

TESIS DOCTORAL

CARACTERIZACIÓN GENÉTICA Y FUNCIONAL  
DE LA HETEROGENEIDAD INTRATUMORAL EN  
UN MODELO MURINO DE CÁNCER DE PÁNCREAS

PHD THESIS

GENETIC AND FUNCTIONAL  
INTRATUMORAL HETEROGENEITY  
IN A PANCREATIC MOUSE CANCER MODEL

AUTORA

**Laura Quevedo Palacio**

DIRECTOR

**Dr. Ignacio Varela Egocheaga**



UNIVERSIDAD DE CANTABRIA

Escuela de Doctorado de la Universidad de Cantabria

Santander 2022







El Dr. IGNACIO VARELA EGOICHEAGA, Profesor Titular de la Universidad de Cantabria (UC) e Investigador Principal del grupo Genómica Funcional de la Progresión Tumoral en el Departamento de Señalización Molecular y Celular en el Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC)

CERTIFICA: que LAURA QUEVEDO PALACIO ha realizado bajo su dirección la presente Tesis Doctoral titulada “Caracterización genética y funcional de la heterogeneidad intratumoral en un modelo murino de cáncer de páncreas/Genetic and functional intratumoral heterogeneity in a pancreatic mouse cancer model”.

Considero que este trabajo reúne los requisitos de originalidad y calidad científica necesarios para su presentación como Memoria de Doctorado al objeto de optar al grado de Doctor en Biología Molecular y Biomedicina por la Universidad de Cantabria, con mención internacional.

Y para que conste y surta los efectos oportunos, firmo el presente certificado en Santander a 21 de junio de 2022.

Fdo: Ignacio Varela Egocheaga





La presente Tesis Doctoral titulada “Caracterización genética y funcional de la heterogeneidad intratumoral en un modelo murino de cáncer de páncreas”/Genetic and functional intratumoral heterogeneity in a pancreatic mouse cancer model”, ha sido realizada en el Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC) en el laboratorio del grupo Genómica Funcional de la Progresión Tumoral, en Santander (Cantabria).

Esta tesis ha sido financiada a través de fondos procedentes de:

- Fundación Ramón Areces, Proyecto: “Caracterización molecular del papel de la disfunción mitocondrial en el desarrollo tumoral”. Proyectos para la Ciencia de la Vida y a Materia (2014).
- European Research Council, Proyecto: “Molecular characterization of the role of intratumor heterogeneity in cancer progression and metastasis”. ERC-2014-StG-2014-637904.

Para la realización de este trabajo, Laura Quevedo también ha recibido:

- Beca predoctoral de la Universidad de Cantabria (convocatoria: 2018-2022).
- Ayuda de movilidad predoctoral de la Universidad de Cantabria para realizar una estancia breve en la Technical University of Munich (convocatoria: 2021).



European Research Council





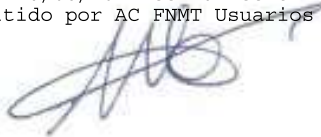
Santander, 16 de Junio de 2022

Durante su trabajo de tesis doctoral, Laura Quevedo ha generado una gran cantidad de información genética y transcriptómica de célula única en un modelo de ratón de cáncer de páncreas basado en la expresión de una versión activada de Kras. Gracias a estudios de última tecnología de secuenciación masiva en múltiples muestras de tumores primarios y metástasis asociadas ha demostrado el origen monoclonal de los tumores así como la presencia de distintos clones genéticos dentro de los tumores de páncreas. Estos clones presentan fenómenos evolutivos de competencia y selección similares a los encontrados en humanos con la característica de que algunos de estos clones presentan capacidad metastática específica de tejido. Ha descrito igualmente que los tumores frecuentemente muestran alteraciones en el cromosoma 6 con la intención de activar la señalización mutante de Kras y que esta señalización es esencial para la generación de metástasis en hígado pero no en pulmón que parece ocurrir de manera muy temprana durante el desarrollo tumoral.

Por último, durante sus estudios transcripcionales de célula única, Laura ha descrito la existencia de diversos grupos transcripcionales en los tumores recurrentes entre muestras con características moleculares que sugieren una colaboración entre distintos grupos celulares. Curiosamente, las células metastáticas de pulmón presentan un perfil transcripcional característico con una transición epitelio-mesénquima parcial, la sobreactivación de la ruta de EGFR y la producción de citoquinas que probablemente están implicadas en la preparación del nicho metastático. Este perfil transcripcional está presente en una población minoritaria de los tumores primarios en lo que podría suponer una población pre-metastática.

Por todo lo expuesto anteriormente, considero que el trabajo de Laura Quevedo es de gran calidad y novedad y que sus resultados supondrán con total seguridad una gran aportación al campo científico del cáncer de páncreas, por lo que apruebo sin reservas el depósito de su tesis doctoral titulada “Caracterización genética y funcional de la heterogeneidad intratumoral en un modelo murino de cáncer de páncreas”.

Firmado por VARELA EGOCHEAGA  
IGNACIO ALEJANDRO - 71649764A el  
día 16/06/2022 con un certificado  
emitido por AC FNMT Usuarios



**Ignacio Varela Egocheaga**  
Profesor Titular  
Departamento de Biología Molecular.







A mis padres



## Resumen

El adenocarcinoma ductal de páncreas (PDAC) es el tipo de cáncer con peor pronóstico debido a su diagnóstico tardío, la diseminación metastásica temprana y a la falta de tratamientos específicos. Esta enfermedad se caracteriza por una extensa heterogeneidad genética y un microambiente tumoral altamente inflamatorio y desmoplástico, lo que contribuye a la agresividad y a la resistencia a tratamientos de esta patología. Con el objetivo de comprender mejor los mecanismos implicados en la progresión y metástasis del cáncer de páncreas, en esta Tesis Doctoral, propusimos caracterizar la dinámica de la heterogeneidad intratumoral en PDAC. Para ello, combinamos la flexibilidad de un modelo de ratón de cáncer de páncreas con la alta sensibilidad de las tecnologías de secuenciación de nueva generación y de célula única, con una estrategia de trazado genético celular, la cual hemos mejorado. A pesar de un origen monoclonal, hemos demostrado la presencia de poblaciones celulares genéticamente distintas dentro de los tumores. Hemos identificado además fenómenos evolutivos de competencia y selección, similares a los descritos en humanos, con la particularidad de que algunos de estos clones presentan una capacidad metastásica específica de tejido. Además, hemos descrito que estos tumores suelen presentar alteraciones en *Kras* y *Cdkn2a*, pero no en *Trp53* o *Smad4*, al contrario que en humanos. También observamos de manera recurrente la activación de la ruta mutada de *Kras*, esencial para la generación de metástasis en el hígado, pero no en la de pulmón, que parece ocurrir de manera temprana. Por otra parte, hemos generado una nueva herramienta de trazado genético que puede ser útil para estudiar la heterogeneidad intratumoral. Por último, utilizando las tecnologías de secuenciación de célula única hemos identificado la presencia de distintos grupos transcripcionales en los tumores y metástasis de nuestro modelo, siendo algunos de ellos recurrentes entre muestras y con características que sugieren una colaboración entre distintos grupos de células. Curiosamente, las células metastásicas de pulmón presentan un perfil transcripcional característico, con una transición epitelio mesénquima parcial, una sobre activación de la ruta de *EGFR* y la producción de citoquinas que probablemente estén implicadas en la preparación del nicho metastático. Este perfil transcripcional está presente en una población minoritaria de los tumores primarios, lo que podría

significar que se trata de una población premetastática. Finalmente, la mejora de conocimiento sobre los genes y las vías moleculares implicadas en la progresión tumoral y la aparición de metástasis puede traducirse en una mejora significativa del tratamiento de los pacientes de PDAC.

## Summary

Pancreatic ductal adenocarcinoma (PDAC) is a highly lethal disease due to its late diagnosis, early metastatic widespread, and the lack of effective treatments. PDAC is characterized by extensive genetic intratumor heterogeneity and by a highly inflammatory and desmoplastic tumor microenvironment, which contributes to the aggressiveness and drug resistance of the disease. To understand the mechanisms involved in PDAC progression and metastasis, in this Doctoral Thesis, we proposed to characterize the dynamics of intratumor heterogeneity in PDAC. To do this, we combined the flexibility of a well-characterized pancreatic mouse model, the high sensitivity of next-generation and single-cell sequencing technologies, and a multi-fluorescent protein lineage tracking model. First, we have demonstrated the monofocal origin of the aggressive pancreatic tumors from our mouse model and the presence of genetically distinct cell populations within tumors. We have also identified evolutionary phenomena of competition and selection, similar to those described in humans, with the particularity that some of these clones have tissue-specific metastatic capacity. In addition, we have described that these tumors recurrently present amplification of the mutant *Kras* signal which seems important for the generation of metastases in the liver, but is disposable for the metastasis in the lung, which seems to be the result of early colonization. Finally, we have identified the presence of different transcriptional groups in the tumors and metastases of our model, some of them being recurrent between samples and with characteristics that demonstrate a collaboration between different groups of cells. Interestingly, metastatic lung cells present a particular transcriptional profile, with a partial epithelial-to-mesenchymal transition, overactivation of the EGFR pathway, and the production of cytokines that are probably involved in the preparation of the metastatic niche. This transcriptional profile is present in a minority population of primary tumors, which could mean that it is a premetastatic population. Finally, the improvement of the knowledge about the genes and molecular pathways involved in tumor progression and the appearance of metastasis could be finally translated to a significant improvement in the treatment of PDAC patients.



*The important thing is not to stop questioning.  
Curiosity has its own reason for existence.*

Albert Einstein





Me gustaría expresar mi más sincero agradecimiento hacia todas aquellas personas que han hecho de estos años una etapa maravillosa y que, sin duda, volvería a repetir.

En primer lugar, a Nacho, por confiar en mí y darme la oportunidad de formar parte de tu equipo y poder trabajar en este proyecto tan emocionante. Por darme siempre todas las facilidades y oportunidades. Por enseñarme y orientarme en este viaje que empezó hace ya más de 6 años.

To Prof. Roland Rad for allowing me to work in your excellent research group. I am very grateful to you, and also to Prof. Dieter Saur, for your valuable scientific discussions. To Sebastian, Niklas, Katha, Anja, Najib, Milad, Montero, Stefanie, Wolle, Rupert, Thomas, Olga, Steven, Özgün, and all of you for the nice time expended together during my short internship in Munich.

A todos los que habéis compartido laboratorio conmigo, ayer y hoy. A Thaidy, Lau, Antonio, Bea, Rosa, David, por vuestra ayuda constante y las horas hablando de ciencia y de la vida. Especial mención a Carlos. Me acuerdo de ti cada día.

A Javier y Chepe, por aportar año tras año su granito de arena en las evaluaciones anuales. A Dolo, Berta, AnaVi, Fernando, Álvaro, Magda, Matxalen y al resto de investigadores del IBBTEC que siempre han estado ahí para aconsejarme, resolver cualquier duda o establecer una colaboración. Por ser una fuente de motivación.

A Mapi, Ana, María, Sheila, Ruli, por apoyarnos en los momentos más difíciles. Por todas las comidas compartidas pre-pandemia.

Al personal técnico del instituto. A Pepi, por lidiar con nuestros pedidos urgentes y por sus golpes de humor y de realidad. A Sandra y Mati por pasar por el labo a primera hora de la mañana para abastecernos de material siempre con una sonrisa. A Víctor por las largas horas evaluando las proteínas fluorescentes. A MariCruz por estar disponible para cuantificar librerías en cualquier momento.

Al personal del animalario, Marta, Patri, Mar, Bea y Miguel, por su ayuda diaria con los ratoncillos, tan necesarios en investigación y sin los que esta tesis no sería posible.

A Charo y a Carmen, por tener siempre su puerta abierta y ayudarnos a todos.

A todos mis compañeros del IBBTEC con los que he tenido el placer de trabajar y de pasar buenos ratos todos estos años. A Antonio, Marta, Emilio, Omar, Jorge, Rocío, Cris, Carol, Ani, Judit, Esther, gracias por los primeros años. A los Piero, los Roses, los Rada, las Calvas y a Javi como una más. En especial a Lore, Ana, Laura, María, Víctor, Lara, Sofía, Endika, Diane, Patri, Silvia, Iván, Miguel, Aurora, Lolu... Gracias a todos. También, a los doctorandos de esta promoción: a David por haber sido el mejor compi de escritura; a Rocío y Marta, que habéis repetido la jugada de Ana y Vini; a AnaB, Fer, Sara y Julia, ya nos queda poco.

A los que se convirtieron en compañeros de vida. Mario, Rocío, Jorge, Yeli, Cande, Vanessa, Lorena, Fer, Tamara, Evi, Adri, David, Helena, Rosa, Meri, Cata, Bea... Por todos los buenos momentos vividos dentro y fuera del labo. Por los cafés compartidos con galletas de chocolate, barbacoas, risas, llantos, confidencias y por lo que está por llegar.

A Lola, por impulsarme desde el colegio. Por seguir siendo alguien fundamental.

A Carlos, Josemari, Gloria y Xose, por ver ese “algo” en mí mientras estudiaba Biología. Por recomendarme a Nacho sin dudarlo. Por demostrar que, con buena ciencia y buenas intenciones, todo acaba en su lugar.

A los goñinines y los elfillos, por el apoyo en la distancia y las reuniones anuales.

A mis amigos de siempre. A Nadia, Ire, Inés, Irene, Ani, Saúl, Javi, Rubén y Dani. MDHF, por ser y estar.

A Víctor, por compartirnos tiempo. Por acompañarme en el día a día. Por seguir construyendo juntos, *Say Yes Munich*.

Por último, a mi familia. A mis hermanos, David y Raquel, y a mis sobrinos, Mario y Sergio, por estar siempre conmigo a pesar de la distancia física. A mis padres, que habéis formado una familia extraordinaria. Por ser mi faro y guiarme en el camino de la vida. Por vuestro apoyo constante y hacerme ser mejor persona. Porque sois lo que más quiero.

Lauri.





## List of abbreviations

ADEX	aberrantly differentiated endocrine exocrine
CA 19-9	carbohydrate antigen 19-9
CAF	cancer-associated fibroblast
ChIP-seq	chromatin immunoprecipitation followed by sequencing
CNA	copy number alteration
CRISPR	clustered regularly interspaced short palindromic repeats
CTC	circulating tumor cells
ddNTP	dideoxynucleotides
DGE	data expression matrix
DNA	deoxyribonucleic acid
dNTP	deoxynucleotides
EGFR	epidermal growth factor receptor
EMT	epithelial-to-mesenchymal transition
Flp	flippase
FOLFORINOX	folinic acid, 5-fluoruracil, irinotecan and oxaliplatin
FP	fluorescent protein
FSF	FRT-Stop-FRT
GSEA	gene set enrichment analysis
HEK	human embryonic kidney
IGV	Integrative Genomics Viewer
Indel	insertion and deletion
IPMN	intraductal papillary mucinous neoplasms
ITH	intratumor heterogeneity
LSL	loxP-stop-loxP
MCN	mucinous cystic neoplasms
MeDIP-seq	methylated DNA immunoprecipitation followed by sequencing

MIP	molecular inversion probe
NET	neuroendocrine tumor
NGS	next-generation sequencing
NGS	next-generation sequencing
PanIN	pancreatic intraepithelial neoplasms
PARP	poly(adenosine diphosphate-ribose) polymerase
PCA	principal components analysis
PDAC	pancreatic ductal adenocarcinoma
QM	quasi-mesenchymal
qRT-PCR	quantitative reverse transcriptase PCR
scATAC-seq	single-cell sequencing assay for transposase accessible chromatin
scDNA-seq	single-cell DNA-sequencing
scRNA-seq	single-cell RNA-sequencing
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
TCR	T-cell receptor
t-SNE	t-distributed stochastic neighbor embedding
TME	tumor microenvironment
UMAP	uniform manifold approximation and projection
UMI	unique molecular identifiers
WES	whole-exome sequencing
WGS	whole-genome sequencing







# Contents

<b>Introduction</b> .....	<b>1</b>
Pancreatic ductal adenocarcinoma .....	3
Mouse models of PDAC .....	13
Intratumor heterogeneity .....	18
Next-generation sequencing .....	25
Single-cell sequencing technologies.....	36
<b>Objectives</b> .....	<b>43</b>
<b>Experimental procedures</b> .....	<b>47</b>
<b>Results</b> .....	<b>67</b>
I. To analyze the tumor clonality in early and late stages using fluorescent lineage tracing tools .....	69
II. To generate new lineage tracing tools to study the involvement of intratumor heterogeneity in cancer progression .....	71
III. Study the role of genetic intratumor heterogeneity in pancreatic cancer progression.....	76
IV. To characterize the presence of transcriptomic intratumor heterogeneity and its role in metastasis.....	86
<b>Discussion</b> .....	<b>99</b>
<b>Conclusions</b> .....	<b>113</b>
<b>References</b> .....	<b>119</b>
<b>Publications</b> .....	<b>149</b>



# Introduction



# Pancreatic ductal adenocarcinoma

## Epidemiology

Pancreatic cancer is the fourth leading cause of cancer-related death in the Western World. For all stages combined, mortality rates have remained stubbornly unchanged for almost four decades, with a 5-year relative survival rate of 11%, the lowest rate among any cancer. Even for the small percentage of people who are diagnosed with local disease, their 5-year survival is only 39%. The majority of patients are diagnosed at an advanced stage, for which the 5-year survival is 3%. Moreover, the incidence rate of pancreatic cancer continues to rise by about 1% per year since 2000 (Siegel *et al.*, 2022). Pancreatic cancer has a poor prognosis due to its early local invasion and distant metastatic dissemination, its lack of early specific symptoms, and its late diagnosis. It typically metastasizes to the liver, lungs, and lymph nodes. Symptoms of pancreatic cancer may include weight loss, abdominal and back pain, and the development of type 2 diabetes. Additionally, some pancreatic tumors may also cause jaundice (yellowing of the skin and/or eyes) leading to earlier diagnosis. Other defining features of pancreatic cancer include progression from distinct types of precursor lesions, an extensive stromal reaction (desmoplasia) resulting in a hypovascular and hypoxic microenvironment, reprogramming of cellular metabolism, and evasion of tumor immunity (Ryan *et al.*, 2014).

## Treatment opportunities for PDAC

At the time of diagnosis, pancreatic cancer patients usually have locally advanced disease and even metastases, precluding surgery, which is currently the only possible curative method (Rosty & Goggins, 2002). Other treatment options are adjuvant chemotherapy and radiation to reduce the risk of recurrence. Targeted therapies, such as the epidermal growth factor receptor (EGFR) inhibitor erlotinib

(Moore *et al.*, 2007) and the poly(adenosine diphosphate-ribose) polymerase (PARP) inhibitor olaparib (Golan *et al.*, 2019), can also be given along with chemotherapy for advanced disease to extend survival or relieve symptoms, but they rarely produce a cure. In current treatment strategies, the combination of platinum agents plus other DNA damage-producing agents or mitotic spindle inhibitors are used. Systematic chemotherapy with gemcitabine or FOLFORINOX (combination therapy of folinic acid, 5-fluoruracil, irinotecan, and oxaliplatin), results in median survival of 5.6 and 11.1 months, respectively (Burriss *et al.*, 1997; Conroy *et al.*, 2011; Cid-Arregui & Juarez, 2015). Due to the combination of four cytotoxic drugs, FOLFORINOX is a highly aggressive treatment option that is in most cases only tolerated in young pancreatic cancer patients. A more recent treatment modality is the combinatorial treatment using nab-paclitaxel (nanoparticle albumin-bound paclitaxel) plus gemcitabine which was shown to prolong the median overall survival to 8.5 months compared to the treatment with gemcitabine alone (Von Hoff *et al.*, 2013). Disappointingly, none of these systematic treatments are close to producing a complete remission and can only minimally extend the life span of pancreatic cancer compared to treatments already available 30 years ago.

Early detection is therefore extremely important to improve patient survival. Imaging methods are not accurate enough to detect early lesions, assess tumor stage or resectability, or even discriminate between benign and malignant lesions (Singhi *et al.*, 2019). The carbohydrate antigen 19-9 (CA 19-9) is currently considered the best-validated serum marker in patients with pancreatic cancer, but it has limited use in the early diagnosis and progression monitoring of the disease (Poruk *et al.*, 2013). More clinically meaningful and efficient diagnostic strategies such as molecular markers capable of identifying pancreatic cancer during the curative stage of the disease are urgently needed. It is expected that the understanding of genetic alterations and molecular pathways involved in tumor progression, combined with the development of high-throughput sensitive techniques, will lead to the discovery

of a panel of biomarkers that will enable an efficient therapy at curable stages of tumor development.

### **Risk factors of PDAC**

Although it is estimated that 5 to 10% of pancreatic cancers have an inherited component, the genetic basis for familial aggregation is still unclear (Klein et al., 2004; Solomon et al., 2012). Some rare genetic syndromes are linked to an increased risk of developing pancreatic cancer. Hereditary breast and ovarian cancer syndrome with mutations in *BRCA1* and *BRCA2* genes are associated with a 3.5 to 10-fold increased risk of developing pancreatic cancer (Breast Cancer Linkage Consortium, 1999), whereas hereditary breast cancer caused by mutations in the *PALB2* gene duplicates the risk (X. Yang et al., 2020). Hereditary pancreatitis with mutations in *PRSS1* gene and Peutz-Jeghers syndrome with mutations in *STK11/LKB1* also increase the risk to 50-fold and 100-fold respectively (Lowenfels et al., 1997; Giardiello et al., 2000). Similarly, familial atypical multiple mole and melanoma syndrome is linked to p16/*CDKN2A* mutations and a 17-fold increased risk (Vasen et al., 2000). Lastly, Lynch syndrome, also known as hereditary non-polyposis colorectal cancer, with mutations in *MLH1*, *MSH2*, and *MSH6*, genes increases the risk up to 8.6-fold (Kastrinos et al., 2009). In all the cases, the risk of pancreatic cancer may be higher if there is also a history of pancreatic cancer in the family.

In the same way, there are also some other non-genetic factors that increase the risk of developing PDAC such as non-hereditary chronic pancreatitis (Duell et al., 2012), smoking (Bosetti et al., 2012), long-lasting diabetes mellitus (Ben et al., 2011), and obesity (Aune et al., 2012).

## **PDAC progression and development**

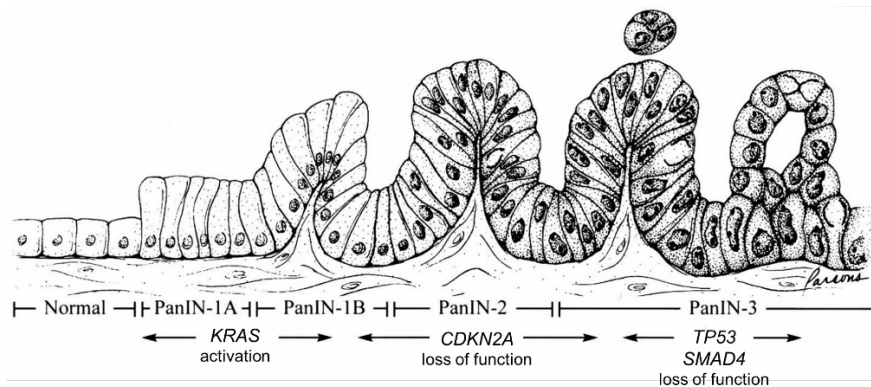
The pancreas is a glandular organ that belongs to the digestive and endocrine systems of vertebrates. In the adult pancreas, the endocrine portion comprises only 1-2% of cells that are contained in the islets of Langerhans. It is composed of alpha cells secreting glucagon to increase blood glucose levels, beta cells secreting insulin to reduce blood glucose levels, delta cells that produce somatostatin to inhibit insulin and glucagon release, and gamma cells, also known as PP, producing pancreatic polypeptide with a role in appetite. The hormones produced by islets are also important for regulating the exocrine tissue. The exocrine portion comprises around 98% of the pancreatic mass and is mainly composed of acinar and ductal cells. Acinar cells secrete digestive enzymes such as proteases (trypsinogen, among others), amylase, lipase, and trypsin inhibitors; while duct cells produce water, bicarbonate, chloride, intrinsic factor, and antibacterial proteins.

The most frequent histological subtype of pancreatic cancer is pancreatic ductal adenocarcinoma (PDAC), accounting for 95% of pancreatic malignancies. It develops in the exocrine tissue of the pancreas. The remaining 5% of pancreatic tumors originate from hormone-producing endocrine cells, also known as islet cells or neuroendocrine tumors (NETs). NETs often have a better prognosis and younger median age of diagnosis.

Histopathological and genetic analyses have defined three distinct pre-malignant lesions of the pancreas: pancreatic intraepithelial neoplasms (PanIN), intraductal papillary mucinous neoplasms (IPMN), and mucinous cystic neoplasms (MCN) (Maitra et al., 2005). PanINs are the best-described precursor lesions of human PDAC. Nevertheless, tumors require multiple genetic abnormalities to progress to metastatic disease (Iacobuzio-Donahue, 2012). PanINs are thought to progress through a stepwise accumulation of specific somatic mutations and cellular atypia. According to the extent of cytological dysplasia, PanINs are sub-stratified



from a low grade to a high grade in PanIN-1A, PanIN-1B, PanIN-2, and PanIN-3 or carcinoma-in-situ (Hezel et al., 2006). The PanIN progression stages are typically paralleled by somatic mutations resulting in *KRAS* activation, as well as inactivation of *CDKN2A*, *TP53*, and genes of the canonical transforming growth factor- $\beta$  (TGF $\beta$ ) signaling pathway such as *SMAD4* (Figure 1) (Hruban et al., 2000). These alterations are present in more than 50% of human PDAC cases and are thus hallmark mutations of the disease, accompanied by other genes mutated at low prevalence that reveals the complex mutational landscape of PDAC (Jones et al., 2008; Waddell et al., 2015).



**Figure 1 | Classical progression model of pancreatic cancer.** Histological and genetic progression from normal duct epithelium to invasive pancreatic cancer (from left to right). Modified from (Hruban et al., 2000).

The occurrence of *KRAS* activating mutations is the first and almost universal event (95% of PDAC) during early PanIN progression. These critical mutations are not randomly distributed along the protein-coding sequence of *KRAS*. There are two main hotspot mutations resulting in changes of amino acids G12 (~90% of cases) or Q61 (~5% of cases) (data from QCMG study, cBioPortal). *KRAS* activating mutations are typically followed by inactivation of the *CDKN2A* locus during PanIN progression. Loss of function of *CDKN2A*

already occurs in low-grade PanIN lesions and is observed in more than 90% of PDAC cases (Hruban et al., 2000; Schutte et al., 1997). The third hallmark of PanIN progression is the inactivation of TP53 protein which typically occurs in more advanced PanIN stages (Hruban et al., 2000). *TP53* governs multiple processes involved in tumor suppression and is mutated in up to 85% of PDAC cases (Yachida et al., 2012). About 66% of all *TP53* mutations are missense mutations that are typically located in the DNA binding domain and result in its functional inactivation (Jones et al., 2008; Yachida et al., 2012). The inactivation of *SMAD4* is the fourth hallmark mutation that occurs during the late stages of PanIN progression to human PDAC (Hruban et al., 2000). *SMAD4* is inactivated in 60% of human PDAC cases, either through somatic mutation (~25%) or through homozygous deletion (~30%) (Hahn et al., 1996; Dardare et al., 2020). *SMAD4* is a co-transcription factor and the central effector of the TGF $\beta$  signaling pathway, which is important for the regulation of cellular growth, differentiation, and tissue homeostasis (Shi & Massagué, 2003).

The study of PanIN precursor lesions supported a stepwise progression model with the gradual accumulation of somatic genetic alterations along with increasing cellular atypia (Hruban et al., 2000). This pattern indicates that the acquisition of driver mutations is associated with waves of clonal expansion. Nevertheless, recent studies challenged this linear progression model, showing an alternative model of punctuated evolution which might be also relevant during the PDAC progression of a subset of human patients (Waddell et al., 2015; Notta et al., 2016). In contrast to linear evolution, this new model proposes the fast accumulation of several genomic and phenotypic changes in a small/unobservable niche, leading to a very fit phenotype that quickly sweeps through the population. One of the observed events that support this model is chromothripsis. This is defined as a phenomenon by which hundreds of chromosomal alterations occur during a single

catastrophic genomic event within a single cell cycle (Stephens et al., 2011; Korbel & Campbell, 2013). The importance of chromothripsis in PDAC is controversial. In some studies, chromothripsis events have been reported at a high frequency in human PDAC (67%), suggesting that PDAC progression occurs in a catastrophic manner (Notta et al., 2016; Real, 2003). In around 16% of cases, the occurrence of chromothripsis caused combined genetic alterations in *KRAS*, *CDKN2A*, *TP53*, and/or *SMAD4* (predominantly affecting two genes) (Notta et al., 2016). By contrast, there are studies in which they have observed chromothripsis in only 10% of PDAC, so, they did not find any evidence of chromothripsis being a major mechanism of driver gene accumulation (Waddell et al., 2015). Besides the controversy of the importance of chromothripsis in PDAC development, chromothripsis has been linked to aggressive tumor behavior in other cancer entities (Rausch, Jones, et al., 2012).

The sequencing of more than 500 pancreatic cancer exomes and more than 100 genomes revealed the extensive mutational heterogeneity of PDAC beyond the four signature mutations (Jones et al., 2008; Biankin et al., 2012; Waddell et al., 2015; Bailey et al., 2016) with mutations frequencies below 10% in many genes involved in DNA damage repair, SWI/SNF mediated chromatin remodeling, axon guidance pathway or in well-known oncogenes (Campbell et al., 2010; Biankin et al., 2012; Waddell et al., 2015). Of note, only some of these genetic alterations, such as in *BRCA1/2*, *PALB2*, *MYC*, or *KDM6A*, could be attributed to certain phenotypes of pancreatic cancer. For instance, genetically unstable human PDACs are significantly associated with homozygous inactivating mutations of *BRCA1/2* or *PALB2* and can be targeted by platinum-based therapy while wild-type tumors do not respond well to that treatment (Waddell et al., 2015). *MYC* amplification, inactivating mutations of *KDM6A*, or upregulated expression of the *TP63ΔN* transcriptional network, are associated with squamous differentiation of human PDAC (Witkiewicz et al., 2015; Bailey et al., 2016). The rest of the mutated genes are not related to clinical, morphological, or biological characteristics of the disease, such as dedifferentiation

or metastatic dissemination (Makohon-Moore et al., 2017). In contrast, the metastatic process has been linked to the epigenetic and metabolic reprogramming of specific pancreatic cancer cells (McDonald et al., 2017; Roe et al., 2017). Accordingly, although significant progress has been made in the genetic characterization of pancreatic cancer in the last few years, the understanding of tumor evolution and phenotypic diversification in PDAC is still limited.

Changes in *Kras* allele dosage are another important event for murine tumor progression and it has been similarly reported in human PDAC (Mueller *et al.*, 2018). Amplification of the mutant *Kras*<sup>G12D</sup> allele in PDAC increases the metastatic potential by inducing epithelial-mesenchymal transition (EMT) signatures, being associated with a worse prognosis. Alternative oncogenic gains in the *Kras* downstream signaling (*Myc*, *Yap1*, or *Nfkb2*) were also identified and described to collaborate with heterozygous *Kras*<sup>G12D</sup> in driving tumorigenesis, but with a lower metastatic potential. This *KRAS* allelic imbalance might help to understand tumor progression and how the tumor acquires major biological features (Mueller *et al.*, 2018).

### **Molecular subtypes of PDAC**

Due to the absence of a clinically relevant classification of PDAC, several studies have classified human PDAC samples according to their transcriptional profiles, which are based on the expression of coding genes (Collisson et al., 2011; Moffitt et al., 2015; Bailey et al., 2016). Importantly, most of the differences between these classifications are probably explained by their different input materials and experimental approaches (Collisson et al., 2019). First, in 2011, Collisson *et al.* microdissected human PDAC from tissue sections away from the stroma and defined three molecular human PDAC subtypes, named classical, quasi-mesenchymal (QM), and exocrine-like (Collisson et al., 2011). These subtypes presented different tumor progression rates and treatment responses in patients. Particularly, the QM subtype correlated with high tumor grade and poor survival. Next, virtual/bioinformatical

microdissection of patient-derived xenografts and human PDAC cell lines, excluding the normal pancreas and tumor microenvironment transcripts from the analysis, allowed to define the classical and basal-like subtypes (Moffitt et al., 2015). The basal-like subtype mostly correlates with the QM subtype from the previous classification. Later, Bailey *et al.* used bulk tumor tissues, including normal and stromal cell types, and proposed four subtypes of human PDAC, denominated squamous, pancreatic progenitor, aberrantly differentiated endocrine exocrine (ADEX), and immunogenic (Bailey et al., 2016).

Recently, the same authors have made efforts to unify the nomenclature between their studies, correlating similar or overlapping molecular subtypes and trying to infer the relationships between them (Collisson et al., 2019). In this context, at least two human PDAC subtypes originally described by Collisson et al. mostly overlap with the subtypes proposed by the following ones. The classical subtype correlates with pancreatic progenitor and immunogenic subtypes, and the QM overlaps with the basal-like and squamous subtypes. It is not clear that the immunogenic subtype is a real subtype of pancreatic cells since it could be caused by contamination with cells of the immune system. Indeed, the exocrine-like signature has been associated with an artifact of acinar cell ‘contamination’ and it is not present in Moffitt’s classification. Moreover, each PDAC subtype has specific gene alterations affecting functionally relevant cancer signaling pathways (Bailey et al., 2016). The pancreatic progenitor subtype preferentially expresses genes involved in early pancreatic development such as *FOXA2/3*, *PDX1*, and *MNX1*. The immunogenic subtype presents immune network upregulation, including pathways involved in acquired immune suppression. It generates a more pronounced immune response. The ADEX subtype presents upregulation of genes that regulate networks involved in *KRAS* activation, exocrine (*NR5A2* and *RBPJL*), and endocrine differentiation (*NEUROD1* and *NKX2-2*). Lastly, the QM, basal-like, and squamous subtypes are associated with a poor prognosis. They present upregulation of the

TP63 $\Delta$ N transcriptional network, mutations in *TP53*, and in genes involved in chromatin modification, such as *MLL2/KMT2D*, *MLL3*, and *KDM6A* (Collisson et al., 2019). They also lose endodermal identity through hypermethylation of genes involved in endodermal and pancreatic cell fate determination, such as *HNF4A* and *GATA6* (Bailey et al., 2016). This makes the transcriptome an optimal method to distinguish the QM/basal-like/squamous subtype from the others (Collisson et al., 2019).

These human transcriptome data sets have been cross-validated with transcriptomes from PDAC mouse models. Genomic and transcriptomic analyses of mouse-derived tumors can be combined with their clinical manifestation to generate associations between genetic mutations and disease phenotypes. Collisson classifier gene sets were projected on human- and mouse-derived tumor data, revealing genomic species differences, but also conserved oncogenic networks between mouse and human molecular subclusters (Collisson et al., 2011; Mueller et al., 2018). The cross-validation separated the human transcriptomes into two subtypes (classical and QM) and the mice ones into three subtypes (classical-equivalent, QM-equivalent, and mesenchymal-equivalent) (Collisson et al., 2011; Mueller et al., 2018). The mesenchymal-equivalent subtype shows a strong upregulation of epithelial-to-mesenchymal transition gene sets and a significant increase of mutant *Kras* dosage compared to the “QM-equivalent” subtype (Mueller et al., 2018). Reliable molecular classifications of PDAC uncover principal oncogenic networks that could potentially improve therapeutic approaches for patients.

## Mouse models of PDAC

Genetically engineered mouse models (GEMMs) have been used in cancer research to study human tumorigenesis molecular mechanisms for decades. Specific mutations can be targeted/introduced to the endogenous locus of genes in mouse embryonic stem (ES) cells by homologous recombination, allowing its transfer through the mouse germ line (Robertson et al., 1986). In addition, conditional and inducible mouse models typically use *Cre/LoxP* or *Flp/FRT* site-specific recombination systems (Sauer, 1998) for tissue- or time-specific gene activation/inactivation.

*KRAS* oncogene is the master driver of PDAC and it is mutationally activated in over 90% of cases (Singhi et al., 2019). One of the *KRAS* hotspot mutations in human PDAC patients is the G>D mutation in codon 12. In 2001, the generation of the *Kras*<sup>LSL-G12D</sup> mouse facilitated for the first time the conditional expression of a hotspot activating *Kras* mutation at physiological levels from the endogenous locus (Jackson et al., 2001). This model was generated through the knock-in of the *Kras*<sup>G12D</sup> mutation silenced by a loxP-stop-loxP (LSL) cassette. The *Kras*<sup>G12D</sup>-allele can be activated by Cre-recombinase expression and subsequent LSL element excision that enables the expression of the downstream oncogene in a tissue- or time-specific manner. This excision can be promoted specifically in the pancreas by breeding *Kras*<sup>LSL-G12D</sup> mice to transgenic mice expressing Cre recombinase from pancreatic-specific promoters. The pancreatic and duodenal homeobox 1 (*Pdx1*) (Gannon et al., 2000) and the pancreas-specific transcription factor 1a (*Ptf1a/p48*) genes (Nakhai et al., 2007) are early developmental transcription factors in the mouse pancreas. Both of them have been widely used to achieve pancreas-specific conditional expression of the Cre recombinase. The *Pdx1* gene is expressed earlier at mouse embryonic day E8.5, while the *Ptf1a* gene is expressed at E9.5. Both genes support the LSL excision in the exocrine pancreatic compartment (Offield et al., 1996; Kawaguchi et al., 2002). Unfortunately, the activity of both promoters in non-pancreatic tissue has been also

described. Thus, *Pdx1* expression has been described in the rostral duodenum, bile duct, and antral stomach during foregut development (Gannon et al., 2000). Additionally, the *Ptf1a* gene product can also be detected in the retina and the brain (Nakhai et al., 2007).

Expression of endogenous levels of *Kras*<sup>LSL-G12D</sup> in the pancreas with either promoter (KC model), drives the formation of the full spectrum of PanIN lesions with slow progression to PDAC and a median tumor latency of 67 weeks (Hingorani et al., 2003). The number and grade of these PanINs increase with the age of the mice. These PanINs spontaneously progress at low frequency to both locally invasive PDAC and metastatic spread to the liver, lungs, and lymph nodes, as found in humans. These observations revealed that activating *Kras* mutations are essential for pancreatic carcinogenesis. Moreover, it allowed for the first time the functional testing of hallmark mutations in PDAC and its progression patterns in an *in vivo* context through the combination of this KC model with conditionally deleted or mutated alleles of tumor suppressor genes known to be mutated during PanIN-to-PDAC progression in humans, such as *CDKN2A*, *TP53*, and *SMAD4*.

Whereas *Cdkn2a* knock-out alone in the mouse pancreas did not lead to the development of any malignancy, the homozygous deletion of the *Cdkn2a* locus in addition to *Kras*<sup>G12D</sup> in the mouse pancreas accelerated the early appearance of PanIN lesions and the rapid progression to highly invasive and metastatic spread mainly to the lymph nodes (Aguirre et al., 2003). The survival of *Kras*<sup>G12D</sup>;*Cdkn2a*<sup>Flox/Flox</sup> mice was dramatically reduced compared with the KC model, from more than 12 months to 2 months. This rapid tumorigenesis could explain the low metastasis frequency in distant organs such as the liver and lungs. Therefore, *Kras* activation seems to be an initiating step in PanINs development, and *Cdkn2a* loss accelerates the malignant progression of the disease.



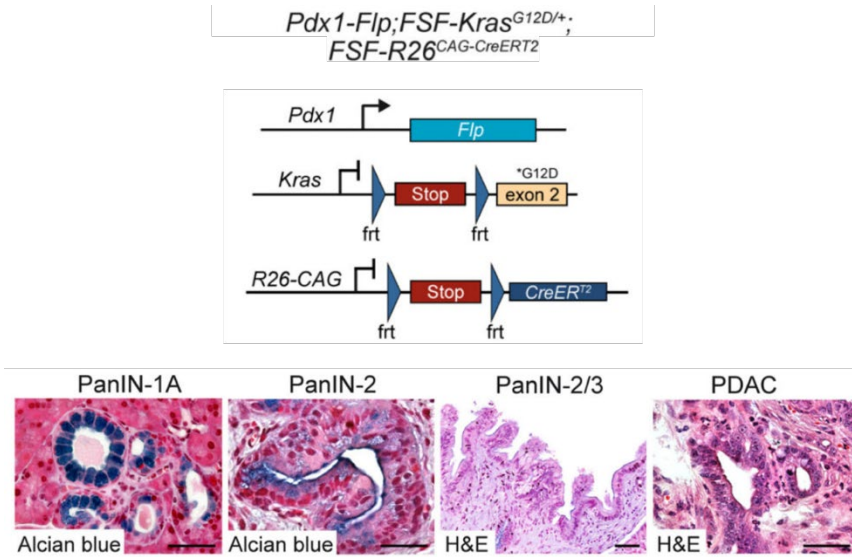
*Trp53* is typically inactivated through homozygous deletion and/or somatic mutations such as the R172H substitution mutation (mouse ortholog of human R175H). To study the role of *Trp53* inactivation during PanIN progression and PDAC development in mice, conditional *Trp53* knock-out (*Trp53*<sup>Flox</sup>) and/or mutant (*Trp53*<sup>LSL-R172H</sup>) alleles were crossed into the Pdx1-Cre;*Kras*<sup>LSL-G12D</sup> background (*Kras*<sup>G12D</sup>;*Trp53*<sup>Flox/Flox</sup> and *Kras*<sup>G12D</sup>;*Trp53*<sup>LSL-R172H/LSL-R172H</sup>) (Hingorani et al., 2005; Bardeesy, Aguirre, et al., 2006). Without expression of oncogenic *Kras*<sup>G12D</sup>, homozygous deletion of *Trp53* and heterozygous expression of R172H mutation did not result in any morphological changes or any signs of pancreatic carcinogenesis. Nevertheless, *Kras*<sup>G12D</sup>;*Trp53*<sup>Flox/Flox</sup> mice resulted in rapid tumorigenesis in the pancreas and a severely reduced median survival of 2 months as compared to 12 months of the *Kras*<sup>G12D</sup> model. The *Kras*<sup>G12D</sup>;*Trp53*<sup>LSL-R172H/LSL-R172H</sup> mice promoted an acceleration of invasive cancer formation, frequent liver or lung metastases and reduced the survival of the mice to 5 months (Hingorani et al., 2005; Bardeesy, Aguirre, et al., 2006). These primary tumours and metastasis exhibited a high degree of genomic instability as seen in human PDAC. However, mutations in other important cancer genes were not detected highlighting a strong oncogenic cooperation between *Kras* and *Trp53*. Interestingly, the metastatic burden was found to be increased in *Kras*<sup>G12D</sup>;*Trp53*<sup>LSL-R172H/LSL-R172H</sup> compared to *Kras*<sup>G12D</sup>;*Trp53*<sup>Flox/Flox</sup> mice, although they developed primary pancreatic cancer in a similar manner (Morton et al., 2010).

Two different models were developed to investigate the role of *Smad4* deletion for PDAC initiation and progression. As already observed for complete inactivation of *Cdkn2a* or *Trp53*, the pancreas-specific homozygous knock-out of *Smad4* alone did not result in morphological abnormalities or the induction of any malignant lesions in the mouse pancreas (Bardeesy, Cheng, et al., 2006). In contrast, the combination of pancreas-specific expression of oncogenic *Kras*<sup>G12D</sup> and complete deletion of *Smad4* resulted in a dramatically reduced median survival of 80 days.

*Kras*<sup>G12D</sup>;*Smad4*<sup>Flox</sup> mice predominantly developed pancreatic lesions resembling intraductal papillary mucinous neoplasia (IPMN, a PDAC precursor in humans) and shows higher expression of epithelial markers instead of the frequent epithelial-to-mesenchymal transition observed in tumors with intact *Smad4* (Bardeesy, Cheng, et al., 2006).

Importantly, in all mouse models where oncogenic *Kras* is activated during embryogenesis, low-grade PanINs take 8-12 months to develop to high-grade lesions and PDAC. In this context, tumor latency can be reduced to 4 months by inducing acute pancreatitis with caerulein (Carrière et al., 2009, 2011)

Recently, a new mouse model of pancreatic cancer has been developed (Schönhuber et al., 2014). It consists on an inducible dual-recombinase system that combines the flippase-FRT (Flp-FRT) and the Cre-loxP recombination technologies for controlling tumor initiation and secondary genetic manipulation, respectively (Figure 2). It expresses an oncogenic allele of *Kras* specifically in the pancreas ( $Tg^{Pdx1-Flp}:Kras^{FSF-G12D}$ ). After pancreatitis induction with caerulein, this model develops PDAC that fully recapitulates all types of lesions observed in human PDAC progression as well as metastatic spread to lymph nodes, liver, and lungs before the age of 8 months. Additionally, this model expresses CreER<sup>T2</sup> (R26<sup>FSF-CreERT2</sup>), a tamoxifen-inducible version of Cre recombinase. CreER<sup>T2</sup> is controlled by an upstream FSF (FRT-Stop-FRT) cassette, which allows CreER<sup>T2</sup> expression to tissue expressing Flp recombinase (in this background specific to the pancreas). This scenario allows the introduction of any genetically engineered allele that depends on Cre expression.



**Figure 2 | Genetic strategy to induce PDAC and introduce a second genetic manipulation.** Left, genetic strategy to activate oncogenic *Kras<sup>G12D</sup>* in the pancreas using the *Flp-FRT* recombination system (*Pdx1-Flp* activation) and to secondly activate any allele by tamoxifen-mediated activation of *CreERT2* in the *Flp* lineage through the *Cre-loxP* recombination system. Right, alcian blue and hematoxylin and eosin (H&E) stained sections of PanIN lesions and PDAC of *Pdx1-Flp;FSFKras<sup>G12D/+</sup>* mice. Modified from (Schönhuber et al., 2014).

Finally, in the last decade, tumorigenic strategies have been developed to confirm and identify relevant pancreatic cancer genes in mice. These new systems, are based either on the random mobilization of transposons, such as the Sleeping Beauty or PiggyBac transposon systems (Dupuy et al., 2005; Ding et al., 2005) or on the simultaneous edition of multiple genes using the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 technology, have unraveled the role of new tumor suppressor genes such as *Uspx9* or *Foxp1* (Pérez-Mancera et al., 2012; Rad et al., 2015)

## **Intratumor heterogeneity**

The simple model of linear cancer progression, in which tumors arise as the result of the progressive accumulation of genetic alterations, each one followed by a selection of the fittest clone with the metastases as the final fatal stage, has been challenged in the last decades. Instead, in 2008 Campbell and collaborators identified the simultaneous presence of different genetic clones in chronic lymphocytic leukemia (Campbell et al., 2008). Later, this observation was also detected in many solid tumors (McGranahan & Swanton, 2017). Therefore, tumors seem to contain different cell populations forming complex ecosystems in constant evolution after malignant transformation. It is also difficult to reconcile this new view of tumor progression with the presence of a small population of cancer stem cells, identified in some tumor types and postulated to support the growth of the tumor bulk (Magee et al., 2012). According to this, it is possible that one or other model predominates in some tumor types, whereas a combination of both is needed to explain the behavior of others.

### **Clinical implications of intratumor heterogeneity**

The presence of ITH has profound clinical implications. It has been postulated that tumor heterogeneity plays a major role in treatment resistance and metastasis, the two major causes of cancer-associated death. In terms of diagnosis, the main source of human tumor samples comes from single biopsies. This material is not sufficient to capture the whole molecular heterogeneity in the primary tumors (Fisher et al., 2013). This challenges our ability to correctly classify the molecular features of a tumor and determine the subtype of the disease from a single sample (Gerlinger et al., 2012, 2014). Interestingly, the presence of clones that could be categorized in different molecular groups has been described inside the same ovarian primary tumor (Bashashati et al., 2013).

Treatment of cancer patients is also hampered by the presence of ITH. As a correct molecular characterization determines a preferred treatment, targeted therapies might affect only some cell clones inside the tumors, while others are insensitive to them. As a consequence of this clonal selection induced by therapies, responses to therapy are expected to be reduced with time. This could explain the frequent cancer relapses arising after targeted therapies. Moreover, the presence of treatment-resistant cell clones inside the primary tumor has been detected even before treatment. These cells are postulated to be the main cause of tumor relapse. Additionally, there is evidence that the spread of tumor cells, as well as seeding metastases, might be sometimes an early event in tumor progression, even at a pre-malignant stage. If metastases are seeded by some of these minority clones, the limited success of the same therapy in metastatic growths can be expected (Rhim et al., 2012). Consequently, the grade of ITH has been associated with a worse prognosis in cancer patients (Fisher et al., 2013; Almendro et al., 2014).

### **Genetic intratumor heterogeneity**

Genetic intratumor heterogeneity (ITH) is likely the result of an ongoing genetic instability followed by independent branched evolution of the generated more competitive subclones (Jamal-Hanjani et al., 2017; de Bruin et al., 2014; Campbell et al., 2010; Swanton, 2012; Hiley et al., 2014; Yachida et al., 2010; Gerlinger et al., 2014). Interestingly, the assumption that clonal subpopulations must always be in competition has been challenged by the results of studies showing that cooperation between clearly different subclones might be required for tumor propagation in cancers with non-cell-autonomous initiating events, and that metastatic sites can be seeded by genetically distinct subclones from the same primary tumor (known as polyclonal-seeding) (Cleary et al., 2014; Gudem et al., 2015; Hong et al., 2015; Maddipati & Stanger, 2015). The exchange of tumor material

between different metastatic sites (known as cross-metastatic seeding) and recolonization of the primary tumor by circulating tumor cells (known as tumor self-seeding) (Gudem et al., 2015; Hong et al., 2015; M.-Y. Kim et al., 2009) have been also been reported.

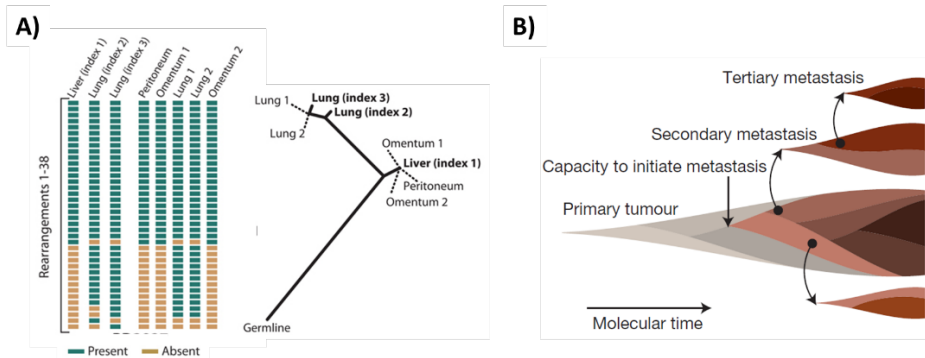
How this genetic heterogeneity is generated and maintained is not well understood. It could be either an ongoing process consequence of a general genetic instability or a sporadic event happening at a specific moment during tumor progression. This second scenario can derive from the occurrence of a catastrophic event that should produce a massive genetic variation in a very short period of time. Examples of this kind of event, such as *chromothripsis* or *kataegis*, have been described by several authors (Nik-Zainal et al., 2012; Stephens et al., 2011).

One of the most cited studies of ITH in cancer was based on multiregional sequencing of renal carcinomas and associated metastatic sites (Gerlinger et al., 2012). This multisampling strategy was based on biopsy sampling of multiple regions within the primary tumor as well as associated metastasis which were sequenced (Jamal-Hanjani et al., 2017; Gerlinger et al., 2014). By multiregional sequencing, two categories of somatic single-nucleotide mutations are distinguished by their presence in all the regions (common or *founder* mutations) or in some of them (specific or *progressor* mutations). The common mutations are assumed to accumulate in the early stage of cancer evolution, indicating a common genetic origin. The common ancestor clone that has acquired all the founder mutations then branches into subclones, which accumulate specific *progressor* mutations and contribute to the formation of ITH. The evolutionary history of the tumor can be inferred through these multiregional mutational profiles by constructing a phylogenetic tree (Gerlinger et al., 2012, 2014). Interestingly, in this study, it can be observed that all metastatic regions emerged from a single evolutionary branch that diverged quite early from the branch responsible for the expansion of the primary tumor.

**Intratumor genetic heterogeneity in PDAC**

A high degree of ITH has been reported in PDAC (Yachida et al., 2010; Campbell et al., 2010). Yachida and colleagues demonstrated that clonal populations that give rise to distant metastases in a branched evolutionary pattern are already present in the primary tumor. These clones genetically evolve from the original non-metastatic parental clone. Accordingly, the genetic heterogeneity of metastases reflects heterogeneity already present in the primary carcinoma (Yachida et al., 2010). The timing of genetic evolution in pancreatic cancer has been studied. Mathematical modeling data suggested a timeframe of at least a decade between the occurrence of the initiating mutation and the appearance of the parental non-metastatic founder cell. Cancer required an additional five years to acquire metastatic potential and patients died an average of two years later (Yachida et al., 2010). In this case, the degree of genetic differences between the primary tumor and the metastatic sites indicated that metastasis emerged late during tumor development.

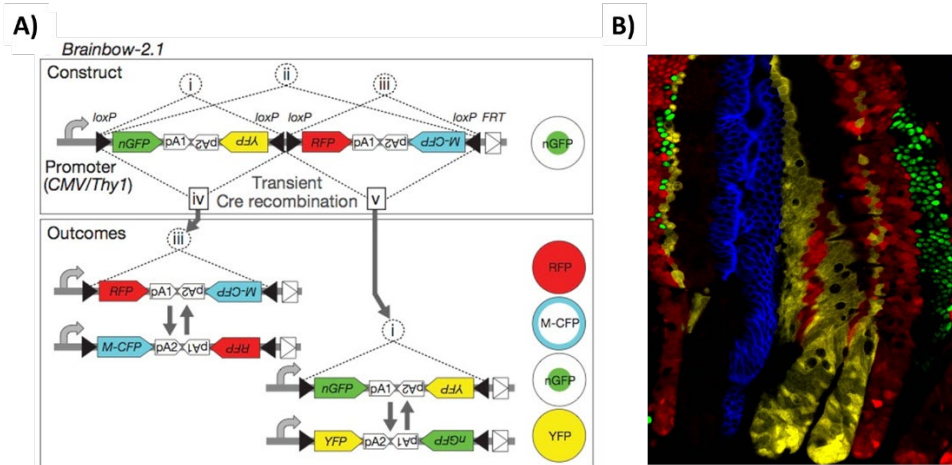
In 2010, Campbell and colleagues also demonstrated that genome instability occurred early in their samples of pancreatic cancer. Additionally, they found that metastasis-initiating cells were genetically heterogeneous, that seeding metastasis may require driver mutations beyond those required for primary tumors, and that phylogenetic trees across metastases show convergent evolutionary paths and organ-specific relationships between metastases (Figure 3) (Campbell et al., 2010). Later, this polyclonal-seeding, in which genetically different subclones from the primary tumor colonize the metastatic sites, was confirmed (Maddipati & Stanger, 2015). Importantly, no recurrently mutated metastasis driver gene has been found yet in human PDAC (Makohon-Moore et al., 2017).



**Figure 3 | Phylogenetic relationships of different metastases from a PDAC patient and a derived model of clonal evolution of metastases. A)** Left, presence (green) or absence (yellow) of somatic rearrangements across the metastases from a patient. Right, inferred phylogenetic tree of relationships for metastases, showing a clade of abdominal metastases and a more evolved clade of lung metastases. The genetic distance between nodes is proportional to the length of black lines. **B)** Model for the clonal evolution of metastases derived from the observed phylogenetic relationships. Molecular time is represented from left to right and is associated with the acquisition of genetic alterations, subclonal evolution, and expansion during the progression of the disease, even within metastases, observing tertiary metastases seeding. Modified from (Campbell et al., 2010).

To study the development of this ITH *in vivo*, the generation of mouse models, such as the Confetti mice, could be very useful (Schepers et al., 2012). This model relies on a lineage-specific genetic labeling system, named Brainbow (Livet et al., 2007). It is based on the random and inheritable expression of different fluorescent markers. Thus, a variable number of different fluorescent proteins, surrounded by loxP sites, are located downstream of a constitutive promoter. After the action of Cre recombinase, one of these proteins is located immediately after the promoter. The election of the recombination event is mainly random and once it occurs, it is permanent and inherited by all descendant cells (Figure 4). This mouse model could be combined with the previously described PDAC mouse model ( $Tg^{Pdx1-Flp}:Kras^{FSF-G12D}$ ) (Schönhuber et al., 2014) to study the tumor clonality in different stages of PDAC progression.





**Figure 4 | Genetic tracing system based on the stochastic expression of fluorescent proteins. A)** Brainbow-2.1 construct containing four fluorescent proteins flanked by loxP sites in opposite orientations. Cre recombination leads to a random inversion or deletion of the proteins which are expressed in different cellular locations: red and yellow fluorescent proteins in the cytoplasm, cerulean in the membrane, and the green protein in the nucleus. From (Livet et al., 2007). **B)** Labelled stem cells from the small intestine of the Confetti mouse (expressing the Brainbow-2.1 allele). From (Schepers et al., 2012).

### Functional intratumor heterogeneity

ITH has also been revealed at the transcriptional level inside the tumors and metastases due to the recent advances in single-cell technologies. In particular, tumor cells with different transcriptional profiles have been found inside the tumors (Patel et al., 2014; Puram et al., 2017). These tumor cells are able to establish mutually beneficial exchanges (Tabassum & Polyak, 2015) or, alternatively, hierarchical relationships through, for example, the potential presence of cancer stem cells (CSCs). They have been described as a minority group of cells with extraordinary capabilities to promote tumor progression, metastatic spread, and treatment resistance (Prasetyanti & Medema, 2017). Importantly, malignant cells from the primary tumor present the ability to modify the environmental conditions

of the metastatic sites through the production of extracellular vesicles (Peinado et al., 2017). For example, it has been shown that derived exosomes from PDAC favor the formation of the pre-metastatic niche in the liver (Costa-Silva et al., 2015).

Besides the diverse malignant cells, we found local stromal components and immune cells. This complex diversity within tumors is known as tumor microenvironment (TME) and affects the aforementioned processes by activating fibrotic pathways and establishing a pro-inflammatory environment. In this context, cancer-associated fibroblasts (CAFs) are an abundant component of the stroma in most solid tumors and actively participate in the formation of a desmoplastic tumor niche (T. Liu et al., 2019). These are the main characteristics of human PDAC. Moreover, several types of CAFs have been identified in different cancers, including PDAC (Elyada et al., 2019). Importantly, each group of CAFs has specific functions, recruiting immune cells or inducing epithelial-mesenchymal transition in tumor cells (Tirosh, Izar, et al., 2016; Puram et al., 2017). Moreover, immune cells are also frequently infiltrated within tumors. Their presence also influences the outcome of the patients, such as the better outcome of patients with non-small cell lung cancer presenting a higher proportion of active CD8+T lymphocytes versus the exhausted ones (Guo et al., 2018). So, the complex relationship between all the cell types present in the tumors determines a tolerant or non-tolerant TME.

## Next-generation sequencing

### The origins of DNA sequencing

Since the description of the double helix structure of the deoxyribonucleic acid (DNA) in 1953 by Watson, Crick, and Franklin, many attempts were made to sequence DNA. The first-generation sequencing technologies emerged in 1977 and include the Maxam-Gilbert method, based on chemical cleavage (Maxam & Gilbert, 1977), and the Sanger method, based on a chain-termination strategy (Sanger, Nicklen, et al., 1977). Following the latter approach, Sanger and colleagues successfully sequenced the first DNA genome belonging to bacteriophage PhiX174 with a whole length of 5375 base pairs (bp) (Sanger, Air, et al., 1977). Since then, the Sanger method was popularized as Sanger sequencing and became the sequencing gold standard for the next 30 years. In this approach dye-labelled deoxynucleotides (dNTPs) and dideoxy-modified dNTPs (ddNTPs) are mixed and supplied to a DNA polymerase. During the elongation, some strands incorporate a ddNTP, thus terminating elongation because ddNTPs lack the 3'-hydroxyl group, which is required to form a bond with the 5'-phosphate of the next dNTP (Chidgeavadze et al., 1984). Then the strands are separated by electrophoresis and the terminal base label of each strand is identified according to the fragment length. A significant number of modifications were implemented to Sanger sequencing over the years. Thus, a set of four fluorescent labels reduced the number of required independent reactions, this, together with fragment size discrimination through capillary gel electrophoresis, allowed the simplification and automation of the process (L. M. Smith et al., 1986). In 2001, the first human genome map with their approximately  $3 \times 10^9$  nucleotides was completed using the improved Sanger method (Lander et al., 2001).

### **Next-generation sequencing technologies**

Since 2005, the emergence of next-generation sequencing (NGS) technologies, also known as massively parallel or high-throughput sequencing technologies, revolutionized genomic research. The advantages of NGS compared with traditional sequencing methods include the analysis of hundreds of independent sequences generated from a specific region and the consequently higher sensitivity to detect low-frequency variants, the speed-up of sequencing even for high sample collections, and the high production of sequence in a single experiment reducing very significantly the cost per sequenced base (Goodwin et al., 2016). Nowadays, NGS technologies facilitate, for instance, the sequence of a complete human genome within a couple of days for less than \$1000 compared with the sequencing of the first human genome which cost about \$1 billion and took 13 years to be completed using Sanger sequencing (Schwarze et al., 2020).

The first commercially successful NGS platforms were based on the Roche 454 technology, which later led to the Ion Torrent/Proton system (Rothberg et al., 2011).

