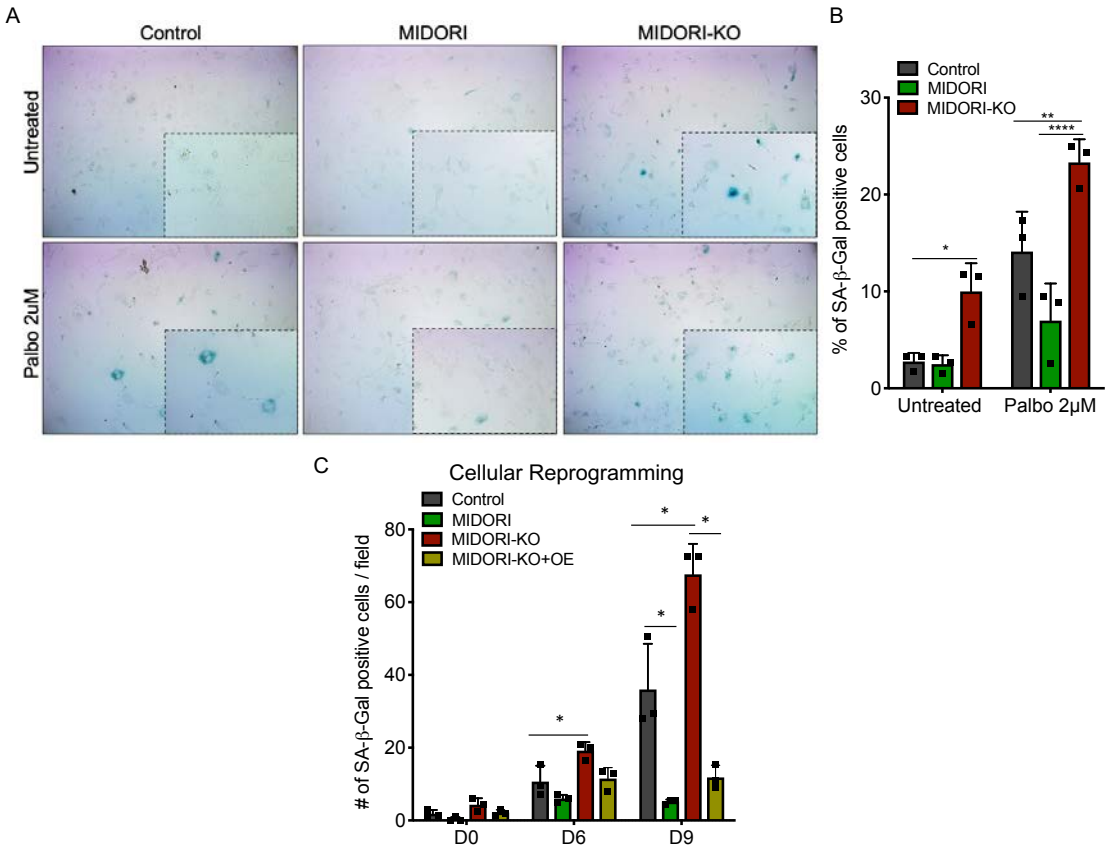


Figure 68. Short MIDORI overexpression does not affect senescence in basal condition and upon Palbociclib treatment but reduces it during cellular reprogramming. (A) Representative images of SA-β-Gal staining in MEFs of the indicated genotypes untreated (DMSO) and treated with 2μm of Palbociclib for 7 days. Pictures were taken under bright-field microscopy with 10X magnification. (B) Percentage of the SA-β-Gal positive cells in the cultures explained in (A). Each point represents a single biological replicate (n=3). **P*<0.05, ** *P*<0.01 and **** *P*<0.0001 using two-way ANOVA test for statistics. (C) Number of the SA-β-Gal positive cells per field at day 0, 6 and 9 of reprogramming of MEFs with the indicated genotypes. WT and MIDORI-KO MEFs were transduced with a doxycycline-inducible lentiviral vector encoding the OSKM factors together with the murine short MIDORI-inducible vector or the empty as control. Each point represents a single biological replicate (n=3). **P*<0.05, using Student's T-Test for statistics.

On top of this, we performed the secretome analysis of short MIDORI-overexpressing MEFs and we observed the same factors up- and down-regulated (Figure 65). Altogether, these experiments let us conclude that the two mouse isoforms of MIDORI behave in the same way during cellular reprogramming.

3.4.6. *In vivo* reprogramming of MIDORI-KO mice

To study the role of MIDORI in reprogramming *in vivo*, we crossed MIDORI-KO mice with the reprogrammable mice (i4F mice) (Abad et al., 2013). We induced *in vivo* reprogramming following two different schemes of doxycycline administration. On the one hand, we added 0.2 mg/ml of doxycycline to the drinking water for 3 weeks, we withdrew the treatment and monitored the mice until they reach the human end-point due to the presence of teratomas. With this experimental setting we observed that teratomas

appeared in different organs, with a similar distribution to that previously described (Abad et al., 2013), and with no obvious differences between i4F and i4F; MIDORI-KO mice. (Figure 69A). However, to our surprise, we saw that MIDORI-KO mice had to be euthanised significantly earlier due to teratomas (Figure 69B), indicating that MIDORI deficiency increases reprogramming *in vivo*.

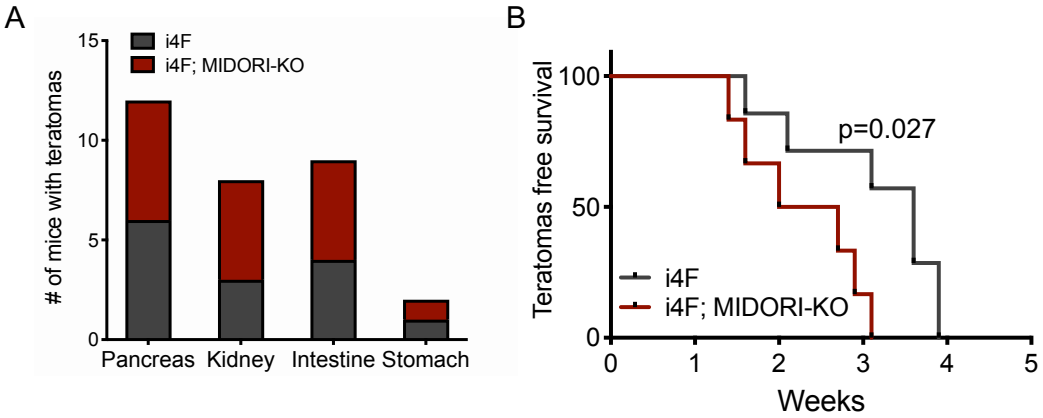


Figure 69. i4F; MIDORI-KO mice show increased reprogramming *in vivo*. (A) Number of mice with the indicated genotype that developed teratoma in the indicated organs. N=6 animals per group. (B) Kaplan-Meier graph showing teratoma free survival of i4F and i4F; MIDORI-KO mice (N=6 mice per group) after inducing reprogramming by adding 0.2 mg/mL of doxycycline in the drinking water for 3 weeks with. P=0.027 using Mantel-Cox test for statistics.

On the other hand, we induced reprogramming by adding 1 mg/ml of doxycycline to the drinking water for 1 week. Right after, we sacrificed the mice and analysed the tissue dysplasia as a readout for *in vivo* reprogramming. We focused on the pancreas since it is one of the organs that is reported to reprogramme the most (Abad et al., 2013). In line with the results of the previous experimental setting, we found that the pancreas of i4F; MIDORI-KO mice present more dysplasia compared to reprogrammed i4F mice (Figure 70A-B), suggesting again that MIDORI deficiency in an *in vivo* context favours reprogramming. Moreover, we also looked at Ki67 positive cells by IHC in the reprogrammed pancreata (Figure 70A) to measure the de-differentiated proliferating cells and we saw a significant increase in MIDORI-KO mice (Figure 70C), further suggesting an increased reprogramming.

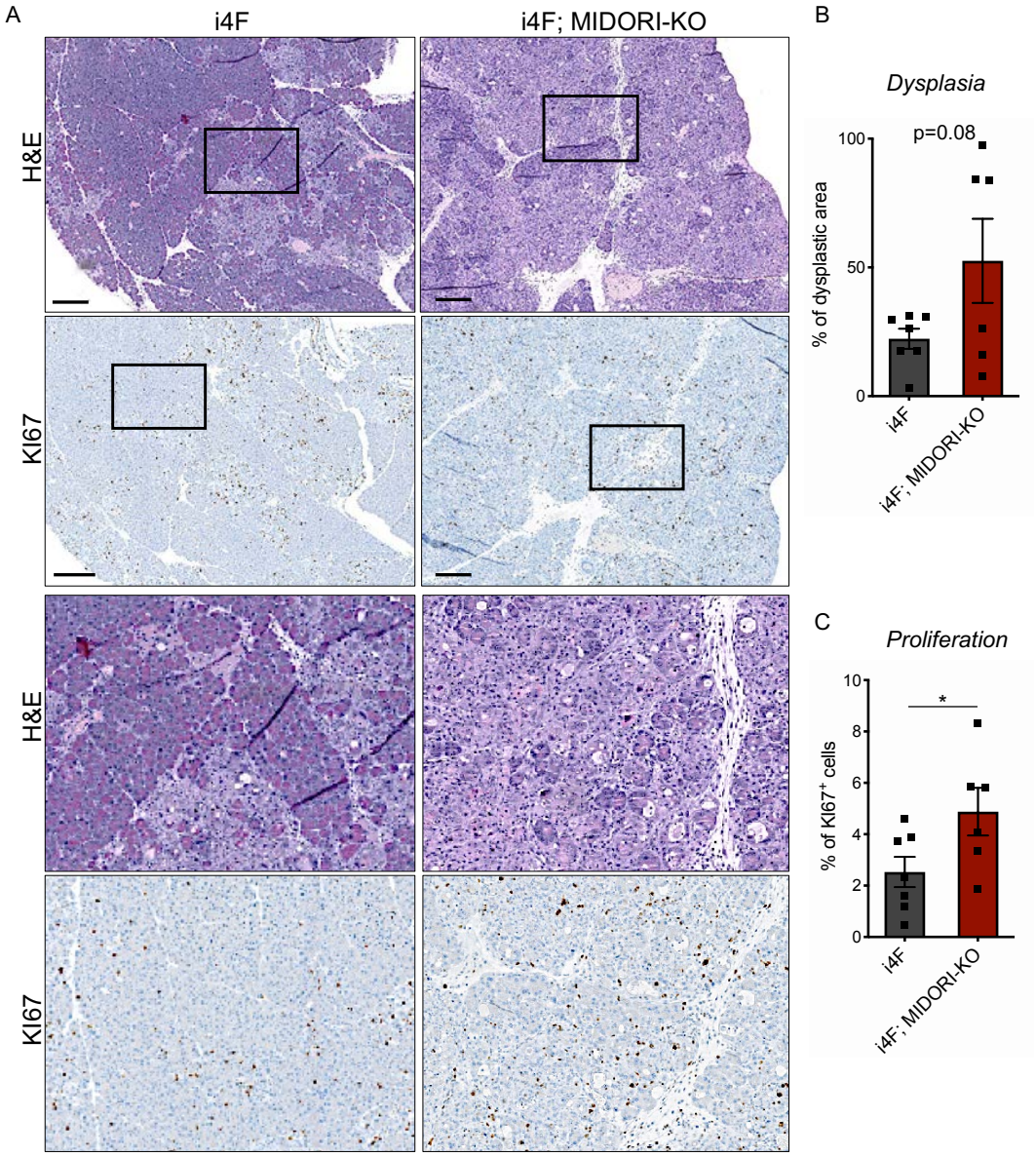


Figure 70. MIDORI-KO mice show more dysplasia and proliferating cells in the pancreas upon *in vivo* reprogramming. (A) (Up) H&E and Ki67 stainings in the pancreas of i4F and i4F; MIDORI-KO mice after reprogramming for one week with 1 mg/ml of doxycycline. Scale bar represents 250μm (Marked areas are blown up in Bottom). (Bottom) Digital magnification of H&E and KI67 representative pictures. (B) Quantification of the dysplastic area (C) percentage of KI67 positive cells in the pancreas of i4F and i4F; MIDORI-KO mice after reprogramming one week with 1 mg/ml of doxycycline. Each dot represents a single animal. Bars represent the mean ± SEM of n=7 mice per group (WT) and n=6 mice per group (MIDORI KO). .*P<0.05 using Student's T-Test for statistics.

As we mentioned before, cellular senescence is a cell intrinsic barrier for reprogramming *in vitro*. However, *in vivo*, the expression of OSKM induces tissue damage and cellular senescence, which provides critical factors for *in vivo* reprogramming (Mosteiro et al., 2016). In fact, in *p16INK4a*-KO mice, whose tissues do not undergo senescence, the reprogramming is impaired (Mosteiro et al., 2016). Since our data show that MIDORI-deficient MEFs undergo an exacerbated senescence upon OSKM, this could be consistent with an increase in *in vivo* reprogramming in MIDORI-KO mice. To further

corroborate our hypothesis, we looked at $p21^{CIP1}$ expression by IHC as a surrogate of senescent cells in the pancreas of MIDORI-KO mice undergoing reprogramming (Figure 71A-B). MIDORI-deficient mice display more $p21^{CIP1}$ positive cells in the pancreas upon *in vivo* reprogramming, which might imply more cellular senescence. We are currently measuring additional senescence markers (including SA-β-Gal staining) to reinforce this result.

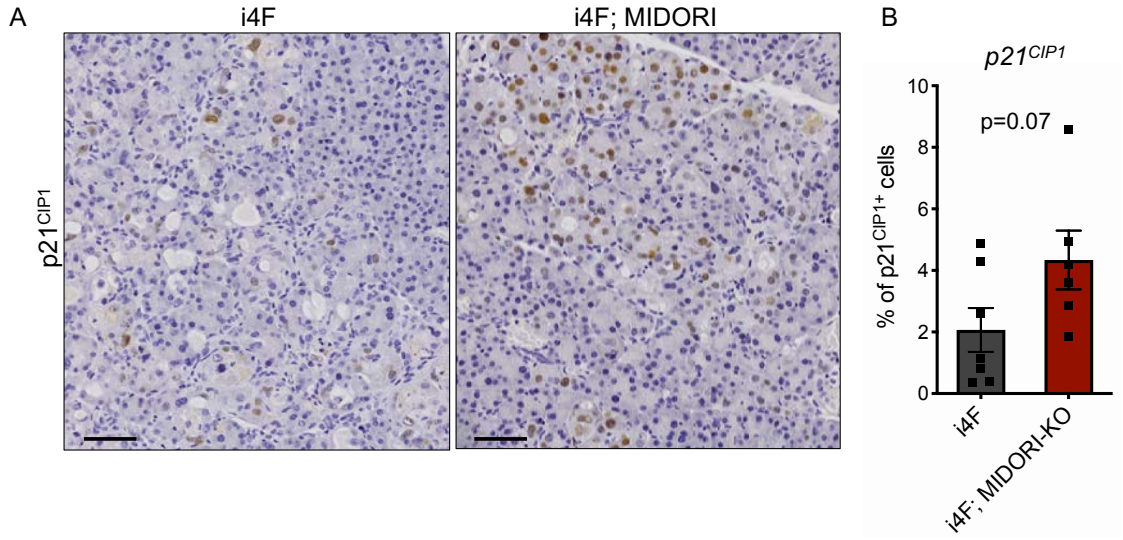


Figure 71. MIDORI-KO mice show more $p21^{CIP1}$ positive cells in the pancreas upon *in vivo* reprogramming. (A) Representative pictures of $p21^{CIP1}$ immunostaining in the pancreas of i4F and i4F; MIDORI-KO mice after reprogramming one week with 1mg/ml of doxycycline. Scale bar represents 100μm. (B) Quantification of the percentage of $p21^{CIP1}$ positive cells in the pancreas of i4F and i4F; MIDORI-KO mice after reprogramming one week with 1mg/ml of doxycycline. Each dot represents a single animal. Bars represent the mean ± SEM of n=7 mice per group (WT) and n=6 mice per group (MIDORI KO). .**P*<0.05 using Student's T-Test for statistics.

In conclusion, while *in vitro* MIDORI deficiency impairs cellular reprogramming by the dysregulation of EMT markers and the induction of senescence, *in vivo* MIDORI promotes cellular senescence and favours reprogramming.
These results are consistent with the observed effect of MIDORI-deficiency upregulating the expression of *p16Ink4a* and *p19Arf* in MEFs, and the fact that p16INK4A-null mice do not undergo senescence and present an impaired *in vivo* reprogramming. As a next step we will test the expression of *p16Ink4a* and *p19Arf* in MIDORI-KO mice undergoing reprogramming.

Despite more research is needed to ascertain the context-dependent functions of MIDORI and the intricacies of its molecular mechanisms, our results suggest that MIDORI is a critical regulator of cellular plasticity that controls the transitions between mesenchymal and epithelial phenotypes and prevents the onset of cellular senescence. Thus, MIDORI emerges as a new key player of cellular identity, which may have important implications in cancer and regenerative medicine.

DISCUSSION

Historically, genome annotation has relied on several criteria that were set to define what an ORF was. Among them, it was assumed that ORFs should be shorter than 100 codons, limited by conventional start and stop codons, and that eukaryotic transcripts are monocistronic (Basrai et al., 1997). All these premises helped the gigantic task of annotating whole genomes. However, they extremely restricted our view of what should be considered meaningful in the genome. In fact, only a small part of the genomes was found to be coding, whereas the rest was considered as "junk DNA". Fortunately, this assumption has changed over the years. In fact, recent advances in techniques such as ribosome profiling (Ribo-seq) and mass spectrometry (MS) have revealed that many previously misannotated non-protein-coding transcripts actually code for evolutionarily conserved microproteins (Makarewich and Olson, 2017). To date, many microproteins have been identified in mammals and they been revealed as a new class of fine-tuning molecular regulators with crucial functions in many cellular processes (Makarewich and Olson, 2017; Merino-Valverde et al., 2020).
The hypothesis that motivated this work is that microproteins can be novel regulators of cellular plasticity and cancer. In this thesis, we have focused on finding novel microproteins coded by lncRNAs as a source of novel regulators of cancer cell plasticity. We have identified MIDORI, a novel microprotein that regulates the transition between mesenchymal and epithelial phenotypes. We have found that MIDORI is translated upon stress and acts as a negative regulator of the EMT programme affecting the expression of EMT-related genes, cell motility and metastatic properties both *in vitro* and *in vivo*. In agreement, we have shown that MIDORI play a role also in cellular reprogramming, a process known to require a MET. Thus, we propose that MIDORI is a novel microprotein encoded by a gene annotated as noncoding which acts as a negative regulator of epithelial-mesenchymal plasticity.

1. Identification of MIDORI microprotein

To identify novel microproteins with a role in cancer cell plasticity, we generated a list of lncRNAs deregulated in cancer and in cellular differentiation-dedifferentiation processes and analysed their coding potential (Figure 10A). Using PhyloCSF algorithm, we identified *ZEB2-AS1* as a possible transcript misannotated as a long non-coding RNA that, in fact, contains a previously uncharacterised sORF that we named MIDORI (Figure 10A). Our phylogenetic analysis showed that MIDORI is conserved across placental mammals. However, in primates occurred a single nucleotide deletion that led to a frameshift and changed the amino acid sequence of the C-terminal part of MIDORI (Figure 10B). In humans, MIDORI sORF produces a microprotein that is totally conserved through primates (Figure 11A). Importantly, although the ORF does not start in the first lncRNA exon, MIDORI's start codon is the first ATG of the transcript. In mouse, the orthologue locus (*Zeb2os*) produces two main splicing isoforms that differ in the first exon (Figure 11B) and could encode two MIDORI protein isoforms: short MIDORI and long MIDORI (Figure 11B). Long MIDORI contains short MIDORI, and both isoforms are highly conserved in placental mammals lower than primates (Figure 11B). Despite these differences in the sequence, our experiments show that human MIDORI have the same function than both mouse MIDORI isoforms, suggesting that the N-terminal, which is

conserved in placental mammals, is the most important part for the function of the protein.

To experimentally validate the coding potential of *ZEB2-AS1*, we analysed publicly available Ribo-seq data and performed *in vitro* translation of the lncRNA. Importantly, those analyses provided us experimental evidence of the lncRNA translation in mouse (Figure 14A-B) and point to *Zeb2os* isoform 1 and long MIDORI as the prevalent isoform. Supporting this notion, we found that *Zeb2os* isoform 1 is the most upregulated isoform upon TGFβ treatment (Figure 14C). Nevertheless, we could not exclude that *Zeb2os* could be upregulated as well upon other stimuli.
We also analysed publicly available mass spectrometry data to find further experimental evidence of MIDORI translation, but we were not able to identify MIDORI with this approach. It is worth mentioning that MIDORI's sequence includes many arginine and lysine residues, targets of trypsin digestion, which is the first step in mass spectrometry analysis (Laskay et al., 2013). Therefore, it could be that the low sensitivity of the technique together with the lack of available tryptic peptides makes MIDORI identification really difficult by this method.

