Exploring the roles of cell plasticity and tumor heterogeneity in renal cell carcinoma : implications on development, progression and response to immunotherapy

Guillaume Davidson

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

Institut de Génétique et de Biologie Moléculaire et Cellulaire

THÈSE présentée par :

Guillaume DAVIDSON

soutenue le : **14 septembre 2023**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Sciences de la vie et de la Santé

Exploring the roles of cell plasticity and tumor heterogeneity in renal cell carcinoma: implications on development, progression and response to immunotherapy.

THÈSE dirigée par :

M. Gabriel Malouf	PU-PH, IGBMC, Illkirch

RAPPORTEURS :

Mme. Sophie Postel-Vinay	Médecin des CLCC, Gustave Roussy, Paris
M. Aurélien De Reyniès	PU-PH, Université de Paris-Cité, Paris

---

AUTRES MEMBRES DU JURY :

Mme. Angela Giagrande	Directeur de recherche, IGBMC, Illkirch
M. François Radvanyi	Directeur de recherche, Institut Curie, Paris

"*Science is the poetry of reality*"

Richard Dawkins

"*Only when our clever brain and our human heart work together in harmony can we achieve our true potential*"

Jane Goodall

"*The love for all living creatures is the most noble attribute of man*"

Charles Darwin

I dedicate this manuscript to my family, friends and all the people who inspired me to follow this path.

Remerciements (acknowledgements)

Alors que j'achève l'écriture de ce manuscrit, je tiens à remercier sincèrement tous les gens qui ont accompagné mon parcours.

Tout d'abord, je remercie Dr**. Sophie Postel-Vinay** et Dr. **Aurélien De Reyniès** pour l'évaluation de mon travail de thèse. Je remercie également Dr. **Angela Giagrande** et Dr. **François Radvanyi** d'avoir en plus aidé ma progression au cours de réunions annuelles, vos conseils m'ont été très précieux et les discussions scientifiques passionnantes !

Ensuite, je tiens à remercier mes anciens collègues. **Maud Contrant** pour m'avoir soutenu pendant ma licence. **Isabelle Michel**, **Dominique Kobi** et **Thomas Strub** pour m'avoir formé à la biologie moléculaire. **Marie Ennen**, notre travail ensemble était le point de départ d'un nouveau parcours pour moi, j'espère que tu trouves le bonheur dans la voie que tu as choisie. **Sébastien Coassolo**, merci pour ton amitié, le travail sur ton projet était très enrichissant pour moi. **Shilpy Joshi**, it was really pleasant working with you and I learned a lot about ChIP-seq, I hope you find your new work fulfilling. **Thomas Kleiber**, merci pour ta gentillesse, j'ai appris beaucoup sur le scRNA-seq dans ton projet et je te souhaite beaucoup de réussite. **Bujamin Vokshi**, merci pour ton amitié et de m'avoir encouragé à poursuivre en thèse, le projet RMC était un tournant pour moi. Thanks to my friends and past colleagues **Giovanni Gambi** and **Pietro Berico**.

Je remercie mes amis de la plateforme GenomEast pour m'avoir accueilli et formé à la bio-informatique. **Damien Plassard**, **Tao Ye**, **Mathieu Jung**, **Stéphanie Le Gras**, **Céline Keime**, **Amandine Velt**, **Stéphanie Levon**, **Constance Vagne**, **Christelle Thibault-Carpentier** et **Bernard Jost**, merci infiniment à tous, j'ai appris énormément à vos côtés !

Je tiens à remercier tous mes collègues actuels. Un énorme merci à **Alexandra Helleux**, j'ai été très chanceux de pouvoir réaliser ma thèse à tes côtés, ta confiance dans le projet m'a aidé à garder le cap même dans les nombreux moments d'adversité, sans tes nombreuses compétences et ton amitié qui m'est très précieuse je n'aurais pas pu aller au bout ! Merci à mes deux amis de bureau pour leur humour et leur bonne humeur, **Alexandre Haller** et **Antonin Lallement**, accrochez-vous les gars, vous avez de belles thèses et carrières en perspective. Merci à **Léane Seno** pour sa grande gentillesse, bon courage pour ton projet qui est en très bonne voie, tu as beaucoup de qualités et un bel avenir devant toi. Merci aux autres membres de l'équipe pour leur aide et disponibilité, **Igor Martianov**, **Gabrielle Mengus** et **Sehrish Khan Basai**.

Je remercie tous les médecins que j'ai pu rencontrer. **Véronique Debien**, ton aide m'a été très importante et les discussions avec toi m'ont beaucoup ouvert aux aspects de la recherche translationelle, j'ai rarement rencontré quelqu'un d'aussi dévouée à son art et c'est une source d'inspiration pour moi ! **Philippe Baltzinger**, merci beaucoup pour ton aide, ton amitié, ton soutien et ta bonne humeur. Merci **Justine Gantzer**, tu as toujours été disponible quand j'ai eu besoin d'aide, je te souhaite beaucoup de courage et de réussite pour la fin de ton parcours scientifique. Merci à **Clara Vacheret** et **Emilie Holder** pour leurs conseils éclairés. Merci aux autres médecins que j'ai eu la chance de côtoyer, **Jonathan Thouvenin**, **Alice Kieny** et **Antonin Fattori**.

Je remercie également mes amis du reste de l'IGBMC. **Maguelone Nogaret**, merci pour ta grande amitié, ta positivité et ta bonne humeur sans failles. **Clémence Elly** pour son agréable compagnie et gentillesse. Merci à **Dounia Chraa** et **Marine Guivarch** pour la collaboration avec leur équipe. Merci à mes amis **Olga Kolosova**, **Julie Lafouge**, **Amélie Zachayus**, **Max Cigrang**, **Jolian Obid**, **Annabelle Klein**, **Qingshuang Cai**, **Dasha Yanushko**, **Julie Terzic** **Manon Boivin** et **Abbas Kazrani** pour leur soutien durant ces trois années.

Merci à **Gabriel Malouf** d'avoir accepté d'être mon encadrant de thèse.

Enfin, je remercie profondément ma famille et mes amis. Un grand merci à mon père, **Irwin**, sans qui le projet n'aurait pas pu aboutir, merci d'avoir supervisé le travail et de m'avoir fait confiance. Merci à **Baptiste Hornecker** pour son amitié qui m'a inspirée depuis de nombreuses années. Je remercie ma mère et mon frère, **Christine**, **Maxime**, merci pour votre support !

List of Abbreviations

APC	Antigen-presenting cell
apCAF	Antigen-presenting CAF
ATL	Ascending thin limb
ATP	Adenosine triphosphate
BAF	BRG1/BRM-associated factor
bHLH	Basic helix-loop-helix
C/EBP	CCAAT enhancer-binding protein
CAF	Cancer-associated fibroblast
ccRCC	Clear-cell renal cell carcinoma
CDC	Collecting ducts carcinoma
CDS	Coding sequence
chRCC	Chromophobe renal cell carcinoma
CM	Mesenchymal capping cell
CNT	Connecting tubule
C-SB	Comma-shaped body
CSC	Cancer stem cell
CXCL	C-X-C motif chemokine ligand
DCT	Distal convoluted tubule
DNA	Deoxyribonucleic acid
DTL	Descending thin limb
EC	Endothelial cell
ECM	Extra-cellular matrix
EdMT	Endothelial-to-mesenchymal transition
EGF	Epidermal growth factor
EMP	Epithelial-mesenchymal plasticity
EMT	Epithelial to mesenchymal transition
ERK	Extracellular signal-regulated kinase
FA	Fatty acid
FGF	Fibroblast growth factor
FOX	Forkhead box
GoF	Gain of function
GSEA	Gene set enrichment analysis
GTP	Guanosine-5'-triphosphate
H&E	Hematoxylin and eosin
HGF	Hepatocyte growth factor
HIF	Hypoxia-induced transcription factor
HIV	Human immunodeficiency virus
HNSC	Head and neck squamous cell carcinoma
HOX	Homeobox
HRE	Hypoxia response element
IC	Intercalated cell
iCAF	Inflammatory-CAF
ICB	Immune checkpoint blockade
ICI	Immune checkpoint inhibitor
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
IMDC	Inner medulla collecting duct
ISUP	International society of urological pathology
JAK	Janus kinase
KIRC	Kidney renal clear cell carcinoma
LD	Lipid droplet

lincRNA	Long intergenic RNA
lncRNA	Long non-coding RNA
LoF	Loss of function
MAPK	Mitogen-activated protein kinase
MD	Macula densa
MEL	Melanocytic
MES	Mesenchymal-like
MET	Mesenchymal to epithelial transition
MHC	Major histocompatibility complex
MiT	Microphthalmia transcription factor
MM	Metanephric mesenchymal cell
MMP	Matrix metalloproteases
MRT	Malignant rhabdoid tumor
MVAC	Methotrexate, vinblastine, doxorubicin, and cisplatin
myCAF	Myofibroblastic-CAF
ncBAF	Non-canonical BAF
NK	Natural killer
ORF	Open reading frame
OxPhos	Oxidative-phosphorylation
PAX	Paired-box
PBAF	Polybromo-associated BAF
PC	Principal cell
PCT	Proximal convoluted tubule
PDGF	Platelet derived growth factor
pH	Potential of hydrogen
PIM	posterior intermediate mesoderm
POU	Pit-oct-unc domain
PRC	Polycomb repressor complex
pRCC	Papillary renal cell carcinoma
PST	Proximal straight tubule
PT	Proximal tubule
PTA	Pre-tubular aggregate
RAS	Rat sarcoma virus
RCC	Renal cell carcinoma
RMC	Renal medullary carcinoma
RNA	Ribonucleic acid
RTK	Receptor tyrosine kinase
RTqPCR	Reverse transcription quantitative polymerase chain reaction
RV	Renal vesicle
scRNA-seq	Single-cell RNA-sequencing
SKCM	Skin cutaneous melanoma
SMAD	Small and mothers against dpp
SOX	SRY-related HMG-box
S-SB	S-shaped body
STAT	Signal-transducer and activator of transcription
SWI/SNF	Switch/sucrose non-fermentable complex
TAG	Triacylglycerol
TAL	Thick ascending limb
TAM	Tumor-associated macrophage
TCGA	The cancer genome atlas
TEC	Tumor endothelial cell
TEM	Trans-endothelial migration
Tfh	T-follicular helper
TGF	Transforming growth factor

Th	T-helper
TKI	Tyrosine kinase inhibitor
TLS	Tertiary lymphoid structure
TMB	Tumor mutational burden
TME	Tumor microenvironment
TNF	Tumor necrosis factor
tRCC	Translocation renal cell carcinoma
Treg	Regulatory T-cell
TSC	Tuberous sclerosis
UB	Uretic bud
UV	Ultraviolet
VEGF	Vascular endothelial growth factor
VHL	Von Hippel-Lindau
WHO	World health organization
WNT	Wingless-related integration site

Table of contents

Chapter I: General introduction 23

1. Cancer 23

A. Origins and hallmarks of cancer 23

B. Invasion and metastasis: the epithelial to mesenchymal transition 25

a. Type 1 EMT 25

b. Type 2 EMT 25

c. Type 3 EMT 27

i. Molecular changes 27

ii. Transcriptional regulation of EMT 29

iii. Signals inducing EMT 31

C. Cell plasticity and phenotype switching: the example of melanoma 31

a. Oncogenesis and classification of melanoma 31

b. Dynamic phenotype switching of melanoma cells 33

D. Non-mutational epigenetic reprogramming: the role of long non-coding RNAs 35

E. Tumor microenvironment 37

a. $CD8^+$ T-lymphocytes 39

b. Natural killer cells 41

c. $CD4^+$ T-lymphocytes 41

d. Macrophages 43

e. Other immune cells 43

f. Fibroblasts 45

g. Vascularization 47

2. Renal cell carcinoma 49

A. Classification of renal cell carcinoma 49

B. Origin of renal cell carcinoma: the kidney nephron 51

a. Cellular composition of the nephron 51

b. Transcriptional control of kidney development, identity and regeneration 53

C. Clear-cell renal cell carcinoma 55

a. Oncogenesis of clear-cell renal cell carcinoma 57

i. The VHL tumor suppressor 57

ii. The HIF transcription factor 59

iii. The PBAF chromatin remodeler 61

iv. The BAP1 deubiquitinase 63

v. The SETD2 methyltransferase 63

b. Standard of care 65
i. Tyrosine kinase inhibitors 65
ii. Immune checkpoint inhibitors 65
iii. The staging system in ccRCC 67
c. Heterogeneity and classification of clear-cell renal cell carcinomas 69
i. Composition of the microenvironment 69
ii. Epithelial to mesenchymal transition in ccRCC 69
iii. The grading system in ccRCC 71
iv. Transcriptomic classifications of ccRCC 71
D. Renal medullary carcinoma 73
a. Clinical characteristics 73
b. Molecular alterations 73
c. Oncogenesis of renal medullary carcinoma 73
3. Objectives of the research 74
A. Describe ccRCC cancer cell plasticity and TME at single-cell level 74
B. Establish relationship between ccRCC tumor composition and patient survival 74
C. Identify new biomarkers for survival and resistance to immunotherapy in ccRCC 75
D. Elucidate genesis and vulnerabilities of RMC 75
Chapter II: Results 78
1. Mesenchymal-like tumor cells and myofibroblastic cancer-associated fibroblasts are associated with progression and immunotherapy response of clear-cell renal cell carcinoma 79
A. Introduction 81
B. Material and Methods. 82
C. Results 91
a. Single-cell RNA-seq profiling of ccRCC tumours 91
b. Plasticity of ccRCC tumour cells 94
c. Comparison of tumour cell states to proximal straight tubules 96
d. Heterogeneity of the mesangial and CAF populations 100
e. Association between cancer cell and stromal cell states with patient outcome 102
f. Spatial association of ccRCC.mes and myCAFs 104
g. Signaling between ccRCC.mes and myCAFs reveals potential therapeutic targets. 108
h. MyCAFs associate with resistance to first-line nivolumab +/- ipilimumab treatment. 108
i. Enrichment for ccRCC.mes and myCAFs at metastatic sites 111
D. Discussion. 111
a. Tumour cell state association with ccRCC disease outcome. 111

b. MyCAF enrichment at the tumour/NAT interface and association with resistance to ICI therapy. 113
E. References 116
F. Supplementary figures 121
G. Summary of single-cell profiled populations 138
2. Tumor heterogeneity and immune evasion 141
A. Tumor composition analysis by stage 141
B. Analysis of ligand-receptor interactions involving immune cells 141
a. Interaction of ccRCC with immune cells 141
b. Interaction of myCAFs with CD8 T-cells 143
c. Expression of PD-L1 143
3. Biomarkers of survival and resistance to immunotherapy 147
A. Epithelial/mesenchymal populations as markers of survival 147
B. Signature genes of myCAFs as biomarker for ICI resistance 147
C. Use of lncRNAs as biomarkers 151
a. Unsupervised clustering of the BIONIKK cohort based on lincRNAs 151
i. Clustering based on immune-infiltration 151
ii. Clustering based on lincRNAs 151
iii. Identification of candidate biomarkers 155
b. The mesenchymal tumor cell marker LINC01615 157
i. LINC01615 in ccRCC 157
ii. LINC01615 in melanoma 160
4. SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance. 167
A. Abstract. 168
B. Introduction. 168
C. Results. 169
a. RMC ontogeny and molecular characterization of tumour cell states. 169
b. Tumour cell state of a patient derived RMC xenograft. 172
c. Characterization of the RMC microenvironment. 172
d. Cultured RMC cells recapitulate the EMT gradient. 173
e. SMARCB1 re-expression in RMC cells represses the oncogenic program. 174
f. SMARCB1 re-expression in RMC cells induces ferroptotic cell death. 175
g. SMARCB1 re-expression promotes genomic SWI/SNF re-localization to enhancers with TFCP2L1 motifs. 176
h. SMARCB1 re-expression remodels MYC genomic binding. 178

D. Discussion. 179
a. Oncogenic transformation of TAL cells into RMC. 179
b. Distinct cells-of-origin and oncogenic mechanisms in RMC and RT. 180
c. A link between RMC ferroptosis and sickle cell trait. 181
E. Methods. 182
F. Bibliography 189
G. Main figures 194
H. Supplementary figures 206
5. Material and methods 228
Chapter III: Concluding discussion 235
A. Malignant transformation of proximal tubules into ccRCC 235
B. Dedifferentiation of ccRCC and disease progression 236
a. Transcriptional control of EMT in ccRCC 236
b. Induction of EMT in ccRCC and metastatic spread 237
c. Impact of EMT on tumor heterogeneity during disease progression 239
i. EMT and vascularization 239
ii. EMT and immune evasion 240
C. Using myCAFs as biomarkers for immunotherapy resistance 241
D. Using LINC01615 as a marker of the mesenchymal state in cancer cells 241
E. Parallels between pathogenesis of RMC and ccRCC 242
F. Future perspectives 243
Chapter IV: Bibliography 246
1. References 246
2. My publications and oral communications 260
Chapter V: Annexes 264

Résumé en Français (French summary)

Introduction

Le cancer du rein représente approximativement 2% de l'ensemble des cancers. Le carcinome rénal du rein correspond à un groupe composé d'au moins 16 sous-types histologiques différents dont le cancer du rein à cellules claires (cRCC) est le plus fréquent (~75% des cas). Sur le plan génétique, le cRCC est caractérisé par l'inactivation du gène suppresseur de tumeur von Hippel-Lindau (VHL), qui est combinée à des mutations récurrentes dans des gènes de remodelage de la chromatine, notamment PBRM1 (~50%), BAP1 (~15%) et SETD2 (~15%). Le cRCC est un type de cancer agressif dont ~30% des patients sont diagnostiqués avec une tumeur ayant envahie les ganglions lymphatiques ou des organes distants et pour les patients avec formes métastatiques la survie à 5 ans est inférieure à 15%. Plusieurs études ont montré l'intérêt de l'immunothérapie par inhibiteurs des checkpoints immunitaires en première ligne ou après échec de thérapies ciblées, cependant une grande partie des patients ne présentent pas de bénéfice clinique démontrant l'importance de comprendre la biologie de ces tumeurs afin de déterminer les mécanismes de résistance au traitement et de trouver des biomarqueurs prédisant la réponse thérapeutique. Récemment, des études se sont penchées sur ces questions en utilisant le séquençage d'ARN en cellule unique (scRNA-seq), mais les auteurs se sont focalisés sur les cellules immunitaires laissant les autres compartiments, notamment le stroma et les cellules cancéreuses, relativement inconnus.

Afin de répondre à ces problématiques, notre équipe a généré un ensemble de données de séquençage. J'ai réalisé les analyses bioinformatiques de 7 tumeurs et 2 tissus adjacents issus de scRNA-seq dont deux ont été séquencées en transcriptomique spatiale pour décrire l'hétérogénéité des cellules cancéreuses, la composition du microenvironnement tumoral et l'interaction entre les deux. J'ai ensuite intégré ces données avec 495 échantillons issus de séquençage à ARN (RNA-seq) présents dans la base de donnée du « The Cancer Genome Atlas – Kidney Renal Cell Carcinoma » (TCGA-KIRC), afin d'établir les profils de composition des tumeurs ainsi que de faire le lien entre les types cellulaires présents et le pronostic vital des patients. Enfin, j'ai intégré les données de scRNA-seq avec 97 échantillons de RNA-seq provenant de l'étude clinique BIONIKK dont une grande partie des patients ont été traités par inhibiteurs des checkpoints immunitaires afin de déterminer les mécanismes de résistance et de nouveaux biomarqueurs prédisant la réponse à ce type de traitement. Les expériences ayant permis de générer les données et toutes les vérifications expérimentales ont été réalisées par Alexandra Helleux.

Un deuxième axe de mes recherches concerne un autre sous-type de cancer du rein : le carcinome médullaire du rein (CMR). Ce cancer très rare touche des personnes jeunes et de descendance africaine présentant tous la drépanocytose ou un trait drépanocytaire. Ce type de cancer est très agressif avec une survie médiane de 13 mois et peu d'options thérapeutiques existent en dehors de la chimiothérapie. Ce cancer trouve son origine dans la médulla du rein et est caractérisé par la perte d'une sous-unité du complexe de remodelage de la chromatine SWI/SNF encodée par le gène SMARCB1. L'oncogenèse est relativement méconnue, le

modèle actuel propose que les conditions hypoxiques de la médulla couplées avec la fragilité des hématies causée par la drépanocytose génèrent des conditions favorables à la mutation de SMARCB1 qui va causer l'apparition du cancer. Pour mieux comprendre cette maladie, j'ai réalisé l'analyse de 3 échantillons de CMR issus de scRNA-seq ainsi que d'une petite série d'échantillons RNA-seq provenant de tumeurs de patients et de données de ChIP-seq. Les expériences de scRNA-seq ont été faites par Alexandra Helleux et l'ensemble des autres expériences sur ce projet a été réalisé par Bujamin Vokshi.

Résultats

L'analyse des données de scRNA-seq m'a permis d'établir les profils transcriptomiques de 46 sous-populations cellulaires qui composent les cRCC, ces populations appartiennent aux grands groupes suivants : cellule cancéreuses, fibroblastiques, myéloïdes, lymphoïdes, endothéliales et tubules rénaux. J'ai analysé plus en détails les compartiments cancéreux et fibroblastique qui sont peu décrits dans la littérature. Les cellules de cancer se sont révélées caractérisées par 4 états qui s'inscrivent sur un gradient de transition épithelio-mésenchymateuse (TEM) : un état épithélial (ccRCC.epi) avec un programme proche des tubes proximaux, un état très dédifférencié (ccRCC.mes) avec un programme mésenchymateux, un état intermédiaire (ccRCC.int) avec un programme hybride épithélial et mésenchymateux et un état peu différencié enflammé (ccRCC.inf) avec un programme de réponse aux interférons et une expression du CMH de classe II. J'ai comparé les profils des cellules de cancer et des tubules rénaux issus des NATs, cette analyse a confirmé une proximité entre les cRCC et les tubes proximaux. Une analyse d'activité des facteurs de transcriptions m'a permis d'identifier notamment que les activités de TP53 et de PPARG sont perdues dans les cellules de cancer au profit d'un gain d'activité de c-MYC. De plus, j'ai pu identifier des facteurs perdus tels que PAX8 et des facteurs progressivement activés lors de la dédifférenciation des cellules de cancer. Ensuite, l'analyse du compartiment fibroblastique a révélé la présence de deux types de fibroblastes associés au cancer (CAFs) précédemment décrits dans d'autres cancers : les CAFs myofibroblastiques (myCAF) caractérisés par un programme de réponse à TGF-β et des gènes contractiles, et les CAFs de type enflammé (iCAFs) caractérisés par un programme de réponse à IL-1α. Les analyses de trajectoires indiquent que ces CAFs seraient dérivés des cellules mésangiales (péricytes du rein).

J'ai utilisé les données de scRNA-seq pour réaliser la déconvolution de 495 échantillons de RNA-seq issus du TCGA-KIRC afin d'étudier la composition tumorale et son lien avec la survie des patients. Ces analyses ont montré une forte corrélation entre les populations mésenchymateuses ccRCC.mes/myCAF qui anti-corrèlent avec la population ccRCC.epi. Une analyse de la composition des tumeurs par grade a révélé différents profils : les tumeurs de bas grade (G1-G2) sont enrichies en composantes épithéliales ccRCC.epi et vasculaires (endothélium et mésangium) alors que les tumeurs de grade plus avancés sont enrichies progressivement en infiltration immunitaire avec les tumeurs de plus haut grade (G4) présentant un fort enrichissement des populations dédifférenciées et mésenchymateuses ccRCC.inf, ccRCC.mes et myCAF. Une analyse similaire de composition des tumeurs en fonction des stades de la maladie m'a donné un résultat similaire. En accord avec ces profils, mes analyses de survie ont mis en évidence que les niveaux de présence de ccRCC.inf,

ccRCC.mes et myCAF sont fortement associés à une mauvaise survie des patients tandis que ccRCC.epi est associé à une bonne survie.

Afin d'étudier l'organisation spatiale des tumeurs, j'ai intégré les profils transcriptomiques définis par scRNA-seq avec des données de transcriptomique spatiale comprenant deux échantillons déjà séquencés en scRNA-seq et trois échantillons disponibles publiquement. Ces analyses ont révélé que les fibroblastes infiltrant les tumeurs délimitent des compartiments et forment une interface avec le tissu sain, de plus, les cellules de cancer mésenchymateuses ccRCC.mes sont fortement colocalisées avec les fibroblastes de type myCAF à cette interface. Notant cette proximité, j'ai étudié la communication entre ces deux types cellulaires par une analyse ligands-récepteurs. Ce résultat montre que ccRCC.mes peut induire le phénotype myCAF dans les cellules mésangiales et à leur tour, myCAF peut induire et maintenir la TEM dans les cellules de cancer à travers les voies TGF- β, AXL et NOTCH.

Ayant noté l'enrichissement des populations mésenchymateuses ccRCC.mes et myCAF dans les tumeurs avancées également enrichies en cellules immunitaires, j'ai exploré les interactions entre ces deux types cellulaires et les lymphocytes cytotoxiques T-CD8. Ces analyses m'ont permis de mettre en évidence différents mécanismes d'immunosuppression. Les cellules de cancer ccRCC.epi et ccRCC.int peuvent recruter des macrophages de type M2 via MIF et atténuer la réponse cytotoxique via SPP1 tandis que les cellules ccRCC.mes peuvent en plus stimuler le checkpoint immunitaire PD-1 via le ligand FAM3C, une interaction retrouvée également dans les myCAFs. J'ai aussi retrouvé une corrélation significative entre la proportion de ccRCC.mes ou de myCAF et l'expression de PD-L1 dans la cohorte BIONIKK, ce qui indique que ces deux populations peuvent potentiellement exprimer ce ligand.

Ensuite, j'ai effectué la déconvolution des échantillons de RNA-seq provenant de l'étude clinique BIONIKK pour étudier la composition des métastases, le lien entre composition et réponse à l'immunothérapie et pour dériver des biomarqueurs. Une analyse de la composition des métastases par rapport aux tumeurs primaire révèle que seules les populations ccRCC.mes et myCAF sont enrichies significativement dans les métastases établissant leur rôle important dans ce processus. Mon analyse de survie des patients traités par nivolumab (anti-PD1) seul et nivolumab+ipilimumab (anti-CTLA4) montre que trois populations stromales (myCAF, iCAF et macrophages) sont associées à la résistance à ces traitements mais myCAF est la population la plus fortement associée à une mauvaise survie sans progression et à une mauvaise survie globale. Basé sur ce résultat, j'ai dérivé une petite signature de 5 à 10 gènes marqueurs des myCAFs qui est significativement associée à une mauvaise survie globale dans BIONIKK, elle pourrait être mesurable par qPCR et potentiellement prédictive de la réponse au traitement.

Pour aller plus loin dans l'exploration des biomarqueurs, j'ai réalisé un clustering des échantillons de la cohorte BIONIKK en fonction de l'expression des longs ARN non-codants intergéniques (lincRNA). Cette classification révèle 5 groupes (LINC1 à LINC5) dont un groupe (LINC5) est significativement associé à une mauvaise survie. En accord avec les résultats précédents, ce groupe LINC5 est composé de tumeurs mésenchymateuses enrichies pour les marqueurs de fibroblastes. J'ai croisé les gènes marqueurs de ce groupe de tumeurs avec les profils d'expression issus du scRNA-seq afin d'identifier des potentiels biomarqueurs. Ceci m'a permis d'isoler plusieurs gènes d'intérêt : SFTA1P exprimé dans les populations fibroblastiques ainsi que LINC01638 et LINC01615 tous les deux spécifiques des cellules de

cancer. J'ai retrouvé qu'une expression haute du gène SFTA1P est significativement associée à une mauvaise survie des patients traités par immunothérapie, ce gène pourrait constituer un biomarqueur de la présence de CAF prédisant la réponse à ce type de traitement. De manière intéressante, le gène LINC01615 est très spécifique de la population ccRCC.mes. J'ai donc analysé l'expression de ce lincRNA dans d'autres jeux de données. Dans le TCGA-KIRC, son expression est significativement enrichie dans les tumeurs par rapport aux tissus sains et une expression haute est très fortement associée à une mauvaise survie des patients. Comme ce lincRNA est décrit comme associé à un programme invasif dans les cancers du foie et de l'estomac, j'ai également vérifié son expression dans le mélanome qui est un cancer dont les transitions phénotypiques sont bien connues. Dans ce cancer, j'ai bien retrouvé une expression de LINC01615 qui est associée à l'état mésenchymateux dans les données publiques du TCGA-SKCM (« Skin Cutaneous Melanoma ») et dans les données de RNA-seq générées par nôtre équipe sur les lignées cellulaires de mélanome. La detection de LINC01615 dans les lignées cellulaire de mélanome a pu être confirmée par RTqPCR qui montre un enrichissement conséquent du gène dans la lignée mésenchymateuse MM099 par rapport à la lignée mélanocytique 501M.

Dans un deuxième projet, j'ai analysé les données de CMR. Le scRNA-seq a permis de mettre en lumière une probable origine du cancer à partir de la branche épaisse ascendante de l'anse de Henlé (TAL) et d'identifier différents sous-types de cellule de cancer allant d'épithélial à mésenchymateux. De manière similaire aux cRCC, j'ai analysé les facteurs de transcriptions intervenant lors de la transformation et dédifférenciation. Un des facteurs de transcription clé est le suppresseur de tumeur TFCP2L1 dont l'activité est perdue lors de la transformation au profit d'un gain de NFE2L2 qui possède une activité anti-ferroptose. Les analyses de lignées cellulaires ré-exprimant SMARCB1 ont permis de mettre en évidence que ce dernier antagonise la fixation génomique du complexe MYC/MAX. De plus, l'analyse du microenvironnement indique que ces tumeurs sont prépondérantes en cellules immunosuppressives comme les CAFs et les macrophages associés au cancer.

Conclusions et perspectives

Dans ce travail, j'ai établi une caractérisation détaillée de la composition des cRCC qui permet de décrire pour la première fois la plasticité des cellules cancéreuses et le compartiment fibroblastique dans ce type de tumeur. Les cellules cancéreuses montrent différents états de différenciation qui définissent un gradient de TEM et un état enflammé analogue à ce qui a été décrit dans le mélanome. Il m'a été possible de retracer les étapes de l'oncogenèse depuis les tubes proximaux jusqu'à l'état le plus indifférencié ccRCC.mes et d'identifier les facteurs de transcriptions actifs à chaque étape. Lors de la transformation, des facteurs de maintien de l'identité épithéliale proximale dont TP53, PPARG et FOXP4 sont perdus au profit d'un gain d'activité de c-MYC, CEBPD et ATF4. Au cours de la dédifférenciation des cRCC, l'expression de PAX8 est progressivement perdue avec gain d'activité pour les facteurs des familles POU, SOX, ZEB et SNAI. En accord avec ce qui est publié dans d'autres cancers, l'état mésenchymateux est associé à une mauvaise survie des patients et ccRCC.mes est la seule population cancéreuse enrichie dans les métastases à distance, ce qui établit son rôle clé dans l'invasion et la dissémination métastatique. L'analyse du compartiment fibroblastique révèle

que les péricytes rénaux peuvent se convertir en myCAFs et iCAFs, deux phénotypes décrits par ailleurs dans les tumeurs pancréatiques.

L'analyse par déconvolution du TCGA-KIRC et de BIONIKK à partir des résultats de scRNA-seq m'a permis d'établir des profils de composition des tumeurs qui varient au cours de l'évolution de la maladie : les tumeurs de bas grade sont plus différenciées et vascularisées alors que les tumeurs de haut grade sont plus infiltrées par les cellules immunitaires et enrichies en composantes mésenchymateuses. On retrouve une forte association entre ccRCC.mes et myCAF qui est confirmée par les données de transcriptomique spatiale. Ces deux populations co-localisent principalement à l'interface tumeur/tissu-sain et sont capables de communiquer. Par la prédiction des interactions ligands-récepteurs, j'ai identifié des voies de communication qui indiquent que ccRCC.mes/myCAF peuvent soutenir leurs phénotypes respectifs et qui constituent des cibles thérapeutiques potentielles. J'ai également décrit des mécanismes d'immunosuppression dont le ligand FAM3C capable de stimuler le checkpoint immunitaire PD-1. Au sein des cellules de cancer, FAM3C est principalement exprimé par les cellules ccRCC.mes et constitue donc une potentielle nouvelle cible thérapeutique.

L'analyse de la cohorte BIONIKK m'a permis de mettre en lumière une corrélation significative entre une proportion élevée de myCAF et la résistance à l'immunothérapie. J'ai pu dériver une petite signature basée sur les gènes marqueurs de myCAFs qui serait utilisable en pratique clinique pour potentiellement prédire la réponse aux inhibiteurs de checkpoints immunitaires. Enfin, par le clustering des patients de cette cohorte basé sur les lincRNA, j'ai pu montrer que des marqueurs épigénétique peuvent refléter la composition tumorale de manière suffisamment précise pour être utilisés comme biomarqueurs de survie et de réponse au traitement. Notamment, le gène SFTA1P est exprimé dans les populations fibroblastiques et son expression est associée à une mauvaise survie des patients traités par immunothérapie, ce gène constitue un potentiel biomarqueur pour prédire la réponse à ce type de traitement. Aussi, le gène LINC01615 est un marqueur de ccRCC.mes dont l'expression est fortement associée à une mauvaise survie dans le TCGA-KIRC. Ce gène est aussi retrouvé dans les lignées de mélanome mésenchymateuses et *in-vivo* dans les tumeurs les plus dédifférenciées de la cohorte du TCGA-SKCM. LINC01615 est aussi rapporté comme étant associé à la TEM dans d'autres cancers. Il constitue donc un biomarqueur et une potentielle cible thérapeutique très intéressante avec peut-être des applications pan-cancer.

L'analyse des CMR a permis d'étendre le modèle d'oncogenèse de ce type de carcinome. Le cancer surviendrait des cellules TAL normalement sensibles à la ferroptose, la transformation implique une perte d'activité de TFCP2L1 conséquente à la perte de SMARCB1 avec un gain de c-MYC et NFE2L2 activant ainsi un programme anti-ferroptotique. On constate des parallèles importants entre la transformation des CMR et des cRCC dans lesquels on observe aussi une perte d'activité de facteur épithéliaux tel que PAX8 accompagnée d'une perte d'un membre de SWI/SNF, VHL/PBRM1 étant la combinaison de mutation la plus fréquente dans ces tumeurs. Cette perte de facteurs de maintien d'identité épithéliale aboutit également sur une activation oncogénique de c-MYC. En se basant sur les résultats dans le cRCC, la présence de CAFs dans le microenvironnement tumoral des CMR pourrait indiquer que ces tumeurs seraient peu propices à l'utilisation de l'immunothérapie sans contre-mesures dirigées aussi contre ce type de cellule.

Dans l'ensemble, ce travail permet de mieux comprendre la biologie des cRCC et des CMR. Il ouvre de nouvelles voies pour comprendre l'oncogenèse de ces deux types de cancer, pour déchiffrer les mécanismes de dédifférenciation menant à la progression de la maladie, pour mieux personnaliser le traitement des patients et pour développer de nouvelles solutions thérapeutiques.

Chapter I: General introduction

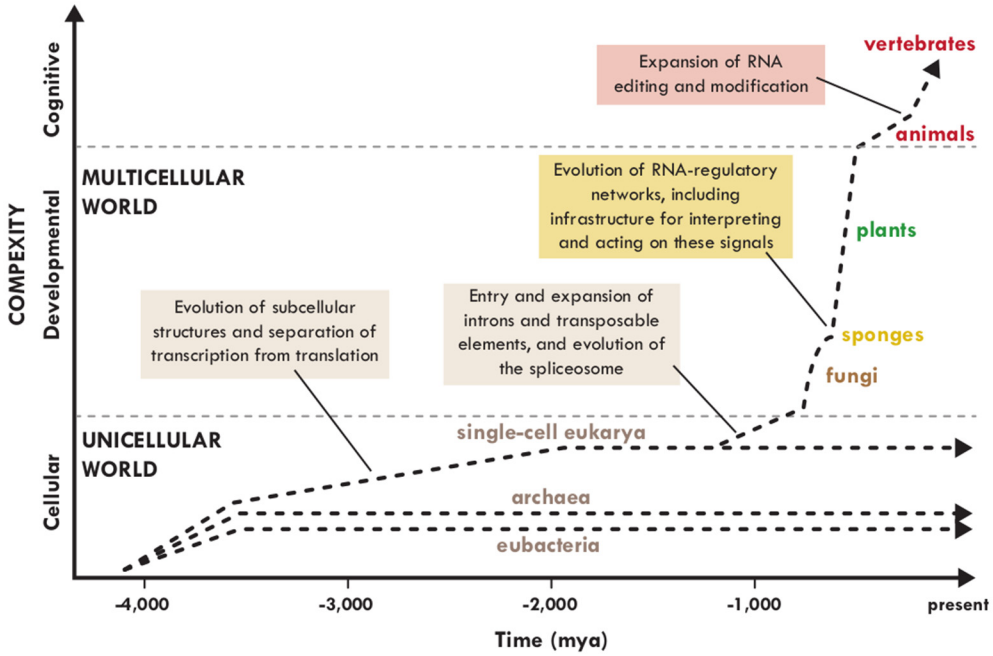


Figure 1. The evolution of complexity

From Mattick et al. (2022)

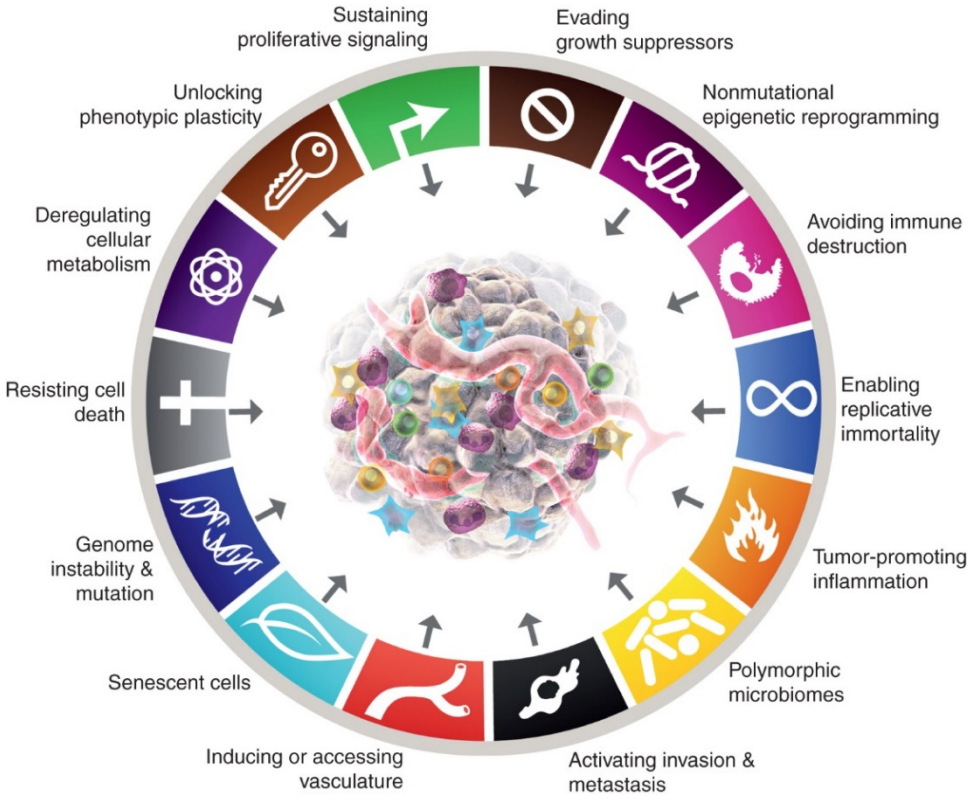


Figure 2. The hallmarks of cancer

From Hanahan et al. (2022)

Chapter I: General introduction

1. Cancer

A. Origins and hallmarks of cancer

Life is thought to have emerged as RNA molecules which eventually evolved into mono-cellular organisms with eukaryotes estimated to have appeared two billion years ago and then into multi-cellular oceanic plants dating back 1.2 billion years[1] (Figure 1). The shift into complex multi-cellular organisms was achieved through numerous gene regulation mechanisms that allow different cells to acquire various specialized sets of features and functions (called "phenotypes") from the same genetic information. Unfortunately, one problem could not be completely solved by evolution: some cells can undergo changes to their DNA allowing them to bypass normal control mechanisms in order to grow and divide endlessly at the expense of the host organism. Evidence show that these cancer cells have plagued multi-cellular organism since their appearance with osteosarcoma being found on a 1.7 million year old humanoid fossil[2] and even on a 77 million year old dinosaur fossil[3]. In current times, cancer has become one of the leading causes of death worldwide[4].

Cancer arises through alterations in key genes which can be defined into two types: those that promote and drive cancer through their activation are called "oncogenes" while protective genes that facilitate cancer through their inactivation are deemed "tumor suppressors". Well studied oncogenes include *SRC* that is highly similar to the *v-src* version carried by the Rous Sarcoma Virus that can induce sarcoma in birds and *MYC* (c-MYC) that plays a role in multiple cancers but was initially studied in Burkitt's lymphoma. On the other hand, some well-known tumor suppressors include retinoblastoma transcriptional corepressor 1 (*RB1*) and tumor protein p53 (*TP53*) both encoding proteins with transcription factor activity. Mutations conferring a cell selective advantages that lead to clonal expansion therefore promoting the development of cancer are called "driver mutations". One such famous mutation is the *BCR-ABL1* gene fusion which is responsible for the development of acute myeloid leukemia and the targeting of its resulting fusion protein has drastically improved patient prognosis for this type of cancer[1,5].

Cancer is a very heterogeneous disease, there are as many cancers as there are cell types in the human body which makes it very challenging to tackle but it is possible to find some common principles. The biological properties of cancer cells were initially described into six hallmarks: enable replicative immortality, sustain proliferation, resist cell death, avoid growth suppression, induce vascularization and activate invasion. These original hallmarks were later extended into fourteen accounting for the latest research into tumor microenvironment (TME) and epigenetic mechanisms[6] (Figure 2). In this introduction we will expand on three of these hallmarks: metastasis, phenotypic plasticity and non-genetic reprogramming as well as introduce the major actors of the TME and how they influence these hallmarks.

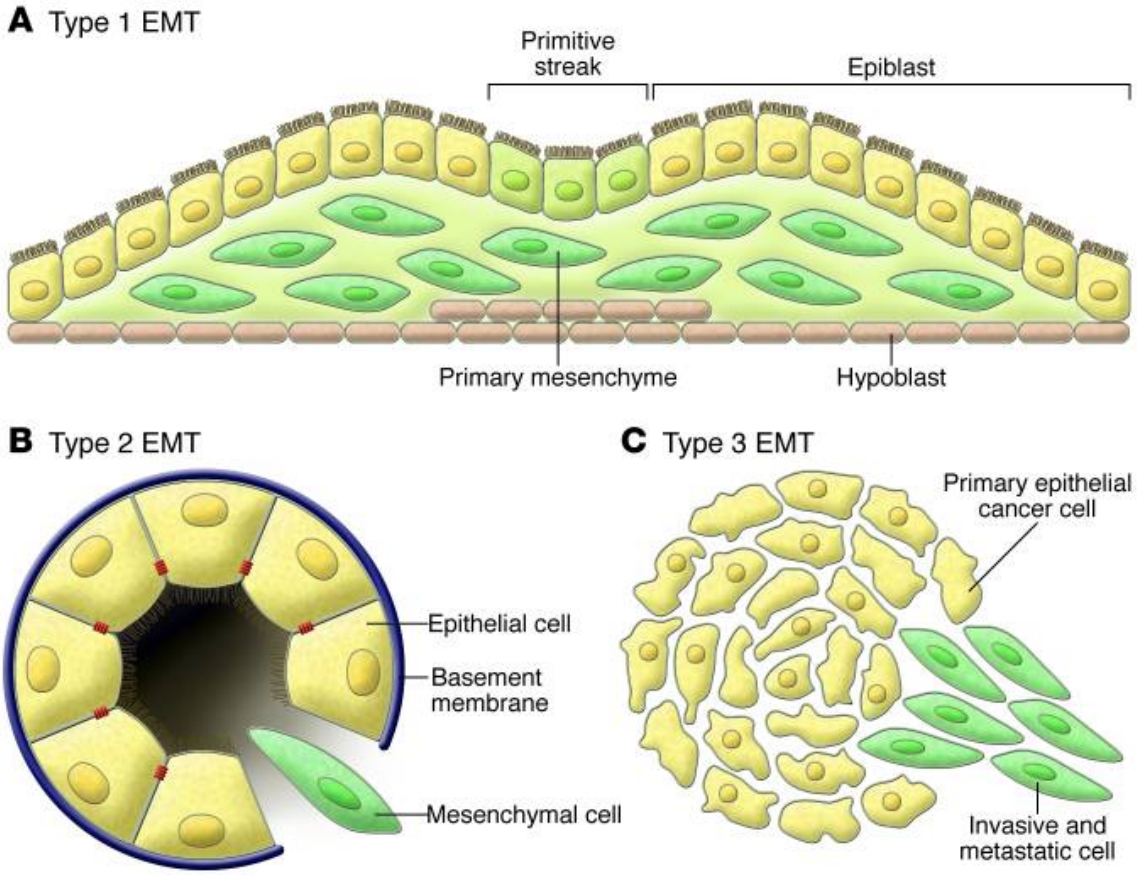


Figure 3. The three classes of EMT.

From Kalluri et al. (2009)

B. Invasion and metastasis: the epithelial to mesenchymal transition

The leading cause of death from cancer is the ability of cancer cells to invade and colonize distant organs. As mentioned above, every cell possesses essentially the same genetic information giving cancer cells the possibility to tap into a variety of different programs. Tumor progression and metastasis can be achieved mainly by hijacking a developmental and wound-healing process called the epithelial to mesenchymal transition (EMT) and the reverse mesenchymal to epithelial transition (MET)[7,8]. During EMT, packed polarized epithelial cells will transition into more motile mesenchymal cells. The process is governed in an epigenetic manner and is fully reversible. It has been proposed to categorize EMT into three types[9] which will be described in this section (Figure 3).

a. Type 1 EMT

Type 1 EMT occurs during implantation, embryo formation and organ development and does not cause fibrosis or systemic spread of mesenchymal cells. Three weeks into human development, the embryo will reorganize itself into a tri-layered structure called "gastrula" that will lay the foundations for subsequent organogenesis[10]. The first sign of gastrulation is the generation of the "primitive streak". This tissue is formed by epithelial cells expressing E-cadherin (*CDH1*) and will generate the mesendoderm which will subsequently form the mesoderm and endoderm using type 1 EMT. This process is regulated by canonical wingless-related integration site (Wnt) signaling and involves proteins of the transforming growth factor beta (TGF-β) superfamily as well as transcription factors such as Snail, Eomes and Mesps that will repress expression of *CDH1*. Other embryonic development steps require type 1 EMT, notably the generation of migratory neural crest cells and formation of the heart valve.

b. Type 2 EMT

Type 2 EMT occurs during wound healing, tissue regeneration and organ fibrosis. The goal is to generate new fibroblastic cells in order to repair damaged tissues. Here the process is mediated by inflammatory queues and is supposed to stop once inflammation subsides but chronic damage or inflammation will cause a harmful condition called organ fibrosis. This condition leads to conversion of epithelial cells into fibroblasts which hinders organ functionalities and tends to occur notably in kidney, lung, liver and intestine[11]. This dysfunction of EMT is mainly mediated by fibroblasts which release extra-cellular matrix (ECM) components as well as inflammatory signals. In kidney fibrosis, a process well studied in mice, it was estimated that about one third of fibroblasts are derived from epithelial kidney tubule cells that underwent type 2 EMT[12]. The process is mediated via growth factors such as TGF-β, platelet derived growth factor (PDGF), epidermal growth factor (EGF) and fibroblast growth factors (FGF) but also by chemokines and matrix metalloproteases (MMPs). There is a particular importance of TGF-β in this process as using an antagonist allowed reversal of fibrosis with repopulation of the organ by epithelial cells and restoration of kidney functions.

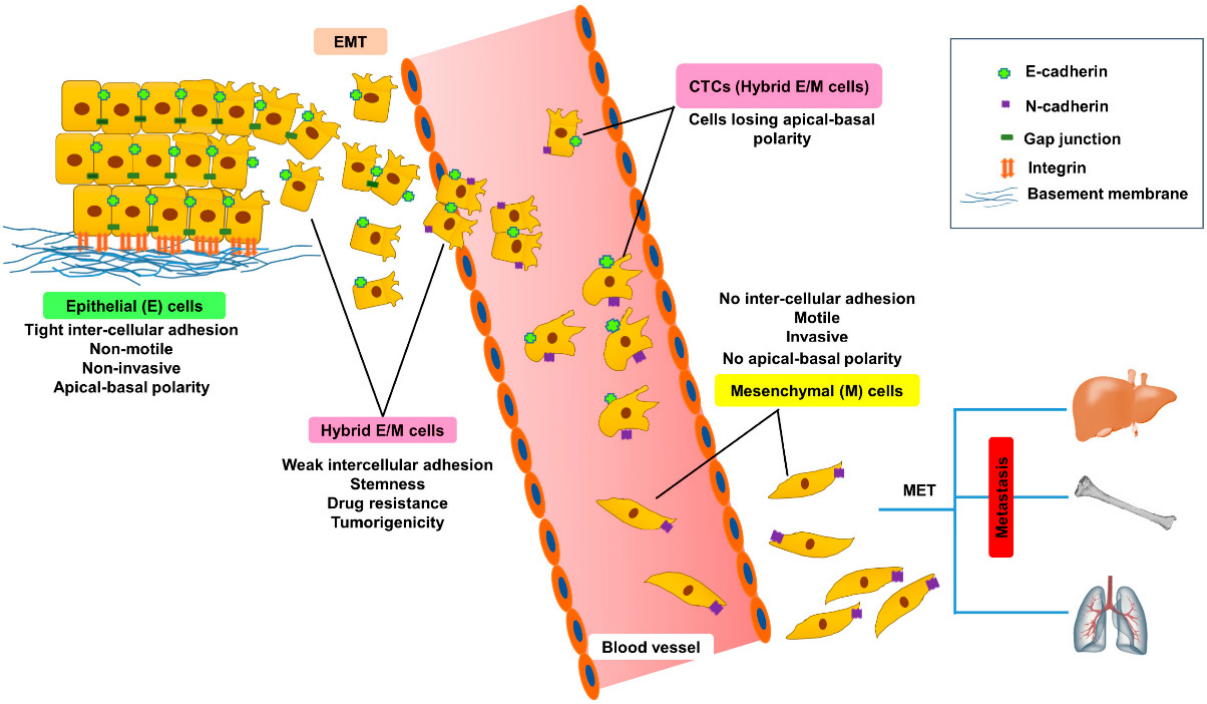


Figure 5. Type 3 EMT used by cancer cells.

From Sinha et al. (2020)

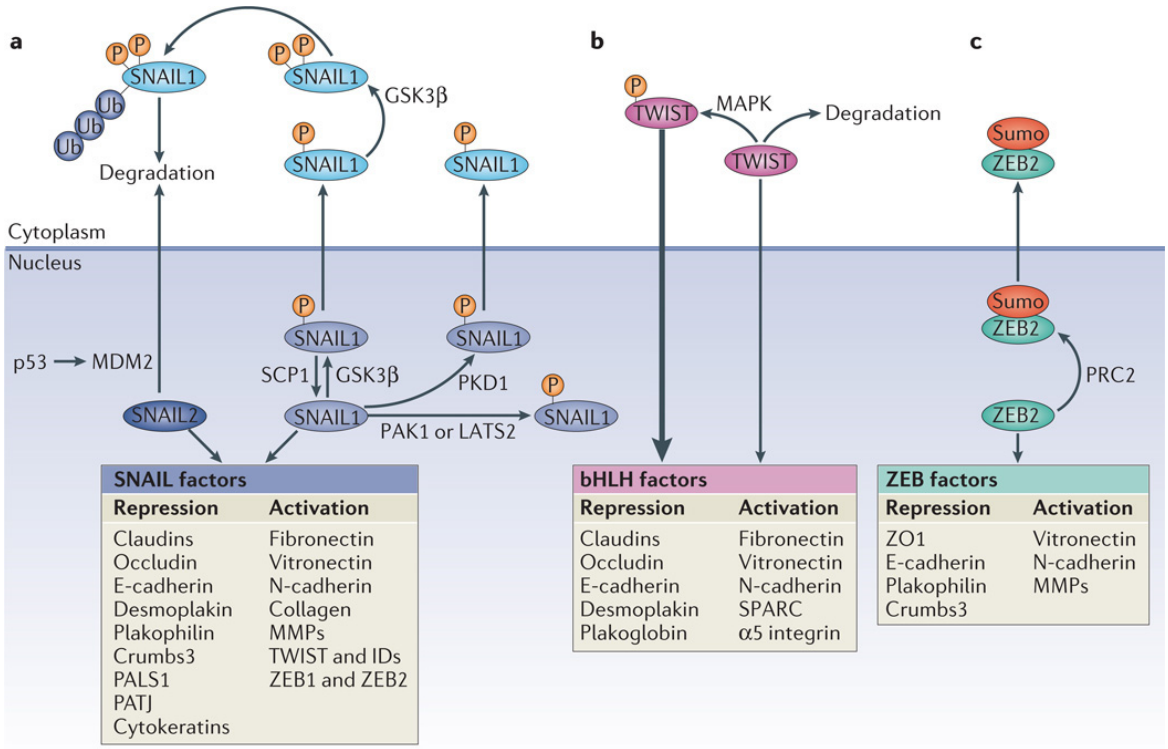


Figure 6. Transcriptional control of EMT.

From Lamouille et al. (2014)

c. Type 3 EMT

Type 3 EMT occurs in epithelial cancer cells, contrary to the other two types which are tightly regulated processes, type 3 EMT is uncontrolled and grants cancer cells a diverse set of properties including increased invasiveness, immune-escape mechanisms, stemness and resistance to treatments[13]. In cancer cells, we do not view EMT as a direct jump from epithelial to mesenchymal state but as a continuum with cells progressively losing features from one state while acquiring features from the other. From this spectrum, we generally define three main states: epithelial, intermediate and mesenchymal. The epithelial state is the most differentiated state and its properties depend on the cell of origin so they can be very varied for different cancer types. The intermediate state also labelled partial or hybrid-EMT (hEMT), as its name suggests, has both mesenchymal and epithelial properties. Lastly, in the mesenchymal state cells will have mostly lost their identity and acquired varied mesenchymal properties, some akin to fibroblasts[14] (Figure 5). We will discuss the molecular bases of type 3 EMT and describe some of the key players below.

i. Molecular changes

During EMT cells will lose their epithelial characteristics including junctions, apical-basal polarity and the cytoskeleton organization shaping their morphology. One crucial gene to be down-regulated is *CDH1*, a member of the cadherin superfamily responsible for cell-cell adhesion in a calcium dependent manner[15]. As tight-junctions need to be deconstructed, members from the claudin and occludin classes of protein as well as the scaffolding protein that links them encoded by *TJP1* are suppressed. Component of desmosomes and gap junctions, desmoplakin encoded by *DSP* and the connexin family of proteins are also repressed[16]. Of note, the cytoskeleton is affected by alteration of cytokeratin expression which can direct E-cadherin towards the plasma membrane, some key members are *KRT8* and *KRT18*[17].

Cells will also gain mesenchymal characteristics with directional motility and an elongated morphology. As *CDH1* is repressed, there is a switch towards expression of neural cadherin *CDH2* which gives them an affinity towards mesenchymal cells enabling easier migration[18]. Signaling though receptor tyrosine kinases (RTKs) such as fibroblast growth receptors (FGFRs) and the SRC family of proteins is facilitated through *CDH2* and one of its interactor: neural cell adhesion molecule 1 (*NCAM1*)[19]. The cell-surface glycoprotein encoded by *CD44* also plays a major role in mesenchymal cell adhesion and migration[20]. One key member of the cytoskeleton to be induced is vimentin (*VIM*) which allows the change of shape enabling increased motility[21]. ECM modulation is essential for EMT as it acts as a substrate for cell adhesion and as a reserve of signaling factors. One class of protein involved in ECM degradation are MMPs with representatives such as *MMP2* and *MMP9* facilitating invasion[22]. In addition, mesenchymal cells will also secret ECM component molecules like collagens such as *COL1A1* that has been linked to metastasis[23], secreted protein and rich in cysteine (*SPARC*)[24] and fibronectin 1 (*FN1*) which is a glycoprotein involved in cell migration[25] (Figure 6).

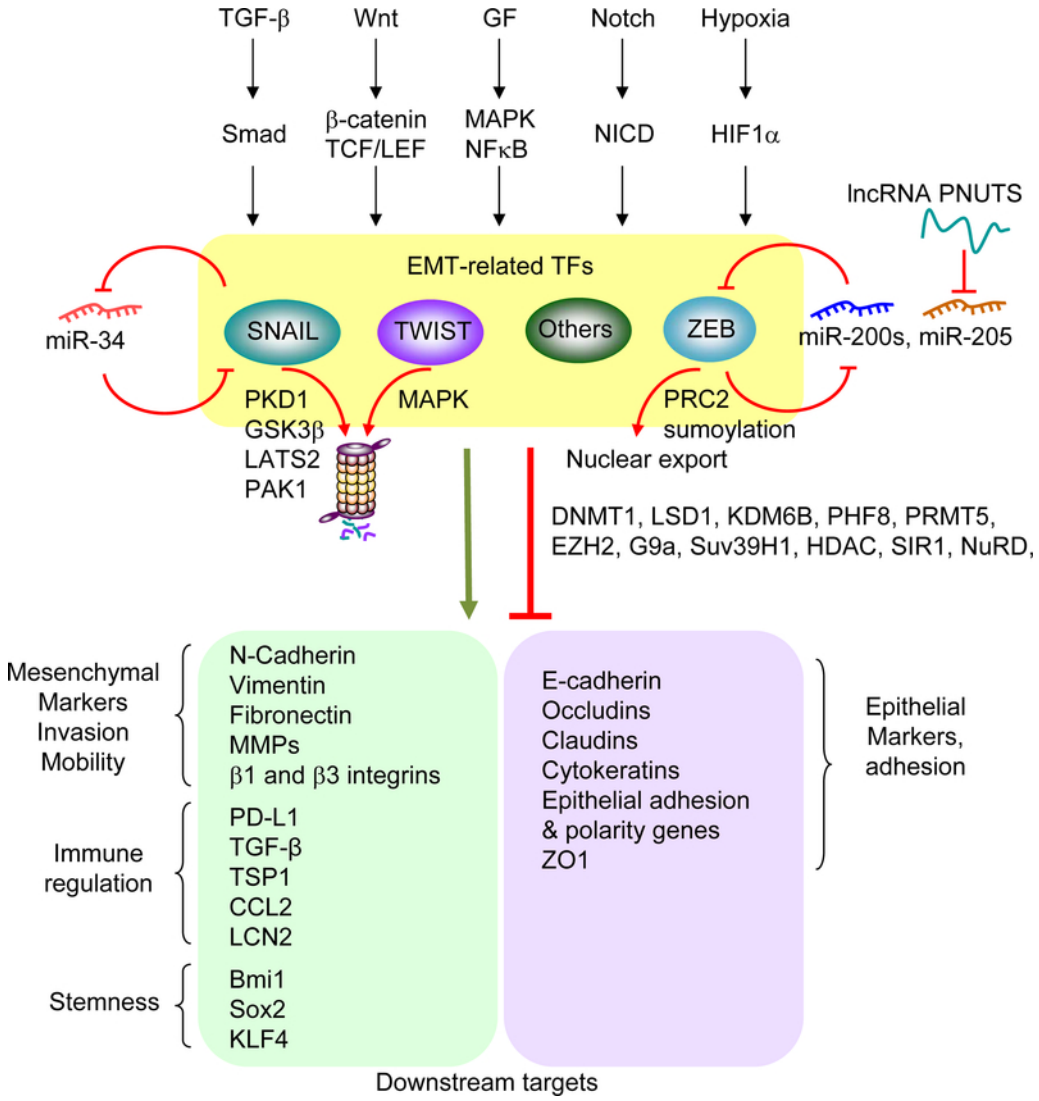


Figure 7. Regulatory networks of EMT.

From Lu et al. (2019)

ii. Transcriptional regulation of EMT

The molecular changes described above are orchestrated by a large variety of transcription factors. Among the best studied are SNAIL (*SNAI1*) and SLUG (*SNAI2*), part of a transcription factor family of zinc finger repressors. These factors will recruit multiple chromatin remodeling factors: the polycomb repressive complex 2 (PRC2) with methyltransferase activity through enhancer of zeste 2 polycomb repressive complex 2 subunit (*EZH2*), the co-repressor encoded by *SIN3A* and the histone demethylase lysine demethylase 1A (LSD1A/*KDM1A*) which acts with the repressor element-1 silencing transcription factor (REST/CoREST) complex[26]. These factors will deposit repressive methylation marks on histones H3K9 and H3K27 in epithelial genes, in particular silencing *CDH1* and *KRT8* expression[27]. In addition they will also deposit activating histone marks such as methylation of H3K4 and acetylation of H3K9 to induce expression of mesenchymal genes like *FN1* and *CDH2*[16]. The expression of these transcriptions factors have been linked to metastasis, notably in breast cancer[28].