Roche 454 sequencing platform is the most closely related to the original pyrosequencing technology (Margulies et al., 2005). The process is carried out in fragments immobilized into beads captured inside emulsion droplets. One of the four fluorescently labeled dNTPs is sequentially added in each sequencing cycle. The nucleotide incorporation in a molecule releases a pyrophosphate that is converted into ATP by an ATP-sulfurylase. In the presence of ATP, luciferase converts luciferin to oxyluciferin to generate light, which is then detected and captured identifying those molecules that incorporate the specific nucleotide in each cycle. This platform was very successful between 2005 and 2008 due to a large number of sequences generated and the low cost of sequencing. In 2012, the 454 technology became non-competitive and Roche ceased supporting this platform in 2016.

Following the same principle, Life Technologies developed in 2010 the Ion Torrent/Proton system (Rothberg et al., 2011). It follows the same strategy as 454 but records the incorporation of each nucleotide with a semiconductor that detects pH changes in the reaction solution produced by the release of a hydrogen ion (H<sup>+</sup>) when dNTPs are incorporated into the DNA. Consequently, Ion Torrent/Proton instruments do not require modified nucleotides or optical methods, such as fluorescence or chemiluminescence, which promised to greatly reduce the sequencing reaction time and cost. However, the process is time-consuming (1 hour/100 bp) (Kchouk et al., 2017).

Both strategies produce an output of reads of variable length that is determined by the template sequence. The lack of terminators in both systems was a major limitation that generates high error rates in homopolymer repeats and ultimately limits the resolution of polynucleotide tracks (Kchouk et al., 2017).

In 2006, the Solexa sequencing platform was commercialized and acquired by Illumina in 2007. Illumina sequencing, as it was popularly known, is also based on *sequencing by synthesis* approach, using a molecular clustering technique. As a novelty, four reversible terminator nucleotides labeled with a different fluorophore are present in each sequencing cycle (Bentley et al., 2008). The identity of the incorporated nucleotide in each cluster is determined by recording the fluorescence. Then, the terminator with the fluorescent label is eliminated and the chamber is ready for a new incorporation cycle. In this system, the template is immobilized on a glass surface which is densely coated with oligonucleotides complementary to the adaptors of the library. In paired-end sequencing mode, after the production of the first read, the adapter on the opposite side of the read hybridizes with another set of oligonucleotides that are present on the glass, creating a bridge that allows the reverse sequencing. Then, a new set of sequencing cycles is performed producing the second read from the opposite extreme of the fragment (Morozova & Marra, 2008). Currently, Illumina is the most popular technology in the NGS market, offering

scalable options to complement the requirements of each study, the cost of sequencing, and the intended use of the sequencing data (Buermans & den Dunnen, 2014). It offers low throughput cheap benchtop platforms (MiniSeq and MiSeq series) as well as high throughput more expensive production systems (NextSeq, HiSeq, and NovaSeq series).

Oxford Nanopore technology emerged in 2009 offering significant differences from other NGS technologies (Clarke et al., 2009). This novel system measures electrical current fluctuations as different molecules go through  $\alpha$ -hemolysin or a *Mycobacterium smegmatis* porin A (MspA) nanopores (Kircher & Kelso, 2010; Y. Wang et al., 2014). It allows DNA, RNA, and protein sequencing. The distinguishing features of this new technology are (1) the sequencing of single molecules avoiding the need for DNA fragments amplification (L. Liu et al., 2012), (2) the fast sequencing in real-time (instead of being paused after each base incorporation) (Schadt et al., 2010), and (3) the generation of long sequencing reads solving problems related with repetitive regions (Kchouk et al., 2017). Longer sequence reads allow sequencing through extended repetitive regions and facilitate closing gaps in current reference assemblies generated from short reads. However, there is a major issue of a high error rate (Mikheyev & Tin, 2014). In 2014, following the previous approach, the first commercially available sequencer in portable size appeared, termed MiniOn. Despite its high sequencing error rate (around 14%), it is particularly attractive for surveillance and clinical diagnostic applications where resources are limited (Lu et al., 2016).

### **NGS applications**

The high flexibility and power of NGS technologies enable more than 200 name-specific applications. Here I will summarize only some of them. Importantly, the type of biological information extracted from each of them is determined by the unique combination of the starting biological material and the specific protocols

related to library preparation, sequencing, and analysis (Metzker, 2010). To extract meaningful information from the sequencing data, specific software have been designed for each NGS application. Accordingly, a new generation of bioinformatics tools has emerged as a requisite to analyze sequencing data. Additionally, researchers frequently need to generate small pieces of code to adapt the analysis to their specific requirements. As it would be very difficult to do a comprehensive review of all available NGS applications and bioinformatic tools, I would focus on those that are especially relevant in the context of the present thesis.

### **Identification of genetic mutations (DNA-seq applications)**

The unbiased sequencing of the complete genome (whole-genome sequencing or WGS) is the most comprehensive method to identify differences versus a reference genome (re-sequencing) or to determine the full sequence of a previously unknown genome (*de novo* sequencing). This can be applied to identify the complete list of nucleotide sequence changes present in the entire genome including single nucleotide variants (SNV) or substitutions, small insertion and deletions (indel), copy number alterations (CNA), and large-scale reorganizations such as translocations, deletions, insertions, and inversions.

Despite the reduction in sequencing costs, the study of the whole genome at high coverage involves a large economic investment. In this sense, several approaches have been developed to focus sequencing efforts on specific regions of interest of the genomic DNA (targeted sequencing), either performed through region capture with designed soluble probes or through PCR-based strategies, both constitute cheaper alternatives. The first strategy consists of the use of synthetic probes complementary to these regions of interest that can be hybridized on a solid support or in a liquid phase. These probes are usually combined with biotin, which allows the fragments to be purified with paramagnetic beads combined with streptavidin and a magnet. The most extended application is the targeted

sequencing of all the protein-coding exons of the genome (whole exome), which corresponds to only 1% of the whole human genome (Hodges et al., 2007; Teer & Mullikin, 2010). Additionally, several companies offer the customer the opportunity to design a pool of probes to purify any combination of selected genomic regions (Mamanova et al., 2010). The second strategy for the enrichment of specific regions is based on their amplification by PCR. Commercial platforms such as Haloplex (Agilent) or Ampliseq (Life Technologies) have exploited this strategy, which offers the advantage of requiring a smaller amount of starting material, as well as being able to work with lower quality material such as that extracted from fixed tissue. However, PCR-based strategies have several limitations, such as potential errors introduced by the polymerase, strands bias, or the low uniformity of the sequenced regions (Q. Wang et al., 2017).

In recent years, new alternatives have been developed to easily generate libraries of fragments of the regions of interest avoiding the amplification bias. Some of them, such as the use of molecular inversion probes (MIPs) or the primer extension capture (PEC) techniques, allow the reduction of PCR errors and biases while maintaining a simple protocol without the requirement of expensive reagents or infrastructures (Niedzicka et al., 2016; Briggs, 2011).

Regarding data analysis, it is convenient to perform an initial quality control of the sequencing data before extracting biological information from it. Tools such as FastQC allow extracting basic parameters directly, like the abundance of specific repetitive sequences, the homogeneity of the sequence quality through the read length, or the GC content of the sequencing data, from the raw read data. Other useful parameters such as the diversity of the sequenced fragment library (amount of PCR duplicates) or the distribution of insert sizes can be extracted only after read alignment using tools like SAMTools (H. Li et al., 2009) or PICARD (Broad Institute. Picard. <http://broadinstitute.github.io/picard>). Most of the DNA-seq applications require the alignment of the generated short sequences (reads) to a reference



genome. This step is typically the most resource-consuming process and the most critical for the subsequent steps. Several aligners have been specifically designed for the use of NGS data. Almost all of them try to identify all potential locations on the genome for a specific part of the read (seed). Subsequently, this seed is extended and the number of mismatches or gaps required for the alignment of the complete read is computed by the aligner to choose the most probable location of the read. This is finally reported and accompanied by a quality score representing the certainty of the alignment. Following the previous strategy, SOAP (R. Li et al., 2008) and MAQ (H. Li et al., 2008) were the first aligners developed. These tools were posteriorly improved using Burrows-Wheeler indexing of the reference genome by BWA (H. Li & Durbin, 2009) and Bowtie (Langmead et al., 2009). This alignment is standardized in the SAM/BAM/CRAM format, and several software suites have been developed for its manipulation like sorting, cleaning, or indexing. Some examples are SAMTools (H. Li et al., 2009), PICARD (Broad Institute. Picard. <http://broadinstitute.github.io/picard>), or GATK (McKenna et al., 2010). Additionally, some tools like the Integrative Genomics Viewer (IGV) allow the graphical visualization of the alignments (Thorvaldsdóttir et al., 2013).

After the quality control and the read alignment, small sequence changes, genomic structural variants, and copy number changes can be identified. Among the different computer strategies to call mutations, the identification of base substitutions is probably the most studied one. Many of the tools available are based on a Bayesian model, initially described in MAQ, to compute the different probabilities associated with each potential genotype in a specific genomic position. This approximation works well for germline substitutions. Nevertheless, in the detection of cancer somatic mutations, where typically, the tumor sample is compared to a non-tumoral DNA sample; the presence of normal DNA contamination, copy number alterations, and intratumor heterogeneity make it difficult to build a model of expected frequency. In this context, new software has

been created specifically to detect somatic single nucleotide variants (sSNVs). Some examples are Strelka (Saunders et al., 2012), VarScan 2 (Koboldt et al., 2012), Seurat (Christoforides et al., 2013), MuTect (Cibulskis et al., 2013), and Ramses (Martínez et al., 2014). Detecting indels have been proved to be a more difficult task and very few tools show good sensitivity and specificity. The main problem relies on the alignment of reads containing this type of mutation. For that reason, some tools like Pindel (Ye et al., 2009) and Dindel (Albers et al., 2011) use a second alignment step on singleton reads to identify insertions and deletions. Nevertheless, these tools are not very specific and generally require additional filtering steps.

Interestingly, in Illumina platforms paired-end sequencing can be used to identify structural variants like insertions, deletions, duplications, inversions, and translocations. Chromosome mapping, read orientation, or insert size of the read pair are used to identify genomic breakpoints. Additionally, breakpoints can also be identified in high coverage sequencing data in split-apart reads in which different segments of the same read map to different parts of the genome. Some tools that can be used to identify these kind of mutations are BreakDancer (Chen et al., 2009) and DELLY (Rausch, Zichner, et al., 2012).

Finally, the quantitative nature of NGS technologies can also be used to detect copy number changes. Thus, the number of reads coming from a specific genomic region depends on the number of DNA copies for that region. Once the genome has been segmented and the data normalized, some tools allow the identification of deletions or amplifications. Some tools available for this are CODEX (Jiang et al., 2015), CNV-seq (Xie & Tammi, 2009), Control-FREEC (Boeva et al., 2012), and ExomeCNV (Sathirapongsasuti et al., 2011). The processing of targeted sequencing data needs to consider the potential bias produced during enrichment. To solve it, CopywriteR software was designed to use only off-target data, theoretically not affected by enrichment biases (Kuilman et al., 2015).

## Identification of epigenetic alterations (DNA-seq applications)

Other DNA-seq applications provide information about the DNA structure through the combination of NGS with the isolation of specific genome regions based on its accessibility to specific enzymes. Thus, DNase-seq (sensitive regions to DNase I digestion), FAIRE-seq (sensitive regions to cross-link with paraformaldehyde), or ATAC-Seq (accessible regions by Tn5 transposase) techniques show the open or closed structure of the DNA genome-wide (Meyer & Liu, 2014). The specific DNA-DNA contacts can also be elucidated using NGS, allowing a higher order of structure resolution. For probing the three-dimensional architecture of whole genomes, Hi-C strategy combines the chemical cross-linking of nearby DNA sequences with fragmentation by restriction digestion and random ligation of the generated fragments prior to their genome-wide massively parallel sequencing (Lieberman-Aiden et al., 2009). Furthermore, genome-wide profiling of DNA-binding proteins, histone modifications, or nucleosomes can also be achieved through chromatin immunoprecipitation followed by sequencing (ChIP-seq). This technique is based on the ultra-sequencing of immunoprecipitated DNA using a specific antibody against a DNA-binding protein. In that way, all protein-DNA binding sites in a specific moment can be identified (Johnson et al., 2007). Finally, DNA methylation can also be detected through ChIP-seq with anti-5-methylcytosine (5-mC) antibodies, known as methylated DNA immunoprecipitation followed by sequencing (MeDIP-seq) (Down et al., 2008), and through the application of DNA-seq technologies to bisulfite-modified DNA (Y. Li & Tollesbol, 2011).

The ChIP-seq, FAIRE-seq, ATAC-seq or DNase-seq data analysis relies upon the identification of specific regions with a significant increase in sequence coverage compared to the background “noise” of the genome (enriched region identification). These regions or peaks can be identified by different statistical methods, after genome segmentation and coverage normalization of the sequencing data. The most popular tool is MACS (Zhang et al., 2008), but other tools such as FindPeaks (Fejes et

al., 2008), F-seq (Boyle et al., 2008), and QuEST (Valouev et al., 2008) are also used. DESeq2 also allows the identification of enrichment differences among samples (Love et al., 2014).

### **Identification of transcriptional alterations (RNA-seq applications)**

NGS technologies generate a number of reads proportional to the initial number of template molecules in the sample. This characteristic allows the use of cDNA converted from RNA as starting material (RNA-seq) to perform gene expression studies of all types of transcripts, including messenger RNA (mRNAs), non-coding RNAs, and small RNAs, with higher sensitivity and reliability than the obtained by array-based technologies. Additionally, as the RNA-seq strategies are not based on a previously generated gene model, new genes, fusion genes or splice variants, and other post-transcriptional modifications can be identified with the appropriate analysis software (Z. Wang et al., 2009). Moreover, specific library preparation methods retaining the molecule strand orientation during the cDNA conversion have been developed. (Levin et al., 2010).

In order to align RNA-seq data, first TopHat and after HiSat tools were developed (D. Kim et al., 2013, 2019; Zhang et al., 2021). They work in two stages. First, all the exons of the genome are identified by a stringent direct alignment of the reads to the reference genome; unaligned reads are subsequently used to find the junctions (splicing events) between the exons. Gene expression change detection relies on the quantitative nature of NGS technologies. Specifically, the number of reads generated from a specific transcript is proportional to the number of copies of this transcript in the original sample, the length of the transcript, and the total amount of sequence generated from this sample. The expression data is usually normalized as RPKM (reads per transcript kb and per million total reads) or FPKM (fragments/read-pairs per transcript kb and per million total reads) to compare the gene expression across different sequencing reactions. Two of the most

broadly used software to detect gene expression changes in RNA-seq data are DESeq2 (Love et al., 2014) and Cufflinks (Trapnell et al., 2010, 2013). Additionally, Cufflinks suite counts with specific tools to identify new genes, transcript-specific expressions, or new splice variants. Furthermore, if the two alleles of a specific gene have different nucleotide sequences, RNA-seq data can be used to detect the specific expression of each one of the different alleles.

## Single-cell sequencing technologies

The recent development of technologies based on sequencing individual cells provides unprecedented opportunities to analyze the cancer complexity at the single-cell resolution (reviewed in (González-Silva et al., 2021)). This strategy facilitates the characterization of the internal architecture of the tumor during disease progression at the genetic, epigenetic, and transcriptomic levels. At the genetic level, single-cell DNA-sequencing (scDNA-seq) technologies offer higher sensitivity than NGS technologies in the detection of minority clones, the reconstruction of clone structure, and the identification of common or exclusive alterations in the same cells. To analyze chromatin accessibility at single-cell resolution, single-cell sequencing assay for transposase accessible chromatin (scATAC-seq) has been developed, revealing the epigenomic landscape. At the functional level, single-cell RNA-sequencing (scRNA-seq) technologies provide a precise cell type annotation of complex cellular samples from primary samples due to the recent generation of single-cell transcriptome atlases comprising normal and pathological samples from humans and mice (Tabula Muris Consortium et al., 2018; Regev et al., 2017).

In the present research work, I would focus on scRNA-seq. Despite the existence of scDNA-seq and scATAC-seq strategies, scRNA-seq technologies are the most developed ones and have provided more insight into cell types and states, allowing a deeper understanding of gene regulatory mechanisms. Regarding cancer research, scRNA-seq experiments have revealed the presence of multiple cell populations within the tumors belonging to different molecular groups according to the actual molecular classifications (Roerink et al., 2018; Puram et al., 2017). The presence of this functional heterogeneity inside a tumor might prevent, or at least bias, the tumor molecular classification from a single biopsy. Moreover, scRNA-seq technologies offer a great opportunity to identify and study those populations that are supposed to be present in very low numbers, in a quiescent or dormant state,

and to design more specific anticancer treatments (Winterhoff et al., 2017). Thus, scRNA-seq experiments have shown the presence of populations with stem-like and treatment-resistance properties in oligodendroglioma and melanoma (Tirosh, Izar, et al., 2016; Tirosh, Venteicher, et al., 2016). Additionally, scRNA-seq technologies could facilitate the correct characterization of the complex relationship between the great diversity of immune cells inside a tumor, which will determine an overall tolerant or non-tolerant environment. Tumor neoantigens were also successfully identified by single-cell characterization of the T-cell receptor (TCR) repertoire, which might be useful in the diagnosis and treatment of cancer (Guo et al., 2018). Finally, scRNA-seq experiments have revealed the presence of heterogeneous circulating tumor cells (CTCs) populations with both epithelial and mesenchymal markers, highlighting that isolation methods based on epithelial markers are not adequate to capture all CTCs (Ting et al., 2014; Miyamoto et al., 2015). The high throughput of scRNA-seq technologies offers an opportunity to reduce the extensive purification required in previous sequencing strategies. A recent study on prostate cancer CTCs identified the activation of the non-canonical Wnt signaling pathway anticipating the appearance of drug resistance (Miyamoto et al., 2015). This observation supports the inclusion of the study of CTCs in the therapeutic decision-making in oncological practice. In summary, all this new data generated from scRNA-seq experiments may have direct clinical implications, enabling a better diagnosis and treatment of cancer patients.

A major limitation in the application of scRNA-seq technologies to primary solid tumor samples is the requirement of dissociation protocols to obtain viable individualized fresh cells. Several studies warn against the potential transcriptional changes arising from tissue manipulation during the time between sample collection and processing (Tung et al., 2017). Some authors have avoided this limitation by working with cell lines or organoids. Although they provided useful information, they failed to mimic complex interactions between cancer cells and the

microenvironment. Additionally, to understand the molecular basis of tumor evolution, it will be preferable to obtain several samples, or even serial samples, from the same patient, which is not straightforward in solid tumors. recent techniques to extract biopsies in a minimally invasive way, such as fine-needle aspiration (FNA), are not useful for traditional genomic analysis because little material is obtained, but they are useful for analysis with scRNA sequencing technologies in clinical research (Baslan & Hicks, 2017). Fortunately, many platforms are compatible with specific protocols of cell fixation and storage. Transcriptomic programs obtained from these cells seem very similar to those of freshly processed cells (Alles et al., 2017; Guillaumet-Adkins et al., 2017). If these protocols are optimized for all applications and broadly set up, we could anticipate a great increase in the application of scRNA-seq technologies in clinical research where coordinating sample collection and processing is not easy. Moreover, the recent development of scRNA-seq strategies on isolated single nuclei sometimes obtained even from paraffin-embedded (FFPE) material, removes the need to obtain viable cells and facilitates the study of long-stored samples (Grindberg et al., 2013; Martelotto et al., 2017).

### **Development of single-cell RNA sequencing technologies**

In just a few years since the emergence of scRNA-seq technologies, the ability to obtain single-cell expression profiles increased from very few cells to thousands in a single experiment (Tang et al., 2009; Zheng et al., 2017). The initial protocols used micromanipulation techniques to isolate the cells. This approach was laborious, expensive, and required a cDNA amplification step that produced a bias in the data which was also characterized by low throughput (Xue et al., 2013; Grindberg et al., 2013). The posterior introduction of nunique molecular identifiers (UMIs), which are random sequences that label individual molecules, significantly removed the amplification bias (Islam et al., 2014). Further developments in STRT-seq or CEL-seq protocols, added an individual barcoding step on isolated cells followed by a



single retrotranscription reaction reducing batch artifacts (Islam et al., 2011; Hashimshony et al., 2012). In 2015, microfluidic devices were introduced to isolate the cells in Drop-seq or InDrop protocols, enabling the processing of thousands of cells at once for a fraction of the cost of the traditional methods (Macosko et al., 2015; A. M. Klein et al., 2015). Following this strategy, 10x Genomics automated equipment characterized recently 1.3 million cells at the single-cell level (Zheng et al., 2017). Unfortunately, microfluidic-based methods are not efficient in the removal of the abundant ribosomal RNA. Unfortunately, microfluidic-based methods do not efficiently remove the abundant ribosomal RNA, so they sequence the end of poly-A-tailed RNAs through the use of poly-T oligonucleotides. This is very useful to generate expression profiles in this group of RNAs but do not provide comprehensive cell transcriptomic information. Other strategies, such as Split-seq and Sci-seq, avoid the physical isolation of the cells. Instead, they use a combinatorial barcode system to tag more than 100,000 single-cell transcriptomes in an individual and unique manner (Cao et al., 2017; Rosenberg et al., 2018). These techniques do not require any expensive microfluidic infrastructure because the cells themselves are fixed and used as 'containers'/nanochamber reactors. Finally, single-cell multi-omics approaches that allow the study of genetic, epigenetic, and transcriptomic profiles in the same cell have been developed, enabling a much deeper characterization of the tumor ecosystem (Stuart & Satija, 2019; Hou et al., 2016).

### **Single-cell RNA sequencing data analysis**

scRNA-seq data analysis is a great challenge. Due to the great variety of sequencing strategies and biological questions, there are many different reported analysis workflows, some of them already publicly available for researchers with limited bioinformatics resources (Tung et al., 2017; Stegle et al., 2015). In general, the computational workflow can be split into two stages: data processing and data analysis. Data processing involves the manipulation of the sequencing data from a fastq or bcl file to a data expression matrix (DGE) of single-cell gene expression

values. This first stage includes sequence quality control (QC) and poly(A) trimming to discard the low-quality cells, read alignment to the reference genome for transcript identification, deduplication to remove duplicate copies of redundant data which significantly decreases storage capacity requirements, de-multiplexing of barcodes and UMIs, and digital quantification of single-cell gene expression levels. The DGE data frame is used as an input for the second stage which is the most variable in methods as it greatly depends on the nature of the sample and the biological questions. However, most computational pipelines include tools for data normalization and QC as the first steps. Afterward, most pipelines perform a normalization step to select the most relevant features, remove uninformative genes, and reduce the number of dimensions to perform the downstream analysis. Next, a linear transformation ('scaling' the data) is applied prior to dimensional reduction techniques such as principal components analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE) (Maaten & Hinton, 2008), or preferable the uniform manifold approximation and projection (UMAP) approach which is able to represent the complex data structure with a lower computation time than the others (Becht et al., 2018), and lastly, the cell clustering. Depending on the potential applications, additional steps are usually required including subpopulation identification, differential gene expression analysis, functional signatures identification, pseudo-timing modeling, cell trajectory inference, or network reconstruction. An additional level of complexity in the computational methods consists of overlaying multi-omics data from the same biological entity. This approach is in increasing demand and improves the possibilities for a comprehensive cell characterization (Valdes-Mora et al., 2018).





# Objectives



Intratumor heterogeneity is postulated to play a critical role in tumor progression, metastasis spread and treatment resistance. In this Doctoral Thesis, we proposed to combine the use of mouse models, an innovative genetic cell lineage tracking system based on fluorescent proteins, and next-generation and single-cell sequencing technologies to further characterize the role of intratumor heterogeneity in pancreatic cancer progression and metastasis. To this end, we proposed the following specific objectives:

- I. To analyze the tumor clonality in early and late stages using fluorescent lineage tracing tools.
- II. To generate new lineage tracing tools to study the involvement of intratumor heterogeneity in cancer progression.
- III. To study of the role of genetic intratumor heterogeneity in pancreatic cancer progression.
- IV. To characterize the presence of transcriptomic intratumor heterogeneity and its role in metastasis.





# Experimental procedures



## Molecular Biology methods

### Molecular cloning methods

Different genetic constructs were generated to test several fluorescent proteins. Specific primers were designed to amplify the coding regions of multiple fluorescent proteins from commercial vectors acquired in *Addgene* repository (**Table 1**). Other primers were designed to contain in their tail the sequences of restriction sites to allow their insertion by cohesive digestion and ligation into the desired digested expression vector (pCDNA3.1-fWPRE) (**Table 2**). This vector was previously prepared to contain a farnesylation signal to direct the expression of the protein to the cell membrane and the RNA stabilization signal from the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element. Typically, 15-20 units of each appropriate restriction enzyme or combination were used to digest either vectors or PCR products for 1h at 37°C. The digested PCR products were purified with Thermo Scientific GeneJET PCR Purification Kit (#K0702). Digested vectors were purified by gel excision and the Thermo Scientific GeneJET Gel Extraction Kit (#K0692).

**Table 1.** Name, use, and reference of the vectors acquired in *Addgene*.

Vector	Use	Reference
pThy1-Brainbow-3.2	Plasmid used to amplify mKate2 and mOrange2	#45179
pCAG-YFP	Plasmid used to amplify YFP	#11180
YPet-C1	Plasmid used to amplify YPet	#54648
Citrine-C1	Plasmid used to amplify Citrine	#54715
AAV-EF1a-BbTagBY	Plasmid used to amplify Tag-BFP	#45185
LI D-E T-Saph + linker c1	Plasmid used to amplify t-Sapphire	#54412
Sirius/pCDNA3	Plasmid used to amplify Sirius	51957

**Table 2.** Primers designed to amplify the coding sequences of interest from the original vectors with restriction sites HindIII-BamHI to clone in the pCDNA3.1-fWPRE vector.

Fluorescent protein	Primer names	Primer sequences
<b>mOrange2</b>	HindIII-mOrange2-F	5'-ACGTAAGCTTGCCGTGCTAGCATAACTTCG-3'
	BglIII-mOrange2-R	5'-ACGTAGATCTCAGGGTCAGCTTGCCGTAG-3'
<b>t-Sapphire</b>	tSapphire-F-HindIII	5'-ACTGAAGCTTATGGTGAGCAAGGGCGAGGAGC TGTT-3'
	tSapphire-R-BglIII	5'-ACTGAGATCTCTGTACAGCTCGTCCATGC-3'
<b>mKate2</b>	HindIII-mKate2-F	5'-ACGTAAGCTTCGCTATGGTGAGCGAGCTGAT-3'
	BglIII-mKate2-R	5'-ACGTAGATCTTGCACTTCTGTGCCCCAGTT-3'
<b>Tag-BFP</b>	TagBFP-R-BglIII	5'-ACTGAGATCTCTGTGCACTTCTGTGCCCCAGT-3'
	TagBFP-F-HindIII	5'-ACTGAAGCTTGCTCTCTGATACCGTTCGT-3'
<b>YPet</b>	YPet-F-HindIII	5'-ACTGAAGCTTATGGTGAGCAAAGGCGAAG-3'
	YPet-R-BglIII	5'-ACTGAGATCTCTATAGAGCTCGTTCATGCCCT-3'

As a general rule, 1 unit of T4 DNA Ligase (Thermo Scientific #EL0014) was used to ligate the digested vector and the digested insert. In general, approximately 200 ng of total DNA was used in the ligation reaction, with between 5 to 10 molecules of insert for each molecule of the vector. The ligation reaction was incubated at 22°C for 1h. After ligation, the ligation mix is desalted for 20 minutes. 2µl of microdialyzed ligation reaction were transformed by electroporation in DH5α competent cells. They, they were recuperated for 1h at 37°C in LB without antibiotic and plated with ampicillin which is, in this case, the correct antibiotic according to the resistance gene contained in the vector. We analyzed the transformants by PCR directly on some cells taken from the colony or by extracting the DNA (Thermo Scientific GeneJET Plasmid Miniprep Kit #K0503) and performing restriction analysis. Finally, all potential constructs were sequenced by the company STAB VIDA (Caparica, Portugal).

### Sanger sequencing

We validated the fluorescent protein constructs by Sanger sequencing. We first performed PCR amplification of the region of interest by using oligonucleotides in Table 2. Each PCR reaction was then subjected to Sanger sequencing analysis. Sequencing reactions were performed by the company STAB VIDA (Caparica, Portugal) and the resulting electropherograms were visualized using SnapGene Viewer.

### Fluorescent proteins evaluation

To analyze the cells by confocal microscopy, we plated the cells over a glass coverslip previously washed with ethanol and coated with poly-d-lysine to promote the attachment of the cells. In the case of FACS analysis, cells were trypsinized, washed, and resuspended in PBS. The maximum signal is reached 48h after transfection. To evaluate the fluorescent proteins, we used a fluorescent confocal microscopy Leica TCS SP5 with five lasers of 405 nm, 488 nm, 514 nm, 543 nm, and 549 nm of wavelength. Different excitation windows were designed in order to recover uniquely as much light as possible for each of the individual proteins. To separate the fluorescent proteins, we used a FACS Aria III system (fluorescence-activated cell sorting) from Becton Dickinson with the following configuration: 407 nm laser (filters: 450/50, 610/20, 710/50, 780/60, 660/20, 525/50), 488 nm laser (filters: 488/10, 695/40, 530/60), 561nm laser (filters: 610/20, 710/50, 780/60, 670/14 and 582/15) and 633 nm laser (filters: 670/30, 780 / 60, 730/45). Laser and filter combination used to register the signals for the FPs are summarized in **Table 3**.

**Table 3.** Laser and filters combination of different FPs for the confocal microscopy and the sorter. Emission ranges refer to the detection filters used to detect each protein.

Fluorescent protein	Confocal microscopy		Sorter/FACS	
	Excitation	Emission	Excitation	Emission
Tag-BFP/Sirius	405 nm	445/25 nm	407 nm	405/50 nm
EGFP	488 nm	505/10 nm	488 nm	530/30 nm
mOrange2	543 nm	575/15 nm	561 nm	582/15 nm
YFP/Citrine/YPet/Phi-YFP	514 nm	535/10 nm	670 nm	530/30 nm
mKate2	594 nm	670/30 nm	505 nm	670/30 nm
tSapphire	405 nm	505/10 nm	407 nm	527/50 nm

### Electrocompetent cells

For the generation of electrocompetent cells, we grow DH5 $\alpha$  *Escherichia coli* strain in LB until it reaches an optical density at 600 nm (OD<sub>600</sub>) between 0.4 and 0.6. Subsequently, we performed two washes with cold water and two washes with a cold 10 % glycerol solution keeping the cells at 4 °C during the whole procedure. Finally, cells were snap-frozen in dry ice and kept at -80 °C until used.

### DNA genotyping

Genomic DNA was isolated from mouse tail biopsies using alkaline lysis buffer (NaOH 25 mM, EDTA 0.2 mM, pH=8) followed by 99 °C incubation for 90 min and posterior neutralization (Tris 40 mM, pH=7.4). PCR was performed using DreamTaq DNA polymerase (ThermoFisher) in all the cases and using the oligonucleotides summarized in **Table 4**. *Confetti* allele was genotyped under the following conditions: denaturation at 95 °C for 30 sec, annealing at 60 °C for 30 sec, and extension at 72 °C for 30 sec, with 40 cycles of amplification. The size for the mutant band was 300 bp and 386 bp for the wild-type band. *Cre* allele genotyping

was performed under the following conditions: denaturation at 95 °C for 30 sec, annealing at 60 °C for 30 sec, and extension at 72 °C for 30 sec, with 35 cycles of amplification. The size for the wild-type band was 226 bp and 316 bp for the mutant allele. *Kras* allele was genotyped under the following conditions: denaturation at 95 °C for 30 sec, annealing at 55 °C for 40 sec, and extension at 72 °C for 40 sec, with 40 cycles of amplification. The size for the mutant band was 351 bp and 270 bp for the wild-type band. *Flippase* allele genotyping was performed under the following conditions: denaturation at 95 °C for 30 sec, annealing at 56 °C for 50 sec, and extension at 72 °C for 50 sec, with 40 cycles of amplification. The size for the wild-type band was 300 bp and 620 bp for the mutant allele. *Tpr53* allele genotyping was performed under the following conditions: denaturation at 95 °C for 30 sec, annealing at 60 °C for 40 sec, and extension at 72 °C for 40 sec, with 35 cycles of amplification. The size for the wild-type band was 258 bp and 292 bp for the mutant allele.

**Table 4.** Primers sequences to genotype our PDAC mice.

Allele	Primer names	Primer sequences
<i>Confetti</i>	11341	5'-GAATTAATTCCGGTATAACTTCG-3'
	oIMR8545	5'-AAAGTCGCTCTGAGTTGTTAT-3'
	oIMR8916	5'-CCAGATGACTACCTATCCTC-3'
<i>Cre</i>	ROSA26-3ARM-Fwd	5'-CCCGACAAAACCGAAAATCTG-3'
	ROSA26-WT-Rev	5'-GGAGTTCTCTGCTGCCTCCT-3'
	R26-Dual/CRE-Rev	5'-GGCTGCAAACAGCTAATGCAC-3'
<i>Kras</i>	Kras-Common-Fwd	5'-CACCAGCTTCGGCTTCCTATT-3
	Kras-WT-Rev	5'-AGCTAATGGCTCTCAAAGGAATGTA-3'
	Kras-FSF-MUT-Rev	5'-GCGAAGAGTTTGTCTCAACC-3'
<i>Flippase</i>	Pdx1-Flp-Fwd	5'-AGAGAGAAAATTGAAACAAGTGCAGGT-3'
	Flp-Rev	5'-CGTTGTAAGGGATGATGGTGAAC-3'
	Gabra-Fwd	5'-AACACACACTGGAGGACTGGCTAGG-3'
	Gabra-Rev	5'-CAATGGTAGGCTCACTCTGGGAGATGATA-3'

<b><i>Trp53</i></b>	p53-frt1	5'-CAAGAGAACTGTGCCTAAGAG-3'
	p53-frt2	5'-CTTTCTAACAGCAAAGGCAAGC-3'

### **DNA extraction**

DNA was extracted from fresh frozen tissues or cell lines using the Agencourt DNAdvance Beckman Coulter kit (#A48705, Beckman Coulter, Brea, CA, USA), following the manufacturer's instructions. After extraction, samples were quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and evaluated for purity (260/280 nm ratio).

### **Exome-DNA libraries generation**

To generate whole-exome DNA libraries, SureSelect<sup>XT</sup> Mouse All Exome kit (Agilent Technologies) was used following the SureSect<sup>XT</sup> Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library protocol (Agilent Technologies). For all cleaning steps, we used Agencourt AMPure XP (Beckman Coulter, #082A63881), following the manufacturer's protocol. Size distribution of the library was measured with the 4200 TapeStation using DNA 1000 kit or D1000 ScreenTape Assay (Agilent Technologies). DNA was quantified using the Qubit dsDNA BR Assay (Life Technologies, #Q32851) a qPCR reaction with primers designed to target the Illumina adapters. Massively parallel sequencing was carried out in an Illumina sequencing machine with a 100 bp paired end (PE) protocol aiming at a 90x average coverage depth.



### **Somatic mutation orthogonal validation**

To confirm the presence of the mutations found in the whole-exome sequencing experiments, we designed to amplify a DNA fragment around each identified mutation, and this DNA fragment was amplified and sequenced at high depth in all the samples extracted from the same mouse. DNA was extracted from fresh frozen tissues as has been previously indicated, and quantified using the Qubit dsDNA BR Assay (Life Technologies, #Q32851). All DNA fragments amplified from the same sample were mixed before performing DNA libraries. For all cleaning steps, we used Agencourt AMPure XP (Beckman Coulter, #082A63881), following the manufacturer's protocol. To prepare the sequencing libraries, we performed some enzymatic steps including: end-repair and adenylation (DNA Rapid End Repair module, NEXTFLEX), generation of the paired-end adaptor by hybridizing two phosphorylated complementary synthetic oligonucleotides, ligation of the PE adaptor with a T4 DNA Ligase (Thermo Fisher Scientific, #EP0062), and PCR indexing amplification (Phusion high fidelity DNA polymerase (Thermo Fisher Scientific, #F530L). Purity of the libraries were tested using the Nanodrop (260/280 nm ratio), size distribution was measured with the 4200 TapeStation using DNA 1000 kit or D1000 ScreenTape Assay (Agilent Technologies), and quantified using the Qubit dsDNA BR Assay (Life Technologies, #Q32851) and a qPCR reaction with primers designed to target the Illumina adapters. Massively parallel sequencing was carried out in an Illumina sequencing with a 100 bp paired end (PE) protocol and a 10.000x average coverage.

### **Single-cell RNA libraries generation**

To perform single-cell capture and barcoding of thousands of single cells from the mouse primary tumor and metastases, we used the BD Rhapsody™ Single-Cell Analysis System which includes the Scanner and the Express Single-Cell Analysis

System. Both of them were used following the manufacturer's user guides (BD Rhapsody, Doc ID 214062, Doc ID 214063).

To create single-cell whole transcriptome mRNA libraries up to 10.000 cells after cell capture, we used a 3' whole transcriptome analysis (WTA) approach, following the Whole transcriptome analysis alpha protocol (BD Rhapsody). In summary, the cDNA of mRNA targets is first encoded on the magnetic capture beads during reverse transcription. Then, a random priming approach is performed, followed by an index PCR step.

To generate the Sample Tag sequencing libraries, after labelling the cells with the Sample Tags, we used the mRNA Whole Transcriptome Analysis (WTA) and Sample Tag Library Preparation protocol (BD Rhapsody, #23-21712-00). This protocol was used to combine several samples in the same cartridge (2305\_1Tumor and 2305\_2Lung; 2849\_1Total, 2849\_1Liver1, 2849\_3Liver2, and 2849\_4Lung). The extended Sample Tags were first denatured from the BD Rhapsody Cell Capture Beads. Later, they were amplified with a series of PCR steps. Meanwhile, the whole transcriptome amplification library is generated directly from the BD Rhapsody Cell Capture Beads using a random priming approach and followed by an index PCR step. Both the whole transcriptome mRNA and Sample Tag libraries were combined for sequencing on an Illumina sequencer generating an average of 40.000 reads per cell.

### **RNA isolation and quantitative Reverse-Transcriptase PCR**

Total RNA was isolated and purified using Extract Me Total RNA Kit (Blirt, DNA Gdansk, Poland) according to the manufacturer's instructions. RNA quality was measured using RNA ScreenTape® (4200 TapeStation Instrument – Agilent Genomics). Reverse transcription was performed using the Takara PrimeScript cDNA Synthesis kit (Takara Bio, Inc., Dalian, Japan) according to the manufacturer's

instructions. mRNA expression was measured in triplicate for each sample by quantitative reverse-transcriptase PCR (qRT-PCR) analysis, using Luminaris Color HiGreen qPCR Master Mix (Thermo Scientific) with StepOnePlus™ real-time PCR system (Applied Biosystems, Foster City, CA). As an internal control, gene expression was normalized to the  $\beta$ -actin gene (used as a housekeeping gene). The  $\Delta\Delta C_t$  method was used for quantification and comparison. A list of the primers used for the qRT-PCR experiments can be found in **Table 5. Supplementary information** including the results of the qRT-PCR can be found in the following link: [https://unican-my.sharepoint.com/:f:/g/personal/varelaei\\_unican\\_es/EuBHR9bq21tEqbjbR4q0XiwBd6G0uSSgAXt4YcKqrN4GAQ?e=Y1rcsv](https://unican-my.sharepoint.com/:f:/g/personal/varelaei_unican_es/EuBHR9bq21tEqbjbR4q0XiwBd6G0uSSgAXt4YcKqrN4GAQ?e=Y1rcsv).

**Table 5.** Sequences of the primers to amplify around 110 bp of the coding sequences of the genes of interest.

Gene	Primers	Primer sequences
<b><i>Cd74</i></b>	Cd74_Fwd	TGACCAACGCGACCTCATCT
	Cd74_Rev	CCAGGACAGAGACACCGGTG
<b><i>Cxcl2</i></b>	Cxcl2_Fwd	TTGCCTTGACCCTGAAGCCC
	Cxcl2_Rev	CGTTGAGGGACAGCAGCC
<b><i>Il1b</i></b>	Il1b_Fwd	CTCGTGCTGTCGGACCCATA
	Il1b_Rev	ATTTTGTCGTTGCTTGGTTCTCC
<b><i>Srgn</i></b>	Srgn_Fwd	CCTTCGTCCTGGTTTGGGGA
	Srgn_Rev	GGTCCCTTCTCCTCGATGCA
<b><i>Ctss</i></b>	Ctss_Fwd	GCATCGAGGAGAAGGGACCA
	Ctss_Rev	TTTCCCAGATGAGACGCCGT
<b><i>H2.AB1</i></b>	H2.AB1_Fwd	TGGAGGCTCAGTGATATGGTGC
	H2.AB1_Rev	TCTCGGTACCTCTGCGTTGG
<b><i>Lyz2</i></b>	Lyz2_Fwd	GAGCTGTGAATGCCTGTGGG
	Lyz2_Rev	CATGCTCGAATGCCTTGGGG
<b><i>Ccl6</i></b>	Ccl6_Fwd	GGATGAGAACTCCAAGACTGCC
	Ccl6_Rev	GGAGGGTTATAGCGACGATCTTC

## **Cell Biology methods**

### **Cell culture**

HEK-293T, HeLa, and mouse pancreatic primary cells were routinely maintained in Dulbecco's modified Eagle's medium (DMEM) DMEM (Lonza, Verviers, Belgium), supplemented with 10% FBS (HyClone Victoria, Australia), 1% Gentamycin and 1% Ciprofloxacin at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub>.

### **Isolation of primary mouse PDAC cell cultures and obtention of single-cell suspensions**

Cell lines of primary pancreatic tumors from our mouse model were obtained by cutting into small pieces the primary tumor, which was mechanically and enzymatically digested using the mouse tumor dissociation kit (Miltenyi Biotec Tumor Dissociation kit, mouse, #130-096-730). After, tumor cells from mouse tumors were enriched with a tumor cell isolation kit (Miltenyi Biotec Tumor cell isolation kit, mouse, #130-110-187), following the manufacturer's instructions. The same protocol was used to obtain single-cell suspensions from the primary tumor and metastases of our mouse model.

### **Mycoplasma PCR**

For testing mycoplasma contamination, cells were cultured for 3 days to 100% confluency in DMEM supplemented with 10% FBS without Gentamycin and Ciprofloxacin. 1 ml of cell culture supernatant was harvested in a 1.5 ml Eppendorf tube and stored at -20°C. Harvested cell culture supernatant was defrosted on ice and 1 µl of added to 10 µl of a MasterMix containing 5.5 µl of 2x KAPA Genotyping

Mix (Sigma-Aldrich), 0.22  $\mu$ l of primer mix (containing 5  $\mu$ M of each primer; 7x forward plus 3x reverse primer) and 4.28  $\mu$ l ddH<sub>2</sub>O was prepared for each PCR reaction (resulting in a total volume of 10  $\mu$ l MasterMix per sample). PCR reaction was performed with the following conditions: 98 °C for 180 sec; 65 °C for 60 sec; 72 °C for 60 sec; 35 cycles of amplification at 98°C for 15 sec, 65°C for 30 sec, and 72 °C for 20 sec; a final extension at 72 °C for 300 sec; and holding at 16 °C. PCR reaction was analyzed on a 1.5% agarose gel (the presence of a PCR band at 500 bp indicates mycoplasma contamination).

### **DNA transfection into eukaryotic cell lines**

Experiments of fluorescent proteins expression were done in the human embrionic kidney 293T cell line (HEK-293T). DNA was introduced into the host cells by transfection with polyethylenimine (PEI) (Sonawane et al., 2003). We used 1'5mg of DNA and 6 $\mu$ l of PEI to transfect the cells in 6-well plates (ratio DNA:PEI 1:4). For best transfection efficiency, cells were transfected at 60% of confluence.

### **Immunofluorescence**

For immunofluorescence analysis, HeLa cells were washed with PBS 1X and fixed in 4% formaldehyde solution for 15 min at room temperature. Then, cells were rinsed three times with PBS 1X, permeabilized using 0.5% Triton X-100 for 5 min at room temperature, and blocked with 3% BSA diluted in PBS plus 0.5% Triton X-100 (PBT) for 1 h at room temperature. Incubation with diluted primary antibodies was performed for 1 h at room temperature in a humidity chamber. Primary antibodies are specified in **Table 6Table 4**. Then, cells were washed three times with PBT 1X for 5 min and secondary antibody incubation was carried out for 1 h at room temperature in a humidity chamber. 3xHA-fmOrange2 and Myc-fmKate2 were

incubated with a donkey anti-mouse IgG Alexa 488 secondary antibody (Thermo, #A21202); 6xHis-ftSapphire, 3xFLAG-fTagBFP, and fYPet-V5 were incubated with a donkey anti-mouse IgG Alexa 647 secondary antibody (Thermo, #A31571). After cells were washed three times with PBT 1X for 5 min and coverslips were mounted with Prolong™ Gold Antifade Mountant with DAPI (Invitrogen, Thermo Scientific).

**Table 6.** Primary antibodies used in the immunofluorescences.

Tag-FP	Primary antibody	Used dilution	Source (catalog number)
<b>3xHA-fmOrange2</b>	Mouse anti-HA IgG2a	1:50	Santa Cruz Biotechnology (#sc-7392/F7)
<b>6xHis-ftSapphire</b>	Mouse anti-His IgG1	1:50	Santa Cruz Biotechnology (#sc-8036)
<b>Myc-fmKate2</b>	Anti-cMyc Monoclonal	1:100	OriGene (#TA150121)
<b>3xFLAG-fTagBFP</b>	Mouse anti-FLAG M2 IgG1	1:500	Sigma (#1804)
<b>fYPet-V5</b>	Mouse anti-V5 IgG2a	1:250	BIO-RAD (#MCA1360GA)

### Fixation and embedding fresh Confetti mouse tissues

To preserve the fluorescent colors of Confetti, it is essential to store the tissues preserving the fluorescence and tissue morphology. Tissues were fixed in formalin 4% for 2 h at room temperature, washed twice with PBS 1X, transferred to a 15% (w/v) sucrose solution using PBS until the tissue sinks, and transferred to a 30% (w/v) sucrose solution using PBS until it sinks. At this point, the permeation was complete. Then, the tissue was embedded in OCT and frozen at -80°C until ready for sectioning. To process the samples, the frozen tissue block was transferred to a cryotome cryostat and both were allowed to equilibrate to -20 °C. Cuts of 14 µm were made on SuperFrost Plus slides. To fix the frozen sections, slides were immersed in acetone for 10 min at room temperature and washed twice with PBS 1X prior to dispensing a small drop volume (25 µl per 22 mm x 22 mm coverslip) of

the VETASHIELD mounting medium onto the sample for preserving fluorescence. A coverslip was placed over the sample and the VECTASHIELD was allowed to disperse throughout the section before observing it in the Confocal Microscopy.

### **Cell preparation and staining for the FACS experiment**

To ensure a negative and positive control of the experiment, 150.000 HEK-293T cells were seeded in each well of a 6-well plate 72h before the FACS experiment. To obtain the positive control, the cells from 2 wells were transfected 48h before the FACS experiment with a pCDNA3.1-HisA-Cd74 construct (obtained by cloning the *Cd74* gene in pCDNA3.1-HisA expression vector). To prepare the single-cell suspensions, the cells from the mouse primary cell lines were trypsinized, washed twice with PBA 1X, and resuspend in 200  $\mu$ l of FACS buffer (previously prepare as follows: 20 ml of fetal calf serum, 10 ml of Na<sub>3</sub>N 10%, 100 ml of PBS 10X, 870 ml of H<sub>2</sub>O).

To perform the cell staining, the cell suspension was incubated with 1  $\mu$ l of the blocker (only the tubes that will be posteriorly stained) for 10 min at 4 °C, centrifuge at 1500 rpm for 5 min at 4 °C, and the supernatant was discarded, the cells were resuspended in 200-500  $\mu$ l of cold-FACS buffer, incubated 3 min at 4 min, and centrifugated at 1500 rpm for 5 min at 4 °C, the supernatant was discarded, cells were resuspended in 100  $\mu$ l of the antibody mixture (mouse CD74 APC-conjugated antibody, monoclonal rat IgG1, R&D systems #FAB7478A; 1:20 dilution in a final volume of 100  $\mu$ l per sample), incubated for 30 min at 4 °C and dark, and centrifugated at 1500 rpm for 5 min at 4 °C, the supernatant was discarded, cells were washed twice with FACS buffer, resuspended in an appropriate volume FACS buffer, and transferred to 5 ml tube by passing through a nylon mesh (or a filter of 30  $\mu$ m of diameter) before performing the FACS sorting. Then, cells were isolated by FACS based on Cd74 expression.

## Animal model procedures

### Animal care

All animal procedures were approved and performed in accordance with the guidelines of the Committee of Animal Experimentation of the University of Cantabria. Animals were housed in a pathogen-free facility under a photoperiod of 12 h light/12 h dark,  $22 \pm 2$  °C of temperature,  $50 \pm 10\%$  of relative humidity and *ad libitum* access to water and food.

### Experimental PDAC mouse model and PDAC induction

We acquired a well-characterized PDAC mouse model developed by Dr. Dieter Saur (Schönhuber et al., 2014).  $Tg^{Pdx1-Flp} : Kras^{FSF-G12D/+} : R26^{FSF-CreE2T2 /FSF-CreERT2}$  (KFC) mice were used as parental mice for colony maintenance. Experimental mice were obtained crossing KFC mice with 129Sv mice to generate a F1 heterozygous for all 129Sv SNPs in order to detect loss of heterozygosity (LOH).

For PDAC induction, pancreatitis was induced through the administration of seven hourly intraperitoneal injections of caerulein (50 µg/kg of body weight) in 2-month-old-mice on two consecutive days. When the mice show clear signs of disease (ascites jaundice or weight loss), they were sacrificed and a careful inspection of all organs was performed to detect pre-malignant lesions, primary tumors and metastases. Tail samples are taken as normal samples for further experiments.



### ***In vivo* metastasis protocol**

Unanaesthetised mice were warmed with a heat lamp for tail vein injections to allow venous dilation. Mice were then placed into a plastic retraining apparatus, and 500.000 tumor cells were injected via the lateral tail vein of C57BL/6 mice strain. After 3 months, or before if there were symptoms of disease, mice were sacrificed. Then, the lungs and livers were fixed embedded, and frozen until the fluorescence analysis, or alternatively, they were dissociated into single-cell suspension to perform single-cell RNA-seq libraries.

## **Data analysis**

### **DNA-sequencing data analysis**

Raw sequence data was quality controlled using FastQC. Then, they were aligned to the mouse genome from Ensembl Genome Reference Consortium mouse 38 (GRCm38) using BWA-MEM (H. Li & Durbin, 2009). Format transformation, sorting, and indexing of the bam files were done with SAMTools (Li et al., 2009). The alignment was fixed and cleaned, and PCR duplicates reads were marked and excluded from the analysis with Picard. Local realignment around indels was done with GATK (McKenna et al., 2010). We calculated the enrichment statistics and target coverage with Bedtools. Next, paired tumor and normal bam files were used to identify putative somatic single nucleotide variations (SNVs) using Realignment Assisted Minimum Evidence Spotter (RAMSES), which is an in-house written algorithm ) (Martínez et al., 2014). We selected those mutations with a confidence score higher than 2 and a mutational frequency higher than 0.05. PINDEL was used to detect small insertions and deletions (indels) that were reported with a minimum of 5 reads in the sample and being absent in the control DNA (Ye et al., 2009). Functional consequence of the mutations was annotated our own in-house written

software (MutAnn, unpublished), that uses the information from Ensembl GRCm38. Big rearrangements such as copy number alterations (CNA) and loss of heterozygosity were found with Control-FREEC (Boeva et al., 2012). **Supplementary information** including tables (such as the list of protein-coding mutations in the different mice) and figures can be found in the following link: [https://unican-my.sharepoint.com/:f:/g/personal/varelaei\\_unican\\_es/EuBHR9bq21tEqbjbR4q0XiwBd6G0uSSgAXt4YcKqrN4GAQ?e=Y1rcsv](https://unican-my.sharepoint.com/:f:/g/personal/varelaei_unican_es/EuBHR9bq21tEqbjbR4q0XiwBd6G0uSSgAXt4YcKqrN4GAQ?e=Y1rcsv).

### Single-cell RNA-sequencing data processing and data analysis

We followed the Seurat guided tutorial from Dr. Satija's lab ([https://github.com/satijalab/seurat/blob/master/vignettes/pbmc3k\\_tutorial.Rmd](https://github.com/satijalab/seurat/blob/master/vignettes/pbmc3k_tutorial.Rmd)). First, a quality control and poly(A)trimming of the fastq files is performed to discard the cells with low quality which are those with a higher percentage of mitochondrial reads than 20%. Next, genes that are present in less than three cells are also filter out. Then, we normalized the data, identified the highly variable features, scaled the data and performed linear dimensional reduction with the uniform manifold approximation and projection (UMAP) approach (Becht et al., 2018), and lastly, the cell clustering. Specific cellular subpopulations were identified by looking for the expression level of specific known markers, such as *Sox9* and *Krt18* for the tumor cells, *Ptprc/Cd45* for the immune cells, *Ctss* for the macrophages, *Col1a1* for the fibroblasts, *Acta2* for the myofibroblastic cancer-associated fibroblasts (CAFs), *Fabp4* for the lipofibroblasts, and *Cd74* for the antigen-presenting CAFs. Gene set enrichment analysis (GSEA) was used to do differential gene expression analysis (Subramanian et al., 2005). Moreover, Slingshot transcriptional trajectories was used to infer cell trajectories (Street et al., 2018). Lastly, InferCNV analysis was used to extract CNA from the single-cell RNA-seq data allowing to analyze the potential relationship between the genetic and the transcriptomic intratumor heterogeneity.

Supplementary information including tables (such as the list of the genes that are upregulated and downregulated in each cluster from Seurat analysis) and figures can be found in the following link: [https://unican-my.sharepoint.com/:f:/g/personal/varelaei\\_unican\\_es/EuBHR9bq21tEqbjbR4q0XiwBd6G0uSSgAXt4YcKqrN4GAQ?e=Y1rcsv](https://unican-my.sharepoint.com/:f:/g/personal/varelaei_unican_es/EuBHR9bq21tEqbjbR4q0XiwBd6G0uSSgAXt4YcKqrN4GAQ?e=Y1rcsv).



## Results

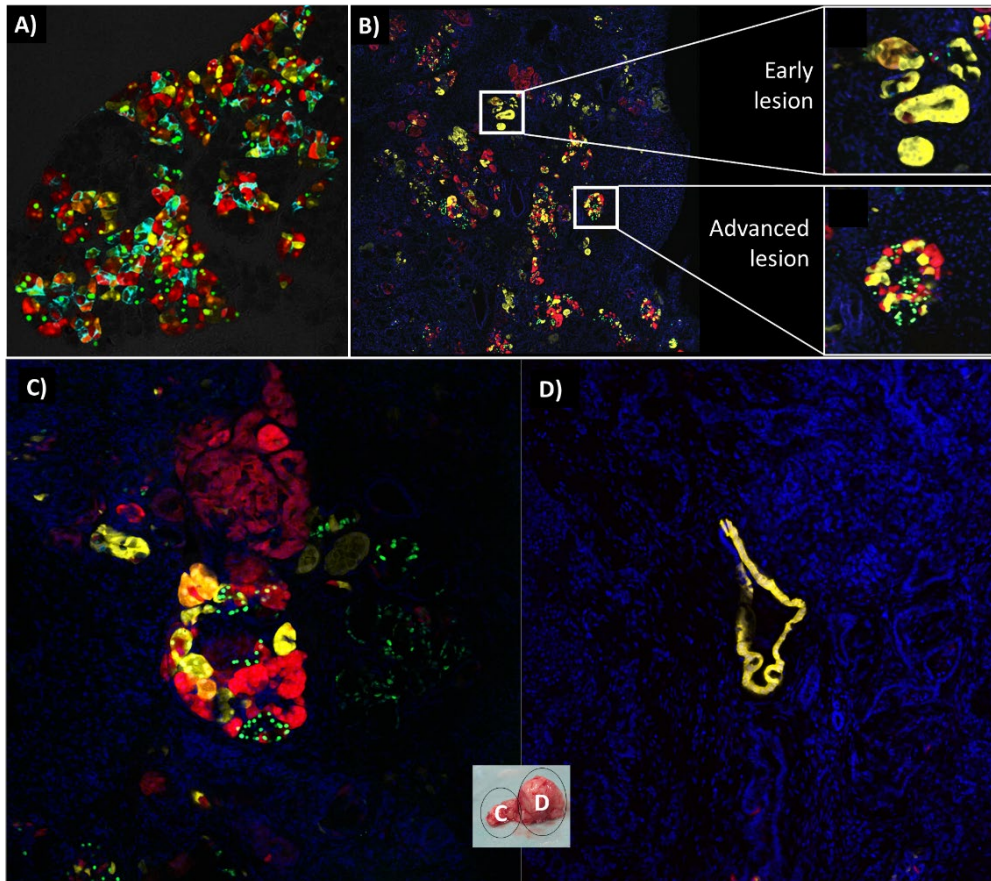


## I. To analyze the tumor clonality in early and late stages using fluorescent lineage tracing tools

To study PDAC intratumor heterogeneity (ITH) *in vivo*, we crossed a well-characterized mouse model of pancreatic cancer (Tg<sup>Pdx1-Flp</sup>:Kras<sup>FSF-G12D</sup>) (Schönhuber et al., 2014) with transgenic mice expressing a conditional and inducible allele of Cre (R26<sup>FSF-CreERT2</sup>). The resulting mice express a tamoxifen-dependent Cre recombinase in the same pancreatic cells expressing the oncogenic version of Kras. This approach allows an inducible dual/double-recombinase strategy that combines the flippase-FRT (Flp-FRT) and the Cre-loxP recombination technologies for tumor initiation and secondary genetic manipulation, respectively.

In order to study the genetic clonality in the early and late stages of murine pancreatic tumors, we took advantage of a multi-fluorescent protein lineage tracking mouse strain (Confetti mice) based on the Brainbow-2.1 allele (Scheper et al., 2012; Livet et al., 2007). First, to check the function of the Brainbow allele, we induced the recombination of the allele with tamoxifen. As expected, we clearly observed a proportional representation of the four fluorescent markers, demonstrating that the system perfectly works in our hands (**Figure 5A**). Next, we crossed our mouse model to the Confetti mice and induced the fluorescent labeling after pancreatitis induction. Similarly, in this context, we also observed the four markers, especially in early stage premalignant lesions of the pancreatic disease (**Figure 5B**). Interestingly, early-stage lesions are multiclonal, with multiple labels observed in each lesion (**Figure 5C**); but this clonality is progressively lost in more advanced lesions as evidenced by the presence of a single label (**Figure 5D**). Surprisingly, the labeling was systematically lost in the most advanced stages of the disease (**Figure 5D**). Recently, a collaborated work led by Dr. Roland Rad and Dr. Dieter Saur has demonstrated that *Kras* wt allele is frequently lost or silenced in advanced stages of mouse pancreatic cancer (Mueller et al., 2018). According to the chromosomal location of the *Rosa26* locus and our mouse crossing designs, the Brainbow-2.1 allele is located very close

to the *Kras* wt allele in our mice. So, a plausible explanation for the loss of the labeling could be the combined loss of both alleles together during the later stages of PDAC development.



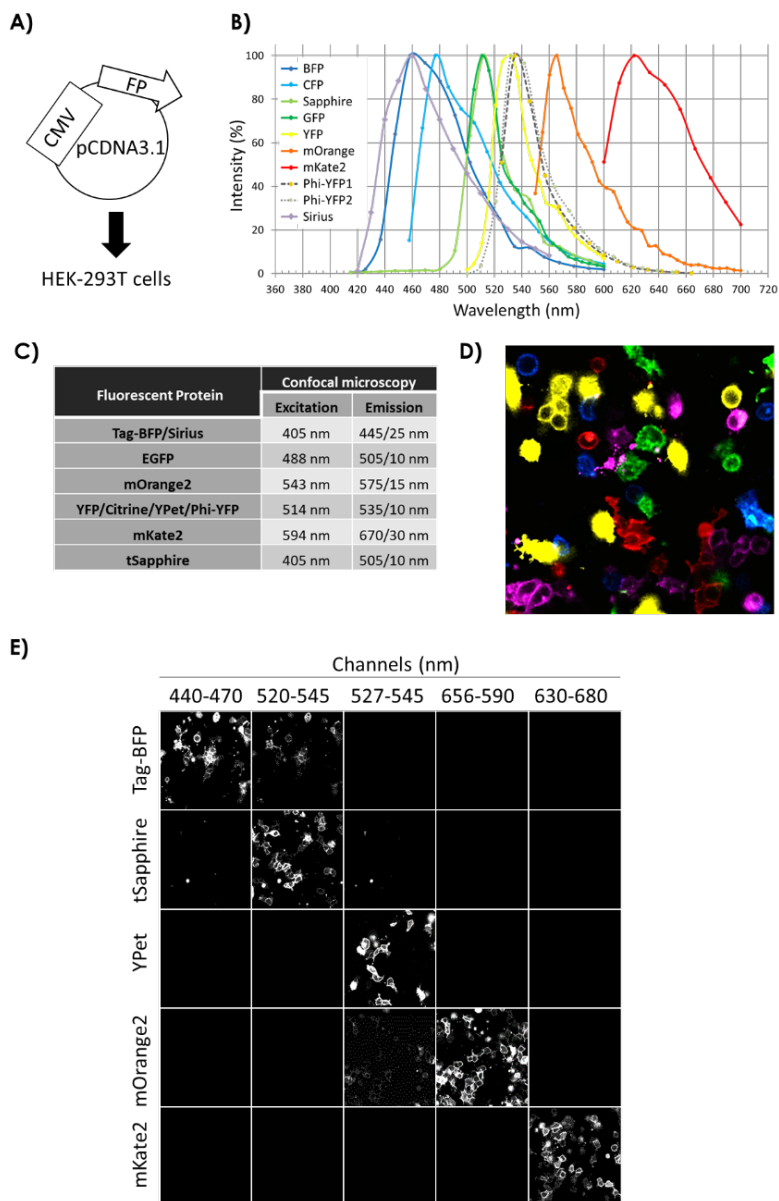
**Figure 5 | Representative pictures of the confetti allele activated in the pancreas. A)** Confetti labeling in the normal pancreas shows an even representation of the four potential labels of the Brainbow-2.1 allele. **B)** Confetti labeling in a pancreatic tumor reveals **(C)** the presence of all the fluorescent proteins in the premalignant lesions, in contrast to **(D)** the most advanced stages presenting a monofocal/clonal origin of the lesions (even with the loss of fluorescence).