1.1. Analysis of *ZEB2-AS1* and MIDORI expression

Although many lncRNAs are expressed in a tissue-specific manner (Cabili et al., 2011), we found that *ZEB2-AS1* and its mouse orthologue do not show a clear expression pattern in tissues, where a basal, low level of expression can be detected in many organs (Figure 20A-B). Interestingly, by using a custom-made antibody we found endogenous MIDORI to be naturally expressed in kidney (Figure 21A-B). Given that *Zeb2os* expression in kidney is very low, it seems that MIDORI translation is uncoupled from the lncRNA transcription, as it has been already described for other genes (Koussounadis et al., 2015).
Remarkably, we found that MIDORI expression is upregulated upon TGFβ and different stresses such as doxorubicin treatment, which induces genotoxic stress, and upon tunicamycin-induced ER stress (Figure 22B-C). Of note, TGFβ treatment and the subsequent induction of the EMT programme are known to induce ER stress and to activate the Unfolded Protein Response (UPR) (Liu et al., 2019; Maiers et al., 2017). Hence, as previously described for other microproteins (Chu et al., 2019; Matsumoto et al., 2017), MIDORI protein translation might be induced upon stress as a mechanism for the cell to cope with damage. Although the regulation of MIDORI expression needs to be further explored upon other types of stimuli, our data suggest that MIDORI could be a novel microprotein involved in the stress response.

1.2. MIDORI subcellular localisation

In silico analyses predicted that hMIDORI and the two mMIDORI isoforms are nuclear microproteins (Figure 15B). In line with this, a NLS is found to be conserved in the three microproteins (Figure 15C), further suggesting a nuclear localisation. When ectopically overexpressed in different cell lines, both hMIDORI and the two mMIDORI are detected both in the nucleus and the cytoplasm by immunofluorescence (Figure 16C-D).

Accordingly, subcellular fractionation showed hMIDORI's distribution in the cytosolic, nuclear and chromatin-bound fractions (Figure 16E). However, the forced overexpression could give a false hint about MIDORI's localisation, given that overexpressed proteins often passively diffuse to every cell compartment because of their abundance. Thus, we looked at the subcellular localisation of endogenous MIDORI (Figure 23). Although we saw nuclear staining, MIDORI was predominantly detected in the cytoplasm, where it shows a punctuated perinuclear pattern that in some cells acquire a spheroidal shape (Figure 23B). Those spheroids do not colocalise with lysosomes or with the Golgi apparatus (Figure 24). Although further experiments are needed to address if MIDORI could colocalise with the ER or with mitochondria, given that MIDORI is induced by stress and interacts with many ribosomal proteins (Figure 42C-D), it is possible that these aggregates are stress granules (see further discussion in *Section 2.4*). Of note, by subcellular fractionation and Western blot analysis MIDORI was detected mainly in the nucleus (Figure 25A), in agreement with the *in silico* prediction. Moreover, endogenous MIDORI can be detected in the nucleus by immunofluorescence after pre-treatment to reduce cytosolic background (Figure 25B), further confirming this last result.
Thus, our experimental data suggest that MIDORI localisation is not restricted to the nuclear compartment. Indeed, we have shown that MIDORI interacts with proteins whose localisation is not restricted to the nuclear compartment (Figure 42C-D). Therefore, we conclude that MIDORI localises both in the nuclear and cytoplasmatic compartment. It is possible that MIDORI shuttle from the nucleus to the cytoplasm and vice versa in response to certain cues, in order to interact with different proteins and, in turn, carry out its molecular function.

1.3. Short mouse MIDORI and long mouse MIDORI

Our experimental evidences (Figure 14C) suggest that *Zeb2os* isoform 1 and subsequently long MIDORI is the predominant expressed isoform. On the one hand, we cannot exclude that *Zeb2os* isoform 2 might be the preferentially expressed isoform upon cues different from TGFβ (i.e., genotoxic or ER stress). Additionally, although Ribo-seq analysis provided us an essential piece of data, we could not conclude whether short MIDORI can be translated independently or just as part of long MIDORI. It is important to mention that both long and short MIDORI have a good Kozak's sequence around their starting sites, meaning that they can both provide the right context for ribosomes to begin the translation. Indeed, by performing *in vitro* translation we observed two bands that, according to their molecular weight, could correspond to long MIDORI and short MIDORI, further suggesting that the translation can start at both methionines (Figure 14B). Finally, in kidney MIDORI is detected by Western blot as a doublet (figure 21A). Even though we cannot exclude that the additional band corresponds to post-translational modifications of MIDORI, it could be that both isoforms are produced or, alternatively, in different tissues and upon different cues one isoform could be more expressed than the other.
Importantly, our results in *Section 2.5* and *4.5* showed that the exogenous expression of long MIDORI and short MIDORI have the same effect. Thus, despite long MIDORI having an extended N-terminal, our data suggest that MIDORI's function could rely on the conserved core of amino acids that are also shared with the human MIDORI orthologue.

Additionally, it is possible that long MIDORI is cleaved in the cell and the active form of the microprotein is represented by short MIDORI.
Concluding, although other studies are needed to decipher the relative relevance of the two isoforms, our evidences demonstrated that MIDORI's function is conserved across evolution, despite the change in its aminoacidic sequence.

2. Functional characterisation of MIDORI in EMT

The expression of *ZEB2-AS1* correlates with cancer progression (Lan et al., 2016; Zhuang et al., 2015) and poorer prognosis in breast cancer, PDAC and ovarian cancer (Figure 12B), in line with its role as an effector of the EMT (Beltran et al., 2008). Thus, we started to investigate MIDORI's function in EMT. To study the function of the microprotein independently of the lncRNA's function, we used a gain-of-function approach in which only MIDORI sORF - a small portion of *ZEB2-AS1*- is exogenously expressed (Figure 16A). . Since the activity of lncRNAs relies on their molecular structure (Graf and Kretz, 2020; Zampetaki et al., 2018), it is unlikely that the RNA transcribed from the sORF could be mimicking the entire lncRNA, and we consider that his strategy allowed us to uncouple the function of the lncRNA from the function of the microprotein.

2.1. MIDORI is a negative regulator of the mesenchymal programme

Surprisingly, when we ectopically expressed MIDORI in human and mouse primary and cancer cell lines, we saw a general downregulation of EMT-related genes (Figure 26) such as EMT transcription factors, ECM protein and remodelling factors, and the epithelial marker E-CADHERIN was found to be upregulated in breast cell lines. Moreover, we observed that MIDORI overexpression does not affect cell proliferation (Figure 27) but it reduces cell migration, invasion (Figure 28) and matrix degradation capacity (Figure 29). Importantly, MIDORI overexpression not only downregulates EMT-related markers and features, but seems to trigger an "epithelialisation" of highly mesenchymal cells. We observed that triple negative breast cancer MDA-MB-231 cells acquire a different morphology upon MIDORI overexpression and start to form clusters (Figure 30A) while re-expressing β-CATENIN at the cell membrane (Figure 30B-C). Although we could not observe E-CADHERIN at the cell membrane of MIDORI-overexpressing MDA-MB-231 cells, its upregulation at the transcriptional level suggests a reduction in the epigenetic silencing of this locus. In fact, it is known that E-CADHERIN locus becomes hypermethylated during EMT and in highly mesenchymal cells like the MDA-MB-231 cells (Herranz et al., 2008; Lin et al., 2010). Thus, our data indicate that the overexpression of MIDORI for just four days induces profound transcriptional changes in metastable mesenchymal cells that, in turn, start the transition towards a more epithelial phenotype.
Then, we wondered whether MIDORI overexpression was sufficient to prevent the acquisition of mesenchymal features when cells are forced to undergo an EMT. Again, we saw that MIDORI overexpression prevents the upregulation of EMT genes upon TGFβ treatment (Figure 31), decreases cell invasion (Figure 34) and prevents the loss

of E-CADHERIN and β-CATENIN at the adherent junctions (Figure 32-33) in NMuMG cells. In agreement with the results observed *in vitro*, we demonstrated that MIDORI overexpression decreases breast cancer metastasis *in vivo,* by using a model of lung colonisation (Figure 47-48) and an orthotopic model of metastatic colonisation after primary tumour removal (Figure 49). While we did not see any effect on metastasis tropism (since both MIDORI-overexpressing and control MDA-MB-231 cells seem to colonise the same organs) (Figure 49B), we did observe a significant decrease in the number of organs with metastasis and a reduction in the metastasis per organ (Figure 49C y D). Thus, MIDORI does not affect the tropism of metastatic cells but affects their metastatic potential.

Mechanistically, we observed that MIDORI overexpression not only decreases SNAIL protein level (Figure 37A), but also decreases the activation of the canonical TGFβ pathway branch, regulated by the phosphorylation of SMAD2/3 (Vincent et al., 2009) and the non-canonical branch, regulated by the phosphorylation of ERK1/2 (Gingery et al., 2008) (Figure 37A). Additionally, we found that MIDORI overexpression decreases the phosphorylation of NF-κB in senescent cells (Figure 35C) and its acetylation upon TNFα treatment (Figure 45C). Although we have not studied yet whether MIDORI decreases NF-κB activation upon TGFβ, our results might indicate that MIDORI negatively regulate EMT independently from the activation stimulus by decreasing the activation level of NF-κB.

Concluding, our results suggest a new mechanism of regulation of the mesenchymal program mediated by MIDORI. While *ZEB2-AS1* is upregulated upon the activation of the TGFβ pathway to promote EMT (Beltran et al., 2008), our data strongly support an opposite effect of its encoded microprotein. Thus, MIDORI might be the effector of a negative feedback loop that shuts down the mesenchymal programme to revert EMT (Figure 37B).

2.2. MIDORI cell extrinsic effect

Once we studied the role of MIDORI in a cell intrinsic context, we sought to investigate if the inhibition of the mesenchymal programme may also impact the secretion of factors important for EMT induction in the neighbouring cells. We specifically looked at the expression of proinflammatory cytokines and chemokines known to be important EMT boosters secreted by mesenchymal cells (Coppe et al., 2010; David and Massague, 2018; Kumari et al., 2016). Of note, MIDORI overexpression not only downregulates cytokines expression in basal conditions, but also upon TGFβ treatment (Figure 35A) and in therapy-induced senescent cells (Figure 35B), a paradigmatic example of secretory cells in which cytokines are upregulated as part of the SASP. These results could indicate a role of MIDORI as a senomorphic agent. Senomorphics are agents that can modulate the phenotypes of senescent cells through interfering with the SASP without the induction of apoptosis (Gasek et al., 2021). In line, MIDORI overexpression decreases the phosphorylation of NF-κB in senescent cells (Figure 25C), which is the main transcription factor in charge of SASP induction (Salminen et al., 2012).

Finally, we studied if the reduced cytokine expression might have any functional effect. On the one hand, we could not observe any difference in the invasion of MDA-MB-231 cells when MIDORI-expressing MDA-MB-231 CM was used as an attractant (Figure 36A-C). However, it is possible that MDA-MB-231 cells, which are highly invasive, might be insensitive to cytokines fluctuations in the CM. Indeed, we should check cytokines levels in the CM to ensure that the decreased mRNA expression is reflected in a decreased secretion of those cytokines. As a next step, given the role of senescent cancer cells in secreting factors important for cancer progression (Coppe et al., 2010), it would be interesting to evaluate the effect of MIDORI-overexpressing senescent cells CM on invasion.
On the other hand, we observed that the CM of MIDORI-overexpressing CAFs is less proficient at inducing cancer invasion in a paracrine manner (Figure 36D-F). Importantly, we saw that MIDORI overexpression decreases the level of the fibroblast activation marker α-SMA (Figure 26 and 38A). This piece of data suggests that MIDORI-overexpressing CAFs could be less activated and, thus, less secretory. This evidence, together with the decreased expression of ECM proteins and remodelling factors that we observed upon MIDORI overexpression (Figure 26 and 38A), could explain the reduced invasion of cells exposed to the CAFs CM.

Concluding, we have demonstrated that MIDORI overexpression affects the cell secretome. However, the functional effect of MIDORI-derived secretome seems to be context-dependent. Further studies are needed to totally address the cell extrinsic effect of MIDORI in EMT. As next step, we want to expose different cancer cell lines to MIDORI-overexpressing cells CM during short periods (24 hours) or long periods of time (4 days to one week). By treating during short periods, we plan to investigate rapid changes in the activation of TGFβ pathway effectors in the recipient cells. By contrast, treating the recipient cells for longer period of time, we hope to decipher if MIDORI-derived secretome could “educate” the cells and eventually induce long term transcriptional changes in a cell extrinsic manner.

2.3. Effect of MIDORI deficiency in cancer, wound healing and fibrosis

Although we have not been able to analyse the effect of MIDORI deficiency in metastasis yet (Figure 52A-B), our *in vitro* data show that cancer cells isolated from the MMTV-PyMT; MIDORI-KO mice migrate faster compared to their WT counterpart (Figure 53). In line, MIDORI-KO mice heal faster compared to WT mice in an assay of *in vivo* wound healing (Figure 54A-B). Interestingly, differences in the wound closure between WT and MIDORI-KO mice are displayed already in the first days post-wound and then are maintained until the wound is closed. It is known that the process of *in vivo* wound healing is characterised by four different phases, which may overlap in part: the haemostasis phase, the inflammatory phase, the proliferation phase and the maturation and remodelling phase. In particular, after the wound stops bleeding in the haemostasis phase, the production of proinflammatory cytokines and growth factors is induced. Proinflammatory cytokines induce an EMT, allowing the cells at the edge of the wound to migrate and re-establish the normal tissue architecture and the growth factors favour

the proliferation of keratinocytes in order to reform the multiple layers of the skin (Shaw and Martin, 2016). Thus, if the lack of MIDORI could make the cells more prone to undergo EMT and may potentially upregulate the expression of proinflammatory cytokines needed in the process of tissue repair, promoting a faster wound healing.
Nonetheless, some EMT-healing responses may also be deleterious, such as the exacerbated healing responses that lead to fibrosis and scarring. This is the case of kidney fibrosis, where kidney tubular cells undergo a partial EMT and, in a paracrine manner, contribute significantly to fibrogenesis and inflammation (Nieto et al., 2016). However, we did not observe significant differences either in the transcription of EMT-related genes, proinflammatory cytokines (Figure 55B-C) or in the fibrotic tissue deposition in our MIDORI-KO mice compared to the WT mice upon folic acid-induced kidney injury (Figure 55D-E). On one hand, the induction of fibrosis with this protocol is very fast and it could be worth investigating more intermediate timepoints rather than late timepoints, when the fibrotic process already occurred and, perhaps is more difficult to see significant differences. It is also possible that MIDORI does not participate in kidney fibrosis; the partial EMT induced by folic acid does not provide the right signals for MIDORI expression (i.e., the accumulation of stress in the cells undergone EMT) and, therefore, its deficiency does not affect the fibrotic onset.

2.4. Molecular mechanisms behind the role of MIDORI in EMT

To understand the molecular mechanism underlying MIDORI's phenotype, we performed a MIDORI pull-down followed by mass spectrometry analysis to identify MIDORI's interactome. Surprisingly, Gene Ontology analysis of MIDORI's interactome showed an enrichment in GO terms related with rRNA processing, ribosome biogenesis and translation, as well as with ribosomal or nucleolar localisation (Figure 42C-D). Importantly, these data could fit with MIDORI's subcellular localisation. As shown in Figure 16C-D, exogenously expressed MIDORI is found both in the nucleus and cytoplasm. Additionally, Figure 23 shows endogenous MIDORI as predominantly detected in the cytoplasm, where it displays a spheroidal pattern, like in aggregates. Given that we did not observe any colocalisation of MIDORI either with Golgi apparatus or lysosomes (Figure 24) and we showed that MIDORI possibly interacts with many ribosomal proteins, we hypothesise that that MIDORI could localise in stress granules. It is known that upon several stresses, ribosomes, normally aggregated in polysomes during translation, are disassembled and converge in structures known as stress granules (SGs) (Anderson and Kedersha, 2008). SGs are non-membranous foci that are composed by 40S ribosomal subunits, translation initiation factors, poly(A)+ mRNAs and RNA-binding proteins (RBPs). Importantly, SGs have been demonstrated to regulate gene expression by controlling the processing, sequestering and/or degradation of specific RNA transcripts (Decker and Parker, 2012; Mahboubi and Stochaj, 2017). Thus, given that MIDORI expression is normally induced under stress conditions (Figure 22) and it interacts with many proteins that are sequestered into stress granules upon stress, it is possible that the regulation of EMT-related genes upon MIDORI overexpression could be mediated, at least in part, in the cytosolic compartment via the formation of stress granules. Although this an interesting hypothesis, additional studies are needed to define if this could actually be a good model to explain MIDORI's phenotype in EMT.

We validated some interactors, and we mostly focused on candidates that could help explaining the functional changes that we see upon MIDORI overexpression.

Given the role of MIDORI in inhibiting invasion and matrix degradation (Figure 28-29), we validated ITGB1BP1 (Figure 43A). As described in *Results*, this protein is an inhibitor of the Rho GTPase CDC42 by blocking its binding to GTP (Degani et al., 2002). This way, it negatively regulates the formation of focal adhesion and, consequently, cell spreading on the ECM and cell migration (Bouvard et al., 2003; Degani et al., 2002). Thus, if MIDORI enhances ITGB1BP1's function, this could affect the interaction with the ECM, its degradation and eventually cell migration and invasion. To further study this possibility, it would be interesting to perform CDC42 pull down studies upon MIDORI overexpression to specifically look at its activation status.

The second validated candidate is HMGA1 (Figure 44A). This is a chromatin-associated protein involved in the regulation of gene transcription in many cellular processes (Cleynen and Van de Ven, 2008). Of note, HMGA1 can also be secreted, for example upon TGFβ signalling (Zhong et al., 2017), and it has been demonstrated to increase cell invasion and breast cancer metastasis (Mendez et al., 2018). Remarkably, we found that while HMGA1 is increased in the chromatin fraction upon MIDORI overexpression, it decreases in the secreted fraction (Figure 44B). While the decrease in HMGA1 secretion is in line with MIDORI's role reducing breast cancer invasion and metastasis, further analyses are needed to test if the increased HMGA1 bound to the chromatin has any effect on gene expression. For instance, HMGA1 acts as an oncogene in cancer being involved in the transcriptional regulation of genes related with proliferation, inflammation and metastasis (Wang et al., 2019) by the formation of the "Enhanceosome" (Benecke and Eilebrecht, 2015). Moreover, it has been recently reported that a microprotein called SEBHP acts as a transcriptional regulator by interacting with HMGN1 and HMGN3 (Koh et al., 2021). Therefore, since MIDORI interacts with a protein from the same family and induces transcriptional changes upon its overexpression, it is worth investigating if MIDORI interaction with HMGA1 could also affect its role in gene transcription, as it does with its secretion.

Last candidate we validated is MYBBP1A, the top one in our list according to its FC and BDFR (Figure 45A). This is a corepressor that upon stress shuttles from the nucleolus to the nucleus where it acts inhibiting a plethora of transcription factors (Felipe-Abrio and Carnero, 2020). Similarly to MYBBP1A's role in inhibiting rRNA transcription, we saw a significant decrease in rRNAs transcription when overexpressing MIDORI upon TGFβ treatment (Figure 45B). Furthermore, it is known that rRNA transcription is increased during EMT (Prakash et al., 2019). Thus, MIDORI might be affecting EMT by decreasing rRNA transcription and ribosome biogenesis via MYBBP1A. Finally, it is known that MYBBP1A reduces NF-κB transcription activity by competing with p300 for its acetylation, reducing in turn NF-κB acetylation (Owen et al., 2007). Importantly, we observed a decreased acetylation of NF-κB upon MIDORI overexpression (Figure 45C). This result, together with the decrease in cytokines expression (Figure 35A-B) upon MIDORI overexpression, suggest that MIDORI may be acting as a negative regulator of NF-κB through MYBBP1A. To test this hypothesis, on the one hand, we are performing

ChIP experiments to mechanistically investigate if less acetylated NF-κB is bound at the promoters of proinflammatory cytokines upon MIDORI overexpression. On the other hand, we are performing pull-down experiments to determine if, upon MIDORI overexpression, NF-κB binding to p300 is decreased. We hope that these experiments will help us to understand the molecular mechanisms by which MIDORI regulates EMT.