Another well characterized pair of transcription factors promoting EMT are encoded by *TWIST1* and *TWIST2*. They belong to the basic helix-loop-helix (bHLH) transcription factor class that binds E-box elements and governs lineage specification and differentiation. These factors will recruit a methyltransferase encoded in *KMT5A*[29] and induce recruitment of PRC2[30] to modulate gene expression. In head and neck cancer cells for example, these factors will suppress expression of *CDH1* and of the cyclin dependent kinase inhibitor 2A (*CDKN2A*) tumor suppressor gene while promoting the expression of *CDH2*[31]. These factors are also associated with metastasis in breast cancer and squamous cell carcinoma[32].

The ZEB family of zinc-finger transcription factors contains two EMT regulators encoded in *ZEB1* and *ZEB2*. These can bind E-box elements to directly repress expression of *CDH1*, tight-junction and cell polarity complex genes while inducing expression of *VIM* and *CDH2*. These factors will once again make use of chromatin remodeling complexes to mediate their action, notably they can recruit the Switch/sucrose non-fermentable complex (SWI/SNF) ATPase encoded by brahma-related gene 1 (BRG1/*SMARCA4*) and the demethylase encoded by *KDM1A*[16]. Though ZEB factors have been shown to be indispensable for metastasis in pancreatic cancer[33], it should be noted that their action are context dependent as it has been demonstrated that the two genes *ZEB1* and *ZEB2* can have opposite effects in melanoma[34].

Although the three classes of transcription factors described above are the main actors promoting type 3 EMT, others should be mentioned. The TEAD family transcription factors normally involved in muscle development[35] promote EMT in cancer cells. They are activated downstream of Hippo signaling through Yes associated protein 1 (YAP) and Taffazin (TAZ) and help cancer progression[36]. Another muscle development regulator, the paired family of homeobox member encoded in *PRRX1* is capable of promoting EMT in cancer cells[37,38]. The SRY-related HMG-box (SOX) family of transcription factors can play a similar role in cancer development and progression but the effect of these transcription factors are very context dependent[39]. Lastly, EMT inducing roles have been described for forkhead box (FOX) and GATA-binding (GATA) transcription factors[40,41] (Figure 7).

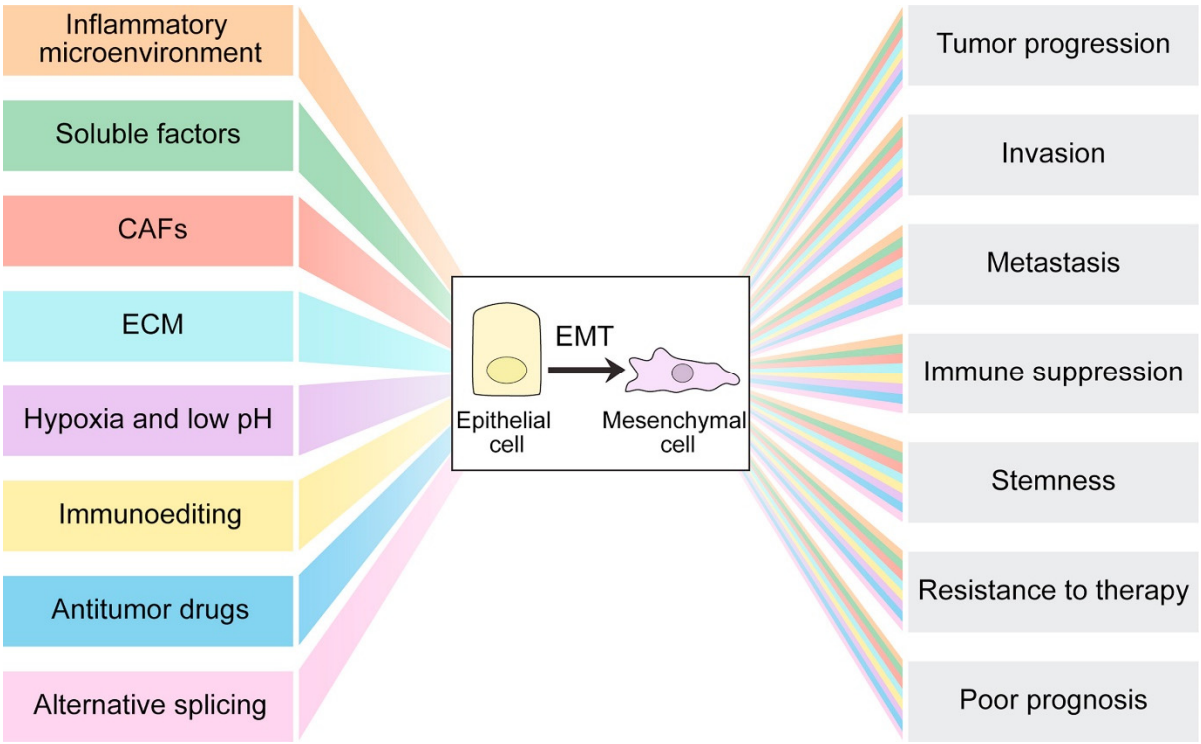


Figure 8. Stimuli inducing EMT.

From Sistigu et al. (2017)

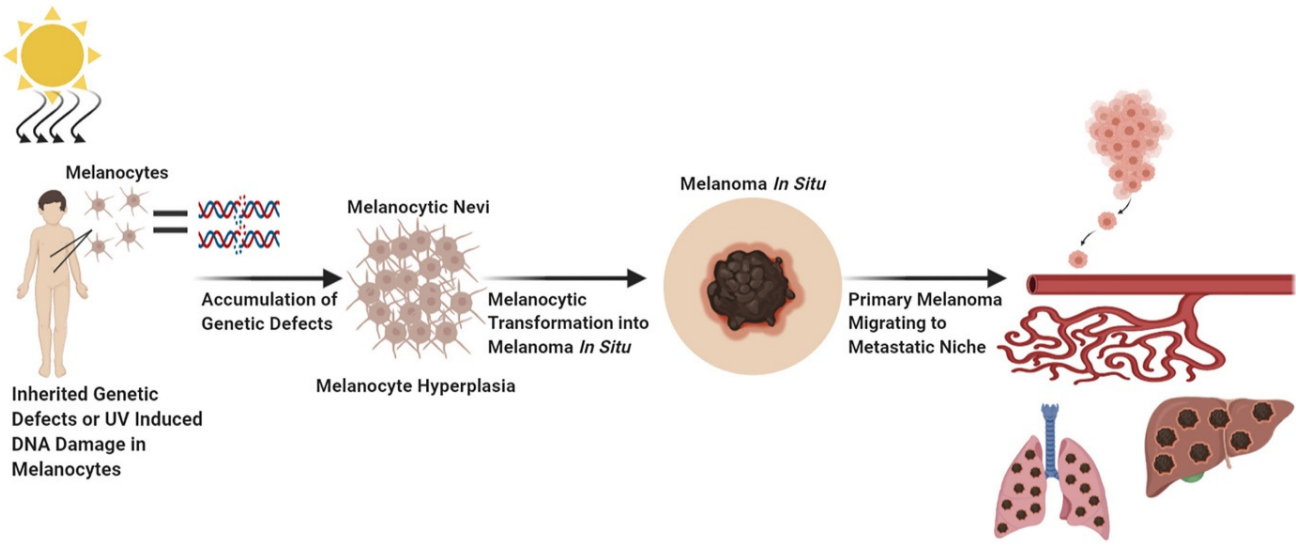


Figure 9. The genesis of melanoma.

From Eddy et al. (2020)

iii. Signals inducing EMT

Type 3 EMT can be induced by a large variety of extracellular signals, some are produced mainly by the TME including several soluble inflammation cytokine[13] (Figure 8). One of the most studied is TGF-β (*TGFB1*), a multi-functional cytokine capable of binding *TGFBR1* and *TGFBR2* encoded cell surface RTKs. Stimulation of these receptors will lead to activation of the small and mothers against dpp (SMAD) family of genes to induce expression of SNAIL, ZEB and TWIST transcription factors. TGF-β can also mediate EMT in a SMAD-independent manner through the mitogen-activated protein kinase also named "extracellular signal-regulated kinase" (MAPK/ERK) signaling pathway and through crosstalk with the Notch, Wnt/β-catenin and integrin pathways[42]. The interleukin-6 (IL-6/*IL6*) is a key pro-tumoral cytokine that is able to promote expression of SNAIL and TWIST, notably through the Janus kinase/signal-transducer and activator of transcription (JAK/STAT) pathway[43]. Interleukin-11 (IL-11/*IL11*) can have similar effects as IL-6[44]. Tumor necrosis factor alpha (TNF-α) is a cytokine acting through nuclear-factor-kappa-B (NF-κB) signaling and capable of stabilizing SNAIL and β-catenin[45]. Interferon-γ (*IFNG*) is a cell-killing molecule secreted by cytotoxic cells during anti-tumor response but it can also induce EMT in exposed cancer cells[46].

In addition to inflammatory molecules, multiple signals can trigger EMT. Mechanical stress disrupts ECM-cytoskeleton interactions and can induce cell dissemination[47]. For example, matrix stiffness and ECM components like *FN1* and *MMP3* are able to promote expression of SNAIL and TWIST[48,49,50]. Additionally, treatments like chemotherapeutic agent doxorubicin can induce or select full-EMT cancer cell[51,52]. Lastly, hypoxia is another important trigger[13] that will be discussed in a later section on renal cancer.

C. Cell plasticity and phenotype switching: the example of melanoma

Cell plasticity refers to the ability of cancer cells to dynamically change their phenotypic characteristics, one example of this is the epithelial-mesenchymal plasticity (EMP) described in the section above. This phenomenon is well studied in melanoma. In this section, we will briefly introduce this type of cancer and overview the impact of phenotypic plasticity.

a. Oncogenesis and classification of melanoma

Melanoma is an aggressive skin cancer that has been on the rise over the past decades[53]. It originates from melanocytes, a type of pigment producing cell derived from the neural crest located in the skin (cutaneous melanoma), eye (uveal melanoma) and different mucosae (mucosal melanoma)[54]. The most common melanoma is the skin cutaneous melanoma (SKCM) found mainly in sun-exposed spots (Figure 9). This type can be classified into multiple subtypes based on its histological features: superficial spreading melanoma (SSM) accounting for 70% of cases, nodular melanoma (NMM), lentigo melanoma (LLM), acral lentiginous melanoma (ALM) and desmoplastic melanoma (DM)[55].

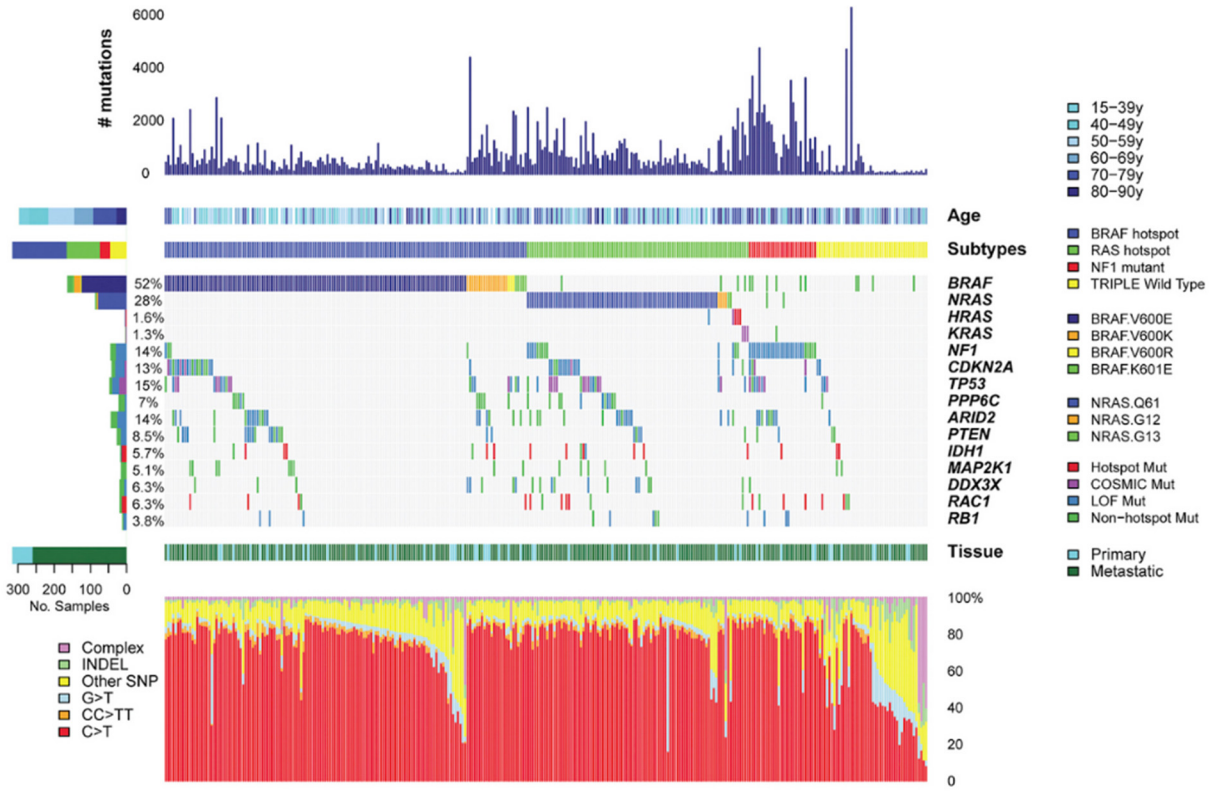


Figure 10. Genomic subtypes of melanoma.

From Akbani et al (2015)

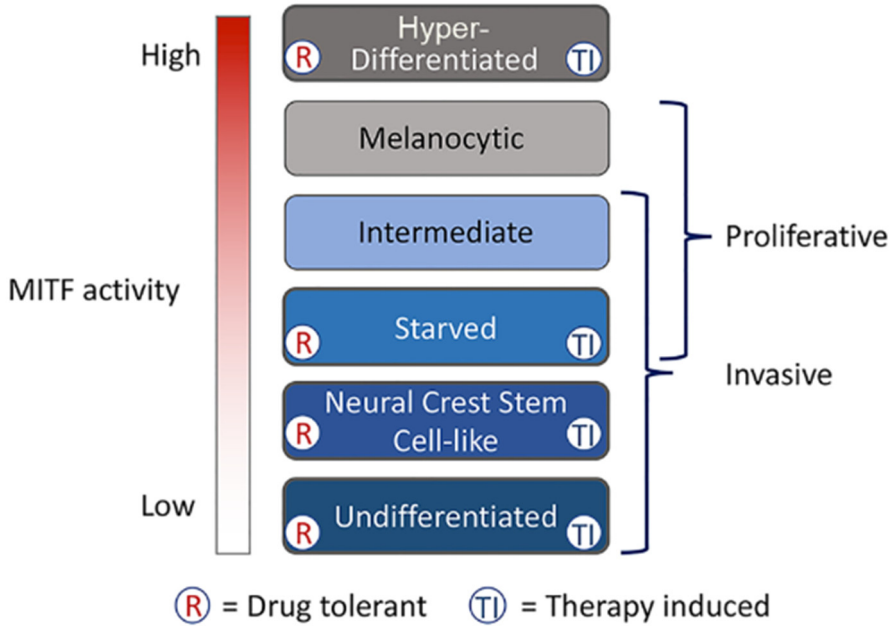


Figure 11. The MITF rheostat model

From Rambow et al. (2019)

At the genetic level multiple recurrent driver mutations can be found thus defining four genomic melanoma subtypes[56] (Figure 10). All these mutations affect the MAPK/ERK pathway which can normally be activated by growth factors and controls numerous processes including cell survival and proliferation[57].

One the most frequently mutated gene is the serine/threonine kinase gene *BRAF* which commonly harbors the V600E mutation favored by ultraviolet (UV) radiations[58]. Gain of function (GoF) mutations in this gene define the *BRAF* subtype in which the protein is constitutively active leading to phosphorylation of MAP2K1 and continuous activation of MAPK signaling.

The rat sarcoma virus gene family *(NRAS*, *KRAS* and *HRAS*) is altered in melanoma and constitutes a second melanoma subtype. RAS proteins have intrinsic GTPase activity and GoF mutations in these genes will lead MAPK/ERK activation through translocation of RAF proteins to the plasma membrane and phosphorylation of MEK1/2 which in turn phosphorylate ERK1/2[59].

The third melanoma subtype is defined by alteration in the neurofibromin 1 (*NF1*) gene. NF1 is another protein with GTPase activity which can inactivate RAS proteins thus suppressing the MAPK/ERK pathway. Consequently, loss of function (LoF) mutations in *NF1* will lead to activation of this pathway.

The last subtype is the heterogeneous triple wild-type (triple-WT) group defined by absence of the three driver mutations presented above. Additionally, melanoma is frequently mutated in previously mentioned tumor suppressor genes *TP53*, *RB1*, *CDKN2A* and in phosphatase and tensin homolog (*PTEN*)[56]. As more genomic data are gathered, recurrent mutations will allow the definition of new subtypes[60].

Because of this genetic landscape, targeted therapies include BRAF inhibitors vemurafenib and dabrafenib as well as MEK inhibitor trametinib. Despite successful results in combination with immunotherapy, tumors can eventually adapt and acquire resistance to treatments[61].

b. Dynamic phenotype switching of melanoma cells

Beyond genetic mutations, melanoma cells display EMP. As melanocytes are not true epithelial cells, in melanoma we refer to the epithelial state as “epithelial-like”, to the mesenchymal state as “mesenchymal-like” and to the transition process as “pseudo-EMT”. Despite this nomenclature, the process is very similar to EMT which is not surprising since the neural crest cells from which melanocytes are derived use EMT after gastrulation as described in the type 1 EMT section[62]. One of the critical gene involved in this plasticity is the melanocyte inducing transcription factors (*MITF*, originally called microphthalmia-associated transcription factor). MITF is a bHLH transcription factor essential for melanocyte development that controls melanogenesis through the specific MITF-M isoform[63]. In melanoma, *MITF* is highly expressed in the epithelial-like state and its activity level will gradually go down along dedifferentiation[64,65]. This dynamic switch of MITF activity led to the “MITF rheostat model” where the level of MITF determines melanoma cell properties[61] (Figure 11).

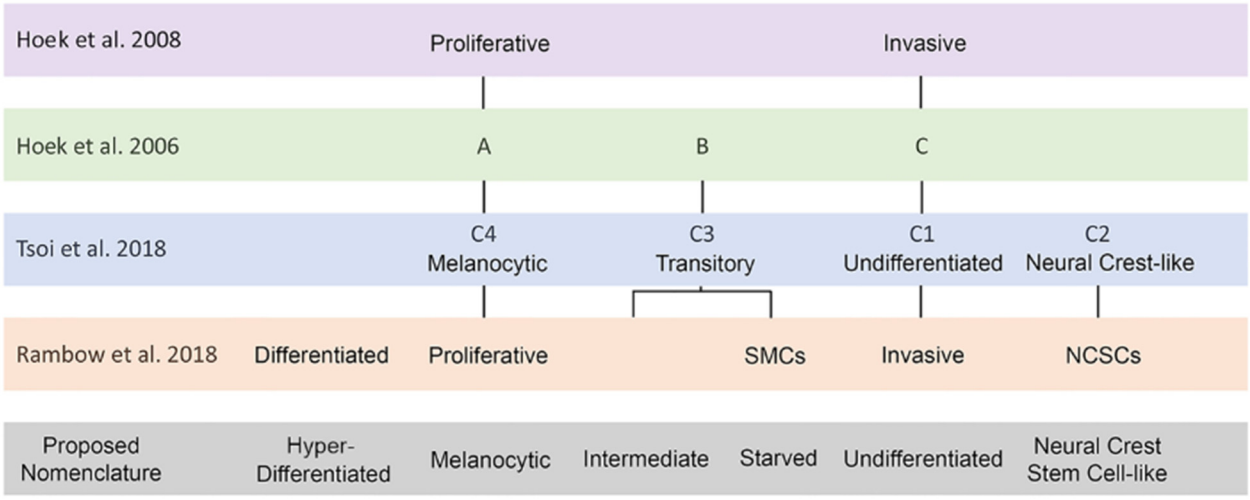


Figure 12. Phenotypic diversity of melanoma cells.

From Rambow et al. (2019)

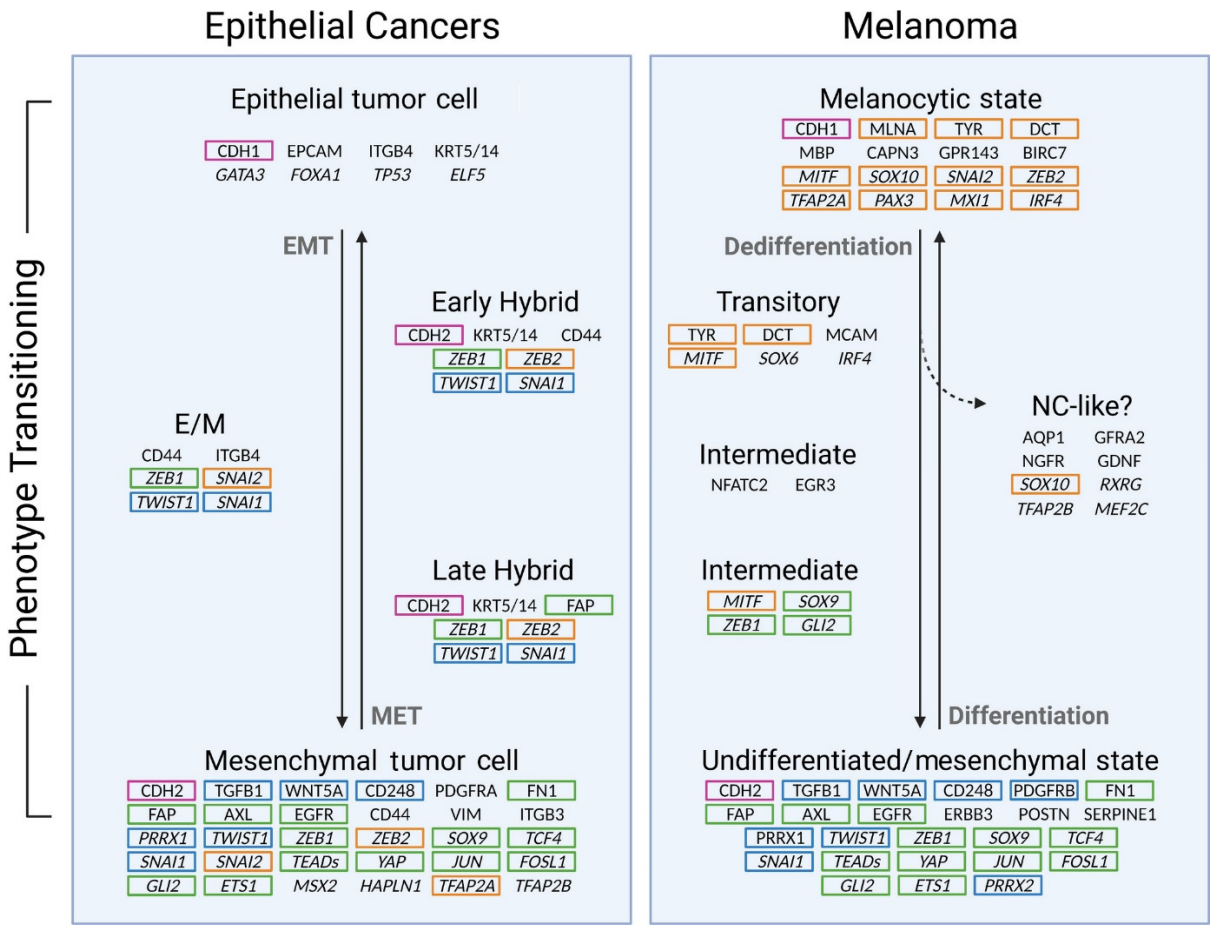


Figure 13. Molecular bases of EMT and melanoma phenotype switching.

From Pedri et al. (2021)

The epithelial-like state was initially named "proliferative" (Figure 12) because of the propensity of these cells to quickly proliferate and their low invasiveness properties but now this state is referred to as "melanocytic" (MEL). This state shows a high differentiation level with strong activity of both MITF and another melanocyte lineage transcription factor encoded by *SOX10*. Cells in this state express the epithelial hallmark *CDH1* and can use oxidative-phosphorylation (OxPhos) for energy production as confirmed by single-cell analysis[66]. Contrary to many carcinomas, both *ZEB2* and *SNAI2* expression are associated with the melanocytic state[34,67] (Figure 13).

The mesenchymal-like state was called "invasive" because of the cells high motility and low proliferation but is now referred to as "undifferentiated" or "mesenchymal" (MES). This state is characterized by low activity of MITF/SOX10 which is replaced by high activity of transcription factors like SOX9, ZEB1, activator protein 1 (AP-1) and PRRX1. The pit-oct-unc domain (POU) transcription factor BRN2 (*POU3F2*) which is important during embryogenesis for neural crest cells seems implicated in melanoma phenotype switching *in-vivo* as well[68]. Cells in this state display mesenchymal hallmark genes like *CDH2*, *FN1*, *CD44* and *AXL* which encodes an RTK implicated in therapy resistance[69]. Like for EMT, the MEL to MES switch is not binary and different intermediate phenotypes have been reported including an $MITF^{low}/SOX10^{high}$ "neural-crest like" phenotype and an "immune-related" phenotype characterized by inflammation response genes[70]. Undifferentiated melanoma cells are capable of forming their own vasculature in a process called "vasculogenic mimicry"[71]. Some evidence show endothelial trans-differentiation to be possible as well with implications in dormancy and relapse[72].

All in all, combination of these different phenotypes confers melanoma tumors adaptability to different scenarios that allows them to progress and resist treatments.

D. Non-mutational epigenetic reprogramming: the role of long non-coding RNAs

Cancer cell plasticity is achieved by hijacking various gene regulation mechanisms. Most of the human genome is dedicated to this regulation allowing the development of an organisms with trillions of cells and thousands of phenotypes from a single cell. We already saw examples of gene regulation during EMT through the histone code with remodeler factors depositing activating or repressing marks on chromatin but several other mechanisms exist including alternative splicing, DNA methylation, enhancer elements and regulation by RNAs[1]. In this section, we will overview the regulation of gene expression by long non-coding RNAs (lncRNAs) with examples in cancer.

The GENCODE initiative is a project aimed at delivering an annotation of functional elements in the mouse and human genomes[73]. This consortium classified genes into four main biotypes, the first two are protein coding genes characterized by a coding sequence (CDS) and pseudogenes that are derived from protein coding genes but do not appear to be transcribed. The third biotype comprises the small non-coding RNAs (sncRNAs) which do not contain an open reading frame (ORF) and are smaller than 200 nucleotides, this class contains microRNAs (miRNA), small interfering RNAs (siRNAs), small nuclear RNAs (snRNAs), PIWI-interacting RNAs (piRNAs) and tRNA-derived small RNAs (tRFs).

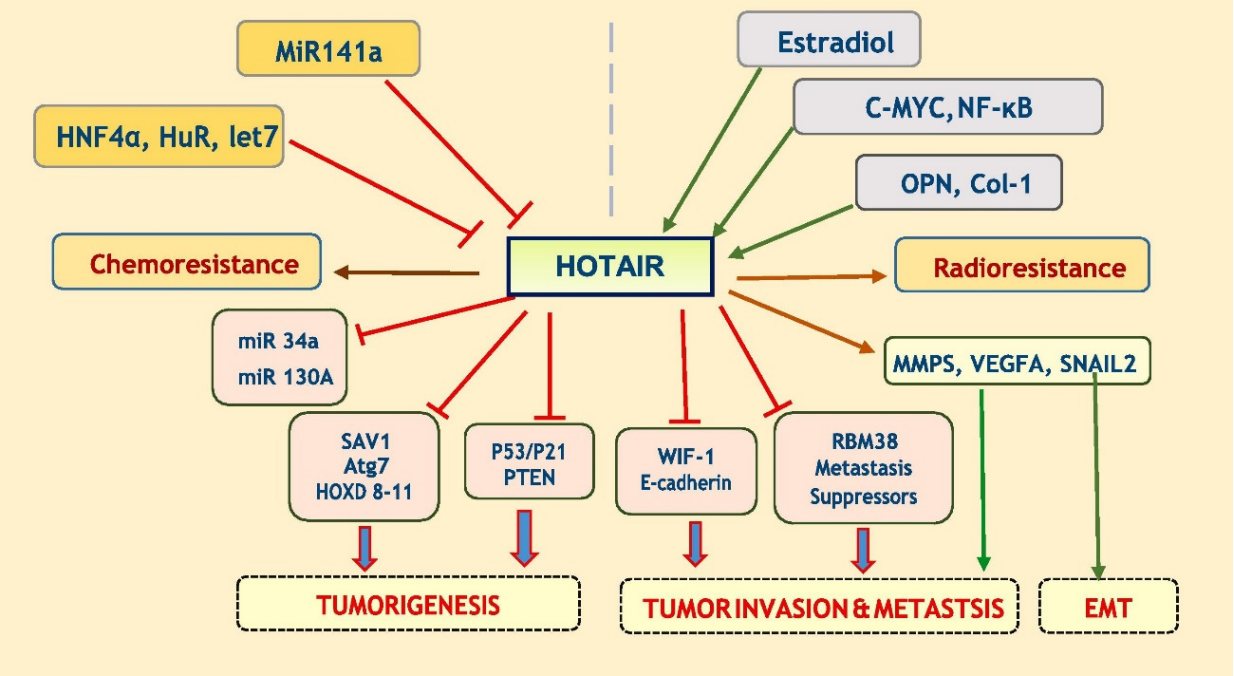


Figure 14. Involvement of lncRNA HOTAIR in oncogenic processes.

From Rajagopal et al. (2020)

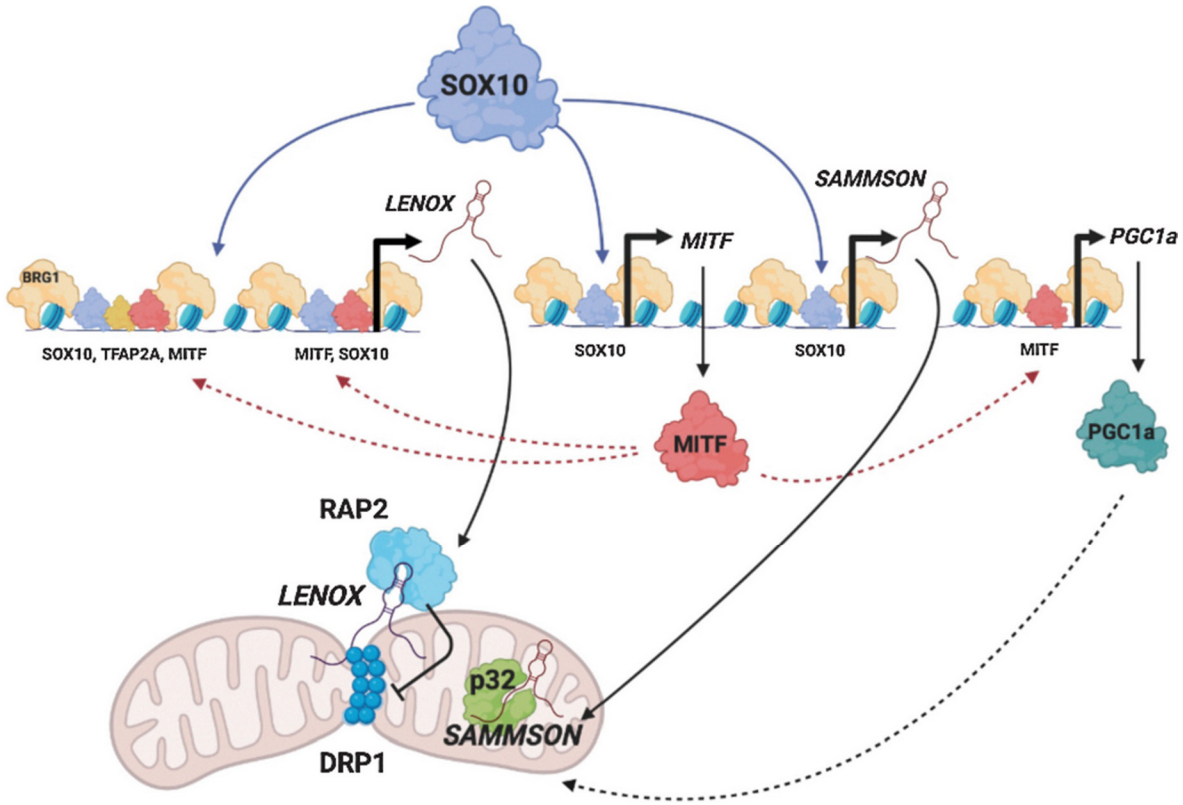


Figure 15. Regulation of OxPhos by lncRNA LENOX and RAP2 in melanoma.

From Gambi et al. (2022)

The last gene biotype is lncRNAs that do not contain an ORF and are longer than 200 nucleotides. These are grouped into four categories depending on their localization on the genome: long intergenic RNAs (lincRNAs) do not overlap genes, anti-sense RNAs (asRNAs) are transcribed on the opposite strand of a coding gene, sense overlapping lncRNAs are transcribed on the same strand of a coding gene and sense intronic lncRNAs are contained within an intron of a coding gene. Functionally these genes regulate a large array of biological processes like cell differentiation, cell cycle, metabolism, stress responses and can be implicated in pathogenesis. For this, they use a vast array of either close-proximity (*cis*) or distant (*trans*) regulation mechanisms ranging from control of transcription to translation and epigenetics[74]. We will briefly explore two examples of lncRNAs relevant to cancer.

The homeobox transcript antisense intergenic RNA (*HOTAIR*) is a heavily studied lncRNA with roles in development[75] and cancer progression[76]. The 5' region of HOTAIR interacts with the repressor complex PRC2 through binding of its EZH2 and SUZ12 subunits[77] thus directing the complex to its target gene with the Jumonji and AT-rich interaction domain containing 2 (JARID2) protein[78]. On the other hand, the 3' region can bind LSD1 making HOTAIR a scaffold for LSD1/REST/CoREST as well as for PRC2 recruitment[79] which were both mentioned in the EMT section. Unsurprisingly given its mechanisms, HOTAIR is involved in type 3 EMT for multiple solid cancers. Its expression is associated to SNAIL and TGF- β mediating silencing of *CDH1*. Furthermore, its knockdown reduced *MMP9* expression. Aside from EMT, HOTAIR has additional roles in about 18 cancers including regulation of Notch, Wnt, β-catenin, Hippo, MAPK and NF-kB signaling as well as repression of the *TP53* and *PTEN* tumor suppressor genes[80] (Figure 14).

Another interesting example of lncRNA used by cancer cells is the melanoma specific lincRNA enhancer of oxidative phosphorylation (LENOX) encoded in the *LINC00518* gene. LENOX is expressed across all melanoma phenotypes and is transcribed by MITF/SOX10 or potentially by TFAP2A in undifferentiated states. LENOX is mainly cytoplasmic where it associates with RAS-related GTPase RAP2 (*RAP2A*, *RAP2B*, *RAP2C*) and enables RAP2 association with the large GTPase DRP1 responsible for mitochondrial fission. Interaction of LENOX/RAP2 with DRP1 leads to increased phosphorylation on S637 inhibiting its activity thus promoting mitochondrial fusion leading to optimization and increase of OxPhos energy production. Interestingly, this LENOX dependent mechanism can be targeted to increase vulnerability of cancer cells in situations where glycolysis is inhibited such as during BRAF inhibition therapies[81] (Figure 15).

In summary, lncRNAs constitute highly specific regulatory mechanisms which can be diverted by cancer cells potentially making them ideal biomarkers or therapy targets.

E. Tumor microenvironment

A tumor is a very complex ecosystem populated by cancer cells but also numerous cells from the host organism. As we saw in previous sections the TME is source of many molecular queues affecting cancer cells and can be either an ally or a foe dictating the tumor progression. In this section we will introduce some of the major immune and stromal cells composing the TME as well as their specific roles in cancer.

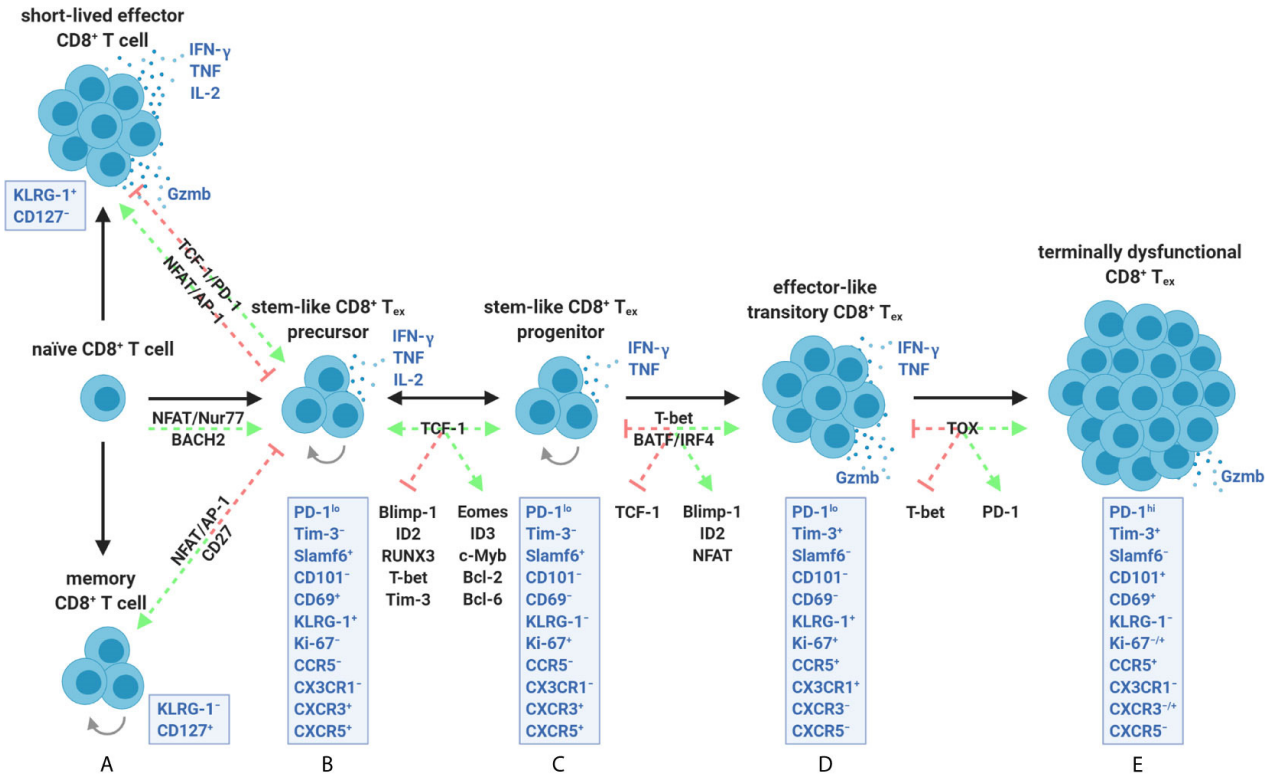


Figure 16. Molecular markers of CD8 T-cell states.

From Dolina et al. (2021)

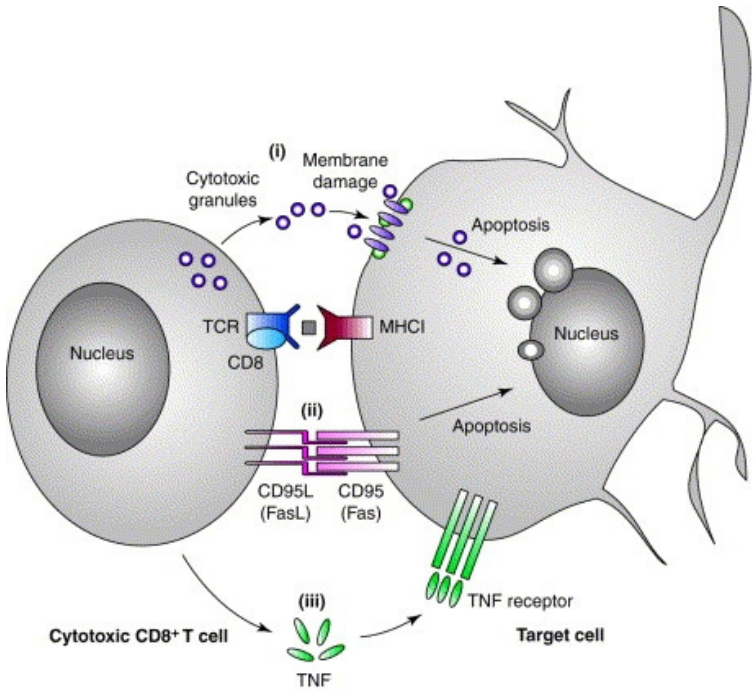


Figure 17. Mechanisms of cell killing by CD8 T-cells.

From Neumann et al. (2002)

a. CD8$^+$ T-lymphocytes

All immune cells are very plastic with a variety of phenotypes allowing an adaptive immune response to threats as well as long term memory. This response is mainly orchestrated by T-lymphocytes (T-cells) expressing the CD3 (*CD3D*, *CD3E*, *CD3G)* receptor, these cells are generally classified into *CD8* (CD8 T-cell) or *CD4* (CD4 T-cell) expressing cells[82]. One of the most crucial immune cell type of the TME is the CD8 T-cell. These cells possessing cytotoxic activity are responsible for killing infected cells during bacterial or viral infections and are also the main defense in anti-cancer response. Multiple phenotypes of CD8 T-cells have been reported including naïve, activated effector, memory and exhausted cells[83] (Figure 16).

Effector cells exert their functions through multiple molecules, relevant for cancer cell killing are granule-associated enzymes (granzymes), perforins and interferon-γ. Once primed by antigen recognition through the major histocompatibility complex class I (MHC-I), cytotoxic cells will release cytolytic granules and perforins at their target[84]. These granules contain granzymes such as GZMA/GZMB, serine-proteases inducing a form of non-apoptotic cell death called pyroptosis[85]. Perforins such as PRF1 will induce pore formation in the plasma-membrane which helps internalization of granules by the target cell but they may also be able to induce cell lysis by themselves[86]. Interferon (IFN)-γ release will impair cystine uptake by target cells causing lipid peroxidation which eventually leads to ferroptotic cell death[87]. Additionally, cytotoxic cells induce apoptosis through secretion of tumor necrosis factor superfamily member 10 (TRAIL/*TNFSF10*) and stimulation of Fas cell surface death receptor (FAS1/*FAS*) by Fas ligand (FasL/*FASLG*). These cytotoxic mechanisms are additive and it is their combination that seems to lead to cancer cell elimination[88] (Figure 17).

CD8 T-cells have a safety mechanism in the form of inhibitory receptors acting as an "off-switch" and representing immune checkpoint proteins. One such receptor is the programmed cell death 1 (PD-1/*PDCD1*) protein which can be stimulated by several ligands including programmed cell death 1 ligand 1 (PD-L1/*CD274*) and programmed cell death 1 ligand 2 (PD-L2/*PDCD1LG2*). Another well characterized receptor is the cytotoxic T-lymphocyte associated protein 4 (CTLA-4/*CTLA4*) activated by its ligands CD80 and CD86. Many more such checkpoints exist including T-cell immunoglobulin mucin receptor 3 (TIM3/*HAVCR2*), lymphocyte activating 3 (LAG-3/*LAG3*) and T-cell immunoreceptor with Ig and ITIM domains (*TIGIT*). In case of prolonged activation after viral persistence such as with human immunodeficiency virus (HIV) or after failure to clear out tumor cells, CD8 T-cells will tend to downregulate cytotoxic effector genes and overexpress immune checkpoint receptors. This non-functional state has been termed "exhausted" state. The switch to this state is epigenetic and governed by transcription factors. After stimulation of immune checkpoint receptors, active effector cells with high T-bet (*TBX21*) activity will switch to the exhausted state with high activity of thymocyte selection associated high mobility group box (*TOX*) and with restriction of cytotoxicity by eomesodermin (*EOMES*)[89]. CD8 T-cells composing the TME will often adopt this hyporeactive state as a consequence of chronic tumoral antigen exposition and reversal of this state has been the basis for therapeutic approaches such as immunotherapy[83]. As is the case with EMP, exhaustion is a continuum involving many factors and leading to different states that are dependent on the context and the disease.

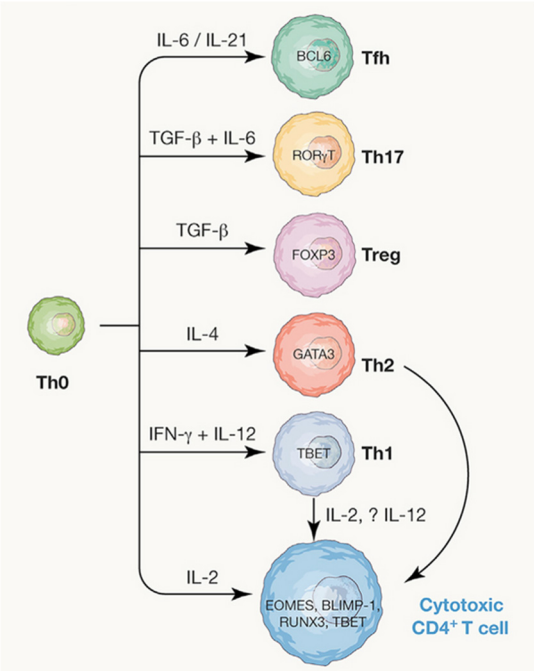


Figure 18. Phenotypes of CD4 T-cells.

From Oh et al. (2021)

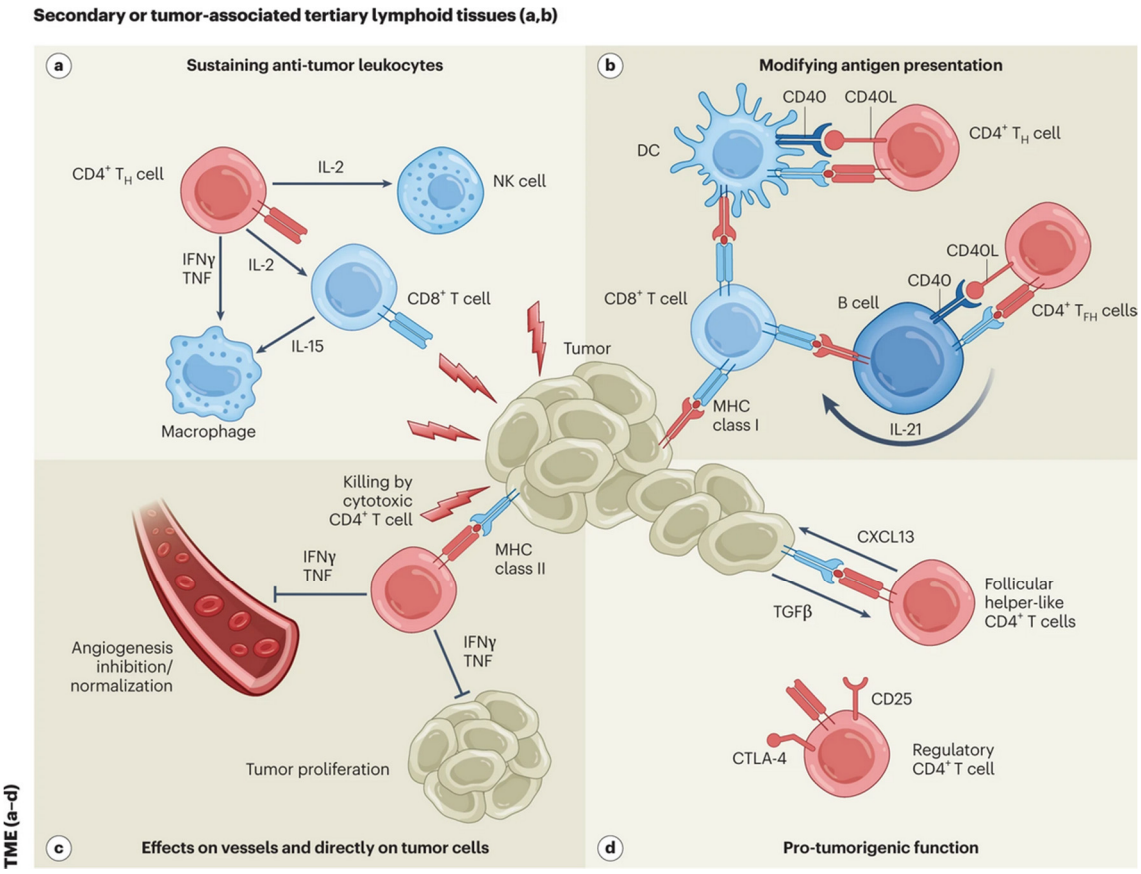


Figure 19. Roles of CD4 T-cells in cancer.

From Speiser et al. (2023)

b. Natural killer cells

The other type of lymphoid cell with cytotoxic cell killing potential is the natural killer (NK) cell. These cells are part of the innate immune system lacking the CD3 receptor and are capable of autonomous cell killing using the same mechanisms as CD8 T-cells described above. One key difference is that NK cells will lyse cells unable to identify themselves through MHC-I. As tumor cells tend to downregulate MHC-I in order to evade CD8 T-cell killing[90], NK cells can potentially counteract this strategy and complement T-cells in anti-cancer response. In the TME, NK cells can contribute to adaptive immunity by producing signals like IFN-γ despite being considered part of innate immunity. Similar to CD8 T-cell, NK cells can be inhibited by immune checkpoints such as PD-1 and they can present dysfunctional states induced by repeated stimulation of their NKG2D (*KLRK1*) receptor[91].

c. $CD4^+$ T-lymphocytes

The second type of T-cells, marked by CD4, possesses multiple phenotypes and play different roles mainly centered on regulation through cytokine signaling but some are also capable of cytotoxic activity[92]. Plasticity of CD4 T-cells has been defined around the T-helper (Th) polarization, they start out as un-polarized Th0 cells and can then adopt Th1, Th2, Th17, T-follicular helper (Tfh) or regulatory T-cell (Treg) phenotypes after recognition of antigens presented by the major histocompatibility complex class II (MHC-II) and depending on activation signals. Th1 polarization activated in response to interleukin (IL)-12 and interferon-γ is important for anti-bacterial and viral responses. It is characterized by production of interferon- γ, TNF-α and IL-2. The Th2 phenotype is induced by IL-4 and is implicated in defense against extracellular parasites with Th2 cells producing specific cytokines IL-4, IL-5 and IL-13. Th17 polarization is triggered by TGF- β and IL-6, it is a determinant in response against fungi by producing other cytokines such as IL-17A, IL-17F and IL-22. The Tfh phenotype activated by IL-6 and IL-12 induces humoral immunity by stimulating B-cells through IL-21 and IL-4[93]. Lastly, the Treg phenotype is induced by TGF- β and has an immunosuppressive role. Tregs express the transcription factor forkhead box P3 (*FOXP3*) and exert suppressive functions notably through CD25 absorbing the stimulation signal IL-2 and through CTLA-4 and LAG-3 that inhibit cytotoxic cells (Figure 18).

In the TME, multiple phenotypes of CD4 T-cells were detected, the majority have anti-tumor activity while some can have pro-tumor effects. They are a major driver of anti-cancer immunity by priming CD8 T-cells, antigen-presenting cells (APCs) and B-cells mainly through stimulation of the CD40 receptor, but can also trigger and maintain active phenotypes for myeloid and NK cells by releasing multiple cytokine signals. On the other hand, immunosuppression by Tregs when they outnumber cytotoxic cells is pro-tumorigenic and correlates with poor prognosis in multiple cancers. Additionally, CD4 T-cells have been shown to nurture cancer cells from B-cell lymphomas, but also solid tumors such as neuroblastoma[94]. The activity of lymphocytes is heavily dependent on metabolism which could be manipulated by the TME to create tumor-favorable conditions[95] (Figure 19).

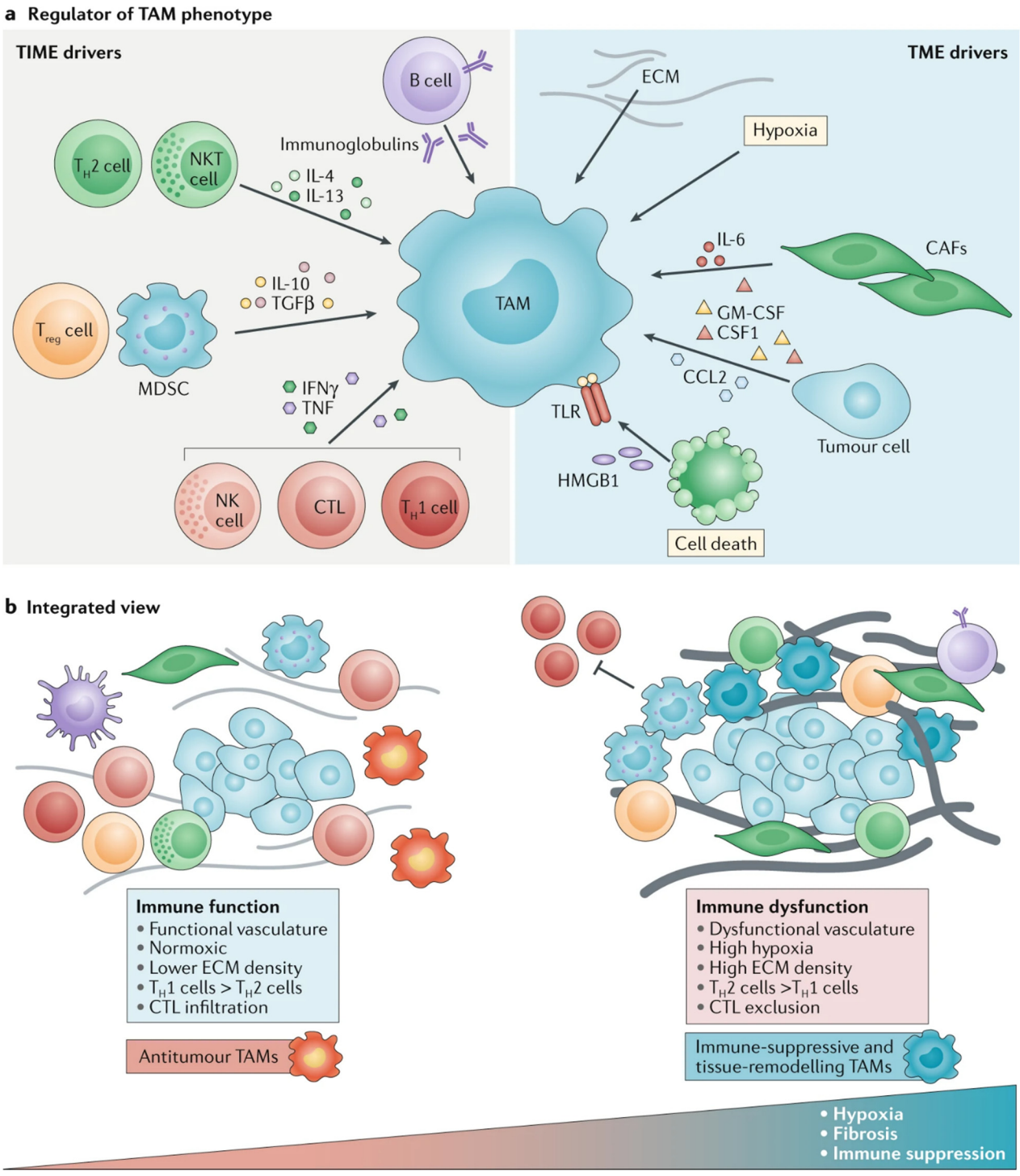


Figure 20. Functions of tumor-associated macrophages phenotypes.

From DeNardo et al. (2019)

d. Macrophages

Antigen-presentation is crucial for establishing an adaptive immune response, this role is undertaken by cells types from the myeloid lineage expressing the MHC-II system: macrophages and dendritic cells. Macrophages are scavenger cells from the innate immune system with multiple functions: phagocytosis of pathogenic agents, antigen-presentation and modulation of inflammatory and wound-healing processes. Phenotypically, these cells have been described around a polarization gradient ranging from inactive M0 then M1 to M2[96]. The M1 activation state is pro-inflammatory both because of production of multiple cytokines like IL-12 and IL-23 and because of a high tendency to present antigens. On the other hand, the M2 activation profile is anti-inflammatory with reduced capacity of antigen-presentation and release of IL-10, IL-4 and IL-13 cytokines. These M2 cells are tasked to clean-up after inflammatory disruption by reducing inflammation and promoting cell proliferation and angiogenesis for tissue repair[97].

Macrophages constituting the TME are called tumor-associated macrophages (TAMs) and are mainly derived from blood-circulating monocytes, but can also originate from tissue-resident populations[98]. Because of their plasticity TAMs can have different roles with M1-like TAMs considered to have anti-tumor effects and M2-like TAMs representing a pro-tumorigenic phenotype. These M2 TAMs have been reported as major players supporting tumor development in multiple cancers. These will negatively regulate T-cells through direct actions like arginine and nitric oxide metabolism or suppression of cytotoxicity through immune checkpoint ligands such as PD-L1. This negative regulation can also be indirect, for example TGF-β secretion impacts recruitment of Tregs while modulation of vasculature and ECM favors protumoral myofibroblasts that will be described below[99]. The M2-associated tissue repair program also favors tumor expansion with production of growth factors like FGF and EGF supporting cancer cell proliferation and with secretion of vascular endothelial growth factor A (*VEGFA*) promoting angiogenesis and neovascularization supplying the tumor with oxygen and nutrients. FGF, EGF and TGF-β are all EMT triggers described previously, combined with angiogenesis providing cancer cells a travelling route to the rest of the body this makes M2-TAMs one of the promoters of the metastatic process[97] (Figure 20).

e. Other immune cells

Additional immune cells that participate to the TME should be mentioned. Dendritic cells are key drivers of immune responses. These APCs have many subtypes with some types like conventional dendritic cells having an essential role in mounting an anti-cancer response while some others like plasmacytoid dendrictic cells promote Treg proliferation and are protumoral[100]. B-cells responsible for humoral response generate antibodies against tumoral antigens and cross-present them to T-cells inside accumulations of T and B-cells called tertiary lymphoid structures (TLS) which are located inside the tumor and resemble secondary lymphoid organs. In this manner, B-cells help the immune response but some phenotypes can also be inhibitory and help the tumor[101]. Lastly, mast cells and neutrophils represent other myeloid types of cells with plastic phenotypes that can either hinder or help tumor progression[102,103].

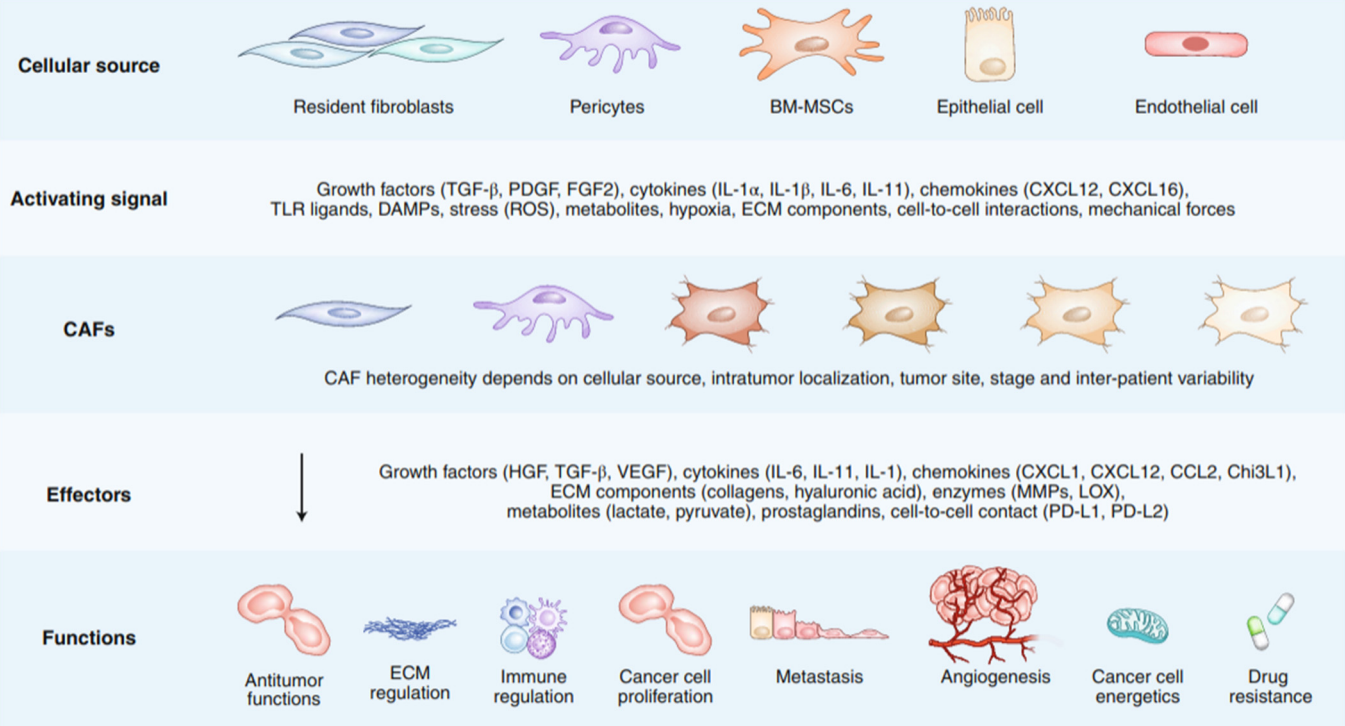


Figure 21. Origin and functions of cancer-associated fibroblasts.