## II. To generate new lineage tracing tools to study the involvement of intratumor heterogeneity in cancer progression

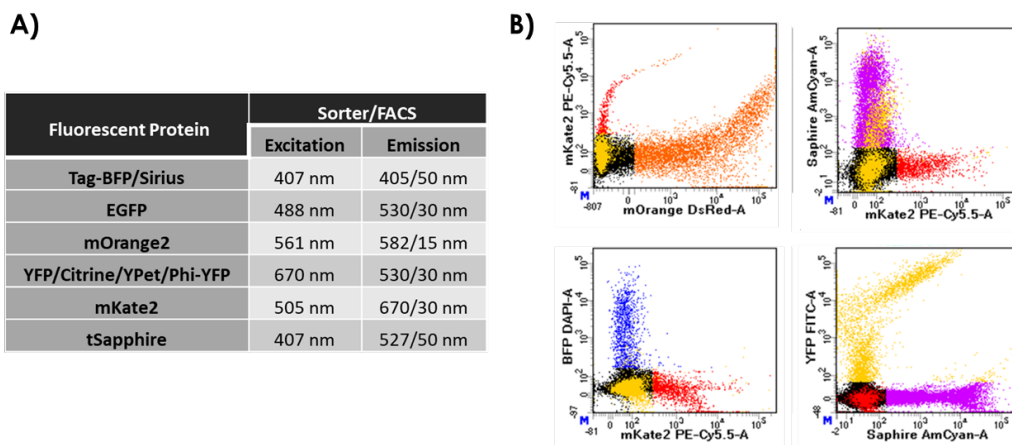
During the first experiments, it became evident that the original proteins selected for the *Brainbow-2.1* construct presented quite a lot of spectra overlapping which hampers the correct identification of the four proteins and the analysis of potential protein combinations. Being aware of this problem an improved construct (called *Brainbow-3.2*) was developed by the original authors using three proteins with almost non-overlapping spectra (Malide et al., 2012). This model offers a better grade of resolution but it is still limited to only three potential outcomes and consequently, to the purification of only three cell populations in each sample. It has been shown that the number of possible combinations produced in cells can be increased by using more different fluorescent proteins or by modifying the same proteins to locate in different cellular compartments (Malide et al., 2012; Loulier et al., 2014).

To overcome these problems, as the second objective of this thesis, we aimed to generate a new cell lineage tracing tool more useful to study ITH in tumor progression. In order to solve the spectral complications of separating many different color combinations, we performed a systematic evaluation of the spectrometric characteristics of different fluorescent proteins in order to construct a new *Brainbow*-based lineage tracking allele with a higher repertoire of colors. For this purpose, we cloned and tested up to twenty fluorescent proteins in a eukaryotic expression vector (pCDNA3.1) (**Figure 6A**). Of them, according to their spectrometric characteristics, we selected ten to transiently transfected the HEK-293T cells and check their excitation and emission spectra. The emission spectra of most of them showed significant overlapping (**Figure 6B**). Nevertheless, we were able to establish unique conditions in the confocal microscope to maximize the identification of five fluorescent proteins (mOrange, tSapphire, mKate2, TagBFP, and YPet) (**Figure 6C-E**).



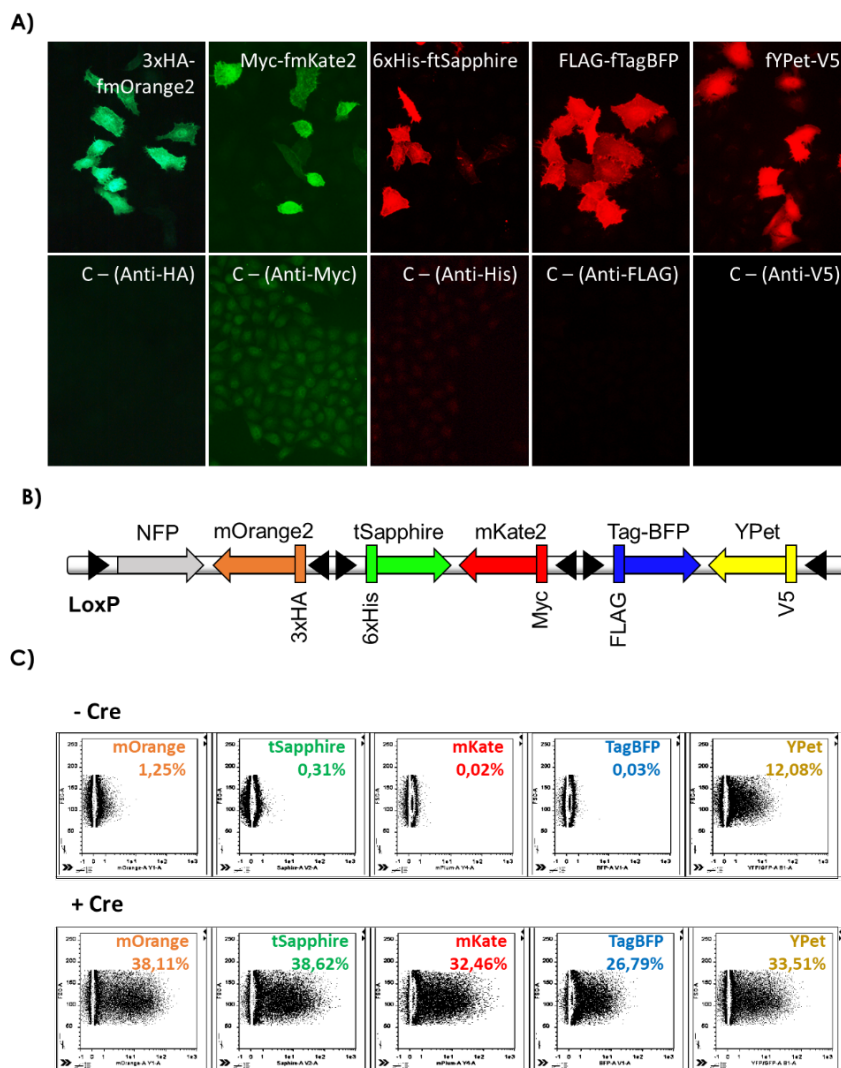
**Figure 6 | Confocal microscopic experiments.** **A)** Schematic representation of the eukaryotic expression vector, based on pCDNA3.1, used to express the fluorescent proteins (FP) in HEK-293T cells. **B)** Emission spectra diagram of several fluorescent proteins analyzed by confocal microscopy. **C)** Excitation and emission range of different fluorescent proteins. **D)** Composite image of a mix of cells individually transfected with the different fluorescent proteins. **E)** Representative images of cells transfected with five fluorescent proteins taken at five different emission intervals in the confocal microscope designed to maximize the unique identification of each protein.

Additionally, due to the different spectra shown by these proteins, we distinguished the expression of each one of these five different proteins and separated the cells expressing them using a FACS Aria III sorter with four spatially separated lasers (**Figure 7**).



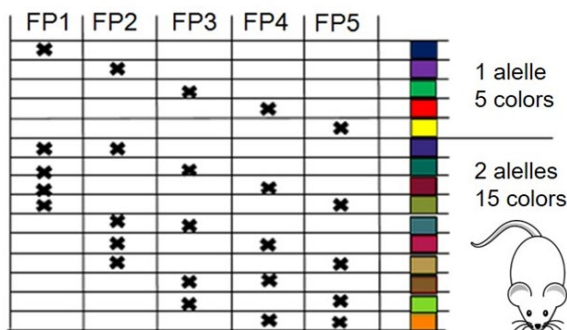
**Figure 7 | FACS experiments. A)** Excitation and emission range of different fluorescent proteins in the sorter. **B)** The 5 different cell populations are separated in the FACS.

To increase the detection versatility of our system, we subsequently incorporated a specific epitope or tag (3xHA, 6xHis, Myc, FLAG, V5) to each fluorescent protein which allows their identification with a specific antibody. We were able to uniquely detect each one of the five different labeling proteins (**Figure 8A**). With all these improvements, we constructed a new allele following the original *Brainbow-2.1* strategy (Livet et al., 2007), with the five fluorescent proteins and their tags surrounded by loxP sites (**Figure 8B**). After the transfection of the HEK-293T cells with this allele and the addition of the Cre recombinase, a homogeneous expression of all the fluorescent proteins was obtained (**Figure 8C**) proving an even number of recombination events among all loxP sites. So, our allele allows the combination and identification of different label combinations, and the same allele can be used in different experimental settings even in the absence of fluorescent detection.



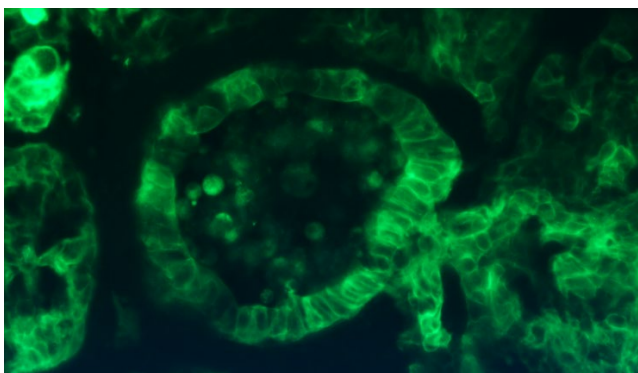
**Figure 8 | Final fluorescent protein allele.** **A)** Representative images of HEK-293T cells transfected with the five fluorescent proteins and their specific tags to uniquely detect their expression with specific antibodies. **B)** Bidimensional representation of the new lineage genetic tracing designed allele with the five fluorescent proteins and their tags surrounded by loxP sites for the study of intratumor heterogeneity in cancer mouse models. **C)** Representation of the results of cytometric analysis of HEK-293T cell line transfected with the new labeling before (up) and after (down) inducing the recombination with tamoxifen. Each graph represents the signal in each of the five selected channels (one for each protein). An even representation of cells labeled with the five different fluorescent proteins can be observed after recombination.

In a hypothetical homozygous mouse model, with two copies of this allele, up to fifteen different color combinations could be uniquely identified which increases significantly the Confetti 4-colour potential (**Figure 9**Figure 8).



**Figure 9 | Potential color combinations in a hypothetical homozygous mouse model.** Homozygous mouse cells would be able to generate up to fifteen different color combinations due to the independent behavior of the two alleles.

Finally, we have generated lentiviral vectors with the individual fluorescent proteins that can be used to generate cell culture *in vitro*. To test this strategy, we generated stably transfected pancreatic cell cultures and performed *in vivo* metastasis experiments. As can be seen in **Figure 10**, these cells retain their labeling after metastasis and produce ductal structures in the lung.



**Figure 10 | In vivo metastasis experiment.** Fluorescent pancreatic ductal structure in the lung after intravenous injection of a pancreatic primary cell line labeled with a fluorescent protein.

### III. Study the role of genetic intratumor heterogeneity in pancreatic cancer progression.

#### Characterization of punctual mutations in KFC mice

In order to study the genetic ITH *in vivo*, we performed multiregional whole-exome sequencing of pancreatic primary tumors and their associated lung- and liver-metastasis from a pancreatic mouse model (Schönhuber et al., 2014). We analyzed 46 tumoral samples from 10 mice. We found 1552 mutations of which 1426 mutations fall inside protein-coding genes. We focussed on them as they are easier to interpret and more likely to functionally impact cell behavior. We identified between 7 and 59 somatic protein-coding mutations per sample (on average, 31 mutations per sample) and we didn't observe any difference between primary and metastatic samples (on average, 29 and 32 mutations per sample, respectively), suggesting that the metastatic potential is not in general accompanied with an increase in genetic instability. We verified the presence of each mutation in all the samples from the same mouse with specific PCR amplification followed by high coverage sequencing.

In terms of the functional impact of the mutations, they are distributed in 62% missense mutations, 21% silent mutations, 8% splice mutations, 4'7% non-sense mutations, 3'7% frameshift mutations and less than 1% of start lost and stop lost mutations. This distribution is similar among mice, pointing out that most of the mutations are probably the result of random mutation accumulation due to genetic instability. Regarding the substitutional profile, the most frequent substitutions are C>T (303 mutations), G>A (250 mutations) and C>A (204 mutations).

Moreover, although we don't find a high number of recurrently mutated genes among different samples (**Table 7**), some of them could constitute new candidates for PDAC cancer genes. Some examples are *Ctnna3*, *Matr3*, or *Lrp1b*.

*Ctnna3* was found mutated in three mice, with three different missense mutations. This gene encodes a cadherin-associated protein alpha 3, which plays an important role in cell-cell adhesion (J. D. Smith et al., 2011). Therefore, its inactivation could contribute to a higher invasive characteristic in the cells. *Matr3* was found mutated in three mice, with three different mutations altering essential splice positions which likely affect the correct mRNA splicing. This gene encodes a nuclear matrix protein, which is involved in differentiation (Cha et al., 2021). Finally, we identified two different missense mutations in *Lrp1b* in two mice. *Lrp1b* gene is a putative tumor suppressor and encodes a member of the low-density lipoprotein (LDL) receptor family. The LDL receptor family interacts with multiple ligands and is proposed to be involved in extracellular signal transduction. Silencing and down-expression of *LRP1B* have been observed in renal cell carcinoma and thyroid cancer. Additional deletion of *LRP1B* has been linked to resistance to chemotherapy in high-grade serous cancers.

**Table 7.** List of the genes that are mutated in more than one mouse. *Kras* and some potential new PDAC cancer genes are colored in yellow.

Gene	N*	Gene	N*	Gene	N*	Gene	N*
<i>Dhfr</i>	4	<i>Car4</i>	2	<i>Lrp1b</i>	2	<i>Rp1</i>	2
<i>Hfm1</i>	4	<i>Cdk20</i>	2	<i>Mapk8ip2</i>	2	<i>Rpap2</i>	2
<i>Ctnna3</i>	3	<i>Col9a1</i>	2	<i>Mn1</i>	2	<i>Sap30bp</i>	2
<i>Dop1b</i>	3	<i>Dgka</i>	2	<i>Mroh2a</i>	2	<i>Serbp1</i>	2
<i>Gm12880</i>	3	<i>Efl1</i>	2	<i>Mrpl49</i>	2	<i>Slc2a4</i>	2
<i>Matr3</i>	3	<i>Fgd6</i>	2	<i>Msh5</i>	2	<i>Snhg4</i>	2
<i>Stam2</i>	3	<i>Foxr2</i>	2	<i>Nbeal2</i>	2	<i>Sycp1</i>	2
<i>Tmcc1</i>	3	<i>Garem1</i>	2	<i>Nfasc</i>	2	<i>Ubap2</i>	2
<i>4930474N05Rik</i>	2	<i>Grin2c</i>	2	<i>Nipbl</i>	2	<i>Ube2k</i>	2
<i>Ankrd45</i>	2	<i>Hmcn2</i>	2	<i>Nup37</i>	2	<i>Unc13d</i>	2
<i>Arpc1b</i>	2	<i>Kras</i>	2	<i>Peak1</i>	2	<i>Utrn</i>	2
<i>B430010I23Rik</i>	2	<i>Krba1</i>	2	<i>Pigo</i>	2	<i>Zan</i>	2
<i>Bnip3</i>	2	<i>Lnx1</i>	2	<i>R3hdm1</i>	2		
<i>Cacul1</i>	2	<i>Loxhd1</i>	2	<i>Rab3c</i>	2		

\*N = Number of mice that present mutations in a specific gene

Interestingly, although KFC mice already have a mutant version of *Kras* (*Kras*<sup>G12D</sup>), we have found *Kras* activating mutations (*Kras*<sup>Q61L</sup>) in the wild-type (WT) allele of two lung metastatic samples from two mice (ID-2365 and ID-1187).

### Characterization of big genomic rearrangements in KFC mice

We studied copy number alterations (CNA) and regions with loss of heterozygosity (LOH) from the exome sequencing data.

We analyzed the CNA present in the 46 tumoral samples from the 10 mice. We found 1680 CNA, of which 1128 were gain alterations. In particular, there were between 17 and 126 CNA per sample (on average, 37 CNA per sample). We observed a slight increase in the number of CNA in the lung metastatic samples (on average, 48 CNA per sample) compared to the pancreatic primary tumor and the liver metastatic samples (on average, 33 and 31 CNA per sample, respectively). This increase was not significant ( $p\text{-valor}_{\text{CNA-LungMet-PPT}} = 0.0504$ ;  $p\text{-valor}_{\text{CNA-LungMet-LiverMet}} = 0.15$ ), in agreement with the commented substitution profiles, emphasizing the idea that the metastatic potential wouldn't be accompanied with an increase in genetic instability.

Additionally, in those samples coming from mice with C57BL/6-129Sv background (26 tumoral samples from 5 mice), we were able to identify regions with loss of heterozygosity (LOH). Similar to what is observed in human PDAC samples, in some cases, we found deletions or LOH in the region where *Cdkn2a* is located (chromosome 4 from 89192710 position to 89212856 position, Chr4:89192710-89212856; **Figure 11A, B**). We also find recurrent alterations in chromosome 6 around the region where *Kras* is located (Chr6:145162425-145195965; **Figure 11B**).



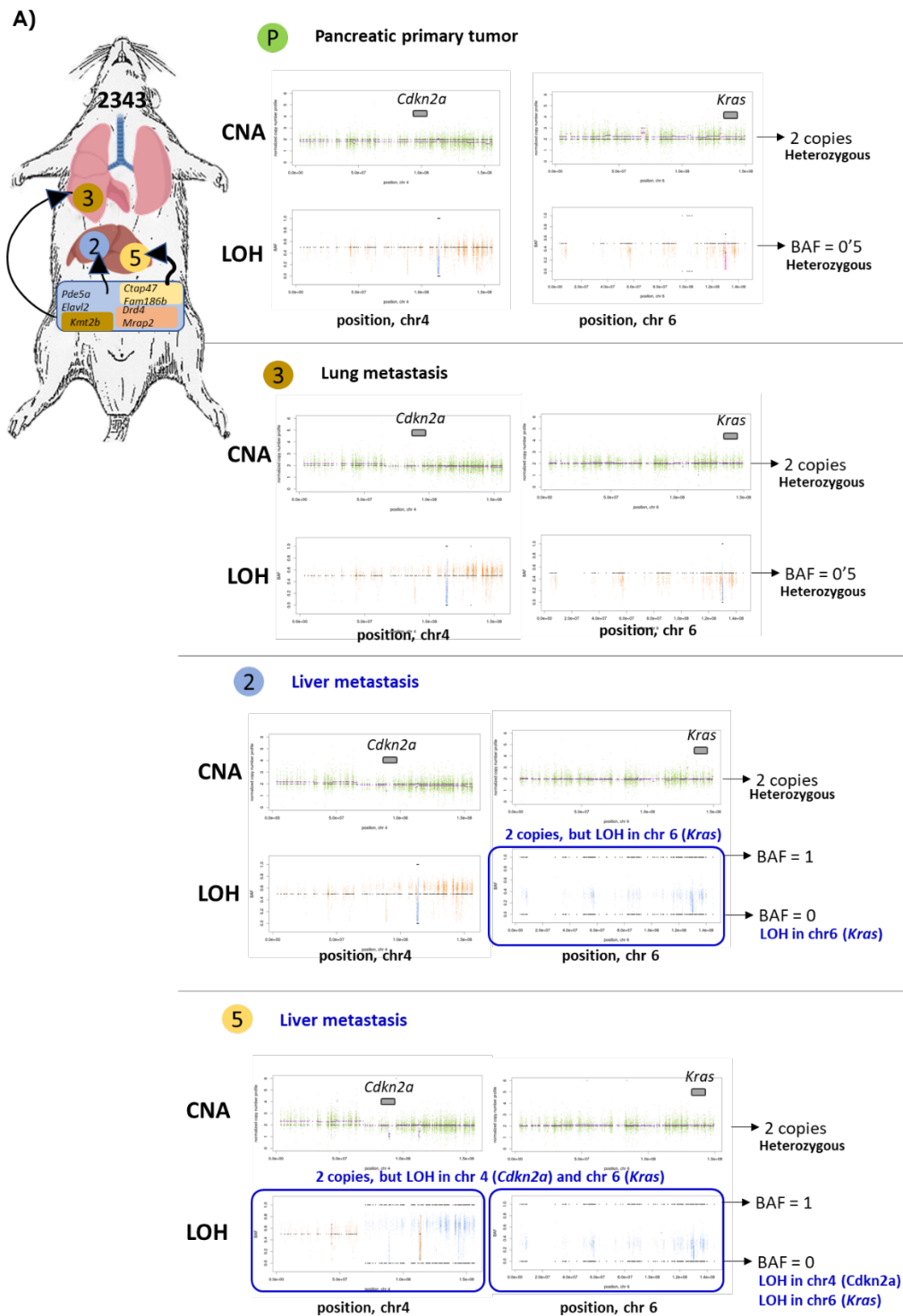
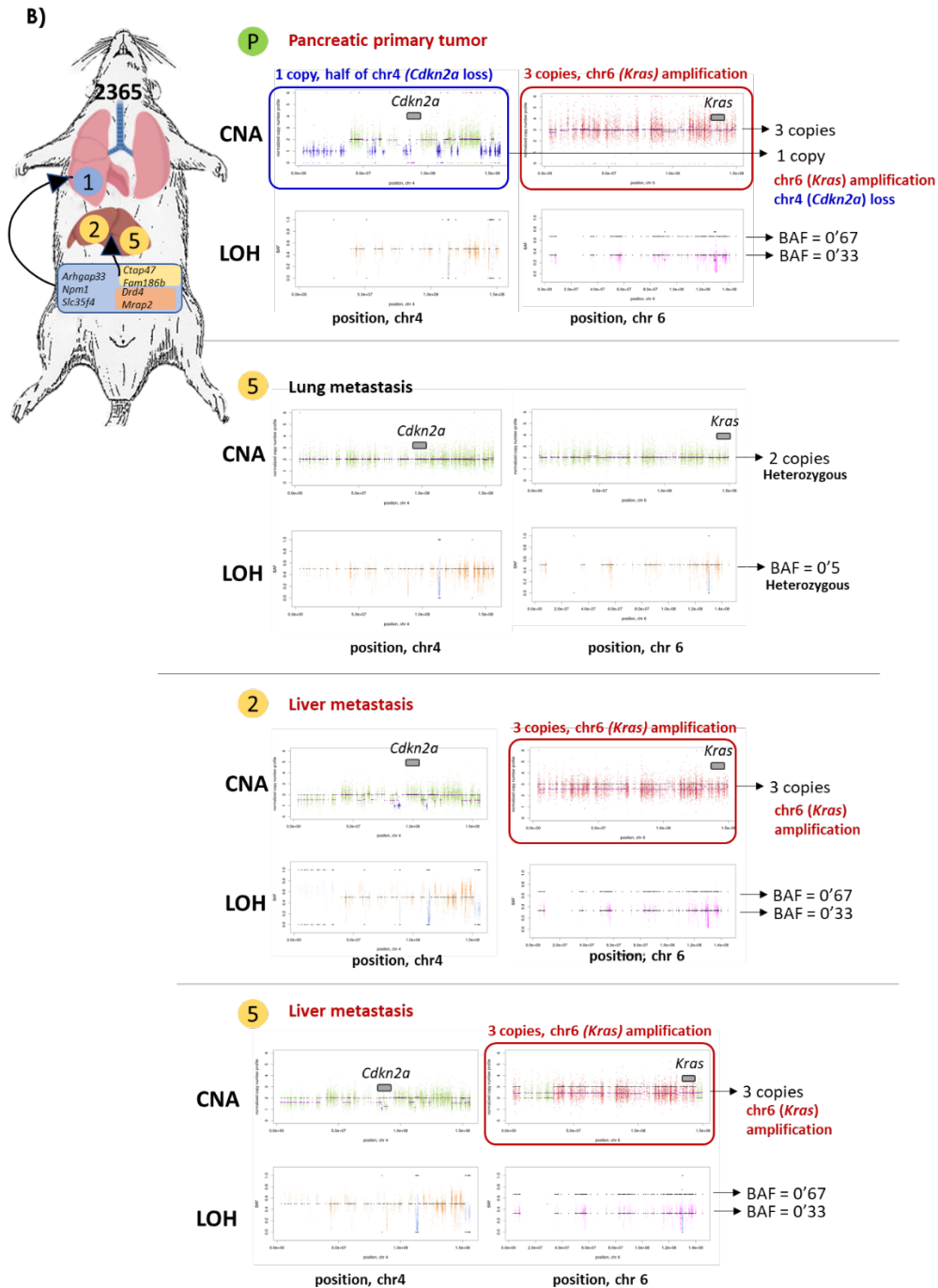


Figure 11 continues on the next two pages.



**Figure 11 | Schematic representation and phylogenetic relationships of the different clones present in the primary pancreatic tumor and their arising metastasis using a multisample sequencing strategy to decipher tumor evolution. The graphs represent copy number alterations (CNA) and loss of**

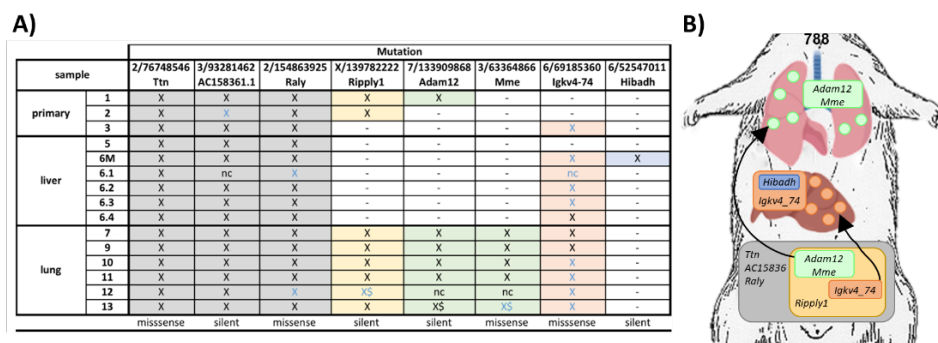
heterozygosity (LOH) from chromosomes 4 and 6 of two different mice (ID-2343 and ID-2365). **A)** Complete LOH in the chromosome 6 (containing *Kras*) of the two-liver metastasis (in blue), is likely the result of the loss of the WT chromosome plus the duplication of the mutant one because there are 2 copies of the chromosome 6. LOH is also present in chromosome 4 (harboring *Cdkn2a*) in one liver metastasis without any CNA (in blue). The lung metastatic sample is perfectly heterozygous without any CNA. **B)** Deletion of one copy of chromosome 4 in the primary pancreatic tumor (containing *Cdkn2a*, in blue). Trisomy of chromosome 6 (harboring *Kras*, in red) in the primary and the liver metastases. The lung metastasis is heterozygous without any CNA.

These alterations likely result in the further activation of the oncogenic *Kras* signal by either loss of *Kras* WT allele or amplification of *Kras* mutant allele, as has been described in a recent collaboration of our group in a project led by Dr. Roland Rad and Dieter Saur (Mueller et al., 2018). The need to amplify the *Kras* oncogenic signal is also shown by the identification of two activating mutations in the *Kras* WT allele (*Kras*<sup>Q61L</sup>) from two lung metastases. Interestingly and in contrast with what is observed in humans, we didn't find alterations in *Trp53* or *Smad4* genes. This indicates that, whereas inactivation of *Cdkn2a* seems to be essential in both human and murine PDAC, *Trp53* and *Smad4* alterations don't seem to be required for a full PDAC progression in mouse models.

Finally, and in accordance with the need to further activate *Kras* oncogenic signals, in two lung samples in which we failed to detect CNA in the *Kras* region, we found two new *Kras* activating mutations (*Kras*<sup>Q61L</sup>) in the WT allele (ID-2365 and ID-1187 mice).

### Characterization of ITH and evolution inference in murine pancreatic tumors

The identification and validation of all the mutations and CNA in multiple samples from the same mice allowed us to confirm the presence of different genetic clones inside the tumors and infer their evolutionary relationships (**Figure 12**).



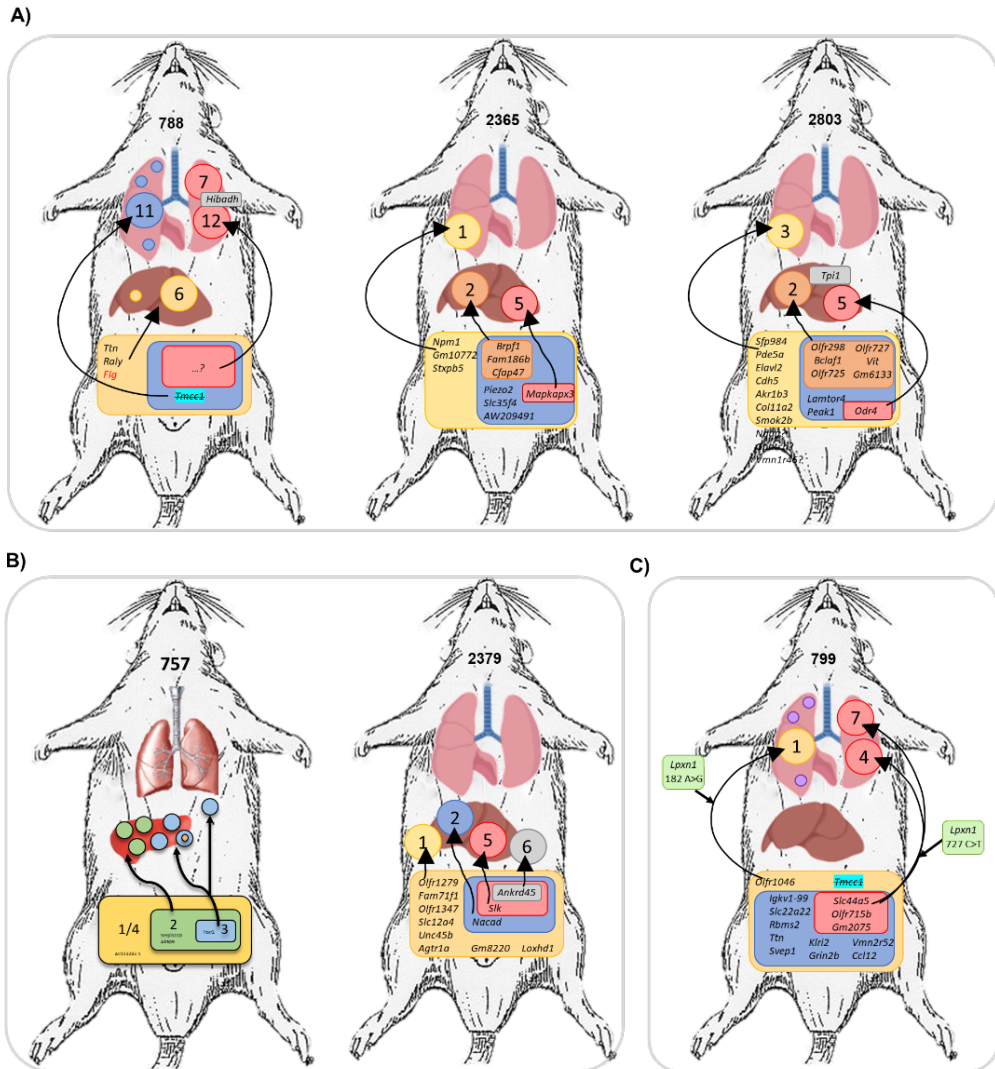
**Figure 12 | Genetic evolution in murine pancreatic tumors. A)** Somatic mutations validated from exome sequencing in pancreatic primary tumors (PPT) and metastatic samples of a mouse (ID-788). Common mutations are shared between all the lesions (in grey), private mutations are specific to a specific tumor region (in yellow, green, and orange) or restricted to a single lesion (liver metastases in blue). There are tissue-specific mutations present in the PPT and the liver metastases (in orange) or in the PPT and the lungs (in green). **B)** Schematic representation and phylogenetic relationships of the different clones present in the PPT and their arising metastasis to decipher tumor evolution from the previously validated mutations.

In all the mice, we found shared mutations in all the taken samples from the same mice, indicating a common genetic origin of the tumors (monofocal tumors) (**Figure 12, Figure 13**), which is in accordance with the Confetti experiments. These common mutations are known as clonal mutations (also as founder/public mutations) and are assumed to have been accumulated in the early stage of cancer evolution. Moreover, the presence of heterogeneity inside the tumors and the metastases was confirmed by the presence of different mutations in one or some samples from the same mouse. This observation reveals the presence of subclones accumulating specific progressor or private mutations that contribute to the formation of ITH (**Figure 13**).

In some cases (ID-757, ID-799, ID-2379), there was evidence of separate waves of metastatic spread from the primary tumor to different tissues through the presence of some metastatic subclones sharing a higher number of mutations with the primary tumor which were not present in other metastatic subclones from the same tissue, suggesting continuous metastatic seeding (**Figure 13B, C**). The CNA also revealed different waves of metastatic spread. This is the case of a mouse (ID-2365) in which we found CNA in chromosomes 4 and 6 in the primary and the liver metastases, but not in the lung metastases. This observation, plus the fact that all the samples shared a number of mutations, indicated that the lung metastasis occurred as an early event (**Figure 13B**). This could also reflect that the alterations required to produce a metastasis in the liver are different from those needed in the lungs.

In a single mouse (ID-799), we observed the *Lpxn* gene, also known as Leupaxin, mutated in two independent manners, suggesting convergent evolution (**Figure 13C**). In humans, this gene acts as a negative regulator in integrin-mediated cell adhesion events and contributes to the regulation of cell adhesion, spreading, and cell migration.

In other mice (ID-788, ID-2365, ID-2343, and ID-2850), we found specific mutations present in some subclones of the primary tumor and the lung metastases, and other private mutations that were present in other subclones from the primary and the liver metastases, revealing tissue-specific metastasis (**Figure 13A**). Curiously, *Kras* amplification is recurrently not present in the lung metastases. The CNA can also indicate tissue-specific metastasis. This was the case of the ID-2850 mouse in which the two lung metastatic samples and a primary tumor sample were very similar in the CNA analysis, while the liver metastases showed a low number of alterations and they were more similar between them, probably indicating tissue-specific metastases



**Figure 13 | Schematic representation and phylogenetic relationships of the different clones present in the primary pancreatic tumor and their arising metastasis using a multisample sequencing strategy to decipher tumor evolution.** In all the cases, the common mutations present in all the samples are colored in yellow, indicating the monofocal origin of the tumors. Genetically different subclones are represented by several colors, confirming the presence of ITH. **A)** Specific subclones of the primary tumor are present in the lung metastases, and other private mutations present in other subclones from the primary are also present in the liver metastases, revealing tissue-specific metastasis. **B, C)** Presence of some metastatic subclones sharing a higher number of mutations with the primary tumor which were not present in other metastatic subclones from the same tissue, indicating separate waves of metastatic spread from the primary tumor to different tissues. **C)** *Lpxn* gene mutated in two independent manners, suggesting convergent evolution.

Unfortunately, we have not found any metastasis-specific recurrent genetic alteration. This observation, in accordance with recent studies in colorectal cancer (Ganesh et al., 2020), suggests that metastatic potential could not be genetically determined, or at least, not by mutations present in protein-coding genes. As any alteration, genetic, epigenetic, or caused by the microenvironment, that affects cellular behavior is expected to affect the cell transcriptome, we decided to characterize the functional intratumor heterogeneity in these tumors.

## IV. To characterize the presence of transcriptomic intratumor heterogeneity and its role in metastasis

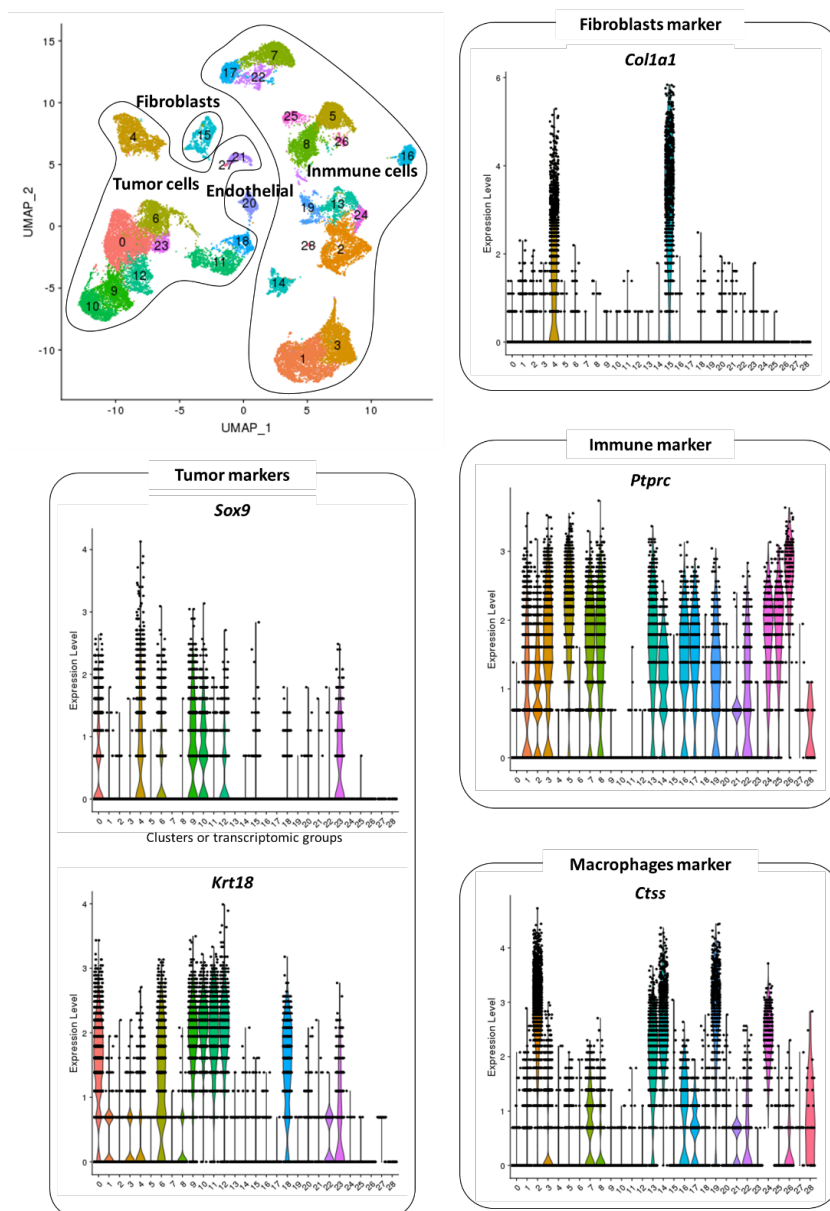
### Characterization of the presence of transcriptomic ITH in KFC mice

In order to study the transcriptomic or functional ITH *in vivo*, we performed single-cell RNA-seq of primary tumors and their associated lung- and liver-metastasis from our pancreatic mouse model. In total, we analyzed 15 samples from 9 different mice including the total fraction of 9 pancreatic primary tumors, the tumor isolated cells of 2 of them, and the total fraction of 2 associated lung-metastasis and 2 associated liver-metastasis.

We obtained transcriptional data from a total of 97432 cells (6495 cells per sample, on average). From them, we obtained good quality data from 60741 cells ranging from 1441 to 10572 cells per sample.

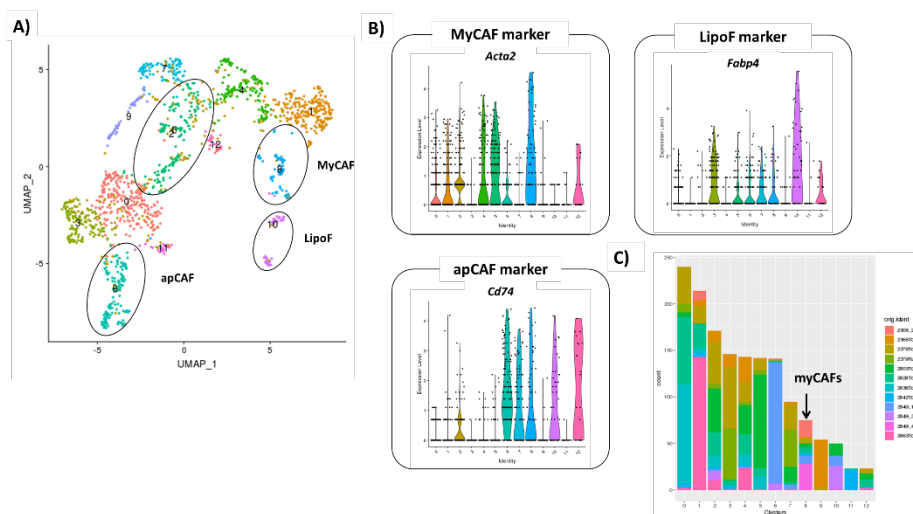
After performing dimensionality-reduction-based UMAP clustering, we identified 28 subgroups/subclusters of cells according to their transcriptomes (**Figure 14**). These subclusters represent the transcriptionally different cell populations present inside the tumors. First of all, we identified the stromal component by looking for the expression level of specific known markers (**Figure 14**). Thereby, immune component of the tumors can be easily identified by a high expression of the *Ptprc/Cd45* gene which is specific for blood cell differentiation. As expected, we found this gene highly expressed in an important fraction of the cells, confirming the huge infiltration of immune cells in the pancreatic tumors. Tumor cells were identified by the overexpression of *Sox9*, *Krt18*, and *Krt19* marker genes. We also identified macrophages and lymphocytes. Next, *Colagen1A1* was used to identify a specific cluster of cancer-associated fibroblasts (CAFs).





**Figure 14 | Unsupervised dimension reduction projection of the transcriptional profile of 60741 cells from 15 samples from 9 PDAC mice.** Machine-derived clustering together with violin plots of specific marker expression, showing normalized expression in each of the subclusters, allows the identification of different cell types and transcriptional groups. *Sox9* and *Krt18* are known markers of tumor cells, *Ptprc/Cd45* is highly expressed in immune cells, *Ctss* is a macrophage marker gene, and *Col1a1* is a fibroblast marker gene.

To further characterize CAF subpopulations, we separately analyzed the cells from the main fibroblast cluster (cluster 15 from **Figure 14**) and clustered them on a UMAP plot, forming 13 subclusters, each with distinct gene signatures (**Figure 15**). We identified at least 3 of the 4 types of fibroblasts that have been previously described in PDAC (Elyada et al., 2019). For instance, subcluster 8 expressed high levels of *Acta2*, corresponding with myofibroblastic CAFs (myCAF). This subcluster is composed of fibroblast from most of the samples, indicating that this subpopulation of CAFs appears recurrently during PDAC progression. Subcluster 6 expressed marker genes, such as *Cd74* (among other antigen-presenting molecules such as TCR or MHCII), that were very similar to the human antigen-presenting CAFs (apCAF) signatures. But, these fibroblasts are only identified in a single primary sample (ID-2849\_1). Subcluster 10 was rich in the *Fabp4* marker like the lypofibroblasts (lipoF). These lipoF were identified in at least two mice. We did not observe any group that could correspond to the called inflammatory CAFs (iCAF) that overexpress cytokines.



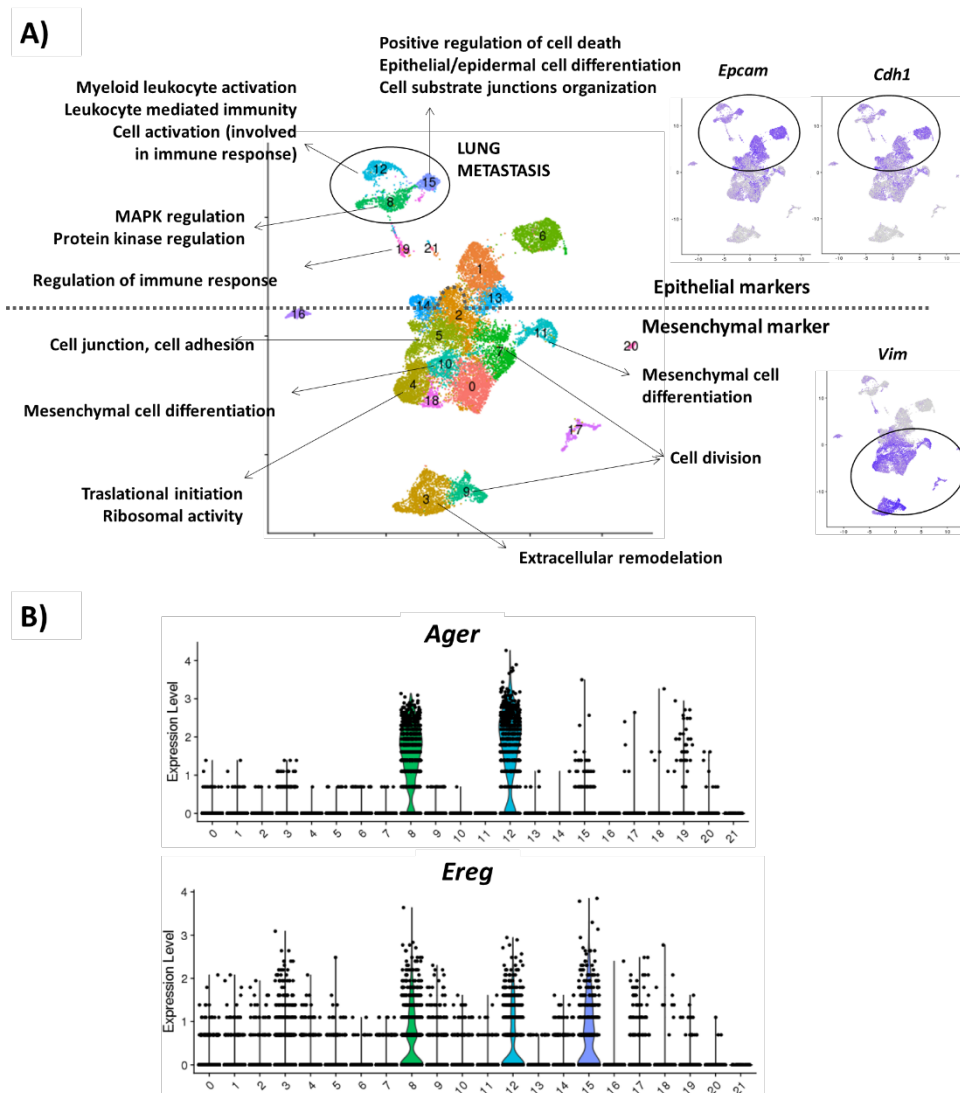
**Figure 15 | Unsupervised UMAP projection of the transcriptional profile of cancer-associated fibroblasts.** **A)** Unsupervised clustering of all cancer-associated fibroblasts from 12 samples from 9 PDAC mice. **B)** Violin plots of selected myofibroblastic CAFs (myCAF), lypofibroblasts (lipoF), and antigen-presenting CAFs (apCAFs) markers showing normalized expression in each of the subclusters. **C)** Composition of each subcluster of fibroblasts among samples.

### Characterization of the presence of recurrent transcriptional tumor cell groups

In order to identify the different transcriptional programs inside the tumor cell component, we extracted tumor cells from the main 11 tumor subclusters (subclusters 0, 4, 6, 9, 10, 11, 12, 18, 21, 23, and 27 from **Figure 14**) and analyzed them separately. Unfortunately, liver samples were so highly infiltrated with immune cells that we failed to obtain good quality data from any tumor cell. Clustering of the cells allowed us to identify 21 subclusters (**Figure 16A**), each with distinct transcriptional signatures.

Interestingly, cancer cells from the two different mice with lung metastasis clustered together with a distinct transcriptional profile (subclusters 8, 12, and 15) (**Figure 16A**). Particularly, they appeared to be involved in positive regulation of cell death, epithelial and epidermal cell differentiation, the organization of cell-substrate junctions, myeloid leukocyte activation, leukocyte mediated immunity, and MAPK and protein kinase regulation. In particular, these metastatic subpopulations overexpressed *Ager* and *Ereg* genes as late metastatic markers (**Figure 16B**).

Moreover, inside the functional ITH, we found recurrent transcriptionally differentiated cancer cell populations, indicating that some transcriptional cancer cell groups appear recurrently during PDAC progression (**Figure 17A**). Some of these subgroups overexpress gene pathways that likely play important functions during tumor progression. For instance, subcluster 4 overexpress genes involved in translational initiation and ribosomal activity, and subcluster 5 is involved in cell junction and cell adhesion processes. The cells in these both subclusters likely have a higher metastatic potential than the rest of the cells. Additionally, subcluster 7 contains cells that overexpress genes involved in cell division which likely correspond to a population of highly proliferative cells. Finally, subcluster 19 overexpress genes involved in the regulation of the immune response which might constitute a group of cells in high communication with tumor stroma.



**Figure 16 | A)** Unsupervised UMAP projection of all the tumor cells from 13 samples from 9 PDAC mice. On the right side, there are graphical representations showing the normalized expression of epithelial (*Epcam* and *Cdh1*) and mesenchymal (*Vim*) markers in the clustering. **B)** Violin plots of selected *Ager* and *Ereg* marker genes as late metastatic markers showing normalized expression in each of the subclusters.

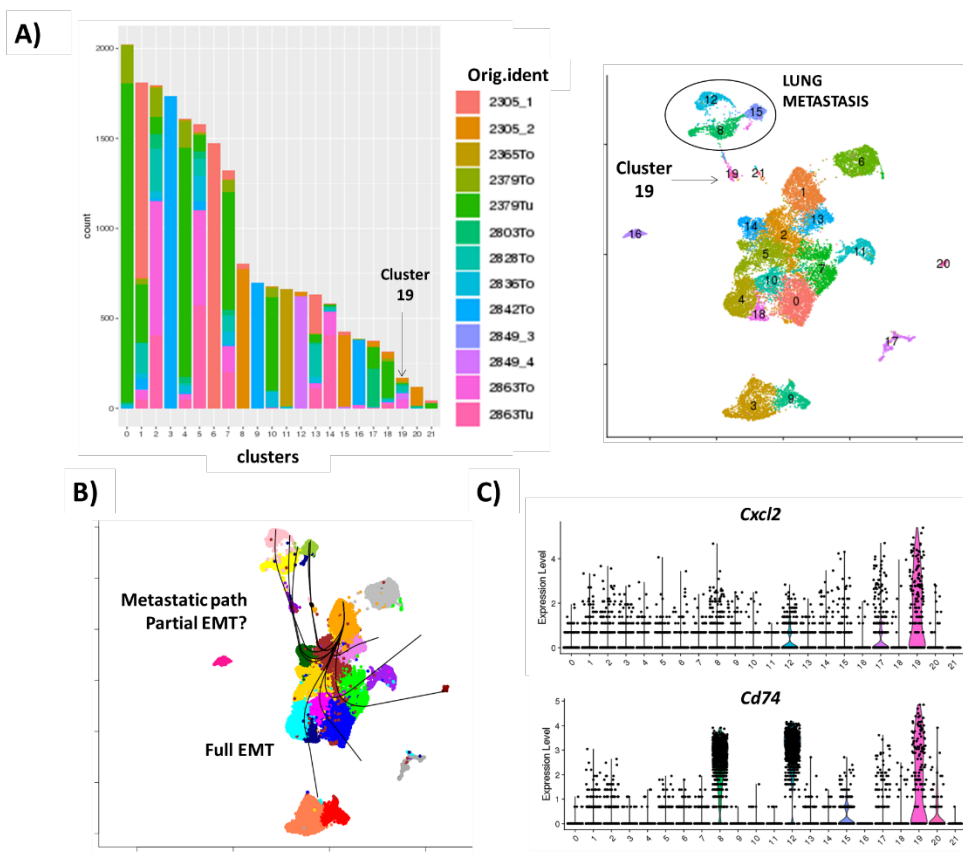
## Epithelial-mesenchymal transition and metastatic trajectories

We also observed a higher expression of epithelial markers, such as *Epcam* or *Cadherin1*, in some subclusters from the primary (subclusters 1, 6, 13, and 14) and the metastatic subclusters (subclusters 8, 12, and 15) compared with a clear reduction of the epithelial identity in others (0, 2, 4, 5, 7, 10, 11, 17, and 18) or even a complete loss of the epithelial identity (subclusters 3 and 9), where *Vimentin* expression was higher, indicating a mesenchymal profile (**Figure 16A**). This observation may reflect a continuous epithelial-mesenchymal transition (EMT) in the tumor cells, which has been reported in many tumor types, including pancreatic cancer, and has been associated with the metastatic potential (Rhim et al., 2012).

Interestingly, trajectory inference analysis identified at least two differentiated transcriptional programs of EMT, with metastatic populations presenting only partial EMT versus primary tumor cells, which shows a complete EMT (**Figure 17B**). While studying the intermediate stages to characterize the transcriptional trajectories, subcluster 19 was proposed as an intermediate point in this transition.

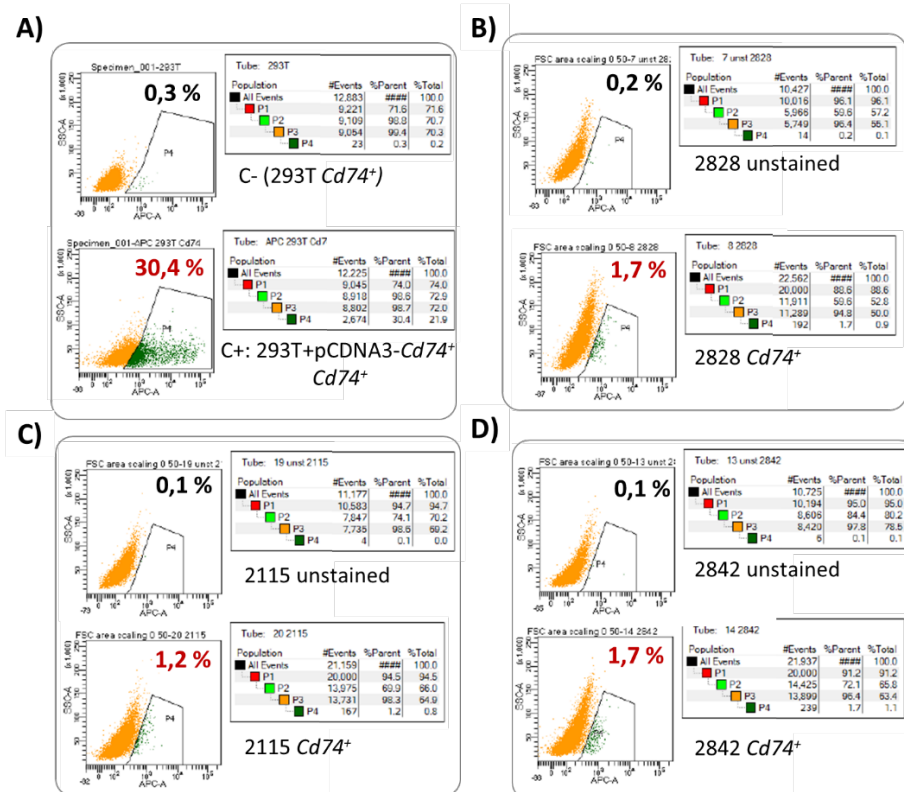
Moreover, among the different recurrent transcriptional groups (**Figure 17A**), the cluster 19, that is located in the middle between the primary tumor and the metastases, is composed of cells extracted from primary and metastatic samples suggesting that it could be a group of pre-metastatic cancer cells inside of the primary tumors with already some transcriptional characteristics of the metastatic cells.

Next, we looked for specific markers of this minor cell population of interest (**Figure 17C**). We observed overexpression of several ligands and receptors of the *TGF $\beta$*  pathway, such as *Cxcl2* and *Il1*. This could indicate that this minority cell population communicates with the normal cells of the tumor microenvironment to promote angiogenesis and invasion of the primary tumor and they might prepare the metastatic niche in the lungs.



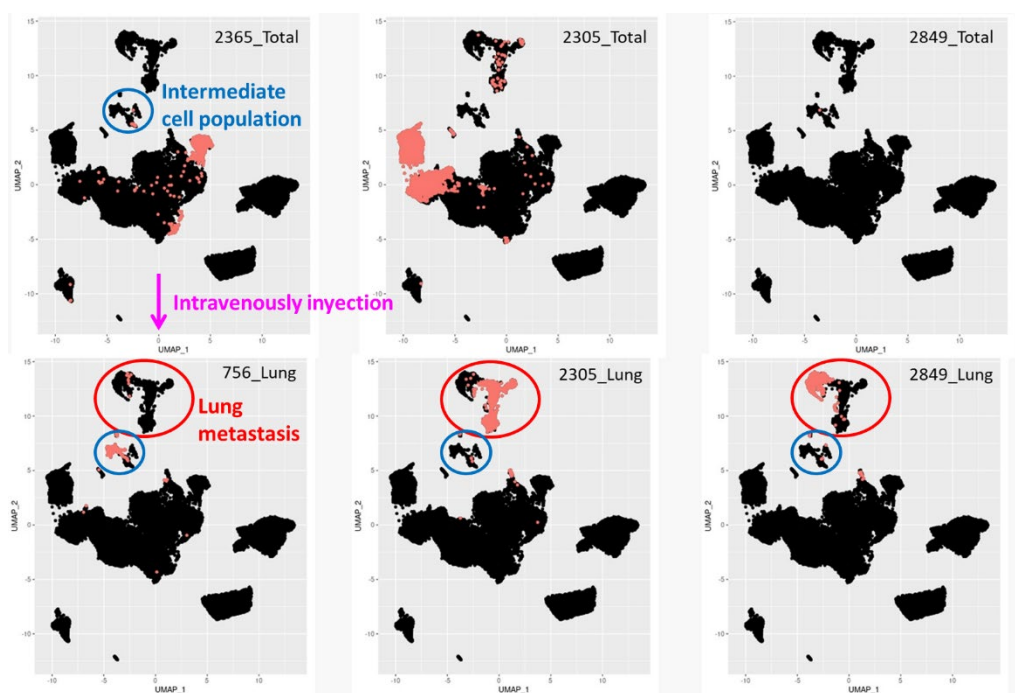
**Figure 17 | A)** Composition of each tumor subcluster among samples. On the right, repeated unsupervised UMAP projection of all the tumor cells. **B)** Trajectories inferred with the InferCNV software. **C)** Violin plots of *Cxcl2* and *Cd74* early markers genes of the pre-metastatic cell population.

In an attempt to characterize this tumor minor cell population, we selected *Cd74* among the highly expressed specific markers, to purify these cells from the primary cell cultures that had been previously established from the tumor cells of the primary tumor. Unfortunately, although we were able to purify this minority cell subpopulation in the FACS with a higher *Cd74* expression (**Figure 18**), after expanding and maintaining the purified cell lines in cell culture, the most expected enriched genes were lost (such as *Cd74*, *Cxcl2*, *Il1b*, *Srgn*, *Lyz2*, *Ctss*, and *Ccl6*). This observation suggests that the transcriptional profile present in these cells in the primary tumors appears as a consequence of the specific tumor microenvironment and it is not maintained when we grow the cells *in vitro*.



**Figure 18 | FACS experiments isolating the *Cd74*<sup>+</sup> minority cell population. A)** HEK-293T cells were used the negative control of the experiment. HEK-293T cells transfected with pCDNA3-*Cd74*<sup>+</sup> expression vector used as a positive control. **B, C, D)** Primary cell culture established from 3 different primary pancreatic tumors showing the *Cd74*<sup>+</sup> minor cell population that we isolated.

To validate our hypothesis, we performed an *in vivo* metastasis assay in which we intravenously injected a primary cell line (from ID-2365) into a C57BL/6-129Sv mouse. These cells recovered their ability to produce lung metastasis. After scRNA-Seq analysis of the tumor cells present in the induced-lung metastasis (ID-756-Lung), we found a transcriptionally similar cell population that clusterized with the original pre-metastatic population (**Figure 19**). These observations confirm that although the transcriptional profile is lost in cell culture, it is recovered in a favorable pro-tumorigenic microenvironment and the metastatic potential is conserved after growing up the cells *in vitro*.