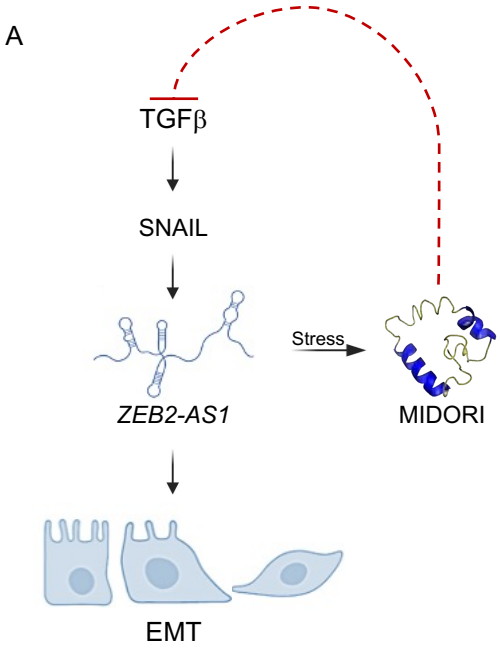
2.5. Proposed model for the role of MIDORI in EMT

Summarising the data obtained in this first part, we have described MIDORI as a novel microprotein encoded by *ZEB2-AS1*. Endogenous MIDORI is translated upon TGFβ as well as upon several stress stimuli (Figure 22). Importantly, MIDORI overexpression induces the downregulation of EMT-related genes (Figure 26) and proinflammatory cytokines (Figure 35A-B). In line with its transcriptional effect, we have demonstrated that MIDORI-overexpressing cells migrate and invade less *in vitro* (Figure 28) and show a decreased metastatic potential *in vivo* (Figure 47 to 49). Moreover, we have observed that MIDORI exogenous expression inhibits the mesenchymal programme also when cells are forced to undergo EMT by TGFβ treatment (Figure 31 to 34). Indeed, we have demonstrated that both the canonical and non-canonical branches of TGFβ pathway are less activated upon MIDORI overexpression (Figure 37). Mechanistically, we have validated three interactors that could help explaining at the molecular level MIDORI's phenotype in EMT at the molecular level (Figure 43 to 45). Especially, the transcriptional repressor role of MYBBP1A could represent the missing link explaining the negative regulation of EMT-related genes and proinflammatory cytokines upon MIDORI overexpression.

Based on the evidence we have so far, we propose a working model in which MIDORI is translated in mesenchymal cells or cells undergoing EMT as a response to the stress induced by the EMT (i.e., ER stress). Once translated, MIDORI acts as a negative regulator of the pathway, decreasing the activation of SMAD, MAPK and NF-κB branches of TGFβ pathway and, in turn, decreasing the expression of EMT-related genes promoting a MET. Thus, MIDORI acts upstream of *ZEB2-AS1* in a negative feedback loop that regulates TGFβ pathway to revert EMT (Figure 72).

Importantly, as described in *Section 3.2.2* of the *Introduction*, even though EMT is required for cancer metastasis (Ye and Weinberg, 2015), invasive tumour cells need to undergo a MET in order to colonise distal organs (Ocana et al., 2012; Padmanaban et al., 2019). Moreover, compelling evidences in the literature demonstrate that cells acquiring a partial-EMT state are the ones crucial for metastatic colonisation, and the acquisition of a full mesenchymal phenotype is dispensable for the metastatic process but provides resistance to therapy (Fischer et al., 2015; Luond et al., 2021). In this scenario, MIDORI might be a new player with different functions through the metastasis process. Although we have described that MIDORI impairs metastasis when overexpressed from the establishment of the primary tumour, it could also represent a mechanism exploited by cancer cells to revert the EMT, or induce a partial-EMT state, thus promoting metastasis outgrowth. Indeed, it is known that invading cells experience several environmental stresses during their metastatic journey (i.e., oxidative stress and

ER stress) (Senft and Ronai, 2016; Tasdogan et al., 2020). Therefore, MIDORI might be expressed under stress conditions in tumour circulating cells (CTCs) and favour distal organs colonisation. In agreement with this hypothesis, the fact that *ZEB2-AS1* expression is a negative prognostic marker for metastasis free survival might be due to the dual role of the lncRNA and its encoded microprotein in the EMT-MET process. As a next step, we are planning to assess MIDORI protein expression by immunohistochemistry in patients samples at different times of the metastatic disease.
Concluding, we have described a new regulator of epithelial-mesenchymal plasticity. Although MIDORI function in cancer might be controversial and exploited in different ways depending on the stages of the disease, this discovery may open the avenue for new diagnostic or therapeutic interventions in cancer.



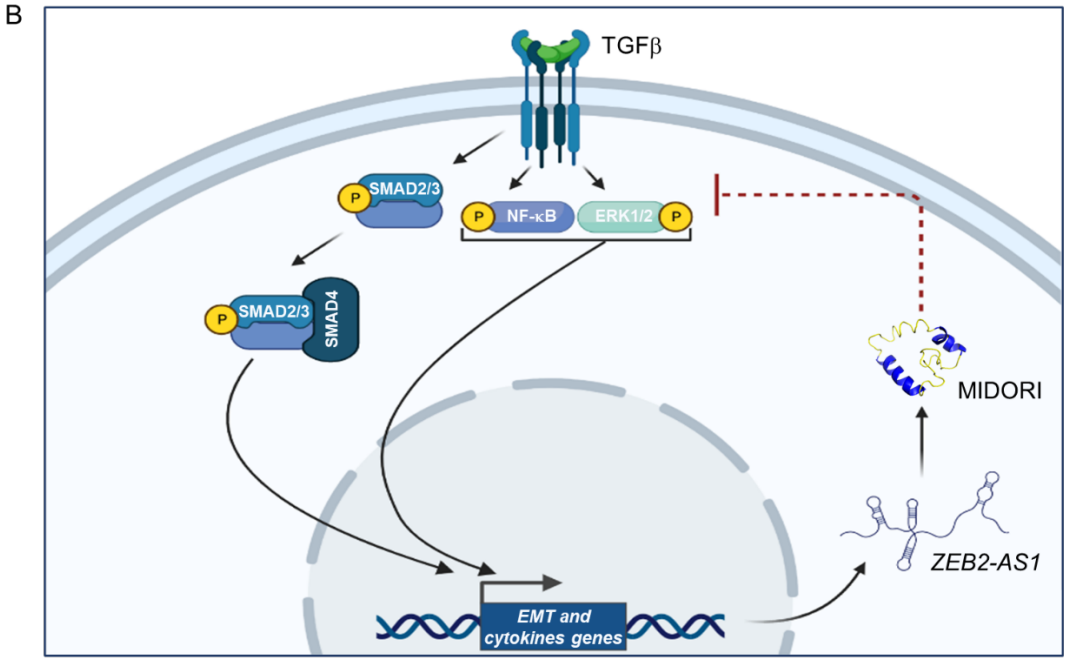


Figure 72. Proposed working model of MIDORI role in EMT. (A) MIDORI is translated downstream TGFβ in cells undergoing EMT as a response to various stress stimuli (i.e., ER stress). Once translated, MIDORI acts as a negative regulator of the EMT programme, in a negative feedback loop that regulates TGFβ pathway upstream of *ZEB2-AS1.* **(B)** MIDORI decreases the activation of SMAD, MAPK and NF-κB branches of TGFβ pathway in a negative feedback loop that in turn downregulates EMT-related genes and cytokines.

3. Functional characterisation of MIDORI in cellular reprogramming

3.1. Endogenous MIDORI is expressed at the early stages of reprogramming

It is well established that a MET at the early stages of reprogramming is crucial for a successful reprogramming to iPSCs (Li et al., 2010; Liu et al., 2013). Moreover, it is known that *Zeb2os* acts as a barrier for cellular reprogramming (Bernardes de Jesus et al., 2018). Given that MIDORI has an opposite role compared to its lncRNA *Zeb2os*, we investigated if this applies to reprogramming too.

Firstly, we observed that MIDORI is upregulated at the early stages of reprogramming at the mRNA and protein levels (Figure 56B-C). Importantly, while *Zeb2os* expression peaks at day 2 and after decays (in line with the required EMT shutdown for reprogramming progression), MIDORI expression is shifted and retained longer (Figure 56C). This expression pattern is consistent with a direct involvement of MIDORI in the MET process that takes place during reprogramming, after *Zeb2os* downregulation. Furthermore, we already showed that MIDORI is translated upon stress (Figure 22B-C), and the expression of the OSKM during reprogramming is known to induce genotoxic stress (Gonzalez et al., 2013; Mosteiro et al., 2016). Thus, it might be that MIDORI is translated upon OSKM-induced stress and, once translated, might favours a MET in reprogramming cells.

3.2. MIDORI cell intrinsic effect in reprogramming

By reprogramming MIDORI-overexpressing MEFs we found that MIDORI ectopic expression significantly increases the reprogramming efficiency (Figure 57), further confirming its role as a negative regulator of EMT. Next, to address whether MIDORI is required during cellular reprogramming, we reprogrammed WT and MIDORI-KO MEFs. Consistent with our previous results, we observed that MIDORI deficiency delays the appearance of iPSC colonies (Figure 59A) and significantly decreases the reprogramming efficiency (Figure 59B-C). Thus, MIDORI seems to be required for a successful reprogramming. In addition, we demonstrated that MIDORI exogenous expression is sufficient to rescue the phenotype of MIDORI-KO MEFs (Figure 61A-B).

By analysing the transcriptional profile of WT and MIDORI-KO MEFs during cellular reprogramming, we noticed that the expression of EMT-related genes is dysregulated in MIDORI-KO MEFs, which expressed *Snail, Slug* and *Tgfβ1* at substantially high levels

when they are clearly down in the controls (Figure 62A). Our results are consistent with previous literature, in which it is reported that the EMT programme needs to be shut down to proceed with the reprogramming process (Li et al., 2010; Liu et al., 2013). Therefore, MIDORI deficiency might impair the downregulation of the EMT programme and, hence, leads to a lower reprogramming efficiency. Of note, we showed that MIDORI endogenous expression starts on day 2, peaks on day 4, and starts to decrease on day 6 of reprogramming (Figure 56C). Indeed, we saw the highest transcriptional differences between WT and MIDORI-KO on day 6 (Figure 62A), suggesting that MIDORI could be acting between days 4 and 6 and that its deficiency results in EMT-genes dysregulation in that time window.

Unexpectedly, we also detected that MIDORI-KO MEFs express significantly higher levels of the tumour suppressor genes *p19^{Arf}* and *p16^{INK4a}* at the intermediate phase of the reprogramming process (from day 6 to day 10) (Figure 62B). These genes represent one of the most important barriers for *in vitro* reprogramming in a cell-autonomous manner by triggering senescence (Li et al., 2009), and, consistent with this, we observed a decreased reprogramming efficiency in our MIDORI-KO MEFs. Therefore, we wondered whether MIDORI deficiency could sensitise the cells to undergo senescence. Indeed, we observed that MIDORI loss-of-function increases SA-β-Gal activity in basal conditions and upon Palbociclib treatment (Figure 63A-B), whereas MIDORI overexpression seems to slightly decrease the number of SA-β-Gal positive cells upon Palbociclib treatment (Figure 63B). These evidences pointed to a possible role of MIDORI in regulating cellular senescence. We then studied the induction of cellular senescence during reprogramming and observed that MIDORI-KO MEFs accumulate more SA-β-Gal positive cells than WT and MIDORI-overexpressing cells (Figure 64A-B). Of note, MIDORI overexpression, both in WT and MIDORI-KO cells, seems to prevent senescence induction (Figure 64B).

Concluding, these data show that the decreased reprogramming efficiency in MIDORI-KO MEFs is in part due to an increase in the induction of senescence, which acts as an intrinsic barrier preventing reprogramming. Therefore, MIDORI seems to play a dual role in reprogramming: controlling both EMT and cellular senescence. Particularly, since MIDORI is expressed in response to stress, it might be that MIDORI deficient MEFs are not able to cope with the stress induced by the OSKM factors and enter in senescence. While MIDORI expression during the reprogramming process may ensure the progression through MET, MIDORI-KO MEFs, which lacks this mechanism, could undergo senescence as an alternative response to damage.

3.3. MIDORI cell extrinsic effect in reprogramming

We then addressed MIDORI's cell extrinsic effect in reprogramming by using the CM collected from MIDORI-overexpressing cells. Surprisingly, we saw that MIDORI-overexpressing cells CM improves reprogramming in a non-cell autonomous manner (Figure 58). In agreement, MIDORI deficiency seems to affect reprogramming efficiency also in a cell extrinsic manner (Figure 60). In addition, we demonstrated that MIDORI exogenous expression in MIDORI-KO MEFs is sufficient to rescue the phenotype (Figure 61C-D). Therefore, as we already demonstrated in the context of cancer cells, MIDORI

could also act in a cell extrinsic manner by modifying the cell secretome, hence impacting on cellular reprogramming. In fact, it is known that microenvironment signals are extremely important in order for reprogramming to occur. On the one hand, as we described in the *Introduction,* different cytokines can have opposite effects in reprogramming both *in vitro* and *in vivo*. For instance, IL-6 has been described as a promoter of reprogramming both in the *in vitro* and in the *in vivo* condition (Brady et al., 2013; Chiche et al., 2017; Mosteiro et al., 2016; Mosteiro et al., 2018), while the members of TGFβ family and TNFα are known to inhibit reprogramming *in vitro* (Li et al., 2010; Liu et al., 2013; Mahmoudi et al., 2019). On the other hand, a large spectrum of small molecules has been identified during the last decades in the attempt to improve the process of reprogramming or even replace the OSKM factors. As an example, Vitamin C improves reprogramming by enhancing the activity of Tet enzymes and others epigenetic modifiers, while valproic acid or butyrate favours reprogramming through the inhibition of histone deacetylases (HDACs) activity (Su et al., 2013). Moreover, recently it has been demonstrated that mouse and human cells can be reprogrammed to pluripotency by using a defined cocktail of small molecules, a process known as chemical reprogramming (Guan et al., 2022; Hou et al., 2013). Thus, identifying new molecules important for reprogramming might help our understanding of the molecular bases of this process and getting rid of the detrimental effects associated with the use of the OSKM as reprogramming factors.

To identify the factors enriched or depleted in MIDORI-derived secretome, we analysed the CM from MIDORI-overexpressing cells by mass spectrometry (Figure 65). Importantly, we found that TGFβ3, a member of the TGFβ family, is downregulated in our MIDORI-overexpressing cells CM. Given the negative effect of TGFβ family members in cellular reprogramming (Li et al., 2010; Liu et al., 2013), the reduced secretion of TGFβ3 could in part explain the enhanced reprogramming.
Although we demonstrated that MIDORI deficiency sensitises MEFs to undergo senescence (Figure 63-64) and the SASP is critical for reprogramming, MIDORI-KO cells-derived secretome (and eventually their SASP) does not seem to induce reprogramming, in fact it impairs reprogramming (Figure 60). Importantly, we observed that MIDORI-KO MEFs express high levels of TNFα at basal levels, and high levels of TGFβ during the reprogramming process (Figure 62A and C). Although they are both components of the SASP (Coppe et al., 2010), TNFα is known to have a detrimental effect on reprogramming (Mahmoudi et al., 2019), as well as TGFβ (Li et al., 2010; Liu et al., 2013). Thus, even though MIDORI-KO cells are more prone to undergo senescence, the particular composition of their SASP could lead to a decrease in reprogramming efficiency. To further explore this hypothesis, we are currently performing a thorough analysis of WT and MIDORI-KO CM by cytokine arrays.

3.4. *In vivo* reprogramming of MIDORI-KO mice

Finally, we studied the effect of MIDORI deficiency during *in vivo* reprogramming. By using two different reprogramming protocols we observed that MIDORI-KO mice die faster due to teratomas (Figure 69B) and display more dysplasia in the reprogrammed pancreas (Figure 70A-B). These data, together with an increased number of proliferating

cells in the dysplastic areas of the pancreas (Figure 70A and C), suggest that MIDORI deficiency increases *in vivo* reprogramming.

As we observed that MIDORI-KO MEFs are prone to undergo senescence upon OSKM induction (Figure 64), we looked at $p21^{CIP1}$ expression in our reprogrammed pancreata. We found an increase in $p21^{CIP1}$ positive cells in the pancreas of MIDORI-KO mice, suggestive of more senescent cells (Figure 71A-B).
It is known that senescent cells provide critical paracrine signals for *in vivo* reprogramming (Mosteiro et al., 2016). In fact, if cells cannot undergo senescence, for instance in a $p16^{INK4a}/p19^{ARF}$ knock-out mice, *in vivo* reprogramming is impaired (Mosteiro et al., 2016). Indeed, MIDORI-KO mice seem to behave in the opposite way to $p16^{INK4a}/p19^{ARF}$ -KO mice. While $p16^{INK4a}/p19^{ARF}$-KO MEFs reprogramme better *in vitro* because they overcome the senescence barrier, $p16^{INK4a}/p19^{ARF}$ -KO mice display an impaired reprogramming *in vivo* because of the lack of paracrine factors secreted by senescent cells (OSKM does not induce senescence in $p16^{INK4a}/p19^{ARF}$-KO mice). In the case of MIDORI, its deficiency triggers the onset of cellular senescence upon OSKM that impairs reprogramming *in vitro* (acting as a cell intrinsic barrier) but promotes reprograming *in vivo*, thus, the improved reprogramming efficiency *in vivo* probably through the SASP.

Our data show that the secretome of MIDORI-KO MEFs does not enhance *in vitro* reprogramming (Figure 60). Although this last piece of data does not fit with the reported effect of SASP in reprogramming (Mosteiro et al., 2016), it is important to highlight that the SASP composition could be different *in vitro* and *in vivo*, or have different effects. For example, while the role of TGFβ/TNFα in *in vitro* reprogramming is well established, nothing is known about their role in *in vivo* reprogramming. Therefore, it is possible that MIDORI-KO-induced secretome is detrimental for *in vitro* reprogramming but could provide the right signals for *in vivo* reprogramming.
Finally, although EMT is known to be detrimental for *in vitro* reprogramming, there is no report exploring its role during *in vivo* reprogramming. In future experiments we will analyse the EMT-MET dynamics during *in vivo* reprogramming, hopefully shedding some light in the black box of the mechanisms of *in vivo* reprogramming.

3.5. Final considerations and proposed model for the role of MIDORI in cellular reprogramming

Concluding, we have firstly demonstrated that MIDORI is endogenously expressed at the early stages of reprogramming (Figure 56C). Of note, this is in line with our previous findings, in which we observed that MIDORI is upregulated upon damage. Our findings have shown that MIDORI overexpression significantly enhances *in vitro* cellular reprogramming in a cell intrinsic manner (Figure 57) and in a cell extrinsic manner (Figure 58). Importantly, by analysing MIDORI-related secretome, we have found that it is reduced in TGFβ3 (Figure 65), a cytokine known to impair reprogramming, thus providing a possible explanation for MIDORI cell extrinsic effect. In agreement, we have shown that MIDORI deficiency impairs reprogramming *in vitro* (Figure 59-60). MIDORI deficiency dysregulates EMT genes dynamics (Figure 62A) and sensitises the cells to

undergo senescence (Figure 63-64), hence impeding reprogramming in a cell intrinsic manner. On the other hand, MIDORI-deficient cells express higher level of TGFβ and TNFα cytokines compared to WT cells (Figure 62A and C), which could be affecting reprogramming in a non-cell autonomous manner.

Based on our results, we propose a model in which MIDORI is upregulated by the stress-induced by OSKM. MIDORI expression induces a MET programme, necessary for cells in order to acquire pluripotency. At the same time, cells overexpressing MIDORI produce secreted factors that promotes reprogramming in a non-cell autonomous manner (Figure 73).

In the absence of MIDORI, the cells cannot activate the MET and cannot proceed further in the reprogramming process. Although this is yet to be confirmed, our hypothesis is that this may force them to enter senescence as a response to the continuous stress of the OSKM expression. Even if senescence has been described to be beneficial for reprogramming in a non-cell-autonomous manner, MIDORI-deficient cells may produce a pool of secreted factors that is not beneficial for cellular reprogramming (Figure 74).

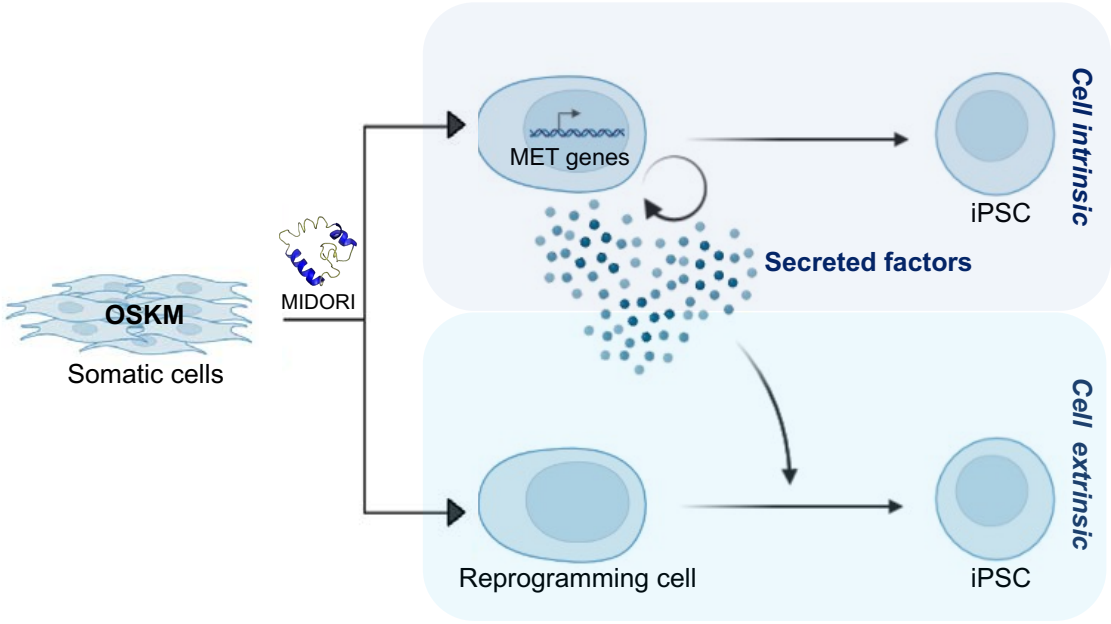


Figure 73. Proposed working model of MIDORI role in reprogramming. MIDORI is upregulated upon OSKM induction during reprogramming. MIDORI expression in a cell-autonomous manner promotes the activation of MET programme, necessary for cells in order to acquire pluripotency. On the other hand, cells overexpressing MIDORI produce secreted factors that help cells during the reprogramming process in a non-cell autonomous manner.