From Koliaraki et al. (2020)

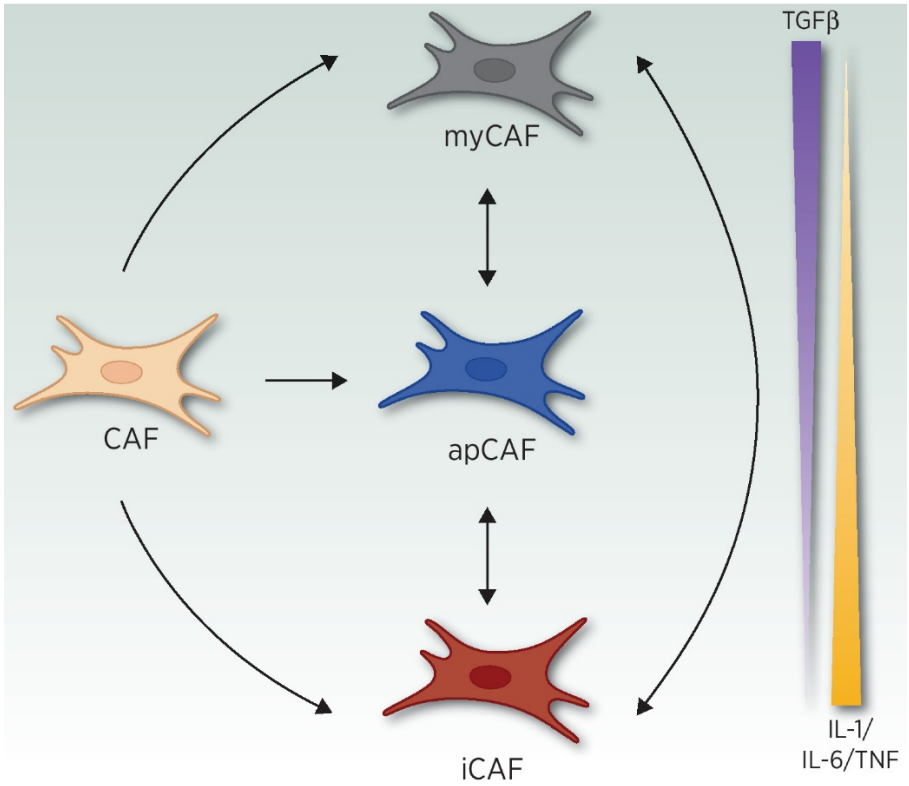


Figure 22. Plasticity of cancer-associated fibroblasts.

From Kennel et al. (2023)

f. Fibroblasts

Aside from immune cells, the TME is composed of stromal cells which are endothelial cells and mesenchymal cells such as fibroblasts and pericytes. Fibroblasts represent a heterogeneous class of cells with multiple functions both in embryonic and adult tissues. During adulthood, they populate connective tissues and have roles in regulation of ECM for tissue homeostasis, orchestration of inflammatory responses and damaged tissue regeneration. These cells can dysfunction which has implications in diseases such as cancer and organ fibrosis which we covered in the section about type 2 EMT. In cancer, fibroblasts are termed cancer-associated fibroblasts (CAFs), are perpetually activated and have a large variety of phenotypes and origins depending on the afflicted tissue[104] (Figure 21). Multiple subtypes have been reported with the advance of single-cell sequencing, for example CAFs have been well studied in pancreatic ductal carcinoma (PDAC) with description of 3 subsets[105] while 6 types of CAFs could also be defined by pooling samples from melanoma, head and neck squamous cell carcinoma (HNSC) and lung cancer[106]. Phenotypically, CAFs can be described using two broad and antagonistic categories: inflammatory-CAFs (iCAFs) and myofibroblastic-CAFs (myCAFs) regulated by IL-1α and TGF-β[107] (Figure 22).

The myCAF phenotype is activated by TGF-β which suppresses IL-1 receptor expression and prevents adoption of the iCAF phenotype[108]. This antagonism likely explains the spatial segregation of the two phenotypes as iCAFs are localized rather far away from the tumor and myCAFs are closer to the heart of the tumor[109]. This phenotype is characterized by expression of fibroblast activation protein alpha (*FAP*), this gene encodes a cell surface protease that participates in ECM degradation[110]. As the name suggests, myCAFs also express contractile genes like α-SMA (*ACTA2*), transgelin (*TAGLN*) and myosin regulatory light chain 9 (*MYL9*). This expression profile is similar to fibroblasts activated during wound healing except that the activation state is always maintained so myCAFs do not revert back to a quiescent state. These cells have dual functions, initial attempts to deplete myCAFs in tumors led to early dissemination and faster metastasis so it is believed that they initially restrain the tumor[111] but later on in the disease they will adopt tumor-promoting roles by regulating the immune system and supporting the metastatic process. The TGF- β activation program contains several signals that can result in growth and type 3 EMT of cancer cells, for example they secrete cytokines like IL-6 and IL-11 and they can activate Notch signaling through expression of jagged canonical notch ligand 1 (*JAG1*)[112]. Also supporting the metastatic process, myCAFs secrete the proangiogenic signal angiopoietin-like 4 (*ANGPTL4)*[113] and the ECM proteolysis coupled with secretion of components like FN1 both increase the invasive potential of cancer cells[114]. These CAFs also have immunosuppressive mechanisms that can help the tumor, notably they can induce the Treg phenotype[115] and IL-6 favors recruitment of M2-TAMs[116].

The iCAF phenotype, on the other hand, is activated by IL-1α. These CAFs downregulate myCAF contractile features and strongly express an inflammatory program characterized by IL-6, IL-8, IL-11, leukemia inhibitory factor (*LIF*), C-X-C motif chemokine ligand (CXCL) 1, CXCL2 and CXCL12. In human data, one subset of iCAF was characterized by pronounced expression of MHC-II genes such as *CD74*, *HLA-DRA*, *HLA-DPA1* and *HLA-DQA1*. This group was thus termed antigen-presenting CAFs (apCAFs)[117]. Like myCAFs, iCAFs can have anti-tumor roles by promoting organization of TLS[118] and they can have tumor

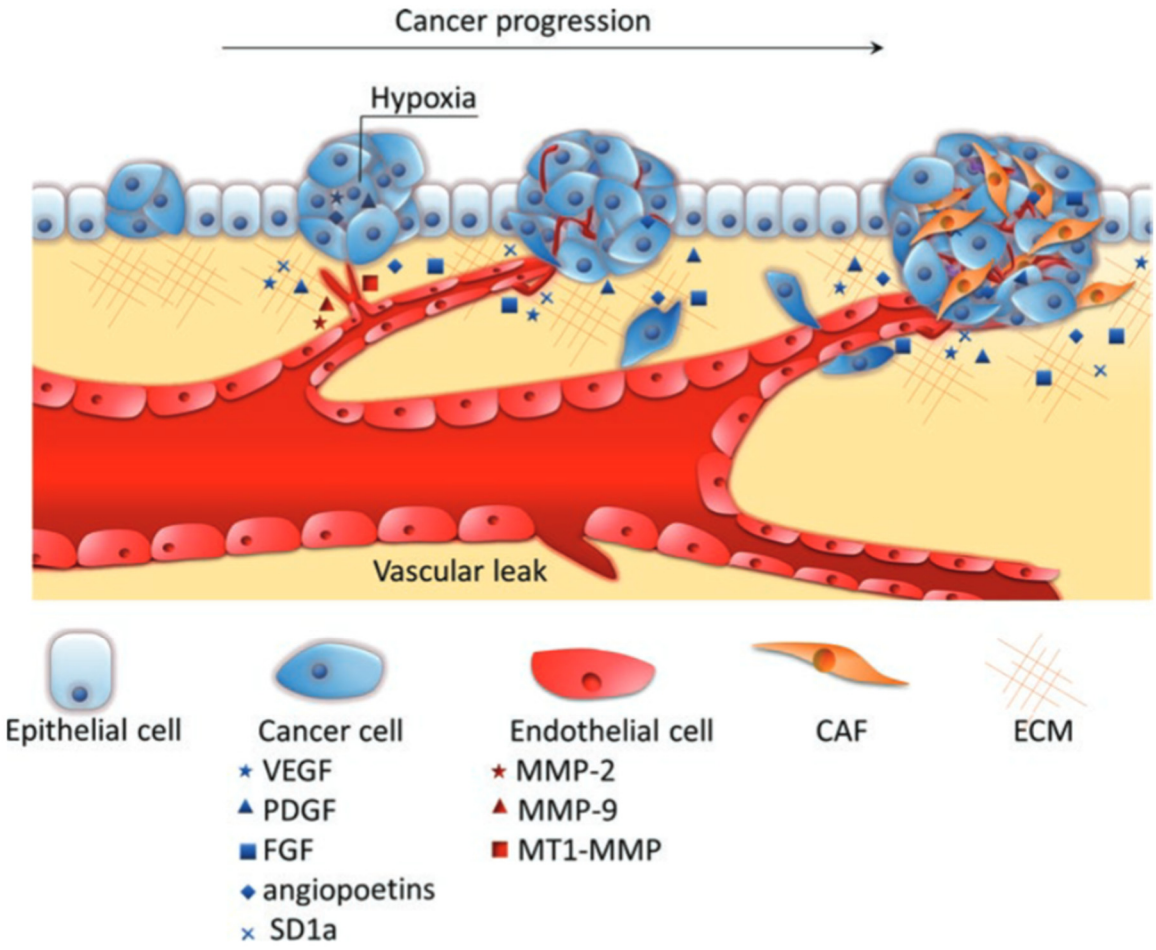


Figure 23. Angiogenesis and cancer progression.

From Sobierajska et al. (2020)

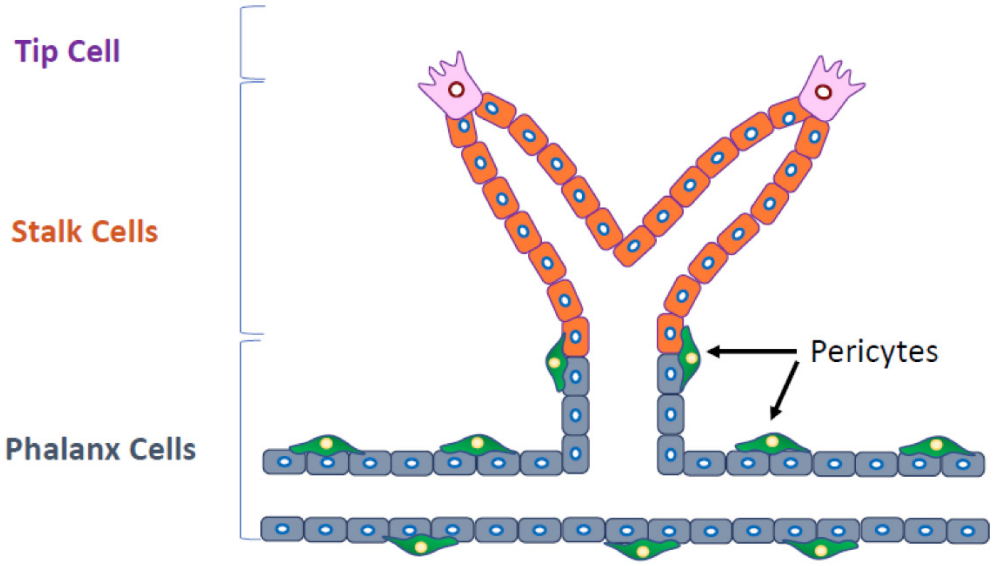


Figure 24. Endothelial phenotypes during new vessel formation.

From Hida et al. (2018)

promoting roles. Mechanisms can be similar such as cytokine signals promoting type 3 EMT and recruitment of an immunosuppressive TME through TAMs and Tregs but some specific mechanisms also exist. For example, when apCAFs present antigens to CD4 T-cells, their lack of co-stimulatory signals tends to induce Treg polarization instead of active CD4 T-cell clonal expansion[119].

g. Vascularization

In addition to fibroblasts, stromal TME contains tumor-associated vasculature composed mainly by pericytes, smooth-muscle and endothelial cells. This vasculature is formed through angiogenesis which expands an existing network of blood vessels. The process is induced by previously mentioned factors like VEGFA (stimulates VEGFR endothelial receptors) produced by the TME such as M2-TAMs but also by hypoxic cancer cells[120] (Figure 23).

Endothelial cells (ECs) are mono-layered cells forming blood vessel walls, they are a multifunctional and very heterogeneous category both in normal conditions and in the TME. During angiogenesis, ECs adopt a mobile phenotype called "tip" cells which drive vessel sprouting, these cells are followed by proliferative "stalk" cells while the more quiescent "phalanx" cells support the newly formed vessels[121] (Figure 24). Tumor ECs (TECs) form vessels with an abnormal morphology characterized by an excess of branching and bulges. They can originate either from normal ECs or from transdifferentiated cancer cells as described in the melanoma section. These TECs acquire dysfunctional properties such as responsiveness to epidermal growth factor receptor (HER1/*EGFR*), resistance to apoptosis and autocrine promotion of their survival. They also have both direct and indirect effects on cancer cells that promote tumor progression and metastasis. They can induce type 3 EMT through IL-6, JAG1, stimulation of NF-κB and ERK signaling[122]. The trans-endothelial migration (TEM) of cancer cells, a process in which cancer cells will pass through individual endothelial cells to disseminate into the blood stream, is facilitated by TECs which have loose immature junctional contacts. Additionally, TECs are capable of endothelial-to-mesenchymal transition (EdMT) triggered mainly by TGF- β which potentially disrupts EC junctions thus facilitating TEM and converts TECs into CAFs[123] that have many pro-tumor roles as described in the previous section.

Pericytes are fibroblastic cells embedded in the vascular basement membrane that participate in stability of vessels and communicate with ECs[124]. In the TME, pericytes display an activated type-2 phenotype which is pro-angiogenic[125] and express regulator-of-G-protein-signaling-5 (*RGS5*) that sustains their survival and expansion in presence of TGF- β[126]. Due to their mesenchymal nature, these cells represent another source of CAFs. Overall, pericytes seem to display tumor promoting roles through abnormal activation.

WHO RENAL CELL CARCINOMA (RCC) CLASSIFICATION

2016 classification

- Clear cell renal cell carcinoma
- Multilocular cystic renal neoplasm of lowmalignant potential
- Papillary renal cell carcinoma
- Hereditary leiomyomatosis and renal cell carcinoma (HLRCC)-associated renal cell carcinoma
- Chromophobe renal cell carcinoma
- Collecting duct carcinoma
- Renal medullary carcinoma
- MIT Family translocation carcinomas
- Succinate dehydrogenase (SDH)-deficientrenal carcinoma
- Mucinous tubular and spindle cell carcinoma
- Tubulocystic renal cell carcinoma
- Acquired cystic disease associated renal cell carcinoma
- Clear cell papillary renal cell carcinoma
- Renal cell carcinoma, unclassified
- Papillary adenoma
- Oncocytoma

2022 classification

- Clear cell renal tumours
  - Clear cell renal cell carcinoma
  - Multilocular cystic renal neoplasm of low malignant potential
- Papillary renal tumours
  - Renal papillary adenoma
  - Papillary renal cell carcinoma
- Oncocytic and chromophobe renal tumours
  - Oncocytoma of the kidney
  - Chromophobe renal cell carcinoma
  - Other oncocytic tumours (Low-grade oncocytic tumor, Eosinophilic vacuolated tumour)
- Collecting duct tumours
  - Collecting duct carcinoma
- Other renal tumours
  - Clear cell papillary renal cell tumour
  - Mucinous tubular and spindle cell carcinoma
  - Tubulocystic renal cell carcinoma
  - Acquired cystic disease-associated renal cell carcinoma
  - Eosinophilic solid and cystic renal cell carcinoma
  - Renal cell carcinoma NOS
- *Molecularly defined renal carcinomas*
  - *TFE3-rearranged renal cell carcinomas*
  - *TFEB-rearranged renal cell carcinomas*
  - *ELOC (formerly TCEB1)-mutated renal cell carcinoma*
  - *Fumarate hydratase-deficient renal cell carcinoma*
  - *Succinate dehydrogenase-deficient renal cell carcinoma*
  - *ALK-rearranged renal cell carcinomas*
  - *SMARCB1-deficient renal medullary carcinoma*

Figure 25. Classification of renal cell carcinomas from 2016 to 2022.

From Rizzo et al. (2023)

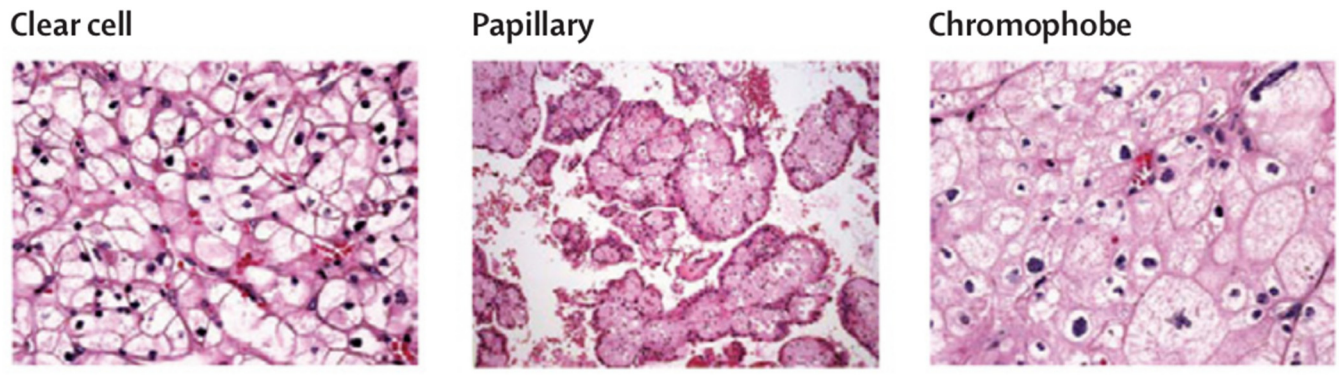


Figure 26. Histologies of renal cell carcinoma.

From Rini et al. (2009)

2. Renal cell carcinoma

A. Classification of renal cell carcinoma

After reviewing general mechanisms pertaining to cancer, in this section we will focus more specifically on one cancer type. Kidney cancer is among the ten most frequent cancers in the United Kingdom[127] and the United States. Renal cell carcinoma (RCC) represents 90% of all kidney cancers and accounted for 2% of all cancer deaths in 2016[128]. RCC is a very heterogeneous disease with 16 subtypes described in the 2016 WHO classification which was recently updated in 2022 with about 21 subtypes[129] (Figure 25). We can distinguish 3 major histologic categories[130]: clear-cell renal cell carcinoma (ccRCC) which represents about 75% of RCC cases, papillary renal cell carcinoma (pRCC) which is about 15-20% of cases and chromophobe renal cell carcinoma (chRCC) accounts for around 5% of cases[131] (Figure 26). As ccRCC will be described in its own section, we will briefly expand other subtypes.

The second most frequent type, pRCC, presents a papillary appearance with heterogeneous cellular morphologies. These tumors were initially classified into type 1 or type 2, but this distinction has been abandoned with some tumors with papillary architecture emerging as their own entities. Genetically, they are mainly characterized by GoF driver mutations in the RTK oncogene *MET* (c-MET) and gain of chromosomes 7/17 in classical pRCC, but rarer recurrent mutations were also reported such as mutations in the tumor suppressor *NF2* or the oncogene *KRAS*[132].

The third most common subtype, chRCC, are tumors containing eosinophilic cells in the center and peripheral smaller "chromophobe" cells which are also called "pale" cells. These tumors are sporadic and generally have a favorable prognosis[133]. Recurrent alterations have been identified such as chromosome losses (1, 2, 6, 10, 13, 17, 21 and X/Y) with tumor suppressor genes *PTEN* and *TP53* emerging as the most frequently mutated[132].

Aside from these main subtypes, other rarer subtypes should be mentioned because of their lethality. Some RCCs are defined by their molecular features, one such example is the microphthalmia transcription factor (MiT) family translocation renal cell carcinoma (tRCC) which represents 1-4% of RCC cases. These tumors present structural rearrangements involving genes of the MiT family: the previously described *MITF*, transcription factor binding to IGHM enhancer 3 (*TFE3*), transcription factor EC (*TFEC*) and transcription factor EB (*TFEB*). The *TFE3* (located at Xp11) tRCCs show recurrent gene fusion partners like proline rich mitotic checkpoint control factor (*PRCC*), non-POU domain containing octamer binding (*NONO*), splicing factor proline and glutamine rich (*SFPQ*), alveolar soft part sarcoma chromosomal region 1 (*ASPSCR1*) and clathrin heavy chain (*CLTC*). The *TFEB* tRCCs have one main fusion partner in metastasis associated lung adenocarcinoma transcript 1 (*MALAT1*) though others have been reported[134] but *TFEB* amplifications were also identified leading to this category of tumors being now called "*TFEB*-altered RCCs"[129]. Both *TFE3* and *TFEB* are regulators of lysosomal activity but their roles in cancer are poorly understood. It is unclear whether their oncogenic properties come from their overexpression, GoF from the fusion or LoF from the partner[135]. Other rare subtypes of aggressive RCC include tumors with LoF alterations for fumarate hydratase (*FH*)[136], succinate dehydrogenase (*SDH*)[137] or the SWI/SNF integrase interactor 1 (INI-1/*BAF47*/*SMARCB1*) that will be described in a later section.

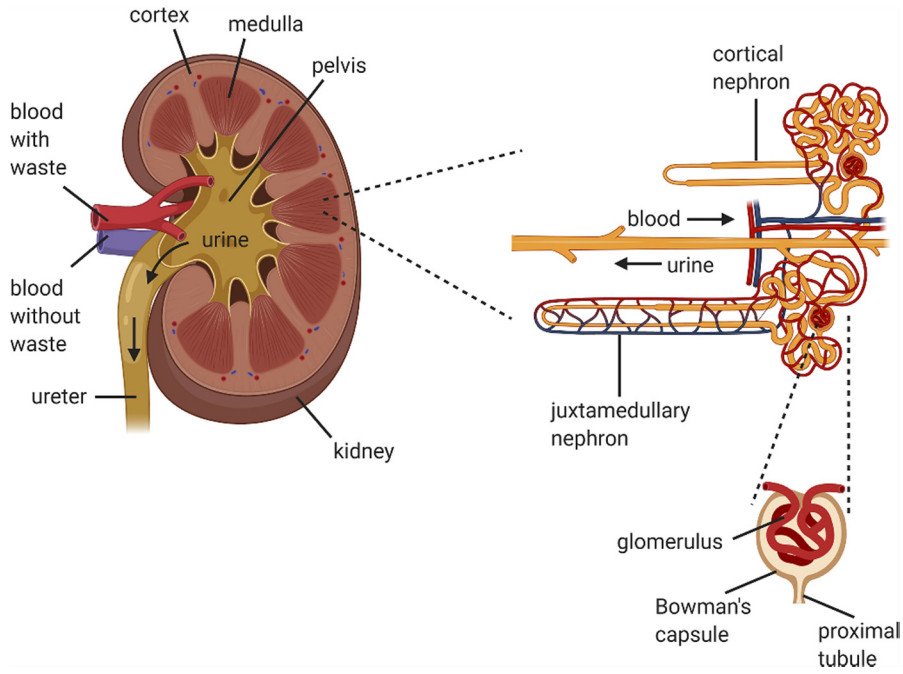


Figure 27. Anatomy of the kidney.

From Fransen et al. (2021)

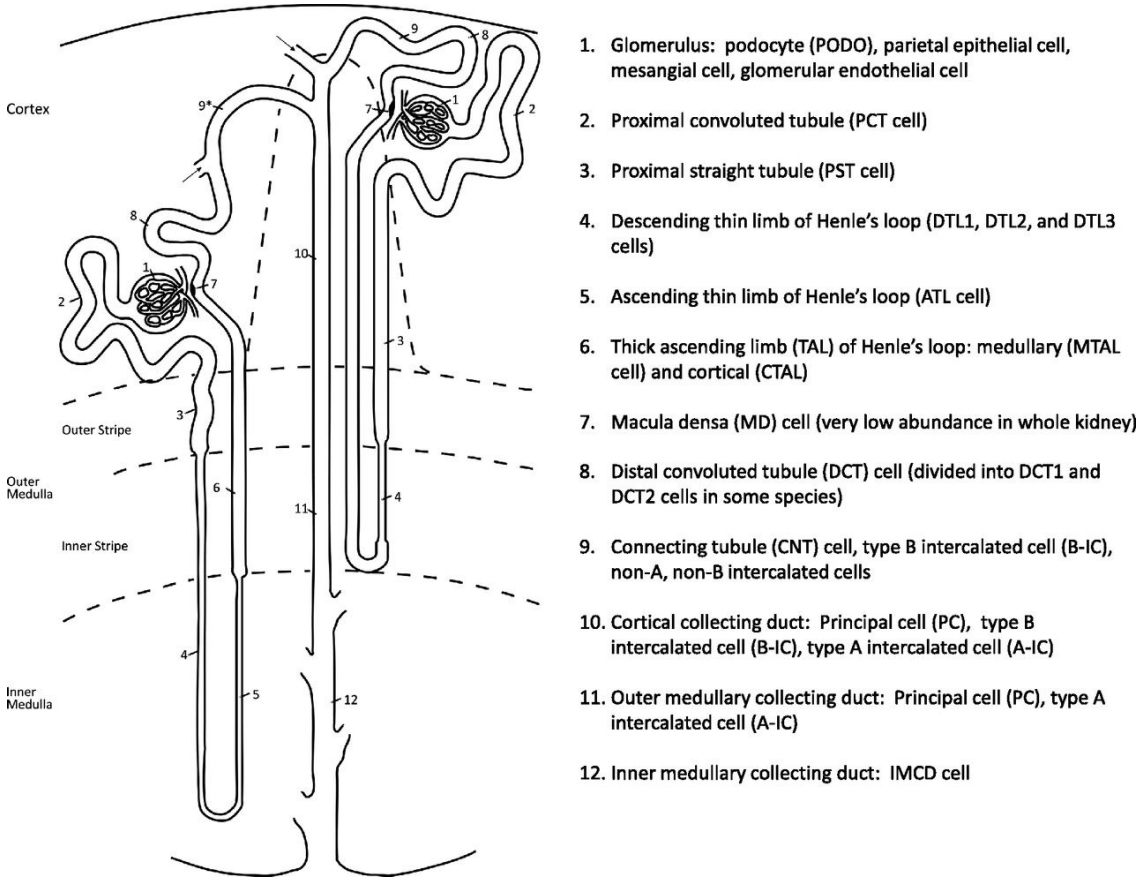


Figure 28. Cell types composing the kidney nephrons.

From Chen et al. (2019)

B. Origin of renal cell carcinoma: the kidney nephron

RCCs originate from the kidney. This organ is responsible for homeostasis of the body by filtering the blood to remove wastes and to balance water, salt and pH levels. The kidneys also release renin, a critical component of the renin-angiotensin-aldosterone system which regulates blood pressure[138], and erythropoietin that stimulates red blood cell production[139]. The kidney is organized into functional units called nephrons, a human kidney contains a million nephrons on average though there is high individual variability in this number[140]. The kidney comprises an outer layer called the cortex and an inner part called the medulla, two types of nephrons exist depending on how far they extend into the medulla (Figure 27). The nephron is a complex structure composed of around 20 cell types[141] that can be cells of origin for different RCC subtypes (Figure 28).

a. Cellular composition of the nephron

Blood filtration occurs in a cortical part of the nephron called the glomerulus which is a heavily branched vascular network of endothelial cells, pericytes called "mesangial cells" and vascular smooth muscle cells. The glomerular capsule known as "Bowman's capsule" contains mesangial cells along particular types of epithelial cells: the podocyte and its progenitor called the parietal epithelial cell. The podocyte has a very unique shape with actin-based "arms" that forms channels regulating the flow of fluid that is passed down the next part of the nephron[142].

The closest part to the glomerulus is formed by proximal tubules (PTs). These come in two types, convoluted (PCTs) and straight proximal tubules (PSTs). These cells are specialized in absorption of important molecules like amino acids, sugars and minerals. Their transcription program reflects this function with expression of membrane ion channels and transporters that require a lot of energy to function. These cells are therefore highly metabolically active and prone to generate an immune response in case of acute or chronic injuries. Because of this, they are implicated in many pathologies including renal fibrosis in which they will undergo EMT[143]. These cells are also the suspected cell of origin for ccRCC, pRCC[144] and tRCC[145].

The next part of the nephron is the loop of Henle that descends into the medulla. This structure is composed by descending thin limb (DTL), ascending thin limb (ATL), thick ascending limb (TAL), macula densa (MD) and distal convoluted tubule (DCT) cells. These cells are implicated in urine concentration mechanisms with distal tubules absorbing calcium and sodium back into the blood and controlling renin release[146]. DCTs are one of the suspected cell of origin for the rare but very aggressive collecting ducts carcinoma (CDC)[147].

The loop of Henle then connects to the collecting duct system that funnels urine out of the kidney to the ureter then to the bladder. This system is composed of connecting tubules (CNTs), intercalated cells (ICs), principal cells (PCs) and inner medulla collecting duct (IMCD) cells. These cells will secrete acid, bicarbonate and absorb chloride therefore they express a distinct set of transporters and channels[148]. The collecting ducts are the putative origin of chRCC[133], some subtypes of CDC[149], and, until recently, of renal medullary carcinoma (RMC)[150].

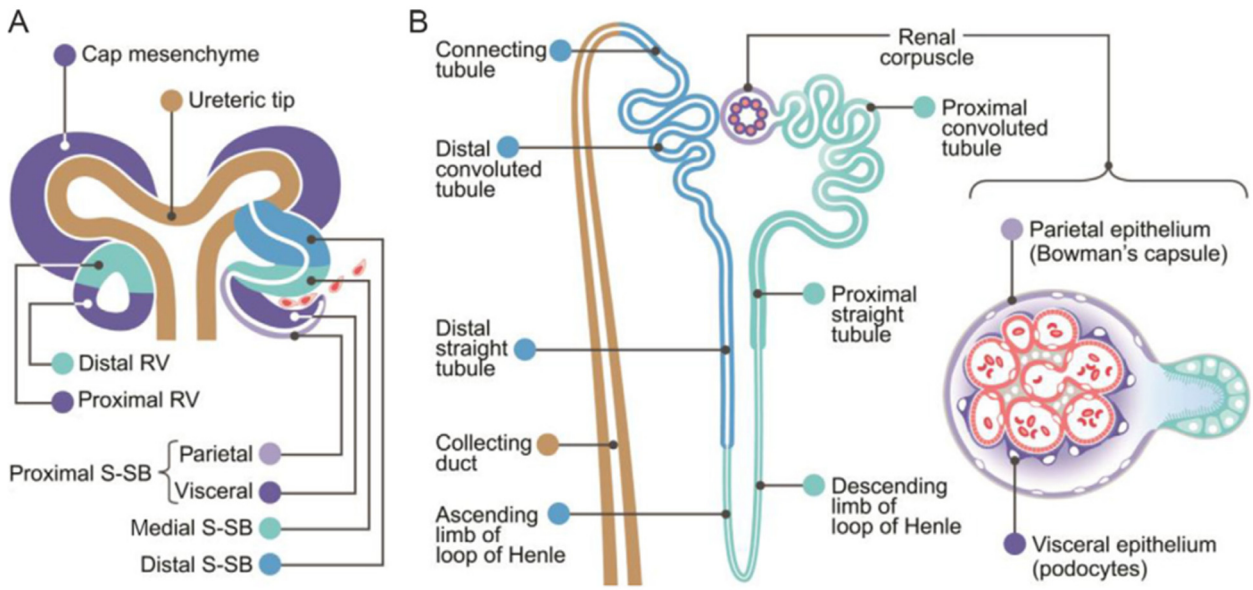


Figure 29. Development of the kidney nephrons in mice.

From McMahon et al. (2016)

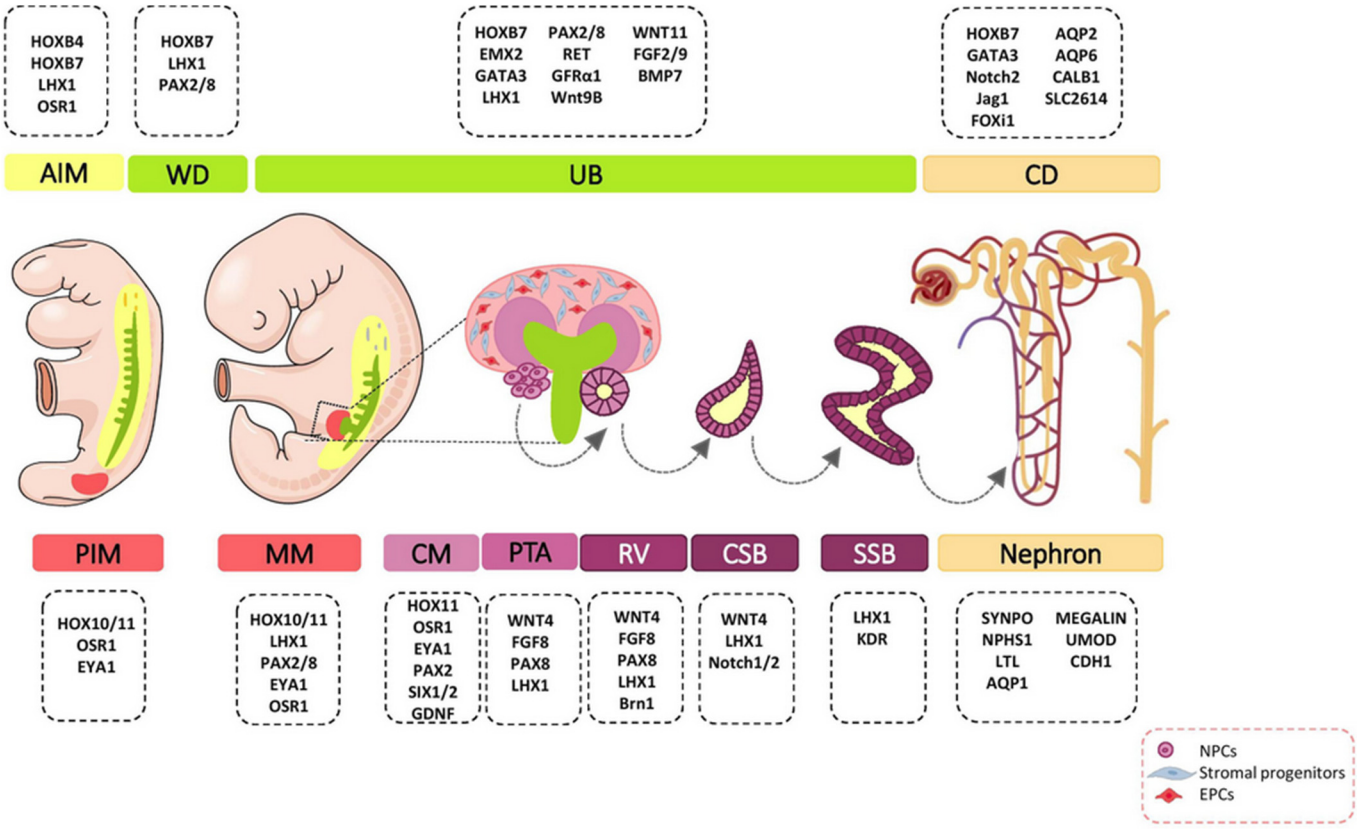


Figure 30. Molecular determinants of nephron development.

From Khoshdel et al. (2020)

b. Transcriptional control of kidney development, identity and regeneration

In mammals, kidney development has been studied mostly in mice and rats. It starts from three structures: the transient pronephros and metanephros which are followed by formation of the metanephros derived from the posterior intermediate mesoderm (PIM) that eventually forms the definitive kidney[151]. Mature kidney development starts when metanephric mesenchymal (MM) cells induce formation of the epithelial outgrowth called the uretic bud (UB) which will branch into a T-like structure. The UB will then continue to branch out for multiple generations thus establishing the collecting duct system. The mesenchymal capping cells (CM) will form pre-tubular aggregates (PTAs) that undergo MET to become renal vesicles (RVs) which will eventually give rise to nephrons. The RVs transform into comma-shaped bodies (C-SBs) then into S-shaped bodies (S-SBs). Different parts of the S-SBs will form the mature nephron with the proximal end generating the glomerulus by initiation of podocyte differentiation and endothelial cell migration through angiogenesis signals. The rest of the S-SB cells will differentiate according to their relative locations to the glomerulus, proximal cells will form the Bowman's capsule, middle cells will form proximal tubules and Henle's loop epithelium, and distal cells will become distal tubule cells[152] (Figure 29).

Patterning and cell specification during nephrogenesis are orchestrated by a large number of regulators and transcription factors. Paired-box (PAX) family members (PAX2/PAX8), homeobox (HOX)-11 (HOX11/*TLX1*), GATA-3, SOX8 and SOX9 are involved in activation of glial-cell-derived neurotrophic factor (GDNF) for development of the UB. CM stem-cell identity is kept by sine oculis homeobox 2 (SIX2) that will be repressed after activation of Wnt signaling for PTA specification. Further epithelization is heavily regulated by Wnt and Notch signaling pathways. In the S-SB, Wilms tumor 1 transcription factor (WT1) and musculoaponeurotic fibrosarcoma basic leucine zipper transcription factor B (MAFB) specify podocyte identity, hepatocyte nuclear factor 1 homeobox B (HNF1B) and BRN1 (*POU3F3*) define proximal tubule and Henle's loop identities while LIN11-Isl1-MEC3 domain homeobox 1 (LIM-1/*LHX1*) and SOX9 help establish distal tubules. Lastly, interstitial cells like mesangial cells are formed from progenitor cells governed by FOXD1[151,152] (Figure 30). Some of these transcription factors are still expressed in the mature kidney and continue to play roles in the adult organism.

PAX2 and PAX8 are members of the PAX family of proteins that diverged from homeobox genes and can promote cell migration, survival and proliferation[153]. PAX2 and PAX8 are expressed very early in kidney development and they have redundant roles as only loss of both factors leads to failure of nephritic structures formation[154]. Their expression is maintained in adult kidney epithelium but their role is poorly understood in this context. One study reported that they regulate solute carriers like urea transporters and aquaporins therefore controlling the water retention and urine concentration roles of collecting ducts. In this same study, PAX8 was shown to activate expression of solute carrier family 41 member 3 (SLC41A3) by recruiting a histone methyltransferase complex to its promoter[155]. As PAX8 is expressed in the adult renal epithelium, its expression was found to be somewhat conserved in all RCCs and can serve as a good target for immunohistochemical staining to diagnose tumors of kidney origins[156].

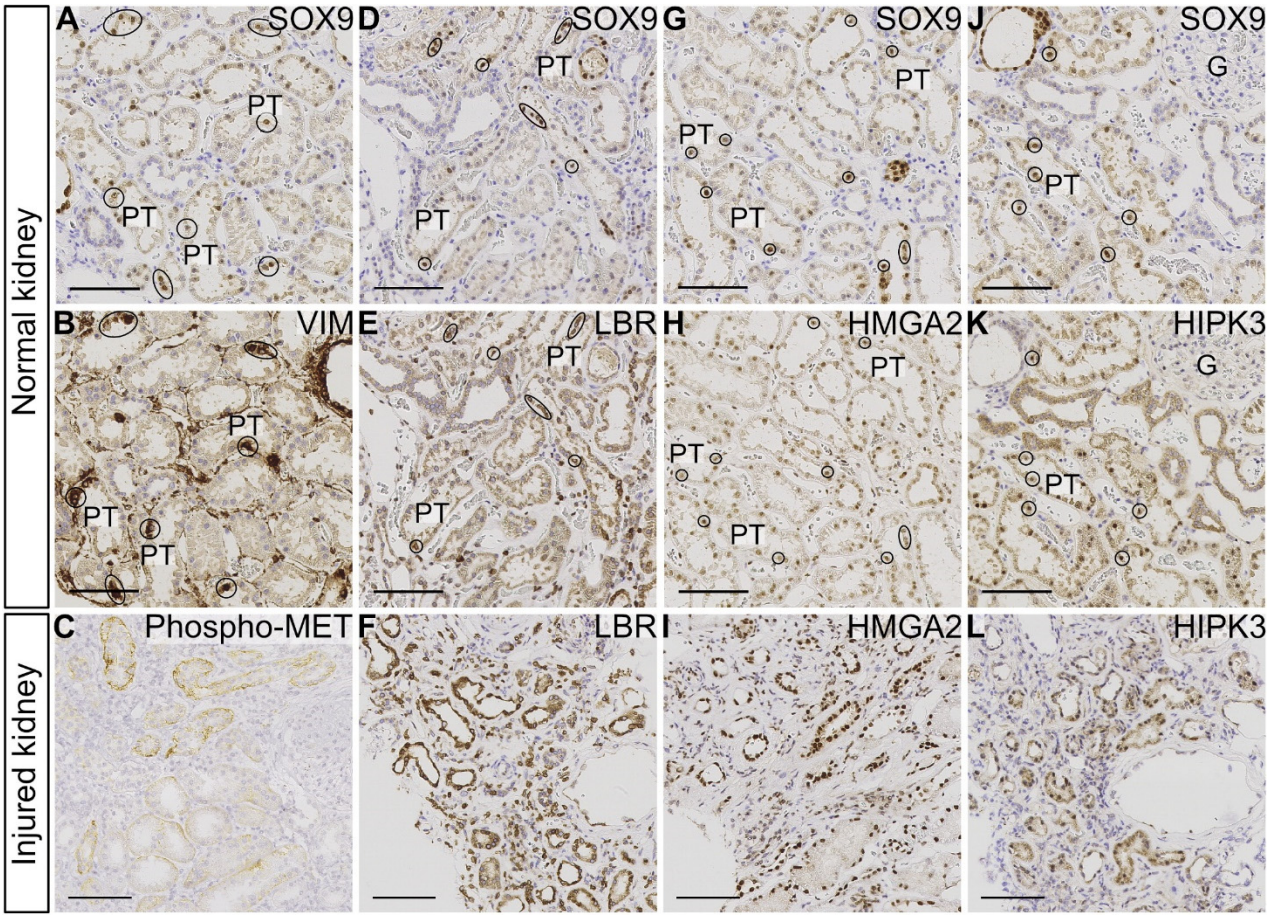


Figure 30. Staining of SOX9 and its target genes in normal and injured kidney.

From Kha et al. (2023)

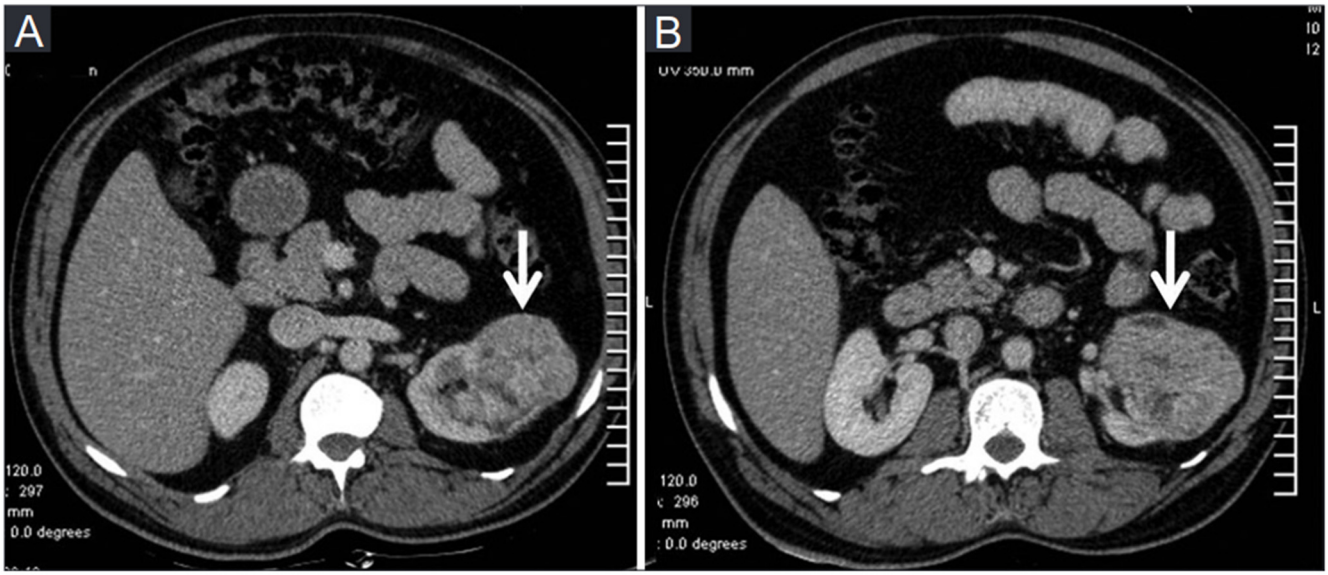


Figure 31. Computerized tomography scan of a ccRCC tumor.

From Mittal et al. (2016)

BRN1 is closely related to BRN2 mentioned in the melanoma section as both are class III of POU transcription factors with very high homology of their DNA binding domains. BRN1 is expressed during kidney development and is essential for the formation of Henle's loop, DCTs and MD. Its expression is maintained in adult TAL cells where it was shown to control expression of key genes such as uromodulin (UMOD)[157], a protein participating in apical membrane organization and secreted into the urine with anti-infection functions[158]. BRN1 is also expressed during development of the central nervous system where it was shown to interact with proteins of the SOX family, in glial cells specifically it was shown to synergize with SOX4 and SOX11[159].

The SOX family of proteins also plays a role in kidney with members such as SOX8 and the SOX9 factor that was mentioned in the melanoma section as a regulator of the mesenchymal-like state. During kidney development, SOX9 participates in formation of the UB and is essential for ureter branching. Its expression is maintained in mature PTs where it was shown to have a role in regeneration after injury. High SOX9 expression occurs after injury, but whether it is induced or a set of scattered kidney tubular stem cells exist is currently debated. $SOX9^{high}$ cells have increased motility, can repopulate PTs after injury and display multi-lineage differentiation capacity[160]. Knockdown of SOX9 in PTs identified lamin B receptor (LBR), high mobility group AT-Hook 2 (HMGA2), homeodomain interacting protein kinase 3 (HIPK3) and c-MET as potential target genes (Figure 30). LBR is involved in liver regeneration, its knockdown hinders cell proliferation and induces senescence. HMGA2 is a cell cycle regulator with roles in differentiation and proliferation during development. HIPK3 is a negative regulator of apoptosis and c-MET is a multifunctional receptor for hepatocyte growth factor (HGF) with an established role in development and tissue repair. SOX9 was shown to be activated downstream of Notch and TGF-β signaling[161].

C. Clear-cell renal cell carcinoma

In this section, we will focus on ccRCC which is the most frequent and one of the most aggressive subtypes of RCC. Known risk factors for development of ccRCC include smoking, hypertension, alcohol consumption, obesity, diet, von Hippel-Lindau (VHL) disease[162], chronic kidney disease, hemodialysis, renal transplantation[163] as well as single nucleotide polymorphisms (SNPs) in SWI/SNF subunit double plant homeodomain fingers 3 (*DPF3*)[164] and bHLH family member E41 (*BHLHE41*). In rare occurrences, patients with Birt-Hogg-Dubé syndrome caused by mutation in the folliculin (*FLCN*) gene[165] and patients with tuberous sclerosis (TSC) caused by mutations in the TSC complex subunits tumor suppressor genes (*TSC1*/*TSC2*)[166] have developed ccRCC.

Macroscopically, ccRCC present as yellow globular lesions growing from the kidney cortex with varying amount of necrosis, cystic changes and hemorrhage[167] (Figure 31). The tumors show an expansive pattern of growth with pushing borders. They are usually surrounded by a fibrous true or pseudo capsule and can sometime show infiltration of adjacent tissue. At the cellular level, ccRCC cells possess clear eosinophilic granular cytoplasm hence the name "clear cell" of the disease. This appearance is due to metabolic changes causing accumulation of cytoplasmic lipid and glycogen rich droplets also conferring the tumor its yellow color[168].

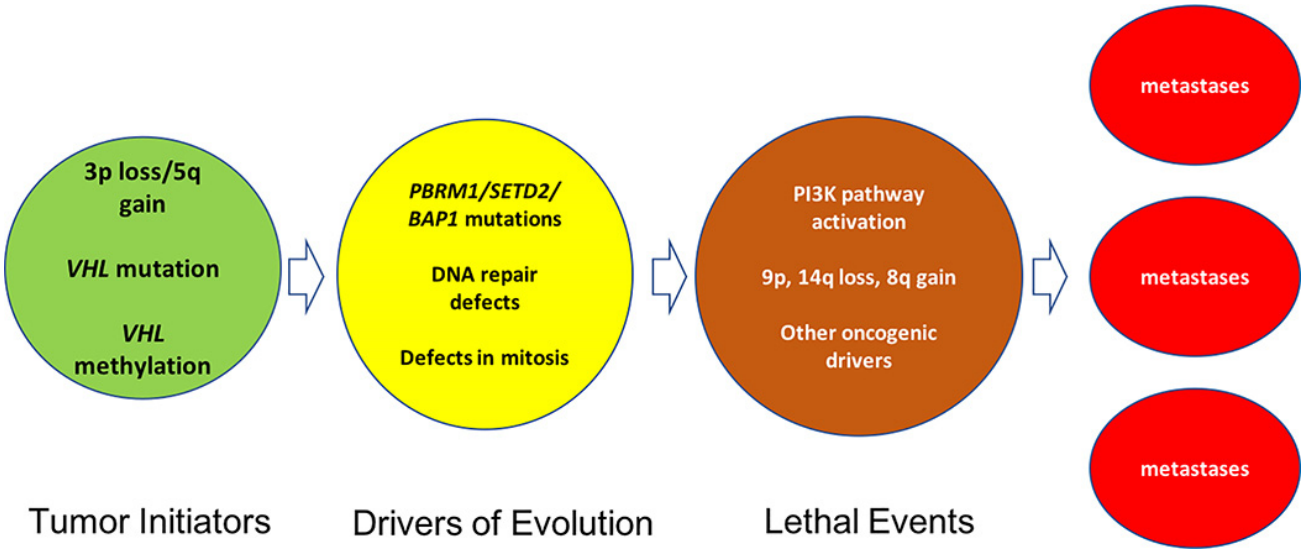


Figure 32. Events leading to clear-cell renal cell carcinoma progression.

From Jonasch et al. (2021)

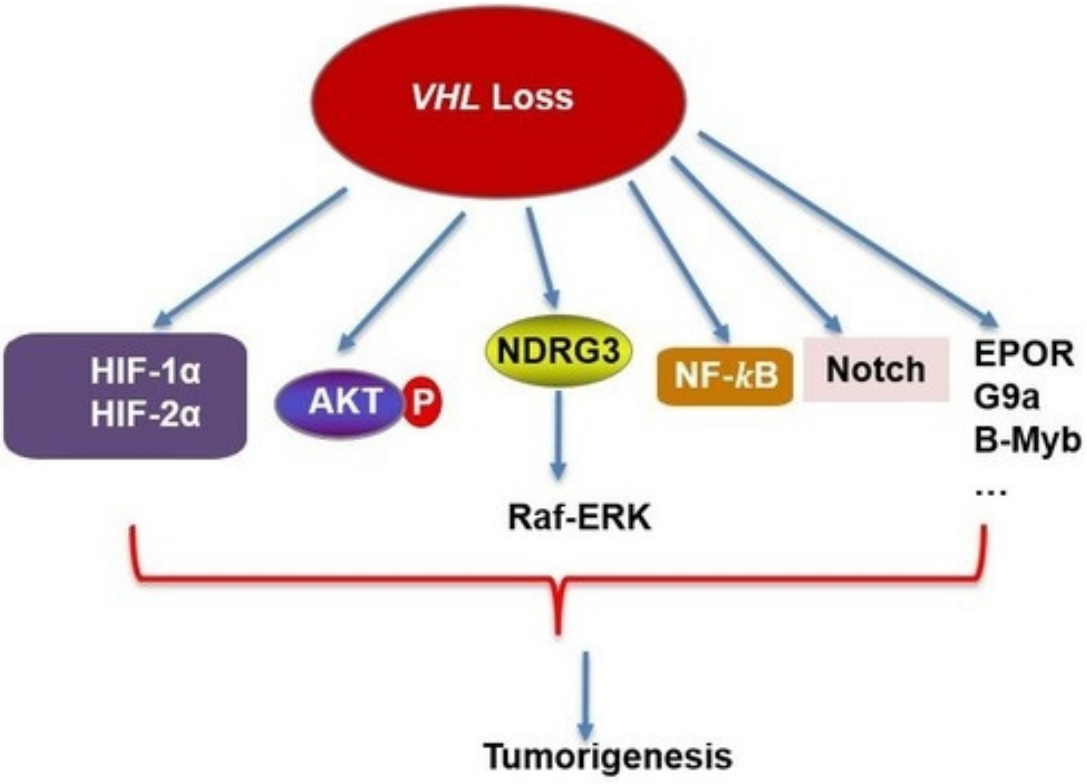


Figure 33. Pathways altered after *VHL* loss in clear-cell renal cell carcinoma.

From Zhang et al. (2018)

a. Oncogenesis of clear-cell renal cell carcinoma

Several genomic studies have identified molecular alterations leading to malignant transformation of proximal tubules in ccRCC. Almost all tumors display loss of the von Hippel-Lindau (pVHL/*VHL*) tumor suppressor gene due to point mutations, DNA methylation or loss of chromosome 3p. Loss of *VHL* by its own is not sufficient to initiate tumorigenesis as demonstrated by experiments in mice[169,170] and by the fact that it is always found in combination with other mutations in human tumor data. These recurrent secondary driver mutations include loss of polybromo 1 (BAF180/*PBRM1*), histone lysine methyltransferase SET domain containing 2 (*SETD2*) and breast cancer type 1 susceptibility protein 1 associated protein 1 (*BAP1*) which are all located on chromosome 3p. Outside of chromosome 3, other notable recurrent mutations are loss of lysine demethylase 5C (*KDM5C*), *TP53* and mechanistic target of rapamycin kinase (*MTOR*)[171,172] (Figure 32).

One study showed that chromothripsis, a mutational phenomenon causing massive clustered genomic rearrangements, is one of the drivers of chromosome 3p loss[173]. After mitosis, DNA can acquire a nuclear envelope thus creating a micronucleus. Inside this structure, the caustic environment is prone to cause DNA damage and chromothripsis. Formation of micronuclei is favored by genotoxic agents and by cellular environmental conditions such as hypoxia and oxidative stress which the proximal tubules are subjected to[174]. Loss of 3p is one of the earliest events in ccRCC oncogenesis, we will oversee the consequences of the different molecular alterations below.

i. The VHL tumor suppressor

The pVHL multiadaptor protein has over 30 binding partners identified. It has E3 ubiquitin ligase activity which marks target proteins for proteasomal degradation. The best known targets are the two subunits of the hypoxia-inducible transcription factor (HIF): HIF-1α (*HIF1A*) and endothelial PAS domain protein 1 (HIF-2α/*EPAS*) which are marked for degradation in normoxic conditions. It has been shown that pVHL regulates several other pathways including the phosphoinositide-3-kinase–protein/protein kinase B (PI3K/Akt), RAF/ERK, NF-κB, Notch signaling and erythropoietin receptor (EPOR) axis[175] (Figure 33).

Under normoxia, pVHL flags N-Myc downstream-regulated gene 3 protein (NDRG3) for degradation. Accumulation of NDRG3 after *VHL* loss or during hypoxia leads to activation of the RAF/ERK pathway through interaction with CRAF (*RAF1*) therefore triggering cell growth and angiogenesis. NDRG3 acts as a lactate-sensor that is stabilized by its binding and represents an alternative way to trigger an hypoxic response independently of HIF[176].

Another pathway normally activated under hypoxia is the PI3K/AKT/mTOR pathway. Phosphorylation of protein kinase B also named AKT serine/threonine kinase 1 (AKT/*AKT1*) at Thr308 and Ser473 leads to an active state which stimulates mTORC1 leading to metabolic changes promoting cell growth and proliferation[177]. AKT is a binding target of pVHL, its loss in cancer cells favors a metabolic switch toward glycolysis[178] and more generally PI3K/AKT/mTOR pathway activation is associated with more aggressive tumors in many cancers[179]. In ccRCC, mTOR activation seems to be a necessary step for oncogenesis[169].

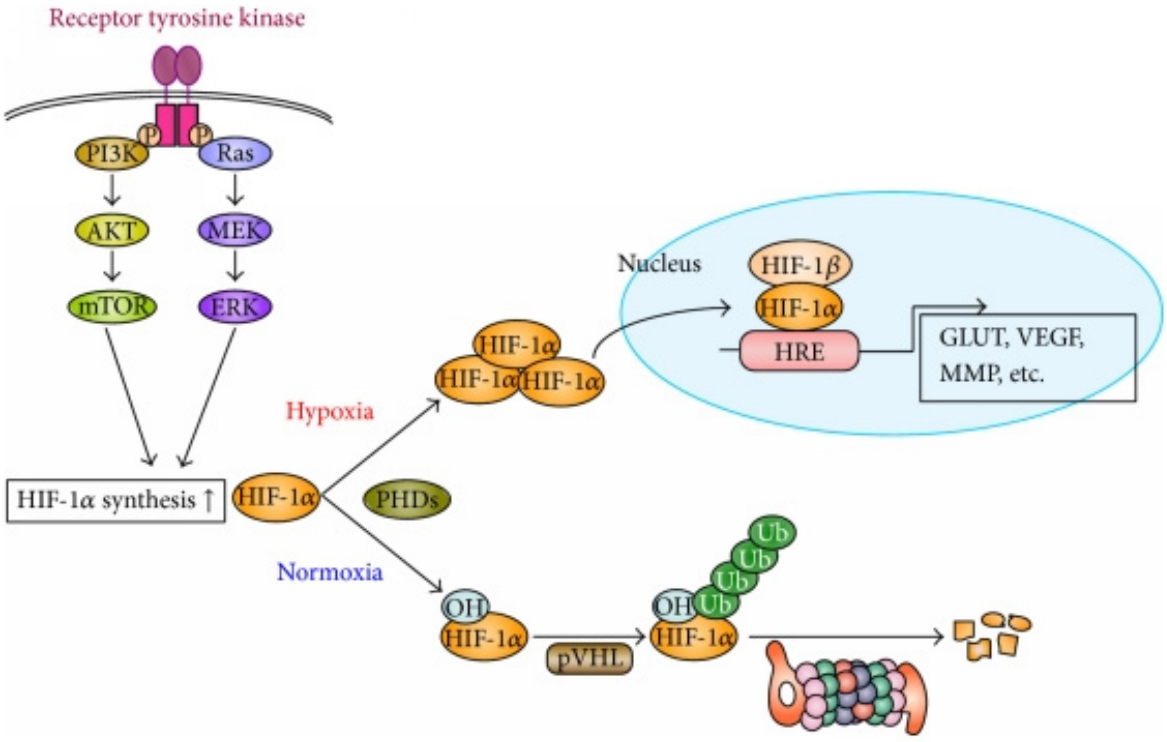


Figure 34. Regulation of hypoxia inducible transcription factor.

From Ogawa et al. (2015)

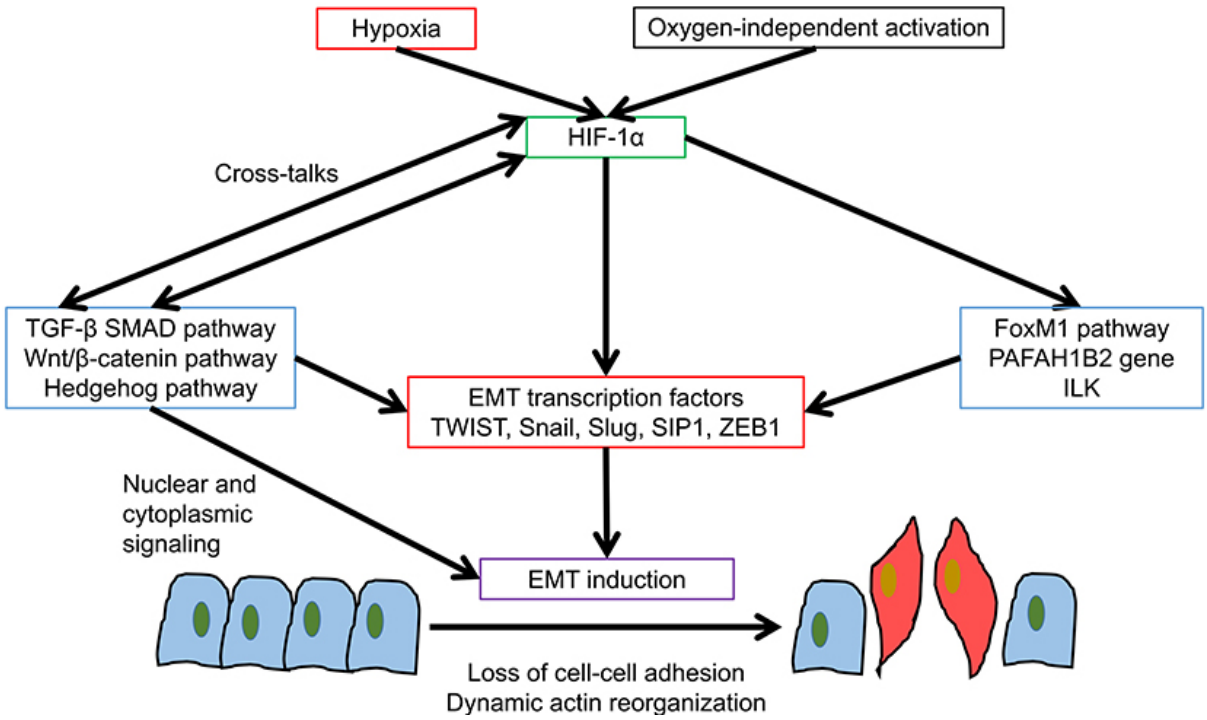


Figure 35. Hypoxia-induced EMT.