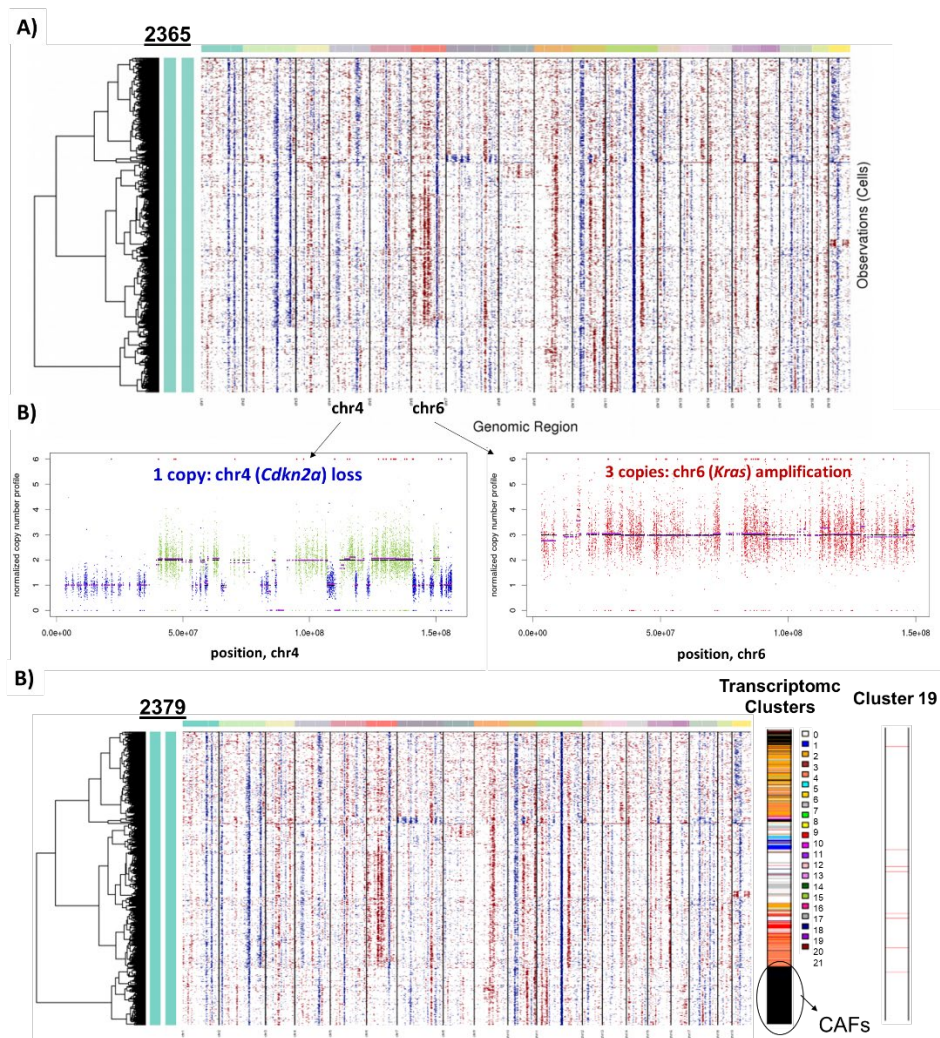


**Figure 19 | Unsupervised clustering of tumor cells.** In each graph, the tumor cells from each specific mouse are colored in pink. 756-Lung metastasis is originated after injecting a primary cell culture established from the ID-2365 mouse. Tumor cells from the 756-Lung metastasis are transcriptionally similar to the tumor cells from other lung metastases and to the pre-metastatic cluster of interest. Lung metastatic clusters and the intermediate pre-metastatic cell population are highlighted with red and blue circles, respectively.



### **Analysis of the potential relationship between the genetic and the transcriptomic ITH.**

Finally, we analyzed the potential relationship between the genetic and the transcriptomic ITH. We identified the presence of CNA from the scRNA-seq data using InferCNV software through the assumption that changes in the expression of several genes that are closely located reflect a change or a CNA in the same part of the chromosome. Interestingly, we recurrently observed changes in chromosome 4 (*Cdkn2a*) and chromosome 6 (*Kras*), supporting our previous findings in the exome data. We clustered the cells according to these genetic alterations predictions and tried to associate them with the previously identified transcriptional groups. As it can be seen in **Figure 20**, there is not any clear correlation between the genetic and the transcriptomic subclusters which indicates that the generation of a specific transcriptional group of those recurrently observed in our mice, is not in general associated with the acquisition of specific genetic alterations. One potential exception is the transcriptomic subcluster 4 which seems to group mainly in a specific subcluster of the genetic tree. Notably, when we introduced the CAFs in the analysis, they grouped together with the absence of any predicted CNA. This observation suggests that these CAFs originate from normal cells and not from a specific group of tumor cells (Yamaguchi et al., 2021). Additionally, the transcriptional minority subcluster 19 is randomly distributed among these genetic subgroups which again supports that the metastatic potential is not in general associated with the acquisition of specific secondary genetic alterations.



**Figure 20 | Analysis of the potential relationship between the genetic and the transcriptomic ITH. A)** Estimation of copy number variants by InferCNV across all the tumor cells from a specific mouse (ID-2365). The rows correspond to the cells. The top color bars indicate different chromosome regions. Genes are ordered from left to right across the chromosomes. Chromosomal region amplifications and deletions are shown in red and blue, respectively, indicating higher or lower expression values of the tumor cells compared with those in control cells. Hierarchical clustering is shown on the left. **B)** Representation of the CNA from chromosome 4 (*Cdkn2a* loss) and chromosome 6 (*Kras* amplification) from the previous mouse. **C)** InferCNV from a second mouse (ID-2379). On the right, all transcriptomic subclusters, including a group of CAFs as a control, and the tumor cells from the transcriptomic subcluster 19 are represented.





**Discussion**



Pancreatic ductal adenocarcinoma (PDAC) is a highly lethal disease accounting for 95% of all pancreatic malignancies. While new treatment options have improved the prognosis of many other cancer types, PDAC has a poor outcome and a rising incidence due to its late diagnosis, early metastatic widespread, and the lack of effective treatment options (Siegel et al., 2022). Taking advantage of next-generation sequencing (NGS), an extensive genetic intratumor heterogeneity (ITH) of PDAC was revealed (Jones et al., 2008; Biankin et al., 2012; Waddell et al., 2015; Bailey et al., 2016). Moreover, PDAC is characterized by a highly inflammatory and desmoplastic tumor microenvironment (TME), which contributes to the aggressiveness and drug resistance of the disease (Whatcott et al., 2013). As a high grade of ITH has been associated with poor prognosis in cancer patients, in this Doctoral Thesis we proposed a proper characterization of the dynamics of ITH to understand the mechanisms involved in PDAC progression and metastasis, with the final aim of improving cancer patient's survival through the design of more efficient treatments. In particular, we proposed for the first time the independent study of different cell clones present in the primary tumor and their associated metastasis. To do this, we combined the flexibility of a mouse model with the high sensitivity of next-generation, single-cell sequencing technologies and a multi-fluorescent protein lineage tracking system.

Mouse cancer models have been used in the past to study tumor-promoting molecular pathways. Nevertheless, the translation of the observations done in mice to human patients has been traditionally difficult due to species-specific mechanisms. In our case, pancreas-specific expression of *Kras*<sup>G12D</sup> in mice induces PDAC that faithfully recapitulates all types of lesions observed in human PDAC progression, from PanINs to metastatic spread to lymph nodes, liver, and lungs, the same tissues of preferent metastasis in human samples. Surprisingly, even when the mutant *Kras* allele is activated in multiple cells, we systematically observed that advanced tumors are monofocal, evidenced by the combination of this mouse model

with an inducible multi-fluorescent protein lineage tracking allele. Unfortunately, the labeling was lost in the most advanced stages of the disease. A potential explanation of this could be the result of an immunogenic reaction to the expression of the fluorescent proteins. According to some authors, GFP expression can deteriorate over time as GFP-labelled cells are prone to death, proving the immunogenicity and toxicity of this fluorescent protein *in vivo* (Ansari et al., 2016). Importantly, adverse immune responses to tumor cells expressing the GFP reporter have been described (Stripecke et al., 1999; Day et al., 2014). Nevertheless, we are able to see fluorescently marked cells in the mice after intravenous infection which proves that there is not a massive immune response against fluorescent proteins in our model. Alternatively, it was demonstrated that the oncogenic *Kras* signal in mouse PDAC is frequently activated by either loss or silencing of the *Kras* wt allele (Mueller et al., 2018). Due to our mouse design, the Brainbow-2.1 allele is located in the *Rosa26* locus in cis with the *Kras* wt allele. Consequently, a plausible explanation could be the concurrent loss of the Brainbow-2.1 allele together with the *Kras* wt allele, losing the expression of the fluorescent protein in the more advanced states. This observation reinforces the requirement for tumor cells to amplify the mutant *Kras* signal in the later stages of PDAC development in our mouse model.

In addition to this limitation of our mouse model setting, the original fluorescent proteins selected for the original Brainbow-2.1 construct presented mostly overlapping spectra, which complicates their identification and the analysis of potential protein combinations, limiting the strategy to only four potential outcomes (Schepers et al., 2012). Being aware of this problem, the same authors improved the construct (called Brainbow-3.2) using three proteins with almost non-overlapping spectra and incompatible loxP sites (Cai et al., 2013). However, the strategy was restricted to only three different markers. At the same time, other authors showed that more different fluorescent proteins could be used or these proteins could be modified to locate them in different cellular compartments and



thus increase the number of potential combinations produced in cells (Malide et al., 2012; Loulier et al., 2014).

For these reasons, we proposed to improve the original Confetti allele and introduce it in another murine locus to avoid the loss of the labeling in the most advanced stages of the disease. In this context, we generated a construct containing five different fluorescent proteins and specific tags (3xHA-mOrange2, 6xHis-tSapphire, Myc-mKate2, FLAG-TagBFP, and V5-YPet) surrounded by loxP sites. We validated the even representation of the five markers after CRE recombination. This new allele could be used in different experimental conditions, even without fluorescent detection due to the use of specific tags. In a hypothetical homozygous mouse model, with two copies of the new allele, up to fifteen color combinations could be uniquely distinguished, which significantly increases the original 4-color potential. Moreover, we have proved that the labeling is maintained after metastases induction through the injection of lentiviral vectors containing each fluorescent protein. This new lineage tracing tool might allow studying the involvement of ITH in cancer progression. We propose to introduce this new allele in the *ColA1* locus (García-Marqués & López-Mascaraque, 2013), which is in a different chromosome than *Kras*, avoiding the loss of the labeling due to the loss/silence of the *Kras* wt allele. Additionally, we have generated individual lentiviral particles containing each one of the new markers independently that can also be used in different experimental settings to mark different cell populations.

Next, we studied the role of genetic ITH in PDAC progression through multisampling whole-exome sequencing experiments in a total of 55 samples from 10 different mice. We found on average 62 protein-coding mutations per sample, including substitutions, small insertions and deletions (indels), and copy number alterations (CNA), showing no statistically significant differences between primary and metastatic samples, which suggests that the metastatic potential wouldn't be accompanied with a higher genetic instability.

It should be noted that this study is the first time in which multisampling sequencing has been carried out in mice to finely characterize the evolution of genetic intratumor heterogeneity in a genetically homogenous mouse model. Interestingly, our KFC mice showed genetic ITH similar to human PDAC patients (Yachida et al., 2010; Campbell et al., 2010), so studies of the evolution of ITH can be done using animal models.

As in humans, we recurrently found alterations in *Cdkn2a* and *Kras*, but alterations in *Trp53* or *Smad4* were not identified. In human PDAC, *KRAS* and *CDKN2A* have been described to be affected in the earlier stages, while *TP53* and *SMAD4* alterations are later events, mainly occurring in invasive PDAC (Hruban et al., 2000; Hosoda et al., 2017). Therefore, our finding could indicate that whereas *Kras* amplification and the loss of *Cdkn2a* seem to be essential in human and mice PDAC, *Trp53* and *Smad4* alterations wouldn't be required for a metastatic progression of PDAC in our mouse model. Moreover, *Cdkn2a* encodes p16<sup>INK4A</sup> and p19<sup>ARF</sup> (p14<sup>ARF</sup> in humans) proteins that act as tumor suppressors by regulating the cell cycle. Considering that p19ARF activates the *Trp53* tumor suppressor gene, perhaps the inactivation of the *Cdkn2a* in mice might have a similar effect as *P53* loss in humans. Lastly, *Smad4* deficiency has no discernible impact on normal pancreas development but it has been associated with increased proliferation of the tumor epithelium and the stromal tissue, accelerating PDAC development of *Kras*<sup>G12D</sup> mice (Bardeesy, Cheng, et al., 2006). The absence of alterations in *Smad4* and the presence of epithelial-mesenchymal transition (EMT) in our mice agree with published works revealing that pancreatic tumors with intact *Smad4* frequently exhibit EMT (Bardeesy, Cheng, et al., 2006). In particular, *SMAD4* plays a complex role in the transforming growth factor-beta (TGFβ) pathway, which is a potent inducer of EMT (Dardare et al., 2020; Massagué, 2008). Recently, it has been suggested that tumor cells undergo EMT through intermediary states that are characterized by greater aggressiveness and metastatic capacity. Indeed, TGFβ modulates immune regulation

and tumor microenvironment modification, processes that cancer cells might exploit to their advantage (Massagué, 2008). This fact plus the differences between mouse and human immune systems might explain the absence of *Smad4* mutations in our mice. Nevertheless, whether *SMAD4* loss has a real influence on the prognosis of the human disease is still controversial (Dardare et al., 2020).

Similar to previous studies, we didn't find a high number of recurrently mutated genes among different samples, but some of them could be new candidates for PDAC cancer genes. In this regard, some examples are *Ctnna3*, *Matr3*, or *Lrp1b*. We found different *Ctnna3* non-silent mutation in three mice. *Ctnna3* codifies a cadherin that is relevant in cell-cell adhesion (J. D. Smith et al., 2011). Therefore, its inactivation could contribute to a higher invasive characteristic in the cells. In this regard, *CTNNA3* has been already described as a tumor suppressor frequently mutated in laryngeal carcinomas (Fanjul-Fernández et al., 2013). Secondly, three different *Matr3* mutations were found in three mice, affecting essential splice positions which likely alter the correct mRNA splicing. *Matr3* encodes a nuclear matrix protein, which is involved in differentiation (Cha et al., 2021). *MATR3* has been recently described as a tumor suppressor in basal-like breast cancer (J. Yang et al., 2020). Thirdly, we found two different missense *Lrp1b* mutations in two mice. *Lrp1b* gene is a putative tumor suppressor proposed to be involved in extracellular signal transduction. Silencing and down-expression of *LRP1B* have been observed in renal cell carcinoma and thyroid cancer (Ni et al., 2013). Further deletion of *LRP1B* has been associated with chemotherapy resistance in high-grade serous cancers (Cowin et al., 2012). Notably, 5% of PDAC patients (7/162) harbor non-synonymous mutations in *LRP1B* according to data from The Cancer Genome Atlas (TCGA). Nevertheless, much work remains to be done in order to confirm that these genes could constitute new PDAC cancer genes.

Interestingly, although KFC mice already have a mutant version of *Kras* (*Kras*<sup>G12D</sup>), we have found *Kras* activating mutations (*Kras*<sup>Q61L</sup>) in the wild-type (WT) allele of two lung metastatic samples.

Next, we compared protein-coding mutations, CNA, and regions with loss of heterozygosity (LOH) among all the samples from each mouse to infer their evolutionary relationships. We found that all the samples from the same mouse shared a number of mutations. This result was in accordance with the labeling experiments, verifying the monofocal origin of the advanced tumors in our mice.

Following this strategy, we also found similar evolutionary processes in our mouse model to those observed in PDAC patients, such as different waves of metastatic spread and convergent evolution (Campbell et al., 2010; Yachida et al., 2010; Maddipati & Stanger, 2015). We found evidence of separate waves of metastases from the primary tumor to different tissues in three mice (ID-757, ID-799, and ID-2379) through the detection of a higher number of shared mutations between some metastatic samples and the primary tumor compared with other metastatic clones from the same tissue. These findings suggested continuous metastatic seeding similar to what was proposed in human PDAC tumors (Campbell et al., 2010). A similar example of different waves of metastases was revealed through the presence of shared CNA in chromosomes 4 and 6 in the primary and the liver metastases, but not in the lung metastases from one mouse (ID-2365). This observation indicates that lung metastatic spread occurred previous to the liver metastasis and it would suggest that specific alterations are needed to produce metastases in different organs. Furthermore, we observed convergent evolution through the identification of the *Lpxn* gene mutated in two independent manners in the lung metastasis from a single mouse (ID-799). This gene is a negative regulator in integrin-mediated cell adhesion processes and contributes to the regulation of cell adhesion, spreading, and cell migration.

Importantly, we found similar evidence of tissue-specific metastasis to those previously described in human PDAC (Campbell et al., 2010; Yachida et al., 2010; Maddipati & Stanger, 2015). In three mice we observed this process through the identification of specific mutations shared by the primary tumor and the lung metastases, and other specific mutations that were present in the primary and the liver metastasis (ID-788, ID-2365, ID-2343, and ID-2850). This tissue-specific metastasis was also revealed at the CNA level. Interestingly, we didn't recurrently find *Kras* amplification in the lung metastases, while it is present in the primary tumor and the liver metastases in most of the cases. These observations might indicate an early lung metastatic spread, which is in accordance with an early tumor cell dissemination proposed by some authors in pancreatic cancer, breast cancer, and melanoma (Rhim et al., 2012; Hüsemann et al., 2008; Eyles et al., 2010). It would also suggest that the amplification of *Kras* might not be essential for producing lung metastases, unlike the liver. So, there may be different possible mechanisms to produce the two types of metastases. Alternatively, it could be that the cells that produce the lung metastases are different from those generating the liver metastases.

In this regard, it has been proposed for decades that the tumor cell colonization is not a random process but that, instead, tumor cells have preferences when metastasizing to specific organs (Paget, 1989). This hypothesis is known as the 'seed and soil' hypothesis and it proposes that the spread of tumor cells depends on the interaction and cooperation between tumor cells (seed) and the host organ (soil). Furthermore, recent studies have described the development of premetastatic niches to facilitate the tumor cell colonization and growth through the accumulation of extracellular matrix proteins and aberrant immune cells at the target organ (Peinado et al., 2017). This favorable microenvironment is firstly promoted by the primary tumor which produces exosomes- and tumor-derived soluble factors prior to tumor dissemination (Peinado et al., 2011). Furthermore, evidence of different

requirements for the establishment of metastases in specific organs has been already provided in breast cancer (Minn et al., 2005). In this case, they identified some specific genes that mediate cancer metastasis to the lungs and not to the bones, the two most frequent metastatic targets of breast cancer.

Unfortunately, we didn't find any recurrent genetic alteration that could explain the metastatic potential. This observation is in agreement with recent studies in colorectal cancer (Ganesh et al., 2020) and suggests that the metastatic potential could not be determined by protein-coding mutations. So, we proposed to characterize the functional/transcriptomic ITH as any alteration, including the influence of the tumor microenvironment, is expected to affect the transcriptome of the cells and, as a consequence, the tumor progression and metastatic potential.

To further explore this question, we analyzed 60741 cells from primary and metastatic samples from 9 different mice with scRNA-seq experiments. We confirmed the presence of functional ITH revealed by the presence of different transcriptomic groups which corresponds with the different cell populations inside the tumors. We superficially characterized the immune and fibroblastic components of our primary tumors. In particular, we found three of the four CAF subpopulations already described in another PDAC mouse model and in human PDAC (Elyada et al., 2019), supporting again the resemblance of our cancer model to human patients. Due to the huge relevance of this component for tumor development, further analysis is required to study the interaction of this component with tumor cells, some such as the study of ligand-receptor co-expression in these cells.

However, in this Doctoral Thesis, we focused on the characterization of the presence of recurrent transcriptional profiles of the tumor cells, which also revealed functional ITH. Although we failed to characterize tumor cells from liver metastasis due to the high immune cell infiltration, interestingly, lung metastatic cells from different mice cluster together indicating similar transcriptional profiles.

Importantly, we found recurrent transcriptional subgroups of cancer cells among mice, suggesting a collaboration between the different transcriptional groups. Some of these recurrent subgroups overexpress gene pathways involved in tumor progression, such as translational initiation and ribosomal activity, cell junction and cell adhesion processes, or cell division. Among them, we also identified a minority subgroup halfway between the primary and the metastases, suggesting that it could be composed of pre-metastatic cancer cells. They overexpressed several ligands and receptors of the TGF $\beta$  pathway, such as *Cxcl2* and *Il1*, which could indicate that they communicate with the normal cells of the TME to promote angiogenesis and invasion of the primary pancreatic tumor and to prepare the metastatic niche in the lungs. Notably, the activated TGF $\beta$  receptor modulates transcription in association with *Smad4* through the phosphorylation of the *Smad2* and *Smad3* proteins, TGF $\beta$  ligands are commonly overexpressed in PDAC, promoting EMT and cancer cell invasion in cell lines (Nolan-Stevaux et al., 2009; Horiguchi et al., 2009). So, our observation is in accordance with published works demonstrating that TGF $\beta$  can also induce angiogenesis, activate tumor-promoting myCAF, and attenuate immune surveillance, creating a pro-tumorigenic TME (Kano et al., 2007; Hinz et al., 2007), although its potential to act as a tumor promoter or suppressor depends on its cross-talk with other pathways (H & K, 2010; Hezel et al., 2012; A. L. Smith et al., 2012).

Indeed, we observed an epithelial-mesenchymal transition (EMT) in the tumor cells, which agrees with what has been reported in many tumor types, including human PDAC where EMT has been reported to be essential for the metastatic dissemination of tumor cells (Rhim et al., 2012; Hotz et al., 2007). We identified two transcriptional programs of EMT: a partial EMT of the metastatic tumor cells and a complete EMT of the primary tumor cell populations. The partial EMT in two different lung metastases agrees with our previous suggestion that the metastatic mechanisms might be different among tissues. More studies of the intermediate stages are required in order to characterize the trajectories.

Finally, in an attempt to establish a relationship between the genetic and the transcriptomic subgroups we identified the presence of CNA from the scRNA-seq data. We recurrently found CNA in chromosomes 4 and 6, probably corresponding to the tumor-promoting alterations of *Cdkn2a* and *Kras*, respectively. Again, we did not find CNA in the chromosomes 11 and 18 that contain *Trp53* and *Smad4*, respectively, supporting our findings in the exome data. Nevertheless, we did not find any clear correlation between the genetic and transcriptomic subgroups or with the EMT. This observation indicates that the generation of recurrent transcriptomic subgroups in our mice is not associated with the acquisition of specific genetic alterations. Interestingly, we did not find CNA in the CAFs, suggesting that their origin is not a tumor cell, as it has been proposed by some authors (Yamaguchi et al., 2021). At this point, much work remains to be done in order to decipher the intrinsic mechanisms of PDAC tumor progression, being one of the major challenges to understanding how these specific transcriptomic changes are involved in PDAC cancer progression.







## Conclusions



1. Lineage tracing and genomic analysis show that advanced pancreatic tumors are monofocal in our mouse model.
2. We have constructed a set of fluorescent labeling tools that allows us to generate up to 15 different color combinations to allow the characterization and isolation of the different genetic clones present in our mouse pancreatic cancer model.
3. Mouse pancreatic tumors present intratumor heterogeneity as well as similar branched and convergent evolution phenomena described in human tumors.
4. Our mouse model presents recurrent genomic alterations in *Kras* and *Cdkn2a* but not in *Tp53* or *Smad4*.
5. We recurrently observed early lung metastatic spread, as well as tissue-specific metastasis of different genetic clones that are incompatible with a progressive mutation accumulation model for tumor progression.
6. Murine pancreatic tumors recurrently present alterations that activate the mutant *Kras* pathway and that are dispensable for lung metastasis.
7. Our mouse model present functional intratumor heterogeneity with the presence of recurrent transcriptionally differentiated cancer cell populations.
8. Lung metastatic cells present a distinct recurrent transcriptional program.
9. We observed at least two differentiated transcriptional programs of epithelial-mesenchymal transition (EMT), with lung metastatic populations presenting only partial EMT versus primary tumor cells.

10. Lung cancer metastatic program is characterized by the upregulation of *EGFR* pathway and the expression of ligands described to be involved in metastatic niche education.
11. Lung metastatic transcriptional program is already present in a minor population of pre-metastatic cells in primary tumors.







## References



- Aguirre, A. J., Bardeesy, N., Sinha, M., Lopez, L., Tuveson, D. A., Horner, J., Redston, M. S., & DePinho, R. A. (2003). Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. *Genes & Development, 17*(24), 3112–3126. <https://doi.org/10.1101/gad.1158703>
- Albers, C. A., Lunter, G., MacArthur, D. G., McVean, G., Ouwehand, W. H., & Durbin, R. (2011). Dindel: Accurate indel calls from short-read data. *Genome Research, 21*(6), 961–973. <https://doi.org/10.1101/gr.112326.110>
- Alles, J., Karaikos, N., Praktijnjo, S. D., Grosswendt, S., Wahle, P., Ruffault, P.-L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., Zinzen, R., Kocks, C., & Rajewsky, N. (2017). Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biology, 15*(1), 44. <https://doi.org/10.1186/s12915-017-0383-5>
- Almendro, V., Cheng, Y.-K., Randles, A., Itzkovitz, S., Marusyk, A., Ametller, E., Gonzalez-Farre, X., Muñoz, M., Russnes, H. G., Helland, A., Rye, I. H., Borresen-Dale, A.-L., Maruyama, R., van Oudenaarden, A., Dowsett, M., Jones, R. L., Reis-Filho, J., Gascon, P., Gönen, M., ... Polyak, K. (2014). Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Reports, 6*(3), 514–527. <https://doi.org/10.1016/j.celrep.2013.12.041>
- Ansari, A. M., Ahmed, A. K., Matsangos, A. E., Lay, F., Born, L. J., Marti, G., Harmon, J. W., & Sun, Z. (2016). Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments. *Stem Cell Reviews and Reports, 12*(5), 553–559. <https://doi.org/10.1007/s12015-016-9670-8>
- Aune, D., Greenwood, D. C., Chan, D. S. M., Vieira, R., Vieira, A. R., Navarro Rosenblatt, D. A., Cade, J. E., Burley, V. J., & Norat, T. (2012). Body mass index, abdominal fatness and pancreatic cancer risk: A systematic review and non-linear dose-response meta-analysis of prospective studies. *Annals of Oncology: Official Journal of the European Society for Medical Oncology, 23*(4), 843–852. <https://doi.org/10.1093/annonc/mdr398>
- Bailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A.-M., Gingras, M.-C., Miller, D. K., Christ, A. N., Bruxner, T. J. C., Quinn, M. C., Nourse, C., Murtaugh, L. C., Harliwong, I., Idrisoglu, S., Manning, S., Nourbakhsh, E., Wani, S., Fink, L., Holmes, O., ... Grimmond, S. M. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature, 531*(7592), 47–52. <https://doi.org/10.1038/nature16965>
- Bardeesy, N., Aguirre, A. J., Chu, G. C., Cheng, K.-H., Lopez, L. V., Hezel, A. F., Feng, B., Brennan, C., Weissleder, R., Mahmood, U., Hanahan, D., Redston, M. S., Chin, L., & Depinho, R. A. (2006). Both p16(Ink4a) and the p19(Arf)-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proceedings of the National Academy of Sciences of the United States of America, 103*(15), 5947–5952. <https://doi.org/10.1073/pnas.0601273103>

Bardeesy, N., Cheng, K.-H., Berger, J. H., Chu, G. C., Pahler, J., Olson, P., Hezel, A. F., Horner, J., Lauwers, G. Y., Hanahan, D., & DePinho, R. A. (2006). Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. *Genes & Development*, *20*(22), 3130–3146. <https://doi.org/10.1101/gad.1478706>

Bashashati, A., Ha, G., Tone, A., Ding, J., Prentice, L. M., Roth, A., Rosner, J., Shumansky, K., Kalloger, S., Senz, J., Yang, W., McConechy, M., Melnyk, N., Anglesio, M., Luk, M. T. Y., Tse, K., Zeng, T., Moore, R., Zhao, Y., ... Shah, S. P. (2013). Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *The Journal of Pathology*, *231*(1), 21–34. <https://doi.org/10.1002/path.4230>

Baslan, T., & Hicks, J. (2017). Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nature Reviews. Cancer*, *17*(9), 557–569. <https://doi.org/10.1038/nrc.2017.58>

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., & Newell, E. W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*. <https://doi.org/10.1038/nbt.4314>

Ben, Q., Xu, M., Ning, X., Liu, J., Hong, S., Huang, W., Zhang, H., & Li, Z. (2011). Diabetes mellitus and risk of pancreatic cancer: A meta-analysis of cohort studies. *European Journal of Cancer (Oxford, England: 1990)*, *47*(13), 1928–1937. <https://doi.org/10.1016/j.ejca.2011.03.003>

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Cheetham, R. K., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., ... Smith, A. J. (2008). Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature*, *456*(7218), 53–59. <https://doi.org/10.1038/nature07517>

Biankin, A. V., Waddell, N., Kassahn, K. S., Gingras, M.-C., Muthuswamy, L. B., Johns, A. L., Miller, D. K., Wilson, P. J., Patch, A.-M., Wu, J., Chang, D. K., Cowley, M. J., Gardiner, B. B., Song, S., Harliwong, I., Idrisoglu, S., Nourse, C., Nourbakhsh, E., Manning, S., ... Grimmond, S. M. (2012). Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*, *491*(7424), 399–405. <https://doi.org/10.1038/nature11547>

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., & Barillot, E. (2012). Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics (Oxford, England)*, *28*(3), 423–425. <https://doi.org/10.1093/bioinformatics/btr670>

Bosetti, C., Lucenteforte, E., Silverman, D. T., Petersen, G., Bracci, P. M., Ji, B. T., Negri, E., Li, D., Risch, H. A., Olson, S. H., Gallinger, S., Miller, A. B., Bueno-de-Mesquita, H. B., Talamini, R., Polesel, J., Ghadirian, P., Baghurst, P. A., Zatonski, W., Fontham, E., ... La Vecchia, C. (2012). Cigarette smoking and pancreatic cancer: An analysis from the International

Pancreatic Cancer Case-Control Consortium (Panc4). *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 23(7), 1880–1888. <https://doi.org/10.1093/annonc/mdr541>

Boyle, A. P., Guinney, J., Crawford, G. E., & Furey, T. S. (2008). F-Seq: A feature density estimator for high-throughput sequence tags. *Bioinformatics*, 24(21), 2537–2538. <https://doi.org/10.1093/bioinformatics/btn480>

Breast Cancer Linkage Consortium. (1999). Cancer risks in BRCA2 mutation carriers. *Journal of the National Cancer Institute*, 91(15), 1310–1316. <https://doi.org/10.1093/jnci/91.15.1310>

Briggs, A. W. (2011). Rapid retrieval of DNA target sequences by primer extension capture. *Methods in Molecular Biology (Clifton, N.J.)*, 772, 145–154. [https://doi.org/10.1007/978-1-61779-228-1\\_8](https://doi.org/10.1007/978-1-61779-228-1_8)

Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica Et Biophysica Acta*, 1842(10), 1932–1941. <https://doi.org/10.1016/j.bbadis.2014.06.015>

Burriss, H. A., Moore, M. J., Andersen, J., Green, M. R., Rothenberg, M. L., Modiano, M. R., Cripps, M. C., Portenoy, R. K., Storniolo, A. M., Tarassoff, P., Nelson, R., Dorr, F. A., Stephens, C. D., & Von Hoff, D. D. (1997). Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: A randomized trial. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 15(6), 2403–2413. <https://doi.org/10.1200/JCO.1997.15.6.2403>

Cai, D., Cohen, K. B., Luo, T., Lichtman, J. W., & Sanes, J. R. (2013). Improved tools for the Brainbow toolbox. *Nature Methods*, 10(6), 540–547. <https://doi.org/10.1038/nmeth.2450>

Campbell, P. J., Pleasance, E. D., Stephens, P. J., Dicks, E., Rance, R., Goodhead, I., Follows, G. A., Green, A. R., Futreal, P. A., & Stratton, M. R. (2008). Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 105(35), 13081–13086. <https://doi.org/10.1073/pnas.0801523105>

Campbell, P. J., Yachida, S., Mudie, L. J., Stephens, P. J., Pleasance, E. D., Stebbings, L. A., Morsberger, L. A., Latimer, C., McLaren, S., Lin, M.-L., McBride, D. J., Varela, I., Nik-Zainal, S. A., Leroy, C., Jia, M., Menzies, A., Butler, A. P., Teague, J. W., Griffin, C. A., ... Futreal, P. A. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*, 467(7319), 1109–1113. <https://doi.org/10.1038/nature09460>

Cao, J., Packer, J. S., Ramani, V., Cusanovich, D. A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S. N., Steemers, F. J., Adey, A., Waterston, R. H., Trapnell, C., & Shendure, J. (2017).

Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science (New York, N.Y.)*, 357(6352), 661–667. <https://doi.org/10.1126/science.aam8940>

Carrière, C., Young, A. L., Gunn, J. R., Longnecker, D. S., & Korc, M. (2009). Acute pancreatitis markedly accelerates pancreatic cancer progression in mice expressing oncogenic Kras. *Biochemical and Biophysical Research Communications*, 382(3), 561–565. <https://doi.org/10.1016/j.bbrc.2009.03.068>

Carrière, C., Young, A. L., Gunn, J. R., Longnecker, D. S., & Korc, M. (2011). Acute pancreatitis accelerates initiation and progression to pancreatic cancer in mice expressing oncogenic Kras in the nestin cell lineage. *PLoS One*, 6(11), e27725. <https://doi.org/10.1371/journal.pone.0027725>

Cha, H. J., Uyan, Ö., Kai, Y., Liu, T., Zhu, Q., Tothova, Z., Botten, G. A., Xu, J., Yuan, G.-C., Dekker, J., & Orkin, S. H. (2021). Inner nuclear protein Matrin-3 coordinates cell differentiation by stabilizing chromatin architecture. *Nature Communications*, 12(1), 6241. <https://doi.org/10.1038/s41467-021-26574-4>

Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L., & Mardis, E. R. (2009). BreakDancer: An algorithm for high-resolution mapping of genomic structural variation. *Nature Methods*, 6(9), 677–681. <https://doi.org/10.1038/nmeth.1363>

Chidgeavdze, Z. G., Beabealashvili, R. S., Atrazhev, A. M., Kukhanova, M. K., Azhayev, A. V., & Krayevsky, A. A. (1984). 2',3'-Dideoxy-3' aminonucleoside 5'-triphosphates are the terminators of DNA synthesis catalyzed by DNA polymerases. *Nucleic Acids Research*, 12(3), 1671–1686. <https://doi.org/10.1093/nar/12.3.1671>

Christoforides, A., Carpten, J. D., Weiss, G. J., Demeure, M. J., Von Hoff, D. D., & Craig, D. W. (2013). Identification of somatic mutations in cancer through Bayesian-based analysis of sequenced genome pairs. *BMC Genomics*, 14, 302. <https://doi.org/10.1186/1471-2164-14-302>

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E. S., & Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, 31(3), 213–219. <https://doi.org/10.1038/nbt.2514>

Cid-Arregui, A., & Juarez, V. (2015). Perspectives in the treatment of pancreatic adenocarcinoma. *World Journal of Gastroenterology*, 21(31), 9297–9316. <https://doi.org/10.3748/wjg.v21.i31.9297>

Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., & Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*, 4(4), 265–270. <https://doi.org/10.1038/nnano.2009.12>

- Cleary, A. S., Leonard, T. L., Gestl, S. A., & Gunther, E. J. (2014). Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature*, *508*(7494), 113–117. <https://doi.org/10.1038/nature13187>
- Collisson, E. A., Bailey, P., Chang, D. K., & Biankin, A. V. (2019). Molecular subtypes of pancreatic cancer. *Nature Reviews. Gastroenterology & Hepatology*, *16*(4), 207–220. <https://doi.org/10.1038/s41575-019-0109-y>
- Collisson, E. A., Sadanandam, A., Olson, P., Gibb, W. J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G. E., Jakkula, L., Feiler, H. S., Ko, A. H., Olshen, A. B., Danenberg, K. L., Tempero, M. A., Spellman, P. T., Hanahan, D., & Gray, J. W. (2011). Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature Medicine*, *17*(4), 500–503. <https://doi.org/10.1038/nm.2344>
- Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Gourgou-Bourgade, S., de la Fouchardière, C., Bennouna, J., Bachet, J.-B., Khemissa-Akouz, F., Péré-Vergé, D., Delbaldo, C., Assenat, E., Chauffert, B., Michel, P., Montoto-Grillot, C., ... PRODIGE Intergroup. (2011). FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *The New England Journal of Medicine*, *364*(19), 1817–1825. <https://doi.org/10.1056/NEJMoa1011923>
- Costa-Silva, B., Aiello, N. M., Ocean, A. J., Singh, S., Zhang, H., Thakur, B. K., Becker, A., Hoshino, A., Mark, M. T., Molina, H., Xiang, J., Zhang, T., Theilen, T.-M., García-Santos, G., Williams, C., Ararso, Y., Huang, Y., Rodrigues, G., Shen, T.-L., ... Lyden, D. C. (2015). Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nature Cell Biology*, *17*(6), 816–826. <https://doi.org/10.1038/ncb3169>
- Cowin, P. A., George, J., Fereday, S., Loehrer, E., Van Loo, P., Cullinane, C., Etemadmoghadam, D., Ftouni, S., Galletta, L., Anglesio, M. S., Hendley, J., Bowes, L., Sheppard, K. E., Christie, E. L., Pearson, R. B., Harnett, P. R., Heinzlmann-Schwarz, V., Friedlander, M., McNally, O., ... Australian Ovarian Cancer Study. (2012). LRP1B deletion in high-grade serous ovarian cancers is associated with acquired chemotherapy resistance to liposomal doxorubicin. *Cancer Research*, *72*(16), 4060–4073. <https://doi.org/10.1158/0008-5472.CAN-12-0203>
- Dardare, J., Witz, A., Merlin, J.-L., Gilson, P., & Harlé, A. (2020). SMAD4 and the TGF $\beta$  Pathway in Patients with Pancreatic Ductal Adenocarcinoma. *International Journal of Molecular Sciences*, *21*(10), 3534. <https://doi.org/10.3390/ijms21103534>
- Day, C.-P., Carter, J., Weaver Ohler, Z., Bonomi, C., El Meskini, R., Martin, P., Graff-Cherry, C., Feigenbaum, L., Tüting, T., Van Dyke, T., Hollingshead, M., & Merlino, G. (2014). 'Glowing head' mice: A genetic tool enabling reliable preclinical image-based evaluation of cancers in immunocompetent allografts. *PloS One*, *9*(11), e109956. <https://doi.org/10.1371/journal.pone.0109956>

de Bruin, E. C., McGranahan, N., Mitter, R., Salm, M., Wedge, D. C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A. J., Grönroos, E., Muhammad, M. A., Horswell, S., Gerlinger, M., Varela, I., Jones, D., Marshall, J., Voet, T., Van Loo, P., ... Swanton, C. (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science (New York, N.Y.)*, *346*(6206), 251–256. <https://doi.org/10.1126/science.1253462>

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., & Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*, *122*(3), 473–483. <https://doi.org/10.1016/j.cell.2005.07.013>

Down, T. A., Rakyán, V. K., Turner, D. J., Flicek, P., Li, H., Kulesha, E., Gräf, S., Johnson, N., Herrero, J., Tomazou, E. M., Thorne, N. P., Bäckdahl, L., Herberth, M., Howe, K. L., Jackson, D. K., Miretti, M. M., Marioni, J. C., Birney, E., Hubbard, T. J. P., ... Beck, S. (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology*, *26*(7), 779–785. <https://doi.org/10.1038/nbt1414>

Duell, E. J., Lucenteforte, E., Olson, S. H., Bracci, P. M., Li, D., Risch, H. A., Silverman, D. T., Ji, B. T., Gallinger, S., Holly, E. A., Fontham, E. H., Maisonneuve, P., Bueno-de-Mesquita, H. B., Ghadirian, P., Kurtz, R. C., Ludwig, E., Yu, H., Lowenfels, A. B., Seminara, D., ... Boffetta, P. (2012). Pancreatitis and pancreatic cancer risk: A pooled analysis in the International Pancreatic Cancer Case-Control Consortium (PanC4). *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, *23*(11), 2964–2970. <https://doi.org/10.1093/annonc/mds140>

Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G., & Jenkins, N. A. (2005). Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature*, *436*(7048), 221–226. <https://doi.org/10.1038/nature03691>

Elyada, E., Bolisetty, M., Laise, P., Flynn, W. F., Courtois, E. T., Burkhart, R. A., Teinor, J. A., Belleau, P., Biffi, G., Lucito, M. S., Sivajothi, S., Armstrong, T. D., Engle, D. D., Yu, K. H., Hao, Y., Wolfgang, C. L., Park, Y., Preall, J., Jaffee, E. M., ... Tuveson, D. A. (2019). Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discovery*, *9*(8), 1102–1123. <https://doi.org/10.1158/2159-8290.CD-19-0094>

Eyles, J., Puaux, A.-L., Wang, X., Toh, B., Prakash, C., Hong, M., Tan, T. G., Zheng, L., Ong, L. C., Jin, Y., Kato, M., Prévost-Blondel, A., Chow, P., Yang, H., & Abastado, J.-P. (2010). Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. *The Journal of Clinical Investigation*, *120*(6), 2030–2039. <https://doi.org/10.1172/JCI42002>

Fanjul-Fernández, M., Quesada, V., Cabanillas, R., Cadiñanos, J., Fontanil, T., Obaya, A., Ramsay, A. J., Llorente, J. L., Astudillo, A., Cal, S., & López-Otín, C. (2013). Cell-cell adhesion genes CTNNA2 and CTNNA3 are tumour suppressors frequently mutated in laryngeal carcinomas. *Nature Communications*, *4*, 2531. <https://doi.org/10.1038/ncomms3531>



Fejes, A. P., Robertson, G., Bilenky, M., Varhol, R., Bainbridge, M., & Jones, S. J. M. (2008). FindPeaks 3.1: A tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics (Oxford, England)*, *24*(15), 1729–1730. <https://doi.org/10.1093/bioinformatics/btn305>

Fisher, R., Puzstai, L., & Swanton, C. (2013). Cancer heterogeneity: Implications for targeted therapeutics. *British Journal of Cancer*, *108*(3), 479–485. <https://doi.org/10.1038/bjc.2012.581>

Ganesh, K., Basnet, H., Kaygusuz, Y., Laughney, A. M., He, L., Sharma, R., O'Rourke, K. P., Reuter, V. P., Huang, Y.-H., Turkecul, M., Emrah, E., Masilionis, I., Manova-Todorova, K., Weiser, M. R., Saltz, L. B., Garcia-Aguilar, J., Koche, R., Lowe, S. W., Pe'er, D., ... Massagué, J. (2020). L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer. *Nature Cancer*, *1*(1), 28–45. <https://doi.org/10.1038/s43018-019-0006-x>

Gannon, M., Herrera, P.-L., & Wright, C. V. E. (2000). Mosaic Cre-mediated recombination in pancreas using the pdx-1 enhancer/promoter. *Genesis*, *26*(2), 143–144. [https://doi.org/10.1002/\(SICI\)1526-968X\(200002\)26:2<143::AID-GENE13>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1526-968X(200002)26:2<143::AID-GENE13>3.0.CO;2-L)

García-Marqués, J., & López-Mascaraque, L. (2013). Clonal identity determines astrocyte cortical heterogeneity. *Cerebral Cortex (New York, N.Y.: 1991)*, *23*(6), 1463–1472. <https://doi.org/10.1093/cercor/bhs134>

Gerlinger, M., Horswell, S., Larkin, J., Rowan, A. J., Salm, M. P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C. R., Martinez, P., Phillimore, B., Begum, S., Rabinowitz, A., Spencer-Dene, B., Gulati, S., Bates, P. A., Stamp, G., Pickering, L., ... Swanton, C. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature Genetics*, *46*(3), 225–233. <https://doi.org/10.1038/ng.2891>

Gerlinger, M., Rowan, A. J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N. Q., Butler, A., Jones, D., Raine, K., Latimer, C., ... Swanton, C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England Journal of Medicine*, *366*(10), 883–892. <https://doi.org/10.1056/NEJMoa1113205>

Giardiello, F. M., Brensinger, J. D., Tersmette, A. C., Goodman, S. N., Petersen, G. M., Booker, S. V., Cruz-Correa, M., & Offerhaus, J. A. (2000). Very high risk of cancer in familial Peutz-Jeghers syndrome. *Gastroenterology*, *119*(6), 1447–1453. <https://doi.org/10.1053/gast.2000.20228>

Golan, T., Hammel, P., Reni, M., Van Cutsem, E., Macarulla, T., Hall, M. J., Park, J.-O., Hochhauser, D., Arnold, D., Oh, D.-Y., Reinacher-Schick, A., Tortora, G., Algül, H., O'Reilly, E. M., McGuinness, D., Cui, K. Y., Schlienger, K., Locker, G. Y., & Kindler, H. L. (2019). Maintenance Olaparib for Germline BRCA-Mutated Metastatic Pancreatic Cancer. *The New England Journal of Medicine*, *381*(4), 317–327. <https://doi.org/10.1056/NEJMoa1903387>

González-Silva, L., Quevedo, L., & Varela, I. (2021). Tumor Functional Heterogeneity Unraveled by scRNA-seq Technologies: (Trends in Cancer 6, 13-19, 2020). *Trends in Cancer*, 7(3), 265. <https://doi.org/10.1016/j.trecan.2021.02.001>

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>

Grindberg, R. V., Yee-Greenbaum, J. L., McConnell, M. J., Novotny, M., O'Shaughnessy, A. L., Lambert, G. M., Araúzo-Bravo, M. J., Lee, J., Fishman, M., Robbins, G. E., Lin, X., Venepally, P., Badger, J. H., Galbraith, D. W., Gage, F. H., & Lasken, R. S. (2013). RNA-sequencing from single nuclei. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49), 19802–19807. <https://doi.org/10.1073/pnas.1319700110>

Guillaumet-Adkins, A., Rodríguez-Esteban, G., Mereu, E., Mendez-Lago, M., Jaitin, D. A., Villanueva, A., Vidal, A., Martínez-Martí, A., Felip, E., Vivancos, A., Keren-Shaul, H., Heath, S., Gut, M., Amit, I., Gut, I., & Heyn, H. (2017). Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biology*, 18(1), 45. <https://doi.org/10.1186/s13059-017-1171-9>

Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L. B., Tubio, J. M. C., Papaemmanuil, E., Brewer, D. S., Kallio, H. M. L., Högnäs, G., Annala, M., Kivinummi, K., Goody, V., Latimer, C., O'Meara, S., Dawson, K. J., Isaacs, W., Emmert-Buck, M. R., Nykter, M., Foster, C., ... Bova, G. S. (2015). The evolutionary history of lethal metastatic prostate cancer. *Nature*, 520(7547), 353–357. <https://doi.org/10.1038/nature14347>

Guo, X., Zhang, Y., Zheng, L., Zheng, C., Song, J., Zhang, Q., Kang, B., Liu, Z., Jin, L., Xing, R., Gao, R., Zhang, L., Dong, M., Hu, X., Ren, X., Kirchhoff, D., Roider, H. G., Yan, T., & Zhang, Z. (2018). Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nature Medicine*, 24(7), 978–985. <https://doi.org/10.1038/s41591-018-0045-3>

H, I., & K, M. (2010). TGFbeta signalling: A complex web in cancer progression. *Nature Reviews. Cancer*, 10(6). <https://doi.org/10.1038/nrc2853>

Hahn, S. A., Schutte, M., Hoque, A. T., Moskaluk, C. A., da Costa, L. T., Rozenblum, E., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H., & Kern, S. E. (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science (New York, N.Y.)*, 271(5247), 350–353. <https://doi.org/10.1126/science.271.5247.350>

Hashimshony, T., Wagner, F., Sher, N., & Yanai, I. (2012). CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Reports*, 2(3), 666–673. <https://doi.org/10.1016/j.celrep.2012.08.003>

Hezel, A. F., Deshpande, V., Zimmerman, S. M., Contino, G., Alagesan, B., O'Dell, M. R., Rivera, L. B., Harper, J., Lonning, S., Brekken, R. A., & Bardeesy, N. (2012). TGF-β and αvβ6

integrin act in a common pathway to suppress pancreatic cancer progression. *Cancer Research*, 72(18), 4840–4845. <https://doi.org/10.1158/0008-5472.CAN-12-0634>

Hezel, A. F., Kimmelman, A. C., Stanger, B. Z., Bardeesy, N., & Depinho, R. A. (2006). Genetics and biology of pancreatic ductal adenocarcinoma. *Genes & Development*, 20(10), 1218–1249. <https://doi.org/10.1101/gad.1415606>

Hiley, C., de Bruin, E. C., McGranahan, N., & Swanton, C. (2014). Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. *Genome Biology*, 15(8), 453. <https://doi.org/10.1186/s13059-014-0453-8>

Hingorani, S. R., Petricoin, E. F., Maitra, A., Rajapakse, V., King, C., Jacobetz, M. A., Ross, S., Conrads, T. P., Veenstra, T. D., Hitt, B. A., Kawaguchi, Y., Johann, D., Liotta, L. A., Crawford, H. C., Putt, M. E., Jacks, T., Wright, C. V. E., Hruban, R. H., Lowy, A. M., & Tuveson, D. A. (2003). Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell*, 4(6), 437–450. [https://doi.org/10.1016/s1535-6108\(03\)00309-x](https://doi.org/10.1016/s1535-6108(03)00309-x)

Hingorani, S. R., Wang, L., Multani, A. S., Combs, C., Deramaudt, T. B., Hruban, R. H., Rustgi, A. K., Chang, S., & Tuveson, D. A. (2005). Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell*, 7(5), 469–483. <https://doi.org/10.1016/j.ccr.2005.04.023>

Hinz, S., Pagerols-Raluy, L., Oberg, H.-H., Ammerpohl, O., Grüssel, S., Sipos, B., Grützmann, R., Pilarsky, C., Ungefroren, H., Saeger, H.-D., Klöppel, G., Kabelitz, D., & Kalthoff, H. (2007). Foxp3 expression in pancreatic carcinoma cells as a novel mechanism of immune evasion in cancer. *Cancer Research*, 67(17), 8344–8350. <https://doi.org/10.1158/0008-5472.CAN-06-3304>

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., Middle, C. M., Rodesch, M. J., Albert, T. J., Hannon, G. J., & McCombie, W. R. (2007). Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, 39(12), 1522–1527. <https://doi.org/10.1038/ng.2007.42>

Hong, M. K. H., Macintyre, G., Wedge, D. C., Van Loo, P., Patel, K., Lunke, S., Alexandrov, L. B., Sloggett, C., Cmero, M., Marass, F., Tsui, D., Mangiola, S., Lonie, A., Naeem, H., Sapre, N., Phal, P. M., Kurganovs, N., Chin, X., Kerger, M., ... Hovens, C. M. (2015). Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nature Communications*, 6, 6605. <https://doi.org/10.1038/ncomms7605>

Horiguchi, K., Shirakihara, T., Nakano, A., Imamura, T., Miyazono, K., & Saitoh, M. (2009). Role of Ras signaling in the induction of snail by transforming growth factor-beta. *The Journal of Biological Chemistry*, 284(1), 245–253. <https://doi.org/10.1074/jbc.M804777200>

Hosoda, W., Chianchiano, P., Griffin, J. F., Pittman, M. E., Brosens, L. A., Noë, M., Yu, J., Shindo, K., Suenaga, M., Rezaee, N., Yonescu, R., Ning, Y., Albores-Saavedra, J., Yoshizawa, N., Harada, K., Yoshizawa, A., Hanada, K., Yonehara, S., Shimizu, M., ... Wood, L. D. (2017).

Genetic analyses of isolated high-grade pancreatic intraepithelial neoplasia (HG-PanIN) reveal paucity of alterations in TP53 and SMAD4. *The Journal of Pathology*, 242(1), 16–23. <https://doi.org/10.1002/path.4884>

Hotz, B., Arndt, M., Dullat, S., Bhargava, S., Buhr, H.-J., & Hotz, H. G. (2007). Epithelial to mesenchymal transition: Expression of the regulators snail, slug, and twist in pancreatic cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 13(16), 4769–4776. <https://doi.org/10.1158/1078-0432.CCR-06-2926>

Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y., & Peng, J. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Research*, 26(3), 304–319. <https://doi.org/10.1038/cr.2016.23>

Hruban, R. H., Goggins, M., Parsons, J., & Kern, S. E. (2000). Progression model for pancreatic cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 6(8), 2969–2972.

Hüsemann, Y., Geigl, J. B., Schubert, F., Musiani, P., Meyer, M., Burghart, E., Forni, G., Eils, R., Fehm, T., Riethmüller, G., & Klein, C. A. (2008). Systemic spread is an early step in breast cancer. *Cancer Cell*, 13(1), 58–68. <https://doi.org/10.1016/j.ccr.2007.12.003>

Iacobuzio-Donahue, C. A. (2012). Genetic evolution of pancreatic cancer: Lessons learnt from the pancreatic cancer genome sequencing project. *Gut*, 61(7), 1085–1094. <https://doi.org/10.1136/gut.2010.236026>

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnerberg, P., & Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*, 21(7), 1160–1167. <https://doi.org/10.1101/gr.110882.110>

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., & Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*, 11(2), 163–166. <https://doi.org/10.1038/nmeth.2772>

Jackson, E. L., Willis, N., Mercer, K., Bronson, R. T., Crowley, D., Montoya, R., Jacks, T., & Tuveson, D. A. (2001). Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes & Development*, 15(24), 3243–3248. <https://doi.org/10.1101/gad.943001>

Jamal-Hanjani, M., Wilson, G. A., McGranahan, N., Birkbak, N. J., Watkins, T. B. K., Veeriah, S., Shafi, S., Johnson, D. H., Mitter, R., Rosenthal, R., Salm, M., Horswell, S., Escudero, M., Matthews, N., Rowan, A., Chambers, T., Moore, D. A., Turajlic, S., Xu, H., ... TRACERx Consortium. (2017). Tracking the Evolution of Non-Small-Cell Lung Cancer. *The New England Journal of Medicine*, 376(22), 2109–2121. <https://doi.org/10.1056/NEJMoa1616288>

Jiang, Y., Oldridge, D. A., Diskin, S. J., & Zhang, N. R. (2015). CODEX: A normalization and copy number variation detection method for whole exome sequencing. *Nucleic Acids Research*, *43*(6), e39. <https://doi.org/10.1093/nar/gku1363>

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)*, *316*(5830), 1497–1502. <https://doi.org/10.1126/science.1141319>

Jones, S., Zhang, X., Parsons, D. W., Lin, J. C.-H., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., Hong, S.-M., Fu, B., Lin, M.-T., Calhoun, E. S., Kamiyama, M., Walter, K., Nikolskaya, T., Nikolsky, Y., Hartigan, J., ... Kinzler, K. W. (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science (New York, N.Y.)*, *321*(5897), 1801–1806. <https://doi.org/10.1126/science.1164368>

Kano, M. R., Bae, Y., Iwata, C., Morishita, Y., Yashiro, M., Oka, M., Fujii, T., Komuro, A., Kiyono, K., Kaminishi, M., Hirakawa, K., Ouchi, Y., Nishiyama, N., Kataoka, K., & Miyazono, K. (2007). Improvement of cancer-targeting therapy, using nanocarriers for intractable solid tumors by inhibition of TGF-beta signaling. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3460–3465. <https://doi.org/10.1073/pnas.0611660104>

Kastrinos, F., Mukherjee, B., Tayob, N., Wang, F., Sparr, J., Raymond, V. M., Bandipalliam, P., Stoffel, E. M., Gruber, S. B., & Syngal, S. (2009). Risk of pancreatic cancer in families with Lynch syndrome. *JAMA*, *302*(16), 1790–1795. <https://doi.org/10.1001/jama.2009.1529>

Kawaguchi, Y., Cooper, B., Gannon, M., Ray, M., MacDonald, R. J., & Wright, C. V. E. (2002). The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. *Nature Genetics*, *32*(1), 128–134. <https://doi.org/10.1038/ng959>

Kchouk, M., Gibrat, J., & Elloumi, M. (2017). *Generations of Sequencing Technologies: From First to Next Generation*. <https://doi.org/10.4172/0974-8369.1000395>

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), R36. <https://doi.org/10.1186/gb-2013-14-4-r36>

Kim, M.-Y., Oskarsson, T., Acharyya, S., Nguyen, D. X., Zhang, X. H.-F., Norton, L., & Massagué, J. (2009). Tumor self-seeding by circulating cancer cells. *Cell*, *139*(7), 1315–1326. <https://doi.org/10.1016/j.cell.2009.11.025>

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing—Concepts and limitations. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *32*(6), 524–536. <https://doi.org/10.1002/bies.200900181>

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A., & Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, *161*(5), 1187–1201. <https://doi.org/10.1016/j.cell.2015.04.044>

Klein, A. P., Brune, K. A., Petersen, G. M., Goggins, M., Tersmette, A. C., Offerhaus, G. J. A., Griffin, C., Cameron, J. L., Yeo, C. J., Kern, S., & Hruban, R. H. (2004). Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. *Cancer Research*, *64*(7), 2634–2638. <https://doi.org/10.1158/0008-5472.can-03-3823>

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., & Wilson, R. K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, *22*(3), 568–576. <https://doi.org/10.1101/gr.129684.111>

Korbel, J. O., & Campbell, P. J. (2013). Criteria for inference of chromothripsis in cancer genomes. *Cell*, *152*(6), 1226–1236. <https://doi.org/10.1016/j.cell.2013.02.023>

Kuilman, T., Velds, A., Kemper, K., Ranzani, M., Bombardelli, L., Hoogstraat, M., Nevedomskaya, E., Xu, G., de Rooter, J., Lolkema, M. P., Ylstra, B., Jonkers, J., Rottenberg, S., Wessels, L. F., Adams, D. J., Peeper, D. S., & Krijgsman, O. (2015). CopywriteR: DNA copy number detection from off-target sequence data. *Genome Biology*, *16*, 49. <https://doi.org/10.1186/s13059-015-0617-1>

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., ... International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. <https://doi.org/10.1038/35057062>

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>

Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., Gnirke, A., & Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods*, *7*(9), 709–715. <https://doi.org/10.1038/nmeth.1491>

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

Li, H., Ruan, J., & Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, *18*(11), 1851–1858. <https://doi.org/10.1101/gr.078212.108>

Li, R., Li, Y., Kristiansen, K., & Wang, J. (2008). SOAP: Short oligonucleotide alignment program. *Bioinformatics (Oxford, England)*, *24*(5), 713–714. <https://doi.org/10.1093/bioinformatics/btn025>

Li, Y., & Tollefsbol, T. O. (2011). DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods in Molecular Biology (Clifton, N.J.)*, *791*, 11–21. [https://doi.org/10.1007/978-1-61779-316-5\\_2](https://doi.org/10.1007/978-1-61779-316-5_2)

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, *326*(5950), 289–293. <https://doi.org/10.1126/science.1181369>

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology*, *2012*, 251364. <https://doi.org/10.1155/2012/251364>

Liu, T., Zhou, L., Li, D., Andl, T., & Zhang, Y. (2019). Cancer-Associated Fibroblasts Build and Secure the Tumor Microenvironment. *Frontiers in Cell and Developmental Biology*, *7*, 60. <https://doi.org/10.3389/fcell.2019.00060>

Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., Sanes, J. R., & Lichtman, J. W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, *450*(7166), 56–62. <https://doi.org/10.1038/nature06293>

Loulier, K., Barry, R., Mahou, P., Le Franc, Y., Supatto, W., Matho, K. S., Ieng, S., Fouquet, S., Dupin, E., Benosman, R., Chédotal, A., Beaurepaire, E., Morin, X., & Livet, J. (2014). Multiplex cell and lineage tracking with combinatorial labels. *Neuron*, *81*(3), 505–520. <https://doi.org/10.1016/j.neuron.2013.12.016>

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>

Lowenfels, A. B., Maisonneuve, P., DiMagno, E. P., Elitsur, Y., Gates, L. K., Perrault, J., & Whitcomb, D. C. (1997). Hereditary pancreatitis and the risk of pancreatic cancer. International Hereditary Pancreatitis Study Group. *Journal of the National Cancer Institute*, *89*(6), 442–446. <https://doi.org/10.1093/jnci/89.6.442>

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, *14*(5), 265–279. <https://doi.org/10.1016/j.gpb.2016.05.004>

Maaten, L. van der, & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(86), 2579–2605.