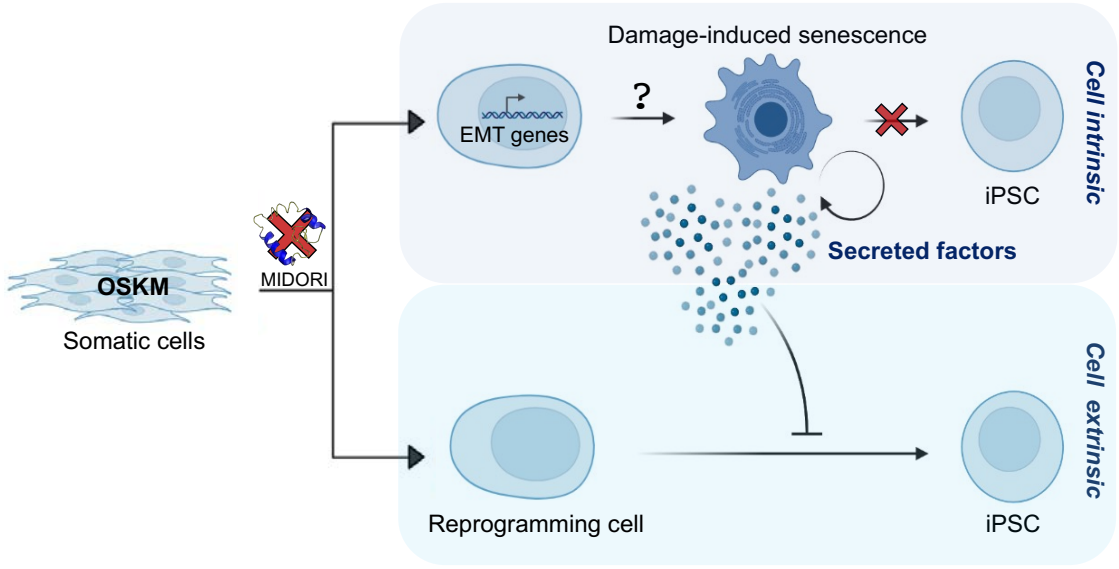


Figure 74. Proposed model of MIDORI deficiency in reprogramming. When MIDORI is lost, the cells cannot activate MET and dysregulate EMT genes instead. This missed activation of the MET programme does not allow the cells to proceed further in the reprogramming process and may force them to enter senescence as a response to the continuous stress of the OSKM expression. In a cell extrinsic manner, MIDORI-deficient cells may produce a dysfunctional pool of secreted factors that also decreases reprogramming in a paracrine way.

In conclusion, microproteins have emerged as novel class of molecular regulators with important roles in different biological processes, including cancer. In this study, we have identified MIDORI, a novel microprotein encoded by a gene annotated as non-coding which acts as a negative regulator of the epithelial to mesenchymal transition.

Our results increase the body of knowledge on the molecular mechanisms behind cellular plasticity adding a new molecular player, MIDORI. From a more general point of view, our results have uncovered that the microproteome regulates the EMP process, and could hide important players with important implications in regenerative medicine and cancer.

CONCLUSIONS

1. *ZEB2-AS1* lncRNA contains a sORF that encondes an evolutionary conserved microprotein that we have named MIDORI.

2. MIDORI expression is upregulated upon damage and TGFβ signalling and it localises both in the nucleus and the cytoplasm.

3. MIDORI overexpression downregulates the mesenchymal programme in primary fibroblasts and several cancer cell lines, and blocks the induction of EMT. In breast cancer cells, MIDORI reduces the secretion of proinflammatory cytokines, and reduces cell migration and invasion.

4. MIDORI impairs metastatic colonisation of breast cancer cells *in vivo*.

5. MIDORI deficiency significantly increases wound healing *in vivo*.

6. Mechanistically, MIDORI acts upstream of *ZEB2-AS1* decreasing the activation of the TGFβ pathway effectors SMAD2, ERK1/2 and NF-κB. The analysis of its interactome revealed that MIDORI interacts with ITGB1BP1, HMGA1 and MYBBP1A.

7. MIDORI is transiently expressed at the early stages of cellular reprogramming from MEFs to iPSCs. Its overexpression increases the efficiency of reprogramming, and its deficiency severely impairs the process.

8. The conditioned medium of MIDORI-overexpressing MEFs promotes cellular reprogramming, probably via decreased TGFb3 in the CM.

9. MIDORI-deficient MEFs display dysregulated EMT transcriptional dynamics and increased expression of the tumour suppressors $p19^{Arf}$ and $p16^{Ink4a}$ during reprogramming.

10. MIDORI-deficiency sensitises MEFs to undergo senescence while MIDORI overexpression impairs the induction of senescence by OSKM.

11. MIDORI deficiency increases cellular reprogramming in vivo.

BIBLIOGRAPHY

Abad, M., Mosteiro, L., Pantoja, C., Canamero, M., Rayon, T., Ors, I., Grana, O., Megias, D., Dominguez, O., Martinez, D.*, et al.* (2013). Reprogramming in vivo produces teratomas and iPS cells with totipotency features. Nature *502*, 340-345.

Acosta, J.C., Banito, A., Wuestefeld, T., Georgilis, A., Janich, P., Morton, J.P., Athineos, D., Kang, T.W., Lasitschka, F., Andrulis, M.*, et al.* (2013). A complex secretory program orchestrated by the inflammasome controls paracrine senescence. Nat Cell Biol *15*, 978-990.

Al Moustafa, A.E., Achkhar, A., and Yasmeen, A. (2012). EGF-receptor signaling and epithelial-mesenchymal transition in human carcinomas. Front Biosci (Schol Ed) *4*, 671-684.

Almagro Armenteros, J.J., Sonderby, C.K., Sonderby, S.K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics *33*, 3387-3395.

An, J., Zheng, Y., and Dann, C.T. (2017). Mesenchymal to Epithelial Transition Mediated by CDH1 Promotes Spontaneous Reprogramming of Male Germline Stem Cells to Pluripotency. Stem Cell Reports *8*, 446-459.

Anderson, D.M., Anderson, K.M., Chang, C.L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R.*, et al.* (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell *160*, 595-606.

Anderson, D.M., Makarewich, C.A., Anderson, K.M., Shelton, J.M., Bezprozvannaya, S., Bassel-Duby, R., and Olson, E.N. (2016). Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. Sci Signal *9*, ra119.

Anderson, P., and Kedersha, N. (2008). Stress granules: the Tao of RNA triage. Trends Biochem Sci *33*, 141-150.

Andrews, S.J., and Rothnagel, J.A. (2014). Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet *15*, 193-204.

Aoi, T., Yae, K., Nakagawa, M., Ichisaka, T., Okita, K., Takahashi, K., Chiba, T., and Yamanaka, S. (2008). Generation of pluripotent stem cells from adult mouse liver and stomach cells. Science *321*, 699-702.

Attalla, S., Taifour, T., Bui, T., and Muller, W. (2021). Insights from transgenic mouse models of PyMT-induced breast cancer: recapitulating human breast cancer progression in vivo. Oncogene *40*, 475-491.

Banito, A., Rashid, S.T., Acosta, J.C., Li, S., Pereira, C.F., Geti, I., Pinho, S., Silva, J.C., Azuara, V., Walsh, M.*, et al.* (2009). Senescence impairs successful reprogramming to pluripotent stem cells. Genes Dev *23*, 2134-2139.

Basrai, M.A., Hieter, P., and Boeke, J.D. (1997). Small open reading frames: beautiful needles in the haystack. Genome Res *7*, 768-771.

Bass, A.J., Watanabe, H., Mermel, C.H., Yu, S., Perner, S., Verhaak, R.G., Kim, S.Y., Wardwell, L., Tamayo, P., Gat-Viks, I.*, et al.* (2009). SOX2 is an amplified lineage-

survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet *41*, 1238-1242.

Beltran, M., Puig, I., Pena, C., Garcia, J.M., Alvarez, A.B., Pena, R., Bonilla, F., and de Herreros, A.G. (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev *22*, 756-769.

Benecke, A.G., and Eilebrecht, S. (2015). RNA-Mediated Regulation of HMGA1 Function. Biomolecules *5*, 943-957.

Bernardes de Jesus, B., Marinho, S.P., Barros, S., Sousa-Franco, A., Alves-Vale, C., Carvalho, T., and Carmo-Fonseca, M. (2018). Silencing of the lncRNA Zeb2-NAT facilitates reprogramming of aged fibroblasts and safeguards stem cell pluripotency. Nat Commun *9*, 94.

Bernat-Peguera, A., Navarro-Ventura, J., Lorenzo-Sanz, L., da Silva-Diz, V., Bosio, M., Palomero, L., Penin, R.M., Perez Sidelnikova, D., Bermejo, J.O., Taberna, M.*, et al.* (2021). FGFR Inhibition Overcomes Resistance to EGFR-targeted Therapy in Epithelial-like Cutaneous Carcinoma. Clin Cancer Res *27*, 1491-1504.

Bernat-Peguera, A., Simon-Extremera, P., da Silva-Diz, V., Lopez de Munain, M., Diaz-Gil, L., Penin, R.M., Gonzalez-Suarez, E., Perez Sidelnikova, D., Bermejo, O., Vinals, J.M.*, et al.* (2019). PDGFR-induced autocrine SDF-1 signaling in cancer cells promotes metastasis in advanced skin carcinoma. Oncogene *38*, 5021-5037.

Bhatta, A., Atianand, M., Jiang, Z., Crabtree, J., Blin, J., and Fitzgerald, K.A. (2020). A Mitochondrial Micropeptide Is Required for Activation of the Nlrp3 Inflammasome. J Immunol *204*, 428-437.

Bi, P., Ramirez-Martinez, A., Li, H., Cannavino, J., McAnally, J.R., Shelton, J.M., Sanchez-Ortiz, E., Bassel-Duby, R., and Olson, E.N. (2017). Control of muscle formation by the fusogenic micropeptide myomixer. Science *356*, 323-327.

Blasco, M.A., Lee, H.W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R.A., and Greider, C.W. (1997). Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. Cell *91*, 25-34.

Bouvard, D., Vignoud, L., Dupe-Manet, S., Abed, N., Fournier, H.N., Vincent-Monegat, C., Retta, S.F., Fassler, R., and Block, M.R. (2003). Disruption of focal adhesions by integrin cytoplasmic domain-associated protein-1 alpha. J Biol Chem *278*, 6567-6574.

Brady, J.J., Li, M., Suthram, S., Jiang, H., Wong, W.H., and Blau, H.M. (2013). Early role for IL-6 signalling during generation of induced pluripotent stem cells revealed by heterokaryon RNA-Seq. Nat Cell Biol *15*, 1244-1252.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell *2*, 151-159.

Brunet, M.A., Leblanc, S., and Roucou, X. (2020). Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. Exp Cell Res *393*, 112057.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev *25*, 1915-1927.

Cahu, J., Bustany, S., and Sola, B. (2012). Senescence-associated secretory phenotype favors the emergence of cancer stem-like cells. Cell Death Dis *3*, e446.

Campisi, J., and d'Adda di Fagagna, F. (2007). Cellular senescence: when bad things happen to good cells. Nat Rev Mol Cell Biol *8*, 729-740.

Cao, X., and Slavoff, S.A. (2020). Non-AUG start codons: Expanding and regulating the small and alternative ORFeome. Exp Cell Res *391*, 111973.

Carbonnelle, D., Vignard, V., Sehedic, D., Moreau-Aubry, A., Florenceau, L., Charpentier, M., Mikulits, W., Labarriere, N., and Lang, F. (2013). The melanoma antigens MELOE-1 and MELOE-2 are translated from a bona fide polycistronic mRNA containing functional IRES sequences. PLoS One *8*, e75233.

Cazzalini, O., Scovassi, A.I., Savio, M., Stivala, L.A., and Prosperi, E. (2010). Multiple roles of the cell cycle inhibitor p21(CDKN1A) in the DNA damage response. Mutat Res *704*, 12-20.

Chai, G.S., Duan, D.X., Ma, R.H., Shen, J.Y., Li, H.L., Ma, Z.W., Luo, Y., Wang, L., Qi, X.H., Wang, Q.*, et al.* (2014). Humanin attenuates Alzheimer-like cognitive deficits and pathological changes induced by amyloid beta-peptide in rats. Neurosci Bull *30*, 923-935.

Chang, D.D., Wong, C., Smith, H., and Liu, J. (1997). ICAP-1, a novel beta1 integrin cytoplasmic domain-associated protein, binds to a conserved and functionally important NPXY sequence motif of beta1 integrin. J Cell Biol *138*, 1149-1157.

Chiche, A., Le Roux, I., von Joest, M., Sakai, H., Aguin, S.B., Cazin, C., Salam, R., Fiette, L., Alegria, O., Flamant, P.*, et al.* (2017). Injury-Induced Senescence Enables In Vivo Reprogramming in Skeletal Muscle. Cell Stem Cell *20*, 407-414 e404.

Chien, Y., Scuoppo, C., Wang, X., Fang, X., Balgley, B., Bolden, J.E., Premsrirut, P., Luo, W., Chicas, A., Lee, C.S.*, et al.* (2011). Control of the senescence-associated secretory phenotype by NF-kappaB promotes senescence and enhances chemosensitivity. Genes Dev *25*, 2125-2136.

Chin, L., Pomerantz, J., and DePinho, R.A. (1998). The INK4a/ARF tumor suppressor: one gene--two products--two pathways. Trends Biochem Sci *23*, 291-296.

Chng, S.C., Ho, L., Tian, J., and Reversade, B. (2013). ELABELA: a hormone essential for heart development signals via the apelin receptor. Dev Cell *27*, 672-680.

Chong, C., Muller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B.J.*, et al.* (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. Nat Commun *11*, 1293.

Chu, Q., Martinez, T.F., Novak, S.W., Donaldson, C.J., Tan, D., Vaughan, J.M., Chang, T., Diedrich, J.K., Andrade, L., Kim, A.*, et al.* (2019). Regulation of the ER stress response by a mitochondrial microprotein. Nat Commun *10*, 4883.

Chu, Q., and Saghatelian, A. (2019). A Hidden ORF Reveals an Immune Protector. Biochemistry *58*, 1022-1023.

Cleynen, I., and Van de Ven, W.J. (2008). The HMGA proteins: a myriad of functions (Review). Int J Oncol *32*, 289-305.

Collado, M., Gil, J., Efeyan, A., Guerra, C., Schuhmacher, A.J., Barradas, M., Benguria, A., Zaballos, A., Flores, J.M., Barbacid, M.*, et al.* (2005). Tumour biology: senescence in premalignant tumours. Nature *436*, 642.

Coppe, J.P., Desprez, P.Y., Krtolica, A., and Campisi, J. (2010). The senescence-associated secretory phenotype: the dark side of tumor suppression. Annu Rev Pathol *5*, 99-118.

D'Lima, N.G., Ma, J., Winkler, L., Chu, Q., Loh, K.H., Corpuz, E.O., Budnik, B.A., Lykke-Andersen, J., Saghatelian, A., and Slavoff, S.A. (2017). A human microprotein that interacts with the mRNA decapping complex. Nat Chem Biol *13*, 174-180.

David, C.J., and Massague, J. (2018). Contextual determinants of TGFbeta action in development, immunity and cancer. Nat Rev Mol Cell Biol *19*, 419-435.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.

DeBerardinis, R.J., and Chandel, N.S. (2020). We need to talk about the Warburg effect. Nat Metab *2*, 127-129.

Decker, C.J., and Parker, R. (2012). P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. Cold Spring Harb Perspect Biol *4*, a012286.

Degani, S., Balzac, F., Brancaccio, M., Guazzone, S., Retta, S.F., Silengo, L., Eva, A., and Tarone, G. (2002). The integrin cytoplasmic domain-associated protein ICAP-1 binds and regulates Rho family GTPases during cell spreading. J Cell Biol *156*, 377-387.

DeNicola, G.M., and Tuveson, D.A. (2009). RAS in cellular transformation and senescence. Eur J Cancer *45 Suppl 1*, 211-216.

Di Mitri, D., Toso, A., Chen, J.J., Sarti, M., Pinton, S., Jost, T.R., D'Antuono, R., Montani, E., Garcia-Escudero, R., Guccini, I.*, et al.* (2014). Tumour-infiltrating Gr-1+ myeloid cells antagonize senescence in cancer. Nature *515*, 134-137.

Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235-239.

Dickson, M.A. (2014). Molecular pathways: CDK4 inhibitors for cancer therapy. Clin Cancer Res *20*, 3379-3383.

Dimri, G.P., Lee, X., Basile, G., Acosta, M., Scott, G., Roskelley, C., Medrano, E.E., Linskens, M., Rubelj, I., Pereira-Smith, O.*, et al.* (1995). A biomarker that identifies senescent human cells in culture and in aging skin in vivo. Proc Natl Acad Sci U S A *92*, 9363-9367.

Dongre, A., and Weinberg, R.A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol *20*, 69-84.

Dou, Z., and Berger, S.L. (2018). Senescence Elicits Stemness: A Surprising Mechanism for Cancer Relapse. Cell Metab *27*, 710-711.

Eggert, T., Wolter, K., Ji, J., Ma, C., Yevsa, T., Klotz, S., Medina-Echeverz, J., Longerich, T., Forgues, M., Reisinger, F.*, et al.* (2016). Distinct Functions of Senescence-Associated Immune Responses in Liver Tumor Surveillance and Tumor Progression. Cancer Cell *30*, 533-547.

Espinoza, I., and Miele, L. (2013). Deadly crosstalk: Notch signaling at the intersection of EMT and cancer stem cells. Cancer Lett *341*, 41-45.

Felipe-Abrio, B., and Carnero, A. (2020). The Tumor Suppressor Roles of MYBBP1A, a Major Contributor to Metabolism Plasticity and Stemness. Cancers (Basel) *12*.

Fernandez-Gonzalez, A. (2021). TGF-beta and NF-kappaB Cross-Talk: Unexpected Encounters in the Developing Lung. Am J Respir Cell Mol Biol *64*, 275-276.

Fickett, J.W. (1995). ORFs and genes: how strong a connection? J Comput Biol *2*, 117-123.

Filipponi, D., Emelyanov, A., Muller, J., Molina, C., Nichols, J., and Bulavin, D.V. (2019). DNA Damage Signaling-Induced Cancer Cell Reprogramming as a Driver of Tumor Relapse. Mol Cell *74*, 651-663 e658.

Fischer, K.R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S.T., Choi, H., El Rayes, T., Ryu, S., Troeger, J.*, et al.* (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature *527*, 472-476.

Flanagan, K.C., Alspach, E., Pazolli, E., Parajuli, S., Ren, Q., Arthur, L.L., Tapia, R., and Stewart, S.A. (2018). c-Myb and C/EBPbeta regulate OPN and other senescence-associated secretory phenotype factors. Oncotarget *9*, 21-36.

Folmes, C.D., Nelson, T.J., Martinez-Fernandez, A., Arrell, D.K., Lindor, J.Z., Dzeja, P.P., Ikeda, Y., Perez-Terzic, C., and Terzic, A. (2011). Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. Cell Metab *14*, 264-271.

Frangogiannis, N. (2020). Transforming growth factor-beta in tissue fibrosis. J Exp Med *217*, e20190103.

Freudlsperger, C., Bian, Y., Contag Wise, S., Burnett, J., Coupar, J., Yang, X., Chen, Z., and Van Waes, C. (2013). TGF-beta and NF-kappaB signal pathway cross-talk is mediated through TAK1 and SMAD7 in a subset of head and neck cancers. Oncogene *32*, 1549-1559.

Freund, A., Patil, C.K., and Campisi, J. (2011). p38MAPK is a novel DNA damage response-independent regulator of the senescence-associated secretory phenotype. EMBO J *30*, 1536-1548.

Friedmann-Morvinski, D., Bushong, E.A., Ke, E., Soda, Y., Marumoto, T., Singer, O., Ellisman, M.H., and Verma, I.M. (2012). Dedifferentiation of neurons and astrocytes by oncogenes can induce gliomas in mice. Science *338*, 1080-1084.