From Tam et al. (2020)

In resting cells, the transcription factor NF-κB is normally associated with inhibitor of nuclear factor kappa B kinase (IKBKB), a pVHL target, which restricts NF-κB binding to DNA[180]. Loss of *VHL* thus leads to activation of NF-κB mediated inflammation which is also due to accumulation of zinc fingers and homeoboxes 2 (ZHX2) normally targeted by pVHL[181]. Activation of this pathway makes ccRCC cells more resistant to cytotoxic agents[182].

Other pathways notable pathways altered by *VHL* loss include deregulation of normal ECM development because of pVHL binding to FN1[183], deregulation of microtubule stability[184] and mitotic spindle formation leading to genomic instability[185].

ii. The HIF transcription factor

The most notable consequence of VHL loss is accumulation of HIF subunits HIF-1α and HIF-2α. HIF is a heterodimeric transcription factor composed by either HIF-1α or HIF-2α with HIF-1β (*ARNT*). When oxygen is available, prolyl hydroxylases (EgIN/PHDs) will promote hydroxylation of HIF proline residues which will be subsequently marked for degradation by pVHL. In the absence oxygen where PHDs are inactive, or after *VHL* loss, HIF will dimerize and translocate to the nucleus where it binds the hypoxia response elements (HREs) motifs which are close to an E-BOX[186]. In this setting, HIF will drive adaptation to low oxygen conditions with production of signals inducing angiogenesis and promotion of a metabolism relying more on glycolysis than on OxPhos[187]. In ccRCC, HIF transcriptional program is activated even in normoxic conditions leading the cancer cells to adopt a state of pseudo-hypoxia (Figure 34). HIF directly regulates angiogenesis genes such as *VEGFA*, angiopoietins (eg: *ANGPTL4*) and MMPs (eg: *MMP2*)[188] causing ccRCC to be a highly vascularized tumor[189].

HIF-1α and HIF-2α regulate both a common and distinct set of genes. HIF-1α regulates the glycolysis program consisting of glucose transporters (GLUT-1/*SLC2A1*, GLUT-3/*SLC2A3*) and glycolytic enzymes. These enzymes are: phosphofructokinase liver type (*PFKL*), aldolase (*ALDOA*), phosphoglycerate kinase 1 (*PGK1*), enolase (*ENO1*, *ENO2*, *ENO3*), pyruvate kinase M1/2 (PKM2/*PKM*)[190] and lactate dehydrogenase A (*LDHA*). Additionally, OxPhos is suppressed in two ways, first by reducing metabolites from the TCA cycle via pyruvate dehydrogenase kinase 1 (*PDK1*) and LDHA that use pyruvate to produce lactate, and second by reducing mitochondrial activity through downregulation of cytochrome oxidase subunit expression and inducement of mitochondrial autophagy[191]. In cancers such as ccRCC, cells use glycolysis over OxPhos even in normoxic conditions. This phenomenon has been termed the Warburg effect and, even though it produces less ATP per molecule of glucose, cancer cells gain an advantage through rapid ATP synthesis, availability of increased concentrations of glycolytic intermediates used for nucleotide and amino acid synthesis, or through disruption of nearby cells[192]. As mentioned in the EMT section, hypoxia is a type-3 EMT-inducing signal. HIF-1α cross talks with EMT pathways such as TGF-β, Wnt/β-catenin, Notch and NF-κB but can also directly modulate EMT transcription factor expression. For example, HIF-1α can bind *TWIST*, *SNAIL* and *ZEB1* through HREs in their proximal promoters and it is possible that HIF-2α can do the same[193] (Figure 35). HIF driven ECM remodeling with increased production of MMPs and FN1 also contributes to type-3 EMT induced by hypoxia[194].

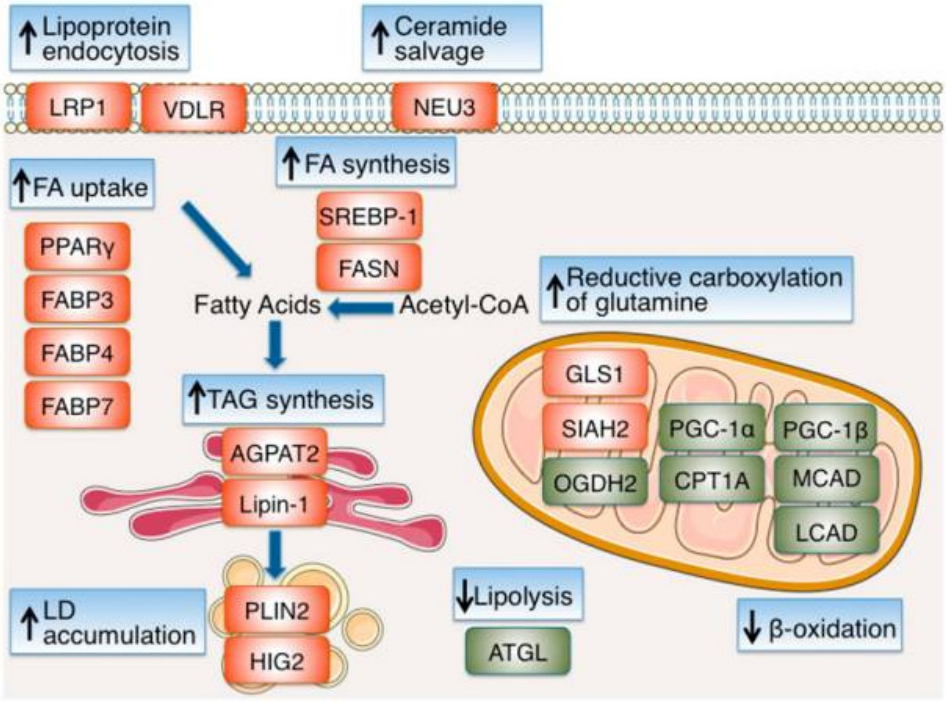


Figure 36. Regulation of lipid metabolism during hypoxia.

From Mylonis et al. (2019)

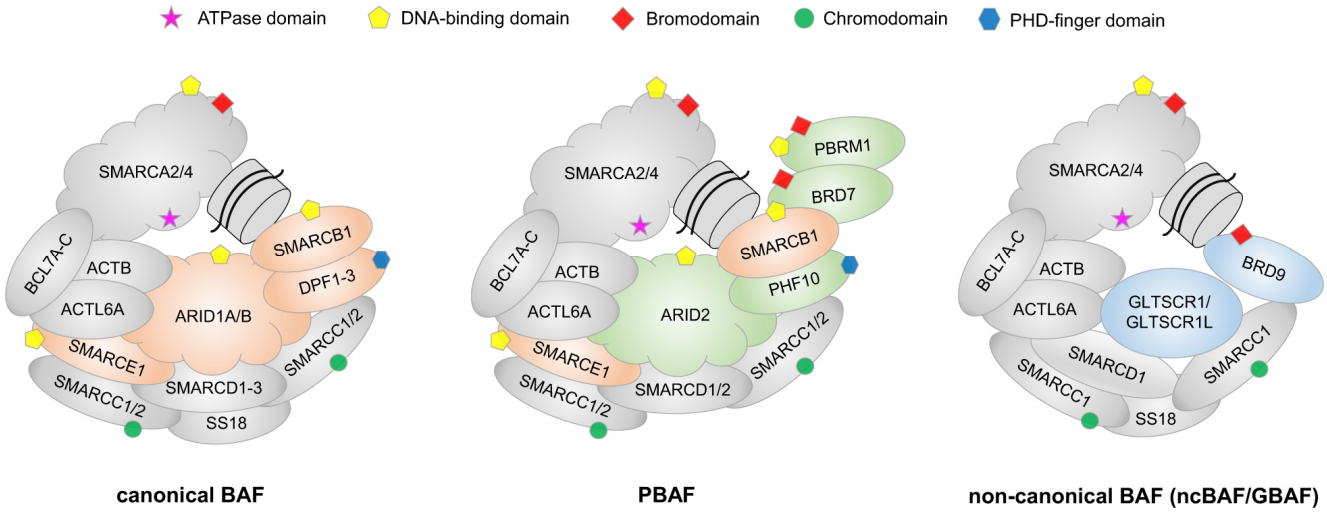


Figure 37. Composition of mammalian SWI/SNF complexes.

From Wanior et al (2021)

On the other hand, HIF-2α regulates protection against oxidative stress, cell cycle progression and vascular remodeling. Regulation of cyclin D1 (*CCND1*) by HIF-2α has oncogenic potential with inhibition of RB1 and altered control of G1/S progression. HIF-2α also regulates the erythropoiesis signal *EPO* and transforming growth factor alpha (TGF-α/*TGFA*) that promotes cell proliferation. Of note, HIF-2α regulates expression of POU class 5 homeobox 1 (OCT4/*POU5F1*), one of the four stem cell pluripotency Yamanaka factors alongside c-MYC, SOX2 and KLF4[195]. It was initially thought that HIF-2α rather than HIF-1α drives ccRCC oncogenesis but this assumption has been contradicted by other studies and mouse models showing that both factors might be required for tumor initiation[174].

Most of the observable characteristics and marker genes of ccRCC derive from the HIF regulated transcriptional program. Glucose metabolism regulation by HIF has been extensively studied but lipid metabolism is also altered in hypoxic conditions. Lipid oxidation for energy production occurs in mitochondria through β-oxidation enzymes that are stimulated by co-activators such as peroxisome proliferator activated receptor gamma coactivator 1 alpha (PCG-1α/*PPARGC1A*). This activity is heavily inhibited under hypoxia and by HIF that will also promote a program that uses lipids for other processes. This program will enhance fatty acid (FA) uptake, promote de novo fatty acid synthesis and transform them into triacylglycerols (TAGs) for storage. To protect cells against lipotoxicity, these TAGs will be stored into lipid droplets (LD) which confer ccRCC cells their characteristic appearance. This altered lipid metabolism supports tumor cell proliferation by protecting against harmful free radical formation and by creating a reservoir of building components for biogenesis of new membranes[196] (Figure 36). HIF regulated genes such as fatty acid binding protein 7 (*FABP7*) involved in fatty acid uptake[197], carbonic anhydrase 9 (*CA9*) involved in pH regulation[198], NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 4-like 2 (*NDUFA4L2*) involved in reduced mitochondrial oxygen consumption[199], *ANGPTL4*[200] and *VEGFA* are all marker genes for ccRCC cell transformation.

iii. The PBAF chromatin remodeler

The SWI/SNF subunit *PBRM1* is the second most frequently mutated gene alongside *VHL* in ccRCC with a mutation rate of over 40%[131]. SWI/SNF is one of the four class of ATP-dependent chromatin remodeling complexes with imitation switch (ISWI), chromodomain helicase DNA-binding (CHD) and inositol requiring mutant 80 (INO80). SWI/SNF complexes are composed of at least 29 proteins and always contain an ATPase subunit which is either Brahma (BRM/*SMARCA2*) or BRG1. These complexes are recruited by transcription factors and use ATP hydrolysis energy to reposition nucleosomes facilitating access to cis-regulatory elements such as transcription start sites for the transcriptional, DNA replication and repair machinery. SWI/SNF subunits have an overall mutation rate of 20% in cancer making this complex one of the most altered entities after TP53[201]. Three complexes have been described in mammals: BRG1/BRM-associated factor (BAF), polybromo-associated BAF (PBAF) and non-canonical BAF (ncBAF). The PBAF complex contains specific subunits like AT-rich interaction domain 2 (*ARID2*), bromodomain containing 7 (*BRD7*) and the PBRM1 subunit that confers the complex its name[202] (Figure 37). PBRM1 is a targeting subunit that contains 6 bromodomains capable of binding acetylated lysines on histones such as the active H3K27ac.

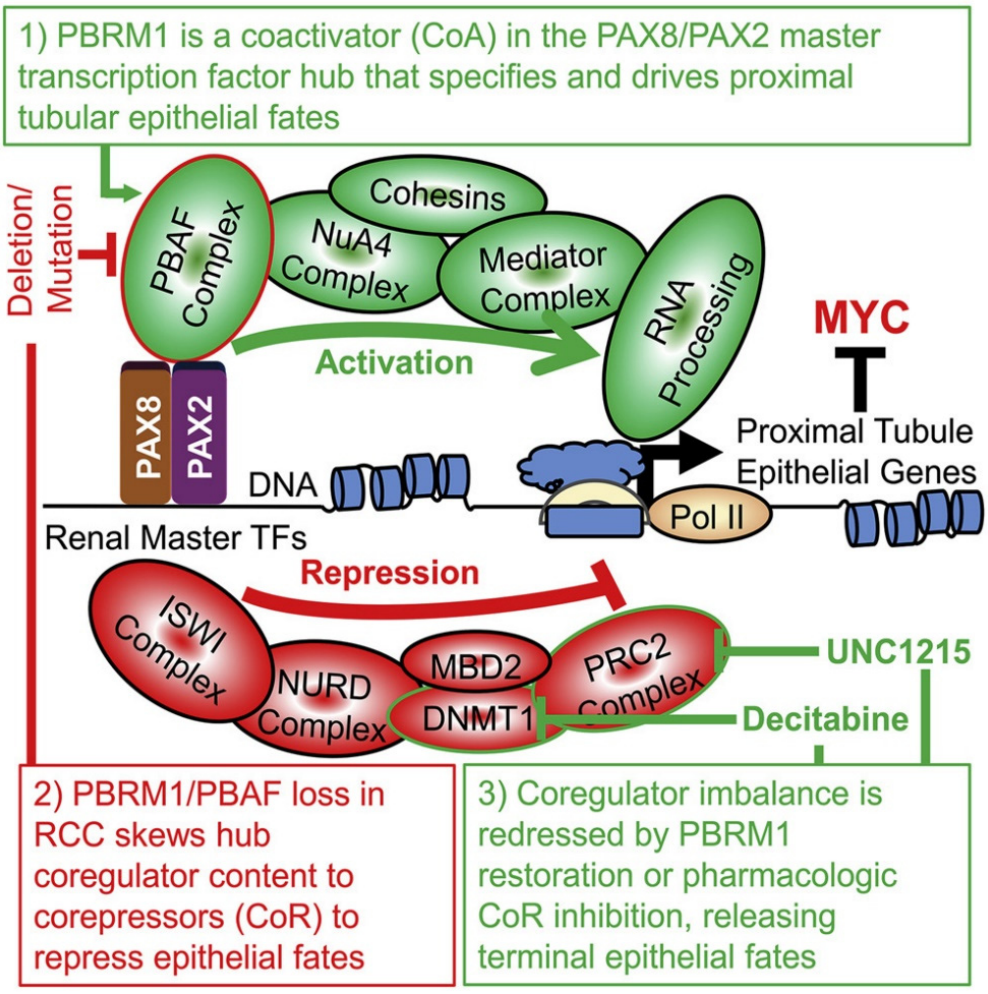


Figure 38. Control of epithelial identity is altered after PBRM1 loss.

From Gu et al. (2021)

The oncogenic consequences of *PBRM1* loss are not fully understood but several studies have provided possible mechanisms. One study showed that *VHL* deficient cells accumulate DNA damage upon replication leading to cell cycle arrest due to pVHL having a DNA maintenance role during conditions of replicative stress. They show that this replication stress can be rescued by loss of *PBRM1* which restored growth and proliferation capacity of the cells by altering trimethylation of H3K9 (H3K9me3). In their model, only *VHL*/*PBRM1* double mutation conferred the cellular fitness advantage necessary for tumorigenesis[170]. Another study found that PBRM1 bromodomain 4 can bind TP53 through its K382ac. It was found that loss of this binding affects the ability of TP53 to transcribe the negative cell cycle regulator cyclin dependent kinase inhibitor 1A (*CDKN1A*). Therefore, *PBRM1* loss hinders TP53 tumor suppressor function and, in line with this result, SWI/SNF mutations seem to be mutually exclusive with TP53 mutations[203]. One last study showed that PAX2/PAX8 can recruit PBAF in order to transcribe epithelial terminal differentiation genes. In consequence, loss of *PBRM1* imbalances a PAX2/PAX8/GATA3/WT1 transcriptional hub leading to expression of co-repressors that allows an oncogenic program driven by c-MYC to take over[204] (Figure 38).

iv. The BAP1 deubiquitinase

The *BAP1* gene is mutated in around 10-20% of ccRCCs and encodes a deubiquitinating enzyme that can target H2AK119, O-linked N-acetylglucosamine transferase (*OGT*) and host cell factor C1 (*HCFC1*)[205]. BAP1 has tumor suppression implications in multiple cancers. For example, decreased levels of *OGT* and *HCFC1* after *BAP1* loss leads to myeloid neoplasia. Additionally, the polycomb repressor complex ubiquitinates H2AK119 to exert its repressor function and this effect is counteracted by BAP1[206]. In ccRCC, *BAP1* loss negatively affects microspherule protein 1 (*MCRS1*) expression which leads to an increase of chromosome instability[207]. In mice models, *VHL*/*BAP1* double deletion produced kidney neoplasms[208].

v. The SETD2 methyltransferase

The *SETD2* gene encodes yet another protein with chromatin modification capacity and is also mutated in around 10-20% of ccRCCs. It has methyltransferase activity specific for H3K36 and is responsible for depositing the active chromatin mark H3K36me3. Similar to *PBRM1* and *BAP1*, *SETD2* and H3K36me3 have implications in DNA repair and genomic stability. Loss of *SETD2* hinders DNA damage repair by preventing recruitment of homologous recombination promoting factors like TP53, replication protein A (RPA) and radiation sensitivity 51 recombinase (RAD51)[209]. Also, this loss increases genomic instability as SETD2 can methylate lysine 40 of the mitotic spindle alpha tubulin and failure to do so promotes formation of micronuclei. Of note, PBRM1 can bind the αTubK40me3 mark deposited by SETD2 indicating a functional convergence of these two factors. As of yet, no mouse model exist for VHL/SETD2 deficient ccRCC[174].

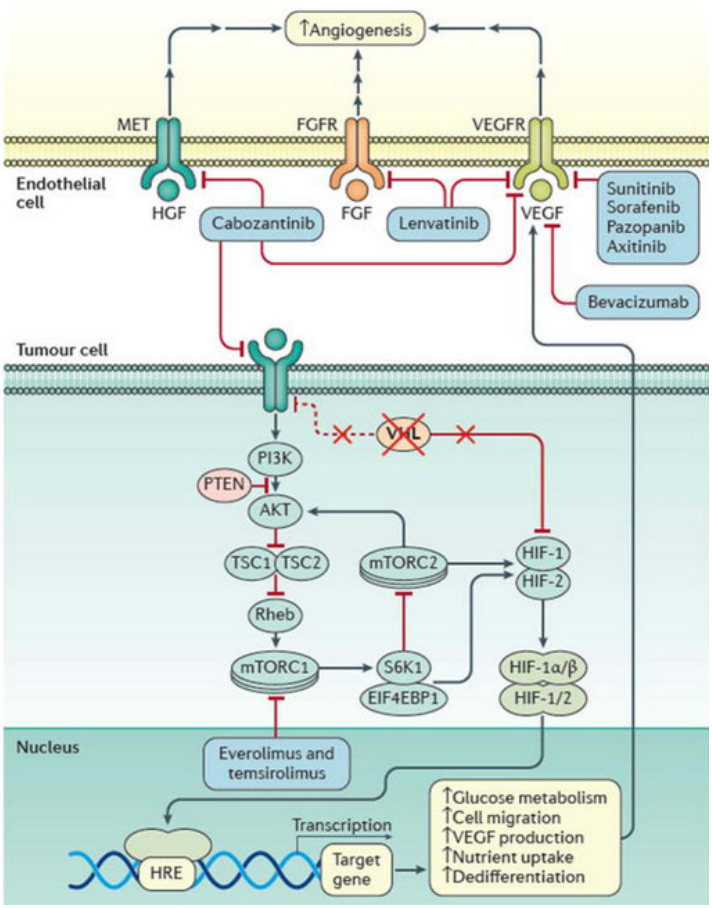


Figure 39. Targeted therapy with tyrosine kinase inhibitors in ccRCC.

From Hsieh et al. (2017)

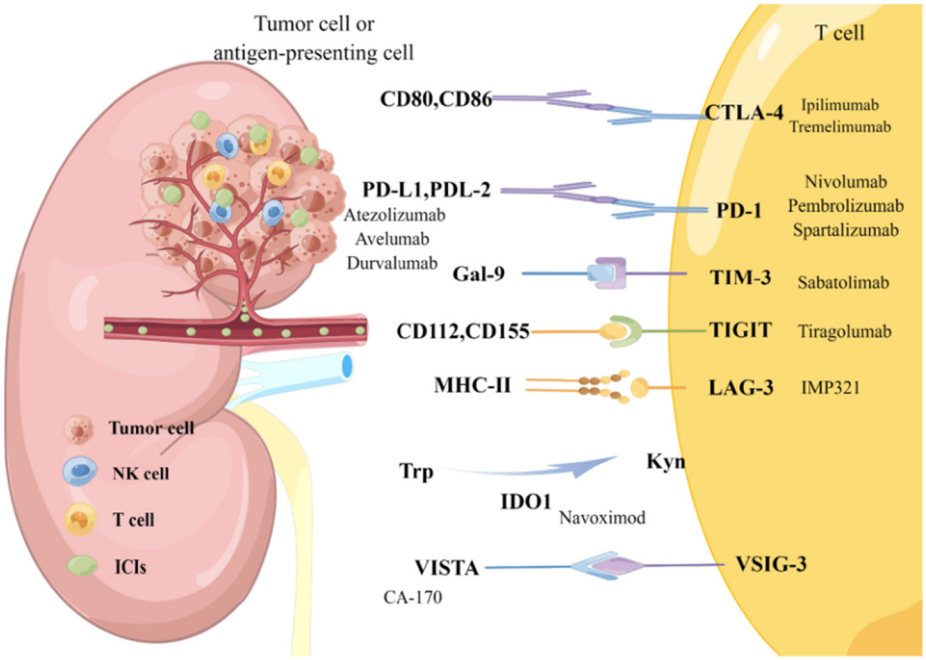


Figure 40. Immune checkpoint inhibition in ccRCC.

From Liu et al. (2022)

b. Standard of care

In the following section, we will overview clinical care of patients with ccRCC. The survival rate of patients with RCC is dependent on how early the disease is detected. For localized disease (the cancer has not spread outside the kidney) the 5-year survival rate is 93%, for regional disease (the cancer has spread in nearby structures or lymph nodes) the survival rate is 72% and for distant disease (the cancer has spread to distant organs) the survival rate drops drastically to 15%[210]. More than half of RCCs are discovered by chance as many patients remain asymptomatic for a long time. Only 30% of patients are diagnosed with RCC based on symptoms, the main ones include flank pain, hematuria and abdominal mass. As a result, about 20-30% of patients are metastatic at the time of diagnosis[211]. For local disease, surgical resection remains the best option, usually radical nephrectomy which removes the entire kidney or nephron-sparing surgery which tries to remove only the afflicted tissue while preserving the normal kidney as best as possible. For metastatic forms, current treatment options include targeted drug therapy with tyrosine kinase inhibitors (TKIs) and immune checkpoint inhibitors (ICIs)[212].

i. Tyrosine kinase inhibitors

Because of HIF activation leading to overexpression of angiogenesis signals, inhibition of the VEGF axis has been the main treatment strategy for targeted therapy. TKIs are multi-target drugs that always inhibit kinase insert domain receptor (VEGFR-2/*KDR*) alongside up to six other RTKs such as PDGF receptor (PDGFR), FGFR, c-MET, BRAF or AXL. Sunitinib (sutent) has been a reference ccRCC treatment for a long time, it targets Fms related receptor tyrosine kinase 1 (VEGFR-1/*FLT*), VEGFR-2, PDGFR, FLT3 and KIT proto-oncogene receptor tyrosine kinase (KIT). Other molecules were approved or tested other the past years notably sorafenib, axitinib, cabozantinib, lenvatinib and pazopanib[213,214] (Figure 39). Although TKIs were first tested for their effect on endothelial receptors that inhibit angiogenesis, evidence has shown that secondary RTK targets have important impacts on cells of the TME such as M2-TAMs[215].

ii. Immune checkpoint inhibitors

The second strategy for treating metastatic ccRCC involves immune checkpoint blockade (ICB). We introduced exhaustion mechanisms in the section discussing T-cells, the ICB strategy aims to inhibit immune checkpoint ligands produced by the tumor and TME or to block T-cell surface receptors for these ligands[216] (Figure 40). Nivolumab is an anti-PD-1 monoclonal antibody that has shown a manageable safety profile and could provide survival benefits in monotherapy[217]. Nivolumab can be combined with the anti-CTLA-4 antibody ipilimumab. Nivolumab with ipilimumab bi-therapy showed a better long-term response than sunitinib but also displayed more toxicity[218]. In the past years, TKI/ICI combinations have also shown efficacy, for example association of the anti-PD-L1 avelumab with the TKI axitinib was approved for treating ccRCC by the US food and drug administration[219].

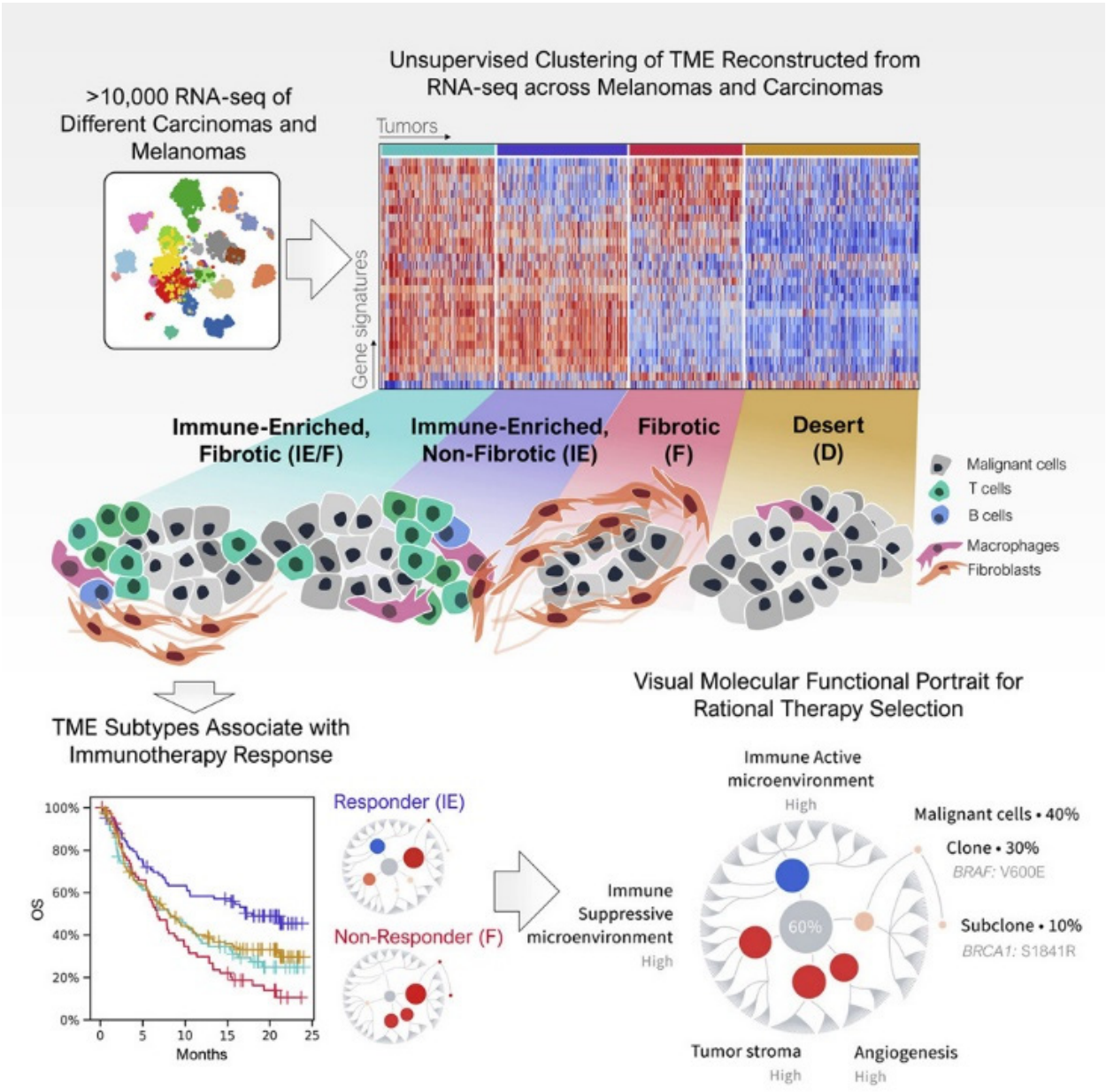


Figure 41. Association between TME subtypes and resistance to immunotherapy.

From Bagaev et al. (2021)

ICB has been used in many cancers and biomarkers for optimizing patient selection have been identified. These markers depend on the type of cancer and can either be intrinsic to cancer cells or to the TME. For example, in melanoma, a high tumor mutational burden (TMB) that may reflect the formation of neoantigens in cancer cells correlates with positive response to ICB. Positive biomarkers rather associated with the TME include high infiltration of CD8 T-cells with low amounts of Tregs, presence of TLS and high expression of an IFN-γ response gene signature. The TME composition can also constitute a biomarker of resistance to ICB, one computational study showed that an immune-cold and fibroblastic TME is associated with poor response to ICIs[220] (Figure 41). CAFs have been implicated in resistance to immunotherapy because of their ability to remodel the ECM, crosstalk with other cell types and produce TFG-β which is associated with immune exclusion of effector CD8 T-cells[221,222]. In ccRCC however, a sizable proportion of patients do not respond to ICIs and no robust biomarkers have been identified yet[223].

iii. The staging system in ccRCC

Choice of treatment depends of the advancement of the tumor which is measured by a staging system as follows[224]:

- T1: the tumor is limited to the kidney and measures less than 4cm (T1a) or range from 4-7 cm (T1b).
- T2: the tumor is limited to the kidney and its size ranges from 7-10 cm (T2a) or is over 10 cm (T2b).
- T3: the tumor has extended into the renal vein, the renal pelvis, or the perirenal sinus fat (T3a), the vena cava below the diaphragm (T3b), the vena cava wall or the vena cava above the diaphragm (T3c).
- T4: the tumor has invaded beyond the renal fibrous capsule (gerota fascia) or into the ipsilateral adrenal gland.

The disease is staged using the tumor node metastasis (TNM) classification which is as follows:

- Stage I: the tumor is T1 with no lymph node/distant metastasis.
- Stage II: the tumor is T2 with no lymph node/distant metastasis.
- Stage III: the tumor is T3 with no lymph node/distant metastasis or the tumor is T1-T3 with regional lymph node metastasis.
- Stage IV: the tumor is T4 with no lymph node/distant metastasis or the tumor is T1-T4 with any distant metastasis.

For patients with Stage I to Stage III disease progression, surgical treatment is recommended. On the other hand, if patients present Stage IV disease, surgical resection of the primary tumor is possible but systemic treatment is usually recommended[225].

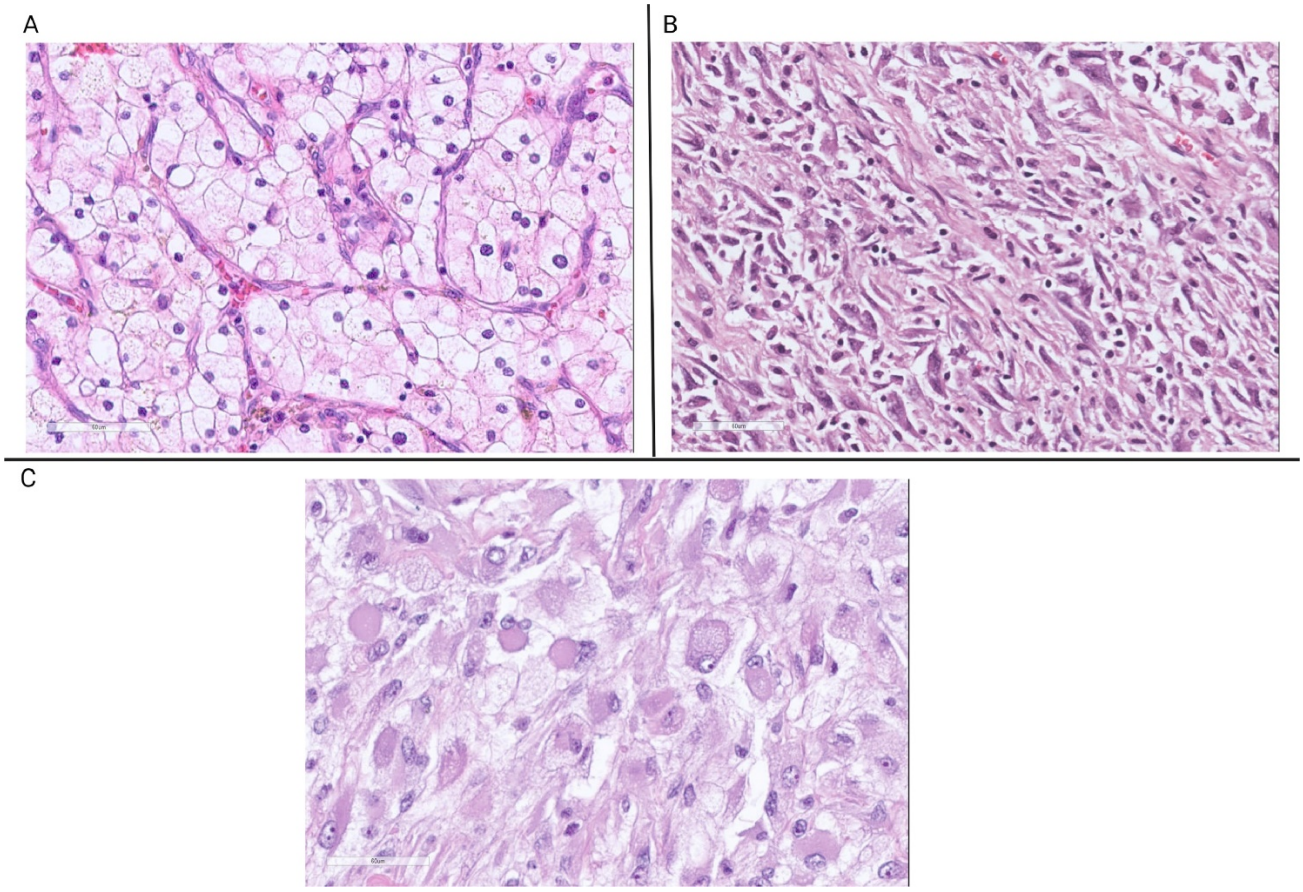


Figure 42. Epithelial (A), sarcomatoid (B) and rhabdoid (C) morphologies in ccRCC.

From Hahn et al. (2023)

c. Heterogeneity and classification of clear-cell renal cell carcinomas

i. Composition of the microenvironment

The distribution and type of cells present in the TME is a great source of heterogeneity among different tumors. Many studies have described the TME in ccRCC and have found it to be highly infiltrated by immune cells with a rich endothelial component making ccRCC one the most immune-hot solid tumor types[226]. One study quantified the immune microenvironment by flow cytometry which revealed it to be composed of around 50% T-cells, 30% myeloid cells, 9% NK cells and 4% B-cells. Among these, they could report many phenotypes with 17 TAM profiles that seemed to be associated with Tregs and exhausted cells[227]. Later, multiple single-cell RNA sequencing studies (scRNA-seq) found similar results[228,229] with one study in particular showing that tumors are gradually enriched in M2-TAMs and exhausted T-cells as the disease progresses from Stage I to Stage IV[230]. Immune cells have a particular spatial distribution with some tumors showing presence of TLS mediating a better immune response[231].

ii. Epithelial to mesenchymal transition in ccRCC

Intra-tumoral heterogeneity of cancer cells has been way understudied until recently. In the clinical setting, histological observations identified undifferentiated RCC cells with mesenchymal phenotypes. These phenotypes were termed sarcoma-like "sarcomatoid" and rhabdomyosarcoma-like "rhabdoid" in reference to the tumors of mesenchymal origins which they resemble. These cells present a spindle-shaped morphology (Figure 42) and show low expression of CDH1 with high expression of CDH2/VIM. These phenotypes can occur in any RCC subtypes and, in ccRCC, sarcomatoid dedifferentiation is detected in 5% of cases usually alongside the epithelial morphology. Some rare tumors can present both sarcomatoid and rhabdoid components. Presence of these cells is associated with a more aggressive disease, a high frequency of distant metastases but a better response to ICIs[232,233]. It was also possible to train a convoluted neural network to classify hematoxylin and eosin (H&E) stained slides into epithelial and mesenchymal subtypes[234].

Further *in-vivo* evidence of EMT in ccRCC was found by RNA-seq analysis from human tumors and mice models. Several studies mentioned identifying cancer cells with stemness features that were called "cancer stem cells" (CSCs). These cells were notably marked by expression of the polycomb gene *EZH2*[235]. Another study identified gene programs associated with invasion and metastatic competence, they find that early invasion is associated with AP-1 activation and distant metastasis are associated with PRRX1 gene expression[236].

Additionally, EMT could be detected in cellular models. Expression of SPARC related modular calcium binding 2 (*SMOC2*) pushed ACHN (pRCC cell line) and 786-O (VHL/TP53 mutated ccRCC cell line) towards a mesenchymal phenotype with increased protein expression of FN1, ACTA2 and VIM accompanied by decreased expression of CDH1[237].

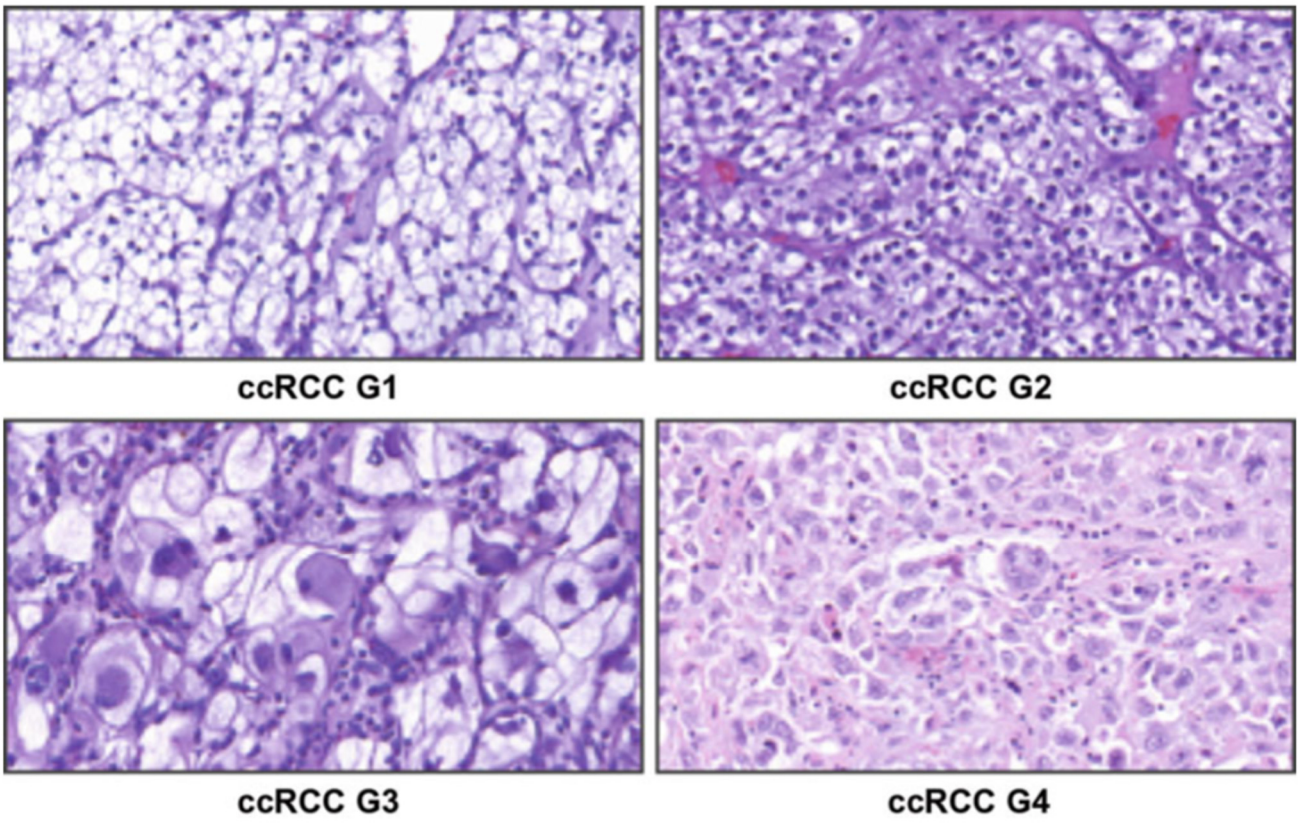


Figure 43. The four different grades of ccRCC tumor cells.

From Offermann et al. (2019)

iii. The grading system in ccRCC

While tumor stage evaluates the progression of the disease, the grading system aims to evaluate the aggressive nature of cancer cells based on their morphological features. In ccRCC the popular Fuhrman grading system based on the nuclear characteristics was adapted by the international society of urological pathology (ISUP) and then endorsed by the world health organization (WHO). The current WHO/ISUP grading system[238,239] (Figure 43) is as follows:

- ISUP grade 1 (ISUP1/G1): cells present absent or basophilic, inconspicuous nucleoli at ×400 magnification.
- ISUP grade 2 (ISUP2/G2): cells present conspicuous and eosinophilic visible nucleoli at ×400 magnification, but not prominent at ×100 magnification.
- ISUP grade 3 (ISUP3/G3): cells present conspicuous and eosinophilic nucleoli at ×100 magnification.
- ISUP grade 4 (ISUP4/G4): cells present rhabdoid or sarcomatoid differentiation, tumor giant cells and extreme nuclear pleomorphism with clumping of chromatin.

iv. Transcriptomic classifications of ccRCC

Several classifications of ccRCC were established based on mRNA expression but none is accepted as consensual. Initially, a dichotomy between two ccRCC subtypes (ccA/ccB) was established based on microarray expression data[240]. Later, the TCGA established a classification into 4 subgroups (m1 to m4) based on 446 patients. They report that the m1 subgroup has an association with chromatin remodeling and a higher frequency of mutations in *PBRM1*. The m3 subgroup which is associated with the worse survival in patients has higher frequency of *PTEN* and *CDKN2A* mutations. The m4 subgroup had higher mutation frequency of *BAP1* and *mTOR*[241].

Another classification into four ccRCC subgroups (ccRCC1 to ccRCC4) emerged from a French project aiming to characterize solid tumors. This classification was established based on 53 microarray samples from patients with Stage IV ccRCC and later, a classifier of 35 genes was built to easily assign future samples to one of the four subgroups. These groups showed associations with clinical response to TKI treatment and it was hypothesized that they might represent a tumor evolution gradient from cRCC3 to ccRCC2 to ccRCC1 then to ccRCC4. The ccRCC3 group was called “normal-like” as it resembled normal tissues and the ccRCC2 group was deemed “classical” as it was characterized by a high pro-angiogenic gene signature. The ccRCC1 group called “c-MYC up” showed tumor with low immune infiltration and less differentiation characterized by higher grade and stem-like features. The last group ccRCC4 called “c-MYC and immune up” contained tumor with high immune infiltration and also showed low differentiation including high grade and sarcomatoid tumors. Due to the differences in angiogenic gene signature and infiltration profile, it was speculated that these groups might respond differently to TKIs or ICIs. This hypothesis was tested in the BIONIKK clinical trial (NCT02960906) which evaluates nivolumab+ipilimumab against sunitinib in ccRCC3/ccRCC2 patients and against nivolumab in ccRCC1/4 patients[242].

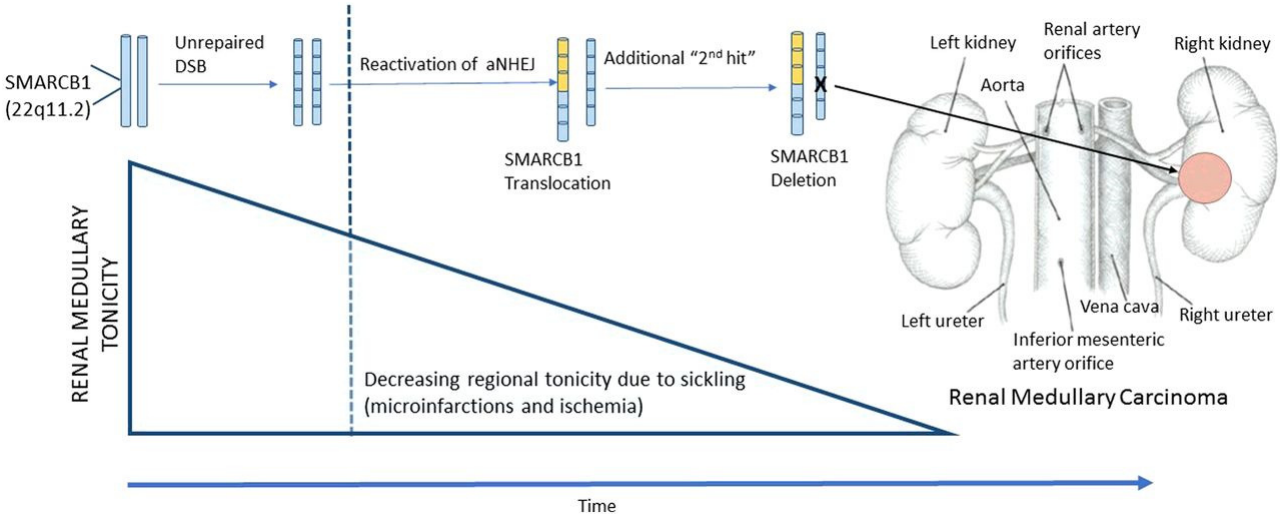


Figure 44. Model of RMC genesis.

From Msaouel et al (2018)

D. Renal medullary carcinoma

a. Clinical characteristics

In this section, we will briefly introduce one of the rare molecularly defined RCC now called SMARCB1-deficient renal medullary carcinoma (RMC). RMC accounts for less than 0.5% of RCC cases. It afflicts adolescent and young adult patients of African descent virtually all presenting either sickle-cell trait or sickle-cell disease. RMC is one of the most aggressive types of RCC with patients usually metastatic at presentation and a median survival time of 13 months. Common metastatic sites include liver, adrenal glands, lungs and regional lymph nodes. Not many treatment options exist, the recommended choice is chemotherapy such as high dose of MVAC (mix of methotrexate, vinblastine, doxorubicin, and cisplatin)[243].

b. Molecular alterations

RMC originates from the kidney medulla. The molecular hallmark and believed to be the driver mutation of RMC is loss of the SWI/SNF subunit SMARCB1 which is also observed in other pediatric tumors like malignant rhabdoid tumors (MRTs), atypical teratoid/rhabdoid tumors and epithelioid sarcomas. SMARCB1, a subunit of BAF and PBAF complexes, is a potent tumor suppressor with implications in cell differentiation, cell proliferation, DNA damage repair and DNA replication[244,245]. In MRTs, SMARCB1 loss causes SWI/SNF to be retargeted from typical enhancer elements maintaining differentiation to super-enhancers leading to deregulation of cell cycle progression[246]. Another recurrent alteration in RMC is gain of chromosome 8q leading to *MYC* amplification[247].

c. Oncogenesis of renal medullary carcinoma

Mechanisms leading to RMC genesis remain to be elucidated. The current model tries to tie in the origin of the disease in the kidney medulla with the sickle-cell anemia. This condition results from mutations in the hemoglobin gene leading to very fragile “sickle-shaped” red blood cells prone to hemolysis. The kidney medulla is thought to be the most hypoxic region in the body thus the current model of RMC development states that the abnormal red blood cells in these extreme hypoxic conditions cause local ischemia and microinfarctions. In turn, mobilization of alternative DNA repair pathways after downregulation of *RAD51* and *BRCA1* during hypoxia lead to conditions favoring deletions and translocations of *SMARCB1* located on the fragile chromosome 22[248] (Figure 44).

3. Objectives of the research

While several scRNA-seq studies of ccRCC were recently published, they heavily focused on analyzing immune cells, T-cells and macrophages in particular. Cancer cell and stromal heterogeneity is still mostly unknown in ccRCC therefore the main goal of my thesis is to unravel the plasticity of cancer cells and how it pertains to progression of the disease and resistance to treatment by immune checkpoint inhibitors. We will oversee the different steps of my project below.

A. Describe ccRCC cancer cell plasticity and TME at single-cell level

The first step of the project is to profile the composition of ccRCC tumors using scRNA-seq. For this we generated scRNA-seq for 7 ccRCC tumor samples and 2 normal adjacent tissues. Out of these, 2 tumors were also sequenced using spatial transcriptomics. Additionally, I had access to 97 RNA-seq samples from the BIONIKK clinical trial and 495 samples from the cancer genome atlas (TCGA). The aims of this step are:

- Establish transcriptomic profiles of cell types composing ccRCC tumors.
- Identify transcription factors governing cancer cell states and transition from normal kidney cells to neoplastic ccRCC cells.
- Identify axes of communications between cells by ligand-receptor interaction.
- Study spatial distribution of these cells.
- Establish tumor composition profiles from deconvolution of the large BIONIKK and TCGA cohorts.

B. Establish relationship between ccRCC tumor composition and patient survival

We saw in the introduction that tumor plasticity and stromal composition has an important effect on disease outcome however this has not been extensively studied in ccRCC. The TCGA dataset has sample clinical annotations with overall survival data and the BIONIKK dataset has survival data for patients treated by ICIs. Therefore, the aims of this step are the following:

- Discover cell populations associated with overall patient survival and disease progression in the TCGA dataset.
- Discover cell populations associated with resistance to ICIs in the BIONIKK dataset.

C. Identify new biomarkers for survival and resistance to immunotherapy in ccRCC

Immunotherapy has improved the treatment of metastatic ccRCC but it still has an overall response rate of about 30%. There is a yet unmet need for biomarkers in ccRCC, the goals of these steps are:

- Identify key genes from cell populations profiled in the previous steps that can serve as biomarkers for survival and resistance to ICIs.
- Use clustering approaches from lincRNA gene expression in the BIONIKK cohort to identify epigenetic biomarkers of treatment resistance.

D. Elucidate genesis and vulnerabilities of RMC

RMC is a rare disease with scarce knowledge and few treatment options as seen in the introduction. I had access to scRNA-seq data from two human samples with one mouse patient-derived xenograft, about 30 RNA-seq samples from a small patient cohort and c-MYC, BRG1, SMARCB1 and H3K27ac ChIP-seq or Cut&Tag data from RMC cellular models. The aims of this part are as follows:

- Uncover cell of origin and oncogenic mechanisms of RMC.
- Discover new vulnerabilities in this type of cancer.

Chapter II: Results

Chapter II: Results

In this section, I will present my results which include two published papers. Section II.1 contains the first paper which I co-authored with Alexandra Helleux and is titled: “Mesenchymal-like tumor cells and myofibroblastic cancer-associated fibroblasts are associated with progression and immunotherapy response of clear-cell renal cell carcinoma”. This paper contains the main results of my thesis work fulfilling the objectives described in sections I.3.A and I.3.B. I performed the main bioinformatics analysis in this study while the main wet-lab experiments and verifications were done by Alexandra.

I performed follow-up analyses to find biomarkers and broaden the scope of the paper which focused essentially on the tumor and fibroblast relationship. I presented these unpublished results in sections II.2 and II.3, these fulfill the objectives described in section I.3.C.

The second paper which I co-authored with Bujamin Vokshi is titled “SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance” and presented in section II.4. This work is on renal medullary carcinoma and fulfills the objectives presented in section I.3.D. I realized the main bioinformatics analyses in this study while Bujamin performed the wet-lab experiments.

1. Mesenchymal-like tumor cells and myofibroblastic cancer-associated fibroblasts are associated with progression and immunotherapy response of clear-cell renal cell carcinoma

Guillaume Davidson[1*], Alexandra Helleux[1*], Yann A. Vano[2], Véronique Lindner[3], Antonin Fattori[3], Marie Cerciat[4], Reza T. Elaidi[5], Virginie Verkarre[6], Cheng-Ming Sun[7], Christine Chevreau[8], Mostefa Bennamoun[9], Hervé Lang[10], Thibault Tricard[10], Wolf H. Fridman[7], Catherine Sautes-Fridman[7], Xiaoping Su[11], Damien Plassard[4], Celine Keime[4], Christelle Thibault-Carpentier[4], Philippe Barthelemy[12], Stéphane M. Oudard[2¤], Irwin Davidson[1¤] and Gabriel G. Malouf[1,12¤]

[1] Department of Cancer and Functional Genomics, Institute of Genetics and Molecular and Cellular Biology, CNRS/INSERM/UNISTRA, 67400 Illkirch, France

[2] Department of Medical Oncology, Hôpital Européen Georges Pompidou, Institut du Cancer Paris CARPEM, APHP, Université Paris Cité, Paris, France

[3] Department of Pathology, Strasbourg University Hospital, Strasbourg, France

[4] Genomeast platform, Institute of Genetics and Molecular and Cellular Biology, CNRS/INSERM/UNISTRA, 67400 Illkirch, France

[5] Association pour la Recherche sur les Thérapeutiques Innovantes en Cancérologie, Paris, France

[6] Department of Pathology, Hôpital Européen Georges Pompidou, Institut du Cancer Paris CARPEM, APHP, Université Paris Cité, Paris, France

[7] Centre des Cordeliers, INSERM, Université de Paris Cité, Sorbonne Université, Equipe labellisée Ligue contre le Cancer, F-75006, Paris, France

[8] Department of Medical Oncology, Institut Universitaire du Cancer Toulouse Oncopole, Toulouse, France

[9] Department of Medical Oncology, Institut Mutualiste Montsouris, Paris, France

[10] Department of Urology, Strasbourg University Hospital, Strasbourg, France

[11] Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[12] Department of Medical Oncology, Strasbourg University, Institut de Cancérologie de Strasbourg, Strasbourg, France

*****These authors should be considered equally as first co-authors.

¤These authors should be considered equally as co-senior authors.

Correspondence: Prof. Gabriel Malouf. Address: IGBMC, 1 Rue Laurent Fries, 67400, Illkirch, France. Phone: + 33 3 68 76 72 17. E-mail: maloufg@igbmc.fr

Running title: Crosstalk between tumour cells and myCAFs in renal cancer

Keywords: clear-cell renal cell carcinoma, cancer associated fibroblasts, EMT, immune checkpoint inhibitors, spatial transcriptomics

Published online in Cancer Research, 19 June 2023

Abstract

Immune checkpoint inhibitors (ICI) represent the cornerstone for treatment of patients with metastatic clear-cell renal cell carcinoma (ccRCC). Despite a favorable response for a subset of patients, others experience primary progressive disease highlighting the need to precisely understand plasticity of cancer cells and their crosstalk with the microenvironment to better predict therapeutic response and personalize treatment. Single-cell RNA sequencing of ccRCC at different disease stages and normal adjacent tissue (NAT) from patients identified 46 cell populations, including 5 tumor subpopulations, characterized by distinct transcriptional signatures representing an epithelial to mesenchymal transition gradient and a novel inflamed state. Deconvolution of the tumor and microenvironment signatures in public datasets and in data from the BIONIKK clinical trial (NCT02960906) revealed a strong correlation between mesenchymal-like ccRCC cells and myofibroblastic cancer-associated fibroblasts (myCAFs), which are both enriched in metastases and correlate with poor patient survival. Spatial transcriptomics and multiplex immune staining uncovered spatial proximity of mesenchymal-like ccRCC cells and myCAFs at the tumor-NAT interface. Moreover, enrichment in myCAFs was associated with primary resistance to ICI therapy in the BIONIKK clinical trial. This data highlights the epithelial-mesenchymal plasticity of ccRCC cancer cells and their relationship with myCAFs, a critical component of the microenvironment associated with poor outcome and ICI resistance.

Significance Single-cell and spatial transcriptomics reveal the proximity of mesenchymal tumor cells to myofibroblastic cancer-associated fibroblasts and their association with disease outcome and immune checkpoint inhibitor response in clear cell renal cell carcinoma.

A. Introduction

Renal cell carcinomas (RCC) are the most frequent malignant neoplasms arising from the kidney. Clear-cell renal cell carcinoma (ccRCC) is the most common RCC representing 75% among all cases, with an estimated 315.000 new patients worldwide per year (1,2). Extensive studies identified genetic and epigenetic alterations driving oncogenesis, with frequent mutations including *VHL* inactivation followed by mutations in various genes involved in chromatin remodelling complexes (*PBRM1, SETD2* and *BAP1*)(3–6). Recent single-cell transcriptomics described the ccRCC tumour microenvironment (TME) identifying co-occurrence between exhausted $CD8^+$ T-cells and M2-like tumour associated macrophages (TAMs) in advanced forms of the disease and their association with ICI resistance (7–9). Furthermore, association between lack of ICI response and enrichment in endothelial cells was also described (10). Nevertheless, the potential plasticity of ccRCC cancer cells and their interactions with the TME have not been investigated. Similarly, the role of cancer-associated fibroblasts (CAFs) and their potential cell of origin are still poorly described in ccRCC as compared with various other cancer types (11,12).

To address these questions, we performed scRNA-seq of 56,421 cells identifying 5 tumour cell states forming an epithelial to mesenchymal gradient together with a novel inflamed population expressing MHC class II genes. We further identified that myofibroblastic (my)CAFs strongly associate with mesenchymal-like ccRCC tumour cells reflecting their spatial proximity within tumours. Notably, these populations were selectively enriched at metastatic ccRCC sites, and the presence of myCAFs strongly associated with resistance to ICI in the BIONIKK clinical trial.

B. Material and Methods.

Patient samples

The seven ccRCC samples and the two NATs subjected to scRNAseq were collected from the pathology department of Strasbourg University Hospital. Sample collection for further research analysis was approved by the local ethics committee and all patient signed an informed written consent for the use of the material for research. The seven ccRCC samples represented various stages/grades of tumour evolution. NAT sample N2 is the associated normal adjacent tissue of tumour sample T4 whereas N1 comprises 3 different NAT samples, one of which matched to T3 and 2 from patients where no corresponding tumour samples were obtained. N1 and N2 were combined to provide more tubule diversity.

Single-cell sample preparation and sequencing

Following resection, samples from the tumour and normal adjacent tissue were conserved in 1mL of MACS Tissue Storage solution (Miltenyi Biotech) at 4°C for up to 24h. Single cell suspensions were then prepared using gentleMACS^{TM} dissociator and human tumour kit dissociation (Miltenyi Biotech) according to manufacturer instructions. Briefly, samples were transferred to C tubes containing 4.7 mL of DMEM at room temperature and minced to segments of <5mm^3. 100µL enzyme H, 50µL enzyme R and 12.5µL enzyme A were added to each C tube and loaded on the dissociator using program h_tumour_01. Tubes were then incubated for 30min at 37°C at 130 rpm and loaded again on the dissociator and program h_tumour_02 was run followed by another identical period of incubation. Finally, tubes were loaded a third time on the dissociator and program h_tumour_03 was run. The cell suspension was applied on MACS SmartStrainer 70µm placed on a 15mL Falcon tube and 10mL DMEM were used to wash C tubes and the filter. Following centrifugation of 10min at 300g, 4°C, cells were resuspended in 40% DMEM, 50% FCS, and 10% DMSO, placed in Cryotubes and gradually frozen to -80°C. On the day of single-cell experiment, cells were quickly thawed at

37°C until only a tiny ice crystal remained and dropwise transferred to 10mL of pre-warmed DMEM+10%FCS media. The cryotube vial was washed with 2mL of DMEM+10%FCS and add to the main 10mL solution. Samples were centrifuged 10min at 300g, 4°C. Dead cells were removed using Dead cell removal kit (Miltenyi Biotech) by resuspending in the correct volume of dead cell removal microbeads incubation for 15min at room temperature and passage through a MS column placed in separator magnetic field. Column was rinsed with binding buffer and un-labelled cells were collected in a 15mL Falcon tube placed in ice as live cell fraction. Cells were centrifuged again to be resuspended in smaller volume. Cell count and viability were assessed mixing cells with trypan blue (1:1 ratio) and using a Malassez counting chamber. 15.000 cells were then loaded on the Chromium 10x Genomics to be captured for subsequent preparation of the 3'-mRNA single-cell libraries following manufacturer instructions. Libraries were sequenced 2x100bp on either NextSeq550 or HiSeq4000 sequencers.