Macosko, E. Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, *161*(5), 1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002>

Maddipati, R., & Stanger, B. Z. (2015). Pancreatic Cancer Metastases Harbor Evidence of Polyclonality. *Cancer Discovery*, *5*(10), 1086–1097. <https://doi.org/10.1158/2159-8290.CD-15-0120>

Magee, J. A., Piskounova, E., & Morrison, S. J. (2012). Cancer stem cells: Impact, heterogeneity, and uncertainty. *Cancer Cell*, *21*(3), 283–296. <https://doi.org/10.1016/j.ccr.2012.03.003>

Maitra, A., Fukushima, N., Takaori, K., & Hruban, R. H. (2005). Precursors to invasive pancreatic cancer. *Advances in Anatomic Pathology*, *12*(2), 81–91. <https://doi.org/10.1097/01.pap.0000155055.14238.25>

Makohon-Moore, A. P., Zhang, M., Reiter, J. G., Bozic, I., Allen, B., Kundu, D., Chatterjee, K., Wong, F., Jiao, Y., Kohutek, Z. A., Hong, J., Attiyeh, M., Javier, B., Wood, L. D., Hruban, R. H., Nowak, M. A., Papadopoulos, N., Kinzler, K. W., Vogelstein, B., & Iacobuzio-Donahue, C. A. (2017). Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nature Genetics*, *49*(3), 358–366. <https://doi.org/10.1038/ng.3764>

Malide, D., Métais, J.-Y., & Dunbar, C. E. (2012). Dynamic clonal analysis of murine hematopoietic stem and progenitor cells marked by 5 fluorescent proteins using confocal and multiphoton microscopy. *Blood*, *120*(26), e105-116. <https://doi.org/10.1182/blood-2012-06-440636>

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, *7*(2), 111–118. <https://doi.org/10.1038/nmeth.1419>



- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bembien, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Ho, C. H., ... Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376–380. <https://doi.org/10.1038/nature03959>
- Martelotto, L. G., Baslan, T., Kendall, J., Geyer, F. C., Burke, K. A., Spraggon, L., Piscuoglio, S., Chadalavada, K., Nanjangud, G., Ng, C. K. Y., Moody, P., D'Italia, S., Rodgers, L., Cox, H., da Cruz Paula, A., Stepansky, A., Schizas, M., Wen, H. Y., King, T. A., ... Reis-Filho, J. S. (2017). Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. *Nature Medicine*, *23*(3), 376–385. <https://doi.org/10.1038/nm.4279>
- Martínez, N., Almaraz, C., Vaqué, J. P., Varela, I., Derdak, S., Beltran, S., Mollejo, M., Campos-Martin, Y., Agueda, L., Rinaldi, A., Kwee, I., Gut, M., Blanc, J., Oscier, D., Strefford, J. C., Martinez-Lopez, J., Salar, A., Sole, F., Rodriguez-Peralto, J. L., ... Piris, M. A. (2014). Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation. *Leukemia*, *28*(6), 1334–1340. <https://doi.org/10.1038/leu.2013.365>
- Massagué, J. (2008). TGFbeta in Cancer. *Cell*, *134*(2), 215–230. <https://doi.org/10.1016/j.cell.2008.07.001>
- Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(2), 560–564. <https://doi.org/10.1073/pnas.74.2.560>
- McDonald, O. G., Li, X., Saunders, T., Tryggvadottir, R., Mentch, S. J., Warmoes, M. O., Word, A. E., Carrer, A., Salz, T. H., Natsume, S., Stauffer, K. M., Makohon-Moore, A., Zhong, Y., Wu, H., Wellen, K. E., Locasale, J. W., Iacobuzio-Donahue, C. A., & Feinberg, A. P. (2017). Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nature Genetics*, *49*(3), 367–376. <https://doi.org/10.1038/ng.3753>
- McGinn, S., & Gut, I. G. (2013). DNA sequencing—Spanning the generations. *New Biotechnology*, *30*(4), 366–372. <https://doi.org/10.1016/j.nbt.2012.11.012>
- McGranahan, N., & Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*, *168*(4), 613–628. <https://doi.org/10.1016/j.cell.2017.01.018>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>

Metzker, M. L. (2010). Sequencing technologies—The next generation. *Nature Reviews. Genetics*, *11*(1), 31–46. <https://doi.org/10.1038/nrg2626>

Meyer, C. A., & Liu, X. S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nature Reviews. Genetics*, *15*(11), 709–721. <https://doi.org/10.1038/nrg3788>

Mikheyev, A. S., & Tin, M. M. Y. (2014). A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources*, *14*(6), 1097–1102. <https://doi.org/10.1111/1755-0998.12324>

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., & Massagué, J. (2005). Genes that mediate breast cancer metastasis to lung. *Nature*, *436*(7050), 518–524. <https://doi.org/10.1038/nature03799>

Miyamoto, D. T., Zheng, Y., Wittner, B. S., Lee, R. J., Zhu, H., Broderick, K. T., Desai, R., Fox, D. B., Brannigan, B. W., Trautwein, J., Arora, K. S., Desai, N., Dahl, D. M., Sequist, L. V., Smith, M. R., Kapur, R., Wu, C.-L., Shioda, T., Ramaswamy, S., ... Haber, D. A. (2015). RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science (New York, N.Y.)*, *349*(6254), 1351–1356. <https://doi.org/10.1126/science.aab0917>

Moffitt, R. A., Marayati, R., Flate, E. L., Volmar, K. E., Loeza, S. G. H., Hoadley, K. A., Rashid, N. U., Williams, L. A., Eaton, S. C., Chung, A. H., Smyla, J. K., Anderson, J. M., Kim, H. J., Bentrem, D. J., Talamonti, M. S., Iacobuzio-Donahue, C. A., Hollingsworth, M. A., & Yeh, J. J. (2015). Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nature Genetics*, *47*(10), 1168. <https://doi.org/10.1038/ng.3398>

Moore, M. J., Goldstein, D., Hamm, J., Figer, A., Hecht, J. R., Gallinger, S., Au, H. J., Murawa, P., Walde, D., Wolff, R. A., Campos, D., Lim, R., Ding, K., Clark, G., Voskoglou-Nomikos, T., Ptasynski, M., Parulekar, W., & National Cancer Institute of Canada Clinical Trials Group. (2007). Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: A phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *25*(15), 1960–1966. <https://doi.org/10.1200/JCO.2006.07.9525>

Morozova, O., & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, *92*(5), 255–264. <https://doi.org/10.1016/j.ygeno.2008.07.001>

Morton, J. P., Timpson, P., Karim, S. A., Ridgway, R. A., Athineos, D., Doyle, B., Jamieson, N. B., Oien, K. A., Lowy, A. M., Brunton, V. G., Frame, M. C., Evans, T. R. J., & Sansom, O. J. (2010). Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(1), 246–251. <https://doi.org/10.1073/pnas.0908428107>

Mueller, S., Engleitner, T., Maresch, R., Zukowska, M., Lange, S., Kaltenbacher, T., Konukiewitz, B., Öllinger, R., Zwiebel, M., Strong, A., Yen, H.-Y., Banerjee, R., Louzada, S., Fu, B., Seidler, B., Götzfried, J., Schuck, K., Hassan, Z., Arbeiter, A., ... Rad, R. (2018). Evolutionary routes and KRAS dosage define pancreatic cancer phenotypes. *Nature*, *554*(7690), 62–68. <https://doi.org/10.1038/nature25459>

Nakhai, H., Sel, S., Favor, J., Mendoza-Torres, L., Paulsen, F., Duncker, G. I. W., & Schmid, R. M. (2007). Ptf1a is essential for the differentiation of GABAergic and glycinergic amacrine cells and horizontal cells in the mouse retina. *Development (Cambridge, England)*, *134*(6), 1151–1160. <https://doi.org/10.1242/dev.02781>

Ni, S., Hu, J., Duan, Y., Shi, S., Li, R., Wu, H., Qu, Y., & Li, Y. (2013). Down expression of LRP1B promotes cell migration via RhoA/Cdc42 pathway and actin cytoskeleton remodeling in renal cell cancer. *Cancer Science*, *104*(7), 817–825. <https://doi.org/10.1111/cas.12157>

Niedzicka, M., Fijarczyk, A., Dudek, K., Stuglik, M., & Babik, W. (2016). Molecular Inversion Probes for targeted resequencing in non-model organisms. *Scientific Reports*, *6*, 24051. <https://doi.org/10.1038/srep24051>

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., ... Breast Cancer Working Group of the International Cancer Genome Consortium. (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell*, *149*(5), 979–993. <https://doi.org/10.1016/j.cell.2012.04.024>

Nolan-Stevaux, O., Lau, J., Truitt, M. L., Chu, G. C., Hebrok, M., Fernández-Zapico, M. E., & Hanahan, D. (2009). GLI1 is regulated through Smoothed-independent mechanisms in neoplastic pancreatic ducts and mediates PDAC cell survival and transformation. *Genes & Development*, *23*(1), 24–36. <https://doi.org/10.1101/gad.1753809>

Notta, F., Chan-Seng-Yue, M., Lemire, M., Li, Y., Wilson, G. W., Connor, A. A., Denroche, R. E., Liang, S.-B., Brown, A. M. K., Kim, J. C., Wang, T., Simpson, J. T., Beck, T., Borgida, A., Buchner, N., Chadwick, D., Hafezi-Bakhtiari, S., Dick, J. E., Heisler, L., ... Gallinger, S. (2016). A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature*, *538*(7625), 378–382. <https://doi.org/10.1038/nature19823>

Offield, M. F., Jetton, T. L., Labosky, P. A., Ray, M., Stein, R. W., Magnuson, M. A., Hogan, B. L., & Wright, C. V. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development (Cambridge, England)*, *122*(3), 983–995. <https://doi.org/10.1242/dev.122.3.983>

Paget, S. (1889). The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Reviews*, *8*(2), 98–101.

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Louis, D. N., Rozenblatt-Rosen, O., Suvà, M. L., Regev, A., & Bernstein, B. E. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, *344*(6190), 1396–1401. <https://doi.org/10.1126/science.1254257>

Peinado, H., Lavotshkin, S., & Lyden, D. (2011). The secreted factors responsible for pre-metastatic niche formation: Old sayings and new thoughts. *Seminars in Cancer Biology*, *21*(2), 139–146. <https://doi.org/10.1016/j.semcancer.2011.01.002>

Peinado, H., Zhang, H., Matei, I. R., Costa-Silva, B., Hoshino, A., Rodrigues, G., Psaila, B., Kaplan, R. N., Bromberg, J. F., Kang, Y., Bissell, M. J., Cox, T. R., Giaccia, A. J., Erler, J. T., Hiratsuka, S., Ghajar, C. M., & Lyden, D. (2017). Pre-metastatic niches: Organ-specific homes for metastases. *Nature Reviews. Cancer*, *17*(5), 302–317. <https://doi.org/10.1038/nrc.2017.6>

Pérez-Mancera, P. A., Rust, A. G., van der Weyden, L., Kristiansen, G., Li, A., Sarver, A. L., Silverstein, K. A. T., Grützmann, R., Aust, D., Rümmele, P., Knösel, T., Herd, C., Stemple, D. L., Kettleborough, R., Brosnan, J. A., Li, A., Morgan, R., Knight, S., Yu, J., ... Tuveson, D. A. (2012). The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. *Nature*, *486*(7402), 266–270. <https://doi.org/10.1038/nature11114>

Poruk, K. E., Gay, D. Z., Brown, K., Mulvihill, J. D., Boucher, K. M., Scaife, C. L., Firpo, M. A., & Mulvihill, S. J. (2013). The clinical utility of CA 19-9 in pancreatic adenocarcinoma: Diagnostic and prognostic updates. *Current Molecular Medicine*, *13*(3), 340–351. <https://doi.org/10.2174/1566524011313030003>

Prasetyanti, P. R., & Medema, J. P. (2017). Intra-tumor heterogeneity from a cancer stem cell perspective. *Molecular Cancer*, *16*(1), 41. <https://doi.org/10.1186/s12943-017-0600-4>

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C. L., Mroz, E. A., Emerick, K. S., Deschler, D. G., Varvares, M. A., Mylvaganam, R., Rozenblatt-Rosen, O., Rocco, J. W., Faquin, W. C., Lin, D. T., Regev, A., & Bernstein, B. E. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell*, *171*(7), 1611-1624.e24. <https://doi.org/10.1016/j.cell.2017.10.044>

Rad, R., Rad, L., Wang, W., Strong, A., Ponstingl, H., Bronner, I. F., Mayho, M., Steiger, K., Weber, J., Hieber, M., Veltkamp, C., Eser, S., Geumann, U., Öllinger, R., Zukowska, M., Barenboim, M., Maresch, R., Cadiñanos, J., Friedrich, M., ... Bradley, A. (2015). A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nature Genetics*, *47*(1), 47–56. <https://doi.org/10.1038/ng.3164>

Rausch, T., Jones, D. T. W., Zapatka, M., Stütz, A. M., Zichner, T., Weischenfeldt, J., Jäger, N., Remke, M., Shih, D., Northcott, P. A., Pfaff, E., Tica, J., Wang, Q., Massimi, L., Witt, H., Bender, S., Pleier, S., Cin, H., Hawkins, C., ... Korbel, J. O. (2012). Genome sequencing of

pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell*, 148(1–2), 59–71. <https://doi.org/10.1016/j.cell.2011.12.013>

Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* (Oxford, England), 28(18), i333–i339. <https://doi.org/10.1093/bioinformatics/bts378>

Real, F. X. (2003). A ‘catastrophic hypothesis’ for pancreas cancer progression. *Gastroenterology*, 124(7), 1958–1964. [https://doi.org/10.1016/s0016-5085\(03\)00389-5](https://doi.org/10.1016/s0016-5085(03)00389-5)

Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., Clevers, H., Deplancke, B., Dunham, I., Eberwine, J., Eils, R., Enard, W., Farmer, A., Fugger, L., Göttgens, B., ... Human Cell Atlas Meeting Participants. (2017). The Human Cell Atlas. *ELife*, 6, e27041. <https://doi.org/10.7554/eLife.27041>

Rhim, A. D., Mirek, E. T., Aiello, N. M., Maitra, A., Bailey, J. M., McAllister, F., Reichert, M., Beatty, G. L., Rustgi, A. K., Vonderheide, R. H., Leach, S. D., & Stanger, B. Z. (2012). EMT and dissemination precede pancreatic tumor formation. *Cell*, 148(1–2), 349–361. <https://doi.org/10.1016/j.cell.2011.11.025>

Robertson, E., Bradley, A., Kuehn, M., & Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature*, 323(6087), 445–448. <https://doi.org/10.1038/323445a0>

Roe, J.-S., Hwang, C.-I., Somerville, T. D. D., Milazzo, J. P., Lee, E. J., Da Silva, B., Maiorino, L., Tiriach, H., Young, C. M., Miyabayashi, K., Filippini, D., Creighton, B., Burkhardt, R. A., Buscaglia, J. M., Kim, E. J., Grem, J. L., Lazenby, A. J., Grunkemeyer, J. A., Hollingsworth, M. A., ... Vakoc, C. R. (2017). Enhancer Reprogramming Promotes Pancreatic Cancer Metastasis. *Cell*, 170(5), 875–888.e20. <https://doi.org/10.1016/j.cell.2017.07.007>

Roerink, S. F., Sasaki, N., Lee-Six, H., Young, M. D., Alexandrov, L. B., Behjati, S., Mitchell, T. J., Grossmann, S., Lightfoot, H., Egan, D. A., Pronk, A., Smakman, N., van Gorp, J., Anderson, E., Gamble, S. J., Alder, C., van de Wetering, M., Campbell, P. J., Stratton, M. R., & Clevers, H. (2018). Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature*, 556(7702), 457–462. <https://doi.org/10.1038/s41586-018-0024-3>

Rosenberg, A. B., Roco, C. M., Muscat, R. A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L. T., Peeler, D. J., Mukherjee, S., Chen, W., Pun, S. H., Sellers, D. L., Tasic, B., & Seelig, G. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* (New York, N.Y.), 360(6385), 176–182. <https://doi.org/10.1126/science.aam8999>

Rosty, C., & Goggins, M. (2002). Early detection of pancreatic carcinoma. *Hematology/Oncology Clinics of North America*, *16*(1), 37–52. [https://doi.org/10.1016/s0889-8588\(01\)00007-7](https://doi.org/10.1016/s0889-8588(01)00007-7)

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., ... Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, *475*(7356), 348–352. <https://doi.org/10.1038/nature10242>

Ryan, D. P., Hong, T. S., & Bardeesy, N. (2014). Pancreatic adenocarcinoma. *The New England Journal of Medicine*, *371*(11), 1039–1049. <https://doi.org/10.1056/NEJMra1404198>

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M., & Smith, M. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, *265*(5596), 687–695. <https://doi.org/10.1038/265687a0>

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>

Sathirapongsasuti, J. F., Lee, H., Horst, B. A. J., Brunner, G., Cochran, A. J., Binder, S., Quackenbush, J., & Nelson, S. F. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics (Oxford, England)*, *27*(19), 2648–2654. <https://doi.org/10.1093/bioinformatics/btr462>

Sauer, B. (1998). Inducible gene targeting in mice using the Cre/lox system. *Methods (San Diego, Calif.)*, *14*(4), 381–392. <https://doi.org/10.1006/meth.1998.0593>

Saunders, C. T., Wong, W. S. W., Swamy, S., Becq, J., Murray, L. J., & Cheetham, R. K. (2012). Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics (Oxford, England)*, *28*(14), 1811–1817. <https://doi.org/10.1093/bioinformatics/bts271>

Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics*, *19*(R2), R227–240. <https://doi.org/10.1093/hmg/ddq416>

Schepers, A. G., Snippert, H. J., Stange, D. E., van den Born, M., van Es, J. H., van de Wetering, M., & Clevers, H. (2012). Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. *Science (New York, N.Y.)*, *337*(6095), 730–735. <https://doi.org/10.1126/science.1224676>

Schönhuber, N., Seidler, B., Schuck, K., Veltkamp, C., Schachtler, C., Zukowska, M., Eser, S., Feyerabend, T. B., Paul, M. C., Eser, P., Klein, S., Lowy, A. M., Banerjee, R., Yang, F., Lee, C.-L., Moding, E. J., Kirsch, D. G., Scheideler, A., Alessi, D. R., ... Saur, D. (2014). A next-

generation dual-recombinase system for time- and host-specific targeting of pancreatic cancer. *Nature Medicine*, 20(11), 1340–1347. <https://doi.org/10.1038/nm.3646>

Schutte, M., Hruban, R. H., Geradts, J., Maynard, R., Hilgers, W., Rabindran, S. K., Moskaluk, C. A., Hahn, S. A., Schwarte-Waldhoff, I., Schmiegel, W., Baylin, S. B., Kern, S. E., & Herman, J. G. (1997). Abrogation of the Rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas. *Cancer Research*, 57(15), 3126–3130.

Schwarze, K., Buchanan, J., Fermont, J. M., Dreau, H., Tilley, M. W., Taylor, J. M., Antoniou, P., Knight, S. J. L., Camps, C., Pentony, M. M., Kvikstad, E. M., Harris, S., Popitsch, N., Pagnamenta, A. T., Schuh, A., Taylor, J. C., & Wordsworth, S. (2020). The complete costs of genome sequencing: A microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 22(1), 85–94. <https://doi.org/10.1038/s41436-019-0618-7>

Shi, Y., & Massagué, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell*, 113(6), 685–700. [https://doi.org/10.1016/s0092-8674\(03\)00432-x](https://doi.org/10.1016/s0092-8674(03)00432-x)

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(1), 7–33. <https://doi.org/10.3322/caac.21708>

Singhi, A. D., George, B., Greenbowe, J. R., Chung, J., Suh, J., Maitra, A., Klempner, S. J., Hendifar, A., Milind, J. M., Golan, T., Brand, R. E., Zureikat, A. H., Roy, S., Schrock, A. B., Miller, V. A., Ross, J. S., Ali, S. M., & Bahary, N. (2019). Real-Time Targeted Genome Profile Analysis of Pancreatic Ductal Adenocarcinomas Identifies Genetic Alterations That Might Be Targeted With Existing Drugs or Used as Biomarkers. *Gastroenterology*, 156(8), 2242–2253.e4. <https://doi.org/10.1053/j.gastro.2019.02.037>

Smith, A. L., Robin, T. P., & Ford, H. L. (2012). Molecular pathways: Targeting the TGF- $\beta$  pathway for cancer therapy. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 18(17), 4514–4521. <https://doi.org/10.1158/1078-0432.CCR-11-3224>

Smith, J. D., Meehan, M. H., Crean, J., & McCann, A. (2011). Alpha T-catenin (CTNNA3): A gene in the hand is worth two in the nest. *Cellular and Molecular Life Sciences: CMLS*, 68(15), 2493–2498. <https://doi.org/10.1007/s00018-011-0728-0>

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, 321(6071), 674–679. <https://doi.org/10.1038/321674a0>

Solomon, S., Das, S., Brand, R., & Whitcomb, D. C. (2012). Inherited pancreatic cancer syndromes. *Cancer Journal (Sudbury, Mass.)*, 18(6), 485–491. <https://doi.org/10.1097/PPO.0b013e318278c4a6>

Sonawane, N. D., Szoka, F. C., & Verkman, A. S. (2003). Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes. *The Journal of Biological Chemistry*, *278*(45), 44826–44831. <https://doi.org/10.1074/jbc.M308643200>

Stegle, O., Teichmann, S. A., & Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews. Genetics*, *16*(3), 133–145. <https://doi.org/10.1038/nrg3833>

Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., Pleasance, E. D., Lau, K. W., Beare, D., Stebbings, L. A., McLaren, S., Lin, M.-L., McBride, D. J., Varela, I., Nik-Zainal, S., Leroy, C., Jia, M., Menzies, A., Butler, A. P., ... Campbell, P. J. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, *144*(1), 27–40. <https://doi.org/10.1016/j.cell.2010.11.055>

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., & Dudoit, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, *19*(1), 477. <https://doi.org/10.1186/s12864-018-4772-0>

Stripecke, R., Carmen Villacres, M., Skelton, D., Satake, N., Halene, S., & Kohn, D. (1999). Immune response to green fluorescent protein: Implications for gene therapy. *Gene Therapy*, *6*(7), 1305–1312. <https://doi.org/10.1038/sj.gt.3300951>

Stuart, T., & Satija, R. (2019). Integrative single-cell analysis. *Nature Reviews. Genetics*, *20*(5), 257–272. <https://doi.org/10.1038/s41576-019-0093-7>

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. <https://doi.org/10.1073/pnas.0506580102>

Swanton, C. (2012). Intratumor heterogeneity: Evolution through space and time. *Cancer Research*, *72*(19), 4875–4882. <https://doi.org/10.1158/0008-5472.CAN-12-2217>

Tabassum, D. P., & Polyak, K. (2015). Tumorigenesis: It takes a village. *Nature Reviews Cancer*, *15*(8), 473–483. <https://doi.org/10.1038/nrc3971>

Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, & Principal investigators. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, *562*(7727), 367–372. <https://doi.org/10.1038/s41586-018-0590-4>

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). MRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*(5), 377–382. <https://doi.org/10.1038/nmeth.1315>



- Teer, J. K., & Mullikin, J. C. (2010). Exome sequencing: The sweet spot before whole genomes. *Human Molecular Genetics*, *19*(R2), R145-151. <https://doi.org/10.1093/hmg/ddq333>
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Ting, D. T., Wittner, B. S., Ligorio, M., Vincent Jordan, N., Shah, A. M., Miyamoto, D. T., Aceto, N., Bersani, F., Brannigan, B. W., Xega, K., Ciciliano, J. C., Zhu, H., MacKenzie, O. C., Trautwein, J., Arora, K. S., Shahid, M., Ellis, H. L., Qu, N., Bardeesy, N., ... Haber, D. A. (2014). Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Reports*, *8*(6), 1905–1918. <https://doi.org/10.1016/j.celrep.2014.08.029>
- Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., ... Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (New York, N.Y.)*, *352*(6282), 189–196. <https://doi.org/10.1126/science.aad0501>
- Tirosh, I., Venteicher, A. S., Hebert, C., Escalante, L. E., Patel, A. P., Yizhak, K., Fisher, J. M., Rodman, C., Mount, C., Filbin, M. G., Neftel, C., Desai, N., Nyman, J., Izar, B., Luo, C. C., Francis, J. M., Patel, A. A., Onozato, M. L., Riggi, N., ... Suvà, M. L. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature*, *539*(7628), 309–313. <https://doi.org/10.1038/nature20123>
- Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, *31*(1), 46–53. <https://doi.org/10.1038/nbt.2450>
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–515. <https://doi.org/10.1038/nbt.1621>
- Tung, P.-Y., Blischak, J. D., Hsiao, C. J., Knowles, D. A., Burnett, J. E., Pritchard, J. K., & Gilad, Y. (2017). Batch effects and the effective design of single-cell gene expression studies. *Scientific Reports*, *7*, 39921. <https://doi.org/10.1038/srep39921>
- Valdes-Mora, F., Handler, K., Law, A. M. K., Salomon, R., Oakes, S. R., Ormandy, C. J., & Gallego-Ortega, D. (2018). Single-Cell Transcriptomics in Cancer Immunobiology: The Future of Precision Oncology. *Frontiers in Immunology*, *9*, 2582. <https://doi.org/10.3389/fimmu.2018.02582>

Valouev, A., Johnson, D. S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R. M., & Sidow, A. (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nature Methods*, *5*(9), 829–834. <https://doi.org/10.1038/nmeth.1246>

Vasen, H. F., Gruis, N. A., Frants, R. R., van Der Velden, P. A., Hille, E. T., & Bergman, W. (2000). Risk of developing pancreatic cancer in families with familial atypical multiple mole melanoma associated with a specific 19 deletion of p16 (p16-Leiden). *International Journal of Cancer*, *87*(6), 809–811.

Von Hoff, D. D., Ervin, T., Arena, F. P., Chiorean, E. G., Infante, J., Moore, M., Seay, T., Tjulandin, S. A., Ma, W. W., Saleh, M. N., Harris, M., Reni, M., Dowden, S., Laheru, D., Bahary, N., Ramanathan, R. K., Tabernero, J., Hidalgo, M., Goldstein, D., ... Renschler, M. F. (2013). Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *The New England Journal of Medicine*, *369*(18), 1691–1703. <https://doi.org/10.1056/NEJMoa1304369>

Waddell, N., Pajic, M., Patch, A.-M., Chang, D. K., Kassahn, K. S., Bailey, P., Johns, A. L., Miller, D., Nones, K., Quek, K., Quinn, M. C. J., Robertson, A. J., Fadlullah, M. Z. H., Bruxner, T. J. C., Christ, A. N., Harliwong, I., Idrisoglu, S., Manning, S., Nourse, C., ... Grimmond, S. M. (2015). Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*, *518*(7540), 495–501. <https://doi.org/10.1038/nature14169>

Wang, Q., Wang, X., Tang, P. S., O'leary, G. M., & Zhang, M. (2017). Targeted sequencing of both DNA strands barcoded and captured individually by RNA probes to identify genome-wide ultra-rare mutations. *Scientific Reports*, *7*(1), 3356. <https://doi.org/10.1038/s41598-017-03448-8>

Wang, Y., Yang, Q., & Wang, Z. (2014). The evolution of nanopore sequencing. *Frontiers in Genetics*, *5*, 449. <https://doi.org/10.3389/fgene.2014.00449>

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, *10*(1), 57–63. <https://doi.org/10.1038/nrg2484>

Whatcott, C., Han, H., Posner, R. G., & Von Hoff, D. D. (2013). Tumor-stromal interactions in pancreatic cancer. *Critical Reviews in Oncogenesis*, *18*(1–2), 135–151. <https://doi.org/10.1615/critrevoncog.v18.i1-2.80>

Winterhoff, B. J., Maile, M., Mitra, A. K., Sebe, A., Bazzaro, M., Geller, M. A., Abrahante, J. E., Klein, M., Hellweg, R., Mullany, S. A., Beckman, K., Daniel, J., & Starr, T. K. (2017). Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells. *Gynecologic Oncology*, *144*(3), 598–606. <https://doi.org/10.1016/j.ygyno.2017.01.015>

Witkiewicz, A. K., McMillan, E. A., Balaji, U., Baek, G., Lin, W.-C., Mansour, J., Mollaei, M., Wagner, K.-U., Koduru, P., Yopp, A., Choti, M. A., Yeo, C. J., McCue, P., White, M. A., & Knudsen, E. S. (2015). Whole-exome sequencing of pancreatic cancer defines genetic

diversity and therapeutic targets. *Nature Communications*, 6, 6744. <https://doi.org/10.1038/ncomms7744>

Xie, C., & Tammi, M. T. (2009). CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics*, 10, 80. <https://doi.org/10.1186/1471-2105-10-80>

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y. E., Liu, J., Horvath, S., & Fan, G. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*, 500(7464), 593–597. <https://doi.org/10.1038/nature12364>

Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R. H., Eshleman, J. R., Nowak, M. A., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., & Iacobuzio-Donahue, C. A. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*, 467(7319), 1114–1117. <https://doi.org/10.1038/nature09515>

Yachida, S., White, C. M., Naito, Y., Zhong, Y., Brosnan, J. A., Macgregor-Das, A. M., Morgan, R. A., Saunders, T., Laheru, D. A., Herman, J. M., Hruban, R. H., Klein, A. P., Jones, S., Velculescu, V., Wolfgang, C. L., & Iacobuzio-Donahue, C. A. (2012). Clinical significance of the genetic landscape of pancreatic cancer and implications for identification of potential long-term survivors. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 18(22), 6339–6347. <https://doi.org/10.1158/1078-0432.CCR-12-1215>

Yamaguchi, J., Kokuryo, T., Yokoyama, Y., Ebata, T., Ochiai, Y., & Nagino, M. (2021). Premalignant pancreatic cells seed stealth metastasis in distant organs in mice. *Oncogene*, 40(12), 2273–2284. <https://doi.org/10.1038/s41388-021-01706-8>

Yang, J., Lee, S. J., Kwon, Y., Ma, L., & Kim, J. (2020). Tumor suppressive function of MatrIn 3 in the basal-like breast cancer. *Biological Research*, 53(1), 42. <https://doi.org/10.1186/s40659-020-00310-6>

Yang, X., Leslie, G., Doroszuk, A., Schneider, S., Allen, J., Decker, B., Dunning, A. M., Redman, J., Scarth, J., Plaskocinska, I., Luccarini, C., Shah, M., Pooley, K., Dorling, L., Lee, A., Adank, M. A., Adlard, J., Aittomäki, K., Andrulis, I. L., ... Tischkowitz, M. (2020). Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 38(7), 674–685. <https://doi.org/10.1200/JCO.19.01907>

Ye, K., Schulz, M. H., Long, Q., Apweiler, R., & Ning, Z. (2009). Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics (Oxford, England)*, 25(21), 2865–2871. <https://doi.org/10.1093/bioinformatics/btp394>

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. <https://doi.org/10.1186/gb-2008-9-9-r137>

Zhang, Y., Park, C., Bennett, C., Thornton, M., & Kim, D. (2021). Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Research*. <https://doi.org/10.1101/gr.275193.120>

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*, 14049. <https://doi.org/10.1038/ncomms14049>





**Publications**

## Opinion

# Tumor Functional Heterogeneity Unraveled by scRNA-seq Technologies

Laura González-Silva,<sup>1</sup> Laura Quevedo,<sup>1</sup> and Ignacio Varela<sup>1,\*</sup>

**Effective cancer treatment has been precluded by the presence of various forms of intratumoral complexity that drive treatment resistance and metastasis. Recent single-cell sequencing technologies are significantly facilitating the characterization of tumor internal architecture during disease progression. New applications and advances occurring at a fast pace predict an imminent broad application of these technologies in many research areas. As occurred with next-generation sequencing (NGS) technologies, once applied to clinical samples across tumor types, single-cell sequencing technologies could trigger an exponential increase in knowledge of the molecular pathways involved in cancer progression and contribute to the improvement of cancer treatment.**

### Multifaceted Heterogeneity and Its Impact on Cancer Progression

Tumors comprise various cell populations in constant evolution. Some of this complex heterogeneity derives from genetic diversification and Darwinian selection of tumor cells as they adapt to variable environments. Next-generation sequencing (NGS; see [Glossary](#)) used for the past decade had enough sensitivity to detect mutations present in minor cell populations and, combined with multi-sampling of human tumors (multisampling sequencing), fostered many studies that characterized intratumor heterogeneity in various cancers [1]. The level of intratumor heterogeneity is considered a main driver of therapy resistance and metastasis and is associated with poor prognosis [2].

In addition, human cancers frequently have tumor cell populations with different transcriptional programs. This functional diversity is likely associated with the genetic heterogeneity described above but is also the result of many other factors. First, the presence of a hierarchical structure, where a group of quiescent stem-like cells fosters the growth of a tumor comprising cells in different differentiation states, was demonstrated in various tumor types [3]. Additionally, different transcriptional programs can be activated in tumor cells as a response to stochastic factors or to a variable tumor microenvironment. This functional diversity provides tumors with a plasticity that grants a high capacity for adaptation [4].

Finally, human tumors comprise not only malignant/transformed cells but also a plethora of different cell types recruited from the surrounding tissue and the immune system. The tumor microenvironment shows also genetic and transcriptional diversity and plays important roles in tumor progression, metastasis, and treatment resistance [1,5].

Fine characterization of these levels of tumor heterogeneity is essential to the successful treatment of cancer patients. The recent development of technologies based on sequencing individual cells (single-cell sequencing technologies) opens new ways for the characterization of tumor heterogeneity. At the genetic level, single-cell DNA-seq technologies offer higher sensitivity in the detection of minority clones, the reconstruction of clone structure, and the identification of concurrent or exclusive alterations in the same cells. However, it is in the study of functional heterogeneity that single-cell RNA-seq (scRNA-seq) significantly improves on previous technologies, increasing our molecular comprehension of cancer progression. A precise cell-type annotation of complex cellular samples from primary tumors is possible thanks to the recent generation of single-cell transcriptome atlases. These comprise normal and pathological samples from human and mouse [6,7].

### The Emergence of scRNA-seq Technologies

In just a few years, the ability to perform single-cell expression profiles increased from a handful of cells to thousands of cells in a single experiment [8]. After the first scRNA-seq experiment in a

### Highlights

Tumors are highly complex entities comprising cell populations with various transcriptional programs.

Single-cell sequencing technologies are evolving fast and have the capacity to finely characterize the huge heterogeneity inside tumors.

New single-cell sequencing protocols do not need special infrastructure and can be applied to a huge multitude of cancer sample types in many research areas.

A fine characterization of liquid biopsies, tumor functional heterogeneity, and the tumor microenvironment will be followed by an exponential increase in our knowledge on tumor progression and will significantly improve cancer treatment.

<sup>1</sup>Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria – CSIC, Santander, Spain

\*Correspondence: [ignacio.varela@unican.es](mailto:ignacio.varela@unican.es)



four-cell-stage blastomere [9], several studies were published based on cell isolation and individual **genomic library** preparation. These initial protocols were laborious and expensive, required RNA amplification steps that introduced bias in the data, and were characterized by reduced throughput [10,11]. The subsequent introduction of unique molecular identifiers (UMIs), which are random sequences that label individual molecules, significantly removed cDNA amplification bias [12]. Further developments in STRT-seq and CEL-seq protocols included the introduction of an individual barcoding step on isolated cells before a single retrotranscription reaction reducing batch artifacts [13,14]. In 2015, the introduction of **microfluidic** devices (Drop-seq [15] and InDrop [16]) enabled the processing of thousands of cells at once. Following this strategy, 10x Genomics automated equipment recently characterized 1.3 million cells at the single-cell level [8]. Unfortunately, microfluidics-based methods are not efficient in the removal of the abundant rRNA. Consequently, they use poly-T oligonucleotides to sequence the end of poly-A-tailed RNAs. This is useful in generating expression profiles in this group of RNAs but does not provide complete transcriptomic information. Split-seq and Sci-seq strategies avoid physical cell isolation, taking advantage of a combinatorial barcoding strategy that permits the individual labeling of more than 100 000 single-cell transcriptomes [17,18]. These techniques do not require expensive microfluidics infrastructure and permit greater control over the number of analyzed cells. Finally, **single-cell multiomics** approaches that allow the study of genetic, epigenetic, and transcriptomic profiles in the same cell have been developed [19,20]. This opens a window of opportunity for comprehensive cell characterization.

Single-cell sequencing data analysis is a great challenge, similar to the early years of the use of NGS technologies. Due to a great variety of sequencing strategies and biological questions, there are many different reported analysis workflows. Analysis tools for subpopulation identification, differential expression, functional signatures, pseudotiming modeling, and network reconstruction are publicly available for researchers with limited bioinformatics resources [21,22].

## Dissecting the Tumor Ecosystem with scRNA-seq Functional Diversity of Tumor Cells

Transcriptional heterogeneity among tumor cells has clear and direct clinical implications. First, molecular classification according to **transcriptional signatures** is commonly used for clinical management in many tumor types. Regarding this, the presence of different transcriptional programs inside the same tumor might prevent, or at least bias, molecular classification from a single biopsy. In this context, scRNA-seq experiments have demonstrated the presence within the tumor of multiple cell populations belonging to different molecular groups according to standard classifications [23–25].

Second, the presence of functional diversity within tumors likely improves their adaptation to hostile environments. Functionally diverse cell populations with symbiotic, mutually beneficial relationships have been reported in tumors [26]. This diversity can also be hierarchical, as described in several tumor types in which a minority of highly specialized cells, termed cancer stem cells (CSCs), might have special capacities to maintain tumor growth, metastasize, and resist antitumor treatments [27]. Nevertheless, the lack of universally accepted CSC markers and properties has generated controversy in these studies. scRNA-seq technologies offer an opportunity for the unbiased identification and study of those populations that supposedly are present in very low numbers and in a quiescent or dormant state, and to design more specific antitumor treatments [28]. scRNA-seq experiments recently demonstrated the presence of populations with stem-like and treatment-resistance properties in oligodendroglioma and melanoma [29,30].

Finally, single-cell technologies can detect minor treatment-resistant cell populations inside complex tumors, which can be used to select appropriate therapies. For instance, the presence of a melanoma cell population expressing high levels of AXL anticipated the occurrence of clonal selection after treatment with RAF or MEK inhibitors and the eventual development of drug resistance [29].

### Glossary

**Circulating tumor cells (CTCs):** cancer cells that have escaped from the primary tumor and travelled through the blood vessels.

**Functional heterogeneity:** presence of cells with different transcriptional programs inside tumors.

**Genetic heterogeneity:** existence of cell clones with different genetic somatic mutations inside human tumors.

**Genomic library:** collection of DNA fragments with common adapters ready to be analyzed by next-generation sequencing technologies.

**Intratumor heterogeneity:** the presence of cell diversity inside human tumors.

**Microfluidics:** group of techniques that allow the manipulation of fluids in the range of microliters to picoliters.

**Multisampling sequencing:** comprehensive analysis of regionally distant samples from the same tumor by next-generation technologies.

**Next-generation sequencing (NGS) technologies:** family of applications that allow the affordable parallel sequencing of hundreds of millions of small fragments in a single reaction.

**Single-cell multiomics:** technologies that allow the simultaneous analysis of different cell molecular characteristics such as genomics, transcriptomics, epigenomics, or proteomics.

**Single-cell RNA-sequencing (scRNA-seq):** analysis of the RNA content of single cells by next-generation sequencing technologies.

**Transcriptional signature:** a specific set of genes expressed by a cell in a given moment under particular circumstances.

### Tumor Microenvironment

Cancer-associated fibroblasts (CAFs) are present in many if not all solid tumors and participate actively in tumor development [31]. The molecular mechanisms behind CAFs' role remain largely unknown and the lack of reliable cell markers to identify CAFs prevents a clear statement of their abundance and importance in solid tumors [32]. The origin of CAFs is also under debate. They can be the result of the transformation of resident fibroblasts previously present in the normal tissue or new cells generated from special cell precursors recruited to the tumor [33]. scRNA-seq reports in the past years have provided useful information in this respect. Different types of CAFs have been reported in breast and colorectal tumors, which is likely to be associated with different cell origins [34–36]. Additionally, each group of CAFs has special functions in the recruitment of immune cells and in the induction of the epithelial–mesenchymal transition (EMT) in tumor cells [24,29,34,36].

Tumors are also frequently infiltrated by immune cells. The activation of the immune system to attack tumor cells is attractive as an antitumoral therapy [37]. Consequently, the so-called immunotherapies have become a promising tool in fighting cancer, although variable responses have been observed when they are applied to cancer patients [38]. There is a great diversity of immune cells with differing, and probably opposite, functions in tumor development. This complexity requires a correct transcriptional characterization of all the different cell types present in the tumor [39]. Here, scRNA-seq studies offer an unprecedented opportunity. A recent study demonstrated that a high proportion of active versus exhausted CD8<sup>+</sup> T lymphocytes is associated with a better outcome in non-small cell lung cancer [40]. By contrast, tumors presenting large proportions of regulatory T lymphocytes or myeloid-derived suppressor cells have a poor prognosis [41–43]. The complex relationship between the different immune cells present in the tumor will determine an overall tolerant or nontolerant environment. Finally, some studies successfully identified tumor neoantigens by single-cell characterization of the T cell receptor (TCR) repertoire, which might be useful in the diagnosis and treatment of cancer [40] (Figure 1, Key Figure).

### Circulating Tumor Cells

The characterization of cells that extravasate into the blood circulation, **circulating tumor cells (CTCs)**, constitute a good and low-invasive alternative for the diagnosis and, more importantly, monitoring of tumors [44]. The utility of this strategy has been widely shown in many tumor types and the quantification of CTCs can be used as a prognostic factor [45]. Whereas many authors claim that CTCs recapitulate intratumor diversity perfectly, others have reported that they resemble metastasis more than primary tumors [46].

The low number of CTCs present in the blood circulation has forced many studies to purify CTCs according to specific epithelial surface markers. Debate about the specificity of these markers calls into question some of the reported observations [45]. Some current platforms for CTC isolation are based on physicochemical properties, but it remains unclear whether this constitutes a less biased isolation method [47]. The high throughput of modern single-cell sequencing technologies offers without doubt an opportunity to reduce the need for extensive purification, which will help to clarify the nature and the source of CTCs (Figure 1).

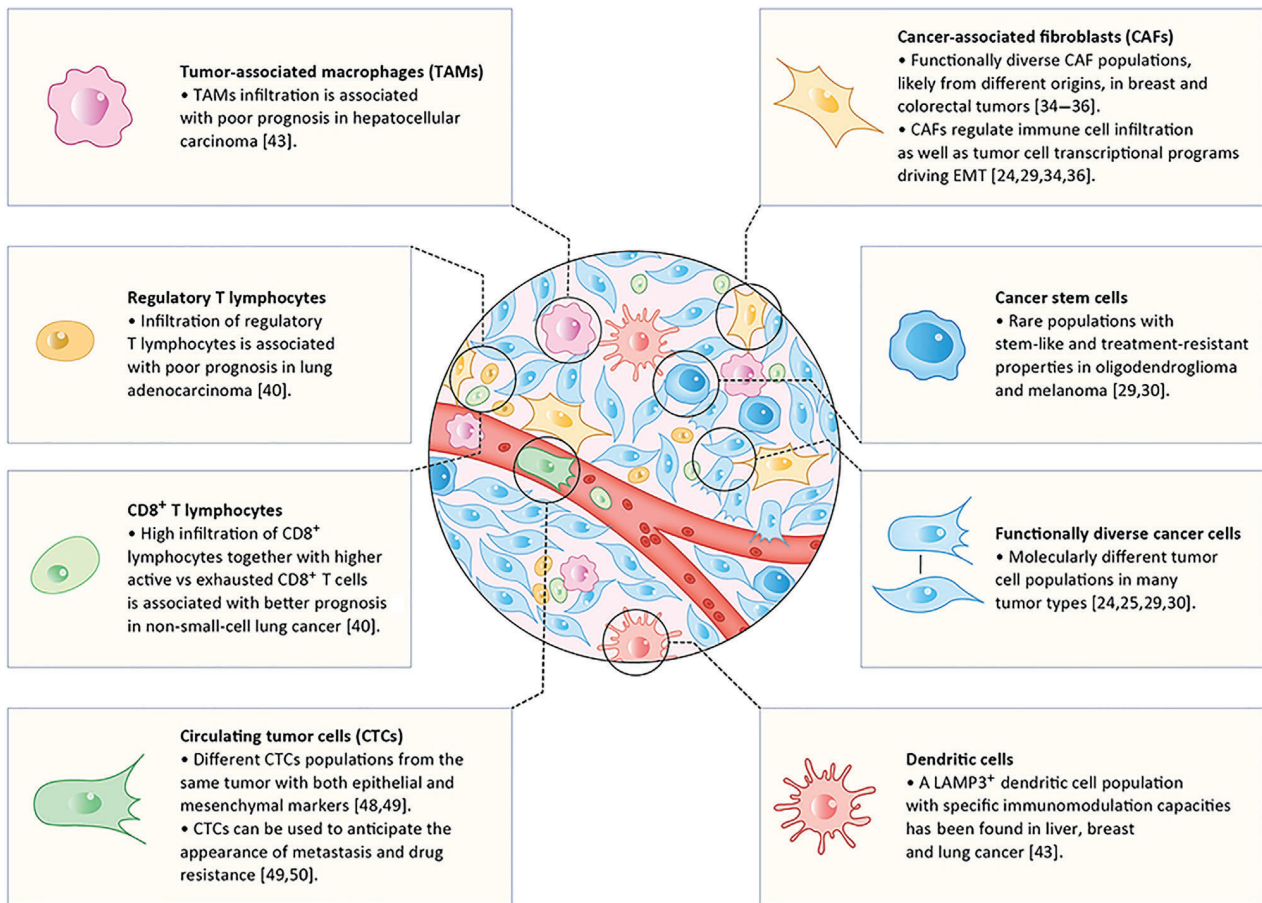
Relevant observations came recently from CTC single-cell sequencing studies. The presence of heterogeneous CTC populations with both epithelial and mesenchymal markers was identified, stressing that isolation methods based on epithelial markers are likely to be inadequate to capture all CTCs [48,49]. Additionally, a recent study on prostate cancer CTCs identified the activation of the noncanonical Wnt signaling pathway, anticipating the appearance of drug resistance [49]. Finally, the presence of plaloglobin in breast cancer CTCs was associated with earlier metastasis appearance [50]. This suggests that we need to include the study of CTCs in therapeutic decision-making in oncological practice.

### Limitations of Single-Cell Technologies in Human Cancer

A major limitation in the application of scRNA-seq technologies to solid tumor samples is the requirement for complex dissociation protocols to obtain viable, individualized fresh cells. This limitation is especially important as several studies have raised caution on potential transcriptional changes

## Key Figure

### Functional Heterogeneity of Human Tumors Revealed by Single-Cell RNA-seq (scRNA-seq) Studies

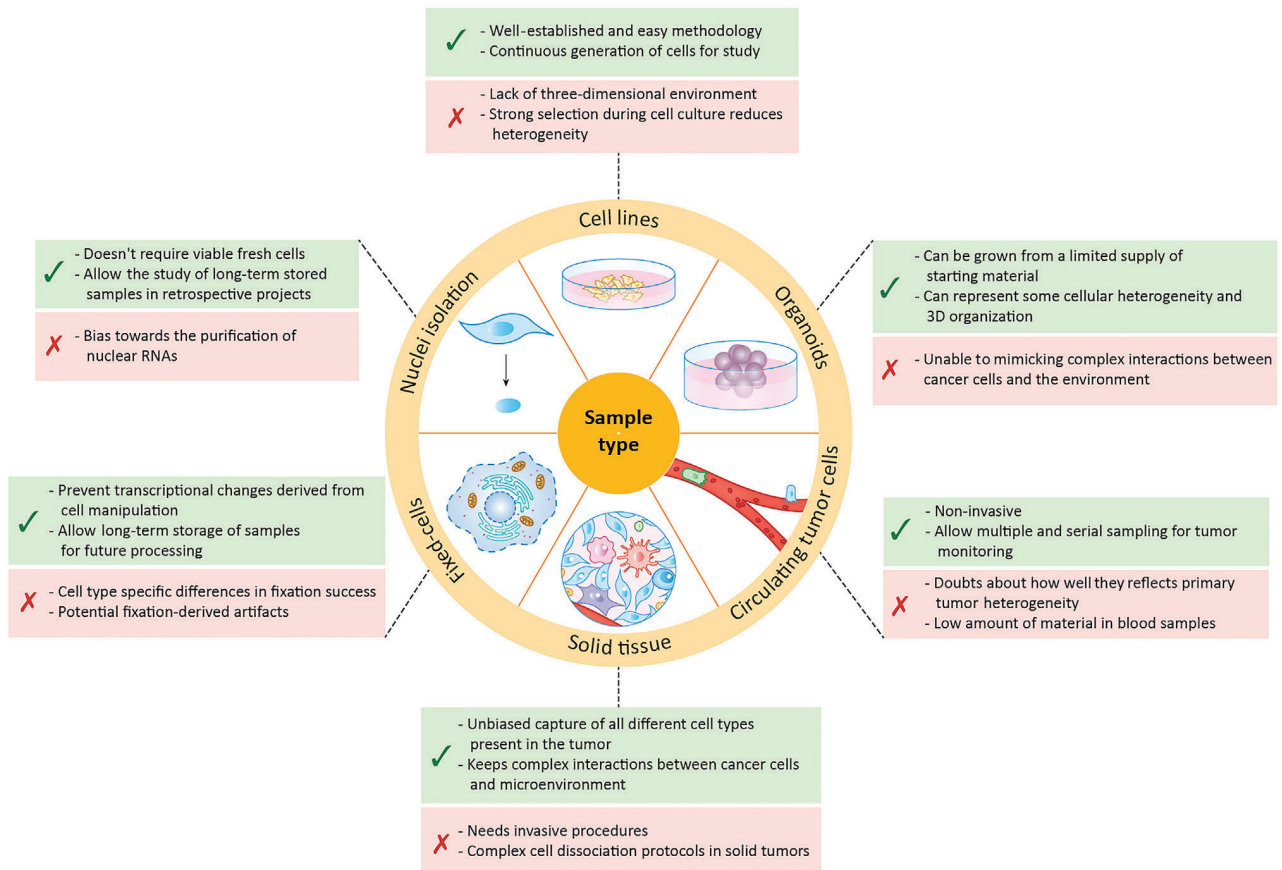


Trends in Cancer

**Figure 1.** Graphical representation of some of the main cell types present in solid tumors. scRNA-seq studies providing useful information about potential functions in cancer progression are specified next to each cell type. Abbreviation: EMT, epithelial–mesenchymal transition. See [24,25,29,30,34–36,40,43,48–50].

arising from tissue manipulation between sample collection and processing [21]. Some authors have bypassed this limitation by working with either cell lines or organoids (Figure 2) [51]. Although these have provided useful information, they fail to mimic the complex interactions between cancer cells and the microenvironment. Additionally, to understand the molecular basis of tumor evolution, it would be important to obtain several samples, or even serial samples, from the same patient, which is not straightforward in solid tumors. Recently developed low-invasive biopsy techniques such as fine-needle aspiration (FNA) are not very practical for traditional genomic analysis due to the low amount of recovered material, but offer a window of opportunity for the application of scRNA-seq technologies in clinical research [52].

Fortunately, many platforms are compatible with cell fixation and storage protocols. Transcriptomic programs obtained from these cells seem similar to those of freshly processed cells [53,54] (Figure 2).



Trends in Cancer

**Figure 2. Different Sample Types Used for Single-Cell Sequencing.**

Different types of samples currently used for single-cell sequencing are represented together with their advantages (green boxes) and disadvantages (red boxes).

If these protocols are optimized for all applications and broadly implanted, it is easy to anticipate a great increase in the application of scRNA-seq technologies in clinical research, where coordinating sample collection and processing is not always easy. Moreover, the recent development of scRNA-seq strategies for isolated single nuclei, sometimes obtained from formalin-fixed paraffin-embedded (FFPE) material, reduces the need to obtain viable cells and facilitates the study of long-stored samples in retrospective projects [11,55].

### Concluding Remarks

Despite overall improvement in the treatment of cancer patients, long-term success of targeted therapies remains limited to specific tumor types. Tumor cellular complexity is likely a key factor in this failure. Consequently, tumors with huge infiltrations of different cell types, like pancreatic adenocarcinoma, have mortality rates that remained stubbornly unchanged.

The recent development of single-cell sequencing technologies brings a revolution in the characterization of tumor heterogeneity, not only at the genetic but also at the epigenetic and transcriptomic level. Despite technical problems that need to be solved, we anticipate the incorporation of these technologies in clinical research extended to many tumor types. Similar to the explosion of genetic data generated following the development of NGS, single-cell sequencing technologies will trigger

an exponential increase in knowledge about tumor architecture and evolution dynamics (see Outstanding Questions). Finally, all of this new data will be translated into better diagnosis and treatment of cancer patients.

## References

- McGranahan, N. and Swanton, C. (2017) Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 168, 613–628
- Almendo, V. et al. (2014) Inference of tumor evolution during chemotherapy by computational modeling and *in situ* analysis of genetic and phenotypic cellular diversity. *Cell Rep.* 6, 514–527
- Magee, J.A. et al. (2012) Cancer stem cells: impact, heterogeneity, and uncertainty. *Cancer Cell* 21, 283–296
- Arozarena, I. and Wellbrock, C. (2019) Phenotype plasticity as enabler of melanoma progression and therapy resistance. *Nat. Rev. Cancer* 19, 377–391
- Junttila, M.R. and de Sauvage, F.J. (2013) Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* 501, 346–354
- The Tabula Muris Consortium et al. (2018) Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 562, 367–372
- Regev, A. et al. (2017) The Human Cell Atlas. *Elife* 6, e27041
- Zheng, G.X.Y. et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049
- Tang, F. et al. (2009) mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382
- Xue, Z. et al. (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593–597
- Grindberg, R.V. et al. (2013) RNA-sequencing from single nuclei. *Proc. Natl Acad. Sci. U. S. A.* 110, 19802–19807
- Islam, S. et al. (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11, 163–166
- Islam, S. et al. (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* 21, 1160–1167
- Hashimshony, T. et al. (2012) CEL-seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep* 2, 666–673
- Macosko, E.Z. et al. (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214
- Klein, A.M. et al. (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201
- Cao, J. et al. (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357, 661–667
- Rosenberg, A.B. et al. (2018) Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 360, 176–182
- Stuart, T. and Satija, R. (2019) Integrative single-cell analysis. *Nat. Rev. Genet.* 20, 257–272
- Hou, Y. et al. (2016) Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 26, 304–319
- Tung, P.-Y. et al. (2017) Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* 7, 39921
- Stegle, O. et al. (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16, 133–145
- Roerink, S.F. et al. (2018) Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* 556, 457–462
- Puram, S.V. et al. (2017) Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171, 1611–1624.e24
- Patel, A.P. et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401
- Tabassum, D.P. and Polyak, K. (2015) Tumorigenesis: it takes a village. *Nat. Rev. Cancer* 15, 473–483
- Prasetyanti, P.R. and Medema, J.P. (2017) Intra-tumor heterogeneity from a cancer stem cell perspective. *Mol. Cancer* 16, 41
- Winterhoff, B.J. et al. (2017) Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells. *Gynecol. Oncol.* 144, 598–606
- Tirosh, I. et al. (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196
- Tirosh, I. et al. (2016) Single-cell RNA-seq supports a developmental hierarchy in human oligodendrogloma. *Nature* 539, 309–313
- Liu, T. et al. (2019) Cancer-associated fibroblasts build and secure the tumor microenvironment. *Front. Cell Dev. Biol.* 7, 60
- Cortez, E. et al. (2014) Functional subsets of mesenchymal cell types in the tumor microenvironment. *Semin. Cancer Biol.* 25, 3–9
- Anderberg, C. and Pietras, K. (2009) On the origin of cancer-associated fibroblasts. *Cell Cycle* 8, 1461–1465
- Anjanappa, M. et al. (2017) Individualized breast cancer characterization through single-cell analysis of tumor and adjacent normal cells. *Cancer Res* 77, 2759–2769
- Bartoschek, M. et al. (2018) Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat. Commun.* 9, 5150
- Bian, S. et al. (2018) Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* 362, 1060–1063
- Chen, D.S. and Mellman, I. (2017) Elements of cancer immunity and the cancer-immune set point. *Nature* 541, 321–330
- Yu, Y. and Cui, J. (2018) Present and future of cancer immunotherapy: a tumor microenvironmental perspective. *Oncol. Lett.* 16, 4105–4113
- Valdes-Mora, F. et al. (2018) Single-cell transcriptomics in cancer immunobiology: the future of precision oncology. *Front. Immunol.* 9, 2582
- Guo, X. et al. (2018) Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* 24, 978–985
- Ruffell, B. et al. (2010) Lymphocytes in cancer development: polarization towards pro-tumor immunity. *Cytokine Growth Factor Rev.* 21, 3–10
- Zheng, C. et al. (2017) Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* 169, 1342–1356.e16
- Zhang, Q. et al. (2019) Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* 179, 829–845.e20

## Outstanding Questions

Do transcriptionally different cell populations collaborate during tumor progression?

Can the equilibrium inside the diverse tumor cell ecosystem be exploited to improve cancer treatment?

Is the functional diversity observed inside tumors the result of a specific hierarchical program?

What is the origin of cancer stem cells? Are they present in all tumor types?

Do circulating tumor cells faithfully recapitulate tumor functional heterogeneity? Can they be efficiently used to anticipate tumor progression, metastasis, or tumor resistance?



44. van den Bos, H. et al. (2018) Single-cell sequencing to quantify genomic integrity in cancer. *Int. J. Biochem. Cell Biol.* 94, 146–150
45. Zhao, X.-H. et al. (2019) Molecular detection of epithelial–mesenchymal transition markers in circulating tumor cells from pancreatic cancer patients: potential role in clinical practice. *World J. Gastroenterol.* 25, 138–150
46. Ren, X. et al. (2018) Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol.* 19, 211
47. Zhu, Z. et al. (2018) Progress and challenges of sequencing and analyzing circulating tumor cells. *Cell Biol. Toxicol.* 34, 405–415
48. Ting, D.T. et al. (2014) Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Rep.* 8, 1905–1918
49. Miyamoto, D.T. et al. (2015) RNA-seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science* 349, 1351–1356
50. Aceto, N. et al. (2014) Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 158, 1110–1122
51. Grün, D. et al. (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525, 251–255
52. Baslan, T. and Hicks, J. (2017) Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat. Rev. Cancer* 17, 557–569
53. Alles, J. et al. (2017) Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* 15, 44
54. Guillaumet-Adkins, A. et al. (2017) Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol.* 18, 45
55. Martelotto, L.G. et al. (2017) Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. *Nat. Med.* 23, 376–385



# ARID2 deficiency promotes tumor progression and is associated with higher sensitivity to chemotherapy in lung cancer

Thaidy Moreno<sup>1</sup> · Beatriz Monterde<sup>1</sup> · Laura González-Silva<sup>1</sup> · Isabel Betancor-Fernández<sup>2</sup> · Carlos Revilla<sup>1</sup> · Antonio Agraz-Doblas<sup>1</sup> · Javier Freire<sup>3</sup> · Pablo Isidro<sup>4</sup> · Laura Quevedo<sup>1</sup> · Rosa Blanco<sup>1</sup> · Santiago Montes-Moreno<sup>3</sup> · Laura Cereceda<sup>3</sup> · Aurora Astudillo<sup>4</sup> · Berta Casar<sup>1,5</sup> · Piero Crespo<sup>1,5</sup> · Cristina Morales Torres<sup>6</sup> · Paola Scaffidi<sup>6</sup> · Javier Gómez-Román<sup>3</sup> · Eduardo Salido<sup>2</sup> · Ignacio Varela<sup>1</sup>

Received: 10 January 2020 / Revised: 23 February 2021 / Accepted: 4 March 2021 / Published online: 19 March 2021  
© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Abstract

The survival rate in lung cancer remains stubbornly low and there is an urgent need for the identification of new therapeutic targets. In the last decade, several members of the SWI/SNF chromatin remodeling complexes have been described altered in different tumor types. Nevertheless, the precise mechanisms of their impact on cancer progression, as well as the application of this knowledge to cancer patient management are largely unknown. In this study, we performed targeted sequencing of a cohort of lung cancer patients on genes involved in chromatin structure. In addition, we studied at the protein level the expression of these genes in cancer samples and performed functional experiments to identify the molecular mechanisms linking alterations of chromatin remodeling genes and tumor development. Remarkably, we found that 20% of lung cancer patients show ARID2 protein loss, partially explained by the presence of ARID2 mutations. In addition, we showed that ARID2 deficiency provokes profound chromatin structural changes altering cell transcriptional programs, which bolsters the proliferative and metastatic potential of the cells both in vitro and in vivo. Moreover, we demonstrated that ARID2 deficiency impairs DNA repair, enhancing the sensitivity of the cells to DNA-damaging agents. Our findings support that ARID2 is a bona fide tumor suppressor gene in lung cancer that may be exploited therapeutically.

Deceased: Carlos Revilla

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41388-021-01748-y>.

✉ Ignacio Varela  
ignacio.varela@unican.es

- <sup>1</sup> Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria-CSIC, Santander, Spain
- <sup>2</sup> Departamento de Patología, Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Tenerife, Spain
- <sup>3</sup> Servicio de Anatomía Patológica y Biobanco Valdecilla, HUMV/IDIVAL, Santander, Spain
- <sup>4</sup> Biobanco del Principado de Asturias (BBPA), Hospital Universitario Central de Asturias, Oviedo, Spain
- <sup>5</sup> Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain
- <sup>6</sup> Cancer Epigenetics Laboratory, The Francis Crick Institute, London, UK

## Introduction

Lung cancer is the major cause of cancer-related deaths worldwide with an average 5-year survival rate below 20% irrespective of the subtype [1]. Consequently, any new knowledge about the molecular mechanisms that drive this disease could have a great impact on the treatment of patients. Recently, large genomic projects have facilitated the identification of major players in this tumor type. Thus, small cell lung cancer (SCLC), which constitutes around 15% of all cases, is mainly driven by mutations in *TP53* and *RBI*, but the role of other genes like *PTEN*, *SLIT2*, or *CREBBP* has been also described [2]. Among non-small cell lung cancer (NSCLC), more than half of the cases are adenocarcinomas, where *TP53*, *KRAS*, *EGFR*, *ALK*, *ROS1*, and *BRAF* are the main recurrently altered genes [3], while squamous cell carcinomas (SCC) are genetically more heterogeneous and poor in actionable mutations so far.

Lately, several members of the SWI/SNF family of chromatin remodeling complexes have been identified recurrently altered in different tumor types adding to the

accumulated compelling evidence on the role of chromatin structure in cancer development. It is estimated that ~20% of all tumors contain alterations in these complexes, a frequency that is only exceeded by mutations in *TP53* [4]. In the case of NSCLC, the expression of any of the two mutually exclusive catalytic ATPase subunits (SMARCA2 or SMARCA4) is lost in 30% of the cases where it is associated with worse prognosis [5]. In addition, *ARID1A*, which encodes for one of the auxiliary subunits of the complex, frequently appears mutated in lung adenocarcinoma [6].