Friedmann-Morvinski, D., and Verma, I.M. (2014). Dedifferentiation and reprogramming: origins of cancer stem cells. EMBO Rep *15*, 244-253.

Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A., and Couso, J.P. (2007). Peptides encoded by short ORFs control development and define a new eukaryotic gene family. PLoS Biol *5*, e106.

Gao, S., Tao, L., Hou, X., Xu, Z., Liu, W., Zhao, K., Guo, M., Wang, H., Cai, T., Tian, J.*, et al.* (2016). Genome-wide gene expression analyses reveal unique cellular characteristics related to the amenability of HPC/HSCs into high-quality induced pluripotent stem cells. Stem Cell Res Ther *7*, 40.

Gasek, N.S., Kuchel, G.A., Kirkland, J.L., and Xu, M. (2021). Strategies for Targeting Senescent Cells in Human Disease. Nat Aging *1*, 870-879.

Ge, Q., Jia, D., Cen, D., Qi, Y., Shi, C., Li, J., Sang, L., Yang, L.J., He, J., Lin, A.*, et al.* (2021). Micropeptide ASAP encoded by LINC00467 promotes colorectal cancer progression by directly modulating ATP synthase activity. J Clin Invest *131*.

Gerardo, H., Lima, A., Carvalho, J., Ramos, J.R.D., Couceiro, S., Travasso, R.D.M., Pires das Neves, R., and Graos, M. (2019). Soft culture substrates favor stem-like cellular phenotype and facilitate reprogramming of human mesenchymal stem/stromal cells (hMSCs) through mechanotransduction. Sci Rep *9*, 9086.

Gingery, A., Bradley, E.W., Pederson, L., Ruan, M., Horwood, N.J., and Oursler, M.J. (2008). TGF-beta coordinately activates TAK1/MEK/AKT/NFkB and SMAD pathways to promote osteoclast survival. Exp Cell Res *314*, 2725-2738.

Gonzalez, D.M., and Medici, D. (2014). Signaling mechanisms of the epithelial-mesenchymal transition. Sci Signal *7*, re8.

Gonzalez, F., Georgieva, D., Vanoli, F., Shi, Z.D., Stadtfeld, M., Ludwig, T., Jasin, M., and Huangfu, D. (2013). Homologous recombination DNA repair genes play a critical role in reprogramming to a pluripotent state. Cell Rep *3*, 651-660.

Goossens, S., Vandamme, N., Van Vlierberghe, P., and Berx, G. (2017). EMT transcription factors in cancer development re-evaluated: Beyond EMT and MET. Biochim Biophys Acta Rev Cancer *1868*, 584-591.

Graf, J., and Kretz, M. (2020). From structure to function: Route to understanding lncRNA mechanism. Bioessays *42*, e2000027.

Grande, M.T., Sanchez-Laorden, B., Lopez-Blau, C., De Frutos, C.A., Boutet, A., Arevalo, M., Rowe, R.G., Weiss, S.J., Lopez-Novoa, J.M., and Nieto, M.A. (2015). Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease. Nat Med *21*, 989-997.

Grigore, A.D., Jolly, M.K., Jia, D., Farach-Carson, M.C., and Levine, H. (2016). Tumor Budding: The Name is EMT. Partial EMT. J Clin Med *5*.

Guan, J., Wang, G., Wang, J., Zhang, Z., Fu, Y., Cheng, L., Meng, G., Lyu, Y., Zhu, J., Li, Y.*, et al.* (2022). Chemical reprogramming of human somatic cells to pluripotent stem cells. Nature *605*, 325-331.

Guerra, C., Schuhmacher, A.J., Canamero, M., Grippo, P.J., Verdaguer, L., Perez-Gallego, L., Dubus, P., Sandgren, E.P., and Barbacid, M. (2007). Chronic pancreatitis is essential for induction of pancreatic ductal adenocarcinoma by K-Ras oncogenes in adult mice. Cancer Cell *11*, 291-302.

Guha, P., Kaptan, E., Gade, P., Kalvakolanu, D.V., and Ahmed, H. (2017). Tunicamycin induced endoplasmic reticulum stress promotes apoptosis of prostate cancer cells by activating mTORC1. Oncotarget *8*, 68191-68207.

Guo, B., Wu, S., Zhu, X., Zhang, L., Deng, J., Li, F., Wang, Y., Zhang, S., Wu, R., Lu, J.*, et al.* (2020). Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. EMBO J *39*, e102190.

Guo, Z.W., Meng, Y., Zhai, X.M., Xie, C., Zhao, N., Li, M., Zhou, C.L., Li, K., Liu, T.C., Yang, X.X.*, et al.* (2019). Translated Long Non-Coding Ribonucleic Acid ZFAS1 Promotes Cancer Cell Migration by Elevating Reactive Oxygen Species Production in Hepatocellular Carcinoma. Front Genet *10*, 1111.

Gupta, P.B., Pastushenko, I., Skibinski, A., Blanpain, C., and Kuperwasser, C. (2019). Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. Cell Stem Cell *24*, 65-78.

Gurdon, J.B., Elsdale, T.R., and Fischberg, M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature *182*, 64-65.

Guy, C.T., Cardiff, R.D., and Muller, W.J. (1992). Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. Mol Cell Biol *12*, 954-961.

Hall, D.B., Wade, J.T., and Struhl, K. (2006). An HMG protein, Hmo1, associates with promoters of many ribosomal protein genes and throughout the rRNA gene locus in Saccharomyces cerevisiae. Mol Cell Biol *26*, 3672-3679.

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discov *12*, 31-46.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458-460.

Hashimoto, Y., Niikura, T., Tajima, H., Yasukawa, T., Sudo, H., Ito, Y., Kita, Y., Kawasumi, M., Kouyama, K., Doyu, M.*, et al.* (2001). A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and Abeta. Proc Natl Acad Sci U S A *98*, 6336-6341.

Heldin, C.H., Vanlandewijck, M., and Moustakas, A. (2012). Regulation of EMT by TGFbeta in cancer. FEBS Lett *586*, 1959-1970.

Hernandez-Segura, A., de Jong, T.V., Melov, S., Guryev, V., Campisi, J., and Demaria, M. (2017). Unmasking Transcriptional Heterogeneity in Senescent Cells. Curr Biol *27*, 2652-2660 e2654.

Herranz, N., Gallage, S., Mellone, M., Wuestefeld, T., Klotz, S., Hanley, C.J., Raguz, S., Acosta, J.C., Innes, A.J., Banito, A.*, et al.* (2015). mTOR regulates MAPKAPK2 translation to control the senescence-associated secretory phenotype. Nat Cell Biol *17*, 1205-1217.

Herranz, N., Pasini, D., Diaz, V.M., Franci, C., Gutierrez, A., Dave, N., Escriva, M., Hernandez-Munoz, I., Di Croce, L., Helin, K.*, et al.* (2008). Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. Mol Cell Biol *28*, 4772-4781.

Hoffding, M.K., and Hyttel, P. (2015). Ultrastructural visualization of the Mesenchymal-to-Epithelial Transition during reprogramming of human fibroblasts to induced pluripotent stem cells. Stem Cell Res *14*, 39-53.

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K., and Yamanaka, S. (2009). Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature *460*, 1132-1135.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K.*, et al.* (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science *341*, 651-654.

Hu, F., Lu, J., Matheson, L.S., Diaz-Munoz, M.D., Saveliev, A., and Turner, M. (2021). ORFLine: a bioinformatic pipeline to prioritise small open reading frames identifies candidate secreted small proteins from lymphocytes. Bioinformatics.

Hu, X., Sood, A.K., Dang, C.V., and Zhang, L. (2018). The role of long noncoding RNAs in cancer: the dark matter matters. Curr Opin Genet Dev *48*, 8-15.

Huang, J.Z., Chen, M., Chen, Gao, X.C., Zhu, S., Huang, H., Hu, M., Zhu, H., and Yan, G.R. (2017). A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. Mol Cell *68*, 171-184 e176.

Hubackova, S., Krejcikova, K., Bartek, J., and Hodny, Z. (2012). IL1- and TGFbeta-Nox4 signaling, oxidative stress and DNA damage response are shared features of replicative, oncogene-induced, and drug-induced paracrine 'bystander senescence'. Aging (Albany NY) *4*, 932-951.

Iglesias, J.M., Gumuzio, J., and Martin, A.G. (2017). Linking Pluripotency Reprogramming and Cancer. Stem Cells Transl Med *6*, 335-339.

Iten, L.E., and Bryant, S.V. (1973). Forelimb regeneration from different levels of amputation in the newt,Notophthalmus viridescens: Length, rate, and stages. Wilhelm Roux Arch Entwickl Mech Org *173*, 263-282.

Ji, Z., Song, R., Regev, A., and Struhl, K. (2015). Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. Elife *4*, e08890.

Johnsson, P., Lipovich, L., Grander, D., and Morris, K.V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim Biophys Acta *1840*, 1063-1071.

Jolly, M.K., Tripathi, S.C., Jia, D., Mooney, S.M., Celiktas, M., Hanash, S.M., Mani, S.A., Pienta, K.J., Ben-Jacob, E., and Levine, H. (2016). Stability of the hybrid epithelial/mesenchymal phenotype. Oncotarget *7*, 27067-27084.

Jopling, C., Boue, S., and Izpisua Belmonte, J.C. (2011). Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. Nat Rev Mol Cell Biol *12*, 79-89.

Jopling, C., Sleep, E., Raya, M., Marti, M., Raya, A., and Izpisua Belmonte, J.C. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. Nature *464*, 606-609.

Jordan, N.V., Johnson, G.L., and Abell, A.N. (2011). Tracking the intermediate stages of epithelial-mesenchymal transition in epithelial stem cells and cancer. Cell Cycle *10*, 2865-2873.

Kang, T.W., Yevsa, T., Woller, N., Hoenicke, L., Wuestefeld, T., Dauch, D., Hohmeyer, A., Gereke, M., Rudalska, R., Potapova, A.*, et al.* (2011). Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. Nature *479*, 547-551.

Kang, Y., Siegel, P.M., Shu, W., Drobnjak, M., Kakonen, S.M., Cordon-Cardo, C., Guise, T.A., and Massague, J. (2003). A multigenic program mediating breast cancer metastasis to bone. Cancer Cell *3*, 537-549.

Kareta, M.S., Gorges, L.L., Hafeez, S., Benayoun, B.A., Marro, S., Zmoos, A.F., Cecchini, M.J., Spacek, D., Batista, L.F., O'Brien, M.*, et al.* (2015). Inhibition of pluripotency networks by the Rb tumor suppressor restricts reprogramming and tumorigenesis. Cell Stem Cell *16*, 39-50.

Katoh, Y., and Katoh, M. (2009). FGFR2-related pathogenesis and FGFR2-targeted therapeutics (Review). Int J Mol Med *23*, 307-311.

Kerosuo, L., and Bronner-Fraser, M. (2012). What is bad in cancer is good in the embryo: importance of EMT in neural crest development. Semin Cell Dev Biol *23*, 320-332.

Kikuchi, K., Holdway, J.E., Werdich, A.A., Anderson, R.M., Fang, Y., Egnaczyk, G.F., Evans, T., Macrae, C.A., Stainier, D.Y., and Poss, K.D. (2010). Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes. Nature *464*, 601-605.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96-105.

Koh, M., Ahmad, I., Ko, Y., Zhang, Y., Martinez, T.F., Diedrich, J.K., Chu, Q., Moresco, J.J., Erb, M.A., Saghatelian, A.*, et al.* (2021). A short ORF-encoded transcriptional regulator. Proc Natl Acad Sci U S A *118*.

Kopp, J.L., Grompe, M., and Sander, M. (2016). Stem cells versus plasticity in liver and pancreas regeneration. Nat Cell Biol *18*, 238-245.

Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc Natl Acad Sci U S A *106*, 10171-10176.

Koussounadis, A., Langdon, S.P., Um, I.H., Harrison, D.J., and Smith, V.A. (2015). Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. Sci Rep *5*, 10775.

Kudo-Saito, C., Shirako, H., Takeuchi, T., and Kawakami, Y. (2009). Cancer metastasis is accelerated through immunosuppression during Snail-induced EMT of cancer cells. Cancer Cell *15*, 195-206.

Kuilman, T., Michaloglou, C., Vredeveld, L.C., Douma, S., van Doorn, R., Desmet, C.J., Aarden, L.A., Mooi, W.J., and Peeper, D.S. (2008). Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. Cell *133*, 1019-1031.

Kumari, N., Dwarakanath, B.S., Das, A., and Bhatt, A.N. (2016). Role of interleukin-6 in cancer progression and therapeutic resistance. Tumour Biol *37*, 11553-11572.

Kumari, R., and Jat, P. (2021). Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype. Front Cell Dev Biol *9*, 645593.

Kurz, D.J., Decary, S., Hong, Y., and Erusalimsky, J.D. (2000). Senescence-associated (beta)-galactosidase reflects an increase in lysosomal mass during replicative ageing of human endothelial cells. J Cell Sci *113 (Pt 20)*, 3613-3622.

Laberge, R.M., Sun, Y., Orjalo, A.V., Patil, C.K., Freund, A., Zhou, L., Curran, S.C., Davalos, A.R., Wilson-Edell, K.A., Liu, S.*, et al.* (2015). MTOR regulates the pro-tumorigenic senescence-associated secretory phenotype by promoting IL1A translation. Nat Cell Biol *17*, 1049-1061.

Lan, T., Chang, L., Wu, L., and Yuan, Y. (2016). Downregulation of ZEB2-AS1 decreased tumor growth and metastasis in hepatocellular carcinoma. Mol Med Rep *14*, 4606-4612.

Land, H., Parada, L.F., and Weinberg, R.A. (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature *304*, 596-602.

Laskay, U.A., Lobas, A.A., Srzentic, K., Gorshkov, M.V., and Tsybin, Y.O. (2013). Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments. J Proteome Res *12*, 5558-5569.

Laumont, C.M., Vincent, K., Hesnard, L., Audemard, E., Bonneil, E., Laverdure, J.P., Gendron, P., Courcelles, M., Hardy, M.P., Cote, C.*, et al.* (2018). Noncoding regions are the main source of targetable tumor-specific antigens. Sci Transl Med *10*.

Lee, B.Y., Han, J.A., Im, J.S., Morrone, A., Johung, K., Goodwin, E.C., Kleijer, W.J., DiMaio, D., and Hwang, E.S. (2006). Senescence-associated beta-galactosidase is lysosomal beta-galactosidase. Aging Cell *5*, 187-195.

Li, H., Collado, M., Villasante, A., Strati, K., Ortega, S., Canamero, M., Blasco, M.A., and Serrano, M. (2009). The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature *460*, 1136-1139.

Li, M., Li, X., Zhang, Y., Wu, H., Zhou, H., Ding, X., Zhang, X., Jin, X., Wang, Y., Yin, X.*, et al.* (2020). Micropeptide MIAC Inhibits HNSCC Progression by Interacting with Aquaporin 2. J Am Chem Soc *142*, 6708-6716.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q.*, et al.* (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell *7*, 51-63.

Lin, A.W., Barradas, M., Stone, J.C., van Aelst, L., Serrano, M., and Lowe, S.W. (1998). Premature senescence involving p53 and p16 is activated in response to constitutive MEK/MAPK mitogenic signaling. Genes Dev *12*, 3008-3019.

Lin, B., Coleman, J.H., Peterson, J.N., Zunitch, M.J., Jang, W., Herrick, D.B., and Schwob, J.E. (2017). Injury Induces Endogenous Reprogramming and Dedifferentiation of Neuronal Progenitors to Multipotency. Cell Stem Cell *21*, 761-774 e765.

Lin, M.F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics *27*, i275-282.

Lin, T., Ponn, A., Hu, X., Law, B.K., and Lu, J. (2010). Requirement of the histone demethylase LSD1 in Snai1-mediated transcriptional repression during epithelial-mesenchymal transition. Oncogene *29*, 4896-4904.

Liu, J., Zhang, C., Hu, W., and Feng, Z. (2019a). Tumor suppressor p53 and metabolism. J Mol Cell Biol *11*, 284-292.

Liu, X., Sun, H., Qi, J., Wang, L., He, S., Liu, J., Feng, C., Chen, C., Li, W., Guo, Y.*, et al.* (2013). Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. Nat Cell Biol *15*, 829-838.

Liu, Z., Li, C., Kang, N., Malhi, H., Shah, V.H., and Maiers, J.L. (2019b). Transforming growth factor beta (TGFbeta) cross-talk with the unfolded protein response is critical for hepatic stellate cell activation. J Biol Chem *294*, 3137-3151.

Luond, F., Sugiyama, N., Bill, R., Bornes, L., Hager, C., Tang, F., Santacroce, N., Beisel, C., Ivanek, R., Burglin, T.*, et al.* (2021). Distinct contributions of partial and full EMT to breast cancer malignancy. Dev Cell *56*, 3203-3221 e3211.

Magny, E.G., Platero, A.I., Bishop, S.A., Pueyo, J.I., Aguilar-Hidalgo, D., and Couso, J.P. (2021). Pegasus, a small extracellular peptide enhancing short-range diffusion of Wingless. Nat Commun *12*, 5660.

Magny, E.G., Pueyo, J.I., Pearl, F.M., Cespedes, M.A., Niven, J.E., Bishop, S.A., and Couso, J.P. (2013). Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. Science *341*, 1116-1120.

Mahboubi, H., and Stochaj, U. (2017). Cytoplasmic stress granules: Dynamic modulators of cell signaling and disease. Biochim Biophys Acta Mol Basis Dis *1863*, 884-895.

Maherali, N., and Hochedlinger, K. (2009). Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. Curr Biol *19*, 1718-1723.

Mahmoudi, S., Mancini, E., Xu, L., Moore, A., Jahanbani, F., Hebestreit, K., Srinivasan, R., Li, X., Devarajan, K., Prelot, L.*, et al.* (2019). Heterogeneity in old fibroblasts is linked to variability in reprogramming and wound healing. Nature *574*, 553-558.

Maiers, J.L., Kostallari, E., Mushref, M., deAssuncao, T.M., Li, H., Jalan-Sakrikar, N., Huebert, R.C., Cao, S., Malhi, H., and Shah, V.H. (2017). The unfolded protein response mediates fibrogenesis and collagen I secretion through regulating TANGO1 in mice. Hepatology *65*, 983-998.

Makarewich, C.A., Baskin, K.K., Munir, A.Z., Bezprozvannaya, S., Sharma, G., Khemtong, C., Shah, A.M., McAnally, J.R., Malloy, C.R., Szweda, L.I.*, et al.* (2018). MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid beta-Oxidation. Cell Rep *23*, 3701-3709.

Makarewich, C.A., and Olson, E.N. (2017). Mining for Micropeptides. Trends Cell Biol *27*, 685-696.

Marion, R.M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M., and Blasco, M.A. (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature *460*, 1149-1153.

Masip, M., Veiga, A., Izpisua Belmonte, J.C., and Simon, C. (2010). Reprogramming with defined factors: from induced pluripotency to induced transdifferentiation. Mol Hum Reprod *16*, 856-868.

Massague, J. (2000). How cells read TGF-beta signals. Nat Rev Mol Cell Biol *1*, 169-178.

Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., Saghatelian, A., Nakayama, K.I., Clohessy, J.G., and Pandolfi, P.P. (2017). mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. Nature *541*, 228-232.

Mendez, O., Peg, V., Salvans, C., Pujals, M., Fernandez, Y., Abasolo, I., Perez, J., Matres, A., Valeri, M., Gregori, J.*, et al.* (2018). Extracellular HMGA1 Promotes Tumor Invasion and Metastasis in Triple-Negative Breast Cancer. Clin Cancer Res *24*, 6367-6382.

Merino-Valverde, I., Greco, E., and Abad, M. (2020). The microproteome of cancer: From invisibility to relevance. Exp Cell Res *392*, 111997.

Merrell, A.J., and Stanger, B.Z. (2016). Adult cell plasticity in vivo: de-differentiation and transdifferentiation are back in style. Nat Rev Mol Cell Biol *17*, 413-425.

Milanovic, M., Fan, D.N.Y., Belenki, D., Dabritz, J.H.M., Zhao, Z., Yu, Y., Dorr, J.R., Dimitrova, L., Lenze, D., Monteiro Barbosa, I.A.*, et al.* (2018). Senescence-associated reprogramming promotes cancer stemness. Nature *553*, 96-100.

Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L., and Massague, J. (2005). Genes that mediate breast cancer metastasis to lung. Nature *436*, 518-524.

Mosteiro, L., Pantoja, C., Alcazar, N., Marion, R.M., Chondronasiou, D., Rovira, M., Fernandez-Marcos, P.J., Munoz-Martin, M., Blanco-Aparicio, C., Pastor, J.*, et al.* (2016). Tissue damage and senescence provide critical signals for cellular reprogramming in vivo. Science *354*.

Mosteiro, L., Pantoja, C., de Martino, A., and Serrano, M. (2018). Senescence promotes in vivo reprogramming through p16(INK)(4a) and IL-6. Aging Cell *17*.