Individual sample processing of scRNA-seq data

After sequencing, raw reads were processed using CellRanger (v3.1) to align on the hg19 human genome, remove unexpressed genes and quantify barcodes and UMIs. Data were then analyzed in R (v4.0.2). Samples are first processed individually using Seurat (v3.2.0) with the recommended workflow. Cells were filtered to keep only cells with feature count ranging from 200 to 4500 and percentage of mitochondrial reads <20%. Potential doublets were removed using the DoubletFinder package assuming a doublet formation rate of 0.8% per 1000 cells reported by CellRanger. Counts were normalized with the "LogNormalize" method and data scaled to remove unwanted sources of variation (number of features and mitochondrial reads). Highly variable genes were determined by the "VariableFeatures()" and PCA was performed on these genes with the "RunPCA()" function. The number of principal components to use was determined from the plots generated from the "JackStraw()" function and generally ranged

from 20 to 30. Clustering was performed using functions “FindNeighbors()” and “FindClusters()” with a resolution of 0.8 on the most significant principal components. Cluster markers are computed with the function “FindAllMarkers()” and cells were broadly labelled by manually validating automated labelling using R package clustifyr (v1.6.0).

Analysis of merged samples

In order to correct batch effect when merging samples, 3 different method were tested: “SCTransform()”, “IntegrateData()” following the Seurat integration vignette and the R package Harmony using sample names as the batch variable. In our dataset, the linear correction approach with “SCTransform()” produced a better result than Seurat integrate and Harmony that showed excessive bias probably due to the similar level of gene capture in samples and their asymmetric compositions. All 9 pre-filtered samples with broad cell labels were merged using the “merge.Assay” function of Seurat. Data were normalized and scaled with batch effect corrected according to sample name using “SCTransform()” then an initial clustering was performed using a resolution of 0.8 and the 30 most significant principal components. Based on cluster marker genes, cells from 6 compartments (cancer cells, normal epithelium, fibroblasts, endothelium, lymphoid cells, and myeloid cells) were extracted with the “subset()” function and submitted to re-clustering. Misclustered cells were removed based on their broad labelling then data were normalized, scaled and re-clustered by using varying principal components and resolutions to find optimal parameters for each compartment. Cells were annotated according to their marker genes in the different sub-clustering analysis and these new cluster labels were transferred to the global analysis.

Cell typing and gene signatures

To phenotype clusters, both automatic (Clustifyr) and manual techniques were used. Known lineage markers for immune populations [CD45 (pan-immune), CD3E (pan T-cells), CD4 ($CD4^+$ T cells), CD8A ($CD8^+$ T cells), CD79A (B cells), NKG7 (NK cells), CPA3 (Mast cells),

CD68/LYZ (macrophages), CD16/CD14 (monocytes)], endothelial population (PLVAP), mesangial population (ACTA2, RGS5), epithelial population (EPCAM, KRT8, CDH16) and tumour population (CA9, ANGPTL4) were projected on the UMAP. Subsequently, each major cell type was re-clustered to identify specific subpopulations and differentially expressed genes for each sub-cluster were determined. Cell identification was based on comparison of the differentially expressed genes of each sub-cluster with markers previously described in the literature, (Dataset S1).

Analysis of public scRNA-seq data

For re-analysis of the Krishna et al. (8) dataset (SRA, accession number: PRJNA705464), we retrieved the available pre-computed RDS Seurat object and set it to v3 architecture with the "UpdateSeuratObject()" function. First, epithelial cells were re-clustered by sub-setting clusters 9 and 16 (labelled as "PAX8+ epithelium" and "CA9+ ccRCC") with a resolution of 0.5 and the 30 most significant components resulting in 10 new clusters. Based on marker genes, some clusters could clearly be identified as normal renal tubules and the remainder marked by *CA9* expression were then isolated and re-clustered with a resolution of 0.5 to obtain the final 7 ccRCC clusters.

For re-analysis of the Braun et al. (13)(2021) dataset (dbGaP, accession number phs002252.v1.p1), we retrieved raw counts and cell annotations in CSV format and used it to create a new Seurat object. We applied an analysis similar to our dataset with normalization and scaling with batch effect correction according to the "Batch" column present in the annotation file. Global clustering was performed with a resolution of 0.8 using the 30 most significant components that yielded 35 clusters. We found a high visual overlap between these new clusters and the originals by plotting the labels "ClusterName_AllCells" on the UMAP. Epithelial clusters with marked expression of *CA9* were re-clustered with a resolution of 0.4 to obtain the final 8 ccRCC clusters.

Functional analysis of scRNA-seq data

Data were represented as single-cell heatmaps with the "DoHeatmap()" function, as custom heatmaps with the pheatmap package and as bubble-plots with the "DotPlot()" function. Data were also visualized using the scope website (https://scope.aertslab.org) after converting the Seurat object to loom format using the SCopeLoomR package. To represent gene expression on the global UMAP and avoid overlapping signal, we used the schex package (https://github.com/SaskiaFreytag/schex). Regulome analyses of active transcription factors were performed using the SCENIC v1.1.2.2 package. Gene signatures were computed and visualized on UMAPs using the R package VISION (https://github.com/YosefLab/VISION). Gene set variation analysis were performed using the r-package GSVA and gene set enrichment analyses were done with the GSEA software v3.0 using the hallmark gene sets of Molecular Signature Database v6.2. Gene Ontology analysis was done using DAVID (http://david.abcc.ncifcrf.gov/). Gene list intersections and Venn diagrams were computed by the web-tool Venny (https://bioinfogp.cnb.csic.es/tools/venny/). Cell state trajectories were defined using the R package SWNE v0.6.18 (https://yanwu2014.github.io/swne/)(14).

Ligand-receptor interactions were inferred using the CellPhoneDB python package v2.0 (https://github.com/Teichlab/cellphonedb) with the "statistical_analysis" argument and default parameters (15). In addition, for predicting ligand-receptor interactions regulating specific programs we used NicheNet (16). First we used cluster MSG as 'receiver', cluster ccRCC.mes as 'sender' and the list of up-regulated genes (log2FC>1, adj.p<5%) from the myCAF versus MSG clusters as 'geneset_oi' arguments. Then we used cluster ccRCC.mes as 'receiver', cluster 'myCAF' as 'sender' and the list of 442 ccRCC.mes marker genes from the tumour cell reclustering as 'geneset_oi' arguments.

Generation of the BIONIKK dataset

RNA from 50 samples from the BIONIKK trial were received for sequencing. In addition, frozen tissues (n=47) were also received and processed with the AllPrep DNA/RNA kit following manufacturer's instructions (Qiagen – ref 80204). Briefly, around 30mg of frozen tissue were plunged in 600µL of RLT Plus buffer supplemented with 40mM of DTT and disrupted using an UltraTurrax for 30seconds. Tubes were centrifuged at 10.000g, at RT for 3 min and supernatant was transferred to an AllPrep DNA spin column. After centrifugation at 10.000g, RT for 30s, 1 volume of 70% EtOH was added to the flow-through, mixed by pipetting and transferred to an RNeasy spin column. After centrifugation, the column was sequentially washed once with RW1 buffer and twice with RPE buffer. After drying, the RNA was eluted twice with 20µL of RNAse-free water and sample quality was assessed by BioAnalyzer. Libraries for total rRNA-depleted RNA were prepared for the 97 qualified samples and paired-end 2X 100 bp sequencing was performed on a HiSeq4000 sequencer.

Analysis of bulk RNA-sequencing

RNA-seq data from the BIONIKK clinical trial was generated by total rRNA-depleted paired end RNA-seq. After sequencing, raw reads were pre-processed in order to remove adapter and low-quality sequences (Phred quality score below 20) using cutadapt version 1.10 and reads shorter than 40 bases were discarded. Residual reads maping to rRNA sequences using bowtie version 2.2.8, were also removed. Reads were mapped onto the hg19 assembly using STAR version 2.5.3a. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.6.1p1, with annotations from Ensembl version 75 and "union" mode. Only non-ambiguously assigned reads were retained for further analyses. Read counts were normalized across samples with the median-of-ratios method and across genes by using the median transcript length. For KIRC-TCGA, the raw data files were downloaded and re-

normalized using the Bioconductor package DESeq2 version 1.16.1. Sample compositions were estimated by deconvolution from our single-cell data using the CIBERSORTx algorithm.

Survival analysis.

Relationships between cell signatures and patient survival were computed in R using survival version 3.2 and survminer version 0.4.9. Patients were stratified using "surv_cutpoint()" and Kaplan-Meier curves were represented with the "survfit()" and "ggsurvplot()" functions. Hazard ratios were determined by univariate Cox proportional-hazards model with the "coxph()" function using continuous values from CIBERSORTx absolute scores. Scores with significant p-values in univariate models were adjusted with the independent clinical co-variates age and stage in multiple regression analyses. Overall Response Rate was defined as the proportion of patients with Partial Response (PR) and Complete Response (CR) while Clinical Benefit Rate was defined as the proportion of patients with PR, CR and Stable Disease (SD) over six months. In the BIONIKK dataset there is one case where two tumours samples were taken from the same patient, one of the two samples was arbitrarily removed for survival analysis.

Analysis of spatial-transcriptomics data

Spatial transcriptomics was performed on Formalin Fixed Paraffin Embedded (FFPE) tissue sections using the 10X Genomics Visium Spatial Gene Expression platform (10X Genomics, Pleasanton, CA, USA). Using a Leica RM2235 microtome (Leica, Germany), 5 µm tissue sections were cut from FFPE tissue blocks and placed within the etched frames of the capture areas on the active surface of the Visium Spatial Slide. Slide was dried in an oven overnight at 37°C then at 60°C during 25 minutes before dehydration in ethanol and Hematoxylin & Eosin (H&E) staining. Slide was finally transferred to water and mounted in 87% glycerol before image acquisition the same day. Image mosaic of each section was acquired at 5X and 20X magnification (Zeiss EC Plan-Neofluar 5x/0.16, Zeiss Plan-Apochromat 20x/0.8) with

transmitted white light and a Zeiss colour camera Zeiss Axiocam Color on a Zeiss Axio Observer microscope to generate a brightfield image of each section including the fiducial markers frame. The final stitched mosaic was created with Zeiss ZEN Stitching module. Subsequently, stained sections were strictly processed as outlined in the demonstrated protocol « Visium Spatial Gene Expression Reagent kits for FFPE, UserGuide » (Part number CG000407, RevC). Quantification and quality control of the final libraries were performed using Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Libraries were sequenced on an Illumina HiSeq 4000 platform as paired-end 2x100 bp reads and trimmed at 28+50 bp reads. Image analysis and base calling were performed using RTA version 2.7.7 and Spaceranger mkfastq v1.1.0. Image and reads processing were performed using Spaceranger v1.3.1 count. Data were then analyzed in R using Seurat. UMI counts were normalized, scaled and the most variable features determined by the function "SCTransform()" then dimension reduction was performed with "RunPCA()". For unsupervised clustering analysis, Seurat objects from both samples were merged before running "RunPCA()", "FindNeighbors()", "FindClusters()" and "FindAllMarkers()" functions using default parameters and the 30 most significant principal components. For integration with scRNA-seq data, the previously analyzed Seurat object was provided as reference with the spatial Seurat object as query to the "FindTransferAnchors()" function then predictions from "TransferData()" were stored as an assay in the spatial Seurat object. Both predictions and gene expression were visualized on the slides using the "SpatialFeaturePlot()" function. To better illustrate ccRCC.mes and myCAF proximity, we generated a multi-coloured representation highlighting spots with a prediction score above 0.1 for myCAF in green, above 0.1 (or 0.01 for T1) for ccRCC.mes in red, while spots with double-positive signal (above the thresholds for both cell types) are in yellow. For analysis of GEO public data (GSE175540) from Meylan et.al. (15), we retrieved three samples corresponding

to one of each available clinical stage: pT1 (GSM5924035), pT3 (GSM5924037) and pT4 (GSM5924031) and applied the same analysis pipeline.

Multiplex immuno-staining.

Tissues were prepared using standard fixation and paraffin embedding techniques. Tissues were sliced at 5µm to have a slide for specific antibody staining and another for isotype antibody control. Multiplex immunostaining was performed using the OPAL^{TM} 4-color kit (Akoya Biosciences) according to the manufacturer instructions. Briefly, slides were baked in the oven for 1 hour, dewaxed for 30min with Sub-X, a xylene substitute, and rehydrated through a graded series of ethanol baths (100%, 10min; 95% 10min; rinse in 70% and rinse in distilled water). Slides were fixed for 40min in 10% neutral buffered formalin, rinsed in distilled water and in appropriate AR buffer. Slides were placed in AR buffer bath, microwaved for 45sec at 900W and for 15min at 350W. After the slides cooled down and were rinsed in water and TBST buffer (25mM Tris-HCl, pH 7.5; 150mM NaCl; 0.05% Tween20), tissue sections were blocked for 10min in a humidified chamber using OPAL kit buffer. Sections were incubated with the primary antibody working solution (antibody diluted in the blocking buffer) for various times or temperatures. After washes, slides were incubated in Polymer HRP Ms+Rb for 10min, washed again and incubated in Opal Fluorophore working solution for 10min. Microwave treatment was again applied to strip the primary-secondary-HRP complex. The entire process was repeated starting at the blocking step to label with the next antibody. First, FN1 primary antibody (1/100, Proteintech) was introduced using pH6 AR buffer, a 2h incubation at room temperature and the OPAL570 as secondary antibody. Then, CA9 primary antibody (1/1000, Abcam) was added with pH9 AR buffer, an overnight incubation at 4°C and the OPAL690. The third primary antibody was TAGLN (1/1000, Proteintech) with pH6 AR buffer, a 2h incubation at room temperature and the OPAL520. Finally, DAPI working solution

was applied for 5min. Slides were washed and mount using ProLong Gold. We also performed negative controls using the same procedure, but omitting the primary antibodies. Immunofluorescence images were acquired on a wide field microscope Leica DM6B using a Hamamatsu ORCA-Flash 4.0LT sCmos camera. Representative images were taken at 20x magnification using the Leica LasX software for acquisition. Data were obtained in .lif format and opened with FIJI to adjust brightness/contrast and convert merged and individual channel images into .tiff format files.

Data Availability.

The single-cell and spatial RNA-seq data generated in this study are publicly available in Gene Expression Omnibus (GEO) at GSE210042. The data that support the findings of BIONIKK trial are not publicly available due to patient privacy requirements but are available upon reasonable request from the study sponsor (contact@association-artic.org). The public scRNA-seq data analyzed in this study were obtained from SRA (accession number: PRJNA705464) and dbGaP (accession number: phs002252.v1.p1)

C. Results

a. Single-cell RNA-seq profiling of ccRCC tumours

We performed single-cell (sc)RNA sequencing on 7 *in situ* ccRCC primary tumour samples (T1-T7) and normal adjacent tissues (NAT, N1 a mix of 3 NAT one associated with T3, N2 associated with T4, Fig. 1A-B and Fig. S1A). Patients were selected to cover a diverse ccRCC spectrum with early to advanced stages and grades at time of surgery, 5 of which were treatment naïve and 2 with metastatic disease underwent delayed nephrectomies after optimal response of distant metastatic lesions to either ICI (ipilimumab-nivolumab combination) or

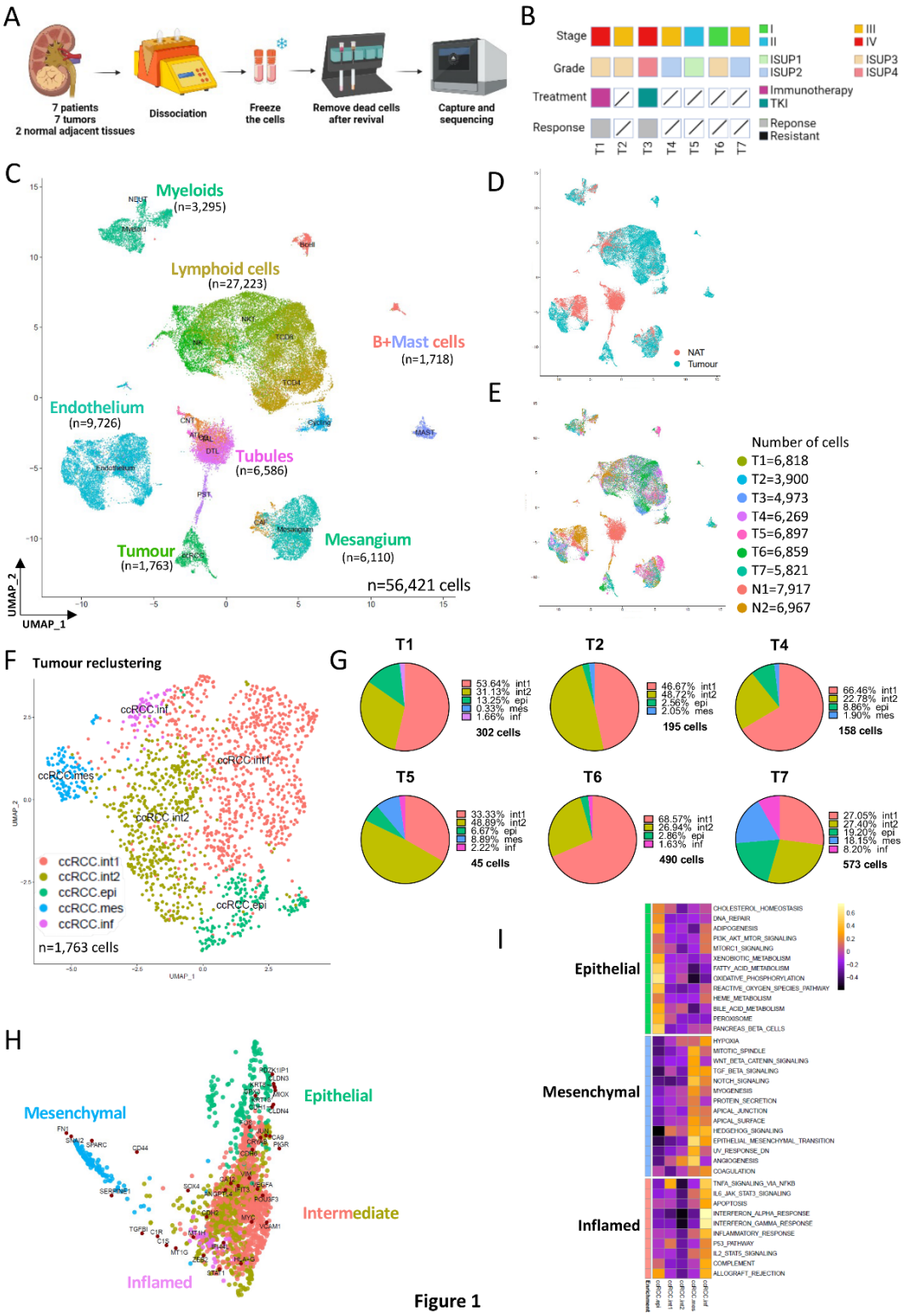


Figure 1

Figure 1. Single-cell RNA sequencing of ccRCC. A. Overall processing strategy. **B.** Clinical information for each sample. **C.** UMAP plot of the 56,421 cells captured and analysed for all 9 tumour and NAT samples. **D-E.** UMAP projections indicating tumour or NAT origin and the number of cells in each cluster used for analysis. **F.** UMAP showing re-clustering of tumour cells with each sub-cluster identified. **G**. Composition of each tumour showing total number of cells captured and the % of each subtype. **H.** SWNE trajectory analysis showing a set of selected markers per cluster. **I.** Heatmap representation of the GSVA analysis showing specific hallmarks enriched in each cluster

tyrosine kinase inhibitor (sunitinib) (Fig. 1B and Fig. S1A). Sequencing data were merged to obtain an integrated UMAP projection of 56,421 single cells broadly categorized by marker gene expression as lymphoid cells, myeloid cells, endothelium, mesangium, renal tubule and tumour cells. (Fig. 1C and Fig. S1B, Dataset S1). Immune cells represented 50% of the total, with a majority of lymphoid cells classified as natural killer-like T cells (NKT; *CD3D, CD8A/B* and *KLRD1*), T cells (*CD3D, CD3E* and *CD4* or *CD8A/B*) and NK cells (*KLRD1*) (Fig. 1C and Fig. S1B). Other immune populations included myeloid cell populations included *LYZ*-expressing monocytes and tumour-associated macrophages (TAM), neutrophils (*FCGR3B*), plasma or follicular B cells (*CD79A* and *MS4A1*) and mast cells (*CPA3*). *MKI67*-expressing cycling cells were mainly lymphocytes and a smaller number of tumour cells. Stromal cells included endothelial cells (*EMCN*), mesangial cells (*RGS5*) and cancer-associated fibroblasts (CAF; *POSTN*) along with proximal and distal renal tubule cells (*CDH16*). Overall, 41,537 cells were derived from tumour samples and 14,884 cells from NAT, the unique source of tubule cells (Fig. 1D-E and Fig. S1C and Dataset S2).

The tumour, stromal and immune cell subtypes were individually re-clustered to obtain 45 sub-clusters (Fig. S2, Dataset S2), 14 from lymphoid cells (Fig. S2A-B) 6 from endothelial cells (Fig. S2C-E), 4 from myeloid cells including classical (MONO.cl) S100A8^{high} and atypical monocytes (MONO.at) along with TAM populations including the recently described CD1C-expressing macrophages (17) (Fig. S2F). Classical monocytes were characterized as *CD14*$^{+}$ and *CD16*$^{-}$ (FCGR3A^{-}), whereas atypical monocytes were *CD14*$^{-}$/*CD16*$^{+}$. TAMs expressed both M1 (*IL1A, CXCL9, CD86*) or M2 (*APOE, C1QA, MRC1*) polarization markers and formed a M1-M2 gradient (Fig. S2F-H) (13,17). The M1 signature was represented in the MONO.at population in agreement with reports that they can differentiate into TAMs (13) and accounting for the residual expression of some genes from the MONO.at signature in the TAM sub-cluster (Fig. S2G).

b. Plasticity of ccRCC tumour cells

We captured 1,763 cancer cells from 6 tumours with none captured from the sunitinib-treated T3 sample, consistent with its good response to TKI therapy (Fig. 1F-G and Fig.S1C). While all tumour cells were characterized by *CA9* and *VEGFA* expression, re-clustering identified 5 populations with differentially expressed genes and pathways: epithelial-like (ccRCC.epi ; n=186 cells, 10.5%), expressing *PDZK1IP1*, *LGALS2* and *GPX4*, mesenchymal-like (ccRCC.mes ; n=116 cells, 6.5%) expressing mesenchymal markers *COL1A1, COL1A2, COL6A3*, fibronectin 1 (*FN1*) and *SERPINE1,* two intermediate states (ccRCCint1/ccRCCint2 ; n=864/536 cells, 49%/30% respectively) expressing stress markers (*ATF3*, *DNAJB1*) or metallothionines (*MT1A*, *MT1M*), a novel inflamed state (ccRCC.inf, n=61 cells, 3.4%) characterized by IL6/JAK/STAT3, IL2/STAT5, interferon signaling pathway and expression of *CD74* and MHC-II genes (*HLA-DRA, HLA-DRB1, HLA-DQB1:* Fig. 1F and Fig. S3A, Dataset S2). Each tumour comprised varying numbers of cells of each phenotype and all phenotypes were present in multiple tumours (Fig. 1G). Acquisition of mesenchymal markers in ccRCC.mes was accompanied by loss of epithelial genes in an EMT gradient (Fig. S3B) confirmed by SWNE trajectory analysis that traced transformation from epithelial-like to mesenchymal-like states via intermediate states (Fig. 1H). Gene Set Variation Analysis (GSVA) of Hallmark pathways showed that ccRCC.epi was enriched for oxidative phosphorylation and fatty acid metabolism, whereas ccRCC.mes was enriched for EMT, hypoxia, TGF-β and Notch signaling (Fig. 1I). KEGG pathway analysis recapitulated the GSVA with ccRCC.epi enriched in oxidative phosphorylation and metabolic pathways and ccRCC.mes in ECM-receptor interaction, focal adhesion and the PI3K-AKT pathway (Fig. S3C).

In contrast, trajectory analysis did not distinguish whether the inflamed state was an intermediate in the EMT gradient or a distinct state adopted by cells from one of the

intermediate populations (Fig. 1H). Analogous 'interferon response' tumour cells were described in a multi-cancer assessment of tumour cell states (18) and in a MAPKi-treated melanoma patient derived xenograft (19). VISION projection of a 129-gene 'interferon response' signature from melanoma (19,20) showed that ccRCC.inf cluster had the highest score confirming the existence of an inflamed/immune-like phenotype in these two different cancers (Fig. S3D).

Characterization of the ccRCC.int sub-clusters was challenging due to low numbers of marker genes compared to other populations (95 for ccRCC.int1, 24 for ccRCC.int2, 351 for ccRCC.epi, 442 for ccRCC.mes and 117 for ccRCC.inf; Dataset S2) and lack of specifically enriched pathways in GSVA and KEGG ontology analyses (Fig. 1I and Fig. S3C). 137 genes were enriched in ccRCC.int1, that appeared more epithelial with higher expression of *EPCAM*, *PDZK1IP1* and *PAX8,* compared with 52 genes in ccRCC.int2 that appeared more mesenchymal with higher expression of *TGFBI*, *SERPINE1*, some collagens and metallothionine genes (Dataset S2). These observations were confirmed by GSEA analysis (Fig. S3E) showing enrichment of mesenchymal phenotype in ccRCC.int2 (Epithelial to mesenchymal transition and Hypoxia pathways), or stress pathways for ccRCC.int1 (e.g. UV response) represented by genes such as *ATF3*, *JUNB, DNAJB* and *DNAJA1*. Despite the heterogeneous aspect of both ccRCC.int1 and ccRCC.int2 visible upon further sub-clustering (Fig. S3F-H), we retained the ccRCC.int1 and ccRCC.int2 populations as they displayed signatures analogous to recently described stress-response and metal response populations in other solid tumours (18).

To validate identification of these ccRCC tumour cell types, we re-analysed datasets reported by Krishna et.al. (8) and Braun et.al. (13). We re-clustered CA9-expressing cells into 6 sub-clusters (CK1-6) in Krishna dataset (Fig. S4A-B) and 8 sub-clusters (CB1-8) in Braun dataset (Fig. S5A-B). Despite the large number of samples, tumour cells were principally

derived from few patients in both cohorts (Fig. S4C-D and Fig. S5C-D). VISION projections of our ccRCC tumour signatures identified the ccRCC.epi, ccRCC.mes and ccRCC.inf populations in both data sets (Fig. S4E and Fig. S5E). Cluster identification was supported by GSVA analyses and heatmaps showing expression of specific signature genes in each cluster (Fig. S4F-G and Fig. S5F-G). Additional ccRCC states with cycling signatures (clusters CK4 and CB5), hypoxia signature (CB6), translation signature (CB3) or apoptosis (CB7) were identified (Fig. S4B and Fig. S5B). A small number of ccRCC cells were present in the "Cycling" cluster of our dataset. Cells with hypoxia signature in the Braun et al. dataset were derived from a single metastasis sample accounting for its absence in our data set coming exclusively from primary tumours (13). The small apoptosis cluster was possibly filtered out in our quality control due to high expression of mitochondrial genes. Otherwise, these analyses confirmed the existence of the major ccRCC tumour cell types in public data sets.

c. Comparison of tumour cell states to proximal straight tubules

To compare tumour cell states with their putative cells of origin, we re-clustered 6,585 renal tubule epithelial cells into 7 populations representing different nephron segments (Fig. 2A): 5 sub-clusters from the proximal and distal tubules, 1 sub-cluster of connecting tubules and 1 sub-cluster of collecting duct cells (Fig. 2A-B and Fig. S6A-B). We defined proximal straight tubules (PST) expressing *GPX3, MIOX* as well as segment 3 markers *PDZK1IP1, MT1G* and *MT1H*. We could not divide the PST cells into 3 segments using the previously defined markers (21,22) preventing a more precise annotation of the PST cluster (Fig. S6C-D). Descending thin limb (DTL) cells displayed highest expression of *SLPI* and ascending thin limb (ATL) cells highest expression of *ITGB6*. Two groups of thick ascending limb (TAL) cells expressed *SLC12A1* and *ENOX1* and could be divided in cortical (cTAL) and medullar (mTAL) based on differential *UMOD* expression. We identified a group of connecting tubules

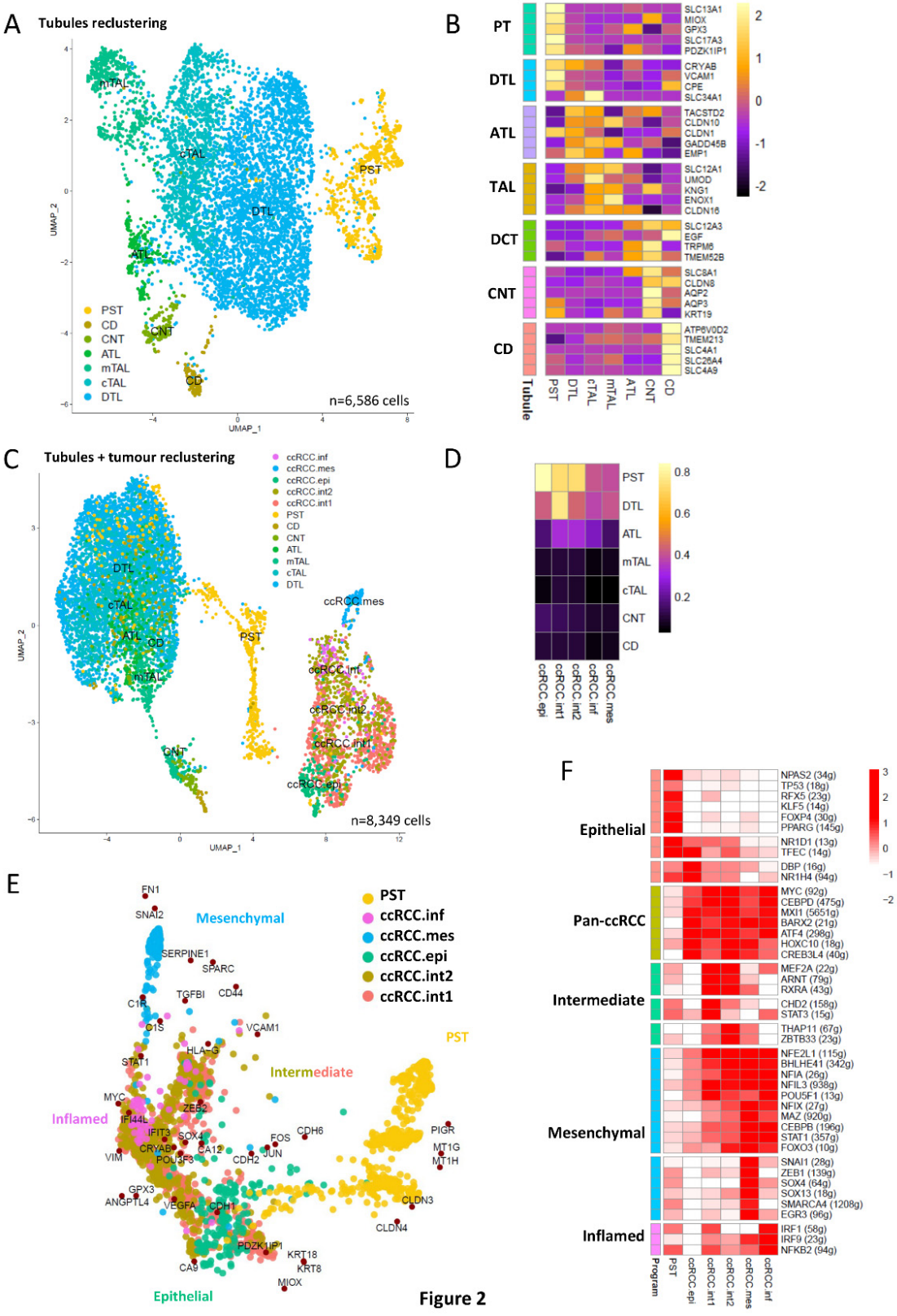


Figure 2. Proximal straight tubules as cell of origin of ccRCC. A. UMAP plot of tubule sub-clusters. **B.** Pseudo-bulk heatmap of identity markers of each renal tubule sub-cluster. **C.** Integrated UMAP plot of ccRCC tumour and tubule clusters. **D.** Pseudo-bulk heatmap of expression for renal tubule cell signatures in each ccRCC sub-cluster. **E.** SWNE trajectory analysis using a set of selected markers per cluster. **F.** SCENIC analysis of PST and tumour cells showing top regulons for each cluster (number of genes per selected regulons are indicated in brackets).

(CNT) marked by expression of *SLC8A1* and *AQP2* and a group of collecting ducts (CD) marked by expression of *ATP6V0D2* and *SLC4A1* (Fig. 2B, Dataset S2).

PST cells most closely resembled ccRCC.epi based on their proximity on the UMAP plot (Fig. 2C) and their gene signature was most represented in ccRCC.epi, but progressively lost in the EMT gradient (Fig. 2D). Trajectory analysis traced PST to ccRCC.epi followed by the EMT gradient (Fig. 2E). A PST origin was also supported by the prominent gluconeogenesis signature (*FBP1, FBP2, PCK1, PCK2, PC, GPI, SGLT1, SGLT2*), the preferential mode of glucose production in proximal renal tubules (23), that persisted in ccRCC.epi and some intermediate cells. Similarly, the OXPHOS signature, prominent in PST cells also persisted in ccRCC.epi cells before diminishing during EMT. In contrast, HIF1a and glycolysis signatures were poorly represented in the PST cells, but were strong in all tumour cells including the ccRCC.epi population. Oncogenic transformation of PST cells therefore involved acquisition of a strong hypoxia signature and a metabolic switch from OXPHOS to glycolysis, with both signatures found in the ccRCC.epi population, but with glycolysis predominant in ccRCC.mes and ccRCC.inf populations (Fig. S6E).

To define the transcriptional regulatory networks underlying the transformation of PST cells into ccRCC and those characterizing the different ccRCC cell states, we performed SCENIC analysis (Fig. 2F). PST cells displayed high activity of several TFs associated with metabolism, cell identity and tumour suppression, either specific to PST (TP53 and PPARG) or gradually lost (TFEC and NR1H4). Transformation was accompanied by activation of multiple regulons linked with stress response, inflammation and angiogenesis shared across all ccRCC clusters (MYC and ATF4), while others were either shared across ccRCC.int clusters (ARNT and RXRA), enriched in ccRCC.int1 (CHD2 and STAT3) or enriched in ccRCC.int2 (THAP11 and ZBTB33). The mesenchymal state was marked by regulons linked with inflammation, Wnt signalling and EMT either gradually activated (STAT1 or FOXO3) or

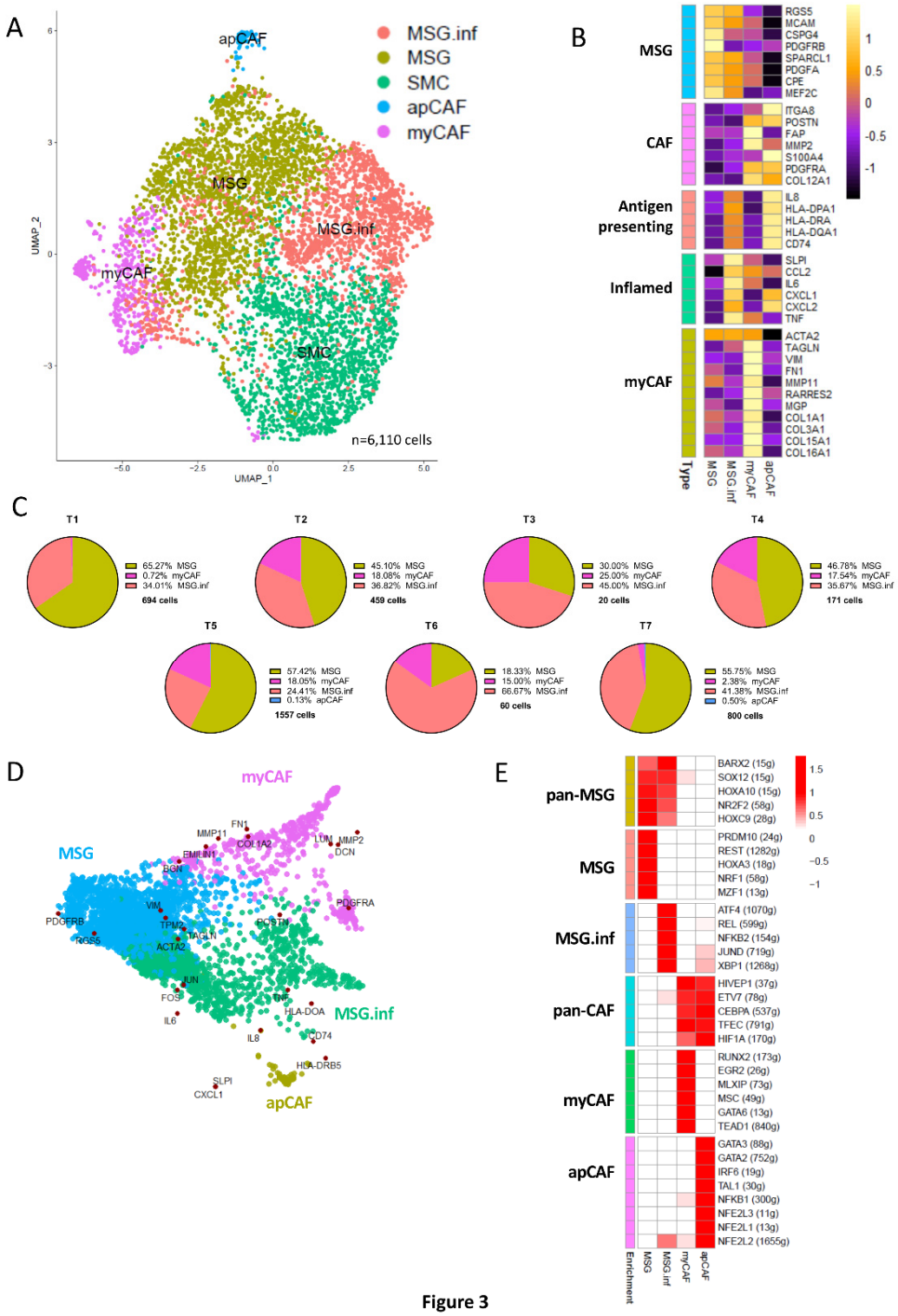


Figure 3

Figure 3. Heterogeneity of CAFs and trajectory analysis. A. UMAP plot of the 5 identified mesangium sub-clusters. **B.** Pseudo-bulk heatmap of identity markers of each mesangial and CAF sub-cluster. **C.** Distribution of the mesangial populations amongst tumours indicating the % of each population and the total number of cells per tumour. **D.** SWNE trajectory analysis using a set of selected markers per cluster. **E.** SCENIC analysis of mesangial populations (excluding SMC) showing top regulons common or specific to each cluster (number of genes per selected regulons are indicated in brackets).

specific to ccRCC.mes (SNAI1 and ZEB1). Furthermore, IRF1 and NFKB2 regulons were enriched in ccRCC.inf consistent with their inflamed/immune-like signature.

These analyses defined how the transcription program of PST cells governed by epithelial identity regulators such as PAX8 and PPARG was replaced by a pan-ccRCC program involving MYC and ATF4 as well progressively acquired or specific programs involving AP1, the SOX and POU family of transcription factors (BRN1 and OCT4) as well as EMT regulators (ZEB and SNAI) defining different cell states (Fig. S6F).

d. Heterogeneity of the mesangial and CAF populations

The mesangial compartment could be sorted in 5 sub-clusters (Fig. 3A-B). A sub-cluster designated SMC represented smooth muscle rather than bone fide fibroblasts with high expression of contractile muscle marker genes such as *ACTA2*, *MYH11* or *MYL9* and *CNN1*, a smooth muscle marker. Classical *RGS5*- and *PDGFRB*-expressing pericytes designated as mesangial cells (MSG) in the kidney were observed along with $RGS5^{+}$ cells expressing stress markers, inflammation-associated cytokines like *IL6, CXCL1* or *TNF* and showing expression of MHC class II genes (MSG.inf). We further identified two clusters of $RGS5^{low}/POSTN^{high}$ and $PDGFRB^{low}/PDGFRA^{high}$ CAFs that segregated into previously described phenotypes of inflammatory/antigen-presenting CAFs (iCAFs/apCAFs) and myofibroblastic CAFs (myCAFs) with myofibroblast markers and TGF-β activated genes like *FAP*, *FN1*, *MMP11* and multiple collagens (Fig. 3B). ApCAFs were distinguished from MSG.inf by loss of *RGS5*, gain of *POSTN* and higher expression of MHC class II genes (Fig. 3B). All tumours comprised varying proportions of MSG and myCAF populations (Fig. 3C), whereas apCAFs were mainly derived from NAT as observed in other tumours like bladder cancer (24). Trajectory analyses suggested MSG.inf as precursors to apCAFs, whereas myCAFs developed from $RGS5^{+}$

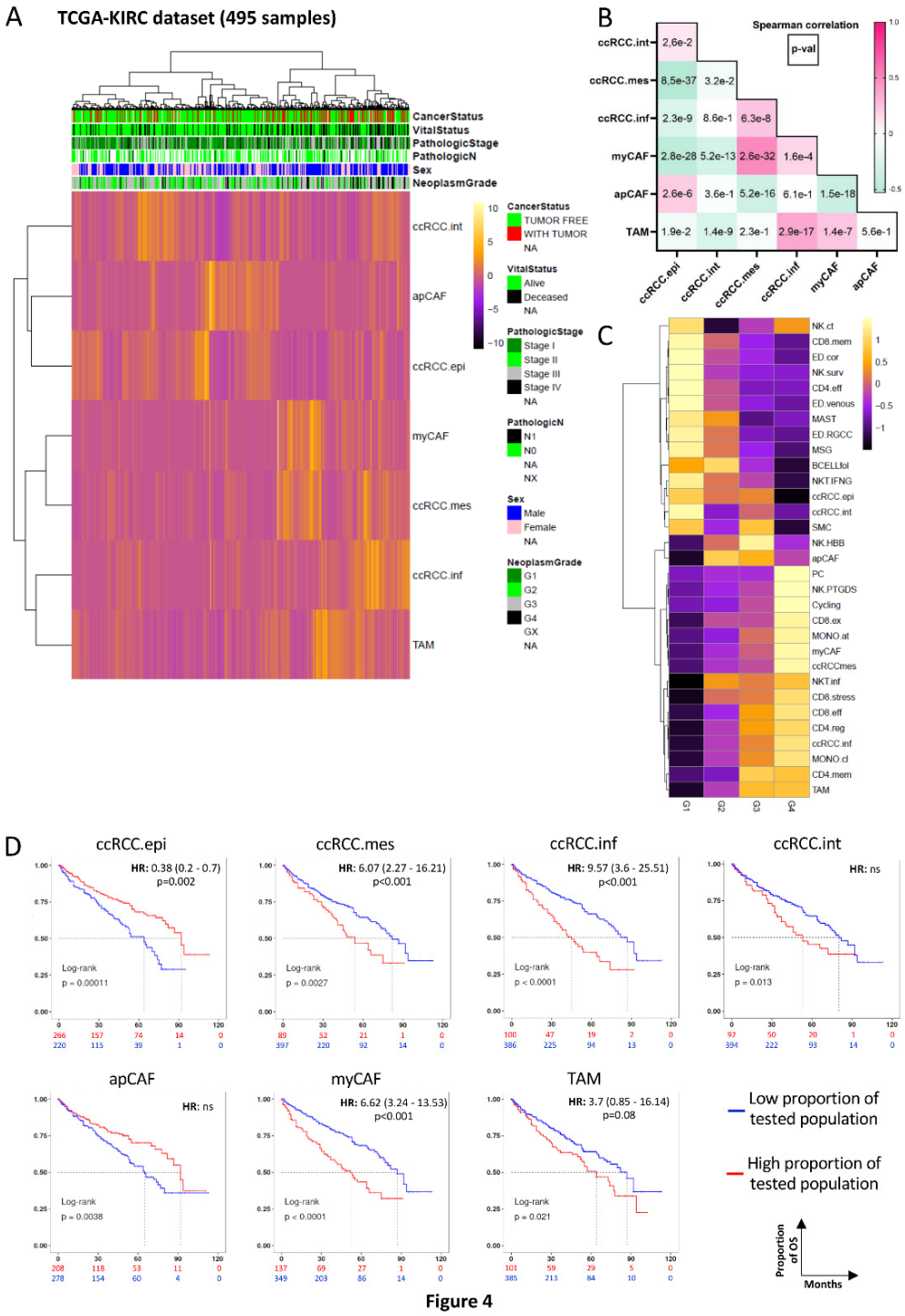


Figure 4

Figure 4. Correlation of ccRCC.mes and myCAF with survival in TCGA-KIRC cohort. **A.** Heatmap showing deconvolution of ccRCC, CAF and TAM cell scores inferred by CIBERSORTx and displayed as row-scaled absolute scores on bulk RNA-seq data of 495 tumour samples. **B.** Spearman correlation coefficient (coloured box) and associated p-value (number in box) between the indicated populations. **C.** Heatmap of each subpopulation average row-scaled absolute score in TCGA-KIRC samples grouped by tumour grades. **D.** Kaplan-Meier curves for overall survival in TCGA-KIRC patients according to calculated proportions of the indicated populations. Patient samples were ranked by proportion of designated populations and divided into two groups according to the optimal cut-off value method with indicated log-rank p-values. Hazard ratios (HR) were calculated using univariate Cox proportional-hazards model with the continuous CIBERSORTx absolute scores for the indicated population. ns: non-significant.

pericytes via a distinct trajectory (Fig. 3D). SCENIC regulon analysis also differentiated pericytes from CAFs with activity of myofibroblast-related transcription factors like MLXIP, TEAD1 or TEAD3 in myCAFs, while apCAFs showed activity for GATA3, NFKB1 and as reported (25) the NFE transcription factors (Fig. 3E). Gain of activity of these regulons was accompanied by the loss of those active in pericytes, although MSG.inf shared reduced activity of 3 regulons strongly active in apCAFs, again consistent with their being a precursor population.

e. Association between cancer cell and stromal cell states with patient outcome

To investigate the presence of the above-described cell populations in the TCGA KIRC patient cohort, we used our scRNA-seq dataset as reference for the CIBERSORTx algorithm to perform deconvolution of bulk RNA-seq data. To facilitate interpretation, ccRCC.int1 and ccRCC.int2 were merged as an intermediate signature designated ccRCC.int.

Unsupervised clustering of TCGA ccRCC cohort (n=495) using deconvolution scores of tumour cell populations, CAF and TAM states, revealed enrichment of ccRCC.mes and myCAF or ccRCC.epi in subsets of tumours in an anti-correlated manner (Fig. 4A-B and Fig. S7A-B, Dataset S3). Spearman correlations confirmed strong anti-correlation of ccRCC.mes and ccRCC.epi (Spearman coefficient r=-0.527; p=8.5e-37), whereas ccRCC.mes and myCAF strongly correlated (r=0.498; p=2.6e-32), and ccRCC.inf showed significant correlation with TAMs (r=0.367; p=2.9e-17) (Fig. 4B and Fig. S7C) in agreement with recent observations (18). MyCAF, ccRCC.mes and ccRCC.inf were strongly enriched in grade (G4) ccRCC (p(myCAF)=3.3e-9; p(ccRCC.inf)=1.7e-5; p(ccRCC.mes)=4.7e-7) along with exhausted CD8.ex and TAM populations as previously described (7) (Fig. 4C). In contrast, ccRCC.epi was strongly enriched in G1. Similarly, myCAF, ccRCC.mes and ccRCC.inf were strongly

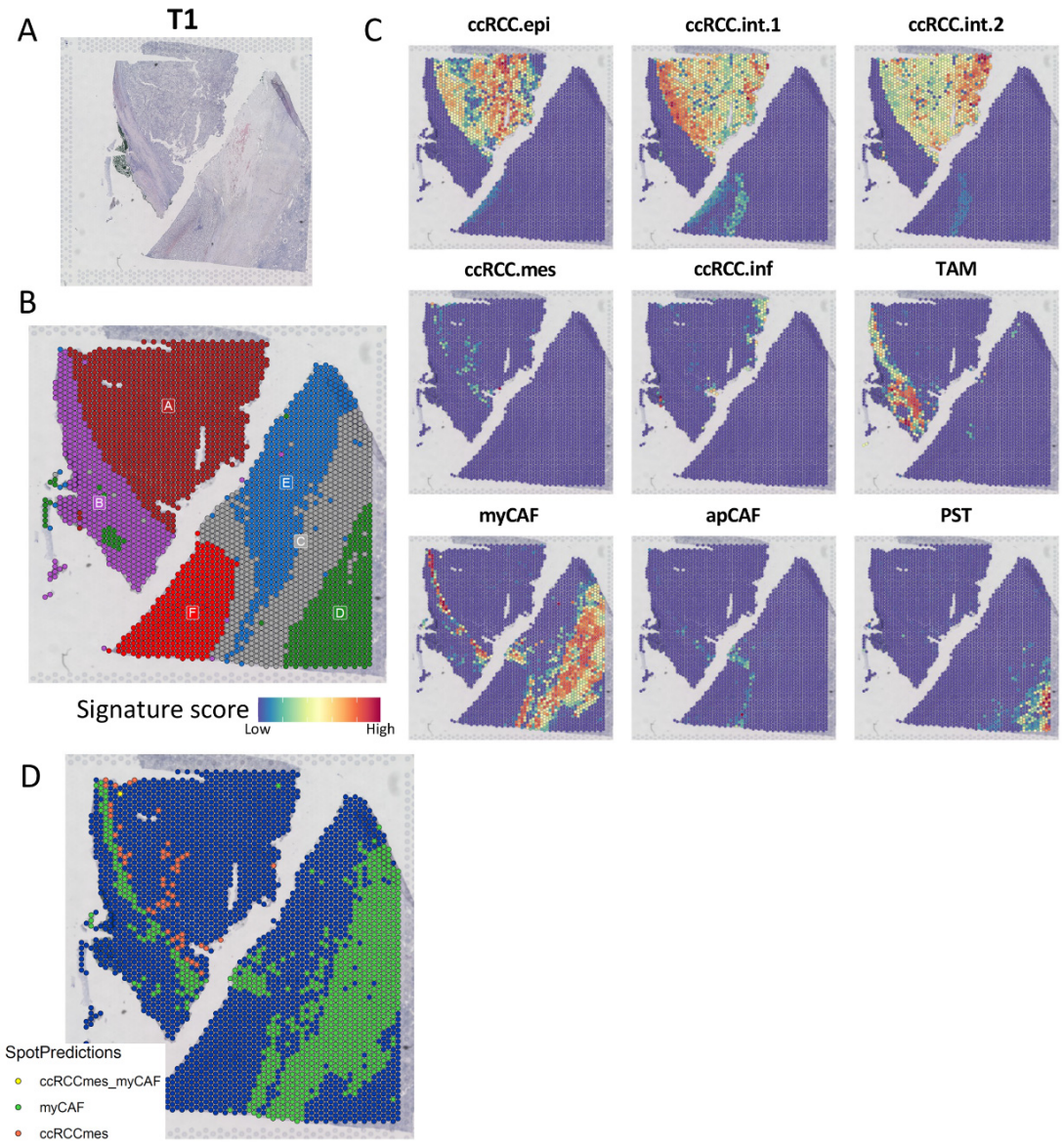


Figure 5

Figure 5. Spatial proximity of ccRCC.mes and myCAFs. A. Haematoxylin/Eosin staining of the T1 tumour section used for spatial transcriptomics. **B.** Spatial plot showing spots clustered by regions (A-F). **C.** Spatial plots showing prediction scores in each spot for indicated scRNA-seq derived signatures. **D.** Dual prediction using color-coded spots myCAF green, ccRCC/mes red, both cell types yellow.

enriched in TCGA m3 molecular subtype, the transcriptional signature in the TCGA classification most strongly associated with poor outcome (26) (Fig. S7D).

We analyzed the relationship between the tumour, CAF and TAM populations with patient survival by stratification using the optimal cut-point value method and computation of univariate Cox regression model using continuous values. Using both methods, only ccRCC.mes (HR=6.07, p<0.001), ccRCC.inf (HR=9.5, p<0.001) and myCAF (HR=6.6, p<0.001) were significantly associated with poor survival, whereas ccRCC.epi was significantly associated with better survival (HR=0.38, p=0.002) (Fig. 4D, Dataset S3). Although not significant using continuous values in the univariate Cox regression model, apCAF showed a trend towards better survival and ccRCC.int and TAM signatures a trend with poor survival using the optimal cut-point value. After Cox multiple regression analysis to correct clinical covariates including age and stage, the ccRCC.mes, myCAFs and ccRCC.inf remained strong negative predictors of survival, with ccRCC.epi a strong positive predictor of survival (Dataset S3).

f. Spatial association of ccRCC.mes and myCAFs

Given the strong ccRCC.mes-myCAFs correlation, we asked if they were spatially associated within ccRCC tumours and thus potentially signaling to one another. We performed 10X visium spatial transcriptomics (Fig. S8A) on sections from tumours T1 and T7 and applied the transcriptional signatures derived from the scRNA-seq to localize $CD4^+$ T cell, CD8+ T cell, NK cell, B cell, Monocytes, Neutrophils, Mast, CAF and tumour cell populations.

Histology and transcriptional analyses of the T1 section revealed several distinct areas. Area A, high grade ccRCC, displayed many ccRCC.epi and ccRCC.int1/2 cells, that although intermixed, were enriched at different locations. CcRCC.int1 cells were enriched at the

interface with fibrotic area B strongly enriched in myCAFs and TAMs (Fig. 5). Strikingly, ccRCC.mes cells localized along the A-B interface adjacent to myCAFs. In contrast, ccRCC.inf cells localized adjacent to ccRCC.int2 in area A. Area E was strongly enriched in B and T cells and separated from the rest of the tumour and NAT by fibrotic area C, strongly enriched in myCAFs, with a small number of apCAFs at their interface (Fig. 5 and Fig. S9A). Area F corresponded to low grade ccRCC, but was more strongly marked by mesangial signature than by tumour signature perhaps due to the abundant capillary vessels around the tumour cells in this region. Region D, separated from the tumour by the fibrotic myCAF-containing area C, corresponded to NAT with the visible presence of tubules and PST signature.

In the T7 section, ccRCC.inf cells located in a specific region of area A with dispersed ccRCC.epi and ccRCC.int1/2 cells in areas A, C and E separated by a fibrotic area B running diagonally through the section and strongly enriched in myCAFs with a small number of adjacent apCAFs (Fig. S8B-D and Fig. S9B). In contrast, ccRCC.mes cells were specifically localized adjacent to myCAFs in area C and small groups of ccRCC.mes closely associated with clusters of myCAFs in area E. Area D, separated from the tumour by the myCAF-containing fibrotic area B, corresponded to NAT, with PST and POD signatures. Although in T1 lymphoid cells were separated from the tumour, areas A, C and E of T7 all showed dispersed lymphoid infiltration (Fig. S9A-B).

To consolidate identification of ccRCC.mes and myCAFs and illustrate their co-localization, we coloured spots with a prediction score above 0.1 for myCAF in green, above 0.1 (or 0.01 for T1) for ccRCC.mes in red and spots with double-positive signal in yellow. Most red spots were close to green spots and a smaller number of yellow spots showed the presence of both cell types in close proximity (Fig. 5D and Fig. S8E). Thus, in contrast to ccRCC.epi/int1/int2 cells localized widely throughout the tumours, ccRCC.mes cells localized adjacent to myCAFs that formed fibrotic areas at the interface with NAT.

These findings were consolidated by analyses of the public dataset of Myelan et al. (27). Based on the provided clinical annotations, we selected one case for each available stage (C20-pT1, C34-pT3 and C3-pT4; Figs. S10A-C, left panel). The ccRCC.epi signature was more widely represented in C20, whereas ccRCC.mes was more abundant in the more advanced stage tumours C34 and C3. However, in all tumours, a spatial proximity of ccRCC.mes and myCAFs was observed (illustrated by colored spots in the dual prediction panel in Fig. S10A-C). Moreover, as seen in T1 and T7, myCAFs formed an interface between the tumour and NAT labelled with the PST signature. In C34, both myCAFs and ccRCC.mes were seen invading the NAT (Fig. S10C) similar to myCAFs invading NAT in T1 and T7.

We further performed multiplex immunostaining on mirrored sections of T1 and T7 used for spatial transcriptomics using FN1 for ccRCC.mes, TAGLN for myCAFs, CA9 for tumour cells and DAP1 staining for nuclei. On T1, CA9 staining delineated the tumour-containing regions A and F previously defined by spatial transcriptomics, TAGLN labelled the fibrotic myCAF regions B and C, whereas ccRCC.mes cells labelled by both CA9 and FN1 were located at the border of the myCAF/TAGLN-labelled region B (Fig. 6A). For T7, CA9 stained tumour regions A and E as defined by spatial transcriptomics, whereas TAGLN labelled the fibrotic myCAF region B invading the NAT in regions D1 and D2 (Fig. 6B). As observed by spatial transcriptomics, CA9 and FN1 labelled ccRCC.mes cells were abundant in region C also infiltrated by dispersed TAGLN expressing myCAFs and bordering myCAF-rich region B. Similarly, dispersed groups of FN1 labelled ccRCC.mes cells localized with TAGLN-labelled myCAFs in region E. Immunostaining therefore confirmed the results of spatial transcriptomics showing proximity of myCAFs and ccRCC.mes within these tumours as well as the propensity of myCAFs to invade the adjacent NAT.

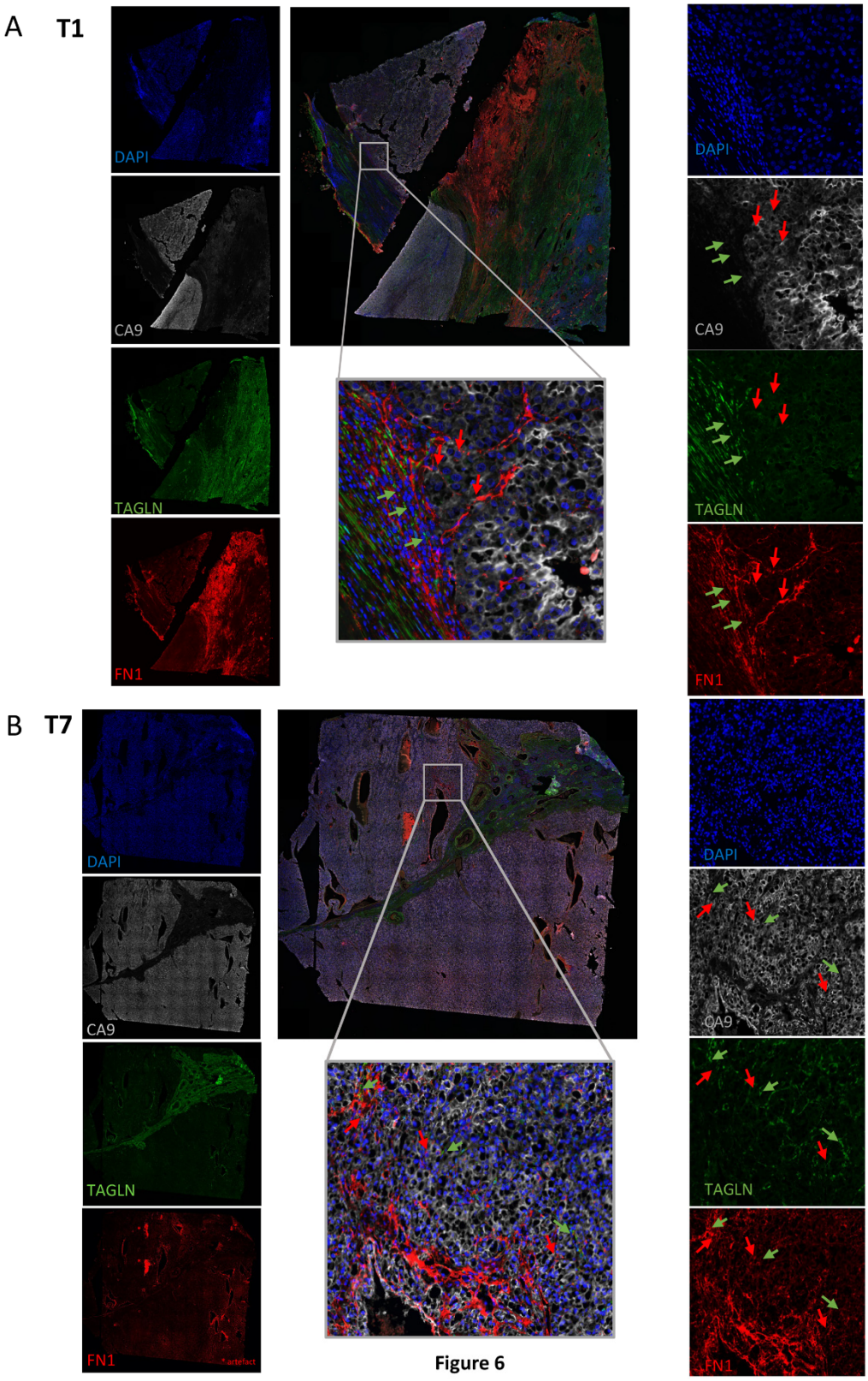


Figure 6

Figure 6. Multiplex immunostaining. A. Tumour T1; **B** tumour T7. Immunofluorescence staining of DAPI (blue), CA9 (gray), TAGLN (green), FN1 (red). Complete section individual colours, left panels and merge upper center panel. Zoom in the region of interest, merge lower center panel, individual colours right panels. Representative CA9 and FN1-labelled ccRCC.mes cells are indicated with red arrows and TAGLN-labeled myCAFs indicated with green arrows on the zoomed region. * Indicates regions of T7 with non-specific FN1 signal.

g. Signaling between ccRCC.mes and myCAFs reveals potential therapeutic targets.