## Results

### Loss of ARID2 protein expression in 20% of lung cancer patients is partially explained by the presence of ARID2 mutations

In order to understand better the role of chromatin remodeling complexes in lung cancer development, we performed a genetic screening on the coding sequences of known cancer genes as well as members of the main chromatin remodeling complexes (Supplementary Table 2). We applied targeted next-generation sequencing technologies in a collection of 81 lung cancer cases (40 lung adenocarcinomas, 12 SCC, and 29 small cell carcinomas) (Supplementary Table 1). Interestingly, we found mutations in *ARID2* in five of the patients, three classified as lung adenocarcinoma and two as small cell carcinoma (Fig. 1a and Supplementary Table 3). In addition, to identify lung cancer driver genes, we run OncodriveFML [7] with our data. This software identifies genes with a number and distribution of predicted deleterious mutations higher than expected by chance, evidencing positive selection. *ARID2* ranked second in the list of genes showing significant positive selection after multi-test correction, just below *TP53* (Fig. 1b and Supplementary Table 4). To validate these results, we sequenced *ARID2* coding sequences in a second cohort of 144 lung adenocarcinoma cases and found mutations in 12 patients. If we consider all analyzed lung cancer patients, irrespective of the subtype, *ARID2* mutations occur at a frequency of 7.5 % (17/225) (Supplementary Table 3). In the case of lung adenocarcinomas (40 and 144 patients from first and second cohorts, respectively), we found *ARID2* non-synonymous mutations in 7% of the samples (13/184), which is near twice the frequency reported in COSMIC database for this tumor type (3.7%, 83/2241) [8] and ranks *ARID2* among the ten genes most commonly mutated in lung cancer. In concordance with a potential role of *ARID2* as tumor suppressor, many of the identified mutations, clustered at the beginning of the protein sequence, are predicted to generate a premature truncation of the protein (Fig. 1c and Supplementary Table 3). Subsequently, to check if the loss of *ARID2* function is a common feature in lung cancer, we performed

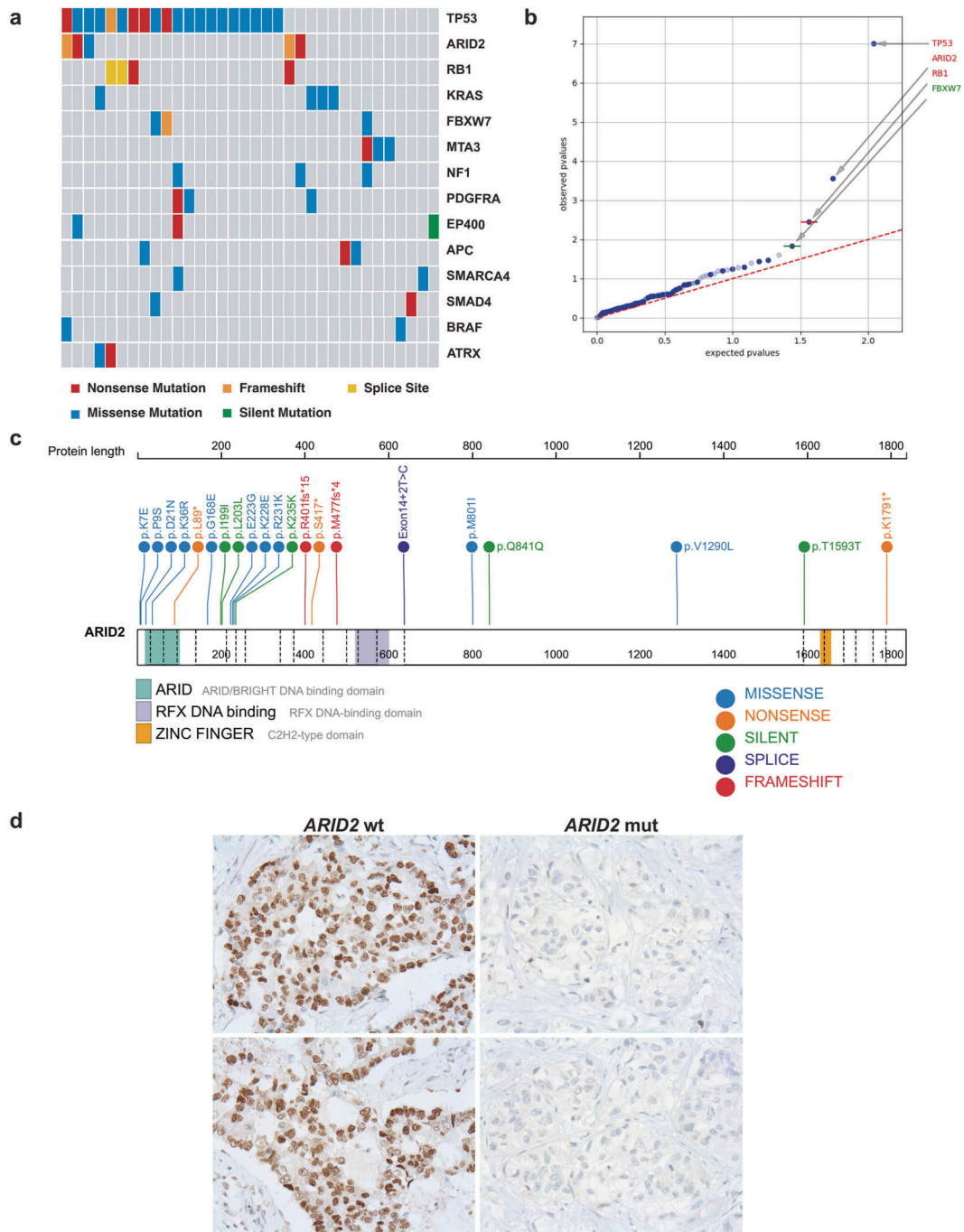
immunohistochemistry analyses in 139 of the studied samples finding loss or low/heterogeneous *ARID2* production in ~20% of the cases (28/139) (Fig. 1d and Supplementary Fig. 1). In addition, loss of *ARID2* expression was significantly more frequent in *ARID2*-mutated patients (6/10 Fisher exact test  $p = 0.0098$ ). Interestingly, this was also true for some samples with missense mutations, which suggests that these mutations might interfere with the correct folding or processing of the protein. Indeed, many of the mutations found are predicted to produce deleterious effects in the protein according to SHIFT or Polyphen algorithms (Supplementary Table 3). We observed a complete loss of *ARID2* signal in many lung cancer samples suggesting a selective pressure to inactivate both *ARID2* alleles. In addition, some non-mutated samples showed also loss of *ARID2* production, suggesting the existence of nongenetic mechanisms that interfere with *ARID2* expression. Since the presence of normal tissue contamination in the tumor samples as well as the sequencing strategy followed for normal tissue, prevented an estimation of the cellularity or the zygosity of the mutations, we could not determine whether full loss of *ARID2* protein is due to genetic loss or silencing of the wild-type allele.

### ARID2 deficiency increases proliferative and metastatic potential in vitro and in vivo

In order to check if alterations in *ARID2* could promote lung cancer development, we knocked down the protein in different *ARID2*-proficient NSCLC cell lines. As it can be observed in Fig. 2a, b, *ARID2* mRNA and protein production was efficiently reduced by two different shRNAs. This reduction was accompanied by an increase in the proliferation of A549 cells, as well as their invasion and migration capacities compared to those cells transduced with the empty vector. Similar results were obtained in NCI-H460 cell line (Fig. 2c, d and Supplementary Fig. 2). Moreover, when these cells were injected into immunocompromised mice, they showed a greater capacity to produce tumors in vivo (Fig. 2e and Supplementary Fig. 3).

RNA-Seq experiments in A549 transduced cell lines showed that loss of *ARID2* was accompanied by significant changes in the expression of 1155 genes (366 upregulated and 789 downregulated), that supported the observed phenotypes in the cells (Fig. 2f and Supplementary Table 5). Thus, we observed a downregulation of genes involved in cellular adhesion and cell differentiation such as *NPNT*, *CDH6*, *FAT3*, *FNI*, *SOX2*, or *SDC2* as well as an upregulation of genes associated with a higher cell-cycle progression such as *CDC45*, *MCM2*, or *HIST1H1E*, which could be associated with the increased proliferation, migration, and invasion capacities of *ARID2*-deficient cells. In addition, we observed downregulation of other tumor suppression genes like *RPS6K2*, *TNFSF10*, *ISM1*, or *LDLRAD4* together with upregulation of



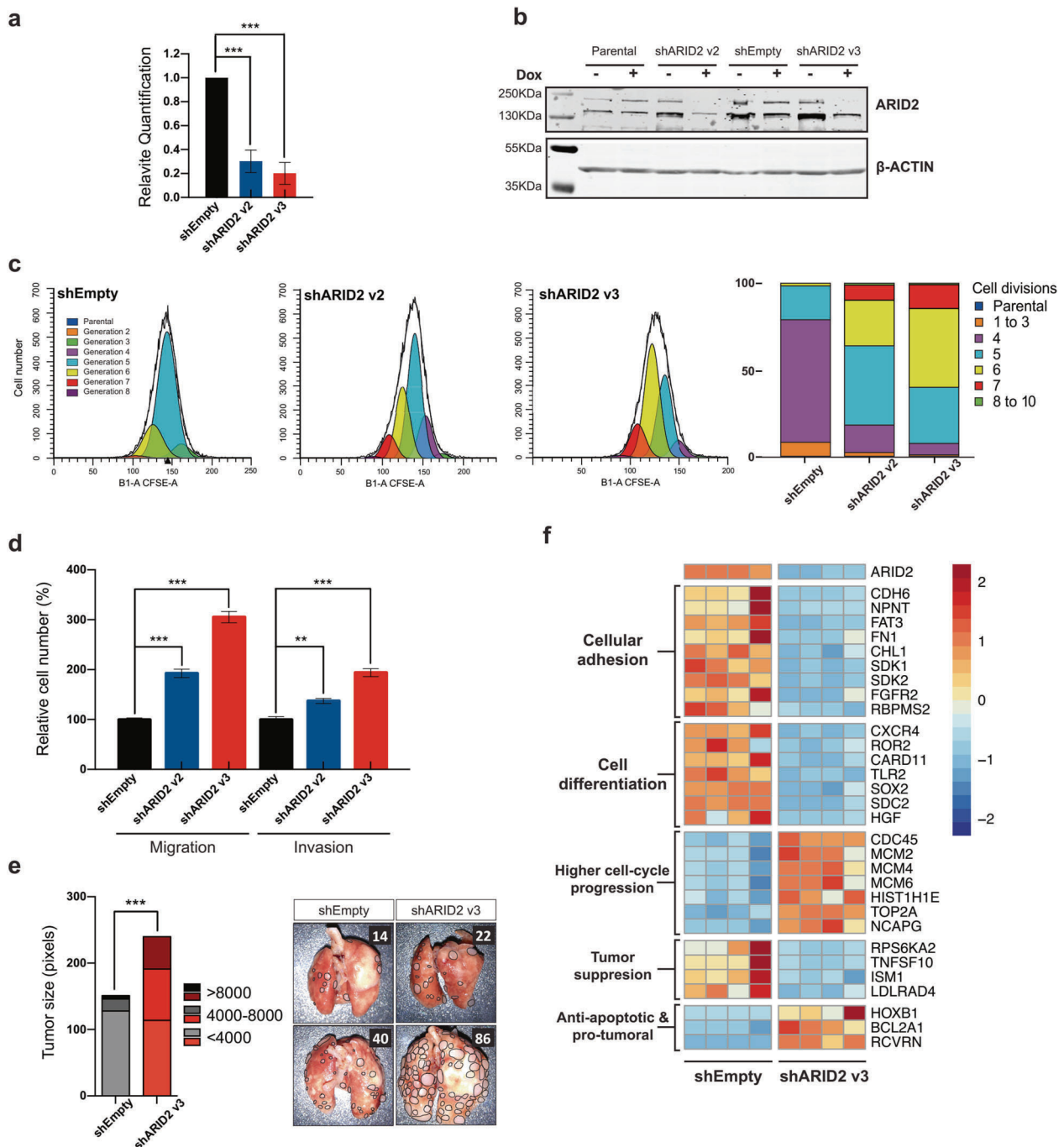


**Fig. 1 Frequent ARID2 mutations associated with protein loss.** **a** Box representation of the mutated patients for the most significantly mutated genes according to OncodriveFML in the lung cancer cohort generated with Maftools. Each box in the central matrix represents an independent patient. Colored boxes represent mutated patients for the corresponding gene in a color code indicating the type of mutation. **b** Representation of the significance analysis of the functional impact

of the mutations found in each gene performed by OncodriveFML. Genes in read showed a  $q$  value  $< 0.1$  after multi-test correction. **c** Visual representation of the location of all identified ARID2 mutations in our discovery and validation lung cancer cohort in relation to the functional protein domains. **d** Representative images of ARID2 immunohistochemistry experiments in two ARID2-mutated (right) and two ARID2-wild-type (left) lung adenocarcinomas.

protumoral and antiapoptotic genes like *HOXB1*, *BCL2A1*, or *RCVRN*. Most of these alterations were not observed when we knocked down ARID1A or ARID1B subunits in the same cells (Supplementary Fig. 4a), which indicates a specific gene-set

regulation by ARID2-containing SWI-SNF complexes. Transcriptional changes in selected genes were further validated by qRT-PCR in independently generated ARID2 knockdown cell lines (Supplementary Fig. 4b).



Altogether, these results prove that *ARID2* plays a tumor suppressor function in lung cancer.

### **ARID2 deficiency is accompanied by widespread chromatin changes, specially affecting enhancers**

We hypothesized that gene expression changes observed in *ARID2*-deficient cells might be the result of changes in SWI/SNF chromatin remodeling activity. To investigate this, we performed ATAC-Seq experiments in *ARID2*

knocked down A549 cells. *ARID2* loss was accompanied by a general loss of chromatin accessibility with 990 regions that showed a significant loss of chromatin accessibility versus 687 regions that showed increased accessibility (Supplementary Table 6). Interestingly, those regions that lost accessibility in the absence of *ARID2* were located distal to gene transcription start sites (Fig. 3a) and showed enrichment of enhancer-specific H3K4me1 and H3K27ac histone marks according to ENCODE project data. An opposite behavior is observed on those regions

◀ **Fig. 2 ARID2 deficiency is associated with an increase in oncogenesis in vitro and in vivo.** **a** Bar representation of ARID2-expression level fold changes measured by qRT-PCR in A549 cells transduced with shARID2 v2 and v3 as well as the empty vector which is used as control. Data are shown as mean  $\pm$  SEM of three independent experiments, relative to control cells A549 Empty vector (black bars), (two-tailed *t*-test \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ ). **b** Representative image of a western blot analysis measuring ARID2 protein levels in A549 parental cells as well as those cell lines transduced with ARID2 shRNAs and the empty vector. In all the cases, the results are shown with and without induction of the shRNA expression by doxycycline (Dox) treatment. **c** Representative experiment of the number of cell divisions suffered by the cells in 48 h estimated by CFSE labeling in A549 cells by flow cytometry. Bar quantification on the number of cells that have suffered each number of cell divisions is represented on the right. **d** Bar representation of quantified cells in destination chamber on migration and invasion assays of A549 cells transduced with two different ARID2 shRNAs (blue and red bars). Data are shown as mean  $\pm$  SEM of three independent experiments, relative to control cells A549 Empty vector (black bars), (two-tailed *t*-test \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ ). **e** Representative images of lung metastasis generated in intravenously injected mice with A549 cells transduced either with shEmpty, or shARID2v3 vectors ( $n = 7$  per group). Individual metastasis is delineated in the image and counted (upper right corner numbers). On the left, a quantification of the number and size of the tumors generated in the two groups is shown. (Fisher exact test \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ ). **f** Heatmap representation of a selection of differently expressed genes in ARID2-deficient A549 cells ( $n = 4$ ) and grouped according to their biological function. Expression differences go from red (upregulation) to blue (downregulation) according to the log<sub>2</sub> of the fold change.

that gained accessibility after ARID2 loss (Supplementary Fig. 5). In addition, AP-1 family transcription factor binding motif, described as abundantly present in enhancers [9], is highly enriched on those regions that showed less accessibility on ARID2-deficient cells (Fig. 3b). In order to explore a special impact of ARID2 loss on enhancers, we analyzed the accessibility of those regions annotated as enhancers in the GeneHancer project [10]. As it can be seen in Fig. 3c, the loss of accessibility is significantly more profound in enhancers than in the rest of the genome, as 87% (1744 of 2001) of the enhancers that showed significant accessibility changes, lose accessibility in ARID2-deficient cells. In addition, many of the target genes of these enhancers showed significant downregulation in the RNA-Seq data (Fig. 3d). All this support that *ARID2* is essential to keep an open chromatin conformation around enhancers, which significantly impacts on the transcriptional regulation of specific gene networks (Supplementary Table 6).

### ARID2 is essential to maintain the expression of the metastasis inhibitor *MTSS1* and the adhesion molecule *SDK1*

In order to identify ARID2 target genes of broad relevance to lung cancer patients, we compared our RNA-Seq results with differential expression analysis

performed on human lung adenocarcinoma patients from TCGA database. Eighteen genes were found upregulated in both ARID2-deficient cells, and in low-ARID2-expressing lung adenocarcinoma patients (Supplementary Table 7). Among them, we found *AREG*, *ERG*, or *NGF* growth factors that might explain the higher proliferation capabilities of ARID2-deficient cells (Fig. 3e and Supplementary Table 7). In addition, we found 133 genes downregulated in both datasets indicating a main gene expression activating role of ARID2 in this cellular context.

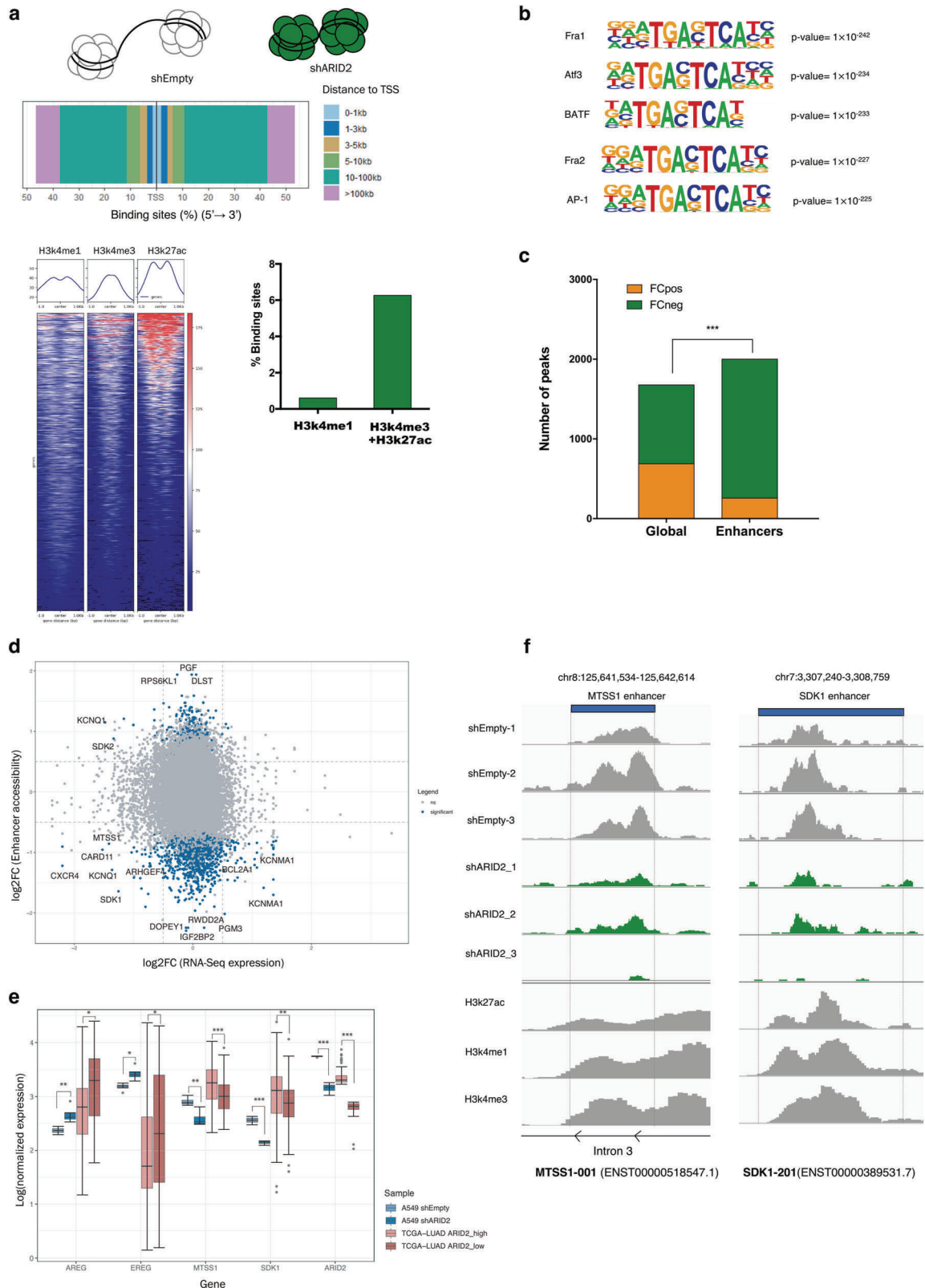
Interestingly, among those genes whose expression rely on ARID2, we found *MTSS1*, a well-described metastasis inhibitor [11], as well as *SDK1*, involved in cell–cell adhesion. The observed significant reduction of *MTSS1* and *SDK1* expression likely explain the higher invasion capabilities of ARID2-deficient cells. In addition, we found a significant reduction of chromatin accessibility on two enhancers regulating these genes in GeneHancer database after ARID2 loss [10]. This observation is concordant with the hypothesis that *MTSS1* and *SDK1* expressions are positively regulated by ARID2 by keeping an open chromatin structure at their enhancers (Fig. 3f).

### ARID2 loss impairs DNA damage repair

The RNA-Seq analysis also revealed deleterious consequences suffered by ARID2-deficient cells that could be exploited therapeutically. Gene-set Enrichment Analyses (GSEA) on the transcriptional alterations observed in ARID2-deficient cells also showed a significant upregulation of genes involved in DNA damage detection and repair, suggesting a defective DNA damage response (Fig. 4a). Supporting a role of ARID2 in DNA repair, analysis of its localization in untransduced A549 cells showed a colocalization with  $\gamma$ H2AX and 53BP1 at the DNA-repair foci (Fig. 4b and Supplementary Fig. 6). In addition to that, ARID2-deficient cells showed a delay in the resolution of DNA damage foci compared to wild-type cells in A549 and NCI-H460 NSCLC cell lines upon treatment with etoposide. Interestingly, this delay was not dependent on TP53 function as we could see a similar delay in ARID2-deficient NCI-H1568 cell line, which is TP53-deficient (Fig. 4c, d). These observations indicate that ARID2 deficiency inhibits efficient DNA repair and suggest that its loss may sensitize cells to DNA-damaging agents.

### ARID2 deficiency increases cell sensitivity to chemotherapy and veliparib

As platinum-based chemotherapy is widely used for the treatment of lung cancer patients [3], we first examined



ARID2-deficient cell lines sensitivity to cisplatin, as well as etoposide. As it can be seen in Fig. 4e and Supplementary Fig. 7, in concordance with defective DNA repair in the

absence of ARID2, ARID2-deficient A549, and NCI-H460 cells exhibited a higher sensitivity to both compounds compared to control cells.



**Fig. 3 Profound chromatin structural changes on enhancers affect gene expression after ARID2 loss.** **a** Analysis of the genomic regions that significantly lost chromatin accessibility after ARID2 loss in A549 cells. In the upper panel the regions are grouped according to their distance to nearest gene transcription start site (TSS). Below, the intensity of H3K4me1, H3K4me3, and H3k27ac histone marks in each identified region is represented by heatmaps (left). In addition, the percentage of identified regions that overlap with regions with histone modification marks are represented in a bar graph (right). **b** Enrichment of sequence motifs identified by HOMER in those regions that lost chromatin accessibility after ARID2 loss in A549 cells. **c** Bar representation of the number of genome-wide (global) regions or enhancer regions, which lost (FCneg green bars) or gained (FCpos yellow bars) after ARID2 loss in A549 cells. (Fisher exact test  $***p < 0.001$ ). **d** Dot plot representing the correlation between the accessibility changes in enhancer regions and expression changes on the target genes for each enhancer. Blue dots represent enhancers that showed significant accessibility changes in ARID2-deficient A549 cells. **e** Boxplot graph of gene expression differences identified in both our ARID2-deficient A549 cells and in lowly *ARID2*-expressing lung adenocarcinoma patients (*ARID2\_low*) versus highly *ARID2*-expressing patients (*ARID2\_high*) from TCGA database, (DEseq2 statistical test  $*p < 0.05$ ,  $**p < 0.01$ , and  $***p < 0.001$ ). **f** Visualization, in two described *MTSS1* and *SDK1* enhancers, of read alignments for the different replicates of our ATAC-Seq experiments in ARID2-deficient A549 cells. In addition, read alignments of ChIP-Seq experiments performed against different histone marks during ENCODE project are also represented.

In addition, in the last decade, many researchers have described a higher sensitivity of PARP inhibitors in tumors harboring defects in DNA-repair mechanisms due to synthetic lethality [12]. Consequently, we checked if this might apply as well to ARID2-deficient cells. As it can be seen in Fig. 4e, ARID2 loss led to a higher sensibility to veliparib, a well-described PARP inhibitor that is under research in several clinical trials in breast, ovarian, and, most importantly, lung cancer. This observation suggests ARID2 deficiency as a useful marker for the stratification of lung cancer patients that may benefit for PARP inhibitor treatment.

## Discussion

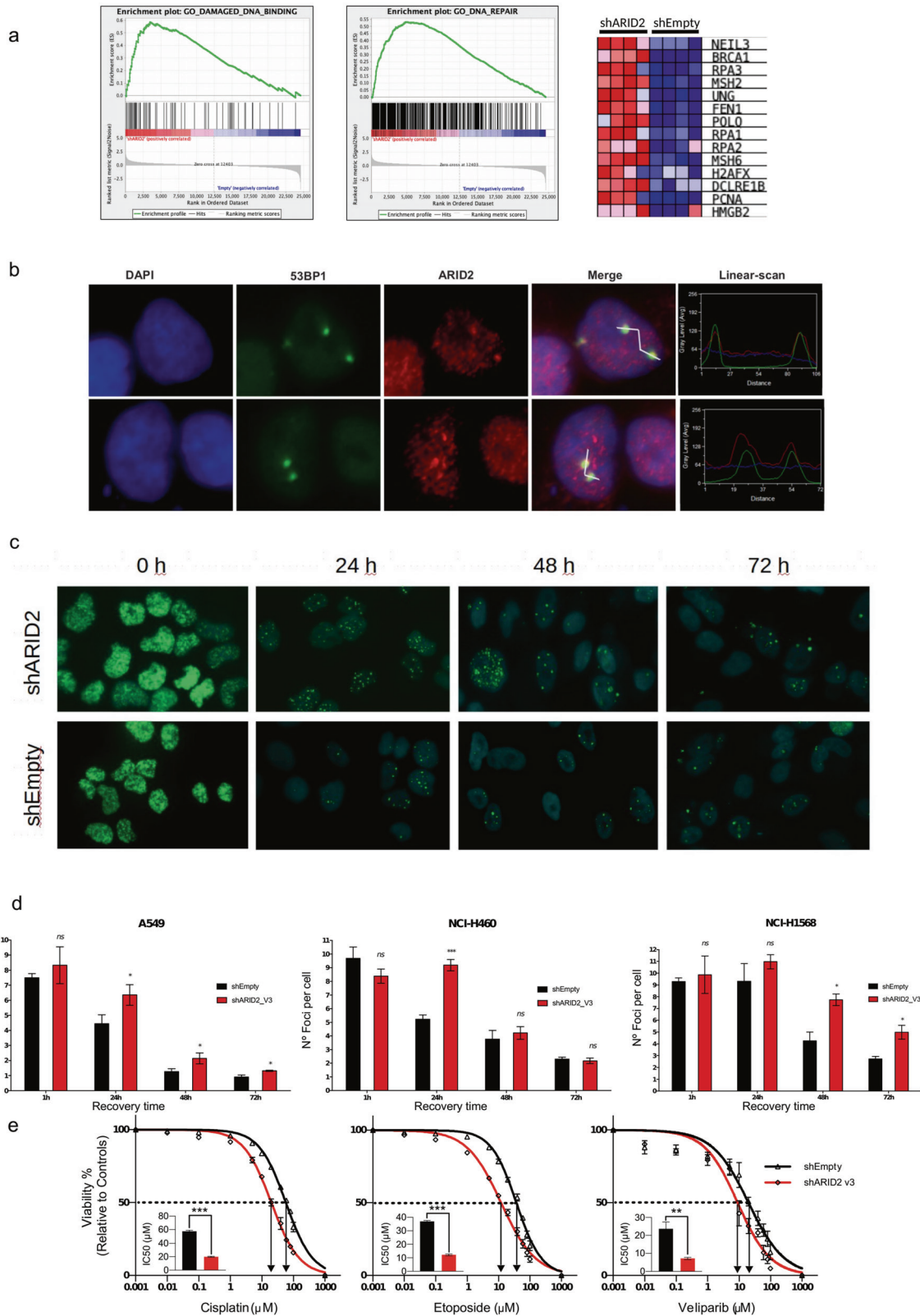
Although some evidence of the presence of *ARID2* alterations in lung cancer have been reported previously [13], the relevance of these alterations for oncogenesis has not been clearly proved. Our results showed an *ARID2* mutation recurrence higher than the one reported in COSMIC database. In addition, the distribution and predicted impact of the mutations found and our in vitro and in vivo experiments provided compelling evidence of the role of *ARID2* as bona fide tumor suppressor in lung cancer. Supporting this, *ARID2* has been already proposed as cancer driver gene in melanoma and hepatocellular carcinoma [14, 15]. We don't have enough data to definitely prove that complete ARID2 activity loss is necessary for tumor progression but several evidences

support this idea. First, in many cases, we observed a complete loss of ARID2 expression in the human adenocarcinoma samples, which suggests a selective pressure to inactivate both *ARID2* alleles. In addition, we observed that V2THS\_74399 shRNA construct was less efficient in abrogating ARID2 expression than V3THS\_347660 and cells transduced with the former typically showed less-pronounced changes than cells transduced with the latter.

The precise molecular mechanisms by which alterations in chromatin remodeling complexes promote cancer development are not sufficiently understood. Interactions with well-described cancer genes like *TP53*, *RB*, or *MYC* have been described [16–18]. In addition to this, they play essential roles in the activation of differentiation and the suppression of proliferative programs of many cellular lineages [19]. In this study, we described a list of near 200 genes that are specifically deregulated after ARID2 loss in both our cellular model and in lung adenocarcinoma patients from the TCGA database. Some of these genes, like *AREG*, *EREG*, or *NGF* growth factors might account for the higher proliferation capabilities of ARID2-deficient cells. In terms of the molecular mechanisms behind this regulation, we show that ARID2 deficiency is associated with widespread chromatin structural changes. Our results prove that ARID2 is essential to keep an open chromatin structure in enhancer regions in agreement with an important role of different SWI/SNF members in regulating enhancer activity [20, 21]. Two of these ARID2-dependent enhancers regulate *MTSS1* and *SKD1* expression that, consequently, showed a significant downregulation in ARID2-deficient cells. This suggests that ARID2 might regulate directly the expression of *MTSS1* and *SKD1*, although further work is necessary to finally confirm this. *MTSS1* is a well-described migration and invasion inhibitor, associated with worse prognosis in several tumor types [11, 22, 23], and *SKD1* plays important roles in cell–cell adhesion. Their deficiency might well explain the higher migration and invasion capabilities of ARID2-deficient cells.

In addition, we observed an active role of ARID2 in the detection and repair of DNA damage in vitro in lung cancer cell lines, as ARID2-deficient cells present important delays in the resolution of DNA damage foci that were not dependent on TP53 activity. In accordance with this view, other members of the SWI/SNF complex have been shown to be involved in different steps of DNA damage repair [24–26].

Finally, any advance in the possibility of exploiting therapeutically any vulnerability associated to deficiency in SWI/SNF complex genes is of great interest, as ~20% of all human cancers are reported to have alterations in this complex. In this study, ARID2-deficient cells showed a higher sensitivity to different DNA-damaging therapies, likely as a result of the ARID2 involvement in DNA repair. Considering that platinum-based chemotherapy is still



widely used in lung cancer patients with high variable success [3], our results suggest that ARID2 expression might be explored as a stratification marker for these

therapies. Moreover, we show that ARID2 deficiency shows synthetic lethality with PARP inhibition using veliparib, an inhibitor that has shown good results in the treatment of

◀ **Fig. 4 ARID2 deficiency affects DNA repair and affects sensitivity to antitumor therapies.** **a** Results of the Gene-Set Enrichment Analysis (GSEA) from RNA-Seq experiments in ARID2-deficient A549 cells showing enrichment of genes involved in different DNA-repair ontologies. On the right, a heatmap representation of the expression of different genes belonging to these ontologies in the different replicates is included. **b** Representative images of immunofluorescence experiments demonstrating colocalization of 53BP1 (green) and ARID2 (red) in A549 cells in DNA damage foci. Colocalization was confirmed on the right through the parallel quantification of red and green signals on a manually selected path through the image using the LineScan tools from Methamorph software. **c** Representative images of DNA-repair foci visualized by H2AX immunofluorescence (green) in transduced nuclei stained with DAPI (blue) at different recovery times after the treatment with etoposide in NCI-H460 cell lines. **d** Bar representation of the foci quantification in each transduced cell line. **d** Bar quantification of the number of foci per cell at different recovery times after DNA damage induced by etoposide, in different ARID2-deficient cell lines, the results are represented as mean  $\pm$  SEM of at least three independent experiments, (two-tailed *t*-test \**p* < 0.05, \*\**p* < 0.01, and \*\*\**p* < 0.001). **e** Representative experiments measuring cell survival to increasing concentrations of cisplatin, etoposide, and veliparib on A549 cells transduced with shEmpty (black), or shARID2v3 (red) vectors. Bar graphs represent the calculated IC50 value for each experiment. In all cases, the results are represented as mean  $\pm$  SEM of at least three independent experiments, (two-tailed *t*-test \**p* < 0.05, \*\**p* < 0.01, and \*\*\**p* < 0.001).

breast cancer [27] and is included in several clinical trials on breast, ovarian, and, most importantly, lung cancer. Our results suggest that the stratification of lung cancer patients according to ARID2 expression might improve the efficiency of PARP inhibitors in NSCLC. In addition, a very recent study has shown that ARID2-deficient melanoma cells are particularly sensitive to immunotherapy through alterations in mTORC1 and IFN $\gamma$  pathways [28]. Interestingly, we observed that some downstream response genes in these pathways such as *GBP2*, *GBP3*, and *SCD5* are significantly downregulated in ARID2-deficient cells (Supplementary Table 5). All these results support the potential use of ARID2 expression as a new stratification marker for personalized treatment in lung cancer patients.

In summary, here we present compelling evidence for the role of ARID2 as tumor suppressor in lung cancer. Although ARID2 has been proposed as a driver gene in other tumor types [14, 15], little has been reported about the molecular mechanisms underlying this involvement. In this work, we propose that its role in lung cancer is exerted by fostering a specific pro-oncogenic transcriptomic program as a result of changes in chromatin structure around enhancers. Importantly, our results indicate that ARID2 deficiency could be exploited for lung cancer patient treatment.

## Materials and methods

Detailed protocols can be found in Supplementary Methods.

## Next-generation sequencing

Cancer patient primary tumor samples and, when available, matched corresponding normal samples, were obtained from different tumor biobanks after the corresponding approval of the hospital ethics committees and patient informed consent. A detailed list of the origin and characteristics of each sample can be found in Supplementary Table 1. DNA was extracted using the Agencourt DNAdvance Beckman Coulter kit (Beckman Coulter, USA), fragmented, and submitted to end-repair and adenylation, adaptor ligation, and PCR indexing amplification. Target capture was performed using a Sure Select<sup>®</sup> user-defined probe kit (Agilent Technologies, USA).

For ATAC-Seq libraries, cell nuclei were extracted using a cold lysis buffer and submitted to tagmentation (Nextera DNA Library Preparation Kit, Illumina, USA). After purification, adapter sequences were used to complete Illumina sequencing adapters by PCR with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, UK).

Total RNA was purified using Extract Me Total RNA Kit (Blirt, USA). Reverse transcription was performed using the Takara PrimeScript cDNA Synthesis kit (Takara Bio Europe, France). Poly-A mRNA was enriched, fragmented, and submitted to cDNA generation using PrimeScript Enzyme for first strand and RNase HI, DNA polymerase I and T4 DNA Polymerase (Thermo Fisher Scientific, UK) for the second strand. Afterwards, genomic libraries were generated as above. Individual mRNA expression was measured by qRT-PCR using Luminaris Color HiGreen qPCR Master Mix (Thermo Fisher Scientific, UK).  $\beta$ -actin was used as housekeeping gene and the  $\Delta\Delta$ Ct method was used for quantification and comparison.

## Sequencing data analysis

DNA sequence data were mapped to the human genome (hg19) using BWA 0.7.3 [29]. In addition, Samtools 0.1.18 [30], Picard 1.61 (<http://broadinstitute.github.io/picard/>), and GATK 2.2.8 [31] were used for format transformation, cleaning, sorting and indexing of the bam files, marking PCR duplicates, and performing indel local realignment. RAMSES [32] and PINDEL 0.2.4d [33] were used for substitutions and small insertion and deletion identification, respectively. All ARID2 mutations were validated by PCR amplification coupled with ultrasequencing at 10,000 $\times$  coverage. In addition, a similar orthogonal validation of more than 180 mutations randomly picked showed a near 80% of specificity in the mutation calling. OncodriveFML software was run to detect genes with evidence of positive selective pressure [7].

ATAC-Seq reads were aligned against the human genome (hg19) using BWA 0.7.3 [29]. Accessible regions were



identified using MACS 2.1.2 [34]. A combined list of all the regions identified in all the samples by MACS, as well as the list of enhancers annotated in the GeneHancer project [10], were used to identify significant changes in region accessibility in ARID2-deficient cells versus control using DESeq2 [35]. Region annotation was performed using ChIPSeeker software [36]. BEDTools [37] was used to estimate the overlapping of the identified regions with ENCODE publicly available histone marks ChIP-Seq A549 data and the results were plotted using deeptools v3.3.1 [38]. Motif enrichment analysis was performed using HOMER [39]. Finally, alignments were visualized using IGV genome browser [40].

RNA-Seq data were aligned using Tophat [41] to the human genome (hg19). Differentially expressed genes (DEG) were identified using HTSeq + DESeq2 [42, 43]. GSEA on gene ontology terms was performed using GSEA software [44]. For the analysis of the TCGA database lung adenocarcinoma patients, raw counts for each patient ( $n = 524$ ) were downloaded and normalized using DESeq2 software. Low-ARID2-expressing lung adenocarcinoma patients ( $n = 64$ ) were defined as those with a normalized expression less than the mean minus two standard deviations calculated from the whole cohort. Similarly, high ARID2-expressing lung adenocarcinoma patients ( $n = 78$ ) were defined as those with a normalized expression higher than the mean plus two standard deviations. DEG between two groups were identified using DESeq2.

### Cell culture and in vitro assays

A549, NCI-H1568, and NCI-H460 lung cancer cell lines were obtained from The Francis Crick Institute common and ATCC repositories, authenticated by STR profiling, and tested for mycoplasma. Tetracycline-inducible pTRIPZ constructs V2THS\_74399 (v2), V3THS\_347660 (v3) (*ARID2*), V2THS-283735 (*ARID1A*), and V2THS-11753 (*ARID1B*) were used for stable cell line generation (Dharmacon/GE Healthcare, USA). The empty vector (RHS4750) was used as control. Virus production was performed by transfecting HEK293T/17clone cells with the pTRIPZ and packaging constructs. Infected cells were selected with 1- $\mu$ g/ml puromycin and isolated by FACS using a FACS-Aria II cell sorter (Becton Dickinson, USA) based on TurboRFP expression after the induction with 1  $\mu$ g/ml of Doxycycline.

Extrapolated growth curves were constructed over a period of 14 days by serial passaging and cell counting with a hemocytometer or by the PrestoBlue<sup>®</sup> assay (Thermo Fisher Scientific, UK). Cell proliferation was also analyzed using the CellTrace<sup>™</sup> CFSE Cell Proliferation Kit (Invitrogen, USA) in cells synchronized by gradual serum deprivation following published protocols [45]. Cells were harvested at 48 h and subjected to division peak resolution

by flow cytometry. The cell proliferation index was analyzed using MODFIT software (Verity, USA). Proliferation index was the sum of the cells in all generations divided by the calculated number of original parent cells.

In vitro cell migration assays were performed by using 8- $\mu$ m pore size transwell chambers (Corning<sup>™</sup> Transwell<sup>™</sup> Multiple Well Plate) in 24-well plates using 10% FBS in the lower chamber as chemo-attractant. For invasion assays, cells were plated on growth factor-reduced Matrigel (BD Biosciences) pre-coated 8- $\mu$ m pore transwell chambers. Filters or invasive cells were quantified by fixing chambers in 4% paraformaldehyde for 10 min and staining with crystal violet.

Growth inhibition assays were performed to determine the half maximal inhibitory concentration (IC<sub>50</sub>) values for different antitumoral drugs. Viability after 48 h was determined by PrestoBlue<sup>®</sup> reagent (Thermo Fisher Scientific, UK). IC<sub>50</sub> value for each drug was determined with Prism software (GraphPad, USA).

### In vivo tumorigenesis assays

Animal studies were conducted in compliance with guidelines for the care and use of laboratory animals and were approved by the Ethics and Animal Care Committee of Universidad de Cantabria. For proliferation assays, five million cells in 500  $\mu$ l of PBS were subcutaneously injected into the flanks of 6–8-week-old female nude mice (Athymic Nude-Foxn1nu, Envigo, UK). Twenty-four days after the injection, mice were euthanizing and tumor tissues were harvested for analyses. For metastasis assays, 2.5 million of cells in 500  $\mu$ l of PBS with 0.1% BSA were tail-injected into 6–8-week-old female nude mice. Hairpin expression in the cells was induced with 1  $\mu$ g/ml of Doxycycline 7 days before injection. After 2 months, mice were euthanizing and tumor tissues were harvested for analyses. In both cases, to keep the hairpin expression, the animals were treated from the day of injection with 2 mg/mL of Doxycycline in the drinking water supplemented with 1% sucrose refreshed every 2–3 days.

### Western blot analysis

Total protein lysates were prepared in RIPA buffer (50-mM Tris-HCl, pH 8.0, 150-mM NaCl, 1 % NP-40, 1-mM Sodium Orthovanadate, 1-mM NaF), separated by SDS-PAGE in 8% polyacrylamide gels, and transferred to nitrocellulose membranes. Subsequently, membranes were washed with TBST (50-mM TRIS + 150-mM Sodium chloride + 0.1% Tween 20, pH 7.4) and blocked using 5% nonfat milk solution in TBS for 1 h at room temperature. Membranes were then incubated with primary antibodies anti-ARID2 (E-3, Santa Cruz) and anti-Actin (I-19, Santa



Cruz), diluted 1:200 and 1: 1000 in TBST with 5% (w/v) BSA at 4 °C overnight, respectively. Donkey anti-mouse or donkey anti-goat secondary antibodies (LI-COR Biotechnology, USA) conjugated to IRDye 800CW (926-32212) or IRDye 680RD (926-68074), respectively, were used as secondary antibodies.

### Immunohistochemistry analysis

For ARID2 detection on paraffin sections, antigen retrieval was performed for 32 min at 97 °C in citrate buffer pH 6, incubated with 1:300–1:500 anti-ARID2 antibody (abcam ab113283) and developed with HRP-polymer secondary antibodies (Optiview, Roche). ARID2 expression was evaluated by two pathologists on coded tissue sections, without information about the *ARID2* mutation status. Only surgical pathology cases with enough material, both tumor and non-neoplastic surrounding tissue, were considered for ARID2 immunohistochemistry. A consensus score was reached viewing the slides by two pathologists at a multiheaded scope.

Immunofluorescence was performed in cells fixed with 4% paraformaldehyde in PBS for 15 min at room temperature. The cells were permeabilized with 0.5% Triton X-100 in PBS and blocked with 3% BSA in PBT (PBS containing 0.05% Triton X-100). Finally there were subjected to immunofluorescence staining with ARID2 antibody (E-3, sc-166117 Santa Cruz, USA or A302-230A, Bethyl Laboratories, USA), anti-phospho-Histone H2A.X Ser139 ( $\gamma$ H2AX, clone JBW301, Merck Millipore, USA), or anti-53BP1 antibody (H-300, sc-22760, Santa Cruz, USA). Cover slides were incubated with Alexa-labeled secondary antibodies and mounted in VECTASHIELD Antifade Mounting Medium with DAPI (Vector Labs, USA). Colocalization of ARID2 with 53BP1 or  $\gamma$ H2AX was performed measuring the variation in intensity across the lines drawn using the linescan tool from MetaMorph® (Molecular Devices, USA). For DNA-repair assays, cells were treated with 10- $\mu$ M Etoposide for 1 h. Subsequently,  $\gamma$ H2AX foci were quantified using ImageJ software at different recovery times after removing the drug from the media.

### Statistical analysis

In all cases, at least three independent experiments were performed in order to assess the statistical significance of all differences. In figure legends, the specific statistical test performed in each case is indicated. In general, for quantitative variables, a one-tailed *t*-test with equal variance was used to identify significant differences between groups. For qualitative variables, a Fisher exact test was used in order to identify significant differences between groups of patients.

The different software used for the identification of mutations, gene expression differences, enrichment transcription factor binding sites, and gene ontology terms have their own statistical models explained in detailed in the references. When multiple tests were performed, the significance is shown corrected for multiple testing.

### Data availability

DNA-Seq, ATAC-Seq, and RNA-Seq data: <https://www.ebi.ac.uk/ena/data/search?query=PRJEB26936>.

### Code availability

All computer code is available upon request.

**Acknowledgements** In memoriam of Carlos Revilla, we will miss you forever. We would like to thank the support of the funding agencies Ministerio de Economía y Competitividad, Fundación Ramón Areces, European Research Council, Asociación Española contra el cáncer, Cancer Research UK, UK Medical Research Council, Wellcome Trust, Servicio de Salud del Principado de Asturias, Instituto de Salud Carlos III and Fundación Bancaria Cajastur (specific grant references are included in the funding support section). We will like to thank as well the technical support of the different institutions and common services as well as the patients that agreed to participate in this study. We finally want to thank Dr Francisco Real, Dr Roland Rad, Dr Jose Pedro Vaque, and Dr Javier Leon for providing critical reagents and advice as well as to all the patients that agreed to participate in this study. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

**Funding** IV is supported by SAF2012-31627 and SAF2016-76758-R grants from the Spanish Ministerio de Economía y Competitividad (MINECO), by a Fundación Ramón Areces grant and by ERC2014-StG637904 grant from the European Research Council. IV has been awardee of the Programa Ramón y Cajal (MINECO, Spain). TM has been awardee of the Ayudas para la contratación de investigadores predoctorales (MINECO, Spain). BM is awardee of the Ayudas para la formación de profesorado universitario (FPU, Ministerio de Educación y Formación Profesional, Spain). PC laboratory is supported by grant SAF-2015-63638R (MINECO/FEDER, UE); by Centro de Investigación Biomédica en Red de Cáncer (CIBERONC) and by Asociación Española Contra el Cáncer (AECC), grant GCB141423113. BC has been supported by a Retos Jóvenes Investigadores grant SAF-2015-73364-JIN (AEI/FEDER, UE) and a grant from Fundación Francisco Cobos. PS is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001152) and the UK Medical Research Council (FC001152). HUCA/IUOPA is jointly financed by Servicio de Salud del Principado de Asturias, Instituto de Salud Carlos III, and Fundación Bancaria Cajastur. This research was funded in part by the Wellcome Trust [FC001152].

### Compliance with ethical standards

**Conflict of interest** The authors declare no competing interests.

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

## References

- Lovly CM, Carbone DP. Lung cancer in 2010: one size does not fit all. *Nat Rev Clin Oncol*. 2011;8:68–70.
- Gelsomino F, Rossi G, Tiseo M. MET and small-cell lung cancer. *Cancers*. 2014;6:2100–15.
- Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer*. 2014;14:535–46.
- Maslah-Planchon J, Bièche I, Guinebretière J-M, Bourdeaut F, Delattre O. SWI/SNF chromatin remodeling and human malignancies. *Annu Rev Pathol Mech Dis*. 2015;10:145–71.
- Reisman DN, Sciarrotta J, Wang W, Funkhouser WK, Weissman BE. Loss of BRG1/BRM in human lung cancer cell lines and primary lung cancers: correlation with poor prognosis. *Cancer Res*. 2003;63:560–6.
- Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012;150:1107–20.
- Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol*. 2016;17:128. <https://doi.org/10.1186/s13059-016-0994-0>.
- Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017;45:D777–83.
- Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res*. 2013;23:777–88.
- Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*. 2017;2017. <https://doi.org/10.1093/database/bax028>.
- Kayser G, Csanadi A, Kakanou S, Prasse A, Kassem A, Stickeler E, et al. Downregulation of MTSS1 expression is an independent prognosticator in squamous cell carcinoma of the lung. *Br J Cancer*. 2015;112:866–73.
- Lord CJ, Ashworth A. PARP inhibitors: synthetic lethality in the clinic. *Science*. 2017;355:1152–8.
- Manceau G, Letouzé E, Guichard C, Didelot A, Cazes A, Corté H, et al. Recurrent inactivating mutations of *ARID2* in non-small cell lung carcinoma. *Int J Cancer*. 2013;132:2217–21.
- Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat J-P, et al. A landscape of driver mutations in melanoma. *Cell*. 2012;150:251–63.
- Li M, Zhao H, Zhang X, Wood LD, Anders RA, Choti MA, et al. Inactivating mutations of the chromatin remodeling gene *ARID2* in hepatocellular carcinoma. *Nat Genet*. 2011;43:828–9.
- Flowers S, Beck GR, Moran E. Transcriptional Activation by pRB and Its Coordination with SWI/SNF Recruitment. *Cancer Res*. 2010;70:8282–7.
- Tordella L, Khan S, Hohmeyer A, Banito A, Klotz S, Raguz S, et al. SWI/SNF regulates a transcriptional program that induces senescence to prevent liver cancer. *Genes Dev*. 2016;30:2187–98.
- Romero OA, Torres-Diz M, Pros E, Savola S, Gomez A, Moran S, et al. MAX inactivation in small cell lung cancer disrupts MYC-SWI/SNF programs and is synthetic lethal with BRG1. *Cancer Discov*. 2014;4:292–303.
- Wilson BG, Roberts CWM. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer*. 2011;11:481–92.
- Alver BH, Kim KH, Lu P, Wang X, Manchester HE, Wang W, et al. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun*. 2017;8:14648.
- Nakayama RT, Pulice JL, Valencia AM, McBride MJ, McKenzie ZM, Gillespie MA, et al. SMARCB1 is required for wide-spread BAF complex-mediated activation of enhancers and bivalent promoters. *Nat Genet*. 2017;46:1613–23. <https://doi.org/10.1038/ng.3958>.
- Giacobbe A, Compagnone M, Bongiorno-Borbone L, Antonov A, Markert EK, Zhou JH, et al. p63 controls cell migration and invasion by transcriptional regulation of MTSS1. *Oncogene*. 2016;35:1602–8.
- Taylor MD, Bollt O, Iyer SC, Robertson GP. Metastasis suppressor 1 (MTSS1) expression is associated with reduced in-vivo metastasis and enhanced patient survival in lung adenocarcinoma. *Clin Exp Metastasis*. 2018;35:15–23.
- Lee H-S, Park J-H, Kim S-J, Kwon S-J, Kwon J. A cooperative activation loop among SWI/SNF,  $\gamma$ -H2AX and H3 acetylation for DNA double-strand break repair. *EMBO J*. 2010;29:1434–45.
- Niimi A, Chambers AL, Downs JA, Lehmann AR. A role for chromatin remodellers in replication of damaged DNA. *Nucleic Acids Res*. 2012;40:7393–403.
- Ray A, Mir SN, Wani G, Zhao Q, Battu A, Zhu Q, et al. Human SNF5/INI1, a component of the human SWI/SNF chromatin remodeling complex, promotes nucleotide excision repair by influencing ATM recruitment and downstream H2AX phosphorylation. *Mol Cell Biol*. 2009;29:6206–19.
- Rugo HS, Olopade OI, DeMichele A, Yau C, van't Veer LJ, Buxton MB, et al. Adaptive randomization of Veliparib–Carboplatin treatment in breast cancer. *N Engl J Med*. 2016;375:23–34.
- Pan D, Kobayashi A, Jiang P, Ferrari de Andrade L, Tay RE, Luoma A, et al. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. *Science*. 2018;359:770–5.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
- Martínez N, Almaraz C, Vaqué JP, Varela I, Derdak S, Beltrán S, et al. Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation. *Leukemia*. 2014;28:1334–40.
- Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25:2865–71.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
- Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382–3.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-seq data analysis. *Nucleic Acids Res*. 2016;44:W160–5.

39. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–89.
40. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the integrative genomics viewer. *Cancer Res*. 2017;77:e31–4.
41. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.
42. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11:R106.
43. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–9.
44. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102:15545–50.
45. Lauand C, Niero EL, Dias VM, Machado-Santelli GM. Cell cycle synchronization and BrdU incorporation as a tool to study the possible selective elimination of ErbB1 gene in the micronuclei in A549 cells. *Braz J Med Biol Res Rev Bras Pesqui Medicas E Biol*. 2015;48:382–91.

# PLC $\gamma$ 1/PKC $\theta$ Downstream Signaling Controls Cutaneous T-Cell Lymphoma Development and Progression

Nuria García-Díaz<sup>1</sup>, Berta Casar<sup>2</sup>, Ruth Alonso-Alonso<sup>3</sup>, Laura Quevedo<sup>2</sup>, Marta Rodríguez<sup>3</sup>, Fulgencio Ruso-Julve<sup>1</sup>, Anna Esteve-Codina<sup>4,5</sup>, Marta Gut<sup>4,5</sup>, Alejandro A. Gru<sup>6,7</sup>, María Carmen González-Vela<sup>8</sup>, Ivo Gut<sup>4,5</sup>, José Luis Rodríguez-Peralto<sup>9</sup>, Ignacio Varela<sup>2</sup>, Pablo Luis Ortiz-Romero<sup>10</sup>, Miguel A. Piris<sup>3</sup> and José Pedro Vaqué<sup>1</sup>

Developing mechanistic rationales can improve the clinical management of cutaneous T-cell lymphomas. There is considerable genetic and biological evidence of a malignant network of signaling mechanisms, highly influenced by deregulated TCR/PLC $\gamma$ 1 activity, controlling the biology of these lesions. In addition, activated signal transducer and activator of transcription 3 is associated with clinical progression, although the alterations responsible for this have not been fully elucidated. Here, we studied PLC $\gamma$ 1-dependent mechanisms that can mediate STAT3 activation and control tumor growth and progression. Downstream of PLC $\gamma$ 1, the pharmacological inhibition and genetic knockdown of protein kinase C theta (PKC $\theta$ ) inhibited signal transducer and activator of transcription 3 activation, impaired proliferation, and promoted apoptosis in cutaneous T-cell lymphoma cells. A PKC $\theta$ -dependent transcriptome in mycosis fungoides/Sézary syndrome cells revealed potential effector genes controlling cytokine signaling, TP53, and actin cytoskeleton dynamics. Consistently, an *in vivo* chicken embryo model xenografted with mycosis fungoides cells showed that PKC $\theta$  blockage abrogates tumor growth and spread to distant organs. Finally, the expression of a number of PKC $\theta$  target genes found in mycosis fungoides cells significantly correlated with that of *PRKCCQ* (PKC $\theta$ ) in 81 human mycosis fungoides samples. In summary, PKC $\theta$  can play a central role in the activation of malignant cutaneous T-cell lymphoma mechanisms via multiple routes, including, but not restricted to, STAT3. These mechanisms may, in turn, serve as targets for specific therapies.

*Journal of Investigative Dermatology* (2021) ■, ■-■; doi:10.1016/j.jid.2021.09.024

## INTRODUCTION

Cutaneous T-cell lymphomas (CTCLs) are a heterogeneous group of extranodal non-Hodgkin lymphomas characterized by the clonal expansion of malignant T cells in the skin

(Willemze et al., 2005). Mycosis fungoides (MF) and Sézary syndrome (SS) account for the vast majority of CTCLs. Clinically, classic MF presents with patches and plaques and may eventually attain a tumoral stage. Some patients progress, their disease involving peripheral blood, lymph nodes, and viscera. SS is an aggressive leukemic subtype characterized by erythroderma, lymphadenopathy, and the presence of clonal cerebriform T cells (Sézary cells) in the skin, lymph nodes, and peripheral blood. Whereas the prognosis of patients with MF depends on stage, particularly the type and extent of skin lesions, in patients with SS it is generally poor, with a median survival of 3 years (Agar et al., 2010). Several skin-directed and systemic therapies are being explored with the aim of improving the clinical management of these malignancies, especially at advanced stages (Oka and Miyagaki, 2019), but a deeper understanding of the main mechanisms controlling the biology of these lesions would boost our ability to diagnose and treat CTCL.

In the last few years, we have learned that multiple genetic alterations affecting TP53, TCR/PLC $\gamma$ 1 (PLCG1 hereafter), NOTCH, NF- $\kappa$ B, and JAK/signal transducer and activator of transcription (STAT) activities may drive the development and progression of CTCL (Choi et al., 2015; da Silva Almeida et al., 2015; Kiel et al., 2015; Mcgirt et al., 2015; Park et al., 2017; Pérez et al., 2015; Prasad et al., 2016; Ungewickell et al., 2015; Vaqué et al., 2014; Wang et al., 2015; Woollard et al., 2016). Although still not

<sup>1</sup>Molecular Biology Department, Universidad de Cantabria-Instituto de Investigación Marqués de Valdecilla, IDIVAL, Santander, Spain; <sup>2</sup>Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria-CSIC, Santander, Spain; <sup>3</sup>Pathology Department, Fundación Jiménez Díaz, CIBERONC, Madrid, Spain; <sup>4</sup>CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain; <sup>5</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain; <sup>6</sup>Department of Pathology, School of Medicine, University of Virginia, Charlottesville, Virginia, USA; <sup>7</sup>Department of Dermatology, School of Medicine, University of Virginia, Charlottesville, Virginia, USA; <sup>8</sup>Division of Pathology, Hospital Universitario Marqués de Valdecilla, Santander, Spain; <sup>9</sup>Department of Pathology, Hospital 12 de Octubre, institute i+12, CIBERONC, Medical School, University Complutense, Madrid, Spain; and <sup>10</sup>Department of Dermatology, Hospital 12 de Octubre, institute i+12, CIBERONC, Medical School, University Complutense, Madrid, Spain

Correspondence: José Pedro Vaqué, Avda. Marqués de Valdecilla s/n, Facultad de medicina-Universidad de Cantabria and IDIVAL-Institute, 39011, Santander, Spain. E-mail: [vaquej@unican.es](mailto:vaquej@unican.es)

Abbreviations: CAM, chorioallantoic membrane; CaN, calcineurin; CTCL, cutaneous T-cell lymphoma; MF, mycosis fungoides; NTC, nontargeting control; PKC, protein kinase C; shPRKCCQ, short hairpin PRKCCQ; SS, Sézary syndrome; STAT, signal transducer and activator of transcription; TPA, 12-O-tetradecanoylphorbol-13-acetate

Received 18 March 2021; revised 7 September 2021; accepted 8 September 2021; accepted manuscript published online XXX; corrected proof published online XXX



fully understood, malignant TCR/PLCG1 activity constitutes an important CTCL mechanism. PLCG1 is a phospholipase that, on T-cell activation, cleaves phosphatidylinositol 4,5-bisphosphate in the plasma membrane into inositol 1,4,5-trisphosphate and diacylglycerol. Whereas inositol 1,4,5-trisphosphate triggers calcium release from the endoplasmic reticulum, leading to NFAT activation via calcineurin (CaN), diacylglycerol can activate specific protein kinase C (PKC) isoforms (Isakov and Altman, 2002). Constitutively activated PLCG1 SNVs (S345F and others), present in 10–20% of cases, promote the activation of important T-cell nuclear effectors, such as NFAT (Park et al., 2017; Patel et al., 2020; Vaqué et al., 2014) and NF- $\kappa$ B (Patel et al., 2020). In addition, deregulated expression and amplification of the *PRKCQ* gene (protein PKC $\theta$ ) have been described in 30% of cases (Choi et al., 2015; Wang et al., 2015; Woollard et al., 2016). PKC $\theta$  is a serine/threonine kinase widely expressed in T lymphocytes (Meller et al., 1998) and a well-known PLCG1 effector (Steinberg, 2008). In contrast, members of the JAK/STAT signaling pathway frequently harbor genetic alterations, such as activating mutations (4%) and amplifications (up to 60%) in CTCL cases (Park et al., 2017). JAKs are tyrosine kinases that phosphorylate and activate STAT transcriptional activity in response to extracellular stimuli (Vainchenko and Constantinescu, 2013). Serine phosphorylation can modulate STAT activities in response to alternative stimuli (Decker and Kovarik, 2000). Whereas PLCG1 mutations and nuclear NFAT and NF- $\kappa$ B accumulation have been found in MF cases at all stages, activated STAT3 has recently been associated with MF in advanced stages (Pérez et al., 2020) and has been detected in SS cells (Erikson et al., 2001).

Thus, acting downstream of PLCG1, we investigated the role of PKC $\theta$  mediating STAT3 activation and studied the mechanisms associated to its potential role at controlling the development and progression of MF lesions.

## RESULTS

### PLC $\gamma$ 1/PKC $\theta$ downstream signaling promotes STAT3 activation and proliferation/survival of CTCL cells

Recent evidence suggests that TCR downstream signaling plays an important role in the biology of CTCL via PLCG1. Because activated STAT3 has been associated with advanced CTCL stages, we decided to study the potential mechanisms underlying PLCG1 and its downstream effector PKC $\theta$  in the activation of STAT3. To this end, we set up a combination of wild type and constitutively activated PLCG1 (PLCG1 S345F) and PKC $\theta$  (PRKCQ A148E; details in Supplementary Materials and Methods and Supplementary Table S1) mutants, transiently expressed alongside specific luciferase-based reporter genes for STAT3 and NFAT in HEK293 cells. We combined this approach with the use of specific inhibitors employed in clinic practice: tacrolimus (a CaN inhibitor), sotrastaurin (a pan-PKC inhibitor mostly for PKC $\theta$ ), and ruxolitinib (a JAK inhibitor), as depicted in Supplementary Figure S1a.