Munoz-Espin, D., and Serrano, M. (2014). Cellular senescence: from physiology to pathology. Nat Rev Mol Cell Biol *15*, 482-496.

Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R., Chen, X., Kavalali, E.T.*, et al.* (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science *351*, 271-275.

Nelson, G., Wordsworth, J., Wang, C., Jurk, D., Lawless, C., Martin-Ruiz, C., and von Zglinicki, T. (2012). A senescent cell bystander effect: senescence-induced senescence. Aging Cell *11*, 345-349.

Nieto, M.A., Huang, R.Y., Jackson, R.A., and Thiery, J.P. (2016). Emt: 2016. Cell *166*, 21-45.

Ocana, O.H., Corcoles, R., Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., Barrallo-Gimeno, A., Cano, A., and Nieto, M.A. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. Cancer Cell *22*, 709-724.

Ohnishi, K., Semi, K., Yamamoto, T., Shimizu, M., Tanaka, A., Mitsunaga, K., Okita, K., Osafune, K., Arioka, Y., Maeda, T.*, et al.* (2014). Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. Cell *156*, 663-677.

Owen, H.R., Elser, M., Cheung, E., Gersbach, M., Kraus, W.L., and Hottiger, M.O. (2007). MYBBP1a is a novel repressor of NF-kappaB. J Mol Biol *366*, 725-736.

Padmanaban, V., Krol, I., Suhail, Y., Szczerba, B.M., Aceto, N., Bader, J.S., and Ewald, A.J. (2019). E-cadherin is required for metastasis in multiple models of breast cancer. Nature *573*, 439-444.

Painter, M.W., Brosius Lutz, A., Cheng, Y.C., Latremoliere, A., Duong, K., Miller, C.M., Posada, S., Cobos, E.J., Zhang, A.X., Wagers, A.J.*, et al.* (2014). Diminished Schwann cell repair responses underlie age-associated impaired axonal regeneration. Neuron *83*, 331-343.

Panopoulos, A.D., Yanes, O., Ruiz, S., Kida, Y.S., Diep, D., Tautenhahn, R., Herrerias, A., Batchelder, E.M., Plongthongkum, N., Lutz, M.*, et al.* (2012). The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. Cell Res *22*, 168-177.

Papp, B., and Plath, K. (2013). Epigenetics of reprogramming to induced pluripotency. Cell *152*, 1324-1343.

Parrinello, S., Coppe, J.P., Krtolica, A., and Campisi, J. (2005). Stromal-epithelial interactions in aging and cancer: senescent fibroblasts alter epithelial cell differentiation. J Cell Sci *118*, 485-496.

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S.*, et al.* (2018). Identification of the tumour transition states occurring during EMT. Nature *556*, 463-468.

Patraquim, P., Mumtaz, M.A.S., Pueyo, J.I., Aspden, J.L., and Couso, J.P. (2020). Developmental regulation of canonical and small ORF translation from mRNAs. Genome Biol *21*, 128.

Pauli, A., Norris, M.L., Valen, E., Chew, G.L., Gagnon, J.A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D.*, et al.* (2014). Toddler: an embryonic signal that promotes cell movement via Apelin receptors. Science *343*, 1248636.

Peeters, M.K.R., and Menschaert, G. (2020). The hunt for sORFs: A multidisciplinary strategy. Exp Cell Res *391*, 111923.

Pei, D., Shu, X., Gassama-Diagne, A., and Thiery, J.P. (2019). Mesenchymal-epithelial transition in development and reprogramming. Nat Cell Biol *21*, 44-53.

Pein, M., Insua-Rodriguez, J., Hongu, T., Riedel, A., Meier, J., Wiedmann, L., Decker, K., Essers, M.A.G., Sinn, H.P., Spaich, S.*, et al.* (2020). Metastasis-initiating cells induce and exploit a fibroblast niche to fuel malignant colonization of the lungs. Nat Commun *11*, 1494.

Philip, B., Ito, K., Moreno-Sanchez, R., and Ralph, S.J. (2013). HIF expression and the role of hypoxic microenvironments within primary tumours as protective sites driving cancer stem cell renewal and metastatic progression. Carcinogenesis *34*, 1699-1707.

Piegari, E., De Angelis, A., Cappetta, D., Russo, R., Esposito, G., Costantino, S., Graiani, G., Frati, C., Prezioso, L., Berrino, L.*, et al.* (2013). Doxorubicin induces senescence and impairs function of human cardiac progenitor cells. Basic Res Cardiol *108*, 334.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J.*, et al.* (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617-1632.

Polycarpou-Schwarz, M., Gross, M., Mestdagh, P., Schott, J., Grund, S.E., Hildenbrand, C., Rom, J., Aulmann, S., Sinn, H.P., Vandesompele, J.*, et al.* (2018). The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation. Oncogene *37*, 4750-4768.

Prakash, V., Carson, B.B., Feenstra, J.M., Dass, R.A., Sekyrova, P., Hoshino, A., Petersen, J., Guo, Y., Parks, M.M., Kurylo, C.M.*, et al.* (2019). Ribosome biogenesis during cell cycle arrest fuels EMT in development and disease. Nat Commun *10*, 2110.

Prensner, J.R., Enache, O.M., Luria, V., Krug, K., Clauser, K.R., Dempster, J.M., Karger, A., Wang, L., Stumbraite, K., Wang, V.M.*, et al.* (2021). Noncanonical open reading frames encode functional proteins essential for cancer cell survival. Nat Biotechnol *39*, 697-704.

Pueyo, J.I., Magny, E.G., Sampson, C.J., Amin, U., Evans, I.R., Bishop, S.A., and Couso, J.P. (2016). Hemotin, a Regulator of Phagocytosis Encoded by a Small ORF and Conserved across Metazoans. PLoS Biol *14*, e1002395.

Puisieux, A., Brabletz, T., and Caramel, J. (2014). Oncogenic roles of EMT-inducing transcription factors. Nat Cell Biol *16*, 488-494.

Quaife, N.M., Chothani, S., Schulz, J.F., Lindberg, E.L., Vanezis, K., Adami, E., O'Fee, K., Greiner, J., Litvinukova, M., van Heesch, S.*, et al.* (2022). LINC01013 Is a Determinant of Fibroblast Activation and Encodes a Novel Fibroblast-Activating Micropeptide. J Cardiovasc Transl Res.

Quelle, D.E., Zindy, F., Ashmun, R.A., and Sherr, C.J. (1995). Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. Cell *83*, 993-1000.

Raff, M. (2003). Adult stem cell plasticity: fact or artifact? Annu Rev Cell Dev Biol *19*, 1-22.

Raghavan, S., Winter, P.S., Navia, A.W., Williams, H.L., DenAdel, A., Lowder, K.E., Galvez-Reyes, J., Kalekar, R.L., Mulugeta, N., Kapner, K.S.*, et al.* (2021). Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. Cell *184*, 6119-6137 e6126.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M.*, et al.* (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65-70.

Redmer, T., Diecke, S., Grigoryan, T., Quiroga-Negreira, A., Birchmeier, W., and Besser, D. (2011). E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. EMBO Rep *12*, 720-726.

Rheinbay, E., Nielsen, M.M., Abascal, F., Wala, J.A., Shapira, O., Tiao, G., Hornshoj, H., Hess, J.M., Juul, R.I., Lin, Z.*, et al.* (2020). Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature *578*, 102-111.

Ritschka, B., Storer, M., Mas, A., Heinzmann, F., Ortells, M.C., Morton, J.P., Sansom, O.J., Zender, L., and Keyes, W.M. (2017). The senescence-associated secretory phenotype induces cellular plasticity and tissue regeneration. Genes Dev *31*, 172-183.

Rodini, C.O., Suzuki, D.E., Saba-Silva, N., Cappellano, A., de Souza, J.E., Cavalheiro, S., Toledo, S.R., and Okamoto, O.K. (2012). Expression analysis of stem cell-related genes reveal OCT4 as a predictor of poor clinical outcome in medulloblastoma. J Neurooncol *106*, 71-79.

Rognoni, E., and Watt, F.M. (2018). Skin Cell Heterogeneity in Development, Wound Healing, and Cancer. Trends Cell Biol *28*, 709-722.

Ruiz-Orera, J., Verdaguer-Grau, P., Villanueva-Canas, J.L., Messeguer, X., and Alba, M.M. (2018). Translation of neutrally evolving peptides provides a basis for de novo gene evolution. Nat Ecol Evol *2*, 890-896.

Saiki, Y., Ishimaru, S., Mimori, K., Takatsuno, Y., Nagahara, M., Ishii, H., Yamada, K., and Mori, M. (2009). Comprehensive analysis of the clinical significance of inducing pluripotent stemness-related gene expression in colorectal cancer cells. Ann Surg Oncol *16*, 2638-2644.

Sakurai, K., Talukdar, I., Patil, V.S., Dang, J., Li, Z., Chang, K.Y., Lu, C.C., Delorme-Walker, V., Dermardirossian, C., Anderson, K.*, et al.* (2014). Kinome-wide functional analysis highlights the role of cytoskeletal remodeling in somatic cell reprogramming. Cell Stem Cell *14*, 523-534.

Salminen, A., Kauppinen, A., and Kaarniranta, K. (2012). Emerging role of NF-kappaB signaling in the induction of senescence-associated secretory phenotype (SASP). Cell Signal *24*, 835-845.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell *7*, 64-77.

Sanges, D., Romo, N., Simonte, G., Di Vicino, U., Tahoces, A.D., Fernandez, E., and Cosma, M.P. (2013). Wnt/beta-catenin signaling triggers neuron reprogramming and regeneration in the mouse retina. Cell Rep *4*, 271-286.

Sapieha, P., and Mallette, F.A. (2018). Cellular Senescence in Postmitotic Cells: Beyond Growth Arrest. Trends Cell Biol *28*, 595-607.

Scheel, C., Eaton, E.N., Li, S.H., Chaffer, C.L., Reinhardt, F., Kah, K.J., Bell, G., Guo, W., Rubin, J., Richardson, A.L.*, et al.* (2011). Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast. Cell *145*, 926-940.

Seki, T., Yuasa, S., Oda, M., Egashira, T., Yae, K., Kusumoto, D., Nakata, H., Tohyama, S., Hashimoto, H., Kodaira, M.*, et al.* (2010). Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. Cell Stem Cell *7*, 11-14.

Senft, D., and Ronai, Z.A. (2016). Adaptive Stress Responses During Tumor Metastasis and Dormancy. Trends Cancer *2*, 429-442.

Serrano, M., Hannon, G.J., and Beach, D. (1993). A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. Nature *366*, 704-707.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. Cell *88*, 593-602.

Shaw, T.J., and Martin, P. (2016). Wound repair: a showcase for cell plasticity and migration. Curr Opin Cell Biol *42*, 29-37.

Slavoff, S.A., Heo, J., Budnik, B.A., Hanakahi, L.A., and Saghatelian, A. (2014). A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. J Biol Chem *289*, 10950-10957.

Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L., and Saghatelian, A. (2013). Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol *9*, 59-64.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994-1004.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230-240.

Stein, C.S., Jadiya, P., Zhang, X., McLendon, J.M., Abouassaly, G.M., Witmer, N.H., Anderson, E.J., Elrod, J.W., and Boudreau, R.L. (2018). Mitoregulin: A lncRNA-Encoded

Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. Cell Rep *23*, 3710-3720 e3718.

Stemmler, M.P., Eccles, R.L., Brabletz, S., and Brabletz, T. (2019). Non-redundant functions of EMT transcription factors. Nat Cell Biol *21*, 102-112.

Su, J.B., Pei, D.Q., and Qin, B.M. (2013). Roles of small molecules in somatic cell reprogramming. Acta Pharmacol Sin *34*, 719-724.

Subramanyam, D., Lamouille, S., Judson, R.L., Liu, J.Y., Bucay, N., Derynck, R., and Blelloch, R. (2011). Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. Nat Biotechnol *29*, 443-448.

Sun, L., Wang, W., Han, C., Huang, W., Sun, Y., Fang, K., Zeng, Z., Yang, Q., Pan, Q., Chen, T.*, et al.* (2021). The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. Mol Cell *81*, 4493-4508 e4499.

Taipale, J., and Beachy, P.A. (2001). The Hedgehog and Wnt signalling pathways in cancer. Nature *411*, 349-354.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Tasdogan, A., Faubert, B., Ramesh, V., Ubellacker, J.M., Shen, B., Solmonson, A., Murphy, M.M., Gu, Z., Gu, W., Martin, M.*, et al.* (2020). Metabolic heterogeneity confers differences in melanoma metastatic potential. Nature *577*, 115-120.

Tata, P.R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B.M., Vinarsky, V., Cho, J.L., Breton, S., Sahay, A.*, et al.* (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. Nature *503*, 218-223.

Teo, G., Liu, G., Zhang, J., Nesvizhskii, A.I., Gingras, A.C., and Choi, H. (2014). SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. J Proteomics *100*, 37-43.

Teslaa, T., and Teitell, M.A. (2015). Pluripotent stem cell energy metabolism: an update. EMBO J *34*, 138-153.

Tetteh, P.W., Basak, O., Farin, H.F., Wiebrands, K., Kretzschmar, K., Begthel, H., van den Born, M., Korving, J., de Sauvage, F., van Es, J.H.*, et al.* (2016). Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. Cell Stem Cell *18*, 203-213.

Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871-890.

Torborg, S.R., Li, Z., Chan, J.E., and Tammela, T. (2022). Cellular and molecular mechanisms of plasticity in cancer. Trends Cancer *8*, 735-746.

Utikal, J., Polo, J.M., Stadtfeld, M., Maherali, N., Kulalert, W., Walsh, R.M., Khalil, A., Rheinwald, J.G., and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature *460*, 1145-1148.

Valadkhan, S., and Valencia-Hipolito, A. (2016). lncRNAs in Stress Response. Curr Top Microbiol Immunol *394*, 203-236.

van Es, J.H., Sato, T., van de Wetering, M., Lyubimova, A., Yee Nee, A.N., Gregorieff, A., Sasaki, N., Zeinstra, L., van den Born, M., Korving, J.*, et al.* (2012). Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. Nat Cell Biol *14*, 1099-1104.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029-1033.

Vincent, T., Neve, E.P., Johnson, J.R., Kukalev, A., Rojo, F., Albanell, J., Pietras, K., Virtanen, I., Philipson, L., Leopold, P.L.*, et al.* (2009). A SNAIL1-SMAD3/4 transcriptional repressor complex promotes TGF-beta mediated epithelial-mesenchymal transition. Nat Cell Biol *11*, 943-950.

Wang, Y., Hu, L., Zheng, Y., and Guo, L. (2019). HMGA1 in cancer: Cancer classification by location. J Cell Mol Med *23*, 2293-2302.

Wang, Z.Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mossinger, K., Bruning, T., Rummel, C., Grutzner, F., Cardoso-Moreira, M., Janich, P.*, et al.* (2020). Transcriptome and translatome co-evolution in mammals. Nature *588*, 642-647.

Weidle, U.H., Birzele, F., and Tiefenthaler, G. (2018). Potential of Protein-based Anti-metastatic Therapy with Serpins and Inter alpha-Trypsin Inhibitors. Cancer Genomics Proteomics *15*, 225-238.

Wells, R.G. (2022). How collagen becomes 'stiff'. Elife *11*.

Wu, S., Zhang, L., Deng, J., Guo, B., Li, F., Wang, Y., Wu, R., Zhang, S., Lu, J., and Zhou, Y. (2020). A Novel Micropeptide Encoded by Y-Linked LINC00278 Links Cigarette Smoking and AR Signaling in Male Esophageal Squamous Cell Carcinoma. Cancer Res *80*, 2790-2803.

Xiao, Y., Jiang, Z., Li, Y., Ye, W., Jia, B., Zhang, M., Xu, Y., Wu, D., Lai, L., Chen, Y.*, et al.* (2015). ANGPTL7 regulates the expansion and repopulation of human hematopoietic stem and progenitor cells. Haematologica *100*, 585-594.

Xie, C., Wang, F.Y., Sang, Y., Chen, B., Huang, J.H., He, F.J., Li, H., Zhu, Y., Liu, X., Zhuang, S.M.*, et al.* (2022). Mitochondrial Micropeptide STMP1 Enhances Mitochondrial Fission to Promote Tumor Metastasis. Cancer Res *82*, 2431-2443.

Xiong, S., Feng, Y., and Cheng, L. (2019). Cellular Reprogramming as a Therapeutic Target in Cancer. Trends Cell Biol *29*, 623-634.

Xu, M., Tchkonia, T., Ding, H., Ogrodnik, M., Lubbers, E.R., Pirtskhalava, T., White, T.A., Johnson, K.O., Stout, M.B., Mezera, V.*, et al.* (2015). JAK inhibition alleviates the cellular senescence-associated secretory phenotype and frailty in old age. Proc Natl Acad Sci U S A *112*, E6301-6310.

Xue, W., Zender, L., Miething, C., Dickins, R.A., Hernando, E., Krizhanovsky, V., Cordon-Cardo, C., and Lowe, S.W. (2007). Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. Nature *445*, 656-660.

Yang, H., Wang, H., Ren, J., Chen, Q., and Chen, Z.J. (2017). cGAS is essential for cellular senescence. Proc Natl Acad Sci U S A *114*, E4612-E4620.

Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G.*, et al.* (2020). Guidelines and definitions for research on epithelial-mesenchymal transition. Nat Rev Mol Cell Biol *21*, 341-352.

Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res *43*, W174-181.

Yao, Y., and Wang, C. (2020). Dedifferentiation: inspiration for devising engineering strategies for regenerative medicine. NPJ Regen Med *5*, 14.

Ye, X., and Weinberg, R.A. (2015). Epithelial-Mesenchymal Plasticity: A Central Regulator of Cancer Progression. Trends Cell Biol *25*, 675-686.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R.*, et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Zampetaki, A., Albrecht, A., and Steinhofel, K. (2018). Long Non-coding RNA Structure and Function: Is There a Link? Front Physiol *9*, 1201.

Zhang, M., Zhao, K., Xu, X., Yang, Y., Yan, S., Wei, P., Liu, H., Xu, J., Xiao, F., Zhou, H.*, et al.* (2018). A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. Nat Commun *9*, 4475.

Zhang, S., Reljic, B., Liang, C., Kerouanton, B., Francisco, J.C., Peh, J.H., Mary, C., Jagannathan, N.S., Olexiouk, V., Tang, C.*, et al.* (2020). Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly. Nat Commun *11*, 1312.

Zhang, Y.E. (2017). Non-Smad Signaling Pathways of the TGF-beta Family. Cold Spring Harb Perspect Biol *9*.

Zheng, X., Chen, L., Zhou, Y., Wang, Q., Zheng, Z., Xu, B., Wu, C., Zhou, Q., Hu, W., Wu, C.*, et al.* (2019). A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling. Mol Cancer *18*, 47.

Zhong, J., Liu, C., Zhang, Q.H., Chen, L., Shen, Y.Y., Chen, Y.J., Zeng, X., Zu, X.Y., and Cao, R.X. (2017). TGF-beta1 induces HMGA1 expression: The role of HMGA1 in thyroid cancer proliferation and invasion. Int J Oncol *50*, 1567-1578.

Zhu, Y., Orre, L.M., Johansson, H.J., Huss, M., Boekel, J., Vesterlund, M., Fernandez-Woodbridge, A., Branca, R.M.M., and Lehtio, J. (2018). Discovery of coding regions in the human genome by integrated proteogenomics analysis workflow. Nat Commun *9*, 903.

Zhuang, J., Lu, Q., Shen, B., Huang, X., Shen, L., Zheng, X., Huang, R., Yan, J., and Guo, H. (2015). TGFbeta1 secreted by cancer-associated fibroblasts induces epithelial-mesenchymal transition of bladder cancer cells through lncRNA-ZEB2NAT. Sci Rep *5*, 11924.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406-3415.

Additionally to the work presented in this thesis, the following publication included as an annex has been published

Experimental Cell Research 392 (2020) 111997

Contents lists available at ScienceDirect

Experimental Cell Research

journal homepage: www.elsevier.com/locate/yexcr

The microproteome of cancer: From invisibility to relevance

Iñaki Merino-Valverde[1], Emanuela Greco[1], María Abad[**]

Vall d'Hebron Institute of Oncology (VHIO), Barcelona, 08035, Spain

ARTICLE INFO

Keywords:
Microprotein
Micropeptide
SEPs
sORF
smORF
alt-ORF
Cancer
Tumor suppressors

ABSTRACT

Recent findings have revealed that many genomic regions previously annotated as non-protein coding actually contain small open reading frames, smaller that 300 bp, that are transcribed and translated into evolutionary conserved microproteins. To date, only a small subset of them have been functionally characterized, but they play key functions in fundamental processes such as DNA repair, RNA processing and metabolism regulation. This emergent field seems to hide a new category of molecular regulators with clinical potential. In this review, we focus on its relevance for cancer. Following Hanahan and Weinberg's classification of the hallmarks of cancer, we provide an overview of those microproteins known to be implicated in cancer or those that, based on their function, are likely to play a role in cancer. The resulting picture is that while we are at the very early times of this field, it holds the promise to provide crucial information to understand cancer biology.