We used CellPhoneDB and NicheNet softwares to predict ligand receptor interactions between myCAF and ccRCC.mes that may contribute to their communication and constitute therapeutic targets. CellPhoneDB revealed a collection of ligand-receptor interactions using either ccRCC.mes ligands with myCAF receptors (Fig. S11A) or vice-versa (Fig. S11B). In both configurations, multiple high confidence interactions were predicted involving secreted (for example, MDK-LRP1 or VEGFA-LRP1) or juxtacrine (JAG1-NOTCH3) ligand receptor pairs. Other bidirectional interactions were predicted involving NRP1 known to promote invasion in cancer cells (28) but also interactions involving MDK, EGFR and AXL.

We used NicheNet to identify ccRCC.mes ligands involved in promoting conversion of pericytes into myCAFs (Fig. S11C). CcRCC.mes-secreted TGF-β was predicted as a principal pathway, but IL6-IL6R, GAS6-AXL and PDK1-PDK2 pathways were further observed. NicheNet also predicted how myCAF-secreted TGF-β may promote or maintain EMT of ccRCC.mes together with GAS6-AXL and CXCL12-CXCR4 (Fig. S11D). A complex network of positive feedback signaling between ccRCC.mes and myCAFs may therefore promote and sustain the phenotypes of these closely associated cell populations.

h. MyCAFs associate with resistance to first-line nivolumab +/- ipilimumab treatment.

To ask if the ccRCC.mes-myCAF correlation is seen in an independent data set, we used CIBERSORTx to deconvolute 97 RNA-seq samples from patients before first line treatment with either sunitinib, nivolumab or the nivolumab/ipilimumab combination from the BIONIKK clinical trial (29,30). Tumours were again heterogeneous with all ccRCC tumour signatures represented and a strong ccRCC.mes-myCAF association (r=0.671; p=5.3e-14; Fig.

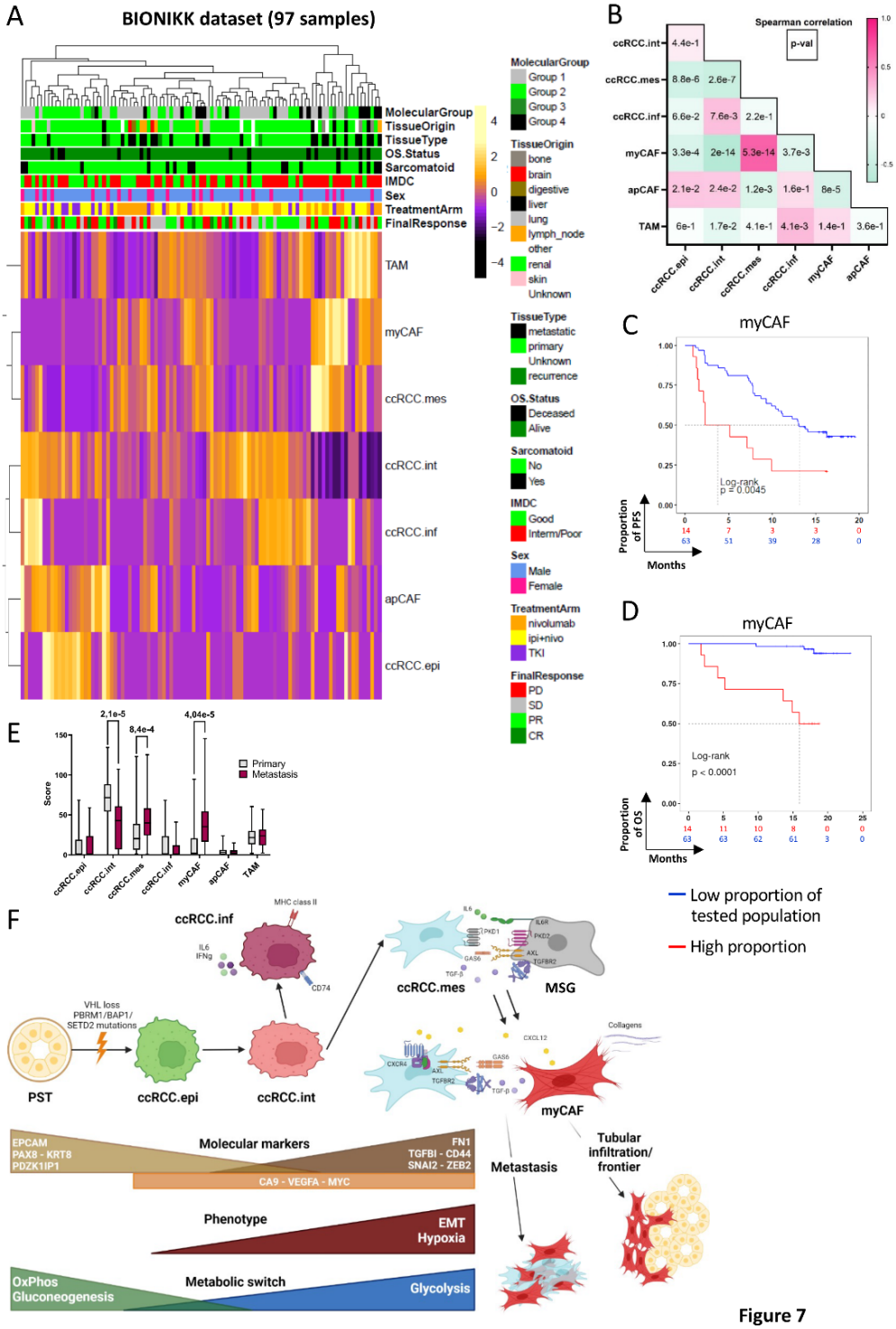


Figure 7

Figure 7. Deconvolution of the BIONIKK dataset. A. Heatmap showing deconvolution of ccRCC, CAFs and TAM signatures inferred by CIBERSORTx and displayed as row-scaled absolute scores on bulk RNA-seq data from 97 BIONIKK tumour samples. **B.** Spearman correlation coefficient (coloured box) and associated p-value (number in box) between the indicated populations. **C-D.** Kaplan-Meier curves for progression-free (C) and overall (D) survival according to myCAF proportions using optimal cut-off value method. **E.** Box-plot representation of differences in deconvolution scores for the indicated populations between metastatic and primary samples. Significant p-values are indicated. **F.** Schematic model showing PST transformation in ccRCC.epi cells, EMT gradient from ccRCC.epi to ccRCC.mes cells through an intermediate state. Correlation of ccRCC.mes and myCAF to promote metastasis, ICI resistance and tubular infiltration. Created with Biorender.com

7A-B and Fig. S12A-C, Dataset S4). BIONIKK was based on tumour classification in 4 groups (ccRCC1-4): ccRCC2 and ccRCC3 designated "pro-angiogenic" compared to less differentiated and immune-rich ccRCC1 and ccRCC4 (31). Computing the average score for each group, with the exception of ccRCC3 not well represented in the sequenced samples, indicated that the "angiogenic" ccRCC2 group was enriched in ccRCC.epi and endothelial ED.cor, whereas ccRCC.mes, myCAFs and TAMs were strongly enriched in the ccRCC4 group and ccRCC.inf in ccRCC1 (Dataset S4).

We then associated cell populations with progression-free survival (PFS) and overall survival (OS) of patients treated by either nivolumab or ipilimumab+nivolumab combination (n=77). Patients with high myCAF signature (n=14) displayed a significantly shorter PFS of 3.7 months (95% CI, 2.1 to NE), compared to 13.1 months (95% CI, 10.4 to NE) for the remaining patients (HR for myCAF-high group, 2.62; 95% CI, 1.32 to 5.12; p=0.005) (Fig. 7C). Furthermore, those patients displayed a shorter median OS of 15.9 months (95% CI, 13.6 to NE) compared to myCAF-low where the median was not reached (HR for myCAF-high group, 15.26; 95% CI, 3.91 to 59.67; p<0.001) (Fig. 7D). Notably, the myCAF-high group showed lower overall response rate (ORR) and clinical benefit rate (CBR) both being 35.71% (5 partial responses, 9 progressive diseases) compared to myCAF-low displaying an ORR of 73.02% and a CBR of 52.38% (8 complete responses, 25 partial responses, 16 stable diseases, 14 progressive diseases).

Despite the association of the ccRCC.mes and myCAF populations, no significant association of high ccRCC.mes score with OS was seen (Fig. S12D). Nevertheless, 11 of the 14 myCAF-high tumours associated with high ccRCC.mes score and represented the 5 mortality events in this group, with no events seen amongst the 17 myCAF-low patients (Fig. S12E). Thus, although there was association of the ccRCC.mes and myCAF populations, only a high myCAF score showed strong association with ICI resistance.

i. Enrichment for ccRCC.mes and myCAFs at metastatic sites

BIONIKK samples comprise RNA-seq data from metastatic tumour sites including liver (n=5), lymph node (n=5), lung (n=4), bone (n=3), brain (n=2), skin (n=1), digestive system (n=1) and unknown locations (n=5). Comparison of primary (n=70) and metastatic tumours (n=26) revealed selective enrichment of only the ccRCC.mes and myCAF populations at the metastatic sites (Fig. 7E). This observation highlighted the importance of these populations and of the ccRCC.mes invasive profile in local infiltration and metastatic dissemination (Fig. 7F).

D. Discussion.

a. Tumour cell state association with ccRCC disease outcome.

We established a comprehensive characterization of ccRCC and the associated TME defining up to 46 cell populations including 5 tumour cell populations. The ccRCC.epi population, enriched in epithelial markers, showed highest signature similarity to the PST population, the proposed cell of ccRCC origin (10,32,33). We defined tumour cell plasticity by characterizing hybrid EMT and fully mesenchymal cells with high activity of EMT-related transcription factors and a population expressing MHC class II genes with transcription factors associated with interferon and inflammatory responses. This latter subtype was observed in other cancers such as melanoma (20, 33,34). Deconvolution of the TCGA KIRC and BIONIKK data sets defined the contribution of the ccRCC cell states to tumour heterogeneity. CcRCC.mes and ccRCC.inf cells were highly enriched in advanced grade tumours and associated with poor prognosis, while ccRCC.epi was enriched in low grade tumours with better outcome. Moreover, in line with previous studies linking EMT to ccRCC metastasis (36), ccRCC.mes was the only ccRCC population enriched at metastatic sites. Our results therefore

highlight how tumour cell states associate with disease progression, with the mesenchymal and inflamed states predicting poor patient prognosis.

Our analyses further defined the transcriptional basis for ccRCC plasticity identifying transcription factors and associated programs involved in oncogenic transformation and EMT. We identified potential tumour suppressors active in PST but lost in ccRCC such as TP53 (37), FOXP4, previously reported as expressed in kidney but diminished in tumours (38), and PPARG shown to maintain epithelial phenotype in renal fibrogenesis (39). Some factors linked to hypoxia response including MYC and CEBPD (40) were active in all ccRCC cells likely resulting from *VHL* inactivation. In contrast, others were linked to a specific cell state, for example ZEB and SNAI families active in ccRCC.mes, or SOX4, SOX13, BRN1 (POU3F3) and OCT4 (POU5F1) active in dedifferentiating ccRCC cells. These results show how loss of factors maintaining terminal differentiation result in MYC activation and loss of epithelial identity in line with a previously proposed model (41).

While several previous ccRCC scRNA-seq studies focused mainly on the immune and myeloid compartments (7,8,14) Bi et al., (7) described 2 tumour cell populations designated TP1 and TP2 whose transcriptional signatures partially overlapped with the ccRCC.epi/int signatures described here. More recently, while this manuscript was under review, Li et.al. (42) identified several tumour cell meta-programs (MPs). The stress-response-related (MP1) resembles ccRCC.int1 characterized by expression of *JUNB, FOSB, DNAJB1* and *ATF3*. MP3 displaying a mesenchymal program is analogous to ccRCC.mes and was found enriched in the TCGA m3 group and characterized by *CEBPB* that we also identified as a key regulon of undifferentiated states. Proximal tubule MP2 is similar to ccRCC.epi with MP2 and MP3 anti-correlated analogous to ccRCC.epi and ccRCC.mes. MP5 shows the characteristics of ccRCC.inf with expression of the MHC-II genes, confirming the existence of this cell type in

ccRCC. These two complementary studies thus converge on ccRCC cell plasticity and how it contributes to ccRCC heterogeneity and disease outcome.

b. MyCAF enrichment at the tumour/NAT interface and association with resistance to ICI therapy.

We defined 2 *RGS5*-expressing pericyte populations one displaying a weak apCAF-type signature. Trajectory analyses suggested that pericytes gave rise to the MSG.inf population, a precursor to *POSTN*-expressing apCAFs, via IL-1α response pathway or to the myCAF population via a TGF-β driven pathway as described in pancreatic cancer (43). Computational analyses predicted ligand-receptor interactions whereby ccRCC.mes may induce pericyte conversion to myCAF, via the TGF-β, IL6-IL6R, PKD1-PKD2 and the AXL-GAS6 pathways. MyCAFs may therefore be derived mainly from the resident pericytes lining the glomeruli and peritubular capillaries of the normal kidney upon interactions with dedifferentiated tumour cells.

We identified myCAFs as a critical component of the ccRCC TME that strongly associated with ccRCC.mes in primary and metastatic tumours and robustly associated with poor ccRCC prognosis. MyCAFs seem also to play a critical role in ccRCC tumour organization. While ccRCC.epi/int cells were widespread within tumours, spatial transcriptomics and multiplex immune-staining revealed a proximity of ccRCC.mes and myCAFs that may account for their strong association and facilitate their signaling to one another. MyCAFs form fibrotic structures seen at the tumour-NAT boundary and sometimes invaded the NAT. CcRCC.mes were also preferentially localized in these regions in agreement with Li et al (41) who localized these cells at the leading edge of the tumour. Li et al., did not however analyze the presence of myCAFs at the tumour-NAT interface, but rather noted enrichment in a subtype of IL-1β-high macrophages. Thus, both studies underscore important

characteristics of ccRCC tumour spatial organization highlighting the potential roles of mesenchymal tumour cells, myCAFs and IL-1β-high macrophages in the invasive process.

Computational analyses identified signaling pathways that could mutually reinforce the myCAF and ccRCC.mes transcriptional programs with a prominent role of the TGF-β pathway with ligands and receptors expressed in both cell types. However, other pathways were identified such as JAG1-NOTCH interaction. NOTCH signaling has been associated with EMT, stemness and cell plasticity (44). In breast cancer, different CAFs promote EMT via TGF-β and CXCL12 or invasion by NOTCH signaling (45). MyCAFs may signal to ccRCC.mes through TGF-β and CXCL12 as well as by NOTCH suggesting they promote both EMT and invasion in agreement with their association with metastases and poor prognosis.

The AXL-GAS6 pathway is potentially a key pathway in myCAF-ccRCC.mes crosstalk and a novel potential therapeutic target. High tumour cell AXL expression is associated with resistance to anti-PD-L1 ICI in ccRCC (46). These observations underscore the potential role of this pathway in ccRCC.mes-myCAF communication.

A key conclusion of our analyses was the robust association of myCAFs, but not ccRCC.mes or other tumour cell states, with resistance to nivolumab or ipilimumab+nivolumab ICI both in terms of OS and PFS in the BIONNIK clinical trial. While the sunitinib-treatment arm of the BIONIKK trial was too small to provide a suitable control, the selective association with poor OS and PFS indicated that high myCAF score was likely predictive of patient response to ICI therapy. Moreover, our data suggest that patients with myCAF-high tumours that are more likely to show ICI resistance may benefit more from a combinatorial TKI-ICI treatment to target the myCAF-ccRCC.mes crosstalk, ideally one targeting both AXL and PDGFR such as sunitinib or axitinib, lenvatinib, cabozantinib or sorafenib. Additional targetable ligand-receptor couples may be relevant, such as the cell surface receptor NRP1 targeted by monoclonal antibody ASP1948 under evaluation in combination with a PD-1

inhibitor in clinical trial NCT03565445 and by the receptor antagonist EG00229 that showed efficacy in a ccRCC xenograft model when combined with everolimus (47). Other potential targets are MDK that has been successfully inhibited by an MDK inhibitor (48) or EGFR targeted by TKIs such as afatinib or by monoclonal antibodies cetuximab and panitumumab.

MyCAFs have been associated with immunosuppressive microenvironment and ICI resistance in other tumours. Analyses of myCAF populations revealed a positive association with PD-1 and CTLA4-expression and primary resistance to ICI therapy in several cancers perhaps forming a barrier excluding immune infiltration (49). Alternatively, myCAFs express high levels of TGF-β known to attenuate anti-PD-L1 response (50). Our data revealed a novel association of myCAFs with ICI resistance in ccRCC further expanding the role of these cells as important modulators of ICI therapy.

Acknowledgements

We thank the staff of the IGBMC common facilities, especially Dr. C. Carpentier and Ms M. Cerciat for their precious help with single-cell techniques, and Dr. E. Guiot for her great guidance within the microscope facility. This work was supported by grants from the ARC foundation (SIGNIT) and ARC-2020-PJA32020070002319, the Ligue Nationale contre le Cancer, the Institut National du Cancer, the ANR-10-LABX-0030 and ANR-10-IDEX-0002-02. ID is an ‘équipe labellisée’ of the Ligue Nationale contre le Cancer. GD was supported by the Région Grand Est, AH was supported by the Ministère de la Recherche. Sequencing was performed by the GenomEast platform, a member of the ‘France Génomique’ consortium (ANR-10-INBS-0009). We would also like to thank Drs V. Debien, J. Gantzer and P. Baltzinger for their insightful comments.

Author Contributions.

Conceptualization: AH, GD, ID, GGM ; Methodology: AH (single-cell suspension, RNA extraction), MC (single-cell and spatial libraries), AF (spatial) ; Sample collection: AH, YAV, VL, RE, VV, CS, CC, MB, HL, TT, WHF, CSF ; Bioinformatics: GD, AH, XS, DP, CK ; Formal analysis: GD, AH, ID, GGM ; Investigations: GD, AH, YAV, VL, AF, MC, RE, VV, CS, CC, MB, HL, TT, WHF, CSF, XS, DP, CK, CTC, PB, SO, ID, GGM ; Writing – original draft: GD, AH, ID; Writing – review and editing: GD, AH, ID, GGM ; Supervision: ID, GGM

E. References

1. Du Z, Chen W, Xia Q, Shi O, Chen Q. Trends and projections of kidney cancer incidence at the global and national levels, 1990–2030: a Bayesian age-period-cohort modeling study. Biomark Res. 2020;8:16.

2. Lindgren D, Sjölund J, Axelson H. Tracing Renal Cell Carcinomas back to the Nephron. Trends Cancer. 2018;4:472–84.

3. Czyzyk-Krzeska MF, Landero Figueroa JA, Gulati S, Cunningham JT, Meller J, ShamsaeI B, et al. Molecular and Metabolic Subtypes in Sporadic and Inherited Clear Cell Renal Cell Carcinoma. Genes. 2021;12:388.

4. Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010;463:360–3.

5. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. Cell. 2018;173:611-623.e17.

6. Turajlic S, Xu H, Litchfield K, Rowan A, Chambers T, Lopez JI, et al. Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. Cell. 2018;173:581-594.e12.

7. Bi K, He MX, Bakouny Z, Kanodia A, Napolitano S, Wu J, et al. Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. Cancer Cell. 2021;39:649-661.e5.

8. Krishna C, DiNatale RG, Kuo F, Srivastava RM, Vuong L, Chowell D, et al. Single-cell sequencing links multiregional immune landscapes and tissue-resident T cells in ccRCC to tumor topology and therapy efficacy. Cancer Cell. 2021;39:662-677.e6.

9. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Forman J, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. Nat Med. Nature Publishing Group; 2020;26:909–18.

10. Zhang Y, Narayanan SP, Mannan R, Raskind G, Wang X, Vats P, et al. Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response. Proc Natl Acad Sci. National Academy of Sciences; 2021;118.

11. Fiori ME, Di Franco S, Villanova L, Bianca P, Stassi G, De Maria R. Cancer-associated fibroblasts as abettors of tumor progression at the crossroads of EMT and therapy resistance. Mol Cancer. 2019;18:70.

12. Li B, Pei G, Yao J, Ding Q, Jia P, Zhao Z. Cell-type deconvolution analysis identifies cancer-associated myofibroblast component as a poor prognostic factor in multiple cancer types. Oncogene. 2021;40:4686–94.

13. Braun DA, Street K, Burke KP, Cookmeyer DL, Denize T, Pedersen CB, et al. Progressive immune dysfunction with advancing disease stage in renal cell carcinoma. Cancer Cell. 2021;39:632-648.e8.

14. Wu Y, Tamayo P, Zhang K. Visualizing and Interpreting Single-Cell Gene Expression Datasets with Similarity Weighted Nonnegative Embedding. Cell Syst. 2018;7:656-666.e4.

15. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat Protoc. 2020;15:1484–506.

16. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17:159–62.

17. Borcherding N, Vishwakarma A, Voigt AP, Bellizzi A, Kaplan J, Nepple K, et al. Mapping the immune environment in clear cell renal carcinoma by single-cell genomics. Commun Biol. 2021;4:1–11.

18. Barkley D, Moncada R, Pour M, Liberman DA, Dryg I, Werba G, et al. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. Nat Genet. Nature Publishing Group; 2022;54:1192–201.

19. Rambow F, Rogiers A, Marin-Bejar O, Aibar S, Femel J, Dewaele M, et al. Toward Minimal Residual Disease-Directed Therapy in Melanoma. Cell. 2018;174:843-855.e19.

20. Berico P, Cigrang M, Davidson G, Braun C, Sandoz J, Legras S, et al. CDK7 and MITF repress a transcription program involved in survival and drug tolerance in melanoma. EMBO Rep. 2021;22:e51683.

21. Lake BB, Chen S, Hoshi M, Plongthongkum N, Salamon D, Knoten A, et al. A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. Nat Commun. Nature Publishing Group; 2019;10:1–15.

22. Young MD, Mitchell TJ, Vieira Braga FA, Tran MG, Stewart BJ, Ferdinand JR, et al. Single cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science. 2018;361:594–9.

23. Lee JW, Chou C-L, Knepper MA. Deep Sequencing in Microdissected Renal Tubules Identifies Nephron Segment–Specific Transcriptomes. J Am Soc Nephrol JASN. 2015;26:2669–77.

24. Chen Z, Zhou L, Liu L, Hou Y, Xiong M, Yang Y, et al. Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. Nat Commun. Nature Publishing Group; 2020;11:5077.

25. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. Cancer Discov. 2019;9:1102–23.

26. TCGA. COMPREHENSIVE MOLECULAR CHARACTERIZATION OF CLEAR CELL RENAL CELL CARCINOMA. Nature. 2013;499:43–9.

27. Meylan M, Petitprez F, Becht E, Bougoüin A, Pupier G, Calvez A, et al. Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. Immunity. 2022;55:527-541.e5.

28. Huang Z, Cheng C, Xiong H, Wang Y, Chen KK, Yang J, et al. NRP1 promotes cell migration and invasion and serves as a therapeutic target in nasopharyngeal carcinoma. Int J Clin Exp Pathol. 2018;11:2460–9.

29. Epaillard N, Simonaggio A, Elaidi R, Azzouz F, Braychenko E, Thibault C, et al. BIONIKK: A phase 2 biomarker driven trial with nivolumab and ipilimumab or VEGFR tyrosine kinase inhibitor (TKI) in naïve metastatic kidney cancer. Bull Cancer (Paris). 2020;107:eS22–7.

30. Vano Y-A, Elaidi R, Bennamoun M, Chevreau C, Borchiellini D, Pannier D, et al. Nivolumab, nivolumab–ipilimumab, and VEGFR-tyrosine kinase inhibitors as first-line treatment for metastatic clear-cell renal cell carcinoma (BIONIKK): a biomarker-driven, open-label, non-comparative, randomised, phase 2 trial. Lancet Oncol. Elsevier; 2022;23:612–24.

31. Beuselinck B, Job S, Becht E, Karadimou A, Verkarre V, Couchy G, et al. Molecular subtypes of clear cell renal cell carcinoma are associated with sunitinib response in the metastatic setting. Clin Cancer Res Off J Am Assoc Cancer Res. 2015;21:1329–39.

32. Lindgren D, Eriksson P, Krawczyk K, Nilsson H, Hansson J, Veerla S, et al. Cell-Type-Specific Gene Programs of the Normal Human Nephron Define Kidney Cancer Subtypes. Cell Rep. 2017;20:1476–89.

33. Polascik TJ, Bostwick DG, Cairns P. Molecular genetics and histopathologic features of adult distal nephron tumors. Urology. 2002;60:941–6.

34. Axelrod ML, Cook RS, Johnson DB, Balko JM. Biological Consequences of MHC-II Expression by Tumor Cells in Cancer. Clin Cancer Res. 2019;25:2392–402.

35. Donia M, Andersen R, Kjeldsen JW, Fagone P, Munir S, Nicoletti F, et al. Aberrant Expression of MHC Class II in Melanoma Attracts Inflammatory Tumor-Specific CD4+ T- Cells, Which Dampen CD8+ T-cell Antitumor Reactivity. Cancer Res. 2015;75:3747–59.

36. Kim K, Zhou Q, Christie A, Stevens C, Ma Y, Onabolu O, et al. Determinants of renal cell carcinoma invasion and metastatic competence. Nat Commun. Nature Publishing Group; 2021;12:5760.

37. Aubrey BJ, Strasser A, Kelly GL. Tumor-Suppressor Functions of the TP53 Pathway. Cold Spring Harb Perspect Med. 2016;6:a026062.

38. Teufel A, Wong EA, Mukhopadhyay M, Malik N, Westphal H. FoxP4, a novel forkhead transcription factor. Biochim Biophys Acta. 2003;1627:147–52.

39. Zhao M, Chen Y, Ding G, Xu Y, Bai M, Zhang Y, et al. Renal tubular epithelium-targeted peroxisome proliferator-activated receptor-γ maintains the epithelial phenotype and antagonizes renal fibrogenesis. Oncotarget. 2016;7:64690–701.

40. Yamaguchi J, Tanaka T, Eto N, Nangaku M. Inflammation and hypoxia linked to renal injury by CCAAT/enhancer-binding protein δ. Kidney Int. 2015;88:262–75.

41. Gu X, Enane F, Tohme R, Schuerger C, Radivoyevitch T, Parker Y, et al. PBRM1 loss in kidney cancer unbalances the proximal tubule master transcription factor hub to repress proximal tubule differentiation. Cell Rep. 2021;36:109747.

42. Li R, Ferdinand JR, Loudon KW, Bowyer GS, Laidlaw S, Muyas F, et al. Mapping single-cell transcriptomes in the intra-tumoral and associated territories of kidney cancer. Cancer Cell. 2022;40:1583-1599.e10.

43. Biffi G, Oni TE, Spielman B, Hao Y, Elyada E, Park Y, et al. IL1-Induced JAK/STAT Signaling Is Antagonized by TGFβ to Shape CAF Heterogeneity in Pancreatic Ductal Adenocarcinoma. Cancer Discov. 2019;9:282–301.

44. Espinoza I, Miele L. Deadly crosstalk: Notch signaling at the intersection of EMT and cancer stem cells. Cancer Lett. 2013;341:41–5.

45. Pelon F, Bourachot B, Kieffer Y, Magagna I, Mermet-Meillon F, Bonnet I, et al. Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. Nat Commun. Nature Publishing Group; 2020;11:404.

46. Terry S, Dalban C, Rioux-Leclercq N, Adam J, Meylan M, Buart S, et al. Association of AXL and PD-L1 Expression with Clinical Outcomes in Patients with Advanced Renal Cell Carcinoma Treated with PD-1 Blockade. Clin Cancer Res. 2021;27:6749–60.

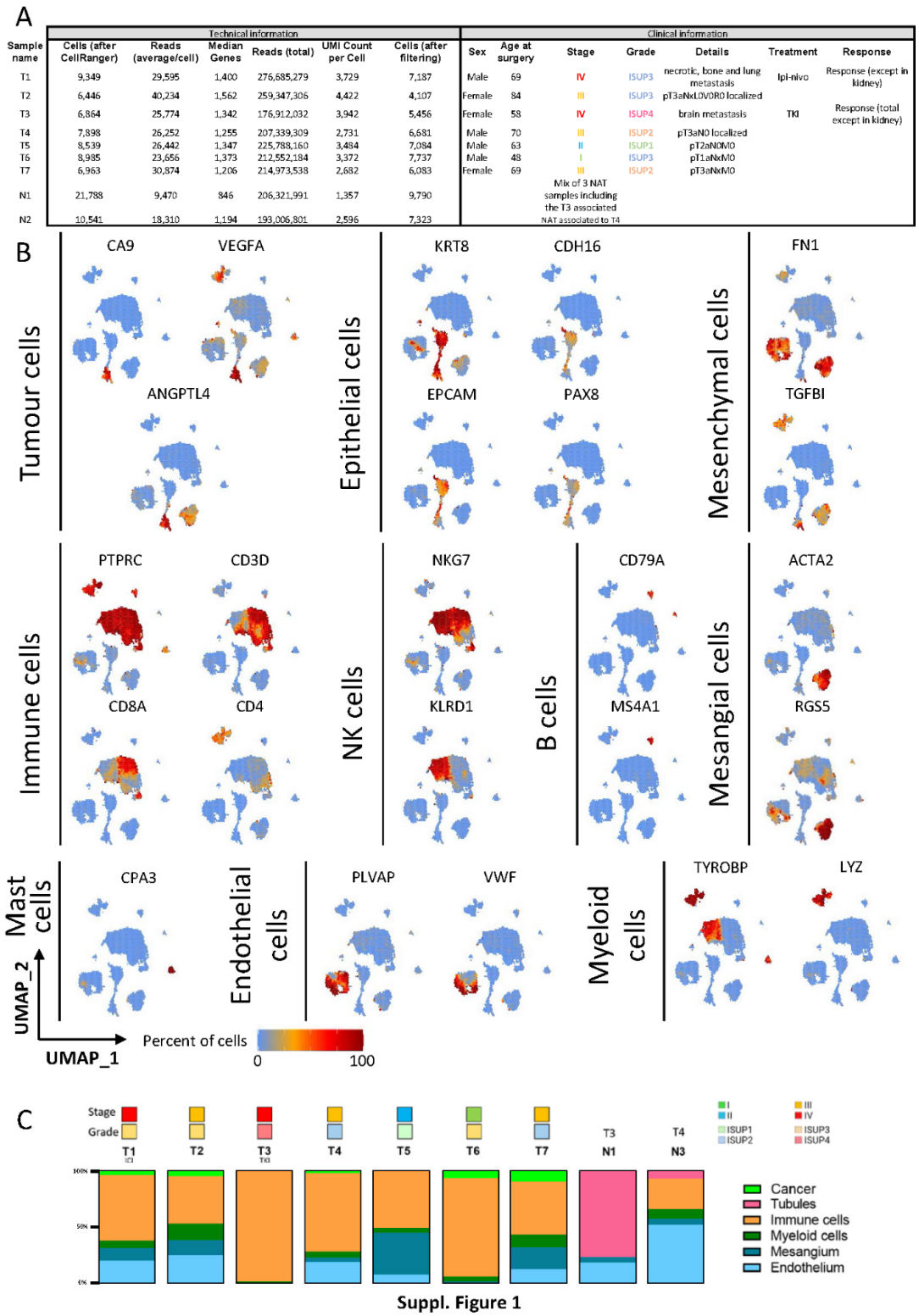
47. Pal K, Madamsetty VS, Dutta SK, Wang E, Angom RS, Mukhopadhyay D. Synchronous inhibition of mTOR and VEGF/NRP1 axis impedes tumor growth and metastasis in renal cancer. NPJ Precis Oncol. 2019;3:31.

48. Erdogan S, Doganlar ZB, Doganlar O, Turkekul K, Serttas R. Inhibition of Midkine Suppresses Prostate Cancer CD133+ Stem Cell Growth and Migration. Am J Med Sci. 2017;354:299–309.

49. Kieffer Y, Hocine HR, Gentric G, Pelon F, Bernard C, Bourachot B, et al. Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer. Cancer Discov. 2020;10:1330–51.

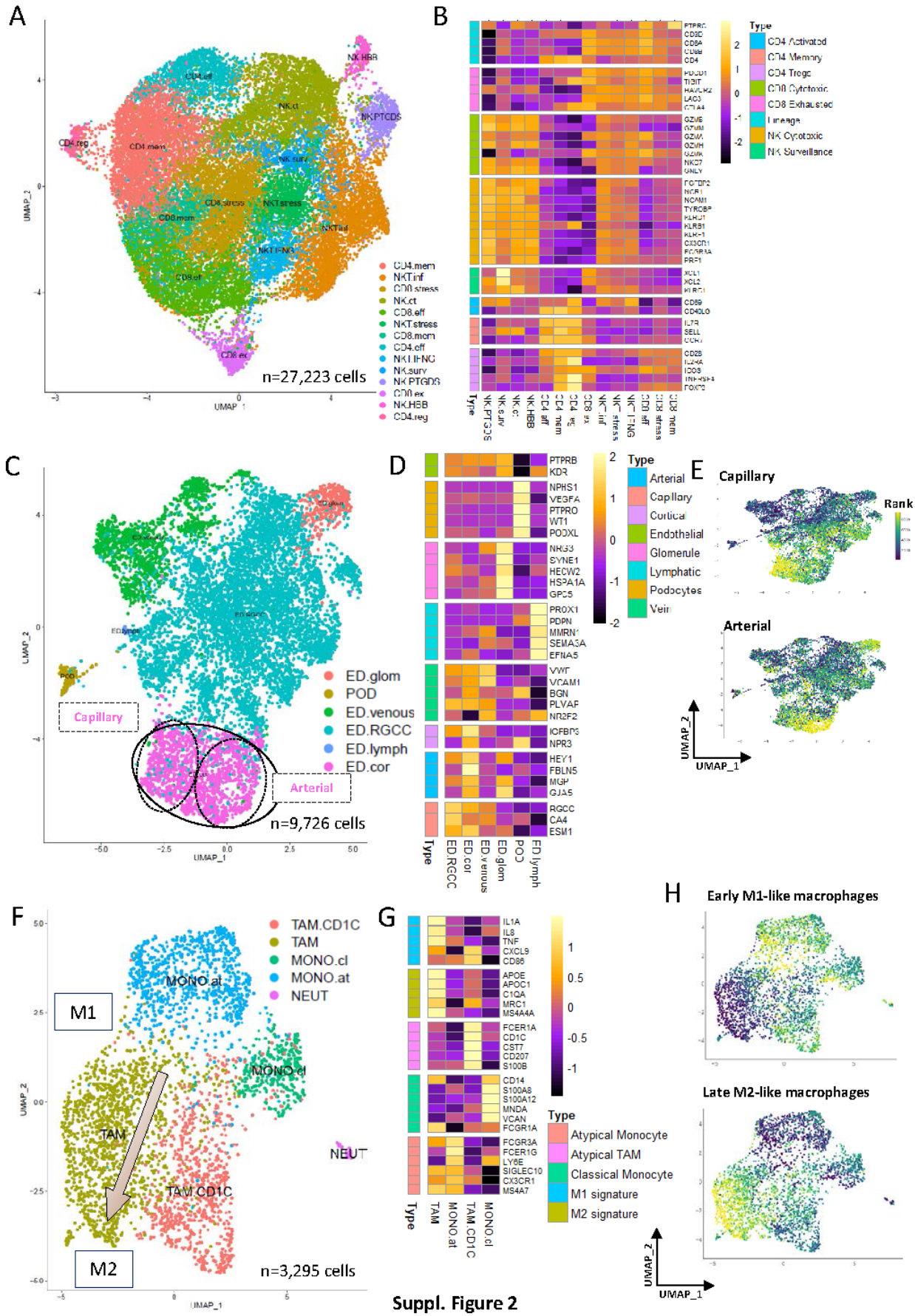
50. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018;554:544–8.

F. Supplementary figures

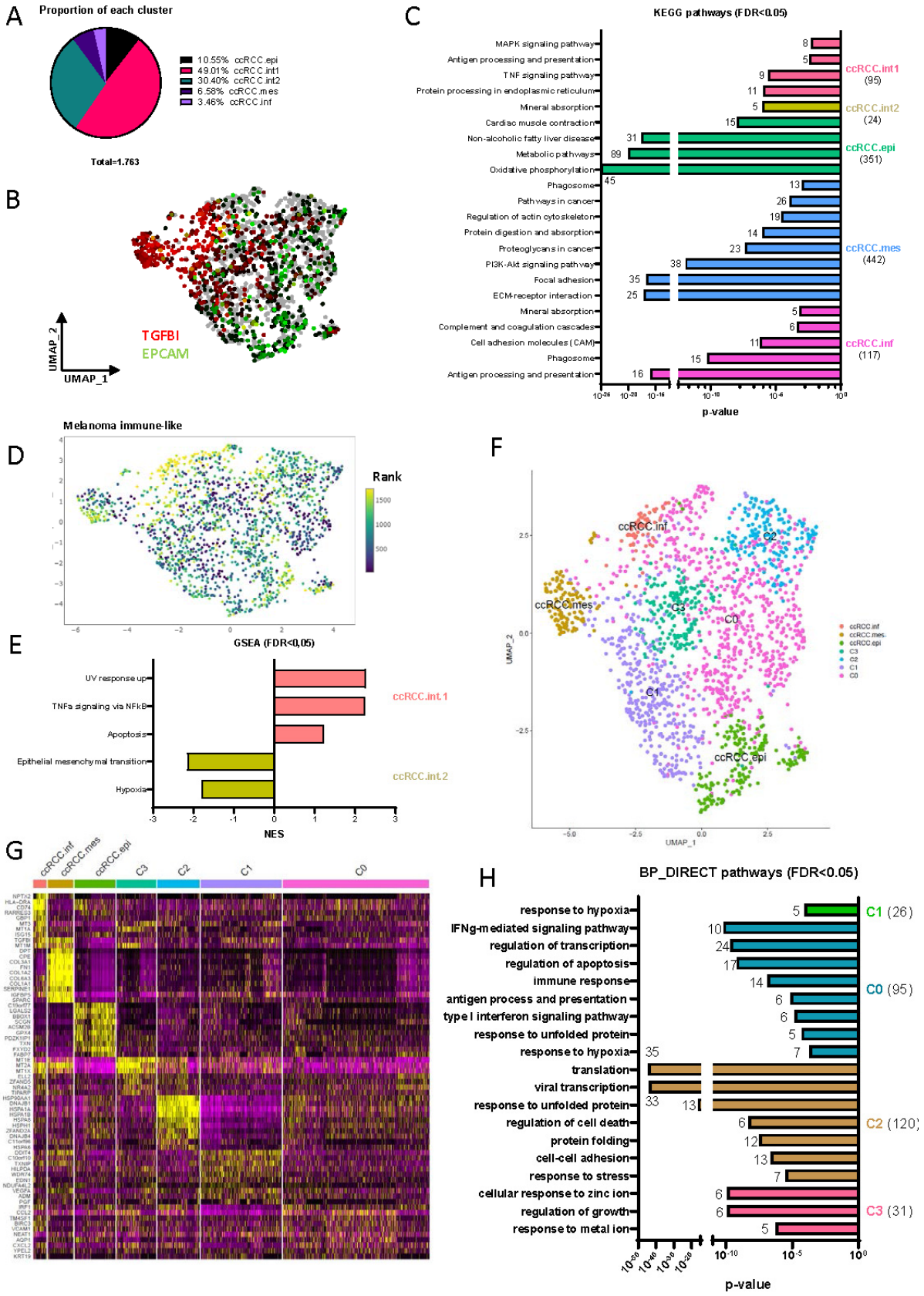


Suppl. Figure 1

Supplementary Figure 1. Phenotyping of main cell populations. A. Metrics (number of cells, reads, median genes, UMI) after sequencing of each sample and their associated clinical characteristics. **B.** Expression of common markers for tumour, immune and stromal cells and of epithelial and mesenchymal states represented as percentage of expressing cells after dividing the UMAP into 80 hexagonal bins. **C.** Proportion of cell types per sample.

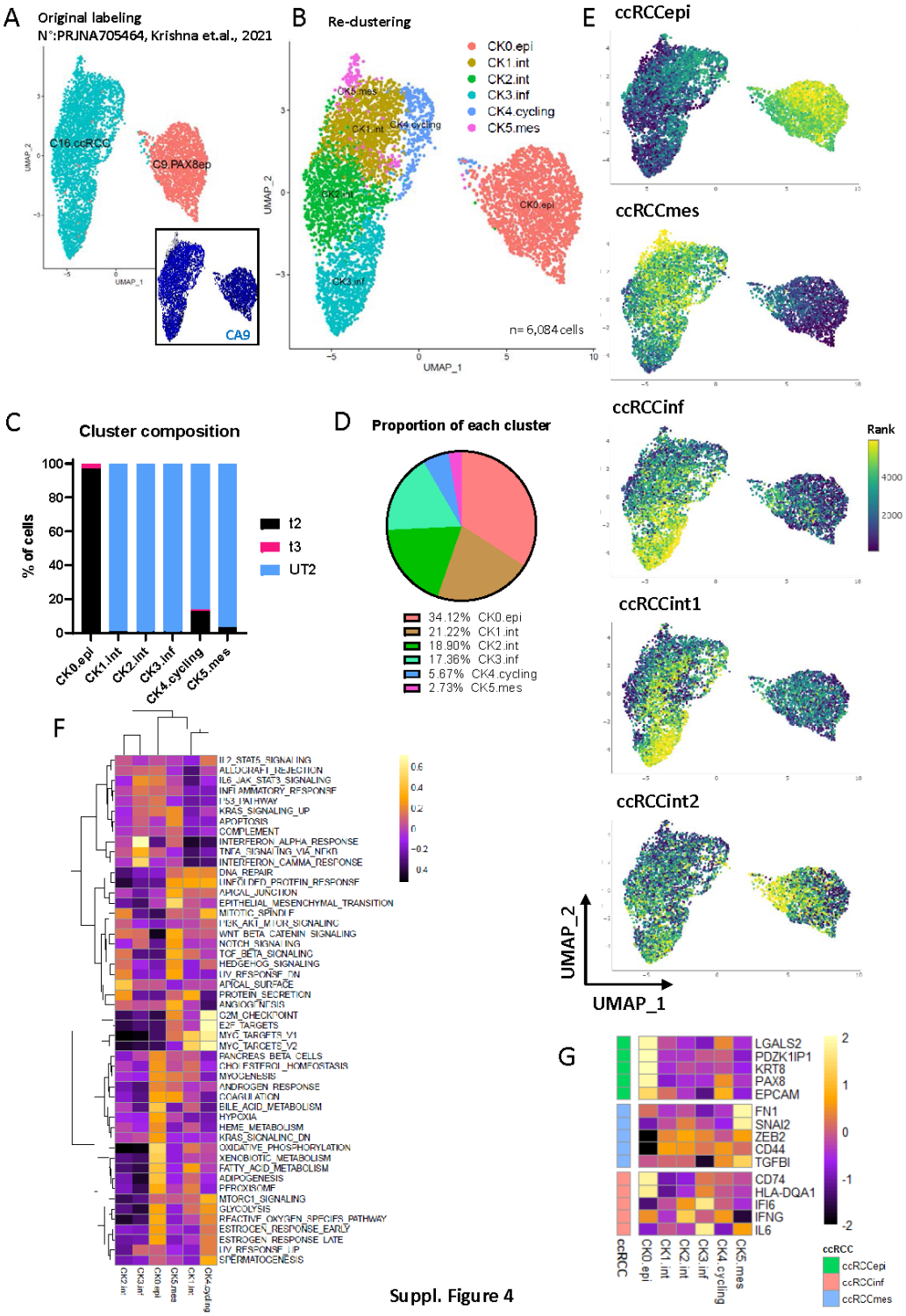

A
B
C
D
E
F
G
H
n=27,223 cells
n=9,726 cells
n=3,295 cells
Capillary
Arterial
M1
M2
TAM
MONO.at
MONO.cl
NEUT
TAM.CD1C
Early M1-like macrophages
Late M2-like macrophages
Rank
Type
Suppl. Figure 2

Supplementary Figure 2. Diversity in lymphoid, myeloid and endothelial populations. **A.** UMAP plot of lymphoid sub-clusters (27,223 cells). **B.** Pseudo-bulk heatmap of identity markers for each lymphoid sub-cluster. *Abbreviations:* CD4.reg: regulatory CD4-T cells; CD4.eff: effector CD4-T cells; CD4.mem: memory CD4-T cells; CD8.mem: memory CD8-T cells; CD8.eff: effector CD8-T cells; CD8.ex: exhausted CD8-T cells; CD8.stress: stressed CD8-T cells; NKT.inf: inflamed natural killer-like T cells; NKT.IFNG: NKT cells expressing IFNG as top marker; NKT.stress: stressed NKT cells; NK.HBB: natural killer cells expressing HBB as top marker; NK.PTGDS: NK cells expressing PTGDS as top marker; NK.surv: surveillance NK cells; NK.ct: cytotoxic NK cells. **C.** UMAP plot of endothelial sub-clusters (9,726 cells). *Abbreviations:* ED.cor: cortical endothelial cells; ED.lymph: lymphatic endothelial cells; ED.RGCC: endothelial cell expressing RGCC as top marker; ED.venous: venous endothelial cells; ED.glom: glomerular endothelial cells; POD: podocytes. **D.** Pseudo-bulk heatmap of identity markers for each endothelial sub-cluster. **E.** VISION projection of capillary and arterial signatures coloured by value ranks showing a division of the cortical endothelial sub-cluster. **F.** UMAP plot of myeloid sub-clusters showing the gradient from M1-like to M2-like macrophages (3,295 cells). *Abbreviations:* NEUT: neutrophils; MONO.cl: classical monocyte; MONO.at: atypical monocytes; TAM: tumour-associated macrophages; TAM.CD1C: TAM expressing CD1c as top marker. **G.** Pseudo-bulk heatmap of identity markers of each monocyte and macrophage sub-cluster. **H.** VISION projection of M1- and M2-like macrophages signatures coloured by value ranks highlighting a gradient in the TAM sub-cluster.



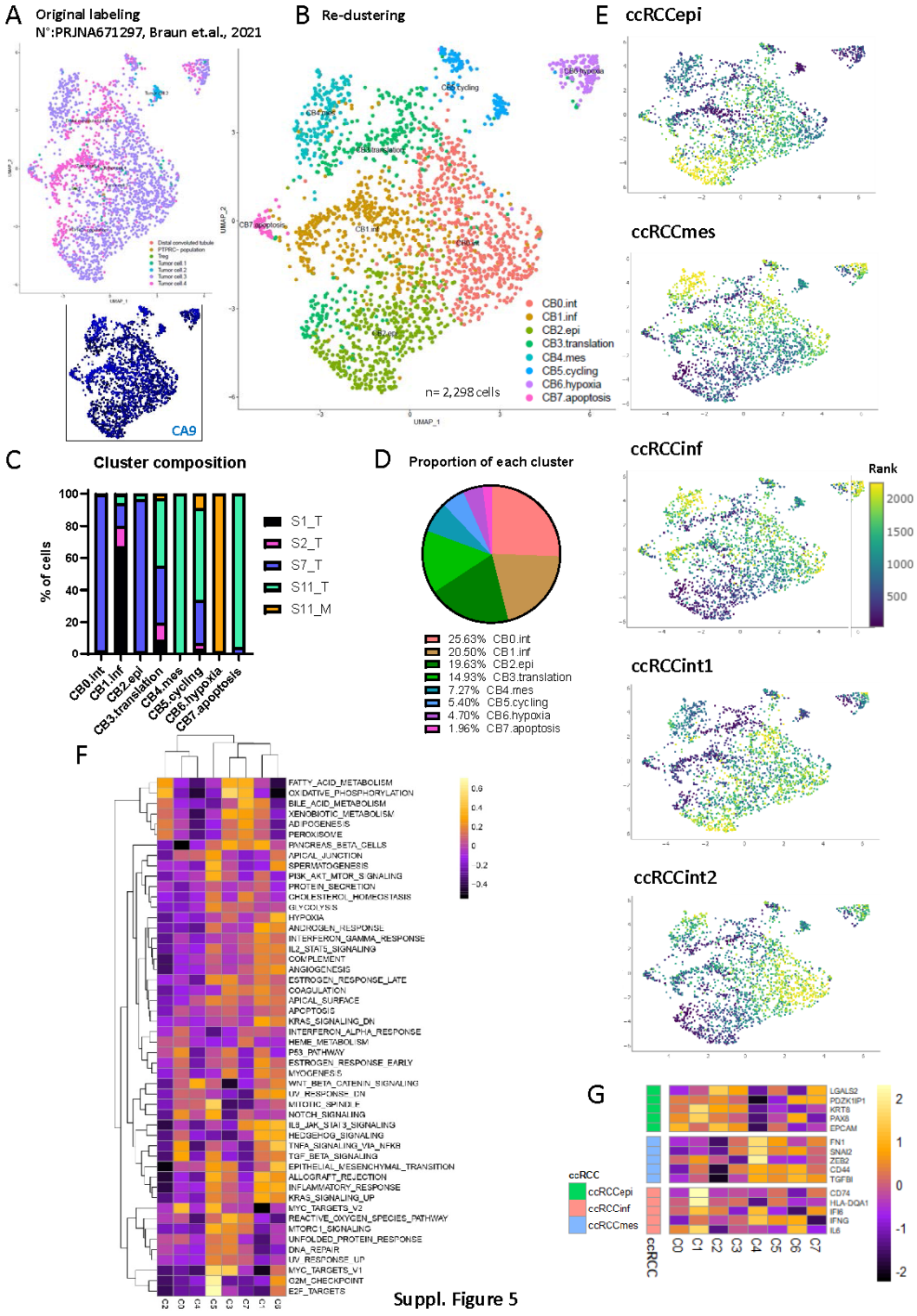
Suppl. Figure 3

Supplementary Figure 3. Heterogeneity in the ccRCC.int populations. A. Pie chart displaying the proportion of each ccRCC subtypes in the total tumour collection. **B.** UMAP projection of gene expression for *TGFBI* (red) and *EPCAM* (green) in ccRCC tumour cells. **C.** KEGG pathway ontology analysis of specific markers for tumour sub-clusters (number of markers for each cluster indicated in brackets; number of genes found in each pathway noted near the bar). **D.** VISION projection of the melanoma immune-like signature. **E.** GSEA analysis of ccRCC.int1 versus ccRCC.int2. **F.** UMAP plot showing ccRCC.int sub-clustering. Each ccRCC.int is divided into two sub-clusters (ccRCC.int1 in C0 and C2; ccRCC.int2 in C1 and C3). **G.** Heatmap of the 10-top markers for each sub-cluster. **H.** DAVID gene ontology analysis of specific markers for the four new formed sub-clusters using BP_DIRECT (number of markers for each cluster indicated in brackets; number of genes found in each pathway noted near the bar).

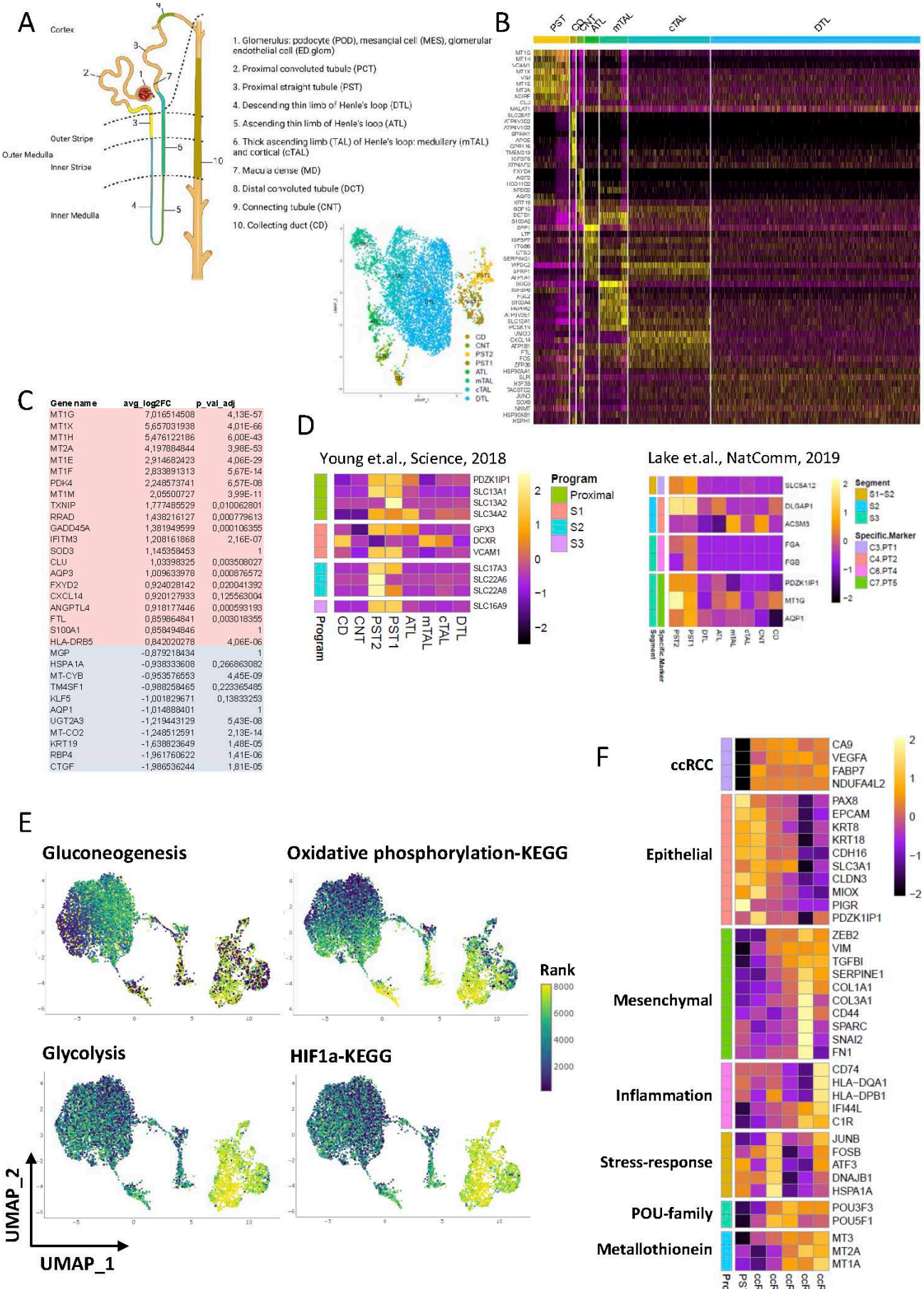


Suppl. Figure 4

Supplementary Figure 4. Validation of ccRCC tumour populations in dataset PRJNA705464 from Krishna et.al. A. Original labelling superimposed on the UMAP plot obtained after re-analysis of CA9-expressing cells. **B.** UMAP plot showing identified sub-clusters. **C.** Bar plot displaying the contribution of 3 tumour samples to each identified cluster. Patient samples comprising less than 20 annotated tumour cells were not included. **D.** Piechart displaying the global proportion of each identified cluster. **E.** VISION projections of indicated ccRCC signatures. **F.** Heatmap representation of the GSVA analysis showing specific hallmarks enriched in each cluster. **G.** Pseudo-bulk heatmap of 5 specific markers for epithelial, mesenchymal and inflamed-state cells.



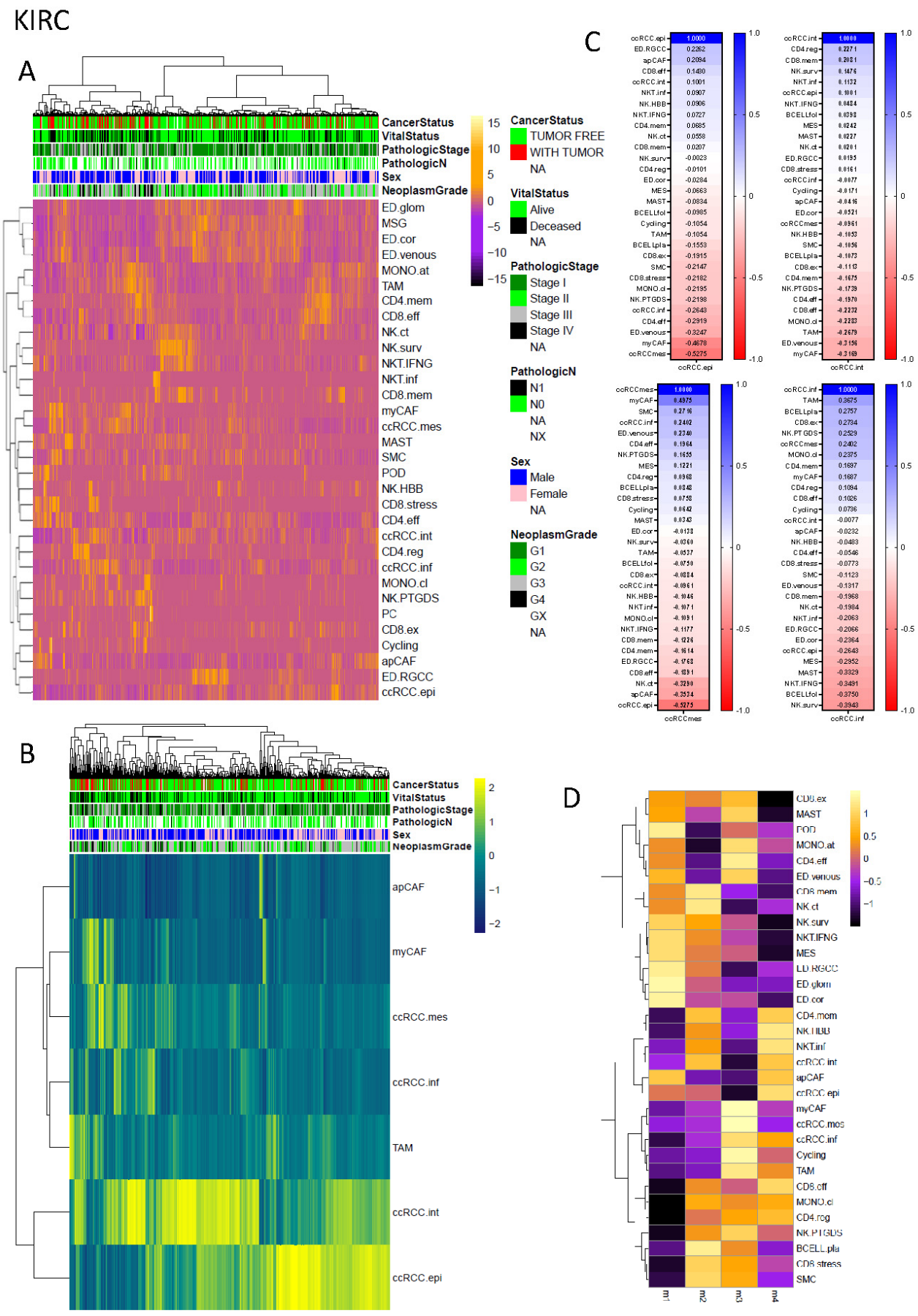
Supplementary Figure 5. Validation of ccRCC tumour populations in PRJNA671297 from Braun et.al. A. Original labelling superimposed on the UMAP plot obtained after our re-analysis and removal of misannotated cells (Dataset: on dbGaP, accession number phs002252.v1.p1). **B.** UMAP plot showing identified sub-clusters. **C.** Bar plot displaying the contribution of 5 tumour samples to each identified cluster. Patient samples comprising less than 20 annotated tumour cells were not included. **D.** Piechart displaying the global proportion of each identified cluster. **E.** VISION projections of indicated ccRCC signatures. **F.** Heatmap representation of the GSVA analysis showing specific hallmarks enriched in each cluster. **G.** Pseudo-bulk heatmap of 5 specific markers of epithelial, mesenchymal and inflamed-state cells.