In these settings, PLCG1 and PRKCQ mutants triggered STAT3 transcriptional activation, which was dependent on PKC $\theta$  and JAK activities (Figure 1a). Their combined

inhibition did not induce a significant STAT3 blockage as compared with each inhibitor used alone. Activated PLCG1 and PRKCQ promoted STAT3 phosphorylation in residues Y705 (a JAK-dependent phosphorylation site) and S727 (JAK independent), which were impaired by sotrastaurin (Figure 1b and Supplementary Figure S1b and c). In contrast, as controls, PLCG1 and PRKCQ mutants elicited NFAT transcriptional activation that was impaired by tacrolimus and sotrastaurin (Supplementary Figure S1d).

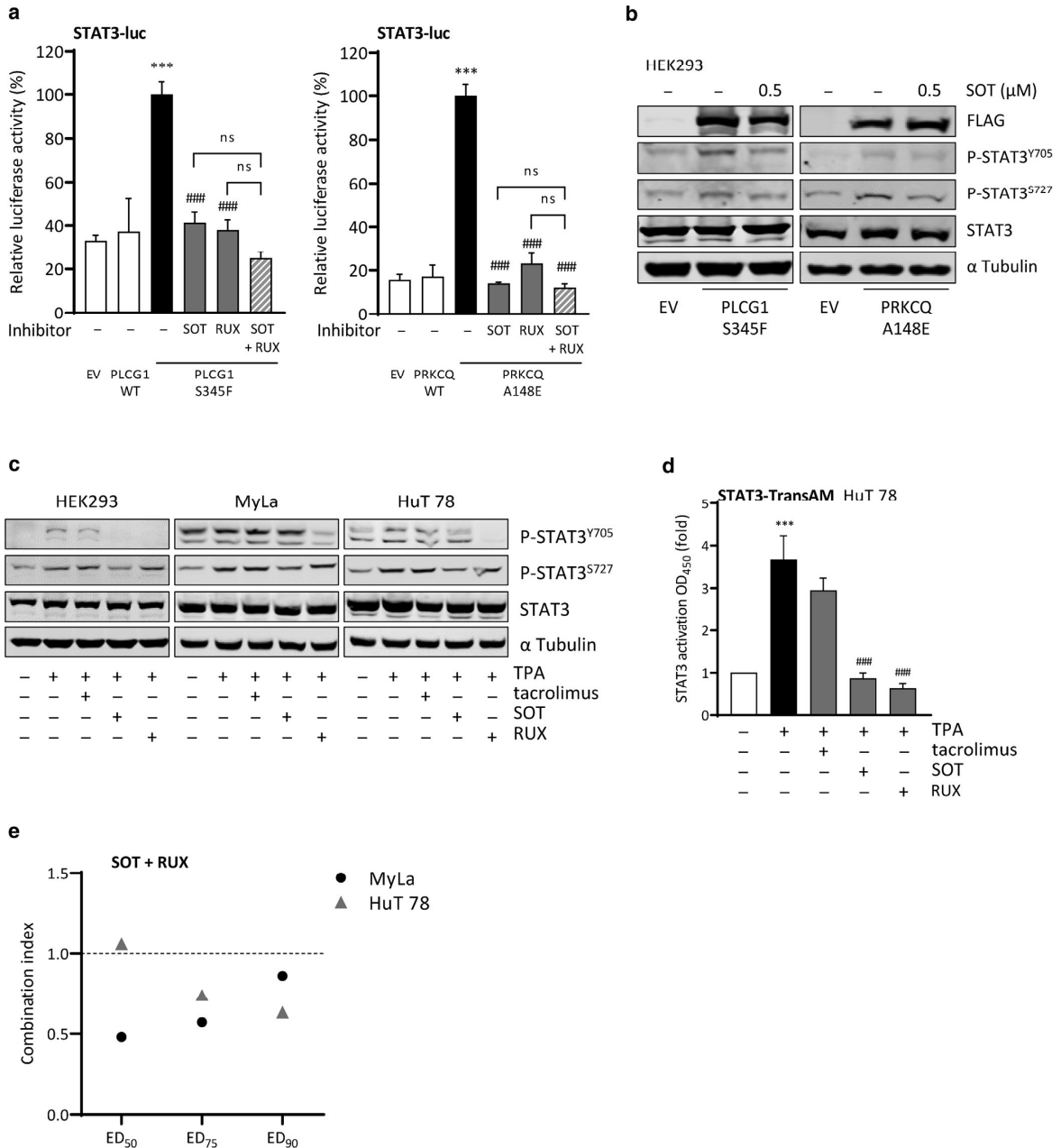
To further explore the mechanisms that, downstream of PKC $\theta$ , can mediate STAT3 activation and malignant proliferation, we took advantage of HEK293, Jurkat, and CTCL (MyLa [MF-derived] and HuT 78 [SS-derived]) cells, which express PKC $\theta$  mRNA and protein at detectable levels (Supplementary Figure S2a). Generally, 12-*O*-tetradecanoylphorbol-13-acetate (TPA) promoted phosphorylation of STAT3 in tyrosine (Y705) and serine (S727) residues. In CTCL cells, whereas MyLa cells displayed a constitutive phosphorylation in Y705, TPA increased S727 both in MyLa and HuT 78 and Y705 in HuT 78 cells. (Figure 1c and Supplementary Figure S2b). Whereas sotrastaurin (PKC $\theta$  inhibitor) impaired phosphorylation in tyrosine and serine residues, ruxolitinib (JAK inhibitor) only abrogated tyrosine phosphorylation. Finally, both inhibitors abrogated the binding of activated STAT3 proteins to a specific STAT3 DNA-binding sequence in HuT 78 cells (Figure 1d).

Biologically, pharmacological inhibition of PKC $\theta$  provoked a concentration-dependent reduction of CTCL cell proliferation and triggered apoptosis (Supplementary Figure S2c and d). Notably, and consistent with our previous results, the combined inhibition of PKC $\theta$  and JAK had a synergistic antiproliferative effect in CTCL cells (Figure 1e).

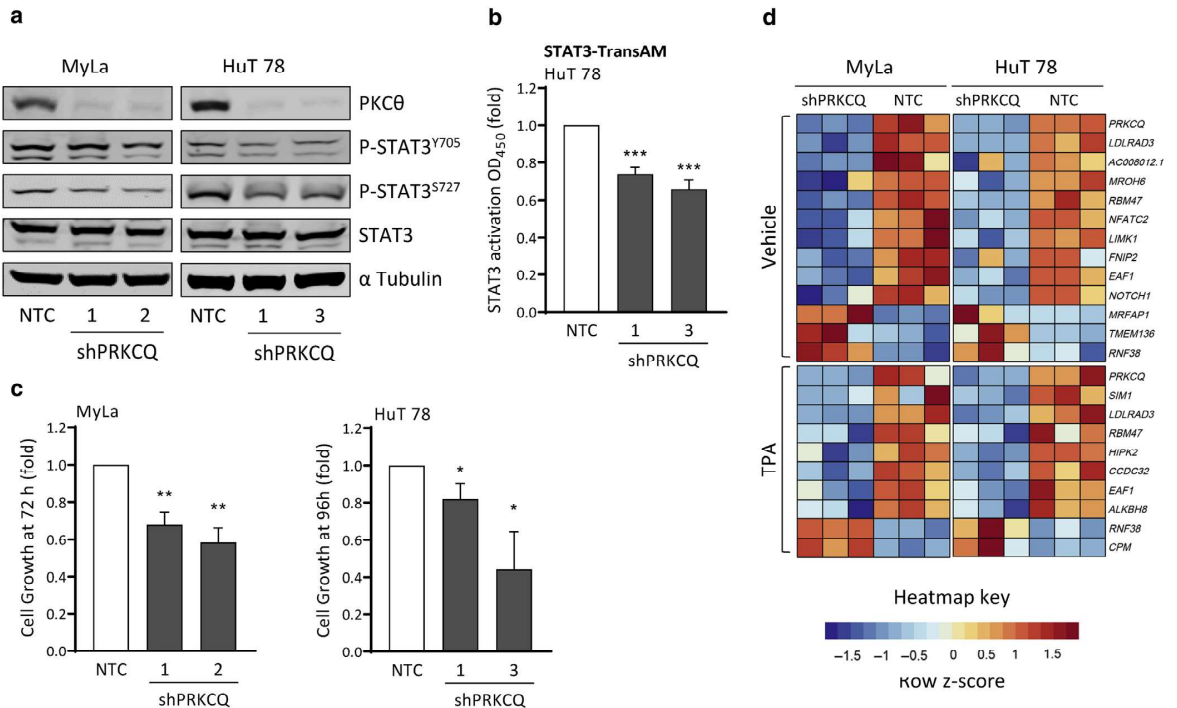
### PKC $\theta$ -dependent transcriptome in MF/SS cells uncovers CTCL disease mechanisms

To further study the role of PKC $\theta$  in CTCL cells, we generated stable MyLa, HuT 78, and HEK293 cells with doxycycline-inducible expression of *PRKCQ* short hairpin RNA (shPRKCQ) and a nontargeting control (NTC). On incubation with doxycycline, PKC $\theta$  protein levels were effectively knocked down in all cells (Figure 2a and Supplementary Figure S2e). PKC $\theta$  knocked down cells showed reduced phosphorylation of STAT3 in Y705 and S727 residues, as well as decreased STAT3 transcriptional activity in HEK293 cells stably transfected with a STAT3 reporter (STAT3-SEAP) (Figure 2a and b and Supplementary Figure S2e and f). PKC $\theta$  deficiency also impaired MyLa and HuT 78 cell proliferation (Figure 2c).

We next searched for specific PKC $\theta$  target genes and pathways that, alongside STAT3 activation, could enlighten its role in CTCL. To this end, we performed mRNA sequencing in vehicle or TPA-stimulated NTC and shPRKCQ MyLa and HuT 78 cells. In our analysis, we identified 23 significant PKC $\theta$ -regulated genes shared between both cell lines: 13 in vehicle (DMSO) and 10 in TPA-treated cells (Figure 2d). Under both conditions, PKC $\theta$  positively controlled the expression of *LDLRAD3* (lipoprotein receptor), *RBM47* (RNA binding/TP53-related), and *EAF1* (transcriptional coactivator) and negatively regulated *RNF38* (ubiquitin ligase/TP53-related). In addition, PKC $\theta$  positively controlled



**Figure 1. PLCG1/PKC $\theta$  downstream signaling triggers STAT3 activation.** (a) STAT3 luciferase reporter activity in HEK293 cells transfected with the indicated vectors and treated with PKC $\theta$  and JAK inhibitors (SOT and RUX, respectively, 1  $\mu$ M, 24 hours) or the combination of both (n = 3). (b) Western blot of HEK293 cells transfected with the indicated vectors, starved, and treated with SOT (0.5  $\mu$ M, 3 hours) and incubated with the indicated antibodies. (c) Western blot of starved HEK293, MyLa, and HuT 78 cells treated with the indicated inhibitors (1  $\mu$ M, 3 hours), stimulated with TPA (10 ng/ml, 1 hour), and incubated with the indicated antibodies. (d) ELISA-based assay showing activated STAT3 bound to specific DNA-binding sequence in HuT 78 cells treated with the indicated inhibitors (1  $\mu$ M, 3 hours) and TPA (10 ng/ml, overnight) (n = 3). (e) CIs at 50%, 75%, and 90% ED in MyLa and HuT 78 cells treated with a combination of SOT and RUX (24 hours). CIs were calculated using the Chou-Talalay method. CI < 1 indicates synergism, CI = 1 is additive, and CI > 1 indicates antagonism. Images are representative of each western blot (n = 3). Data are mean  $\pm$  SEM. Student's *t*-test: \*\*\**P* < 0.001 versus PLCG1/PRKCQ WT (a) or control vehicle (d), ##*P* < 0.01 and ###*P* < 0.001 versus PLCG1 S345F or PRKCQ A148E treated with control vehicle (a) or TPA (d). CI, combination index; ED, effective dose; EV, empty vector; ns, not significant; PKC, protein kinase C; P-STAT, phosphorylated signal transducer and activator of transcription; RUX, ruxolitinib; SOT, sotrastaurin; STAT, signal transducer and activator of transcription; TPA, 12-*O*-tetradecanoylphorbol-13-acetate; WT, wild type.



**Figure 2. PKC $\theta$ -dependent transcriptome in MF/SS cells.** (a) Western blot of cells with inducible expression of NTC or shPRKCQ and incubated with the indicated antibodies. Lanes numbered 1, 2, and 3 refer to different shPRKCQ sequences. Images are representative of each western blot ( $n = 3$ ). (b) ELISA-based assay showing activated STAT3 bound to specific DNA-binding sequence in NTC and shPRKCQ HuT 78 cells ( $n = 4$ ). (c) Proliferation of NTC and shPRKCQ MyLa and HuT 78 cells after *PRKCQ* knockdown induction ( $n = 4$ ). (d) Heatmap representation of a selection of differentially expressed genes in shPRKCQ MyLa and HuT 78 cells compared with NTC cells treated with control vehicle or TPA (10 ng/ml,  $n = 3$ ). Expression differences range from blue (downregulation) to red (upregulation) according to the z-score. Data are mean  $\pm$  SEM. Student's *t*-test: \* $P < 0.05$ , \*\* $P < 0.01$ , and \*\*\* $P < 0.001$  versus NTC. h, hour; MF, mycosis fungoides; NTC, nontargeting control; P-STAT, phosphorylated signal transducer and activator of transcription; PKC, protein kinase C; shPRKCQ, short hairpin *PRKCQ*; SS, Sézary syndrome; STAT, signal transducer and activator of transcription; TPA, 12-*O*-tetradecanoylphorbol-13-acetate.

the expression of *NFATC2* (TCR/PLCG1 effector), *LIMK1* (actin cytoskeleton), and *NOTCH1* under basal conditions and *HIPK2* (kinase/TP53 related) and *ALKBH8* (methyltransferase) in response to TPA. An independent validation of the data is included in [Supplementary Figure S3a](#).

An alternative gene set enrichment analysis was performed to interpret the gene expression data. We identified a number of significantly deregulated gene sets based on Kyoto Encyclopedia of Genes and Genomes pathways ( $P < 0.05$  and false discovery rate  $< 0.25$ ) in shPRKCQ versus NTC cells ([Table 1](#)). Among these, we found potential PKC $\theta$  effectors that can participate in cytokine/cytokine receptor interaction (*CCL22*), hematopoietic cell lineage (*IL6R*), DNA replication (*PCNA*), cell cycle (*PLK1*), adipocytokine signaling (*JAK2*, *STAT3*, and *TNF*), base excision repair (*POLD1*), and TP53 signaling (*CHEK1*) ([Supplementary Table S2](#) and [Supplementary Figure S3b](#)).

#### Blockage of PKC $\theta$ impairs CTCL tumorigenesis and dissemination in vivo

To study the malignant activities carried out by PKC $\theta$  in vivo, we used a chicken embryo model, which offers the possibility of generating primary tumors and studying its spreading

potential in a timely and cost-effective manner. Thus, we generated xenografted tumors derived from MyLa cells implanted on top of the chicken chorioallantoic membrane (CAM). These tumors displayed positive CD30 and phosphorylated STAT3 immunohistochemical staining ([Figure 3a](#) and [Supplementary Figure S4](#)). We blocked PKC $\theta$  using two distinct approaches: pharmacological (i.e., sotrastaurin) and genetic (i.e., short hairpin RNA). To study the effects of sotrastaurin, we seeded  $1 \times 10^6$  MyLa cells into the CAM of day 10 chicken embryos and allowed them to grow for 7 days. In this setting, sotrastaurin was used once (2 days before harvesting) or twice (every 2 days after cell implantation). In parallel,  $1 \times 10^6$  NTC or shPRKCQ MyLa cells, previously incubated with doxycycline, were also seeded into the CAMs of the chicken embryos and allowed to grow under the same conditions. As shown in [Figure 3b](#), PKC $\theta$  blockage greatly impaired primary tumor growth. Moreover, because this system enables the detection of disseminating cells (human cells in this case) in chicken embryo tissue, we detected that control MyLa cells intravasated the distal CAM and disseminated to internal organs (liver and lungs, [Figure 3c](#) and [d](#)). Strikingly, PKC $\theta$  blockage provoked a substantial reduction in the spreading of MyLa cells that occurred alongside a

**Table 1. GSEA of PKC $\theta$ -Dependent Transcriptome**

KEGG PATHWAY	NES	NOM P-Val	FDR q-Val	Representative Enriched Genes
<b>Vehicle</b>				
ANTIGEN PROCESSING AND PRESENTATION	1.54	0.026	0.249	<i>HSPA1A, KIR3DL1, LGMN, KIR2DL3</i>
CYTOKINE-CYTOKINE RECEPTOR INTERACTION	1.73	0.000	0.097	<i>CCL22, IL6R, CCL3, TNFSF13B, IL11, TNFRSF9, IL13, CD40LG, PDGFA, TNFRSF11A, TNFSF9</i>
HEMATOPOIETIC CELL LINEAGE	1.73	0.011	0.116	<i>IL6R, ITGA1, CD37, ITGA6, IL11, ITGB3, CD3G</i>
RIBOSOME	1.87	0.000	0.114	<i>RPS3, RPS27A, RPS15</i>
<b>TPA</b>				
DNA REPLICATION	-2.27	0.000	0.000	<i>RFC1, POLE2, LIG1, PCNA, POLD3, MCM7</i>
CELL CYCLE	-2.19	0.000	0.000	<i>PLK1, PRKDC, CDC25C, SKP2, CDK2, E2F1</i>
MISMATCH REPAIR	-1.68	0.013	0.105	<i>RFC4, MSH2, LIG1, POLD3, PCNA, EXO1</i>
ADIPOCYTOKINE SIGNALING PATHWAY	-1.56	0.032	0.179	<i>TRAF2, JAK2, NFKBIB, CAMKK1, AKT3, STAT3, TNF, SOCS3, PRKCC</i>
BASE EXCISION REPAIR	-1.53	0.039	0.177	<i>PARP1, POLD1, POLE, PCNA</i>
P53 SIGNALING PATHWAY	-1.46	0.038	0.245	<i>CHEK1, CDK1, BID, CCNE2, CDK6</i>
HEMATOPOIETIC CELL LINEAGE	1.63	0.013	0.249	<i>CD38, CD37, FLT3LG, IL11, ITGAM, CD3G, IL11RA, IL9R</i>
RIBOSOME	2.25	0.000	0.000	<i>RPS3, RPS27A, RPS15</i>

Abbreviations: FDR q-Val, false discovery rate q-value; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; NES, normalized enrichment score in shPRKCQ versus NTC cells; NOM P-Val, nominal P-value; PKC, protein kinase C; TPA, 12-O-tetradecanoylphorbol-13-acetate.

reduction in CD30 and phosphorylated STAT3 staining (Figure 3a and Supplementary Figure S4).

#### A PKC $\theta$ expression signature in human MF

Given the role of PKC $\theta$  in MF tumor formation and dissemination, we next questioned whether a PKC $\theta$ -derived expression signature could be detected in a cohort of MF samples. First, 16 PKC $\theta$  target genes, previously identified by mRNA sequencing in MyLa cells, were selected based on significance and their associated biological activities (Figure 4a). This selection included genes involved in the control of the cytoskeleton (*CIT*, *LCP1*, and *LSP1*), signaling oncogenes (*FGFR1*, *FGFR3*, *MAPK13*, or *PRKACB*), cell adhesion (*VCAM*), or angiogenesis (*VEGFB*). Then, using NanoString, we comparatively studied their expression profiles in a cohort of 81 individual samples (including plaques and tumors) from 27 patients with MF and six inflammatory dermatoses used as control (Supplementary Table S3). Analysis of the results showed that *PRKCQ* and its target genes, with the exception of *RHOB*, were differentially and significantly expressed in MF as compared with inflammatory dermatoses (Supplementary Figure S5).

The heatmap showing the general expression pattern of PKC $\theta$  and its target genes in MF cases is shown in Figure 4b. The expression of eight genes significantly correlated with that of *PRKCQ* across the cohort. Whereas *PRKACB*, *LCP1*, *CCDC32*, *LSP1*, and *TNFRSF25* showed a positive correlation, others such as *MAPK13*, *FGFR3*, and *RHOB* displayed a negative correlation (Figure 4c).

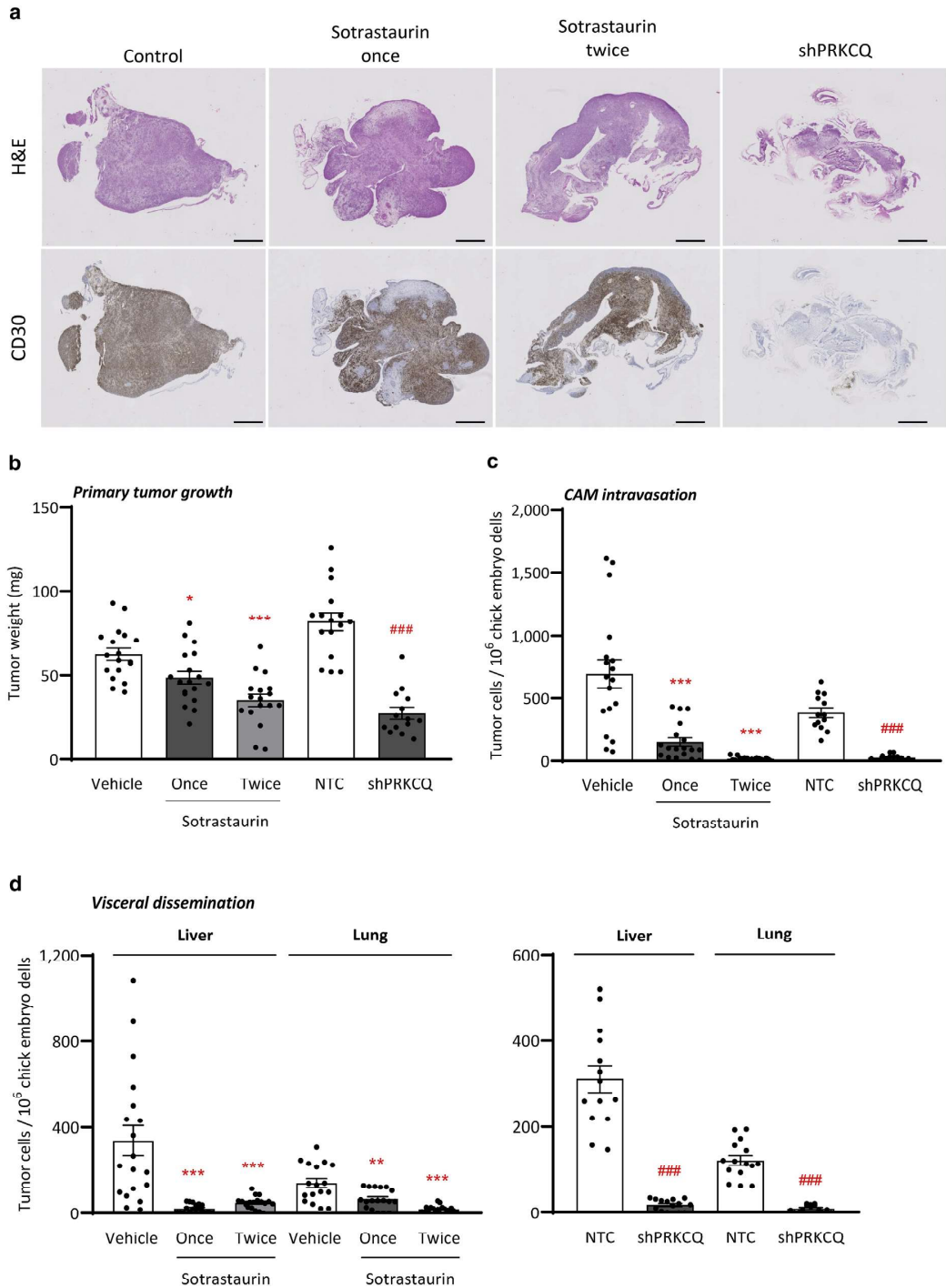
#### DISCUSSION

Clinical management of CTCL requires efficient therapies, especially in advanced stages. Deregulated TCR/PLCG1 downstream signaling constitutes a major alteration, as assessed by the presence of constitutively activated PLCG1 mutants with nuclear NFAT activation (10% of the cases)

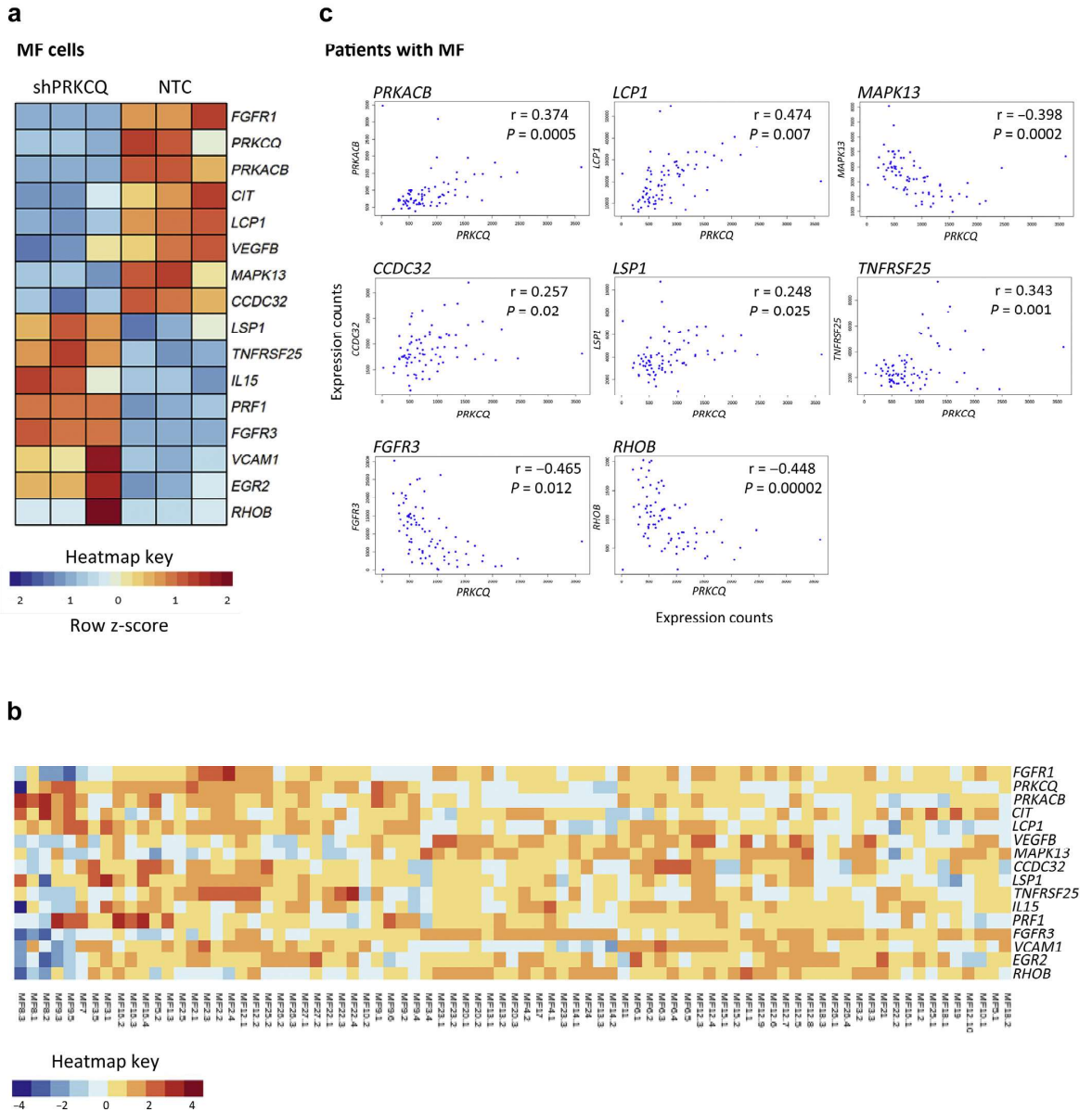
(Choi et al., 2015; Mcgirt et al., 2015; Ungewickell et al., 2015; Vaqué et al., 2014; Woollard et al., 2016) and amplifications in the *PRKCQ* gene (PKC $\theta$ ), in 20–30% of the cases (Choi et al., 2015; Woollard et al., 2016) nonoverlapping with PLCG1 SNVs. Because PLCG1 is still a non-druggable target in the clinic, specific targeting of CaN has the potential to block NFAT activation downstream of TCR/PLCG1 (Vaqué et al., 2014). Based on our previous findings, our team is about to report promising results of a clinical trial using topical pimecrolimus (a CaN inhibitor) in patients with early-stage MF (PimTo-MF study, EudraCT number: 2014-001377-14; Ortiz-Romero et al., unpublished data). Lessons learned from this experience and the various lines of genetic evidence reported in the literature indicate that downstream of TCR/PLCG1, alternative pathways can participate in parallel with CaN/NFAT, acting as mechanisms of resistance to therapy and/or disease progression. In our study, PLCG1-S345F activated NFAT by a mechanism involving CaN and PKC $\theta$ . Additionally, a mutant PKC $\theta$  with constitutive activated kinase activity (A148E) activated NFAT, which was abrogated by sotrastaurin. These preclinical data suggest that combinations of CaN and PKC $\theta$  inhibitors may increase therapy efficacy during early stages and/or provide a rationale for treating nonresponders or even patients with advanced CTCL.

Despite the fact that genetic alterations in the JAK/STAT pathway are frequent in CTCL, the percentage of cases with mutant JAK and STAT proteins (about 8%) cannot fully explain the high proportion of lesions with activated/phosphorylated STAT expression in the nucleus of malignant T cells (Fantin et al., 2008; Pérez et al., 2020; Sommer et al., 2004). Therefore, alternative mechanisms are under discussion and include proinflammatory microenvironments (Kim et al., 2005) or immune responses to bacterial colonization in the compromised skin barriers of CTCL lesions (Fanok et al., 2018; Willerslev-Olsen et al., 2016). It is also





**Figure 3. Blockage of PKC $\theta$  impairs CTCL tumorigenesis and dissemination in vivo.** (a) H&E and CD30 staining of paraffin sections of control, tumors treated once or twice with sotrastaurin, and tumors with PKC $\theta$  deficiency (shPRKCQ). (b–d) Pharmacological and genetic PKC $\theta$  inhibition effects on (b) tumor weight, (c) CAM intravasation, and (d) visceral dissemination to liver and lungs of chicken embryos engrafted with MyLa cells treated once or twice with sotrastaurin (10  $\mu$ M, n = 17–18) or engrafted with NTC or shPRKCQ MyLa cells (n = 14–16). Data are mean  $\pm$  SEM. Student's *t*-test: \**P* < 0.05, \*\**P* < 0.01, and \*\*\**P* < 0.001 versus control vehicle; ###*P* < 0.001 versus NTC. Bar = 1 mm. CAM, chorioallantoic membrane; CTCL, cutaneous T-cell lymphoma; NTC, nontargeting control; PKC, protein kinase C; shPRKCQ, short hairpin *PRKCQ*.



**Figure 4.** Correlation between the expression of PKC $\theta$  and its target genes in MF cells and human samples. (a) Heatmap representation of a selection of differentially expressed genes in shPRKCQ MyLa cells compared with NTC and treated with TPA (10 ng/ml, 24 hours,  $n = 3$ ). Expression differences range from blue (downregulation) to red (upregulation) according to the z-score. (b) Heatmap representation of the expression of the indicated genes in MF samples. (c) Correlation of gene expressions between PRKCQ and PRKACB, LCP1, MAPK13, CCDC32, LSP1, TNFRSF25, FGFR3, and RHOB in MF samples. MF, mycosis fungoides; NTC, nontargeting control; PKC, protein kinase C; shPRKCQ, short hairpin PRKCQ; TPA, 12-*O*-tetradecanoylphorbol-13-acetate.

possible that, downstream of PLCG1, deregulated intracellular mechanisms may also trigger STAT activation. In support of this hypothesis, a combination of targeted sequencing and immunohistochemical staining in a cohort of patients with early and advanced MF detected two PLCG1-S345F-positive patients displaying nuclear phosphorylated STAT3 staining in the absence of mutated JAKs (Pérez et al., 2020). In our work, PLCG1 and PKC $\theta$  mutants promoted the phosphorylation and transcriptional activity of STAT3 by JAK-dependent and JAK-

independent mechanisms. The combination of PKC $\theta$  and JAK inhibitors was not significantly superior to each inhibitor used alone at blocking STAT3, suggesting that these can participate as part of the same signaling axis. Moreover, genetic PKC $\theta$  knockdown also abrogated STAT3 phosphorylation, transcriptional activity, and proliferation in CTCL cells. Following this line of evidence, TPA (a PKC activator) also promoted STAT3 phosphorylation in tyrosine and serine residues alongside STAT3 transcriptional activity in CTCL cells.

Finally, this conjoined PKC $\theta$  and JAK mechanism also proved to be biologically relevant, because their combined inhibition provoked the synergistic abrogation of cell proliferation. Thus, malignant STAT3 activation in advanced CTCL can be triggered by deregulated extracellular stimuli (usually JAK-dependent), mutations in JAKs, and/or alternative mechanisms, such as deregulated PLCG1/PKC $\theta$  downstream signaling.

This PLCG1/PKC $\theta$ –STAT3 connection may have important implications for setting up approaches for treating CTCL, with a special focus on advanced cases. On the one hand, specific JAK inhibitors are being used in the clinical milieu (de Freitas and da Costa Maranduba, 2015; Lee et al., 2014; Malemud, 2018). Therefore, it is conceivable that they may be effective for treating patients with CTCL with deregulated JAK/STAT activity. In this regard, active clinical trials test JAK inhibitors in relapsed T- or NK-cell lymphomas (NCT02974647) and peripheral T-cell lymphoma (NCT04105010), offering the possibility of enrolling patients with CTCL. In contrast, it would be highly beneficial to design future therapies bearing in mind that a variety of mechanisms, such as CaN, PKC $\theta$ , or JAKs, may synergistically drive CTCL progression.

Despite the fact that individual PKC $\theta$  or JAK inhibitors provoked a total blockage of STAT3 activation, their combination still elicited a synergistic antiproliferative effect in CTCL cells. To gain further evidence about the malignant CTCL mechanisms controlled by PKC $\theta$ , other than STAT3, we searched for specific target genes and pathways shared between MF and SS cells. This analysis revealed a number of candidate PKC $\theta$  target genes with potentially relevant malignant activities in CTCL, such as the following: (i) Regulation of *TP53*, the most frequently altered gene in these entities (Chang et al., 2018; da Silva Almeida et al., 2015); *RBM47*, a positive regulator of the TP53/p21 axis that is also associated with lung carcinogenesis (Radine et al., 2020; Sakurai et al., 2016); *HIPK2*, a proapoptotic kinase that regulates and phosphorylates TP53 (Cecchinelli et al., 2006); and *RNF38*, coding a TP53 ubiquitin ligase (Sheren and Kassenbrock, 2013); (ii) Rho/actin cytoskeleton remodeling: *LIMK1*, which controls actin cytoskeleton dynamics and cell shape/movement via cofilin (Hamill et al., 2016); (iii) TCR-PLCG1 downstream signaling: *NFATC2* (NFAT1), acting downstream of CaN, can control cytokine expression such as IL-2 or IL-4 in T cells (Mognol et al., 2016); and (iv) *NOTCH1*, a CTCL oncogene related to cell survival (Kamstrup et al., 2010). Our results argue in favor of PKC $\theta$  acting as part of an intricate TCR/PLCG1 network of signaling mechanisms to play a key mechanistic and biological role, via multiple oncogenic effectors, to control CTCL cell proliferation, survival, and dissemination. Functionally, we carried out an alternative gene set enrichment analysis. This analysis highlighted important biological activities controlled by PKC $\theta$ , such as cytokine/cytokine receptor interaction, hematopoietic cell lineage, cell cycle, and TP53 signaling.

To investigate the biological role of PKC $\theta$  in vivo, we used a chicken embryo xenograft model (Crespo and Casar, 2016; Klingenberg et al., 2014). MF cells deficient in PKC $\theta$  expression had greatly impaired the ability to grow tumors, promote angiogenesis, intravasate blood vessels, and spread to distant organs. Moreover, sotrastaurin provoked similar

effects. These results are consistent with the transcriptome data and, although still in a CTCL preclinical setting, can support a mechanistic role for PKC $\theta$  in controlling multiple malignant cellular activities, such as those explained earlier, promoting tumor development and progression in patients. In this regard, we studied the expression of a specific subset of 16 PKC $\theta$  target genes, chosen among those found in MyLa cells (MF-derived) in a cohort of 81 human MF samples and six inflammatory dermatoses. First, *PRKCQ* and its target genes, except *RHOB*, were differentially expressed in MF cases versus controls, which highlights the potential oncogenic role of PKC $\theta$  and its downstream effectors in CTCL. Regarding MF cases, a heterogeneous gene expression pattern was found between patients and between samples, which aligns with the inter- and intratumoral heterogeneity recently described in CTCL cases, using transcriptomic analyses (Gaydosik et al., 2019; Iyer et al., 2019). Nevertheless, we specifically detected eight genes whose expression was significantly correlated with that of *PRKCQ* in these samples. More in detail, whereas *PRKACB* (protein kinase A, catalytic subunit beta), *TNFRSF25* (a TNF receptor), and the actin binding *LSP1* and *LCPI* correlated positively with *PRKCQ*, a negative correlation was found with *RHOB*, *FGFR3*, and *MAPK13* (p38 MAPK delta). Although further studies will help explain the functional roles of these genes in the biology of MF, in a PKC $\theta$  context, it is plausible that protein kinase A/CREB activity and dynamic regulation of the actin cytoskeleton might play essential roles in the ability of PKC $\theta$  to promote CTCL tumor formation and dissemination.

In summary, this work provides strong mechanistic and biological evidence of the role of a malignant TCR/PLCG1/PKC $\theta$  signaling network controlling the biology of CTCL. From a translational perspective, this study identifies different mechanisms for activating STAT3 downstream of TCR/PLCG1; proposes rational approaches to developing targeted therapies, including CaN, PKC $\theta$ , and JAK inhibitors used alone or in combination; and reveals a number of PKC $\theta$  target genes and pathways to be explored in a translational setting, because they may play essential roles in CTCL.

## MATERIALS AND METHODS

### Patient samples

A total of 87 formalin-fixed, paraffin-embedded tissue samples belonging to 27 patients with MF and six inflammatory dermatoses were used (Supplementary Table S3). The samples were collected in The Hospital 12 de Octubre, Madrid, Spain, in collaboration with the Fundación Jiménez Díaz Hospital, in Madrid, and the study was conducted in accordance with the Declaration of Helsinki. All patients gave written informed consent to be included in this study. All the processes were approved and conducted in adherence with the specific recommendations of the Comité Ético de Investigación clínica del Hospital 12 de Octubre.

### Chicken embryo model for spontaneous tissue colonization

Fertilized hen eggs were obtained from Granja Gibert (Spain) and incubated at 60% humidity and 38 °C in a rotating incubator. Spontaneous metastasis and tissue colonization were performed as described elsewhere (Crespo and Casar, 2016). Briefly, on day 10 of chick development, eggs were windowed and  $1 \times 10^6$  MyLa cells, resuspended in 20% Matrigel (Corning, VA) and 80% serum-free

media, and placed onto the CAM. Cells were allowed to expand for 7 days in a stationary incubator at 60% humidity and 38 °C, and two experiments were performed: (i) doxycycline-induced NTC or shPRKCQ MyLa cells and (ii) MyLa cells topically treated on the upper CAM with vehicle or sotrastaurin (10  $\mu$ M) in serum-free media 2 days before harvesting (once) or every 2 days (twice). Primary tumors were excised, weighed, fixed in 4% buffered formaldehyde, washed with PBS, and embedded in paraffin. Human cells within chick embryo tissues were detected by quantitative *Alu* PCR.

The CAM does not require obtaining ethics committee approval for animal experimentation. Chick embryos are not considered as living animals until day 17 of development in most countries. The CAM is not innervated, and experiments are terminated before the development of centers in the brain associated with pain perception, making this a system not requiring animal experimentation permissions. Thus, according to European law (Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes), the CAM model system does not raise any ethical or legal concerns, thus being an attractive alternative to other animal experiments.

### NanoString gene expression assay

Total RNA isolation was performed on formalin-fixed, paraffin-embedded skin samples using an RNeasy FFPE Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. We used an nCounter Custom CodeSet Design in conjunction with the nCounter Flex Analysis System (NanoString Technologies, Seattle, WA). Gene expression values were normalized with respect to eight house-keeping genes. The data were analyzed by nSolverTM Analysis Software 4.0 (NanoString Technologies) using the nCounter Advanced Analysis (version 2.0.134). Raw data were normalized using internal negative and positive controls. Hierarchical cluster analysis was performed using R 3.6.1 (package heatmap, <https://www.R-project.org>, R Foundation for Statistical Computing, Vienna, Austria).

### Statistical analysis

Unless otherwise specified, all experiments were performed in independent triplicates and numerical data were summarized as mean  $\pm$  SEM using GraphPad Prism6 software (San Diego, CA). Global means of pairs of groups were compared using two-tailed unpaired Student's *t*-tests, with statistical significance concluded for values of  $P < 0.05$ : \* $\# < 0.05$ , \*\* $\#\# < 0.01$  and \*\*\* $\#\#\# < 0.001$ . Correlation of gene expression between *PRKCQ* and the other genes was compared with Pearson's *r* correlation test.

### Data availability statement

For detailed methods and original protocols, please contact [vaquej@unican.es](mailto:vaquej@unican.es). Raw RNA-sequencing files are available at the Gene Expression Omnibus repository under the accession number GSE157442 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>).

### ORCID

Nuria García-Díaz: <http://orcid.org/0000-0002-0511-5590>  
 Berta Casar: <http://orcid.org/0000-0002-3058-5631>  
 Ruth Alonso-Alonso: <http://orcid.org/0000-0001-9503-7268>  
 Laura Quevedo: <http://orcid.org/0000-0001-8902-5818>  
 Marta Rodríguez: <http://orcid.org/0000-0001-8543-6767>  
 Fulgencio Ruso-Julve: <http://orcid.org/0000-0001-6500-6807>  
 Anna Esteve-Codina: <http://orcid.org/0000-0003-0361-2873>  
 Marta Gut: <http://orcid.org/0000-0002-4063-7159>  
 Alejandro A. Gru: <http://orcid.org/0000-0002-2573-8074>  
 María Carmen González-Vela: <http://orcid.org/0000-0002-1695-9203>  
 Ivo Gut: <http://orcid.org/0000-0001-7219-632X>  
 José Luis Rodríguez-Peralto: <http://orcid.org/0000-0002-6578-7153>  
 Ignacio Varela: <http://orcid.org/0000-0002-0969-506X>

Pablo Luis Ortiz-Romero: <http://orcid.org/0000-0003-2985-9639>

Miguel A. Piris: <http://orcid.org/0000-0001-5839-3634>

José Pedro Vaqué: <http://orcid.org/0000-0002-3913-2495>

### CONFLICT OF INTEREST

PLOR has served as an advisor for Takeda, Kyowa, 4SC, MIRAGEN, Helsinn, Recordati Rare Diseases, and Innate Pharma and holds a patent for PLCG1. MAP has served on the advisory board for Millenium/Takeda, Celgene, Gilead, Jansen, Nanotring, and Kyowa Kirin; received lecture fees from Millenium/Takeda and Jansen; and received research funding from Millenium/Takeda, Gilead, and Kura. All other authors state no conflicts of interest.

### ACKNOWLEDGMENTS

This work has been funded by the Instituto de Salud Carlos III (ISCIII)/FEDER (PI16/00156, PI19/00204, and ASOCIACION LUCHAMOS POR LA VIDA to JPV; PI17/0957 to PLOR). NGD has been supported by a predoctoral contract from UC-IDIVAL. BC holds a RyC contract from MICINN (RYC2018-024004). AEC is funded by ISCIII/MINECO/FEDER (PT17/0009/0019).

### AUTHOR CONTRIBUTIONS

Conceptualization: NGD, BC, MAP, JPV; Data Curation: NGD, BC, RAA, FRJ, JPV; Formal Analysis: MCGV, AAG, JLRP, PLOR, MAP; Investigation: NGD, RAA, LQ, MR, AEC, MG, IV, IG; Writing: NGD, BC, MAP, JPV

### SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at [www.jidonline.org](http://www.jidonline.org), and at <https://doi.org/10.1016/j.jid.2021.09.024>.

### REFERENCES

- Agar NS, Wedgeworth E, Crichton S, Mitchell TJ, Cox M, Ferreira S, et al. Survival outcomes and prognostic factors in mycosis fungoides/Sézary syndrome: validation of the revised International Society for Cutaneous Lymphomas/European Organisation for Research and Treatment of Cancer staging proposal. *J Clin Oncol* 2010;28:4730–9.
- Cecchinelli B, Lavra L, Rinaldo C, Iacovelli S, Gurtner A, Gasbarri A, et al. Repression of the antiapoptotic molecule galectin-3 by homeodomain-interacting protein kinase 2-activated p53 is required for p53-induced apoptosis. *Mol Cell Biol* 2006;26:4746–57.
- Chang LW, Patrone CC, Yang W, Rabionet R, Gallardo F, Espinet B, et al. An integrated data resource for genomic analysis of cutaneous T-cell lymphoma. *J Invest Dermatol* 2018;138:2681–3.
- Choi J, Goh G, Walradt T, Hong BS, Bunick CG, Chen K, et al. Genomic landscape of cutaneous T cell lymphoma. *Nat Genet* 2015;47:1011–9.
- Crespo P, Casar B. The chick embryo chorioallantoic membrane as an in vivo model to study metastasis. *Bio Protoc* 2016;6:1–11.
- da Silva Almeida AC, Abate F, Khabanian H, Martinez-Escala E, Guitart J, Tensen CP, et al. The mutational landscape of cutaneous T cell lymphoma and Sézary syndrome. *Nat Genet* 2015;47:1465–70.
- de Freitas RM, da Costa Maranduba CM. Myeloproliferative neoplasms and the JAK/STAT signaling pathway: an overview. *Rev Bras Hematol Hemoter* 2015;37:348–53.
- Decker T, Kovarik P. Serine phosphorylation of STATs. *Oncogene* 2000;19:2628–37.
- Eriksen KW, Kaltoft K, Mikkelsen G, Nielsen M, Zhang Q, Geisler C, et al. Constitutive STAT3-activation in Sezary syndrome: tyrphostin AG490 inhibits STAT3-activation, interleukin-2 receptor expression and growth of leukemic Sezary cells. *Leukemia* 2001;15:787–93.
- Fanok MH, Sun A, Fogli LK, Narendran V, Eckstein M, Kannan K, et al. Role of dysregulated cytokine signaling and bacterial triggers in the pathogenesis of cutaneous T-cell lymphoma. *J Invest Dermatol* 2018;138:1116–25.
- Fantini VR, Loboda A, Paweletz CP, Hendrickson RC, Pierce JW, Roth JA, et al. Constitutive activation of signal transducers and activators of transcription predicts vorinostat resistance in cutaneous T-cell lymphoma. *Cancer Res* 2008;68:3785–94.
- Gaydosik AM, Tabib T, Geskin LJ, Bayan CA, Conway JF, Lafyatis R, et al. Single-cell lymphocyte heterogeneity in advanced cutaneous T-cell lymphoma skin tumors. *Clin Cancer Res* 2019;25:4443–54.
- Hamill S, Lou HJ, Turk BE, Boggon TJ. Structural basis for noncanonical substrate recognition of cofilin/ADF proteins by LIM kinases. *Mol Cell* 2016;62:397–408.
- Isakov N, Altman A. Protein kinase C( $\theta$ ) in T cell activation. *Annu Rev Immunol* 2002;20:761–94.



- Iyer A, Hennessey D, O'Keefe S, Patterson J, Wang W, Salopek T, et al. Clonotypic heterogeneity in cutaneous T-cell lymphoma (mycosis fungoides) revealed by comprehensive whole-exome sequencing. *Blood Adv* 2019;3:1175–84.
- Kamstrup MR, Gjerdrum LM, Biskup E, Lauenborg BT, Ralfkiaer E, Woetmann A, et al. Notch1 as a potential therapeutic target in cutaneous T-cell lymphoma. *Blood* 2010;116:2504–12.
- Kiel MJ, Sahasrabudhe AA, Rolland DCM, Velusamy T, Chung F, Schaller M, et al. Genomic analyses reveal recurrent mutations in epigenetic modifiers and the JAK-STAT pathway in Sézary syndrome. *Nat Commun* 2015;6:8470.
- Kim EJ, Hess S, Richardson SK, Newton S, Showe LC, Benoit BM, et al. Immunopathogenesis and therapy of cutaneous T cell lymphoma [published correction appears in *J Clin Invest* 2007;117:836]. *J Clin Invest* 2005;115:798–812.
- Klingenberg M, Becker J, Eberth S, Kube D, Wilting J. The chick chorioallantoic membrane as an in vivo xenograft model for Burkitt lymphoma. *BMC Cancer* 2014;14:339.
- Lee EB, Fleischmann R, Hall S, Wilkinson B, Bradley JD, Gruben D, et al. Tofacitinib versus methotrexate in rheumatoid arthritis. *N Engl J Med* 2014;370:2377–86.
- Malemud CJ. The role of the JAK/STAT signal pathway in rheumatoid arthritis [published correction appears in *Ther Adv Musculoskelet Dis* 2018;10:225]. *Ther Adv Musculoskelet Dis* 2018;10:117–27.
- Mcgirt LY, Jia P, Baerenwald DA, Duszynski RJ, Dahlman KB, Zic JA, et al. Whole-genome sequencing reveals oncogenic mutations in mycosis fungoides. *Blood* 2015;126:508–19.
- Meller N, Altman A, Isakov N. New perspectives on PKC $\theta$ , a member of the novel subfamily of protein kinase C. *Stem Cells* 1998;16:178–92.
- Mognol GP, Carneiro FR, Robbs BK, Faget DV, Viola JP. Cell cycle and apoptosis regulation by NFAT transcription factors: new roles for an old player. *Cell Death Dis* 2016;7:e2199.
- Oka T, Miyagaki T. Novel and future therapeutic drugs for advanced mycosis fungoides and Sézary syndrome. *Front Med (Lausanne)* 2019;6:116.
- Park J, Yang J, Wenzel AT, Ramachandran A, Lee WJ, Daniels JC, et al. Genomic analysis of 220 CTCLs identifies a novel recurrent gain-of-function alteration in RLTTPR (p.Q575E). *Blood* 2017;130:1430–40.
- Patel VM, Flanagan CE, Martins M, Jones CL, Butler RM, Woollard WJ, et al. Frequent and persistent PLCG1 mutations in Sézary cells directly enhance PLC $\gamma$ 1 activity and stimulate NF $\kappa$ B, AP-1, and NFAT signaling. *J Invest Dermatol* 2020;140:380–9.e4.
- Pérez C, González-Rincón J, Onaindia A, Almaraz C, García-Díaz N, Pisonero H, et al. Mutated JAK kinases and deregulated STAT activity are potential therapeutic targets in cutaneous T-cell lymphoma. *Haematologica* 2015;100:e450–3.
- Pérez C, Mondéjar R, García-Díaz N, Cereceda L, León A, Montes S, et al. Advanced-stage mycosis fungoides: role of the signal transducer and activator of transcription 3, nuclear factor- $\kappa$ B and nuclear factor of activated T cells pathways. *Br J Dermatol* 2020;182:147–55.
- Prasad A, Rabionet R, Espinet B, Zapata L, Puiggros A, Melero C, et al. Identification of gene mutations and fusion genes in patients with Sézary syndrome. *J Invest Dermatol* 2016;136:1490–9.
- Radine C, Peters D, Reese A, Neuwahl J, Budach W, Jänicke RU, et al. The RNA-binding protein RBM47 is a novel regulator of cell fate decisions by transcriptionally controlling the p53-p21-axis. *Cell Death Differ* 2020;27:1274–85.
- Sakurai T, Isogaya K, Sakai S, Morikawa M, Morishita Y, Ehata S, et al. RNA-binding motif protein 47 inhibits Nrf2 activity to suppress tumor growth in lung adenocarcinoma [published correction appears in *Oncogene* 2017;36:5083]. *Oncogene* 2016;35:5000–9.
- Sheren JE, Kassenbrock CK. RNF38 encodes a nuclear ubiquitin protein ligase that modifies p53. *Biochem Biophys Res Commun* 2013;440:473–8.
- Sommer VH, Clemmensen OJ, Nielsen O, Wasik M, Lovato P, Brender C, et al. In vivo activation of STAT3 in cutaneous T-cell lymphoma. Evidence for an antiapoptotic function of STAT3. *Leukemia* 2004;18:1288–95.
- Steinberg SF. Structural basis of protein kinase C isoform function. *Physiol Rev* 2008;88:1341–78.
- Ungewickell A, Bhaduri A, Rios E, Reuter J, Lee CS, Mah A, et al. Genomic analysis of mycosis fungoides and Sézary syndrome identifies recurrent alterations in TNFR2. *Nat Genet* 2015;47:1056–60.
- Vainchenker W, Constantinescu SN. JAK/STAT signaling in hematological malignancies. *Oncogene* 2013;32:2601–13.
- Vaqué JP, Gómez-López G, Monsálvez V, Varela I, Martínez N, Pérez C, et al. PLCG1 mutations in cutaneous T-cell lymphomas. *Blood* 2014;123:2034–43.
- Wang L, Ni X, Covington KR, Yang BY, Shiu J, Zhang X, et al. Genomic profiling of Sézary syndrome identifies alterations of key T cell signaling and differentiation genes. *Nat Genet* 2015;47:1426–34.
- Willemze R, Jaffe ES, Burg G, Cerroni L, Berti E, Swerdlow SH, et al. WHO-EORTC classification for cutaneous lymphomas. *Blood* 2005;105:3768–85.
- Willerslev-Olsen A, Krejsgaard T, Lindahl LM, Litvinov IV, Fredholm S, Petersen DL, et al. Staphylococcal enterotoxin A (SEA) stimulates STAT3 activation and IL-17 expression in cutaneous T-cell lymphoma. *Blood* 2016;127:1287–96.
- Woollard WJ, Pullabhatla V, Lorenc A, Patel VM, Butler RM, Bayega A, et al. Candidate driver genes involved in genome maintenance and DNA repair in Sézary syndrome. *Blood* 2016;127:3387–97.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## SUPPLEMENTARY MATERIALS AND METHODS

### Cell culture

Human HuT 78, Jurkat, and HH cell lines were obtained from ATCC (Rockville, MD). The human MyLa cell line was obtained from the European Collection of Cell Cultures (Salisbury, United Kingdom). They were cultured in RPMI-1640 medium supplemented with 10% heat-inactivated fetal bovine serum (Gibco, Thermo Fisher Scientific, Waltham, MA). HEK293 (ATCC) and HEK-Blue IL-6 cells (HEK-IL6, InvivoGen, San Diego, CA) were cultured in DMEM medium supplemented with 10% fetal bovine serum, and the latter were also supplemented with 100  $\mu$ g/ml Normocin, 200  $\mu$ g/ml Hygromycin B Gold, and 100  $\mu$ g/ml Zeocin (InvivoGen). All cell lines were supplemented with glucose (4.5 g/l), L-glutamine (292 mg/l), streptomycin sulfate (10 mg/l), and potassium penicillin (10,000 U/l) (Lonza, Basel, Switzerland) and maintained in a humidified atmosphere at 37 °C and 5% carbon dioxide.

### Reagents and plasmid constructs

Tacrolimus, sotrastaurin, and ruxolitinib inhibitors were obtained from Selleckchem, Germany. 12-*O*-tetradecanoylphorbol-13-acetate and IL-6 (P1585 and SRP3096, respectively) were obtained from Sigma-Aldrich (St. Louis, MO).

Empty vector and PLCG1 constructs are described elsewhere (Vaqué et al., 2014). Human PRKCQ ORF clone (pCMV6-Entry-PRKCQ Myc-DDK-tagged, RC210910; Origene Technologies, Rockville, MD) was subjected to mutagenesis using QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA), following the manufacturer's instructions. The mutagenic and confirmation primer sequences are listed in [Supplementary Table S1](#).

### Determination of STAT3 activity

Quanti-Blue, luciferase reporter, and TransAM transcription factor assays were used to analyze signal transducer and activator of transcription (STAT) 3 activity in MyLa, HuT 78, HEK293, and HEK-IL6 cells. Briefly, the Quanti-Blue assay determines the activity of secreted embryonic alkaline phosphatase fused to four STAT3 binding sites (STAT3-SEAP). HEK-IL6 cells were cultured in six-well plates and treated under the desired conditions. Then, 20  $\mu$ l of the supernatant were incubated at 37 °C with 180  $\mu$ l of Quanti-Blue solution in a 96-well flat plate for 90 minutes. SEAP levels were quantified at 620–655 nm on a Spark Multimode Microplate Reader (Tecan Trading AG, Switzerland).

A luciferase report assay was performed in HEK293 cells using Dual-Glo Luciferase Assay System (Promega, Madison, WI). Cells were seeded in 12-well plates and, 24 hours later, transiently transfected with Lipofectamine LTX and PLUS reagents (Invitrogen, Waltham, MA) with a mix of DNA plasmids specific for each experiment. General conditions were 0.3  $\mu$ g of firefly STAT3 luciferase reporter (pGL4.47 vector, Promega), 0.1  $\mu$ g of pRL-Null *Renilla* luciferase control reporter vector (Promega), and 1  $\mu$ g of the specific gene or control constructs used for each experiment. Cells without transfected DNA plasmids were used as a blank. At 48 hours after transfection, passive lysis and quantification of *Renilla* and firefly levels were performed following the

manufacturer's instructions. Luminescence was measured with a GloMax-Multi reader (Promega).

To detect and quantify the capacity of activated STAT3 to bind to its DNA consensus binding sites in HuT 78 cells, an ELISA-based assay was performed using a TransAM STAT3 transcription factor assay kit (Active Motif, Carlsbad, CA) following the manufacturer's instructions. Nuclear proteins of cells treated under desired conditions were lysed using a Nuclear Extract Kit (Active Motif) following the manufacturer's instructions. Finally, absorbance was read on a Spark Multimode Microplate Reader (Tecan Trading AG) at 450 nm with an optimal reference wavelength of 655 nm. Wells without nuclear proteins were used as a blank.

### Western blot

Cells were starved overnight, treated under the desired conditions, and lysed with RIPA buffer (Sigma-Aldrich) supplemented with phosphatase and protease inhibitors (Roche, Basel, Switzerland). Whole-cell lysates were subjected to acrylamide SDS-PAGE using standard procedures, transferred onto a nitrocellulose support membrane (GE Healthcare, Chicago, IL), and western blotted. The following antibodies were used:  $\alpha$  Tubulin (Santa Cruz Biotechnology, Dallas, TX); FLAG (DDK, Origene Technologies); phospho-STAT3 Y705, phospho-STAT3 S727, STAT3, and protein kinase C  $\theta$  (Cell Signaling, Danvers, MA); and goat anti-mouse IgG DyLight 800 and goat anti-rabbit IgG DyLight 680 (Invitrogen). Bands were visualized and recorded with an Odyssey Infrared Imaging scanner (LI-COR Biosciences, Lincoln, NE) and were quantified by densitometry using Image Studio Software (LI-COR Biosciences).

### Generation of PRKCQ knockdown cell lines

Protein kinase C  $\theta$  expression was knocked down by stably transducing lentiviral particles carrying turboGFP and doxycycline-inducible nontargeting control short hairpin RNA (shRNA) or shRNA against human PRKCQ mRNA (Dharmacon, Lafayette, CO) in MyLa, HuT 78, Jurkat, and HEK-IL6 cells. Lentiviral particles were produced by cotransfection of 293T cells using the Trans-Lentiviral shRNA Packaging System (Dharmacon) following the manufacturer's instructions. Unless otherwise stated, cells were incubated with doxycycline (1  $\mu$ g/ml, Sigma-Aldrich) for 72 hours to induce GFP and shRNA expression. Transfected cells were selected with puromycin (1  $\mu$ g/ml, Sigma-Aldrich) for at least 7 days.

### Drug synergism assays

To assess drug synergism and generate combination index (CI) values, CalcuSyn software (Biosoft, Cambridge, United Kingdom) was used as previously described (Chou and Talalay, 1984). This determines whether a combination of two drugs produces a synergistic (CI < 1), additive (CI = 1), or antagonistic effect (CI > 1), evaluating the fraction of affected cells (cell viability) of each inhibitor alone compared with the combination of inhibitors.

### Cell viability and apoptosis assays

Cell proliferation was measured as the intracellular ATP content using the CellTiter-Glo Luminescent Cell Viability Assay kit (Promega) following the manufacturer instructions, and luminometric changes were quantified using the Synergy

HTX Multi-Mode Microplate Reader (BioTek, Winooski, VT). The half maximal inhibitory concentration was estimated after 48 hours of drug treatment using GraphPad Prism software (GraphPad Software, San Diego, CA). Cell proliferation, analyzed in nontargeting control or shRNA *PRKCCQ* MyLa and HuT 78 cells, was performed after incubation with doxycycline for 72 hours.

Induction of apoptosis was evaluated using a FlowCelect Annexin Red Kit (Merck Millipore, Burlington, MA) according to the manufacturer's instructions. Data were collected using a CytotFLEX flow cytometer and analyzed with CytExpert software (Beckman Coulter, Brea, CA).

#### Immunohistochemical analysis

Immunohistochemical expression of H&E, CD30, and phospho-STAT3 Y705 was assessed using routine immunohistochemical techniques for primary tumors from chicken embryos. Estimation of the percentage of tumoral cells in patient samples, both in H&E and immunohistochemistry, of every case studied was determined. The proportion of atypical cells among the infiltrate was calculated considering the percentage of cells with large atypical nuclei and the CD4 (or CD8 in selected cases) with loose of pan-T markers.