1. Introduction

Recent advances coming from computational analyses, peptidomics and ribosome profiling have revealed that our proteome includes a new class of small proteins produced by the translation of small open reading frames shorter than 300 bp in length, generating proteins that have been called microproteins (also known as micropeptides or SEPs, from small ORF-encoded peptides) [1,2]. The main reason why microproteins have been overlooked till recently is that most ORF-prediction algorithms -including the one used by FANTOM annotation consortium- placed an arbitrary cut off of 300 bp, missing the proteins below 100 amino acids [3,4]. Although nomenclature has been inconsistent in the field, based on their location, the ORFs encoding microproteins can been classified as 1) small ORFs, "sORFs", when they are located in assumed non-coding transcripts such as lncRNAs, miRNAs and circRNAs, or 2) "alt-ORFs", when they are inside annotated coding genes starting from alternative start codons. Alt-ORFs can be located in the UTRs (typically called "uORFs" when they are in the 5′UTR) or overlapping the reference coding sequence. In this review, we focus on the microproteins derived from sORFs. Although only a subset of them have been functionally characterized, growing evidences demonstrate that microproteins are indeed active proteins playing important functions in a plethora of processes including RNA processing, DNA repair, metabolism regulation and regeneration [5–9]. Microproteins might be particularly well suited to fine-tune complex processes as regulation of enzyme activity, intracellular signal transduction and cell surface signaling, but extensive research in the field is needed to decipher additional functions of the microproteome [1]. These findings open a new category of molecular regulators with implications from basic research to the clinical setting.

Cancer is a complex and multistep disease in which normal cells, through the succession of several genetic and epigenetic events acquire the capacity to grow, invade adjacent tissues, disseminate and ultimately colonize distant organs. Although the specific mechanisms that allow neoplastic transformation and metastasis may vary between different cancer types, there are common regulatory circuits that collectively govern carcinogenesis. In 2000, Hanahan and Weinberg published a seminal paper in which they postulated six capabilities shared by most human tumors [10], which was revisited in 2011 to finally include eight hallmarks of cancer: sustained proliferative signaling, evasion of growth suppressors, resistance to cell death, replicative immortality, induction of angiogenesis, activation of invasion and metastasis, reprogramming of energy metabolism and the evasion of immune destruction (Fig. 1) [10]. Moreover, they proposed two enabling characteristics that represent the mechanisms by which the hallmarks of cancer are acquired: genome instability and the inflammatory state of premalignant lesions (Fig. 1). Although in the last decades we have witnessed a great advance in molecular oncology, we are still far from fully understanding how the hallmarks of cancer are acquired and maintained to sustain tumors. The new information emerging from the study of microproteins suggests that they constitute an important source of cancer regulators implicated in multiple hallmarks of cancer.

* Corresponding author.
E-mail address: mabad@vhio.net (M. Abad).
[1] Authors contributed equally to this work.

https://doi.org/10.1016/j.yexcr.2020.111997
Received 8 January 2020; Received in revised form 2 April 2020; Accepted 3 April 2020
Available online 14 April 2020

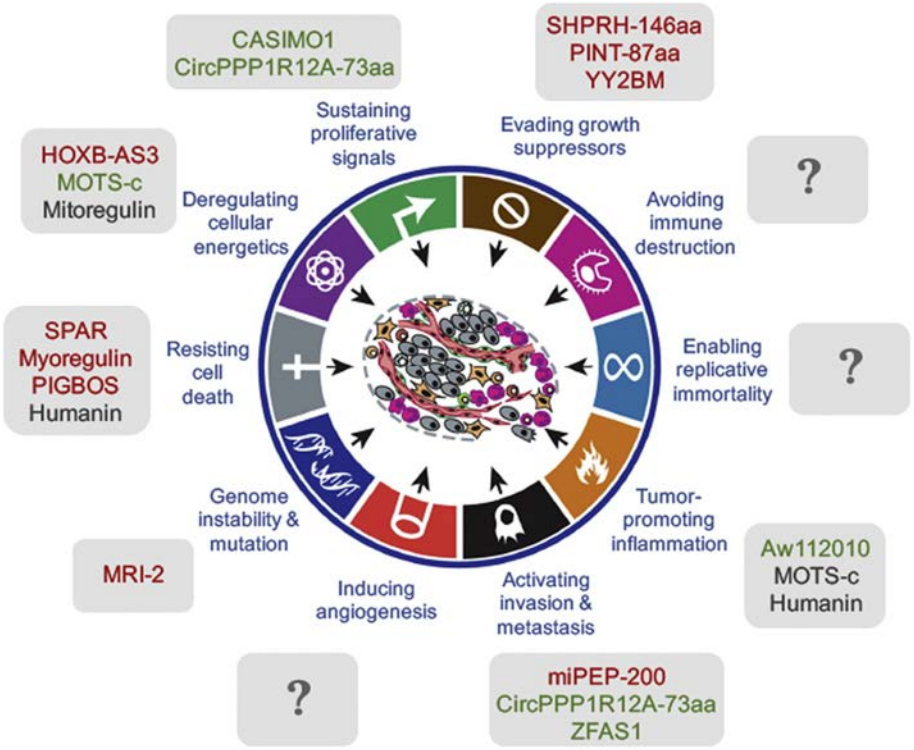


Fig. 1. Microproteins as novel regulators of the hallmarks of cancer. A subset of microproteins have been functionally characterized and have been directly related with cancer; some others, based on their function, are likely to be related with cancer. The figure represents the hallmarks of cancer defined by Hanahan and Weinberg and their related microproteins. In green, the microproteins that promotes or activate the hallmark and, in red, the microproteins that function inhibiting or blocking the hallmark. The ones in grey need further investigation to be classified. Of interest, those depicted in red represent tumor suppressor microproteins with potential pharmacological activity, while those in green are pro-oncogenic peptides that could be targeted in the clinic. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

1.1. Sustaining proliferation signals

In physiologic conditions, mitogenic signals are strictly controlled to ensure the maintenance of normal tissue architecture and homeostasis. By contrast, cancer cells are mitogenically overstimulated. Such mitogenic hyperactivity can be achieved in several ways: First, neoplastic cells can produce their own growth-promoting signals in an autocrine manner or, alternatively, produce paracrine factors that stimulate the release of mitogens by neighbor stromal cells. Second, specific somatic mutations can trigger constitutive activation of growth factor receptors or its downstream components, converting those elements of the signaling cascades in "*bona fide* oncogenes". Recent studies have shown the importance of some microproteins regulating mitogenic signaling. The first sORF described with oncogenic activity has been Cancer-Associated Small Integral Membrane Open reading frame 1 (CASIMO1). CASIMO1 is expressed in hormone-dependent breast cancer during all stages of malignancy, and its deficiency reduces cell proliferation in several breast cancer cell lines [11]. CASIMO1 interacts with squalene epoxidase (SQLE), an oncogene that promotes ERK phosphorylation and MAPK pathway activation [12]. Remarkably, CASIMO1 deficiency reduces ERK phosphorylation while exogenous overexpression of SQLE is sufficient to rescue the loss of CASIMO1, suggesting that CASIMO1 might be modulating ERK activation via SQLE interaction [11]. These observations are in line with a role of CASIMO1 as an oncogene, acting as a positive regulator of MAPK cascade in breast cancer cells. More recently, another microprotein has been proposed to act as a positive regulator of the Hippo-Yap pathway in colon cancer. CircPPP1R12A-73aa is a microprotein encoded by CircPPP1R12A, the most abundant circular RNA (circRNA) in colon cancer, and its overexpression leads to increased cell proliferation [13]. CircPPP1R12A-73aa induces the transcriptional upregulation of Hippo-Yap pathway components and increases YAP1 protein levels, suggesting a possible role of CircPPP1R12A-73aa as an activator of the Hippo pathway [13]. Although CircPPP1R12A-73aa's mechanism of action needs to be further studied, together with CASIMO1 exemplify the relevance of sORFs regulating mitogenic signals, and how they can be exploited by cancer cells to sustain their proliferation needs.

1.2. Evading growth suppressors

In homeostasis, powerful signaling programs block the proliferation of damaged or potentially malignant cells. Signals that activate growth suppression are integrated by the cell in a highly complex manner to decide whether to halt cell-cycle progression, activate apoptotic programs or enter senescence [14,15]. These signaling programs are mainly governed by tumor suppressor proteins and, for neoplastic transformation to occur, these tumor-suppressing mechanisms need to be inactivated [16–18]. Thus, a good understanding of tumor suppression pathways is crucial and might reveal novel therapeutic options in cancer. In this regard, the sORF-encoded proteome can provide new insights on the biology of tumor suppression mechanisms and could represent a novel source of therapeutic agents. Several microproteins have already been shown to play a role in regulating tumor suppression mechanisms through different ways. The *LINC-PINT* gene was already reported to produce a lncRNA regulated by p53 [19], but in its circular form, it contains a sORF that encodes an 87-amino acid microprotein. This microprotein, PINT-87aa, suppresses tumorigenic capabilities of glioblastoma cell lines *in vitro* including cell proliferation, self-renewal and anchorage independent growth, and its deficiency results in increased tumor burden *in vivo*. The authors showed that this microprotein interacts with the polymerase associated factor (PAF1) complex, essential for RNA II polymerase binding and transcription elongation [20]. The interaction of PINT-87aa with PAF1 pauses RNA II polymerase at specific oncogene promoters -such as Cyclin D1, CPEB1, c-MYC and SOX2-impairing their transcription [21].

Although it is bigger than 100 amino acids, and therefore strictly not generated from a sORF, another interesting example of the coding potential of circ-RNAs is SHPRH-146aa. SHPRH is an E3-ubiquitin ligase that promotes the degradation of PCNA (Proliferating Cell Nuclear Antigen), impeding cell cycle progression through S-phase. Recently, Zhang and colleagues have revealed a circular RNA derived from the *SHPRH* primary transcript that codes for SHPRH-146aa. SHPRH-146aa stabilizes SHRPH by preventing its degradation, and thereby promoting the ubiquitination of PCNA. Accordingly, the overexpression of SHPRH-146aa reduces the proliferation of glioma cell lines *in vitro* and *in vivo*

[22]. Importantly, both microproteins, SHPRH-146aa and PINT-87aa are silenced or downregulated in glioblastoma [19][22], further supporting their tumor suppressor potential and opening the possibility of using these small proteins as therapeutic agents.

More recently, the Y-chromosome-linked lncRNA LINC00278 has been shown to encode YY1BM, a microprotein with tumor suppressor activity in esophageal squamous cell carcinoma (ESCC). This microprotein induces apoptosis through the androgen receptor pathway under nutrient deprivation. Interestingly, YY1BM is downregulated by cigarette smoking in human males with ESCC, increasing the survival of cancer cells under nutrient deprivation. Moreover, intratumoral injection of the purified microprotein showed a therapeutic effect in xenograft models, suggesting its potential as a tumor suppressor agent [23].

Last, it is worth to mention that some identified microproteins, like NoBody, have not been directly linked with cancer but they regulate fundamental processes that can impact on cancer cells, such as mRNA decay [5]. Nonsense Mediated Decay is a complex process that can be exploited by cancer cells to degrade the mRNA of tumor suppressor genes. On the other hand, it can also be used as a therapeutic intervention to target oncogene-encoding mRNAs [24]. Thus, the role of this microprotein (and many others) as an oncogene or as a tumor suppressor might be highly tumor specific and dependent on the cellular context.

1.3. Resisting cell death

Regulation of the balance between cell death and survival is critical for maintaining tissue homeostasis. The induction of programmed cell death by apoptosis is a natural barrier for neoplastic transformation [25]. Signals that activate apoptosis are sensed and integrated, among others, by the pro- and anti-apoptotic proteins of the Bcl-2 family. Anti-apoptotic members of the family (Bcl-2, Bcl-x_L, Bcl-w, Mcl1, A1) interact with pro-apoptotic proteins Bax and Bak, suppressing their function. Upon certain pro-apoptotic stimuli, this interaction is broken allowing Bax and Bak to disrupt mitochondrial membrane, which releases cytochrome c to the cytosol, activating in turn the caspases cascade that ultimately disassembles the cell [26]. Interestingly, certain sORFs appear to be important fine-tuning regulators of this process. The microprotein Humanin (HN) is encoded by a sORF in the mitochondrial 16s ribosomal RNA gene, although a nuclear origin of this microprotein has not been ruled out yet due to the presence of similar ORFs in the nuclear genome [27]. HN was first described as a neuroprotective agent in Alzheimer disease [28]. Further studies have shown that its cytoprotective activity is, at least in part, due to its ability to block apoptosis through direct interaction with Bax [29] or by binding and inhibiting the Bax activator BimEL [30]. Regarding cancer, HN is expressed in gastric and bladder cancer and induces chemotherapy resistance [31]. On the other hand, the cytoprotective activity of HN could be beneficial for normal tissues, given that HN also reduces the side-effects of chemotherapy in non-cancer cells [31]. Further studies are needed to clarify the potential role of HN in the clinical setting.

Together with apoptotic suppression, cancer cells activate cell survival mechanisms to avoid cell death. mTOR is a serine/threonine kinase that integrates multiple environmental cues, and it is activated to promote cell growth and survival in favorable conditions [32]. Hyperactivation of mTOR has been reported in more than 70% of cancers [33] and, therefore, mTOR inhibition is an approach currently used in anti-cancer therapies. SPAR is a lncRNA-encoded microprotein that inhibits mTORC1 by its interaction with the lysosomal v-ATPase. Depletion of SPAR *in vivo* has been shown to improve muscle regeneration through higher mTORC1 activity [34]. Among the multiple mechanisms that regulate mTORC1, SPAR appears to reduce the amino-acid sensing route of mTORC1 activation. Although it has not been tested, SPAR expression could have anti-tumor properties in specific cancer types where mTORC1 activation occurs through amino-acid sensing pathways, as it is the case of some lymphomas [35].

On the other hand, the activation of the cellular stress response is an essential mechanism that helps healthy cells to cope with stress and damage, and it is co-opted by malignant cells to thrive in highly adverse conditions and avoid cell death. Accordingly, Unfolded Protein Response (UPR) and endoplasmic reticulum (ER)-stress response have been reported to be upregulated in many cancers, helping them to deal with high protein synthesis demands while protecting them from stress-induced cell death [36,37]. These pro-survival responses are highly dependent on cytosolic Ca^{2+} concentration and the ER is the main responsible of intracellular Ca^{2+} storage [38]. MYOREGULIN (MLN), encoded by a previously annotated long non-coding RNA works as an inhibitor of the Ca^{2+} ATPase pump SERCA [39]. Inhibition of SERCA activity increases cytosolic Ca^{2+} concentration, sensitizing cells to cell death. For this reason, SERCA inhibition is being tested as a potential therapy in tumors with hyperactivation of Ca^{2+} channeling activity [40]. Even though MLN has not been related with cancer so far and in homeostasis its expression is restricted to skeletal muscle, several SERCA-inhibitory microproteins have been reported to control Ca^{2+} signaling in a tissue specific manner [41] and their expression might be useful to tackle cancers with high SERCA activity.

Finally, PIGBOS is a novel microprotein encoded by an antisense of the *PIGB* gene. This microprotein localizes in the mitochondrial outer membrane interacting with the ER through the CLCC1 protein. Downregulation of PIGBOS induces apoptosis by increasing sensitivity to chemically-induced UPR [42]. Whereas the molecular mechanism of PIGBOS function has not been described yet, it is reasonable to think that its inhibition in cancer cells may have therapeutic potential by sensitizing cells to UPR and forcing them to enter apoptosis.

1.4. Activation of invasion and metastasis

Carcinomas are tumors that arise from epithelial tissues. The progression from localized tumor to invasive carcinoma and distal metastasis requires changes in cell morphology and in cell-cell and cell-extracellular matrix (ECM) attachment. The “epithelial-to-mesenchymal transition” (EMT) is a cellular program involved in embryonic morphogenesis and wound healing. Cancer cells can also activate the EMT program to acquire the invasive phenotype needed for metastatic spread [43]. Biological traits acquired during EMT include the loss of adherent junctions, the acquisition of spindle/fibroblastic morphology, the expression of matrix-degradation enzymes, increased motility and resistance to apoptosis. Importantly, while EMT is needed for metastasis, it must be reversed to colonize a new organ through a process known as “mesenchymal-to-epithelial transition” (MET) [43]. Of note, it has been recently proposed that cancer cells may acquire a “plastic-hybrid state” or a “partial EMT” state. According to this vision, cancer cells acquire mesenchymal features while continuing to express epithelial traits, resulting in a selective advantage during the metastatic process [44].

Many non-coding RNAs have been shown to regulate EMT, highlighting the importance of miR-200a and miR-200b as key negative EMT regulators, which are usually epigenetically repressed in cancer cells [45]. Of note, a recent study has identified two potential microproteins encoded by the miR-200a and miR-200b pri-miRNAs, the precursor transcripts of the miRNAs, which have been named miPEP-200a and miPEP-200b. The expression of these sORFs seems to be associated with a decreased migration in wound healing assay and with diminished expression of the mesenchymal marker Vimentin [46]. More recently, the lncRNA ZFAS1 has been shown to translate a microprotein that is proposed to promote cell migration by elevating intracellular reactive oxygen species (ROS) production [47]. Additionally, the microprotein CircPPP1R12A-73aa (described above) increases cancer cell migration and invasion *in vitro* and metastasis development *in vivo* [13], possibly reflecting the activation of EMT by the Hippo-Yap pathway [48,49].

Finally, recent evidences suggest that cell-to-cell fusion events,

especially between cancer cells and immune cells, could contribute to the acquisition of metastatic behaviors [50]. Importantly, microproteins can regulate cell fusion, as it has been shown for MYOMIXER, an 84-amino acid peptide necessary for heterotypic fibroblast-myoblast fusion [51]. Although speculative, it is possible that there are non-identified microproteins playing important functions in metastasis by regulating cell fusion. Although more investigations are needed to have a complete picture, all together these discoveries point to the microproteome as a source of regulators of cancer invasion and metastasis.

1.5. Deregulating cellular energetics

Another important feature of cancer cells is that they reprogram their metabolism in different ways to comply with their highly demanding energetic needs. One of these strategies is to rely on glycolysis rather than on oxidative phosphorylation (OXPHOS) as their primary energy production mechanism, even in the presence of oxygen [52]. The process of metabolic reprogramming results from a complex regulation between mitochondrial genes, tumor suppressors and oncogenic pathways and its targeting is currently being tested as a therapeutic strategy [53]. Importantly, microproteins have emerged as important regulators of this process with potential clinical implications. The lncRNA HOXB-AS3 has been shown to produce a 53-amino acid microprotein involved in RNA splicing. Huang and collaborators showed that HOXB-AS3 is downregulated in colorectal cancer cells, which changes the splicing of Pyruvate Kinase M (PKM) pre-mRNA to re-express the embryonic isoform PKM2, that favors glycolytic activity. By contrast, expression of HOXB-AS3 peptide favors the expression of the adult isoform (PKM1), that promotes oxidative phosphorylation. Collectively, they demonstrate that overexpression of HOXB-AS3 peptide in colorectal cancer cell lines attenuates their oncogenic capacity by altering their use of glucose metabolism [6]. Other microproteins have also been described to play a role in metabolic regulation, although their role in cancer is yet to be investigated. Specifically, MITOREGULIN (also called MOXI or MPM) is a 56-amino acid microprotein encoded by LINC00116, a muscle-enriched lncRNA. The function of this protein has been proposed to rely on its interaction with different inner mitochondrial proteins, increasing respiratory efficiency through the stabilization of supercomplexes in the electron transport chain [54,55] and promoting long-chain fatty acid ß-oxidation [56]. It would be of interest to study the role of this microprotein in cancer cell metabolism. Finally, the mitochondrial genome also plays a role in metabolic activity and, despite its small size, it has been described to contain several sORFs, like the one encoding for MOTS-c (mitochondrial open reading frame of the 12*s* rRNA-c). *In vitro*, MOTS-c increases glucose uptake, glycolytic activity and AMPK activation, while reducing oxygen consumption rate. Accordingly, MOTS-c improves metabolic parameters associated with obesity *in vivo* [57]. Given that MOTS-c favors a glycolytic program and activates AMPK, it might be beneficial for cancer cells, which would be interesting to be addressed in the future.