Gene name	avg_log2FC	p_val_adj
MT1G	7,016514508	4,13E-57
MT1X	5,657031938	4,01E-66
MT1H	5,476122186	6,00E-43
MT2A	4,197884844	3,98E-53
MT1E	2,914682423	4,06E-29
MT1F	2,833891313	5,67E-14
PDK4	2,248573741	6,57E-08
MT1M	2,05500727	3,99E-11
TXNIP	1,777485529	0,010062801
RRAD	1,438216127	0,000779613
GADD45A	1,381949599	0,000106355
IFITM3	1,208161868	2,16E-07
SOD3	1,145358453	1
CLU	1,03398325	0,003508027
AQP3	1,009633978	0,000876572
FXYD2	0,924028142	0,020041392
CXCL14	0,920127933	0,125563004
ANGPTL4	0,918177446	0,000593193
FTL	0,859864841	0,003018355
S100A1	0,858494846	1
HLA-DRB5	0,842020278	4,06E-06
MGP	-0,879218434	1
HSPA1A	-0,938333608	0,266863082
MT-CYB	-0,953576553	4,45E-09
TM4SF1	-0,988258465	0,223365485
KLF5	-1,001829671	0,13833253
AQP1	-1,014888401	1
UGT2A3	-1,219443129	5,43E-08
MT-CO2	-1,248512591	2,13E-14
KRT19	-1,638823649	1,48E-05
RBP4	-1,961760622	1,41E-06
CTGF	-1,986536244	1,81E-05

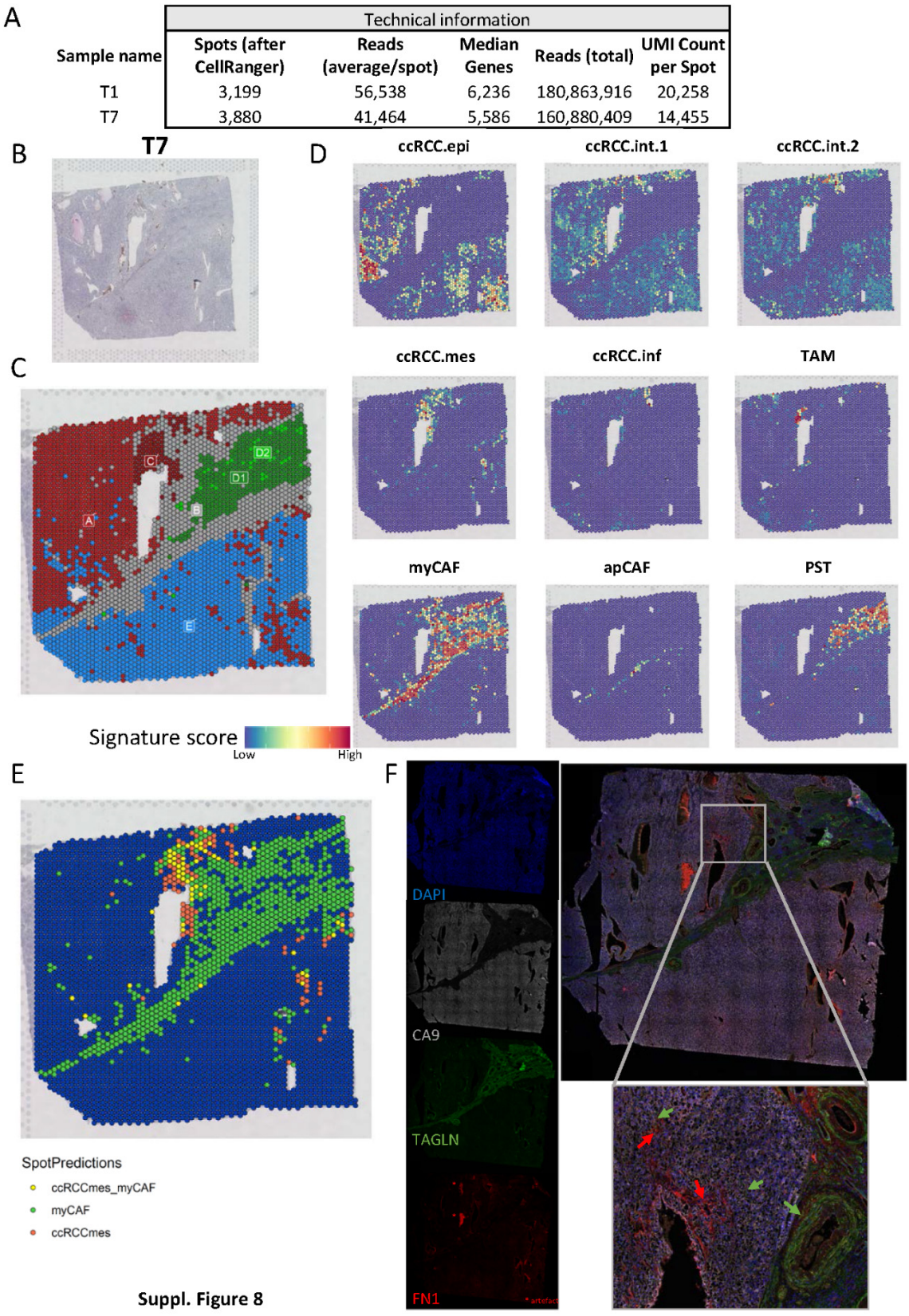
Suppl. Figure 6

Supplementary Figure 6. Description of PST clusters. A. Localisation of each identified cluster with respect to nephron structure. **B.** Heatmap of the 10-top markers for each tubule sub-cluster. **C.** Differentially expressed genes between PST1 (blue) and PST2 (red). **D.** Pseudo-bulk heatmap of S1/S2/S3 PST segment markers by Young et.al., and Lake et.al., to assess if different segments have been captured in our dataset. **E.** VISION projection of gluconeogenesis, glycolysis, oxidative phosphorylation and HIF1a signatures. **F.** Pseudo-bulk heatmap recapitulating expression of key genes for PST and the indicated ccRCC populations.

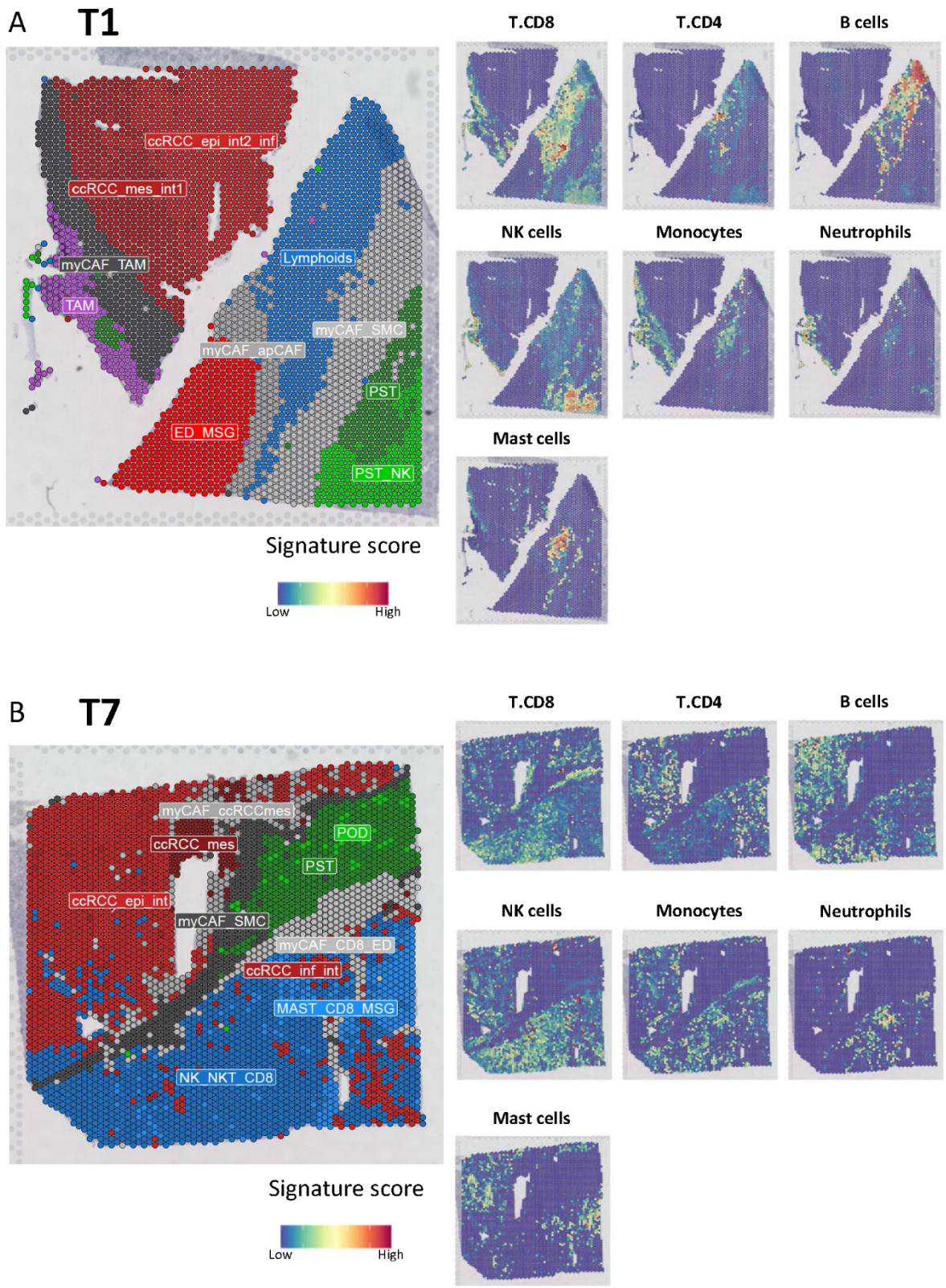


Suppl. Figure 7

Supplementary Figure 7. Analysis of cell populations in TCGA-KIRC cohort. A. Heatmap showing deconvolution results of all identified populations inferred by CIBERSORTx and displayed as row-scaled absolute scores on bulk RNA-seq data of 495 tumour samples. **B.** Reduced heatmap for deconvolution results of indicated populations displayed as column-scaled CIBERSORTx absolute scores. **C.** Sorted Spearman correlation coefficients between the indicated ccRCC populations and all other identified populations. **D.** Heatmap of each subpopulation average row-scaled absolute score with samples grouped according to the m1-m4 TCGA-KIRC classification.

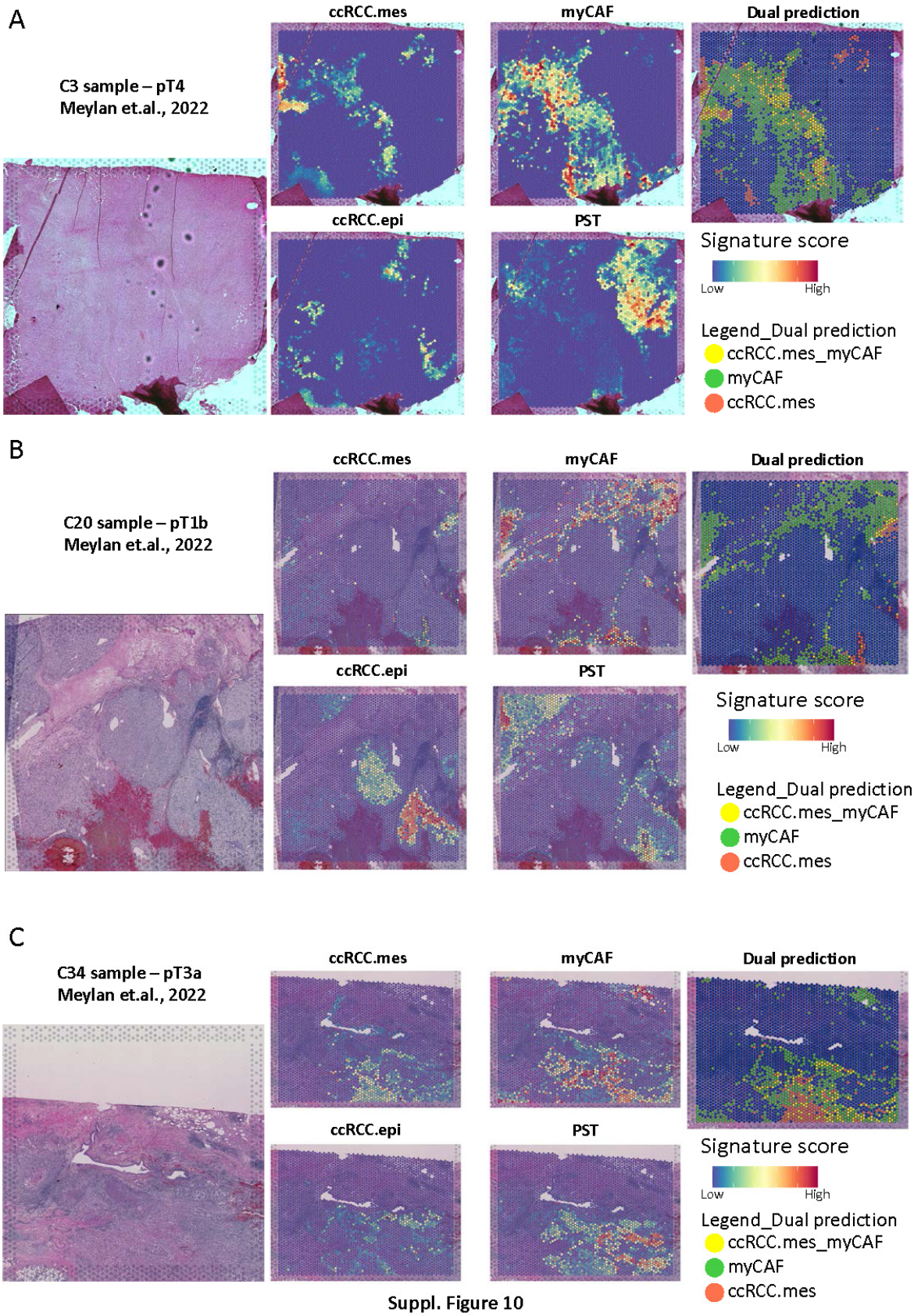


Supplementary Figure 8. Spatial localization of tumour cells and myCAFs. A. Sequencing statistics for T1 and T7 sections. **B**. Hematoxylin/Eosin staining of the T7 tumour section. **C.** Spatial plot showing spots clustered by regions (A-E). **D.** Spatial plots showing prediction scores in each spot for indicated cell signatures. **E.** Dual prediction using color-coded spots myCAF green, ccRCC/mes red, both cell types yellow.

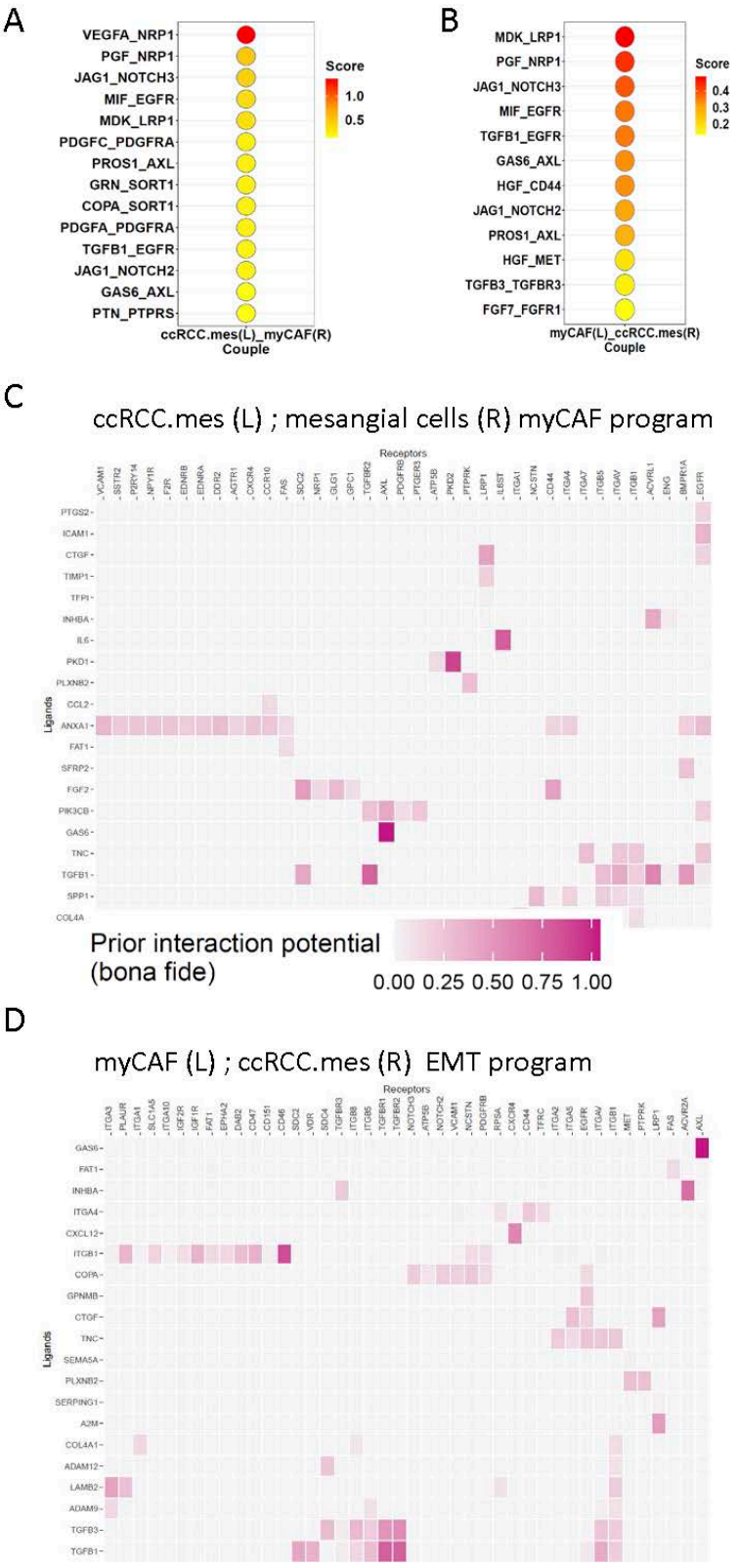


Suppl. Figure 9

Supplementary Figure 9. Spatial transcriptomic localization of lymphoid populations. A-B. Projection of the indicated lymphoid cell populations on T1 and T7 tumour sections as described above.

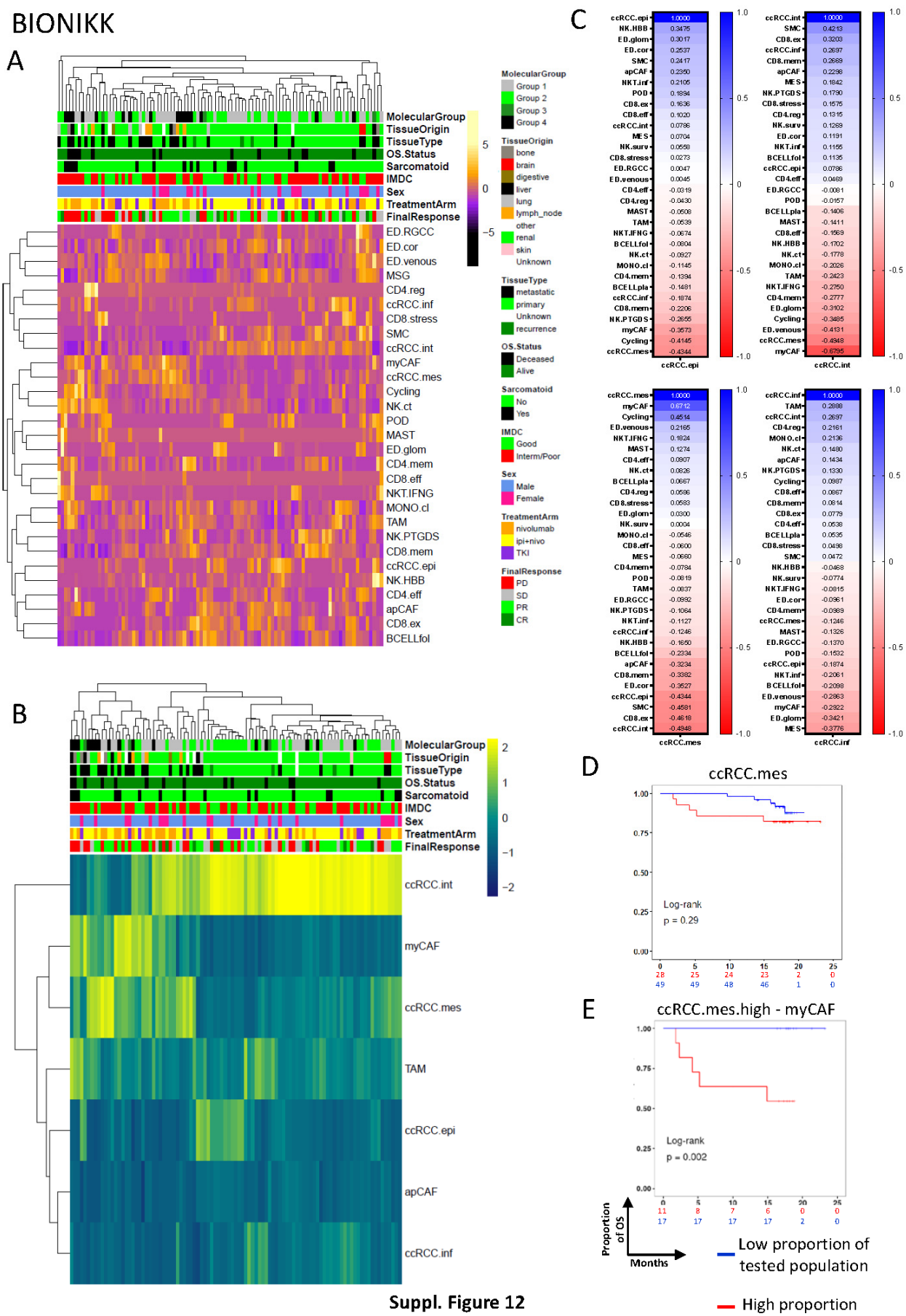


Supplementary Figure 10. Colocalization of ccRCC.mes and myCAF populations. A-C. Hematoxylin/Eosin staining of each indicated tumour section and projection of the indicated cell populations. Dual prediction using color-coded spots myCAF green, ccRCC/mes red, both cell types yellow.



Suppl. Figure 11

Supplementary Figure 11. NicheNet analyses of potential ccRCC.mes and myCAF interactions. A-B. CellPhoneDB predicted interactions with ligand on ccRCC.mes and receptor on myCAFs or vice versa, respectively. **C.** NicheNet predicted interactions inducing the myCAF program by ligands expressed in ccRCC.mes cells targeting receptors expressed by pericyte mesangial cells. **D.** NicheNet predicted interactions inducing the EMT program by ligands expressed in myCAF cells targeting receptors expressed by ccRCC.mes cells.



Suppl. Figure 12

Supplementary Figure 12. Deconvolution analysis of BIONIKK cohort. A. Heatmap with deconvolution results of all cell populations. **B.** Reduced heatmap for deconvolution results of the indicated populations displayed as column-scaled CIBERSORTx absolute scores. **C.** Sorted Spearman correlation coefficients between the indicated ccRCC populations and all other identified populations in scRNAseq. **D.** Kaplan-Meier curve for overall survival according to ccRCC.mes proportions. **E.** Kaplan-Meier curve for overall survival according to myCAF proportion in ccRCC.mes-high tumours.

G. Summary of single-cell profiled populations

Cluster name	Cell type	Cluster description
BCELLfol	B-cell	Follicular B-cells
PC	B-cell	Plasma B-cells
apCAF	Cancer-associated fibroblast	Inflammatory/antigen-presenting CAFs
myCAF	Cancer-associated fibroblast	Myofibroblastic CAFs
ccRCC.inf	ccRCC	Inflamed ccRCC
ccRCC.mes	ccRCC	Mesenchymal-like ccRCC
ccRCC.epi	ccRCC	Epithelial-like ccRCC
ccRCC.int	ccRCC	Intermediate ccRCC
CD4.eff	CD4 T-cell	Effector CD4 T-cells
CD4.mem	CD4 T-cell	Memory CD4 T-cells
CD8.ex	CD8 T-cell	Exhausted CD8 T-cells
CD8.mem	CD8 T-cell	Memory CD8 T-cells
CD8.eff	CD8 T-cell	Effector CD8 T-cells
CD8.stress	CD8 T-cell	Stressed CD8 T-cells
NKT.IFNG	CD8 T-cell natural killer like	IFNG producing NKT cells
NKT.stress	CD8 T-cell natural killer like	Stressed NKT cells
NKT.inf	CD8 T-cell natural killer like	Inflammatory NKT cells
Cycling	Cycling	Proliferating cells
ED.glom	Endothelial	Glomerular endothelial cells
POD	Endothelial	Podocytes
ED.venous	Endothelial	Venous endothelial cells
ED.RGCC	Endothelial	RGCC-expressing endothelial cells
ED.lymph	Endothelial	Lymphatic endothelial cells
ED.cor	Endothelial	Cortical endothelial cells
PST	Kidney tubule	Proximal straight tubules
mTAL	Kidney tubule	Medullary thick ascending limb cells
cTAL	Kidney tubule	Cortical thick ascending limb cells
DTL	Kidney tubule	Descending thin limb cells
ATL	Kidney tubule	Ascending thin limb cells
CNT	Kidney tubule	Connecting tubules
CD	Kidney tubule	Collecting ducts
MAST	Mast cell	Mast cells
MSG.inf	Mesangium/Pericyte	Inflamed mesangial cells
MSG	Mesangium/Pericyte	Mesangial cells
MONO.cl	Monocyte	Classical monocytes
MONO.at	Monocyte	Atypical monocytes
NK.HBB	Natural Killer	HBB-expressing natural killers
NK.PTGDS	Natural Killer	PTGDS-expressing natural killers
NK.surv	Natural Killer	Immune surveillance natural killers
NK.ct	Natural Killer	Cytotoxic natural killers
NEUT	Neutrophil	Neutrophils
SMC	Smooth-muscle like	Vascular smooth muscle like cells
CD4.reg	Treg	Tregs
TAM.CD1C	Tumor-associated macrophage	CD1C-expressing TAMs
TAM	Tumor-associated macrophage	Mainly M2-TAMs

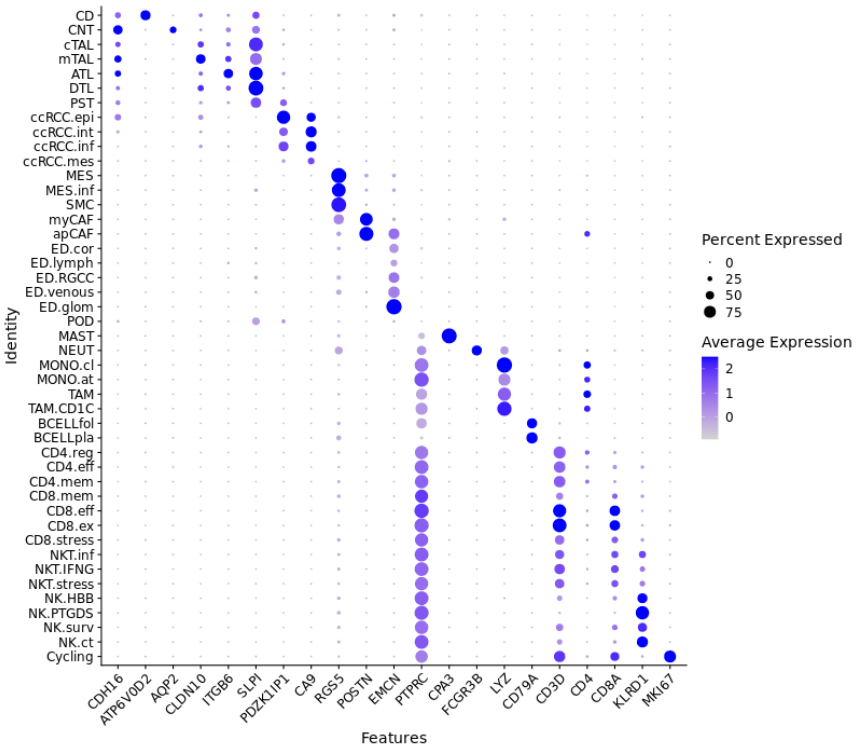


Figure 45. Summary of single-cell clusters.

Table of cluster annotation and bubble-chart of key marker genes expression.

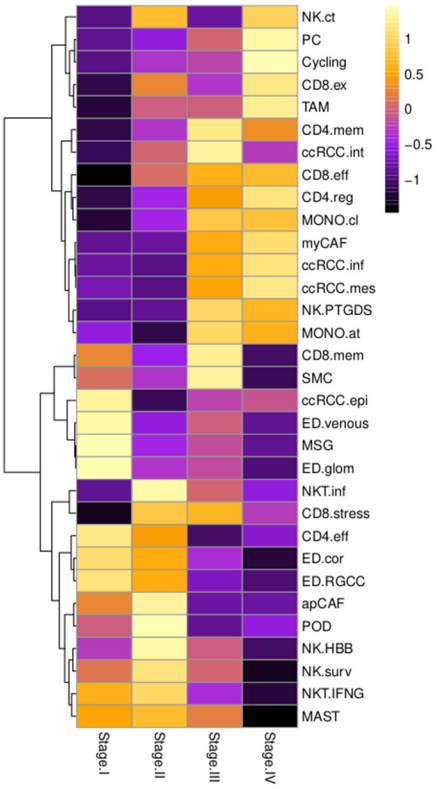


Figure 46. Heatmap of tumor composition grouped by disease stage.

Heatmap showing row-scaled CIBERSORTx absolute scores in TCGA-KIRC samples averaged by disease stage.

	Stage I	Stage II	Stage III	Stage IV	pval (Stage I vs all)	pval (Stage II vs all)	pval (Stage III vs all)	pval (Stage IV vs all)
PC	4.014	5.333	7.955	14.176	0.098	0.756	0.669	0.044
CD4.mem	23.221	28.226	37.552	32.342	0.007	0.917	0.012	0.448
NK.PTGDS	19.970	20.789	44.467	39.542	0.000	0.203	0.000	0.057
CD8.ex	9.981	14.481	12.606	17.412	0.035	0.480	0.872	0.036
NK.HBB	21.286	39.042	23.594	12.671	0.616	0.007	0.744	0.046
CD4.reg	13.786	19.363	26.687	32.169	0.001	0.831	0.070	0.010
NKT.inf	0.926	4.418	2.297	1.428	0.106	0.050	0.487	0.785
CD8.stress	5.701	11.463	11.003	8.525	0.017	0.231	0.083	0.851
NK.ct	84.652	92.794	85.094	93.805	0.187	0.269	0.515	0.101
CD8.eff	90.529	115.536	125.095	126.473	0.000	0.442	0.004	0.020
NKT.stress	0.150	0.000	0.000	0.000	0.178	0.654	0.451	0.572
CD8.mem	13.411	11.917	15.275	10.893	0.937	0.662	0.306	0.341
CD4.eff	46.063	41.191	31.198	33.845	0.002	0.808	0.009	0.175
NKT.IFNG	10.459	11.144	8.890	7.533	0.056	0.198	0.254	0.019
NK.surv	2.667	3.522	2.603	1.467	0.665	0.192	0.907	0.067
myCAF	44.593	47.181	106.142	128.228	0.000	0.151	0.003	0.000
apCAF	16.848	21.541	11.787	11.828	0.146	0.037	0.052	0.150
ccRCC.int	398.883	432.302	468.754	423.927	0.014	0.774	0.008	0.994
ccRCC.epi	422.130	378.133	394.730	396.728	0.266	0.470	0.616	0.755
ccRCC.inf	86.571	80.459	136.678	156.991	0.000	0.085	0.008	0.000
ccRCC.mes	131.004	126.496	165.982	185.556	0.006	0.217	0.085	0.007
Cycling	4.293	10.833	12.813	31.934	0.000	0.901	0.717	0.000
ED.cor	167.911	156.108	134.479	116.357	0.000	0.588	0.021	0.000
ED.lymph	0.547	0.000	0.018	0.004	0.242	0.686	0.527	0.616
ED.RGCC	38.941	32.464	19.147	16.264	0.000	0.690	0.013	0.018
ED.venous	285.958	240.429	252.954	232.661	0.000	0.085	0.170	0.004
MAST	3.702	3.945	3.342	1.463	0.094	0.343	0.882	0.001
MSG	312.988	242.614	252.685	225.164	0.000	0.042	0.017	0.000
SMC	105.498	98.447	130.183	82.862	0.675	0.470	0.002	0.012
MSG.inf	0.051	0.000	0.000	0.113	0.789	0.628	0.417	0.314
MONO.cl	2.039	3.550	6.045	5.981	0.001	0.850	0.016	0.081
MONO.at	139.055	136.284	146.169	144.432	0.449	0.554	0.356	0.661
NEUT	0.108	0.000	0.384	0.000	0.666	0.572	0.133	0.477
POD	3.579	6.374	2.043	2.531	0.585	0.019	0.098	0.443
ED.glom	46.991	24.203	26.736	15.991	0.000	0.064	0.019	0.000
TAM	136.016	160.151	159.801	185.778	0.001	0.571	0.364	0.002

Figure 47. Analysis of tumor composition grouped by disease stage.

Table of CIBERSORTx absolute scores in TCGA-KIRC samples averaged by disease stage followed by statistical analysis of enrichment in each stage.

2. Tumor heterogeneity and immune evasion

As the paper presented in the previous section was focused on the ccRCC.mes/myCAF interactions, in this section I performed follow-up analysis on single-cell data (summarized in section II.1.G, Figure 45) to better understand tumor composition changes during disease progression and crosstalk between ccRCC and the immune system.

A. Tumor composition analysis by stage

In the previous paper, I analyzed tumor composition by grade (Fig 4C), here I performed the same analysis with tumor stage (Figures 46-47). The results of these two analyses are very similar. Early-stage tumors show enrichment for endothelial and mesangial populations associated to vascularization: ED.RGCC ($p<0.001$), ED.venous ($p<0.001$), ED.cor ($p<0.001$) and MSG ($p<0.001$). On the other hand, late-stage tumors show depletion of these vascularization cell populations and present enrichment for undifferentiated ccRCC and myCAF populations: ccRCC.mes ($p=0.006$), ccRCC.inf ($p<0.001$) and myCAF ($p<0.001$). These tumors also show enrichment that can be progressive for certain phenotypes of immune cells such as exhausted/effector CD8 T-cells, Tregs, myeloid and plasma cells populations: CD8.eff ($p=0.02$), CD8.ex ($p=0.03$), CD4.reg ($p=0.009$), TAM ($p=0.001$) and PC ($p=0.04$). Of note, these late-stage tumors are enriched in the 'Cycling' population ($p<0.001$) which is mainly composed by proliferating CD8 T-cells.

B. Analysis of ligand-receptor interactions involving immune cells

a. Interaction of ccRCC with immune cells

As shown in the composition analysis, higher grade/stage tumors are enriched in immune and undifferentiated cancer cells indicating these might better tolerate the immune system. To explore this question, I performed ligand-receptor interaction analysis of mesenchymal cancer cells (ccRCC.mes) with a representative cytotoxic effector cell cluster (CD8.eff) using CellPhoneDB.

I first extracted ligands expressed by ccRCC.mes that can stimulate receptors of CD8.eff cells (Figure 48). This analysis revealed three interactions of interest that were previously described as immunosuppressive. In the first one, macrophage migration inhibitory factor (MIF) stimulates the CD74 receptor leading to anti-inflammatory M2 polarization of macrophages[249], this interaction fortuitously detected with CD8 T-cells applies more to macrophages that highly express CD74. The second interaction involves secreted phosphoprotein 1 (SPP1) that stimulates the CD44 receptor leading to suppression of T-cell proliferation[250]. The third interaction involves stimulation of the immune checkpoint PD-1 by the interleukin-like EMT inducer (ILEI/*FAM3C*) ligand leading to suppression of cytotoxic effector functions. Then, I checked the expression of ccRCC.mes CD8-interacting ligands in all ccRCC clusters (Figure

49). This analysis revealed MIF and SPP1 to be more expressed in epithelial and intermediate ccRCC while FAM3C is a cluster marker enriched in ccRCC.mes (log2FC=1.3, p<0.001).

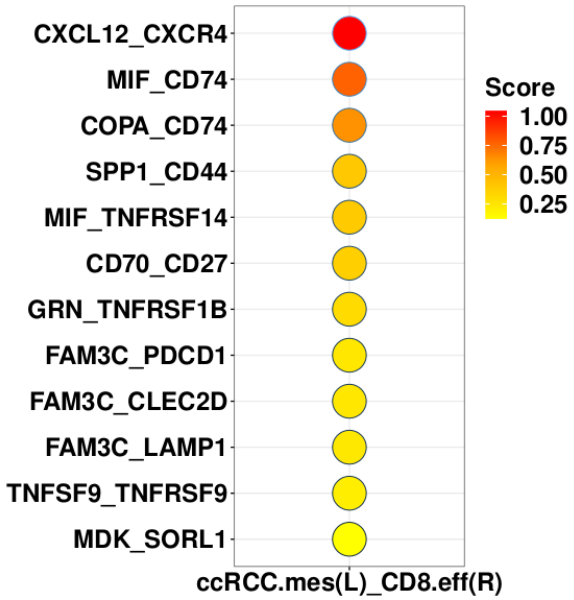


Figure 48. Ligand (ccRCC.mes) – Receptor (CD8.eff) pairs detected by CellPhoneDB.

Bubble-chart of significant (pval<0.05) ligand-receptor interactions colored by interaction strength.

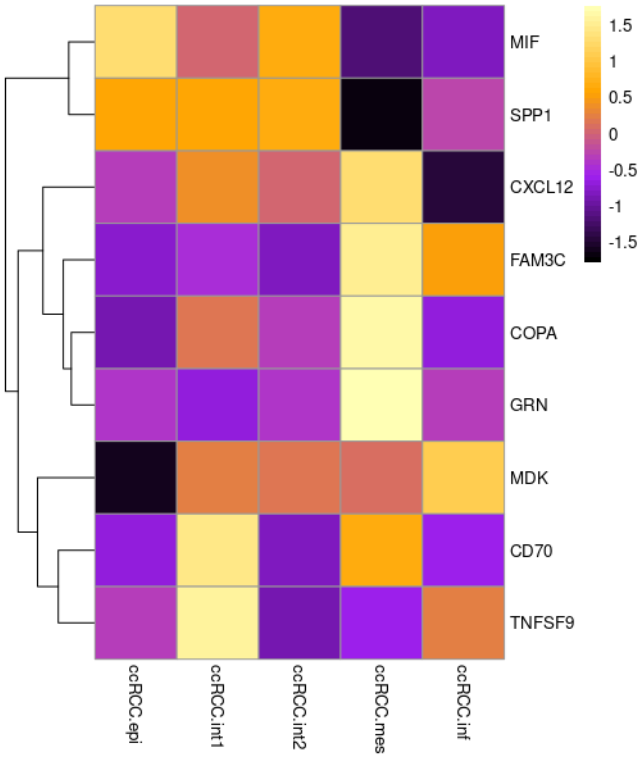


Figure 49. Expression of ccRCC.mes ligands across ccRCC clusters.

Heatmap of row-scaled expression of CD8-stimulating ligand genes in the different ccRCC clusters.

b. Interaction of myCAFs with CD8 T-cells

As myCAFs were found to be associated with ICI resistance as well as enriched in higher grade/stage tumors, I also performed CellPhoneDB analysis against CD8 T-cells (Figure 50). Ligands expressed by myCAFs that could stimulate CD8.eff receptors were mostly similar to the ones from ccRCC.mes. Commonly found interactions include the suppressive FAM3C-PDCD1 couple and the CXCL12 stimulation of C-C motif chemokine receptor 4 (*CCR4*) that leads to chemoattraction of T-cells[251]. Some interactions were more specific to myCAFs but their downstream effect in T-cells is unknown.

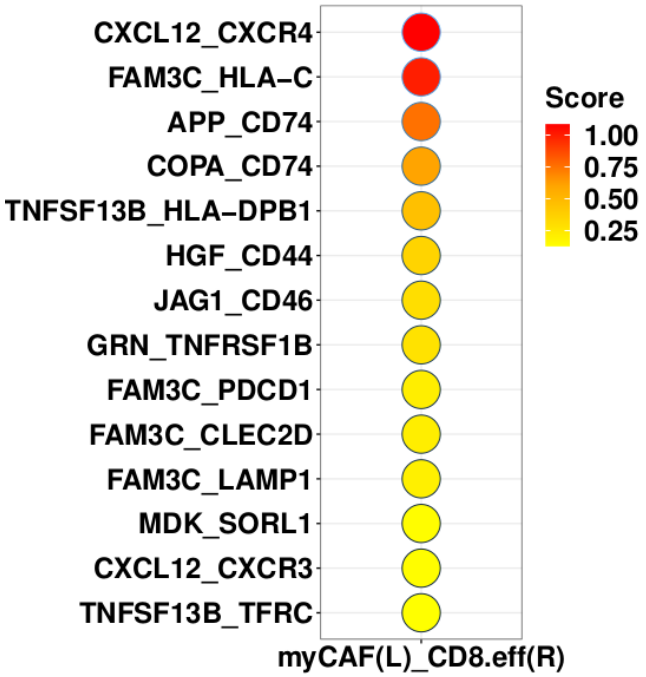


Figure 50. Ligand (myCAF) – Receptor (CD8.eff) pairs detected by CellPhoneDB.

Bubble-chart of significant (pval<0.05) ligand-receptor interactions colored by interaction strength.

c. Expression of PD-L1

The preceding ligand-receptor analyses suggest that ccRCC.mes can autonomously suppress cytotoxicity by stimulating PD-1 through FAM3C, a property that is shared with myCAFs. Another ligand of PD-1 of interest is PD-L1. This very well-known ligand is an ICI target but it is poorly captured in single-cell data. To explore potential expression of PD-L1 in ccRCC.mes and myCAF, I performed a correlation analysis between the expression level of PD-L1 and the proportions of the different scRNA-seq cell populations estimated by CIBERSORTx in the BIONIKK dataset. This analysis revealed significant correlations between proportions of ccRCC.mes (R=0.2, p=0.04) and myCAF (R=0.32, p=0.001) with expression of PD-L1 (Figure 51) suggesting they could be expressing this ligand.

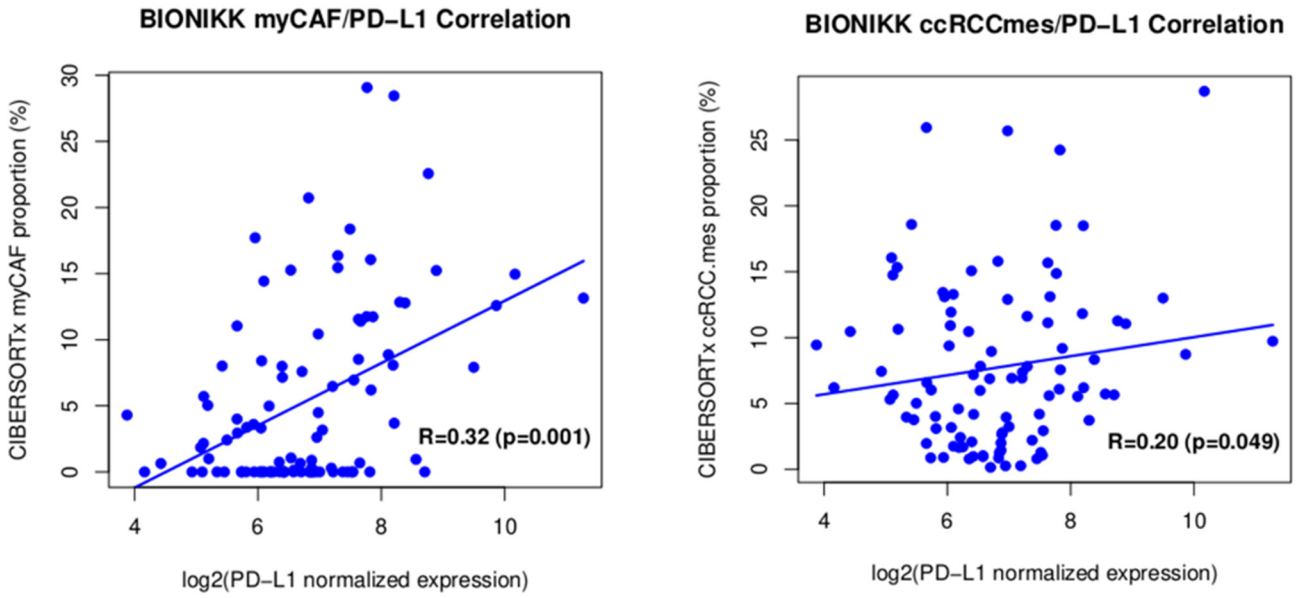


Figure 51. Correlation of myCAF/ccRCC with PD-L1 expression level.

Scatterplot showing PD-L1 expression level (log-normalized RNA-seq counts) and CIBERSORTx-inferred proportions of myCAF or ccRCC.mes (%), Pearson's product-moment correlation and significance are indicated in the bottom right corner.

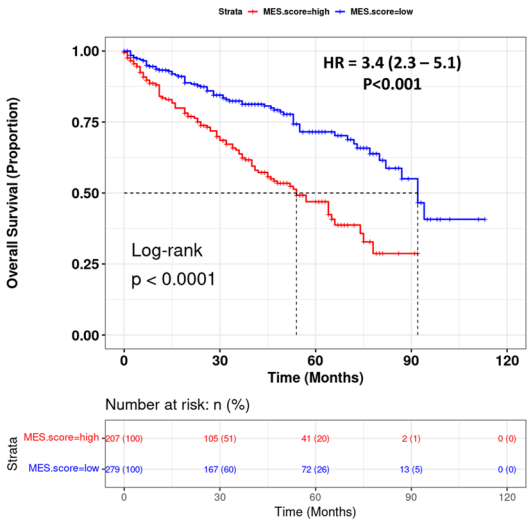


Figure 52. TCGA-KIRC survival according to MES.score stratification.

Kaplan-Meier survival curve of TCGA-KIRC patients stratified by MES.score (optimal cut-point), un-stratified hazard-ratio from univariate COX-regression analysis with 95% confidence interval and p-value are indicated.

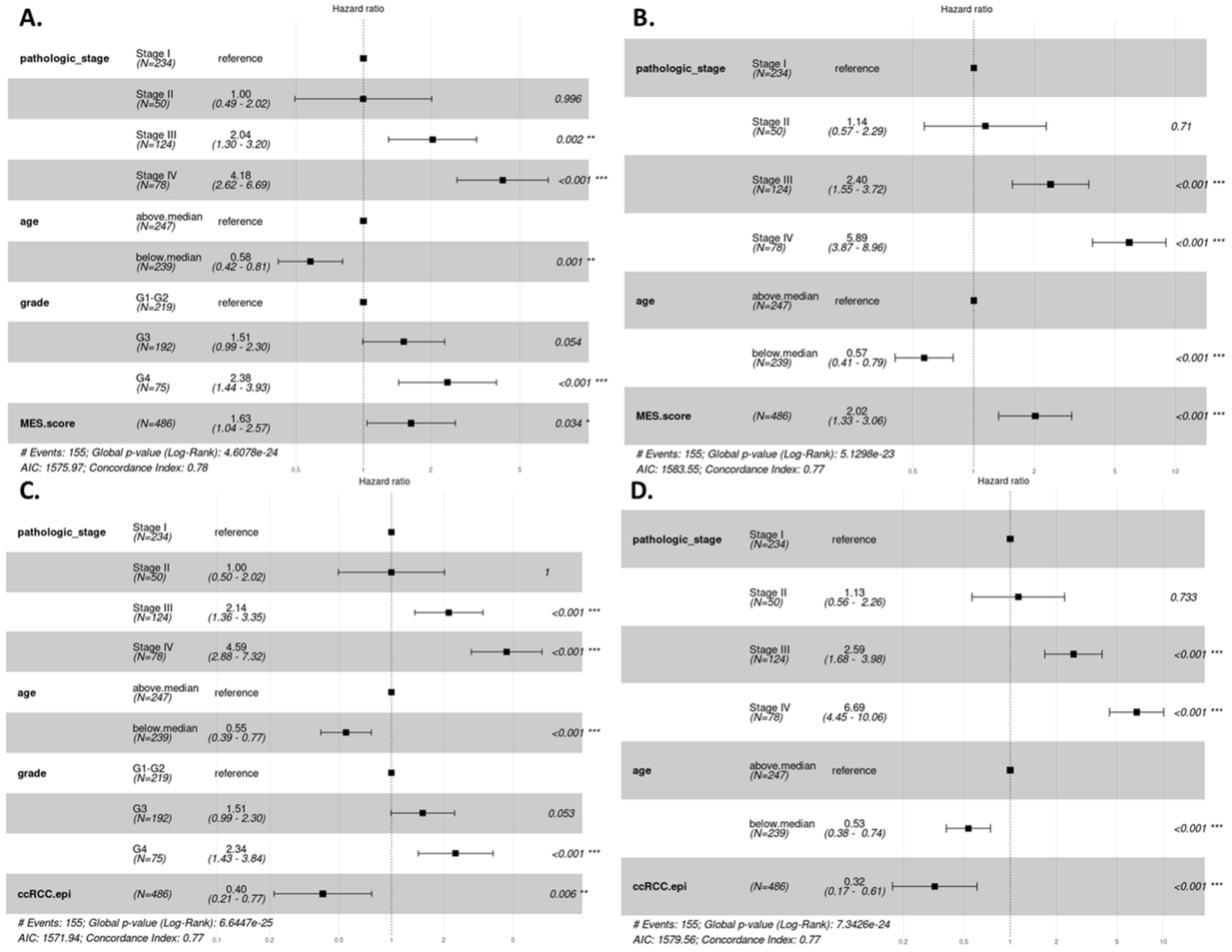


Figure 53. MES.score and ccRCC.epi in multiple COX-regression analyses.

Forest-plots of COX multiple-regression analyses of MES.score with (**A**) or without (**B**) including tumor grade and ccRCC.epi score with (**C**) or without (**D**) including tumor grade.

3. Biomarkers of survival and resistance to immunotherapy

Previous analyses identified relationships between different cell populations and patient survival or resistance to treatment. In this section, I will try to derive useful biomarkers based on these populations.

A. Epithelial/mesenchymal populations as markers of survival

In the paper presented in section 1, I established a relationship between ccRCC.mes, myCAF and ccRCC.inf and poor survival while ccRCC.epi was associated with a better survival. The epithelial component of the tumor is encompassed under one value but the undifferentiated components are scattered across 3 values. Thus, I combined the 3 mesenchymal scores with bad prognosis into one value (MES.score) and explored its overall survival prognostic capacity in the TCGA-KIRC. By stratifying patients into high/low MES.score (Figure 52), the Kaplan-Meier curve reveals that high MES.score is significantly associated with worse survival and this result is confirmed by univariate COX-regression analysis using the continuous value (HR=3.4, $p<0.001$). When challenged in a multiple-regression analysis, the MES.score remains significant (HR=1.63, p=0.034) but collinearity can be observed with tumor grade as G3 (HR=1.51, p=0.054) appears non-significant (Figure 53A). This collinearity is not unexpected as both tumor grade and MES.score are measures of dedifferentiation of the tumor, and indeed removing tumor grade from the model (Figure 53B) affects MES.score significance (HR=2.02, $p<0.001$). The same effect is observed with the ccRCC.epi score (HR=0.4, p=0.006) which is also affected by removal of tumor grade (HR=0.32, $p<0.001$) but to a lesser degree (Figure 53C-D). These analyses indicate that finding marker genes reflecting mesenchymal and epithelial cancer populations can help predict patient survival but this might not be required as routine analyses by pathologists might already provide enough information in a simpler way.

B. Signature genes of myCAFs as biomarker for ICI resistance

In the paper from section 1, we also saw that the myCAF score is significantly correlated with poor survival in patients treated with ICIs in the BIONIKK cohort. I sought to identify a small myCAF gene signature that could be measured by reverse transcription quantitative polymerase chain reaction (RTqPCR) and reflect myCAF tumor infiltration. To do this, I first performed a differential expression analysis between the myCAF-high and myCAF-low BIONIKK tumor groups previously identified (Section 1C.h, Figure 7C-D) in order to identify genes that stand out at the bulk-RNA level. This analysis revealed 477 genes enriched in myCAF-high ($log2FC<-1$, $adj\text{-}pval<0.05$) with ontologies associated with ECM remodeling and 997 genes enriched in myCAF-low ($log2FC>1$, $adj\text{-}val<0.05$) with ontologies associated with kidney transporters and metabolism (Figure 54). Then, I crossed the 477 myCAF-high differentially expressed genes with the 306 myCAF marker genes identified in the scRNA-seq data revealing 45 common genes between the two lists.

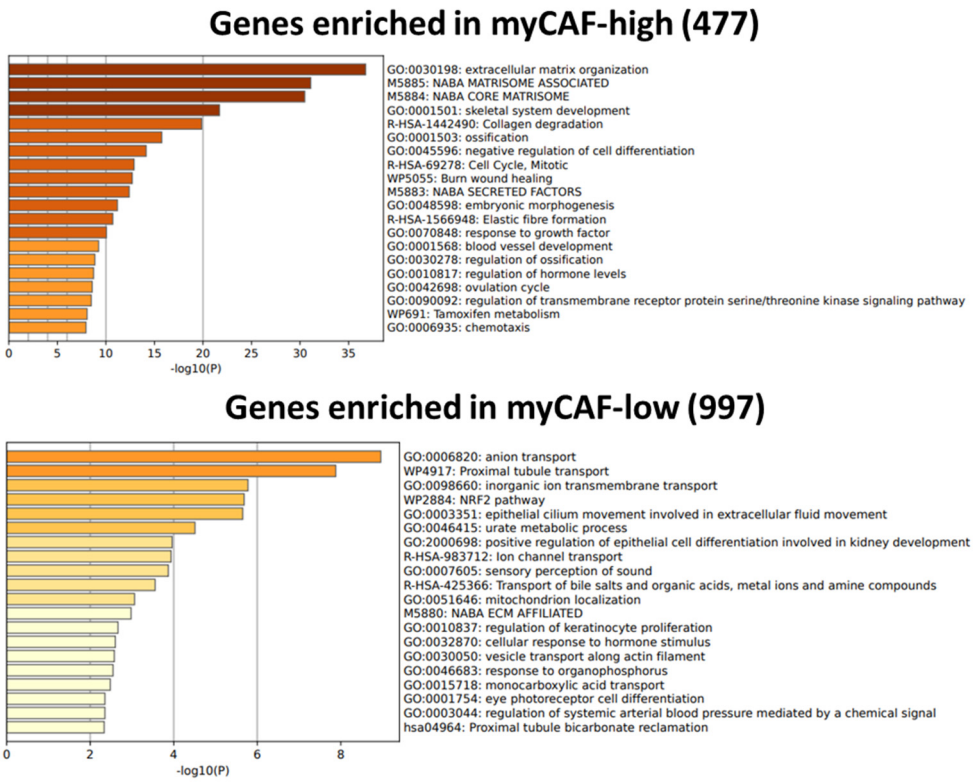
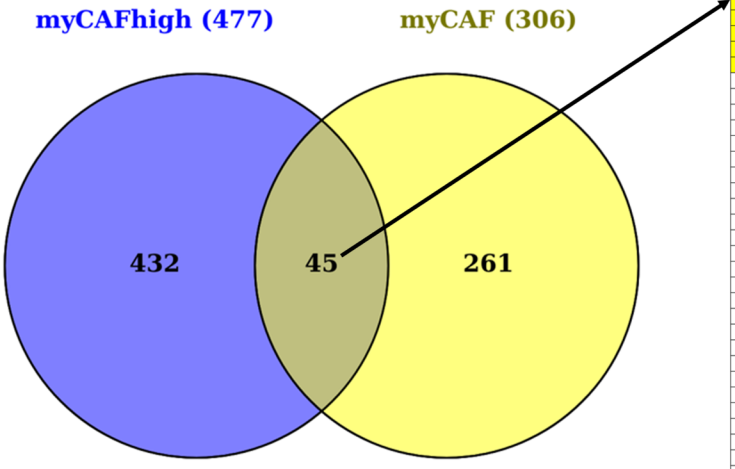


Figure 54. Ontology of genes associated with myCAF-high and myCAF-low tumors.

Metascape gene ontology analysis of significant differentially expressed genes identified by DESeq2 in myCAF-low versus myCAF-high tumors.



Gene name	log2FC
CTSK	-2.73
COL12A1	-2.33
ASPN	-2.23
POSTN	-2.04
FAP	-1.90
COL3A1	-1.83
SERPINF1	-1.80
PDGFRA	-1.79
LUM	-1.76
INHBA	-1.69
CLEC11A	-1.67
DDR2	-1.65
MMP11	-1.53
COL1A1	-1.53
ADAMTS2	-1.50
COL5A2	-1.50
COL5A1	-1.44
EMILIN1	-1.44
PRRX1	-1.41
ITGBL1	-1.39
COL6A3	-1.38
MICAL2	-1.38
COL1A2	-1.38
FBN1	-1.37
CYBRD1	-1.32
COL16A1	-1.31
SULF1	-1.29
ANTXR1	-1.27
GPX8	-1.24
COL14A1	-1.23
DKK3	-1.20
ISLR	-1.20
THBS2	-1.18
PCOLCE	-1.16
CCDC80	-1.16
EFEMP1	-1.15
CTHRC1	-1.10
PALLD	-1.08
EPHA3	-1.08
THY1	-1.07
CYP1B1	-1.03
DCN	-1.02
MRC2	-1.02
LAMA2	-1.01
RAB31	-1.01

Figure 55. Intersection of myCAF marker genes with myCAF-high tumors signature.

Venn diagram showing intersection and table of common genes sorted by fold-changes in descending order, top 5 genes are indicated in orange and top 10 in yellow.

I used the leading-edge genes of the 45 common genes to build a 5 (LE5) or 10 (LE10) genes signature representing myCAF infiltration (Figure 55) and tested their relationship with overall survival of patients treated with ICIs in the BIONIKK cohort. I computed the geometric mean expression of *CTSK*, *COL12A1*, *ASPN*, *POSTN* and *FAP* for the LE5 signature, a high expression level of this signature was significantly associated with worse survival (HR=6.87, p=0.003) by stratification (Figure 56A). I then computed the mean expression of *CTSK*, *COL12A1*, *ASPN*, *POSTN*, *FAP*, *COL3A1*, *SERPINF1*, *PDGFRA*, *LUM* and *INHBA* for the LE10 signature, a high level of this signature was also associated with worse survival (HR= 5.01, p=0.01) by stratification (Figure 56B). These two signatures could have practical applications to select patients for treatment by ICIs.

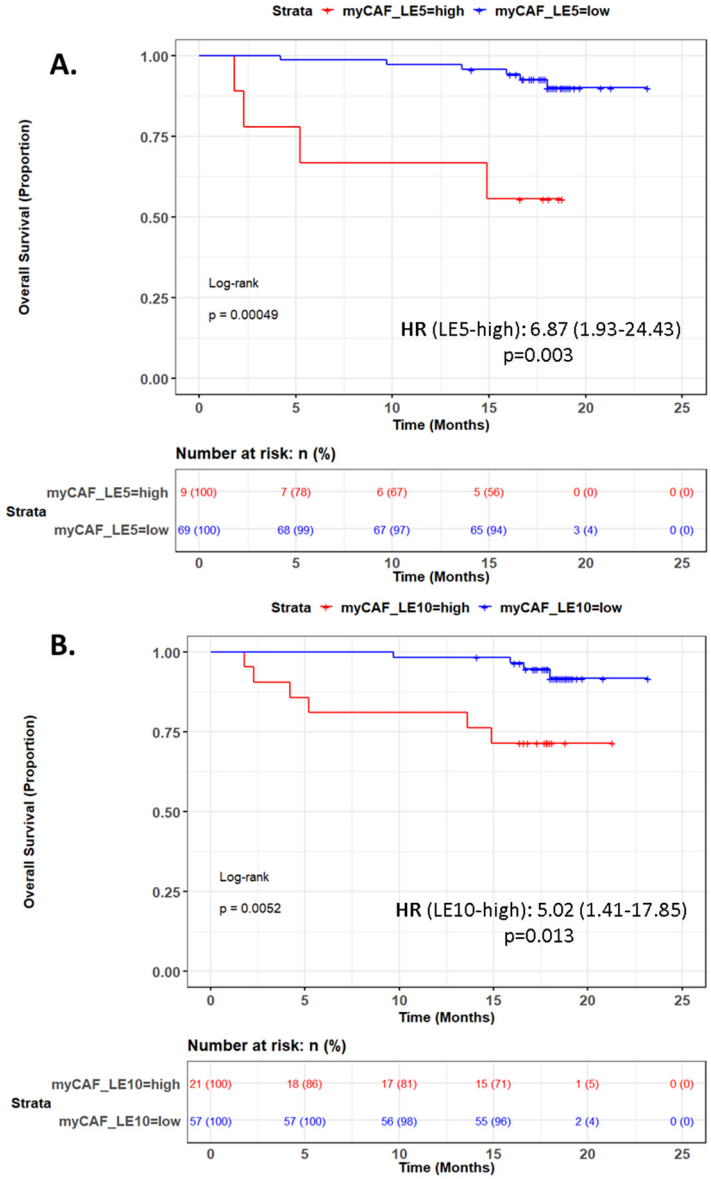


Figure 56. Survival of ICI-treated patients after stratification by myCAF signatures.

Kaplan-meier survival curves of ICI-treated BIONIKK patients stratified by LE5 (**A**) or LE10 (**B**) gene signatures (optimal cut-point), hazard-ratios from univariate COX-regression analysis with 95% confidence interval and p-values are indicated.