#### RT-qPCR

cDNA synthesis was performed using the SuperScript IV Reverse Transcriptase (Invitrogen). cDNA was amplified using the Power SYBR Green PCR Master Mix in a 7300 Fast Real-Time PCR System (Applied Biosystems, Waltham, MA). Specific oligos were designed using Primer-Blast (NCBI). *ACTB* expression was used to normalize values. Gene expression changes were determined using  $2^{(-\Delta\Delta Ct)}$  formula. A melting curve was generated for every run to confirm assay specificity.

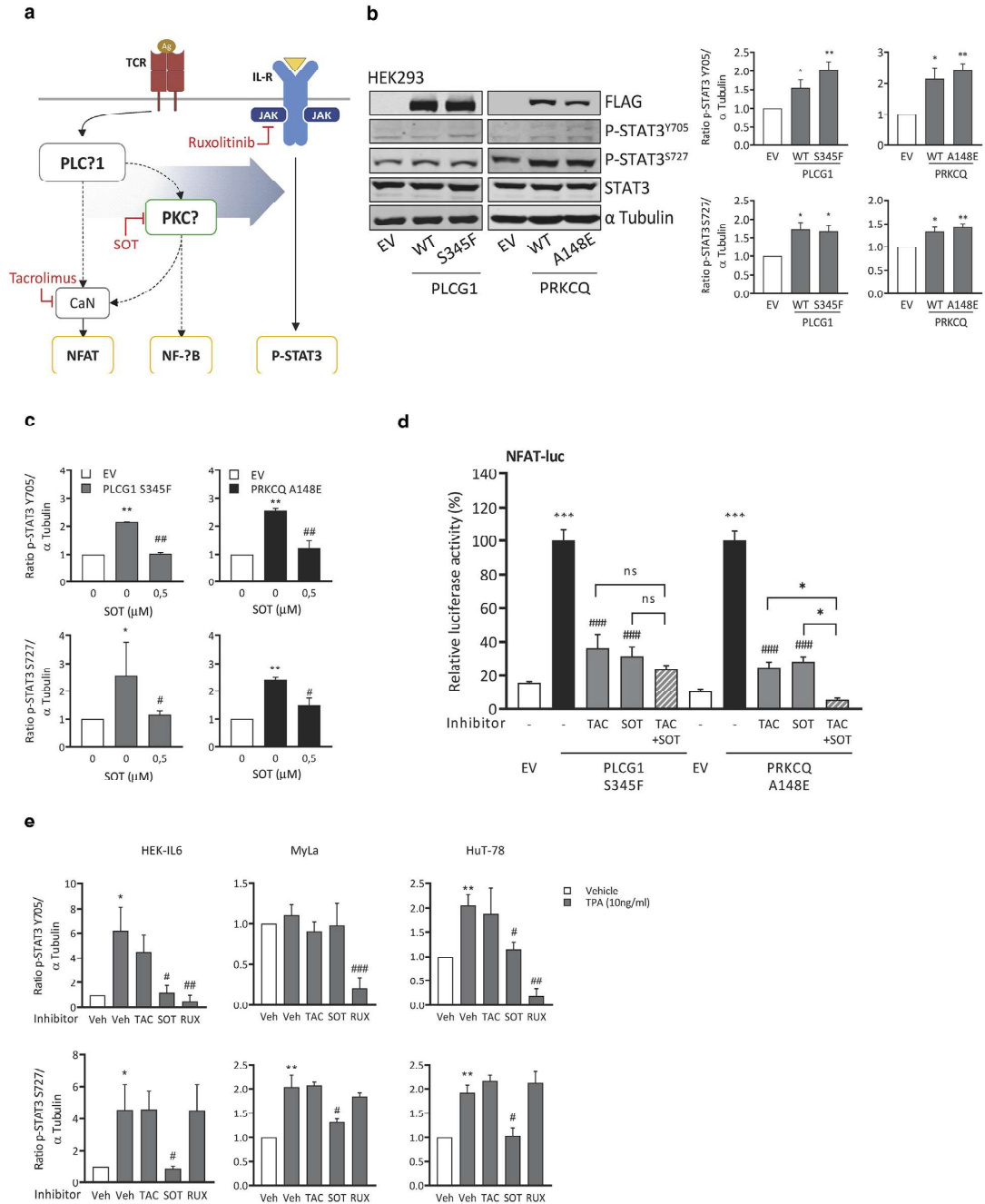
#### RNA-sequencing analysis

High-quality total RNA (RNA integrity number > 8) was isolated using TRIzol reagent (Invitrogen) following the manufacturer's instructions. The RNA-sequencing libraries were prepared following the TruSeq Stranded mRNA LT Sample Prep Kit protocol (Illumina, San Diego, CA). Briefly, total RNA (500 ng) was enriched for the polyA mRNA fraction and fragmented by divalent metal cations at high temperature. To achieve directionality, second-strand cDNA

synthesis was performed in the presence of dUTP. The blunt-ended double-stranded cDNA was 3'-adenylated, and Illumina platform-compatible adaptors with Unique Dual Indexes and Unique Molecular Identifiers (Integrated DNA Technologies, Coralville, IA) were ligated. The ligation product was enriched with 15 PCR cycles and the final library was validated on an Agilent 2100 Bioanalyzer with the DNA 7500 assay (Agilent Technologies). The libraries were sequenced in a HiSeq4000 (Illumina) following the manufacturer's protocol for dual indexing. Image analysis, base calling, and quality scoring of the run were carried out using the manufacturer's Real Time Analysis (RTA 2.7.7) software, after which FASTQ sequence files were generated. RNA-sequencing paired-end reads were mapped against the human reference genome (GRCh38) using STAR version 2.5.3a (Dobin et al., 2013) with ENCODE parameters for long RNA. Annotated genes (GENCODE version 29) were quantified using RSEM version 1.3.0 with default parameters (Li and Dewey, 2014). Differential expression was analyzed with DESeq2 version 1.18.1 (Love et al., 2014). Gene set enrichment analysis was performed to identify significantly altered gene pathways (Mootha et al., 2003; Subramanian et al., 2005).

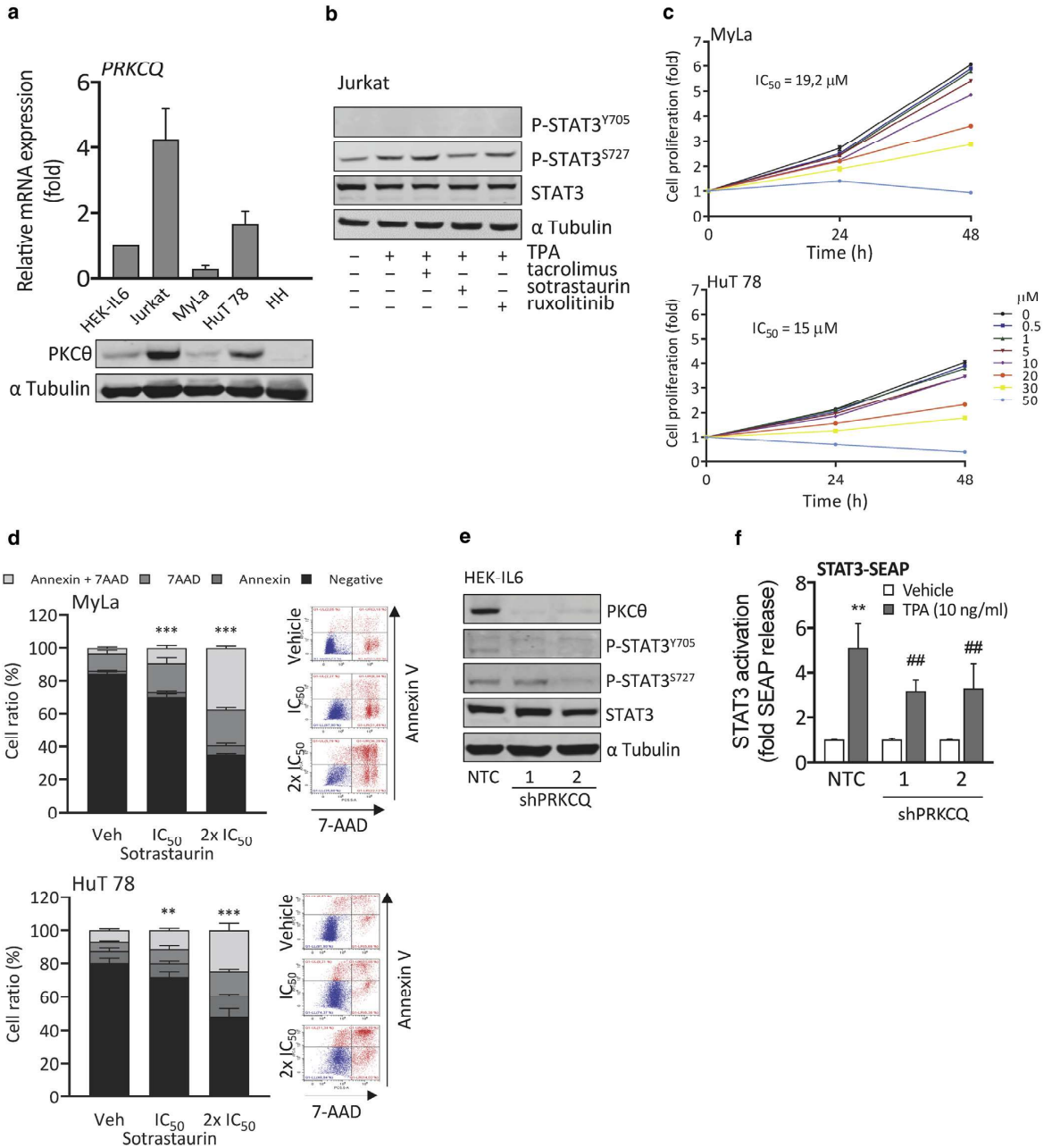
#### SUPPLEMENTARY REFERENCES

- Chou TC, Talalay P. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Adv Enzyme Regul* 1984;22(C):27–55.
- Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *Bioinforma Impact Accurate Quantif Proteomic Genet Anal Res* 2014;4:41–74.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
- Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet* 2003;34:267–73.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
- Vaqu e JP, G omez-L opez G, Mons alvez V, et al. PLCG1 mutations in cutaneous T-cell lymphomas. *Blood* 2014;123:2034–44.

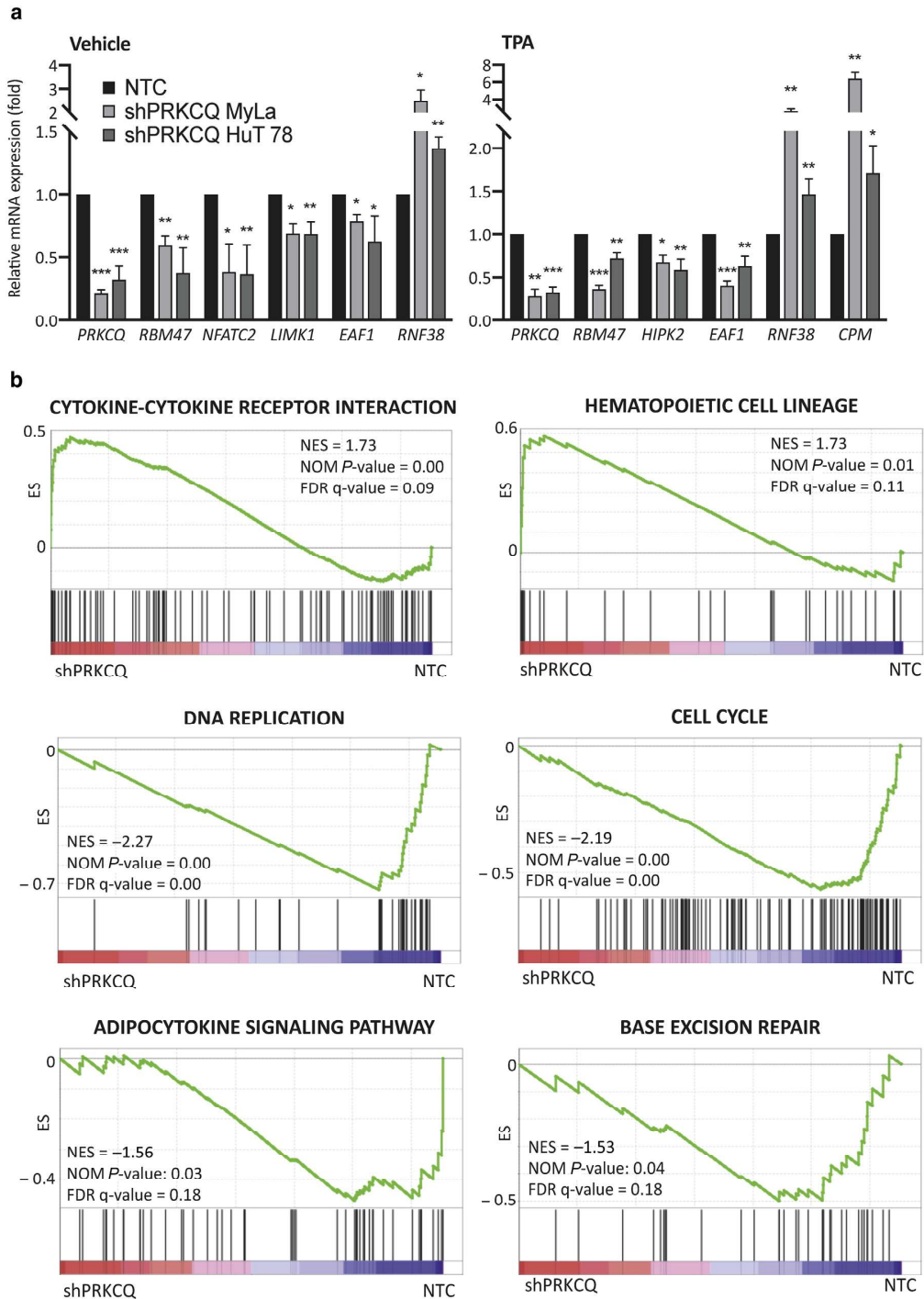


**Supplementary Figure S1. PLC $\gamma$ 1/PKC $\theta$  activates STAT3 and NFAT.** (a) Schematic representation of the signaling network mediated by PLC $\gamma$ 1 and PKC $\theta$  toward NFAT, NF- $\kappa$ B, and STAT transcription factors. (b) Western blot of HEK293 cells transiently transfected with the indicated vectors, starved, and incubated with the indicated antibodies, and quantification of three independent blots by densitometry using Image Studio Software (P-Y705 and P-S727 STAT3 relative to  $\alpha$  tubulin from each blot). Quantification of three independent blots from Figure 1b (c) and Figure 1c (e) by densitometry. (d) NFAT luciferase reporter activity in HEK293 cells transfected with the indicated vectors and treated with the indicated inhibitors (1  $\mu$ M, 24 hours,  $n = 3$ ). Images are representative of each western blot ( $n = 3$ ). Data are mean  $\pm$  SEM. Student's  $t$ -test: \* $P < 0.05$ , \*\* $P < 0.01$ , and \*\*\* $P < 0.001$  versus EV (b–d) or vehicle (e); # $P < 0.05$ , ## $P < 0.01$ , and ### $P < 0.001$  versus PLCG1 S345F or PRKCQ A148E treated with control vehicle (c, d), or TPA (e). CaN, calcineurin; EV, empty vector; ns, not significant; PKC, protein kinase C; P-STAT, phosphorylated signal transducer and activator of transcription; RUX, ruxolitinib; SOT, sotrastaurin; STAT, signal transducer and activator of transcription; TAC, tacrolimus; TPA, 12-*O*-tetradecanoylphorbol-13-acetate; Veh, vehicle; WT, wild type.



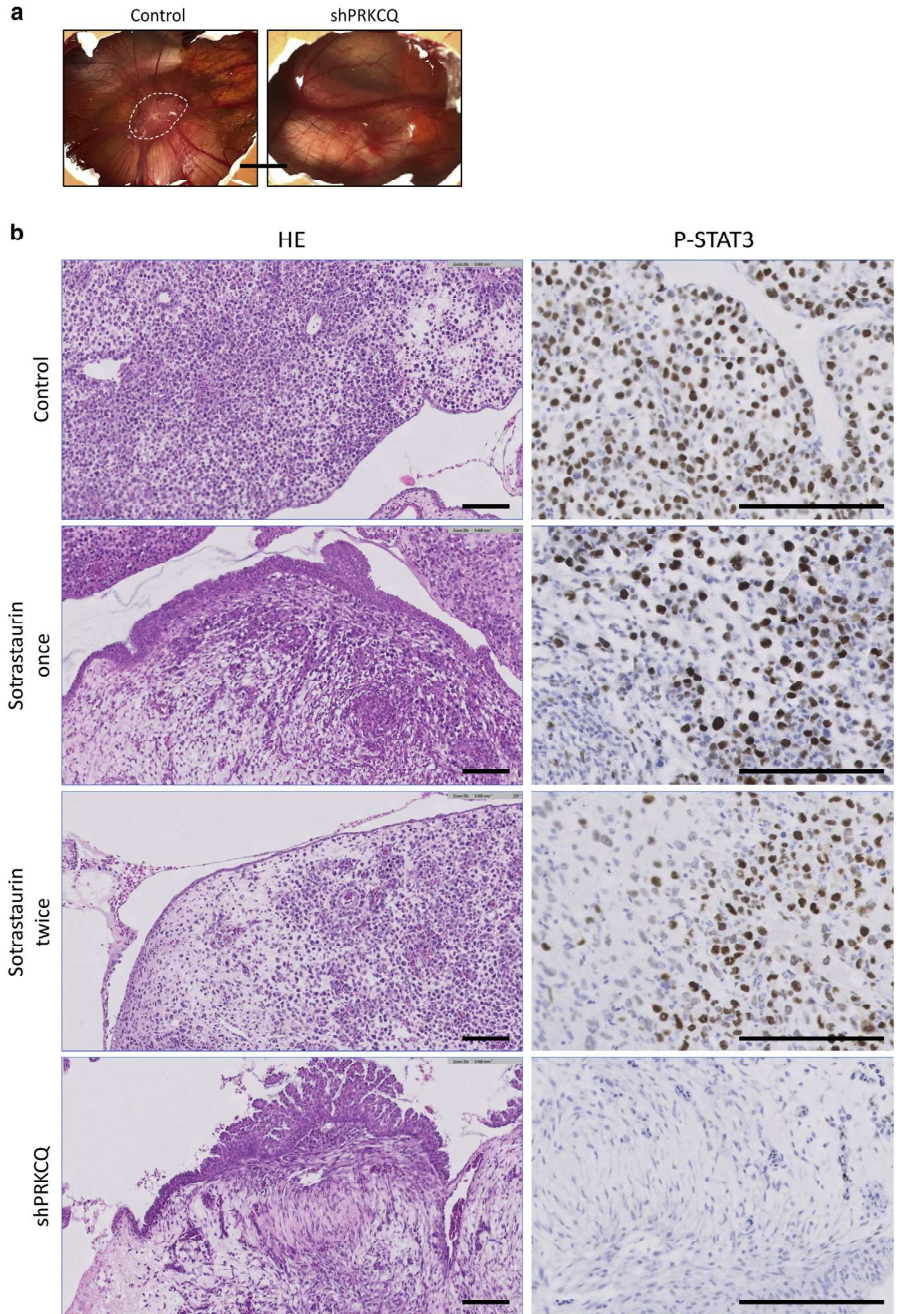


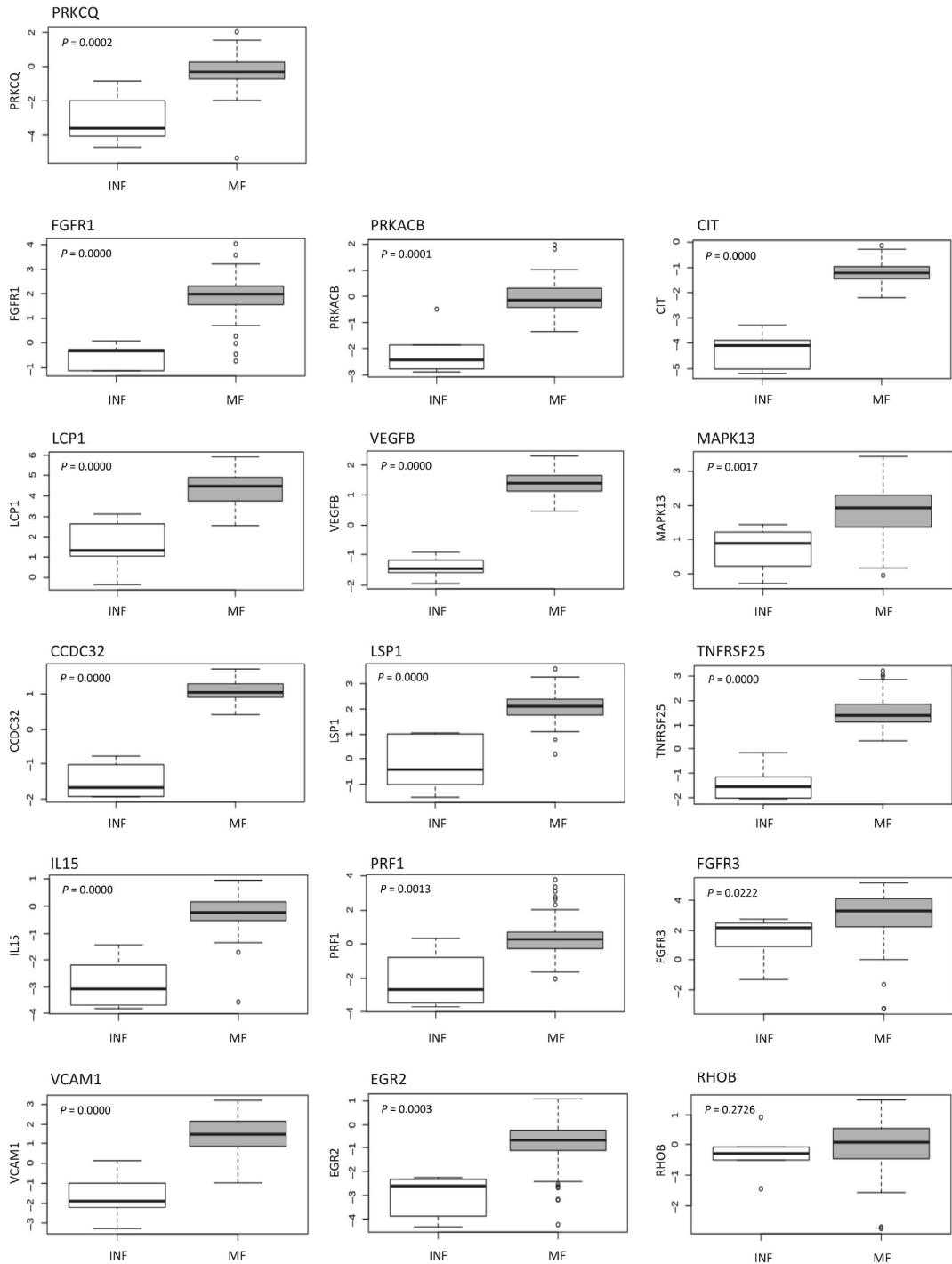
**Supplementary Figure S2. PKC $\theta$  inhibition impairs STAT3 activation and proliferation, and induce apoptosis in CTCL cells.** (a) mRNA expression levels (top) and protein expression (bottom) of *PRKCQ*/PKC $\theta$  in HEK-IL6, Jurkat, MyLa, HuT 78, and HH cells. (b) Western blot analyses of starved Jurkat cells treated with the indicated inhibitors (1  $\mu$ M, 3 hours) and TPA (10 ng/ml, 1 hours) and incubated with the indicated antibodies. (c) Proliferation analyses in MyLa and HuT 78 cells treated with increasing concentrations of sotrastaurin at 0, 24, and 48 hours. (d) Percentage of early (Annexin V) or late (7-AAD) apoptotic MyLa and HuT 78 cells incubated with sotrastaurin (24 hours, n = 3). Representative plots of Annexin V (Y axis) and 7-AAD (X axis) staining data of each condition are shown. Student *t*-test: comparison between viable cells (negative staining) treated with sotrastaurin and viable cells treated with control vehicle. (e) Western blot analyses of inducible NTC or shPRKCQ HEK-IL6 cells incubated with the indicated antibodies. (f) Quantification of SEAP release in NTC and shPRKCQ HEK-IL6 cells treated with TPA (10 ng/ml) overnight, n = 3). Images are representative of each western blot. Data are mean  $\pm$  SEM. Student's *t*-test: \*\**P* < 0.01 versus NTC treated with vehicle (f); ##*P* < 0.01 versus NTC treated with TPA (f). CTCL, cutaneous T-cell lymphoma; IC<sub>50</sub>, half maximal inhibitory concentration; NTC, nontargeting control; PKC, protein kinase C; P-STAT, phosphorylated signal transducer and activator of transcription; shPRKCQ, *PRKCQ* short hairpin RNA; STAT, signal transducer and activator of transcription; TPA, 12-*O*-tetradecanoylphorbol-13-acetate.



**Supplementary Figure S3. PKC $\theta$  target genes and pathways in CTCL cells.** (a) mRNA expression levels of the indicated genes in shPRKCQ MyLa and HuT 78 cells compared with NTC cells treated with control vehicle or TPA (10 ng/ml, 24 hours). (b) Selected GSEA plots shared in shPRKCQ MyLa and HuT 78 cells compared with NTC cells (see Supplementary Table S2 for a complete list of GSEA signatures). Gene sets with a NOM  $P$ -value < 0.05 and FDR  $q$ -value < 0.25 were considered significantly enriched. Data are mean  $\pm$  SEM ( $n = 3$ ). Student's  $t$ -test: \* $P < 0.05$ , \*\* $P < 0.01$ , and \*\*\* $P < 0.001$  versus NTC. CTCL, cutaneous T-cell lymphoma; FDR  $q$ -value, false discovery rate  $q$ -value; GSEA, gene set enrichment analysis; NES, normalized enrichment score; NOM  $P$ -value, nominal  $P$ -value; NTC, nontargeting control; PKC, protein kinase C; shPRKCQ, *PRKCQ* short hairpin RNA; TPA, 12-*O*-tetradecanoylphorbol-13-acetate.

**Supplementary Figure S4. Pharmacological and genetic inhibition of PKC $\theta$  reduces P-STAT3 in vivo.** (a) Representative images of control and shPRKCCQ chicken embryos. Primary tumor is rounded with a dashed circle. (b) H&E and P-STAT3 Y705 staining of paraffin sections of control, tumors treated once or twice with sotrastaurin, and tumors with PKC $\theta$  deficiency (shPRKCCQ). Bar = 100  $\mu$ m. PKC, protein kinase C; P-STAT, phosphorylated signal transducer and activator of transcription; shPRKCCQ, *PRKCCQ* short hairpin RNA; STAT, signal transducer and activator of transcription.





**Supplementary Figure S5. Expression of PRKCQ and its target genes in MF versus inflammatory controls.** Box plots showing the expression of PRKCQ and its target genes in six samples of inflammatory dermatoses and 81 samples of patients with MF. INF, inflammatory dermatoses; MF, mycosis fungoides.



Supplementary Table S1. Primer Sequences Used in this Work

Name	Forward 5'-3' Sequence	Reverse 5'-3' Sequence
PRKCQ-A148E mutagenesis	CTTTCGCTGCTGATCICACCCCGCGTGAT	ATCAGCGCCGGGTGAGATCAAGCAGCAAG
PRKCQ-K409R mutagenesis	CATCTTCTTTAAGGCCCTATTGCGAAAAATTGATTGG TTTTCTTGAATT	AATCAAGAAAACCAATCAATTTTTTCGCAATAA GGGCCTTAAAGAAAGATG
PRKCQ-A148E confirmation	ACTTTCTGGAATGAGTGACACAA	ATGTGGGCTGTGGGAAGAAG
PRKCQ-K409R confirmation	TGGGAAAGGAAGTTTTGGCA	TCCAGGCCAAGGAAAGAAC
Alu	ACGCCTGTAATCCAGGACTT	TCGCCAGGCTGGCTGGGTGCA
GAPDH (Chicken)	GAGGAAAGGTCGCCTGGTGGATCG	GGTGAGGACAAGCAGTGAGGAACG
PRKCQ	CCATGTCGCCATTTCTTCGG	GCCCGTTCTCTGATTCGACA
RBM47	CAGCCATGAGCAGTGACTCG	TCTTGACCATGCTGTAGCC
NFATC2	GTATTACCTGCGGGGTGAC	TCTGATTTCTGGCAGGAGGTC
LIMK1	ACGCTACTTTGTTGCACCTG	ATCATAGATCCTCTGGCCGC
EAF1	GGCCTCCTCCACACTATT	GCCGTTTGTCCCTTGAAC
NOTCH1	GAATGGCGGGAAGTGTGAAGC	TAGTCTGCCACGCCTCTGC
RNF38	GGTGAGACTTCAGAGCCTGTT	GAGAGAGCCGCTGTCTCTTAG
HIPK2	CCCATCTACTCTACCAGCC	GAGTAGCCAGCGTGCTTAGA
CPM	CAGGAAGGATGGAAGCGTT	GTGTTCTTTGAAAACCGCC
ACTB	AGTGTGACGTGGACATCCGCAAG	ATCCACATCTGCTGGAAGTGGAC

Supplementary Table S2. GSEA Data

## Positively Deregulated DMSO

	Gene Set	NES	NOM P-Value	FDR q-Value	FWER P-Value
1	KEGG_RIBOSOME	1.870	0.000	0.114	0.122
2	KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	1.810	0.006	0.115	0.232
3	KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	1.760	0.013	0.117	0.330
4	KEGG_HEMATOPOIETIC_CELL_LINEAGE	1.730	0.011	0.116	0.419
5	KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	1.730	0.000	0.097	0.429
6	KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION	1.570	0.034	0.270	0.846
7	KEGG_BASAL_CELL_CARCINOMA	1.560	0.030	0.250	0.868
8	KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	1.540	0.026	0.249	0.913
9	KEGG_PARKINSONS_DISEASE	1.540	0.011	0.222	0.913
10	KEGG_CHEMOKINE_SIGNALING_PATHWAY	1.490	0.009	0.271	0.955
11	KEGG_SPLICEOSOME	1.490	0.008	0.253	0.956
12	KEGG_OXIDATIVE_PHOSPHORYLATION	1.470	0.015	0.275	0.977
13	KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	1.440	0.052	0.299	0.987
14	KEGG_THYROID_CANCER	1.360	0.108	0.424	1.000
15	KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	1.350	0.068	0.429	1.000
16	KEGG_ALZHEIMERS_DISEASE	1.300	0.064	0.532	1.000
17	KEGG_HUNTINGTONS_DISEASE	1.290	0.075	0.537	1.000
18	KEGG_REGULATION_OF_AUTOPHAGY	1.280	0.161	0.513	1.000
19	KEGG_PROTEASOME	1.270	0.124	0.513	1.000
20	KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_HEPARAN_SULFATE	1.240	0.200	0.563	1.000
21	KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY	1.230	0.207	0.581	1.000
22	KEGG_FOCAL_ADHESION	1.180	0.161	0.681	1.000
23	KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE	1.140	0.297	0.770	1.000
24	KEGG_HEDGEHOG_SIGNALING_PATHWAY	1.110	0.312	0.831	1.000
25	KEGG_MAPK_SIGNALING_PATHWAY	1.110	0.214	0.805	1.000
26	KEGG_GLYCOLYSIS_GLUONEOGENESIS	1.090	0.345	0.843	1.000
27	KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	1.080	0.309	0.846	1.000
28	KEGG_ENDOMETRIAL_CANCER	1.080	0.352	0.834	1.000
29	KEGG_VIRAL_MYOCARDITIS	1.070	0.353	0.822	1.000
30	KEGG_GAP_JUNCTION	1.070	0.365	0.802	1.000

(continued)

## Supplementary Table S2. Continued

## Positively Deregulated DMSO

	Gene Set	NES	NOM P-Value	FDR q-Value	FWER P-Value
31	KEGG_PRIMARY_IMMUNODEFICIENCY	1.060	0.375	0.808	1.000
32	KEGG_ERBB_SIGNALING_PATHWAY	1.050	0.363	0.804	1.000
33	KEGG_RENAL_CELL_CARCINOMA	1.050	0.375	0.788	1.000
34	KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT	1.040	0.418	0.786	1.000
35	KEGG_PATHWAYS_IN_CANCER	1.020	0.395	0.805	1.000
36	KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	1.020	0.402	0.793	1.000
37	KEGG_ECM_RECEPTOR_INTERACTION	1.010	0.425	0.798	1.000
38	KEGG_NEUROTROPHIN_SIGNALING_PATHWAY	0.990	0.470	0.843	1.000
39	KEGG_ARACHIDONIC_ACID_METABOLISM	0.980	0.478	0.837	1.000
40	KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	0.980	0.470	0.822	1.000
41	KEGG_COLORECTAL_CANCER	0.960	0.514	0.847	1.000
42	KEGG_GLYCOSAMINOGLYCAN_DEGRADATION	0.930	0.543	0.924	1.000
43	KEGG_EPITHELIAL_CELL_SIGNALING_IN_H_PYLORI_INFECTION	0.930	0.573	0.905	1.000
44	KEGG_RNA_DEGRADATION	0.900	0.614	0.940	1.000
45	KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY	0.880	0.653	0.971	1.000
46	KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	0.880	0.637	0.966	1.000
47	KEGG_CARDIAC_MUSCLE_CONTRACTION	0.880	0.653	0.948	1.000
48	KEGG_FRUCTOSE_AND_MANNANOSE_METABOLISM	0.840	0.690	1.000	1.000
49	KEGG_ENDOCYTOSIS	0.810	0.847	1.000	1.000
50	KEGG_P53_SIGNALING_PATHWAY	0.800	0.804	1.000	1.000
51	KEGG_BETA_ALANINE_METABOLISM	0.780	0.731	1.000	1.000
52	KEGG_BLADDER_CANCER	0.770	0.797	1.000	1.000
53	KEGG_FC_EPSILON_R1_SIGNALING_PATHWAY	0.760	0.841	1.000	1.000
54	KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	0.750	0.868	1.000	1.000
55	KEGG_NUCLEOTIDE_EXCISION_REPAIR	0.750	0.862	1.000	1.000
56	KEGG_GLUTATHIONE_METABOLISM	0.750	0.836	1.000	1.000
57	KEGG_GLYCEROLIPID_METABOLISM	0.720	0.867	1.000	1.000
58	KEGG_PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	0.710	0.835	1.000	1.000
59	KEGG_LEISHMANIA_INFECTION	0.700	0.917	1.000	1.000
60	KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION	0.690	0.942	1.000	1.000
61	KEGG_BASAL_TRANSCRIPTION_FACTORS	0.680	0.881	1.000	1.000
62	KEGG_CHRONIC_MYELOID_LEUKEMIA	0.680	0.941	1.000	1.000
63	KEGG_RNA_POLYMERASE	0.660	0.916	1.000	1.000
64	KEGG_BASE_EXCISION_REPAIR	0.650	0.936	1.000	1.000
65	KEGG_ACUTE_MYELOID_LEUKEMIA	0.630	0.965	0.998	1.000
66	KEGG_PROTEIN_EXPORT	0.620	0.928	0.989	1.000
67	KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS	0.620	0.998	0.975	1.000
68	KEGG_MTOR_SIGNALING_PATHWAY	0.560	0.991	0.983	1.000

## Negatively Deregulated DMSO

	Gene Set	NES	NOM P-Value	FDR q-Value	FWER P-Value
1	KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	-1.550	0.055	1.000	0.849
2	KEGG_ABC_TRANSPORTERS	-1.550	0.046	1.000	0.850
3	KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY	-1.490	0.031	1.000	0.952
4	KEGG_AMINO_SUGAR_AND_NUCLEOTIDE_SUGAR_METABOLISM	-1.420	0.051	1.000	0.991
5	KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	-1.400	0.057	1.000	0.997
6	KEGG_CITRATE_CYCLE_TCA_CYCLE	-1.390	0.090	0.999	0.997
7	KEGG_NOTCH_SIGNALING_PATHWAY	-1.390	0.087	0.877	0.998
8	KEGG_STEROID_BIOSYNTHESIS	-1.390	0.115	0.771	0.998
9	KEGG_LYSOSOME	-1.380	0.041	0.712	0.998
10	KEGG_MELANOMA	-1.380	0.064	0.651	0.998
11	KEGG_VASOPRESSIN_REGULATED_WATER_REABSORPTION	-1.370	0.101	0.605	0.999
12	KEGG_AUTOIMMUNE_THYROID_DISEASE	-1.360	0.139	0.615	0.999
13	KEGG_LONG_TERM_POTENTIATION	-1.340	0.092	0.635	1.000
14	KEGG_CELL_ADHESION_MOLECULES_CAMS	-1.330	0.079	0.599	1.000
15	KEGG_PPAR_SIGNALING_PATHWAY	-1.330	0.128	0.570	1.000
16	KEGG_CALCIIUM_SIGNALING_PATHWAY	-1.290	0.080	0.659	1.000

(continued)

## Supplementary Table S2. Continued

## Negatively Deregulated DMSO

	Gene Set	NES	NOM P-Value	FDR q-Value	FWER P-Value
17	KEGG_SELENOAMINO_ACID_METABOLISM	-1.280	0.164	0.663	1.000
18	KEGG_DORSO_VENTRAL_AXIS_FORMATION	-1.260	0.174	0.665	1.000
19	KEGG_GNRH_SIGNALING_PATHWAY	-1.250	0.158	0.671	1.000
20	KEGG_PROPANOATE_METABOLISM	-1.250	0.201	0.659	1.000
21	KEGG_ONE_CARBON_POOL_BY_FOLATE	-1.240	0.203	0.633	1.000
22	KEGG_ALLOGRAFT_REJECTION	-1.210	0.218	0.708	1.000
23	KEGG_TYPE_I_DIABETES_MELLITUS	-1.200	0.218	0.712	1.000
24	KEGG_AXON_GUIDANCE	-1.190	0.162	0.710	1.000
25	KEGG_VIBRIO_CHOLERAЕ_INFECTION	-1.180	0.223	0.709	1.000
26	KEGG_GALACTOSE_METABOLISM	-1.180	0.262	0.684	1.000
27	KEGG_PROSTATE_CANCER	-1.150	0.242	0.758	1.000
28	KEGG_TRYPTOPHAN_METABOLISM	-1.140	0.282	0.747	1.000
29	KEGG_N_GLYCAN_BIOSYNTHESIS	-1.140	0.251	0.725	1.000
30	KEGG_ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM	-1.120	0.310	0.762	1.000
31	KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	-1.110	0.263	0.765	1.000
32	KEGG_GRAFT_VERSUS_HOST_DISEASE	-1.110	0.340	0.743	1.000
33	KEGG_TIGHT_JUNCTION	-1.110	0.265	0.726	1.000
34	KEGG_LONG_TERM_DEPRESSION	-1.100	0.316	0.733	1.000
35	KEGG_STARCH_AND_SUCROSE_METABOLISM	-1.090	0.342	0.735	1.000
36	KEGG_WNT_SIGNALING_PATHWAY	-1.070	0.316	0.791	1.000
37	KEGG_PRION_DISEASES	-1.050	0.405	0.830	1.000
38	KEGG_CYTOSOLIC_DNA_SENSING_PATHWAY	-1.040	0.395	0.831	1.000
39	KEGG_SPHINGOLIPID_METABOLISM	-1.040	0.399	0.818	1.000
40	KEGG_DILATED_CARDIOMYOPATHY	-1.030	0.408	0.845	1.000
41	KEGG_PEROXISOME	-1.020	0.398	0.825	1.000
42	KEGG_MELANOGENESIS	-0.990	0.483	0.897	1.000
43	KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	-0.980	0.459	0.913	1.000
44	KEGG_ADHERENS_JUNCTION	-0.980	0.481	0.899	1.000
45	KEGG_ETHER_LIPID_METABOLISM	-0.970	0.489	0.904	1.000
46	KEGG_INOSITOL_PHOSPHATE_METABOLISM	-0.970	0.498	0.886	1.000
47	KEGG_ARGININE_AND_PROLINE_METABOLISM	-0.940	0.526	0.948	1.000
48	KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY	-0.930	0.552	0.963	1.000
49	KEGG_JAK_STAT_SIGNALING_PATHWAY	-0.930	0.601	0.957	1.000
50	KEGG_DRUG_METABOLISM_OTHER_ENZYMES	-0.920	0.568	0.955	1.000
51	KEGG_PYRIMIDINE_METABOLISM	-0.900	0.666	0.989	1.000
52	KEGG_FATTY_ACID_METABOLISM	-0.900	0.607	0.971	1.000
53	KEGG_SMALL_CELL_LUNG_CANCER	-0.870	0.695	1.000	1.000
54	KEGG_APOPTOSIS	-0.870	0.710	1.000	1.000
55	KEGG_NON_SMALL_CELL_LUNG_CANCER	-0.870	0.662	1.000	1.000
56	KEGG_PANCREATIC_CANCER	-0.850	0.704	1.000	1.000
57	KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS	-0.850	0.748	1.000	1.000
58	KEGG_INSULIN_SIGNALING_PATHWAY	-0.840	0.818	1.000	1.000
59	KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	-0.830	0.666	1.000	1.000
60	KEGG_LYSINE_DEGRADATION	-0.830	0.722	1.000	1.000
61	KEGG_CYSTEINE_AND_METHIONINE_METABOLISM	-0.820	0.700	1.000	1.000
62	KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	-0.820	0.741	1.000	1.000
63	KEGG_TGF_BETA_SIGNALING_PATHWAY	-0.810	0.778	1.000	1.000
64	KEGG_TYPE_II_DIABETES_MELLITUS	-0.810	0.755	0.996	1.000
65	KEGG_GLIOMA	-0.790	0.806	1.000	1.000
66	KEGG_PYRUVATE_METABOLISM	-0.780	0.766	1.000	1.000
67	KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	-0.780	0.804	1.000	1.000
68	KEGG_PURINE_METABOLISM	-0.750	0.952	1.000	1.000
69	KEGG_VEGF_SIGNALING_PATHWAY	-0.750	0.885	1.000	1.000
70	KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	-0.740	0.879	1.000	1.000
71	KEGG_TYROSINE_METABOLISM	-0.740	0.835	1.000	1.000
72	KEGG_OOCYTE_MEIOSIS	-0.720	0.964	1.000	1.000
73	KEGG_HOMOLOGOUS_RECOMBINATION	-0.720	0.847	1.000	1.000

(continued)

**Supplementary Table S2. Continued****Negatively Deregulated DMSO**

	Gene Set	NES	NOM P-Value	FDR q-Value	FWER P-Value
74	KEGG_PENTOSE_PHOSPHATE_PATHWAY	-0.710	0.814	0.992	1.000
75	KEGG_BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS	-0.700	0.850	0.997	1.000
76	KEGG_AMYOTROPHIC_LATERAL_SCLEROSIS_ALS	-0.670	0.936	1.000	1.000
77	KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION	-0.650	0.983	1.000	1.000
78	KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	-0.620	0.949	1.000	1.000
79	KEGG_BUTANOATE_METABOLISM	-0.620	0.936	0.999	1.000
80	KEGG_MISMATCH_REPAIR	-0.580	0.943	1.000	1.000
81	KEGG_CELL_CYCLE	-0.570	1.000	0.994	1.000
82	KEGG_DNA_REPLICATION	-0.460	0.998	0.997	1.000

**Positively Deregulated TPA**

	Gene Set	NES	NOM P-Value	FDR q-Value	FWER P-Value
1	KEGG_RIBOSOME	2.250	0.000	0.000	0.000
2	KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION	1.670	0.006	0.265	0.549
3	KEGG_HEMATOPOIETIC_CELL_LINEAGE	1.630	0.013	0.249	0.681
4	KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	1.570	0.030	0.302	0.834
5	KEGG_ARACHIDONIC_ACID_METABOLISM	1.460	0.073	0.536	0.988
6	KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT	1.370	0.087	0.788	0.999
7	KEGG_ECM_RECEPTOR_INTERACTION	1.360	0.085	0.682	0.999
8	KEGG_FOCAL_ADHESION	1.340	0.052	0.669	1.000
9	KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	1.340	0.127	0.618	1.000
10	KEGG_PARKINSONS_DISEASE	1.270	0.120	0.805	1.000
11	KEGG_CARDIAC_MUSCLE_CONTRACTION	1.260	0.144	0.771	1.000
12	KEGG_OXIDATIVE_PHOSPHORYLATION	1.250	0.084	0.721	1.000
13	KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC	1.240	0.184	0.706	1.000
14	KEGG_BETA_ALANINE_METABOLISM	1.230	0.201	0.681	1.000
15	KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION	1.210	0.170	0.713	1.000
16	KEGG_REGULATION_OF_AUTOPHAGY	1.200	0.243	0.710	1.000
17	KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	1.160	0.219	0.791	1.000
18	KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	1.140	0.266	0.831	1.000
19	KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	1.140	0.244	0.789	1.000
20	KEGG_BUTANOATE_METABOLISM	1.110	0.324	0.831	1.000
21	KEGG_VALINE_LEUCINE_AND_Isoleucine_DEGRADATION	1.100	0.335	0.822	1.000
22	KEGG_CHEMOKINE_SIGNALING_PATHWAY	1.100	0.293	0.804	1.000
23	KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	1.050	0.385	0.896	1.000
24	KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	1.050	0.378	0.890	1.000
25	KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE	1.030	0.432	0.905	1.000
26	KEGG_INOSITOL_PHOSPHATE_METABOLISM	1.020	0.411	0.896	1.000
27	KEGG_CALCIIUM_SIGNALING_PATHWAY	1.010	0.417	0.885	1.000
28	KEGG_SPHINGOLIPID_METABOLISM	1.010	0.420	0.864	1.000
29	KEGG_DORSO_VENTRAL_AXIS_FORMATION	1.000	0.448	0.868	1.000
30	KEGG_ARGININE_AND_PROLINE_METABOLISM	0.980	0.470	0.898	1.000
31	KEGG_COLORECTAL_CANCER	0.960	0.521	0.922	1.000
32	KEGG_BASAL_CELL_CARCINOMA	0.960	0.510	0.907	1.000
33	KEGG_EPITHELIAL_CELL_SIGNALING_IN_HELICOBACTER_PYLORI_INFECTION	0.940	0.535	0.922	1.000
34	KEGG_ENDOMETRIAL_CANCER	0.930	0.564	0.926	1.000
35	KEGG_TYROSINE_METABOLISM	0.910	0.554	0.962	1.000
36	KEGG_CYSTEINE_AND_METHIONINE_METABOLISM	0.890	0.605	0.983	1.000
37	KEGG_ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM	0.890	0.620	0.967	1.000
38	KEGG_NOTCH_SIGNALING_PATHWAY	0.880	0.633	0.974	1.000
39	KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	0.860	0.648	0.977	1.000
40	KEGG_NEUROTROPHERIN_SIGNALING_PATHWAY	0.860	0.743	0.968	1.000
41	KEGG_JAK_STAT_SIGNALING_PATHWAY	0.860	0.727	0.950	1.000
42	KEGG_ALZHEIMERS_DISEASE	0.850	0.786	0.934	1.000
43	KEGG_ERBB_SIGNALING_PATHWAY	0.840	0.746	0.933	1.000
44	KEGG_RENAL_CELL_CARCINOMA	0.830	0.786	0.939	1.000
45	KEGG_GLUTATHIONE_METABOLISM	0.790	0.778	0.991	1.000

(continued)



## Supplementary Table S2. Continued

## Positively Deregulated TPA

	Gene Set	NES	NOM P-Value	FDR q-Value	FWER P-Value
46	KEGG_GLYCEROLIPID_METABOLISM	0.780	0.807	0.989	1.000
47	KEGG_ENDOCYTOSIS	0.780	0.910	0.976	1.000
48	KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS	0.770	0.861	0.974	1.000
49	KEGG_HUNTINGTONS_DISEASE	0.750	0.968	0.974	1.000
50	KEGG_LYSOSOME	0.750	0.945	0.965	1.000
51	KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	0.710	0.899	0.991	1.000
52	KEGG_PYRUVATE_METABOLISM	0.640	0.920	1.000	1.000
53	KEGG_GLYCOLYSIS_GLUconeogenesis	0.580	0.985	1.000	1.000
54	KEGG_ACUTE_MYELOID_LEUKEMIA	0.550	0.993	1.000	1.000
55	KEGG_THYROID_CANCER	0.540	0.989	1.000	1.000
56	KEGG_PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	0.470	0.984	0.997	1.000

## Negatively Deregulated TPA

	Gene Set	NES	NOM P-Value	FDR q-Value	FWER P-Value
1	KEGG_DNA_REPLICATION	-2.270	0.000	0.000	0.000
2	KEGG_CELL_CYCLE	-2.190	0.000	0.000	0.000
3	KEGG_HOMOLOGOUS_RECOMBINATION	-1.900	0.002	0.032	0.067
4	KEGG_OOCYTE_MEIOSIS	-1.870	0.000	0.031	0.082
5	KEGG_PYRIMIDINE_METABOLISM	-1.850	0.000	0.031	0.104
6	KEGG_MELANOMA	-1.840	0.002	0.032	0.125
7	KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	-1.810	0.009	0.033	0.155
8	KEGG_MISMATCH_REPAIR	-1.680	0.013	0.105	0.460
9	KEGG_CYTOSOLIC_DNA_SENSING_PATHWAY	-1.670	0.008	0.102	0.489
10	KEGG_TYPE_I_DIABETES_MELLITUS	-1.640	0.018	0.114	0.572
11	KEGG_ALLOGRAFT_REJECTION	-1.560	0.041	0.193	0.794
12	KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY	-1.560	0.032	0.179	0.798
13	KEGG_GALACTOSE_METABOLISM	-1.550	0.029	0.179	0.823
14	KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION	-1.540	0.022	0.181	0.847
15	KEGG_BASE_EXCISION_REPAIR	-1.530	0.039	0.177	0.862
16	KEGG_PROSTATE_CANCER	-1.530	0.011	0.174	0.876
17	KEGG_AMINO_SUGAR_AND_NUCLEOTIDE_SUGAR_METABOLISM	-1.470	0.035	0.239	0.954
18	KEGG_P53_SIGNALING_PATHWAY	-1.460	0.038	0.245	0.961
19	KEGG_AUTOIMMUNE_THYROID_DISEASE	-1.420	0.083	0.301	0.985
20	KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	-1.390	0.065	0.336	0.994
21	KEGG_PRION_DISEASES	-1.380	0.094	0.342	1.000
22	KEGG_PURINE_METABOLISM	-1.370	0.038	0.340	1.000
23	KEGG_GAP_JUNCTION	-1.350	0.084	0.357	1.000
24	KEGG_CELL_ADHESION_MOLECULES_CAMS	-1.350	0.072	0.345	1.000
25	KEGG_MELANOGENESIS	-1.340	0.121	0.364	1.000
26	KEGG_LEISHMANIA_INFECTION	-1.330	0.084	0.362	1.000
27	KEGG_NUCLEOTIDE_EXCISION_REPAIR	-1.330	0.089	0.358	1.000
28	KEGG_LONG_TERM_POTENTIATION	-1.260	0.143	0.483	1.000
29	KEGG_LONG_TERM_DEPRESSION	-1.240	0.176	0.523	1.000
30	KEGG_VASOPRESSIN_REGULATED_WATER_REABSORPTION	-1.220	0.197	0.557	1.000
31	KEGG_LYSINE_DEGRADATION	-1.220	0.179	0.545	1.000
32	KEGG_APOPTOSIS	-1.200	0.153	0.575	1.000
33	KEGG_VIRAL_MYOCARDITIS	-1.190	0.175	0.578	1.000
34	KEGG_STEROID_BIOSYNTHESIS	-1.180	0.264	0.600	1.000
35	KEGG_GRAFT_VERSUS_HOST_DISEASE	-1.160	0.269	0.623	1.000
36	KEGG_HEDGEHOG_SIGNALING_PATHWAY	-1.160	0.275	0.617	1.000
37	KEGG_AXON_GUIDANCE	-1.150	0.220	0.633	1.000
38	KEGG_TIGHT_JUNCTION	-1.120	0.265	0.679	1.000
39	KEGG_CITRATE_CYCLE_TCA_CYCLE	-1.120	0.279	0.684	1.000
40	KEGG_GLIOMA	-1.110	0.277	0.672	1.000
41	KEGG_RNA_DEGRADATION	-1.110	0.289	0.667	1.000
42	KEGG_PANCREATIC_CANCER	-1.090	0.297	0.706	1.000
43	KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	-1.080	0.257	0.712	1.000

(continued)

## Supplementary Table S2. Continued

## Negatively Deregulated\_TPA

	Gene Set	NES	NOM P-Value	FDR q-Value	FWER P-Value
44	KEGG_FRUCTOSE_AND_MANNANOSE_METABOLISM	-1.070	0.356	0.737	1.000
45	KEGG_DILATED_CARDIOMYOPATHY	-1.070	0.339	0.726	1.000
46	KEGG_DRUG_METABOLISM_OTHER_ENZYMES	-1.060	0.387	0.734	1.000
47	KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	-1.060	0.375	0.720	1.000
48	KEGG_ABC_TRANSPORTERS	-1.050	0.380	0.713	1.000
49	KEGG_MAPK_SIGNALING_PATHWAY	-1.050	0.328	0.717	1.000
50	KEGG_TYPE_II_DIABETES_MELLITUS	-1.040	0.407	0.725	1.000
51	KEGG_NON_SMALL_CELL_LUNG_CANCER	-1.030	0.400	0.733	1.000
52	KEGG_ADHERENS_JUNCTION	-1.030	0.386	0.720	1.000
53	KEGG_PRIMARY_IMMUNODEFICIENCY	-1.030	0.416	0.708	1.000
54	KEGG_STARCH_AND_SUCROSE_METABOLISM	-1.020	0.428	0.735	1.000
55	KEGG_PATHWAYS_IN_CANCER	-1.010	0.403	0.746	1.000
56	KEGG_SPLICEOSOME	-1.000	0.440	0.752	1.000
57	KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	-0.990	0.463	0.753	1.000
58	KEGG_PENTOSE_PHOSPHATE_PATHWAY	-0.990	0.454	0.740	1.000
59	KEGG_VIBRIO_CHOLERAЕ_INFECTIОN	-0.980	0.482	0.769	1.000
60	KEGG_PROTEASOME	-0.960	0.510	0.807	1.000
61	KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY	-0.950	0.535	0.832	1.000
62	KEGG_PROPANOATE_METABOLISM	-0.940	0.541	0.822	1.000
63	KEGG_GLYCOSAMINOGLYCAN_DEGRADATION	-0.940	0.502	0.821	1.000
64	KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS	-0.940	0.599	0.820	1.000
65	KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	-0.930	0.535	0.810	1.000
66	KEGG_TRYPTOPHAN_METABOLISM	-0.930	0.567	0.810	1.000
67	KEGG_SMALL_CELL_LUNG_CANCER	-0.930	0.591	0.802	1.000
68	KEGG_GNRH_SIGNALING_PATHWAY	-0.920	0.598	0.818	1.000
69	KEGG_WNT_SIGNALING_PATHWAY	-0.910	0.686	0.835	1.000
70	KEGG_N_GLYCAN_BIOSYNTHESIS	-0.900	0.611	0.833	1.000
71	KEGG_BLADDER_CANCER	-0.890	0.613	0.854	1.000
72	KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	-0.880	0.640	0.851	1.000
73	KEGG_ONE_CARBON_POOL_BY_FOLATE	-0.850	0.621	0.914	1.000
74	KEGG_TGF_BETA_SIGNALING_PATHWAY	-0.840	0.718	0.930	1.000
75	KEGG_ETHER_LIPID_METABOLISM	-0.830	0.695	0.934	1.000
76	KEGG_INSULIN_SIGNALING_PATHWAY	-0.810	0.856	0.976	1.000
77	KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	-0.770	0.907	1.000	1.000
78	KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY	-0.770	0.870	1.000	1.000
79	KEGG_RNA_POLYMERASE	-0.770	0.798	1.000	1.000
80	KEGG_CHRONIC_MYELOID_LEUKEMIA	-0.750	0.905	1.000	1.000
81	KEGG_PROTEIN_EXPORT	-0.730	0.844	1.000	1.000
82	KEGG_AMYOTROPHIC_LATERAL_SCLEROSIS_ALS	-0.710	0.910	1.000	1.000
83	KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY	-0.690	0.925	1.000	1.000
84	KEGG_PEROXISOME	-0.650	0.980	1.000	1.000
85	KEGG_VEGF_SIGNALING_PATHWAY	-0.640	0.974	1.000	1.000
86	KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_HEPARAN_SULFATE	-0.640	0.899	1.000	1.000
87	KEGG_PPAR_SIGNALING_PATHWAY	-0.640	0.950	1.000	1.000
88	KEGG_MTOR_SIGNALING_PATHWAY	-0.580	0.987	1.000	1.000
89	KEGG_FATTY_ACID_METABOLISM	-0.570	0.976	1.000	1.000
90	KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	-0.530	0.995	1.000	1.000
91	KEGG_SELENOAMINO_ACID_METABOLISM	-0.500	0.982	1.000	1.000
92	KEGG_BASAL_TRANSCRIPTION_FACTORS	-0.500	0.998	1.000	1.000
93	KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	-0.440	1.000	1.000	1.000
94	KEGG_BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS	-0.420	0.994	1.000	1.000

Abbreviations: FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes; NTC, nontargeting control; shPRKCQ, *PRKCQ* short hairpin RNA; TPA, 12-*O*-tetradecanoylphorbol-13-acetate.

Positively and negatively deregulated gene sets based on KEGG pathways of DMSO- or TPA-stimulated MyLa and HuT 78 shPRKCQ cells compared with NTC cells.

Supplementary Table S3. Clinical Data Associated with the MF Samples

Patient	Sex	Age at Diagnosis	Sample	Histopathology	% Tumoral Cells
MF1	Female	56	MF1.1	Plaque	5
			MF1.2	Plaque	15
			MF1.3	Tumor	90
MF2	Male	58	MF2.1	Tumor	90
			MF2.2	Plaque	15
			MF2.3	Tumor	80
			MF2.4	Plaque	90
			MF2.5	Tumor	90
MF3	Female	53	MF3.1	Tumor	90
			MF3.2	Plaque	5
			MF3.3	Plaque	10
			MF3.4	Plaque	8
			MF3.5	Tumor	80
MF4	Male	50	MF4.1	Plaque	5
			MF4.2	Plaque	10
MF5	Male	65	MF5.1	Plaque	15
			MF5.2	Tumor	80
MF6	Male	36	MF6.1	Plaque	20
			MF6.2	Plaque	10
			MF6.3	Plaque	15
			MF6.4	Plaque	40
			MF6.5	Plaque	10
MF7	Male	64	MF7	Tumor	30
MF8	Male	71	MF8.1	Plaque	5
			MF8.2	Tumor	90
			MF8.3	Tumor	90
MF9	Female	45	MF9.1	Plaque	20
			MF9.2	Tumor	80
			MF9.3	Tumor	90
			MF9.4	Plaque	20
			MF9.5	Tumor	90
			MF9.6	Plaque	15
MF10	Female	58	MF10.1	Plaque	N/D
			MF10.2	Plaque	N/D
MF11	Female	41	MF11	Plaque	N/D
MF12	Female	70	MF12.1	Plaque	5
			MF12.2	Plaque	10
			MF12.3	Plaque	10
			MF12.4	Plaque	7
			MF12.5	Plaque	10
			MF12.6	Plaque	40
			MF12.7	Plaque	30
			MF12.8	Plaque	15
			MF12.9	Plaque	10
			MF12.10	Tumor	90
MF13	Male	16	MF13.1	Plaque	15
			MF13.2	Plaque	10
			MF13.3	Plaque	5
MF14	Male	32	MF14.1	Plaque	5
			MF14.2	Plaque	5
MF15	Male	37	MF15.1	Plaque	15
			MF15.2	Plaque	10
MF16	Male	54	MF16.1	Plaque	15
			MF16.2	Plaque	20
			MF16.3	Plaque	20
			MF16.4	Tumor	80
MF17	Male	56	MF17	Plaque	5

(continued)

Supplementary Table S3. Continued

Patient	Sex	Age at Diagnosis	Sample	Histopathology	% Tumoral Cells
MF18	Female	76	MF18.1	Plaque	10
			MF18.2	Plaque	20
			MF18.3	Tumor	80
MF19	Female	56	MF19	Plaque	10
MF20	Male	47	MF20.1	Plaque	5
			MF20.2	Plaque	10
			MF20.3	Plaque	10
MF21	Male	32	MF21	Plaque	20
MF22	Female	59	MF22.1	Plaque	20
			MF22.2	Plaque	15
			MF22.3	Plaque	20
			MF22.4	Plaque	15
MF23	Male	14	MF23.1	Plaque	5
			MF23.2	Plaque	10
			MF23.3	Plaque	10
MF24	Male	32	MF24	Plaque	10
MF25	Female	77	MF25.1	Plaque	15
			MF25.2	Tumor	80
MF26	Male	61	MF26.1	Plaque	15
			MF26.2	Plaque	20
			MF26.3	Tumor	95
			MF26.4	Plaque	30
MF27	Female	31	MF27.1	Plaque	15
			MF27.2	Plaque	15
INF1	Male	73	N/A	Inflammatory dermatoses	N/A
INF2	Male	49	N/A	Inflammatory dermatoses	N/A
INF3	Male	74	N/A	Inflammatory dermatoses	N/A
INF4	Male	38	N/A	Inflammatory dermatoses	N/A
INF5	Female	72	N/A	Inflammatory dermatoses	N/A
INF6	Male	36	N/A	Inflammatory dermatoses	N/A

Abbreviations: MF, mycosis fungoides; N/A: not applied; N/D: not determined.