1.6. Enabling characteristic: genomic instability and mutations

As proposed by Hanahan and Weinberg [10], accumulation of genomic alterations during carcinogenesis is an event that enables the acquisition of the core hallmarks discussed above. Tumor progression can be seen as a succession of clonal expansions: mutations are randomly accumulated in the pool of cancer cells and, eventually, advantaged genotypes enabling cell survival and growth are selected under the pressure of environmental stimuli. Due to the adaptive advantage that a high mutational rate confers to cancer cells, the components of the genomic maintenance machinery are often affected in neoplasia. Typically, the accumulation of mutations can be accelerated by defects in the sensors of DNA damage, in components of DNA repair machinery and/or in effectors that force damaged cells to enter senescence or apoptosis [58]. Thus, microproteins involved in the maintenance of genome stability may behave as tumor suppressor genes and are likely to be affected in cancer cells. Slavoff and collaborators identified MRI-2 as a 69-amino acid microprotein coded by *C7orf49* gene, also known as “modulator of retrovirus infection homolog” (MRI) [8]. MRI-2 interactome analysis revealed that this microprotein interacts with Ku70 and Ku80 proteins, the two subunits of the heterodimeric protein Ku, key effector of the non-homologous-end-joining (NHEJ) pathway for DNA double-strand break (DSB) repair. When a DNA DSB occurs, the first protein that binds to the break is the Ku heterodimer, which allows the recruitment of the DNA-dependent protein kinase (DNA-PK) and additional factors which in turn repair the DSB [59]. MRI-2 is accumulated in the nuclei of DBS-induced cells, and recombinant MRI-2 increases NHEJ *in vitro* [8]. The mechanism by which MRI-2 enhances NHEJ has not been fully addressed, but it is possible that the interaction of MRI-2 with Ku proteins may improve their DNA-binding affinity or facilitate the recruitment of other repair complex components. Although in this review we focus in sORF-derived microproteins, it is worth to mention that many alt-ORFs and uORFs can produce functional microproteins with potential roles in genome instability. For example, the AltMRVI1 microprotein is coded by an alt-ORF inside the MRVI gene, and it directly interacts with BRCA1, one of the key effectors of homologous recombination DNA repair machinery [60,61]. The discovery of microproteins involved in genomic stability maintenance suggests that there could be numerous “genomic regulator-microproteins” which help preventing accumulation of genomic alterations in cancer cells.

1.7. Enabling characteristic: tumor-promoting inflammation

Another characteristic that allows the acquisition of the core cancer hallmarks is inflammation [10]. Several microproteins already described in this review might be of particular interest regarding tumor inflammation. In particular, the MDPs MOTS-c and HN have demonstrated their capacity to exacerbate the pro-inflammatory effect of senescence-associated secretory phenotype (SASP) of senescent cells by modulating their mitochondrial activity [62]. Cellular senescence is activated by multiple cellular stressors and it is characterized by a stable cell-cycle arrest and a pro-inflammatory secretome. Senescence acts as a main tumor suppressor barrier that impedes neoplastic transformation of damaged cells and promotes tissue repair [17]. However, the accumulation of senescent cells in tissues could have detrimental effects, mainly because of the inflammatory SASP and, in tumors, the presence of senescent cells promotes cancer cell growth and metastasis [63,64]. In this regard, it would be interesting to study whether the pro-inflammatory cytokines upregulated by MDPs facilitate the immunoclearance of senescent tumor cells or, on the contrary, favor a pro-tumorigenic microenvironment. In addition, ribosome profiling of bone marrow-derived macrophages revealed Aw112010, a non-ATG-initiated microprotein that promotes a pro-inflammatory response increasing canonical inflammatory cytokines like IL-6 and IL-12p40 upon bacterial infection [65]. These findings suggest a role of microproteins in cancer development through regulating inflammation.

2. Concluding remarks

sORF-encoded proteins have expanded our view about the coding potential of the genome, adding a new layer of complexity in the regulation of biological processes. The emergent picture suggests that microproteins allow the fine-tuning of many of these processes to adapt to specific needs and cellular contexts. Here, we have summarized what might be their implication in cancer. Even if only a small subset of microproteins have been functionally characterized so far, there is evidence of many of them as regulators of most of the hallmarks of cancer and its enabling characteristics (Fig. 1). While, so far, microproteins have not been directly related to angiogenesis, replicative immortality and immune evasion, we should take into account that the

proposed hallmarks are interconnected and some of the already identified microproteins, upon further analysis, could be classified in several hallmarks at the same time.

Finally, it is worth mentioning that most sequencing efforts in cancer restrict their analysis to the annotated protein-coding genome, unintentionally ignoring the microproteome. Therefore, it remains unclear if mutations in microproteins are selected during cancer evolution. If this were the case, mutated microproteins could also be a source of cancer neoantigens that can be used to improve the development of personalized immunotherapy [66]. We envision that this is an area that is going to be intensively studied and expanded in the coming years, and will bring crucial information for the clinic.

Here, we have discussed a set of microproteins encoded by lncRNAs, miRNAs, rRNAs, and cirRNAs but many more are yet to be explored, including the ones coded by the so-called alt-ORFs, that we have not addressed in this review. We are at the beginning of a new set of discoveries in which the identification and characterization of the cellular microproteins repertoire -the microproteome- will help us to better understand how physiological and pathological processes are regulated at its finest level. We anticipate that advances in this field will bring new therapeutic opportunities for oncology.

CRediT authorship contribution statement

Iñaki Merino-Valverde: Conceptualization, Writing - original draft, Writing - review & editing. **Emanuela Greco:** Conceptualization, Writing - original draft, Writing - review & editing. **María Abad:** Conceptualization, Writing - original draft, Writing - review & editing, Funding acquisition.

Acknowledgements

M.A. is a recipient of a Ramón y Cajal contract from the Spanish Ministry of Economy (MINECO). I.M. is a recipient of an FPI contract from the Spanish Ministry of Economy (MINECO). Work in the laboratory of M.A. is funded by Fero Foundation and by grants from the Spanish Ministry of Economy (RTI2018-102046-B-I00), from *"la Caixa"* Foundation (HR18-00256), and from Mutua Madrileña Foundation.

References

[1] C.A. Makarewich, E.N. Olson, Mining for micropeptides, Trends Cell Biol. 27 (2017) 685–696.
[2] M.W. Orr, Y. Mao, G. Storz, S.B. Qian, Alternative ORFs and small ORFs: shedding light on the dark proteome, Nucleic Acids Res. 48 (3) (2019) 1029–1042, https://doi.org/10.1093/nar/gkz734.
[3] M.E. Dinger, K.C. Pang, T.R. Mercer, J.S. Mattick, Differentiating protein-coding and noncoding RNA: challenges and ambiguities, PLoS Comput. Biol. 4 (2008) e1000176.
[4] J.W. Fickett, ORFs and genes: how strong a connection? J. Comput. Biol. 2 (1995) 117–123.
[5] N.G. D'Lima, J. Ma, L. Winkler, Q. Chu, K.H. Loh, E.O. Corpuz, B.A. Budnik, J. Lykke-Andersen, A. Saghatelian, S.A. Slavoff, A human microprotein that interacts with the mRNA decapping complex, Nat. Chem. Biol. 13 (2017) 174–180.
[6] J.Z. Huang, M. Chen, X.C. Gao, S. Zhu, H. Huang, M. Hu, H. Zhu, G.R. Yan, A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth, Mol Cell 68 (2017) 171–184 e176.
[7] B.R. Nelson, C.A. Makarewich, D.M. Anderson, B.R. Winders, C.D. Troupes, F. Wu, A.L. Reese, J.R. McAnally, X. Chen, E.T. Kavalali, et al., A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle, Science 351 (2016) 271–275.
[8] S.A. Slavoff, J. Heo, B.A. Budnik, L.A. Hanakahi, A. Saghatelian, A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining, J. Biol. Chem. 289 (2014) 10950–10957.
[9] F. Yeasmin, T. Yada, N. Akimitsu, Micropeptides encoded in transcripts previously identified as long noncoding RNAs: a new chapter in transcriptomics and proteomics, Front. Genet. 9 (2018) 144.
[10] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (2011) 646–674.
[11] M. Polycarpou-Schwarz, M. Gross, P. Mestdagh, J. Schott, S.E. Grund, C. Hildenbrand, J. Rom, S. Aulmann, H.P. Sinn, J. Vandesompele, et al., The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation, Oncogene 37 (2018) 4750–4768.
[12] Z. Sui, J. Zhou, Z. Cheng, P. Lu, Squalene epoxidase (SQLE) promotes the growth and migration of the hepatocellular carcinoma cells, Tumour Biol 36 (2015) 6173–6179.
[13] X. Zheng, L. Chen, Y. Zhou, Q. Wang, Z. Zheng, B. Xu, C. Wu, Q. Zhou, W. Hu, C. Wu, et al., A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling, Mol. Canc. 18 (2019) 47.
[14] B.J. Aubrey, G.L. Kelly, A. Janic, M.J. Herold, A. Strasser, How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? Cell Death Differ. 25 (2018) 104–113.
[15] B.G. Childs, D.J. Baker, J.L. Kirkland, J. Campisi, J.M. van Deursen, Senescence and apoptosis: dueling or complementary cell fates? EMBO Rep. 15 (2014) 1139–1153.
[16] A.R. Delbridge, L.J. Valente, A. Strasser, The role of the apoptotic machinery in tumor suppression, Cold Spring Harb Perspect Biol 4 (2012).
[17] V. Gorgoulis, P.D. Adams, A. Alimonti, D.C. Bennett, O. Bischof, C. Bishop, J. Campisi, M. Collado, K. Evangelou, G. Ferbeyre, et al., Cellular senescence: defining a path forward, Cell 179 (2019) 813–827.
[18] C.J. Sherr, F. McCormick, The RB and p53 pathways in cancer, Canc. Cell 2 (2002) 103–112.
[19] O. Marin-Bejar, F.P. Marchese, A. Athie, Y. Sanchez, J. Gonzalez, V. Segura, L. Huang, I. Moreno, A. Navarro, M. Monzo, et al., Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2, Genome Biol. 14 (2013) R104.
[20] F.X. Chen, A.R. Woodfin, A. Gardini, R.A. Rickels, S.A. Marshall, E.R. Smith, R. Shiekhattar, A. Shilatifard, PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II, Cell 162 (2015) 1003–1015.
[21] M. Zhang, K. Zhao, X. Xu, Y. Yang, S. Yan, P. Wei, H. Liu, J. Xu, F. Xiao, H. Zhou, et al., A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma, Nat. Commun. 9 (2018) 4475.
[22] M. Zhang, N. Huang, X. Yang, J. Luo, S. Yan, F. Xiao, W. Chen, X. Gao, K. Zhao, H. Zhou, et al., A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis, Oncogene 37 (2018) 1805–1814.
[23] S. Wu, L. Zhang, J. Deng, B. Guo, F. Li, Y. Wang, R. Wu, S. Zhang, J. Lu, Y. Zhou, A novel micropeptide encoded by Y-linked LINC00278 links cigarette smoking and AR signaling in male esophageal squamous cell carcinoma, Cancer Res Epub ahead of print (2020), https://doi.org/10.1158/0008-5472.CAN-19-3440.
[24] M.W. Popp, L.E. Maquat, Nonsense-mediated mRNA decay and cancer, Curr. Opin. Genet. Dev. 48 (2018) 44–50.
[25] S.W. Lowe, E. Cepero, G. Evan, Intrinsic tumour suppression, Nature 432 (2004) 307–315.
[26] J.M. Adams, S. Cory, The Bcl-2 apoptotic switch in cancer development and therapy, Oncogene 26 (2007) 1324–1337.
[27] C. Lee, K. Yen, P. Cohen, Humanin: a harbinger of mitochondrial-derived peptides? Trends Endocrinol Metab 24 (2013) 222–228.
[28] Y. Hashimoto, T. Niikura, Y. Ito, H. Sudo, M. Hata, E. Arakawa, Y. Abe, Y. Kita, I. Nishimoto, Detailed characterization of neuroprotection by a rescue factor humanin against various Alzheimer's disease-relevant insults, J. Neurosci. 21 (2001) 9235–9245.
[29] B. Guo, D. Zhai, E. Cabezas, K. Welsh, S. Nouraini, A.C. Satterthwait, J.C. Reed, Humanin peptide suppresses apoptosis by interfering with Bax activation, Nature 423 (2003) 456–461.
[30] F. Luciano, D. Zhai, X. Zhu, B. Bailly-Maitre, J.E. Ricci, A.C. Satterthwait, J.C. Reed, Cytoprotective peptide humanin binds and inhibits proapoptotic Bcl-2/Bax family protein BimEL, J. Biol. Chem. 280 (2005) 15825–15835.
[31] C.F. Zuccato, A.S. Asad, A.J. Nicola Candia, M.F. Gottardo, M.A. Moreno Ayala, M.S. Theas, A. Seilicovich, M. Candolfi, Mitochondrial-derived peptide humanin as therapeutic target in cancer and degenerative diseases, Expert Opin. Ther. Targets 23 (2019) 117–126.
[32] R.A. Saxton, D.M. Sabatini, mTOR signaling in growth, metabolism, and disease, Cell 169 (2017) 361–371.
[33] S.A. Forbes, N. Bindal, S. Bamford, C. Cole, C.Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, et al., COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer, Nucleic Acids Res. 39 (2011) D945–D950.
[34] A. Matsumoto, J.G. Clohessy, P.P. Pandolfi, SPAR, a lncRNA encoded mTORC1 inhibitor, Cell Cycle 16 (2017) 815–816.
[35] A. Ortega-Molina, N. Deleyto-Seldas, J. Carreras, A. Sanz, C. Lebrero-Fernandez, C. Menendez, A. Vandenberg, B. Fernandez-Ruiz, L. Marin-Arraiza, C. de la Calle Arregui, et al., Oncogenic Rag GTPase signaling enhances B cell activation and drives follicular lymphoma sensitive to pharmacological inhibition of mTOR, Nat Metab 1 (2019) 775–789.
[36] J.R. Cubillos-Ruiz, S.E. Bettigole, L.H. Glimcher, Tumorigenic and immunosuppressive effects of endoplasmic reticulum stress in cancer, Cell 168 (2017) 692–706.
[37] M. Wang, R.J. Kaufman, The impact of the endoplasmic reticulum protein-folding environment on cancer development, Nat. Rev. Canc. 14 (2014) 581–597.
[38] R. Bagur, G. Hajnoczky, Intracellular Ca(2+) sensing: its role in calcium homeostasis and signaling, Mol Cell 66 (2017) 780–788.
[39] D.M. Anderson, K.M. Anderson, C.L. Chang, C.A. Makarewich, B.R. Nelson, J.R. McAnally, P. Kasaragod, J.M. Shelton, J. Liou, R. Bassel-Duby, et al., A micropeptide encoded by a putative long noncoding RNA regulates muscle performance, Cell 160 (2015) 595–606.
[40] C. Cui, R. Merritt, L. Fu, Z. Pan, Targeting calcium signaling in cancer therapy, Acta Pharm. Sin. B 7 (2017) 3–17.
[41] D.M. Anderson, C.A. Makarewich, K.M. Anderson, J.M. Shelton, S. Bezprozvannaya,

R. Bassel-Duby, E.N. Olson, Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides, Sci. Signal. 9 (2016) ra119.
[42] Q. Chu, T.F. Martinez, S.W. Novak, C.J. Donaldson, D. Tan, J.M. Vaughan, T. Chang, J.K. Diedrich, L. Andrade, A. Kim, et al., Regulation of the ER stress response by a mitochondrial microprotein, Nat. Commun. 10 (2019) 4883.
[43] K. Polyak, R.A. Weinberg, Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits, Nat. Rev. Canc. 9 (2009) 265–273.
[44] H.C. Lo, X.H. Zhang, EMT in metastasis: finding the right balance, Dev. Cell 45 (2018) 663–665.
[45] L. Vrba, J.C. Garbe, M.R. Stampfer, B.W. Futscher, Epigenetic regulation of normal human mammary cell type-specific miRNAs, Genome Res. 21 (2011) 2026–2037.
[46] S.M. Jinbo Fang, Veena N. Rao, E. Shyam P. Reddy, Decoding of non-coding DNA and non-coding RNA: pri-micro RNA-encoded novel peptides regulate migration of cancer cells, Journal of Pharmaceutical Sciences and Pharmacology 3 (2017) 23–27.
[47] Z.W. Guo, Y. Meng, X.M. Zhai, C. Xie, N. Zhao, M. Li, C.L. Zhou, K. Li, T.C. Liu, X.X. Yang, et al., Translated long non-coding ribonucleic acid ZFAS1 promotes cancer cell migration by elevating reactive oxygen species production in hepatocellular carcinoma, Front. Genet. 10 (2019) 1111.
[48] M. Cordenonsi, F. Zanconato, L. Azzolin, M. Forcato, A. Rosato, C. Frasson, M. Inui, M. Montagner, A.R. Parenti, A. Poletti, et al., The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells, Cell 147 (2011) 759–772.
[49] W. Lehmann, D. Mossmann, J. Kleemann, K. Mock, C. Meisinger, T. Brummer, R. Herr, S. Brabletz, M.P. Stemmler, T. Brabletz, ZEB1 turns into a transcriptional activator by interacting with YAP1 in aggressive cancer types, Nat. Commun. 7 (2016) 10498.
[50] C.E. Gast, A.D. Silk, L. Zarour, L. Riegler, J.G. Burkhart, K.T. Gustafson, M.S. Parappilly, M. Roh-Johnson, J.R. Goodman, B. Olson, et al., Cell fusion potentiates tumor heterogeneity and reveals circulating hybrid cells that correlate with stage and survival, Sci Adv 4 (2018) eaat7828.
[51] P. Bi, A. Ramirez-Martinez, H. Li, J. Cannavino, J.R. McAnally, J.M. Shelton, E. Sanchez-Ortiz, R. Bassel-Duby, E.N. Olson, Control of muscle formation by the fusogenic micropeptide myomixer, Science 356 (2017) 323–327.
[52] N.N. Pavlova, C.B. Thompson, The emerging hallmarks of cancer metabolism, Cell Metabol. 23 (2016) 27–47.
[53] U.E. Martinez-Outschoorn, M. Peiris-Pages, R.G. Pestell, F. Sotgia, M.P. Lisanti, Cancer metabolism: a therapeutic perspective, Nat. Rev. Clin. Oncol. 14 (2017) 11–31.
[54] Y.F. Lin, M.H. Xiao, H.X. Chen, Y. Meng, N. Zhao, L. Yang, H. Tang, J.L. Wang, X. Liu, Y. Zhu, et al., A novel mitochondrial micropeptide MPM enhances mitochondrial respiratory activity and promotes myogenic differentiation, Cell Death Dis. 10 (2019) 528.
[55] C.S. Stein, P. Jadiya, X. Zhang, J.M. McLendon, G.M. Abouassaly, N.H. Witmer, E.J. Anderson, J.W. Elrod, R.L. Boudreau, Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency, Cell Rep. 23 (2018) 3710–3720 e3718.
[56] C.A. Makarewich, K.K. Baskin, A.Z. Munir, S. Bezprozvannaya, G. Sharma, C. Khemtong, A.M. Shah, J.R. McAnally, C.R. Malloy, L.I. Szweda, et al., MOXI is a mitochondrial micropeptide that enhances fatty acid beta-oxidation, Cell Rep. 23 (2018) 3701–3709.
[57] C. Lee, J. Zeng, B.G. Drew, T. Sallam, A. Martin-Montalvo, J. Wan, S.J. Kim, H. Mehta, A.L. Hevener, R. de Cabo, et al., The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance, Cell Metabol. 21 (2015) 443–454.
[58] S.P. Jackson, J. Bartek, The DNA-damage response in human biology and disease, Nature 461 (2009) 1071–1078.
[59] N.R. Pannunzio, G. Watanabe, M.R. Lieber, Nonhomologous DNA end-joining for repair of DNA double-strand breaks, J. Biol. Chem. 293 (2018) 10512–10523.
[60] B. Vanderperre, J.F. Lucier, C. Bissonnette, J. Motard, G. Tremblay, S. Vanderperre, M. Wisztorski, M. Salzet, F.M. Boisvert, X. Roucou, Direct detection of alternative open reading frames translation products in human significantly expands the proteome, PloS One 8 (2013) e70698.
[61] W. Zhao, J.B. Steinfeld, F. Liang, X. Chen, D.G. Maranon, C. Jian Ma, Y. Kwon, T. Rao, W. Wang, C. Sheng, et al., BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing, Nature 550 (2017) 360–365.
[62] S.J. Kim, H.H. Mehta, J. Wan, C. Kuehnemann, J. Chen, J.F. Hu, A.R. Hoffman, P. Cohen, Mitochondrial peptides modulate mitochondrial function during cellular senescence, Aging (Albany NY) 10 (2018) 1239–1256.
[63] A. Calcinotto, J. Kohli, E. Zagato, L. Pellegrini, M. Demaria, A. Alimonti, Cellular senescence: aging, cancer, and injury, Physiol. Rev. 99 (2019) 1047–1078.
[64] D.V. Faget, Q. Ren, S.A. Stewart, Unmasking senescence: context-dependent effects of SASP in cancer, Nat. Rev. Canc. 19 (2019) 439–453.
[65] R. Jackson, L. Kroehling, A. Khitun, W. Bailis, A. Jarret, A.G. York, O.M. Khan, J.R. Brewer, M.H. Skadow, C. Duizer, et al., The translation of non-canonical open reading frames controls mucosal immunity, Nature 564 (2018) 434–438.
[66] A. Garcia-Garijo, C.A. Fajardo, A. Gros, Determinants for neoantigen identification, Front. Immunol. 10 (2019) 1392.