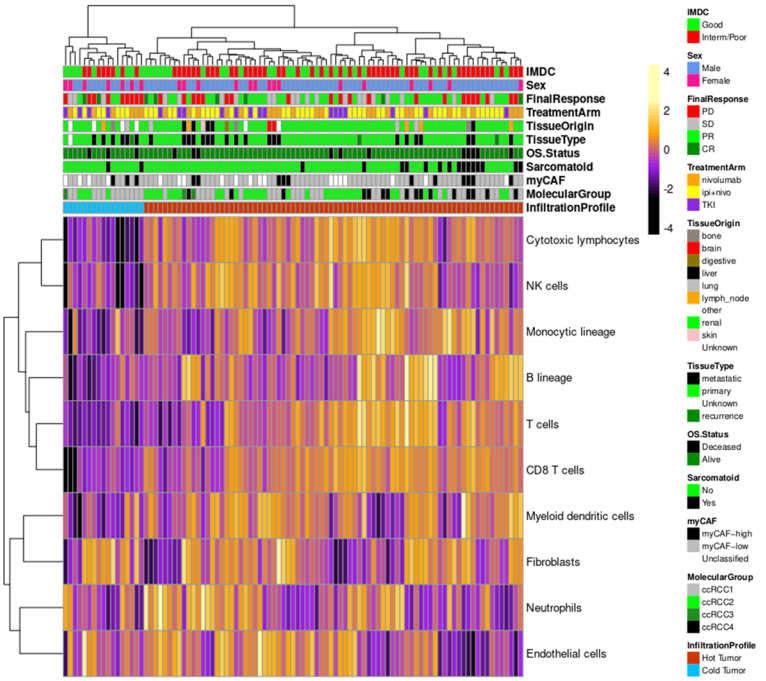


Figure 57. Clustering of the BIONIKK cohort based on stromal cells deconvolution.

Heatmap of row-scaled MCP-counter scores with hierarchical clustering (ward-D2 linkage).

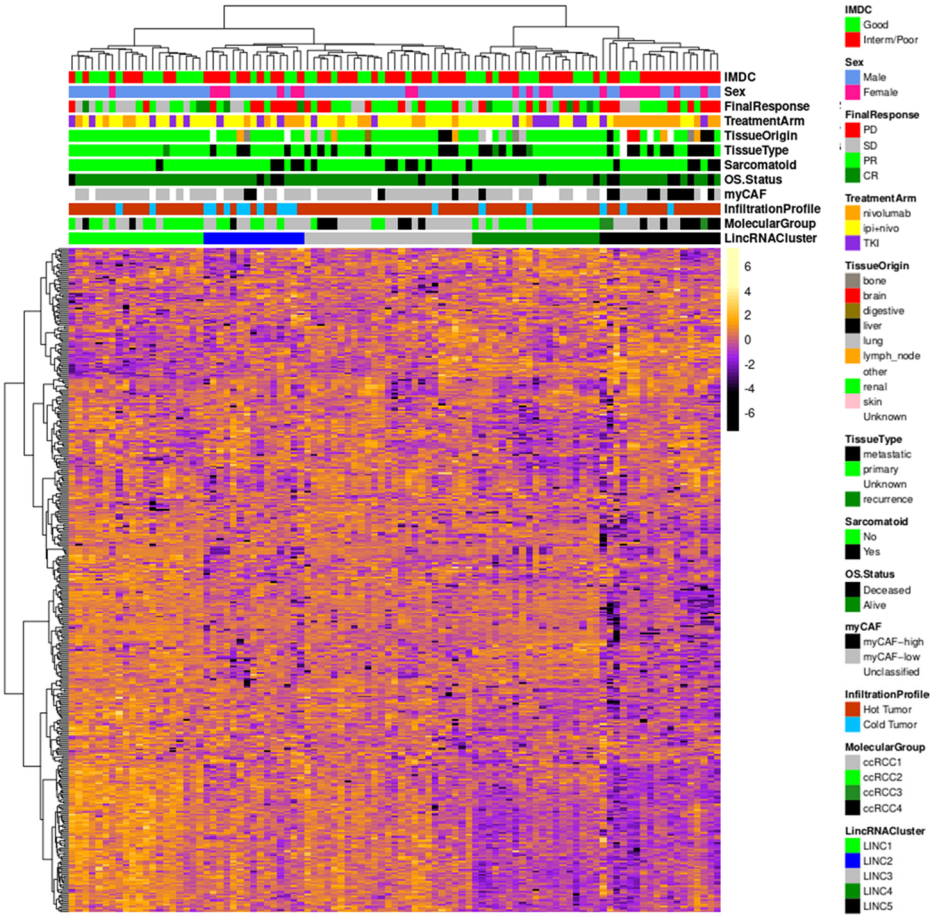


Figure 58. Clustering of the BIONIKK cohort based on the 400 most variable lincRNAs.

Heatmap of row-scaled lincRNA expression with hierarchical clustering (ward-D2)

C. Use of lncRNAs as biomarkers

We saw in the introduction that lncRNAs can have key roles in oncogenesis and possibly present a high degree of specificity. In this part, I focused on the analysis of lincRNAs in the BIONIKK cohort to find potential new biomarkers.

a. Unsupervised clustering of the BIONIKK cohort based on lincRNAs

i. Clustering based on immune-infiltration

In order to assess if a classification based on lncRNAs can be representative of biological properties of the tumors, I gathered tumor characteristics by labelling samples using molecular groups established in the BIONIKK clinical trial to assign patients for treatment (ccRCC1-4, described in the introduction section 2C.c.iv), clinical annotations such as presence of sarcomatoid cells or biopsy/resection origin, and annotations of myCAF-high/low patients established in the paper from section 1. To obtain better information about immune infiltration, I performed deconvolution by MCP-counter followed by clustering (Figure 57). This analysis revealed a group of immune-cold tumors with reduced infiltration of T cells ($p<0.001$), CD8 T cells ($p<0.001$), B lineage ($p=0.01$), NK cells ($p<0.001$), monocytic lineage ($p<0.001$) and myeloid dendritic cells ($p=0.005$) but similar infiltration in fibroblast ($p=0.45$) and endothelial cells ($p=0.78$) according to MCP-counter scores. The hot tumors showed heterogeneous infiltration patterns that could be further sub-clustered. I labelled these two groups “cold” and “hot” tumors.

ii. Clustering based on lincRNAs

Then, I performed unsupervised clustering based on the top 400 most variable lincRNAs as they do not overlap coding genes and thus can be unambiguously quantified (Figure 58). Hierarchical clustering revealed a possible classification into 5 clusters (LINC1-5), I performed differential expression analysis of each cluster against all others in order to identify their characteristics.

LINC1 is mainly composed of ccRCC2 tumors (12/20 LINC1 samples) and DESeq2 detected 486 protein coding marker genes ($log2FC>1$, $adj\text{-}val<0.05$). GSEA analysis did not reveal significantly up-regulated pathways as expression profile was very close to that of LINC4 but ontology analysis of marker genes indicated their involvement in epithelial cilium assembly and ion transporters (Figure 59) in line with the more differentiated nature of ccRCC2 tumors noted in the paper from section 1 (see 1C.h).

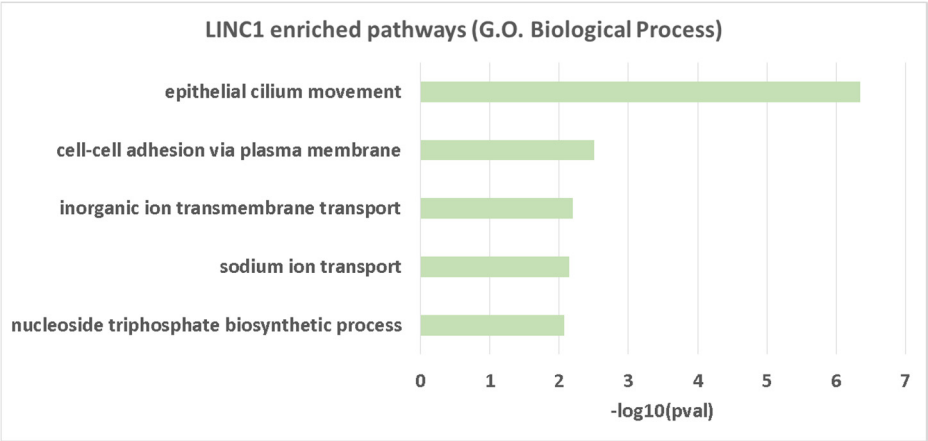


Figure 59. LINC1 enriched pathways.

Chart of significantly enriched ($p<0.05$) non-redundant biological processes after DAVID gene ontology of LINC1 marker genes.

LINC2 is mainly composed of ccRCC1 tumors (9/15 LINC2 samples) and presented 63 protein coding marker genes (log2FC>1, adj-val<0.05). GSEA analysis revealed a gene expression program associated with EMT, hypoxia and TNF-α signaling (Figure 60). Of note, this cluster was significantly enriched in immune-cold tumors (9/15 LINC2 samples) and in fact half of the immune-cold tumors clustered in LINC2 (9/18 cold tumors, $p<0.001$).

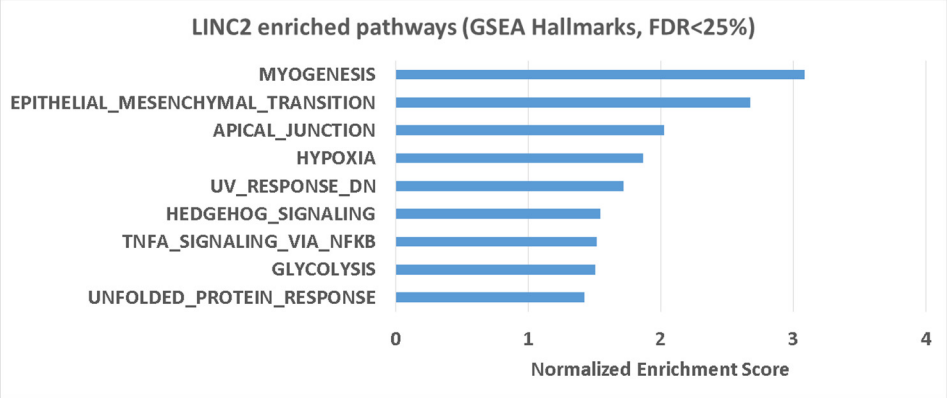


Figure 60. LINC2 enriched pathways.

Chart of enriched human hallmarks following GSEA analysis of LINC2 cluster against all other samples.

LINC3 is composed mainly of ccRCC1 tumors (16/25 LINC3 samples) but also shows presence of about half of the ccRCC4 tumors (6/13 ccRCC4 tumors), it is characterized by 42 protein coding marker genes (log2FC>1, adj-val<0.05). GSEA analysis revealed a gene expression program mainly associated with complement and inflammatory response (Figure 61).

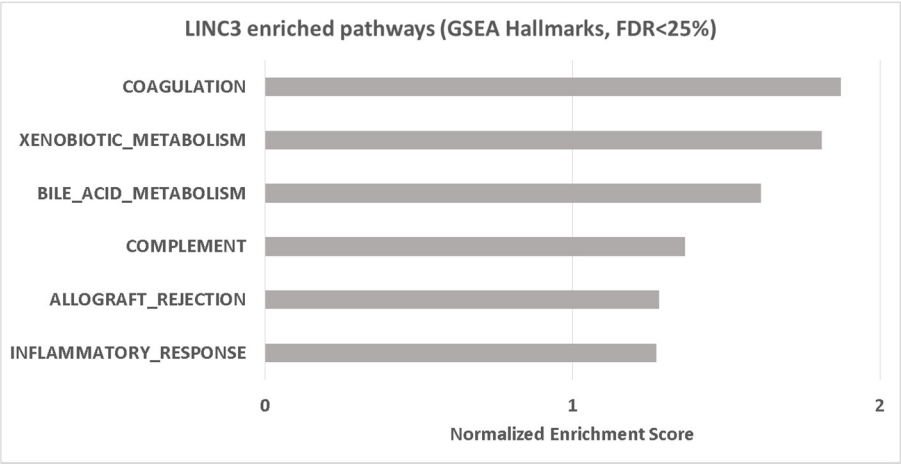


Figure 61. LINC3 enriched pathways.

Chart of enriched human hallmarks following GSEA analysis of LINC3 cluster against all other samples.

LINC4 is composed mainly of ccRCC2 tumors (16/19 LINC4 samples) and displayed 176 protein coding marker genes (log2FC>1, adj-val<0.05). GSEA analysis revealed a gene expression program associated with oxidative-phosphorylation, heme and fatty acid metabolism. This cluster is enriched in a kidney epithelial program similar to LINC1.

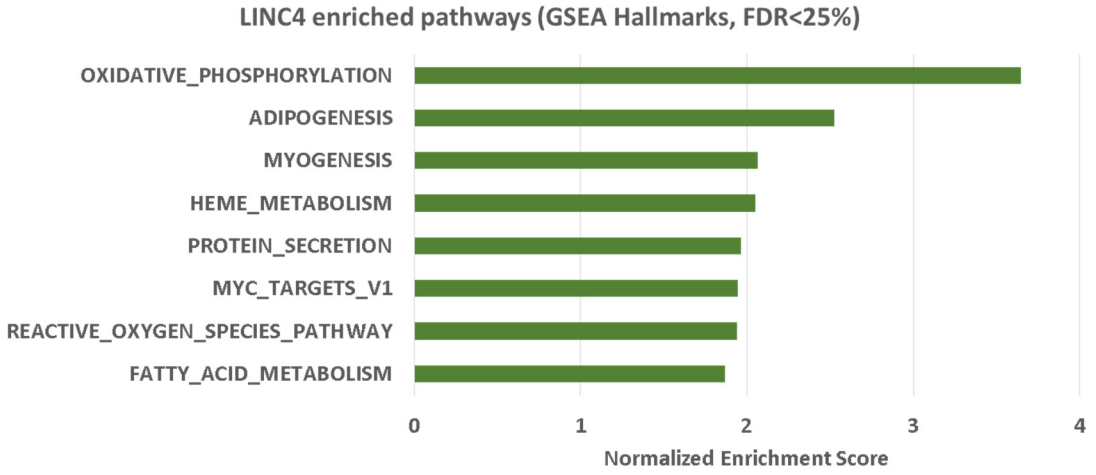


Figure 62. LINC4 enriched pathways.

Chart of enriched human hallmarks following GSEA analysis of LINC4 cluster against all other samples.

LINC5 is composed mainly of both ccRCC1 (8/18 LINC5 samples) and ccRCC4 tumors (7/18 samples) and presented 901 protein coding marker genes (log2FC>1, adj-val<0.05). This cluster is significantly enriched in myCAF-high (9/14 myCAF-high samples, p<0.001), ccRCC4 (7/15 ccRCC4 samples, p=0.01) and distant metastatic tissue samples

(10/26 metastasis, p=0.006). GSEA analysis revealed a gene expression program associated with EMT, cell cycle and inflammation (Figure 63).

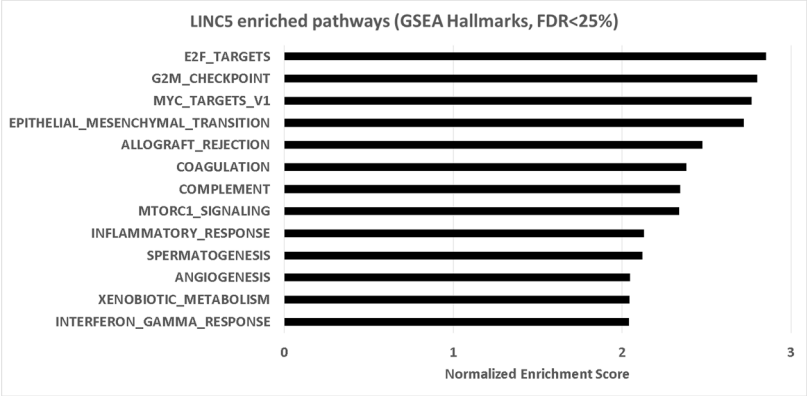


Figure 63. LINC5 enriched pathways.

Chart of enriched human hallmarks following GSEA analysis of LINC5 cluster against all other samples.

To explore the relationship between immunotherapy response and LINC clusters, I analyzed the survival of patients treated by ICIs in each cluster by plotting the Kaplan-Meier curve (Figure 64). This analysis showed that LINC5, the mesenchymal cluster enriched in myCAF-high tumors, was significantly associated with a worse overall survival (p=0.011) showing that lincRNA expression alone can discriminate tumor heterogeneity and TME composition with enough accuracy to be used for prognosis.

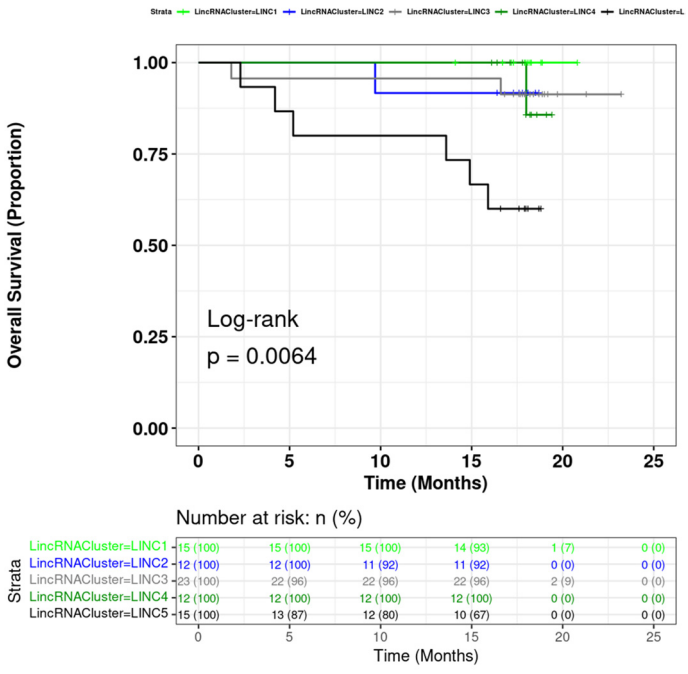


Figure 64. Survival of ICI treated patients according to the lincRNA classification.

Kaplan-meier survival curves of ICI-treated BIONIKK patients grouped by LINC clusters.

iii. Identification of candidate biomarkers

Next, I sought to identify interesting biomarkers by checking gene expression patterns of differentially expressed non-coding genes from the poor prognosis LINC5 cluster in single-cell data. Out of 309 of these genes, 39 were captured in single-cell data and, in line with the LINC5 profile, they were mainly expressed in immune and mesenchymal cells (Figure 65). None of these genes were specific for myCAFs, but lncRNA *SFTA1P* stood out as enriched in all mesenchymal cells. Survival analysis of patients treated with ICIs after stratification by *SFTA1P* expression showed a significantly worse survival in the high-expression group (HR= 5.5, p=0.009) indicating that this lncRNA alone could represent myCAF infiltration well-enough to serve in the clinical setting (Figure 66). Two other genes stood out by showing ccRCC specificity. The first one, *RP5-1172A22.1* (LINC01638), is activated in ccRCC.int and expression is conserved across ccRCC.inf and ccRCC.mes so it could be involved in early dedifferentiation of ccRCC cells. The second one, *RP11-417E7.1* (LINC01615), has highly specific expression in ccRCC.mes which prompted me to study it in more details.

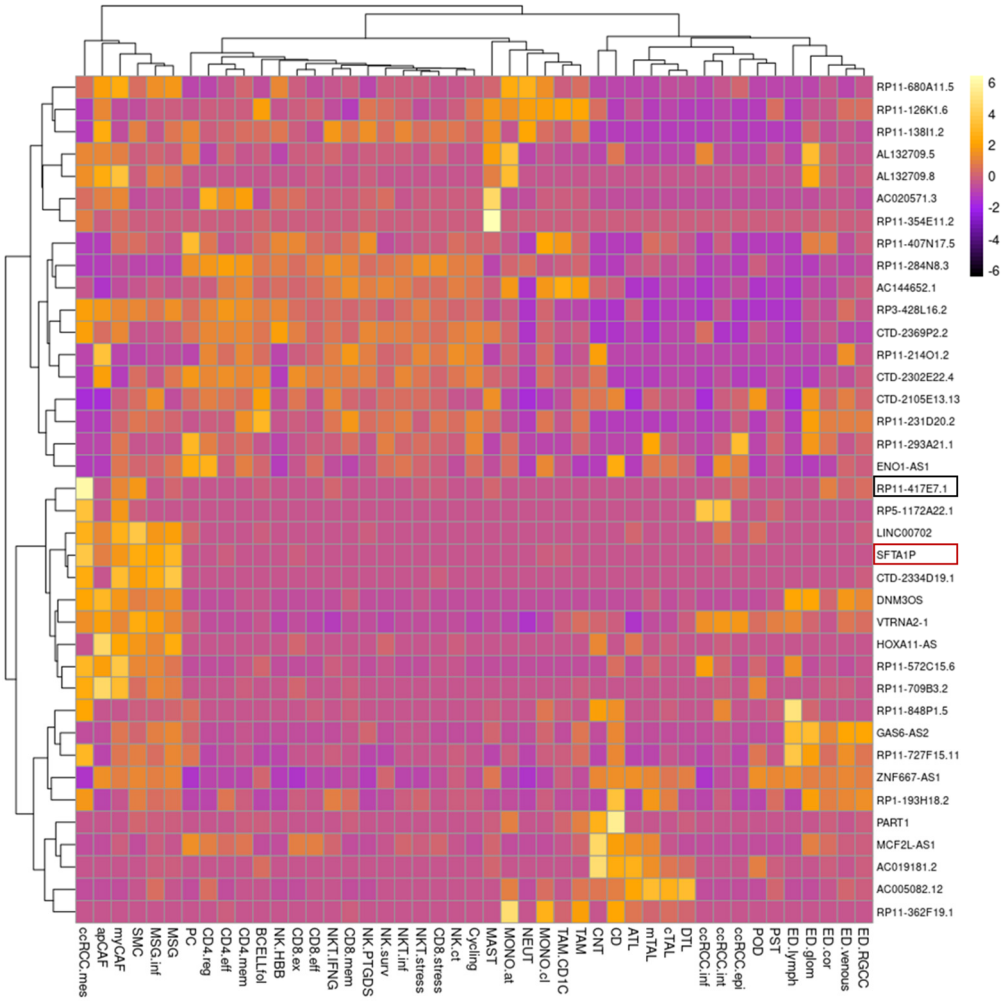


Figure 65. Expression of non-coding LINC5 markers in single-cell clusters.

Heatmap of row-scaled expression for 39 non-coding LINC5 marker genes in single-cell data. Two interesting potential biomarkers are highlighted: *SFTA1P* (red) and *LINC01615* (black).

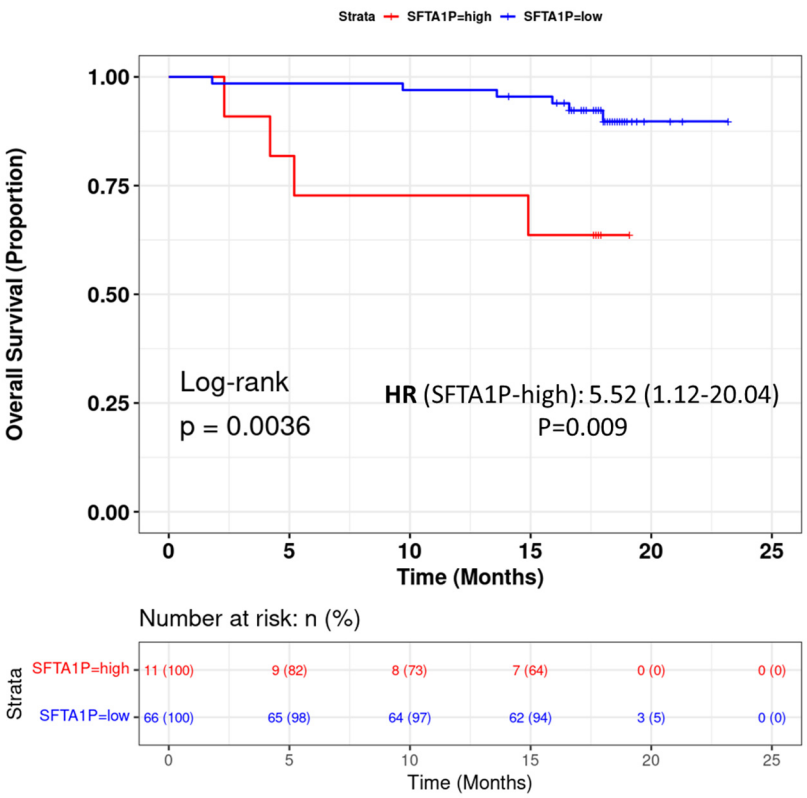


Figure 66. Survival of ICI treated patients after stratification by *SFTA1P* expression.

Kaplan-meier survival curves of ICI-treated BIONIKK patients stratified by *SFTA1P* expression (optimal cut-point), hazard-ratio from univariate COX-regression analysis with 95% confidence interval and p-value are indicated.

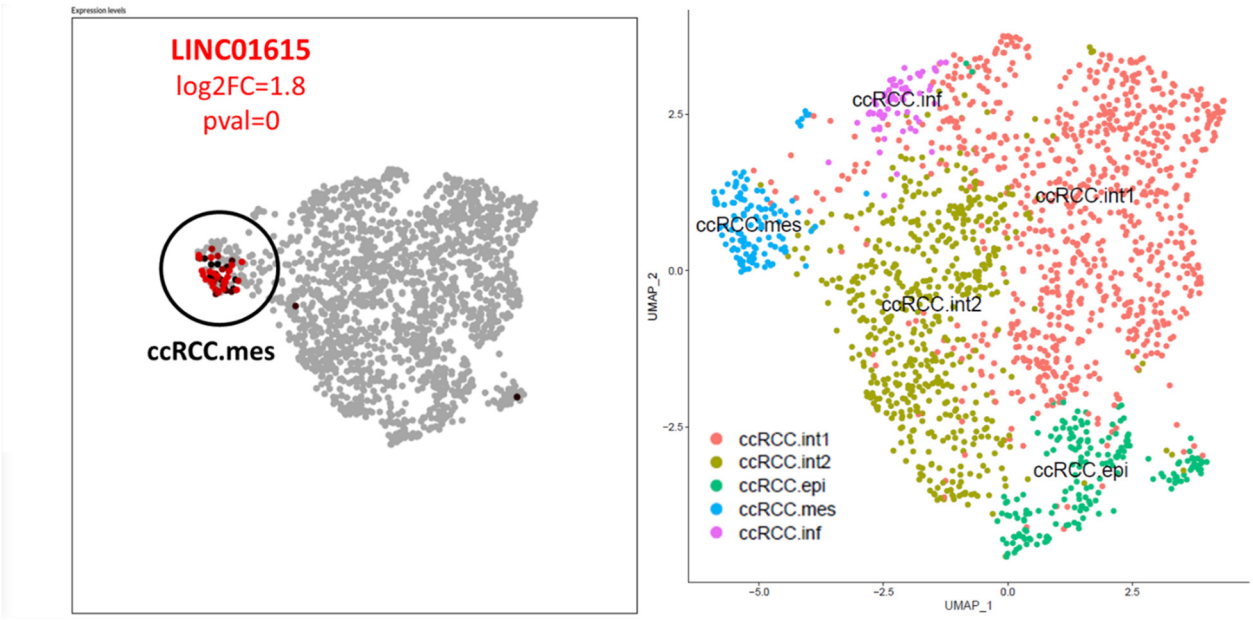


Figure 67. Expression of LINC01615 in ccRCC cells.

Representation of ccRCC cells expressing LINC01615 (black to red gradient) displayed in SCope (left panel) and UMAP showing ccRCC cluster positions (right panel).

b. The mesenchymal tumor cell marker LINC01615

i. LINC01615 in ccRCC

LINC01615 appeared very specific to ccRCC.mes on the heatmap from Figure 65 so I went back to single-cell data and checked its expression in the re-clustering of cancer cells (Figure 67). I found that the gene is actually a cluster marker specific for ccRCC.mes both in this re-clustering (log2FC=1.8, p=0) and in the global analysis (log2FC=1.78, p=0.002) where it showed an expression level around 70-fold higher than in fibroblasts.

I analyzed LINC01615 expression in the TCGA-KIRC samples. First, I performed differential expression analysis between ccRCC tumor and NAT samples. This analysis revealed 5994 genes enriched in tumors and 2602 genes enriched in NATs (log2FC>1, adj-pval<0.05). GSEA analysis showed that the tumor expression program is associated with EMT, cell cycle and inflammation while NATs displayed an epithelial kidney gene expression program linked with oxidative-phosphorylation and metabolism. LINC01615 was found significantly overexpressed (log2FC=3.73, p<0.001) in tumor samples (Figure 68).

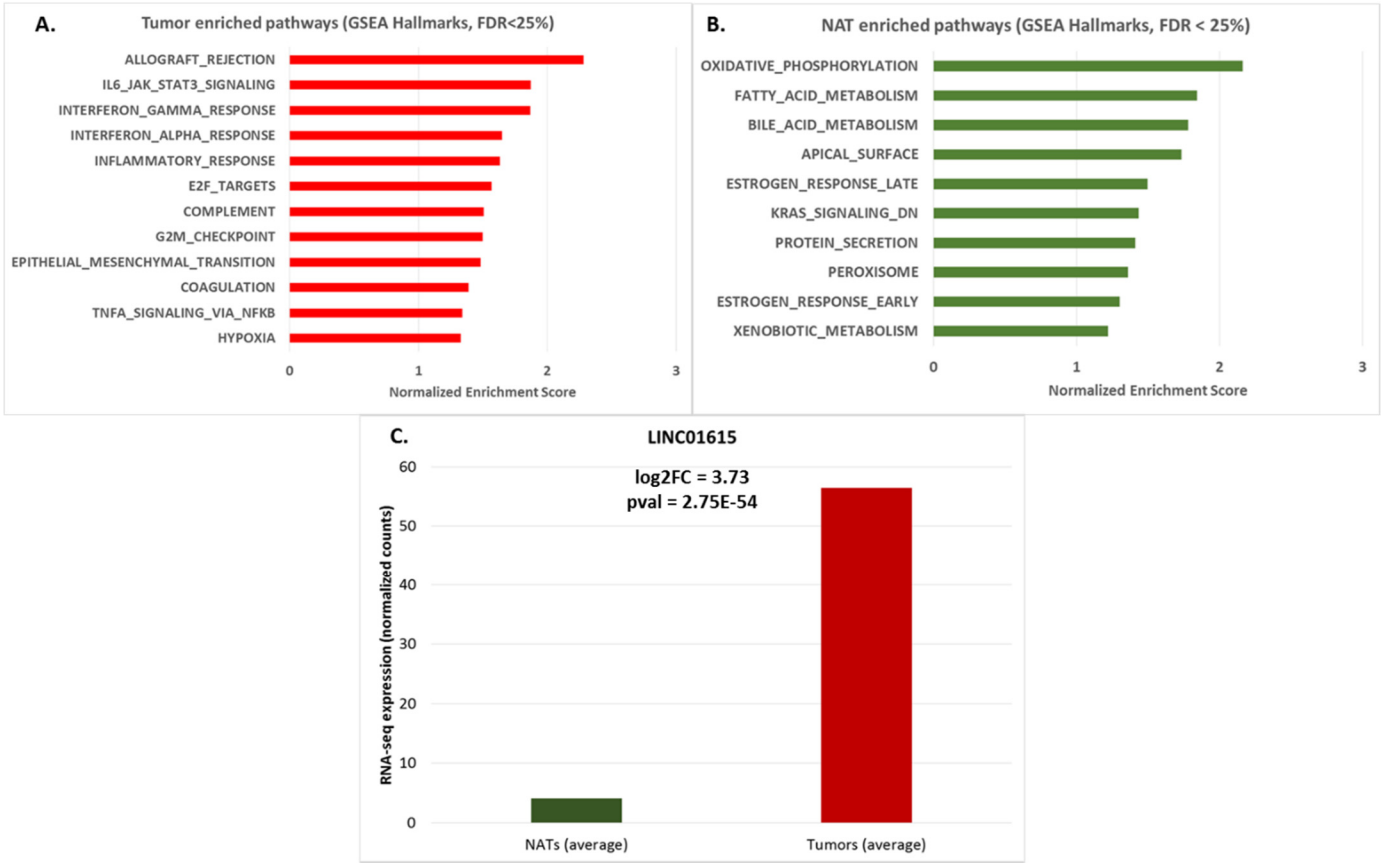


Figure 68. Differential analysis of tumor and normal adjacent tissue samples in TCGA.
Chart of enriched human hallmarks in tumors (**A**) or NATs (**B**) following GSEA analysis. Gene expression levels of LINC01615 with DESeq2 fold-change and adjusted p-value (**C**).

Then, I analyzed the relationship between LINC01615 expression and patient survival (Figure 69). By stratification, patients in the LINC01615-high group displayed significantly worse survival which could be confirmed by univariate COX-regression analysis (HR=27.59, p<0.001). Multiple-regression analysis showed the strong prognostic value of LINC01615 expression (HR=8.24, p=0.004) even after correction with clinical co-variates although, as with MES.score, tumor grade significance is diminished indicating collinearity. Expression of this gene is thus a strong representative of ccRCC.mes presence that can predict patient survival.

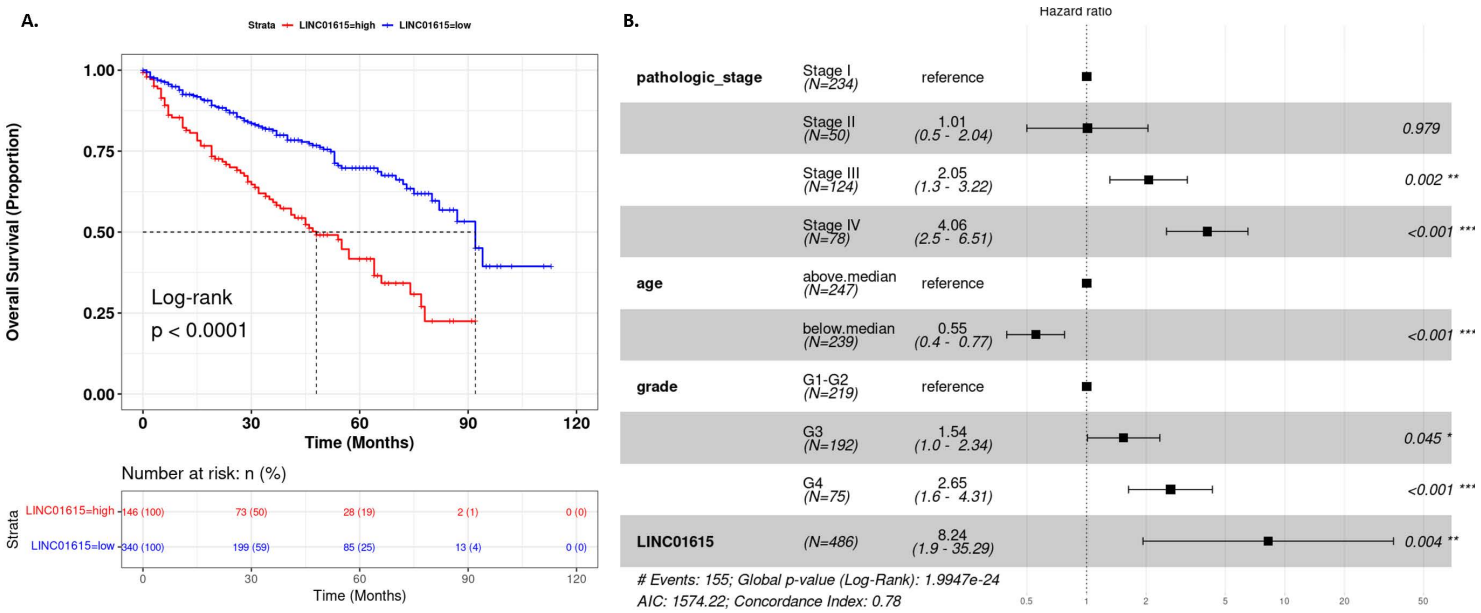
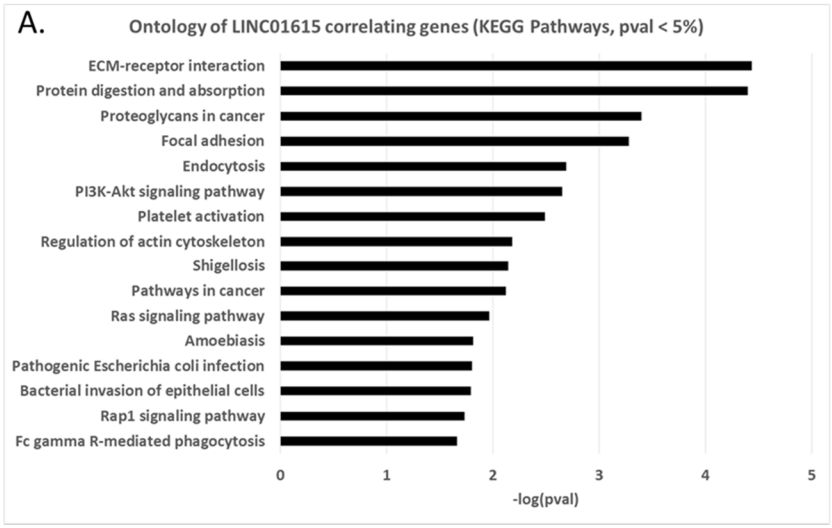


Figure 69. Survival of TCGA-KIRC patients according to LINC01615 expression.

A: Kaplan-Meier overall survival curve of TCGA-KIRC patients stratified by LINC01615 expression (optimal cut-point), un-stratified hazard-ratio from univariate COX-regression analysis with 95% confidence interval and p-value are indicated. **B**: Forest-plot showing results for COX multiple-regression analysis of LINC01615 expression with other clinical co-variates.

Next, I looked for potential regulators of *LINC01615* in ccRCC. First, I performed a co-expression analysis using the BIONIKK samples in order to identify genes correlating with *LINC01615* expression. This analysis revealed 399 correlated genes with ontology associated with ECM modulation and pathways related to invasiveness of cancer cells such as PI3K/Akt[252, 253] and RAS[254] signaling. These genes contained 26 transcription factors including mesenchymal regulators PRRX1, SOX11, ZEB2 and BRN2 (Figure 70). Secondly, I checked cellular models ChIP-seq profile at the LINC01615 locus. I found H3K27ac signal at the promoter for the UOK121 cell line (ccRCC *VHL*/*TP53* mutant) but not in the immortalized kidney proximal tubule cell line RPTEC. This promoter region contained binding sites for MYC, AP-1, AP-2, MAFK, GATA2, STAT3 and CEBPB. Several of these factors are expressed in ccRCC.mes and are good candidate regulators of *LINC01615* expression.



B.

Transcription Factor	Rank (Spearman R)
TSHZ3	1
SHOX2	2
ELK3	3
HMGA2	4
PRRX1	5
SOX11	6
SKIL	7
ETV5	8
SIX4	9
ZEB2	10
ETV4	11
TMF1	12
ZUFSP	13
ZNF197	14
ZNF660	15
KLF17	16
MEIS3	17
ZNF286B	18
ZKSCAN4	19
GSC	20
ZBTB47	21
NKX3-2	22
CGGBP1	23
POU3F2	24
KCMF1	25
ZNF92	26

Figure 70. Genes co-expressed with LINC01615.

A: KEGG pathways ontology of genes correlated with LINC01615 expression (R>0.4) after DAVID analysis. **B**: Table of transcription factors correlating with LINC01615 expression.

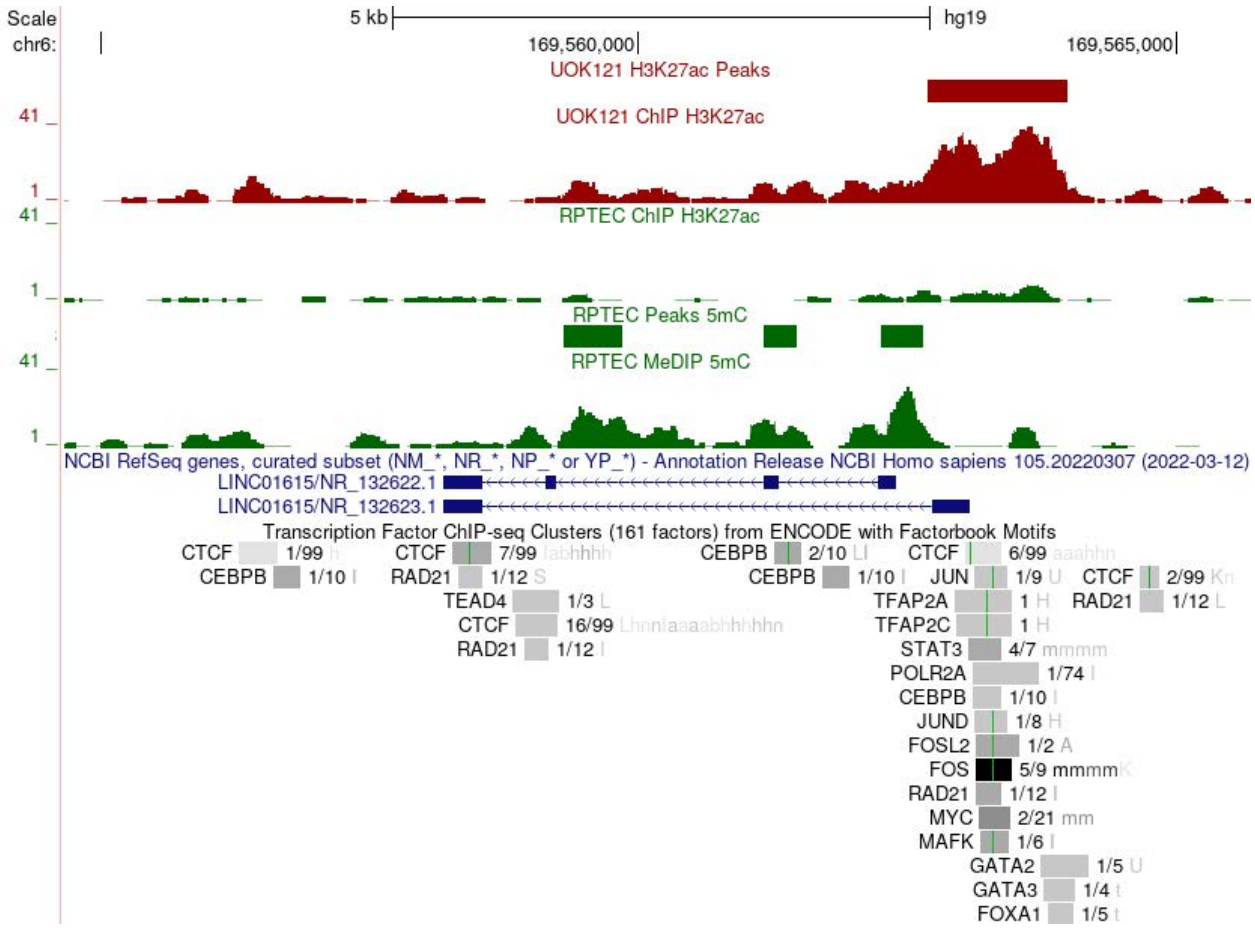


Figure 71. The LINC01615 gene locus.

UCSC view of H3K27ac and 5-methyl-cytosine (5mC) signals colored in red for UOK121 (ccRCC) and in green for RPTEC (proximal tubules) with MACS2 predicted peaks. ENCODE annotation track shows putative binding sites. Experiments performed by Jonathan Thouvenin and Anas Faldoun.

ii. LINC01615 in melanoma

Seeing the strong association of LINC01615 with the mesenchymal state in ccRCC, I checked whether this gene could also play a role in melanoma where dedifferentiation is extensively studied. I first looked into *in-vitro* RNA-seq data from cellular models which are phenotypically well characterized. Our team previously generated data for two melanocytic cell lines (501M, MM074), one intermediate cell line (MM047) and two mesenchymal-like cell lines (MM029, MM099). LINC01615 showed higher expression in intermediate and mesenchymal cell lines (Figure 72). Comparing expression level in the melanocytic 501M cells against the mesenchymal model MM099 revealed significant enrichment of LINC01615 in MM099 (log2FC=6.05, adj-pval<0.001). In a similar manner, the melanocytic MM074 cells adopt an undifferentiated phenotype to resist CDK7 inhibition after THZ1 treatment, mesenchymal resistant cells (MM074R) showed significant up-regulation of LINC01615 (log2FC=2.83, adj-pval<0.001). These observations could be experimentally validated by RTqPCR which detected LINC01615 with a 3-fold increase in MM099 compared to 501M, a level that is comparable to that of FN1 (Figure 73).

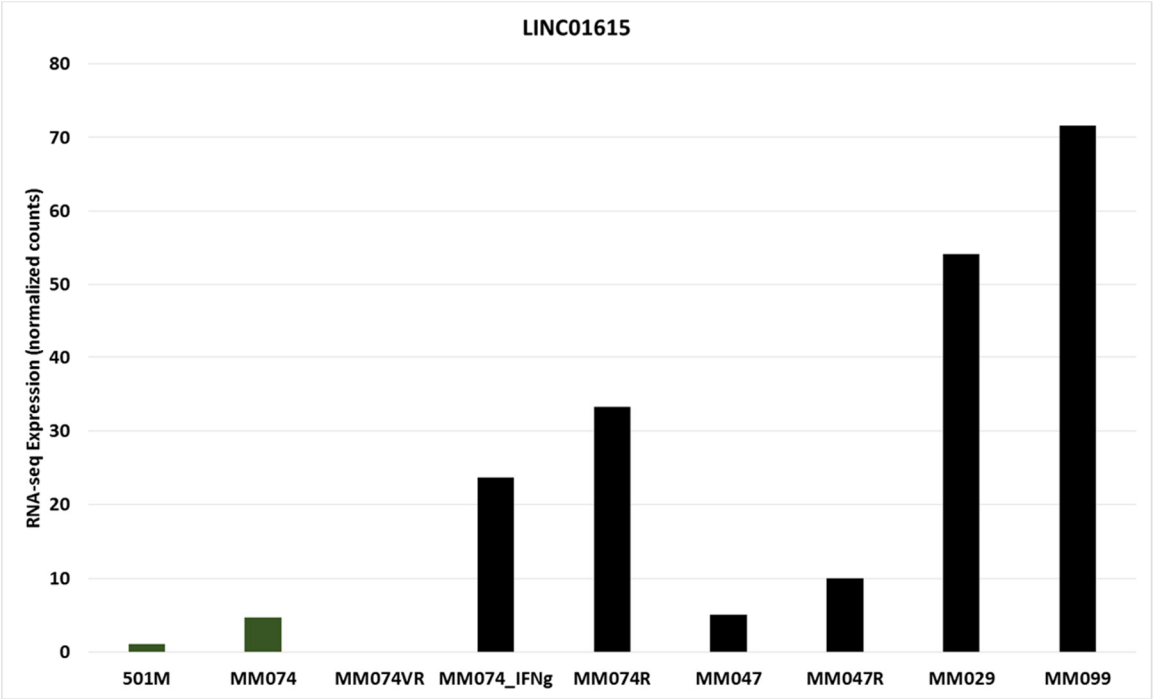


Figure 72. RNA-seq expression of LINC01615 in melanoma cell lines.

Bar chart of RNA-seq LINC01615 expression in baseline, vemurafinib-resistant (VR), THZ1-resistant (R) or treated with interferon-gamma (IFNg) melanoma cell lines. Melanocytic cell lines (501M, MM074, MM074VR) are in green, intermediate (MM047) and mesenchymal-like (MM029, MM099, MM074R, MM074_IFNG) are in black. Experiments performed by Sebastien Coassolo, Bujamin Vokshi and Pietro Berico.

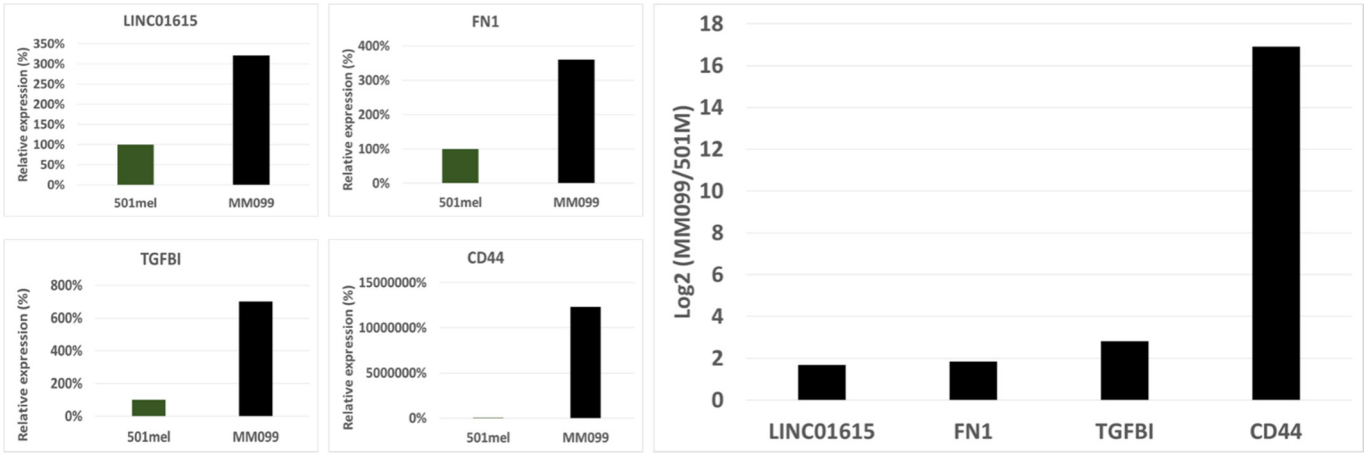


Figure 73. Expression of LINC01615 measured by qPCR in 501M and MM099 cells.

RNA levels of LINC01615 and control mesenchymal markers FN1, TGFBI and CD44 measured by RTqPCR in 501M and MM099 cell lines. The HMBS and GAPDH housekeeping genes were used as normalization control. Experiment performed by Alexandra Helleux.

As LINC01615 was associated with the mesenchymal-like state in cellular models, I then checked whether this would be true *in-vivo* by analyzing RNA-seq data from TCGA skin cutaneous melanoma samples (TCGA-SKCM). I performed two separate consensus clustering of primary and metastatic samples based on the 5000 most differential protein coding genes. Primary tumors were a minority in the SKCM dataset (n=103) and could be clustered into four groups (CCP1-CCP4, Figure 74). I computed marker genes for these clusters by performing DESeq2 differential analysis of each cluster against all other samples. This analysis revealed 678 marker genes for CCP1, 608 for CCP2, 2077 for CCP3 and 389 for CCP4. GSEA analysis (Figure 75) showed CCP1 to be enriched in EMT and Wnt/β-catenin pathways, this appeared to be a more mesenchymal cluster. CCP2 was enriched in cell cycle and OxPhos, this corresponds to a melanocytic signature. CCP3 was enriched in both EMT and inflammation. CCP4 was enriched in inflammatory pathways, this cluster seemed immune-rich as CD45 was among the top enriched gene (log2FC=1.43, adj-pval=0.008). Overall survival analysis of the different CCP clusters showed both mesenchymal clusters CCP1 (HR=2.49, p=0.05) and CCP3 (HR=2.86, p=0.04) to be significantly associated with a worse survival (Figure 76). LINC01615 was significantly enriched in the mesenchymal and inflammatory CCP3 cluster (log2FC=1.07, adj-pval=0.009) which was also the only one with a significant depletion of MITF expression (log2FC=-1.04, adj-val<0.001). Thus, in primary tumors LINC01615 expression appeared associated to the mesenchymal state which correlates with worse patient overall survival though the number of events is rather limited in this cohort (n=29).

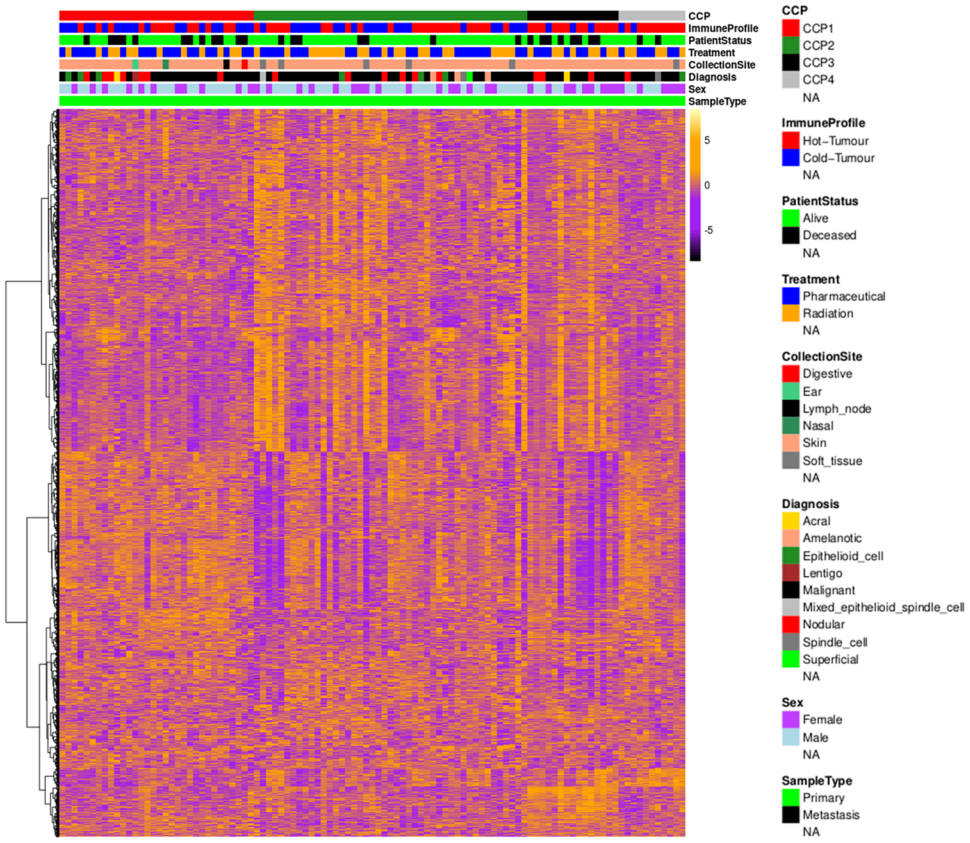


Figure 74. Consensus clustering of melanoma primary tumor samples in TCGA-SKCM.

Heatmap of row-scaled expression for the 5000 most differential protein-coding genes following consensus clustering of TCGA-SKCM primary tumors (clusters named CCP1-CCP4).

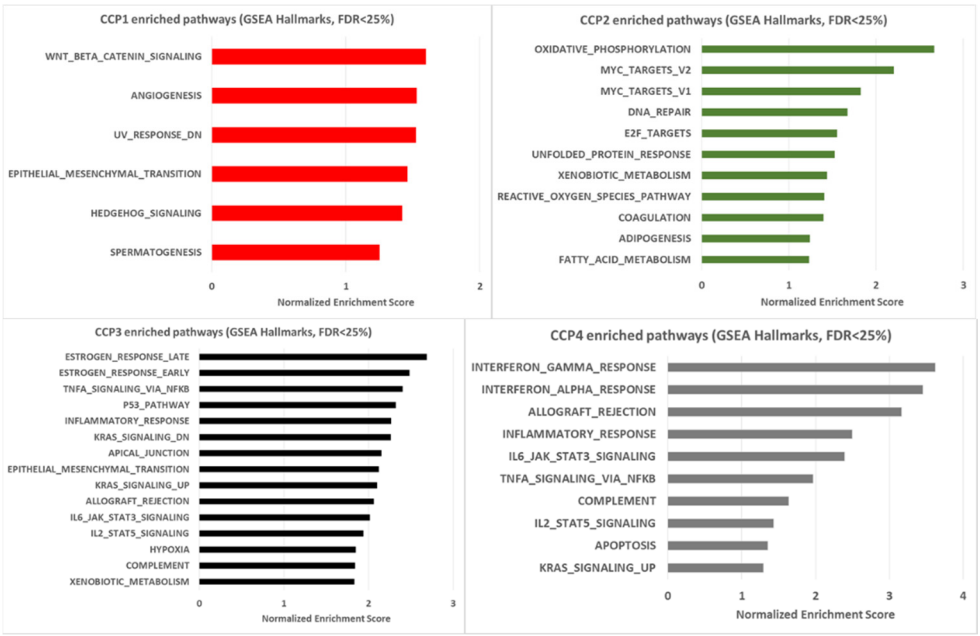


Figure 75. Enriched pathways in CCP clusters.

Chart of enriched human hallmarks following GSEA analysis of each CCP cluster against all other samples.

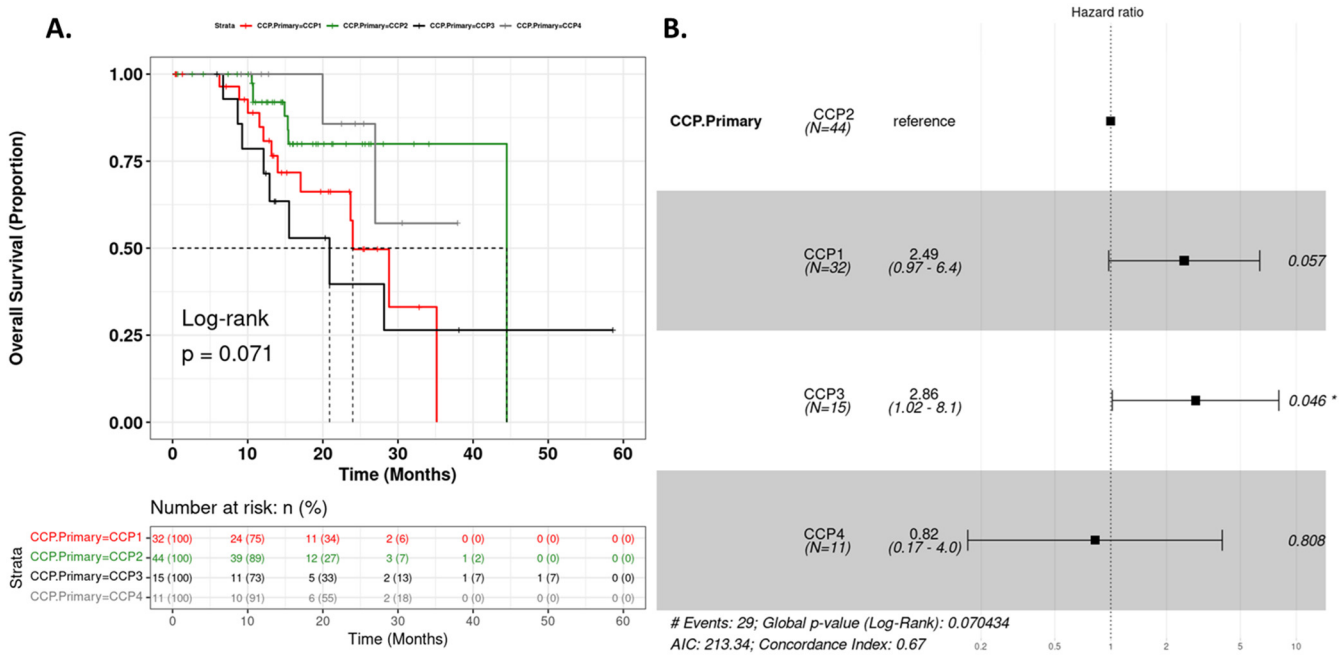


Figure 76. Patient survival in TCGA-SKCM according to CCP clusters.

Kaplan-Meier overall survival curve of TCGA-SKCM patients grouped by CCP cluster (**A**), forest-plot showing result of univariate COX-regression analysis (**B**).

Metastatic tumor samples constitute the majority of TCGA-SKCM cohort (n=367) and could be clustered into 5 groups (CCM1-CCM5, Figure 77). Again, I computed cluster markers for CCM1 (1843), CCM2 (311), CCM3 (839), CCM4 (1136) and CCM5 (1822). GSEA analysis showed CCM1 to be enriched in EMT and cell-cycle, CCM2 in OxPhos, CCM3 in OxPhos and cell-cycle like CCP2, CCM4 in EMT and inflammation like CCP3, and CCM5 was enriched in inflammation like CCP4 (Figure 78). I computed overall survival across these clusters which revealed the melanocytic cluster CCM3 (HR=3.1, p<0.001) to be the one most significantly associated with worse survival closely followed by CCM1 (HR=2.5, p<0.001) and CCM2 (HR=2.4, p=0.005) which both have cell-cycle or OxPhos signatures (Figure 79). This result indicates that higher lethality of metastases is associated with the presence of melanocytic cells, perhaps due to their proliferative nature. Once again, LINC01615 was found significantly enriched in the mesenchymal and inflamed cluster CCM4 (log2FC=1, adj-pval<0.001) that was depleted in MITF expression (log2FC=-1.1, adj-pval<0.001). Of note, LINC01615 showed non-significant positive enrichment in CCM1 (log2FC=0.22, adj-val=0.41) indicative of the probable presence of mesenchymal-like melanoma cells in this cluster.

All in all, TCGA-SKCM analyses show LINC01615 expression to be detectable and related to the mesenchymal state *in-vivo* similar to what was observed in cellular models. Thus, measuring this gene could have practical applications in melanoma as well as ccRCC.

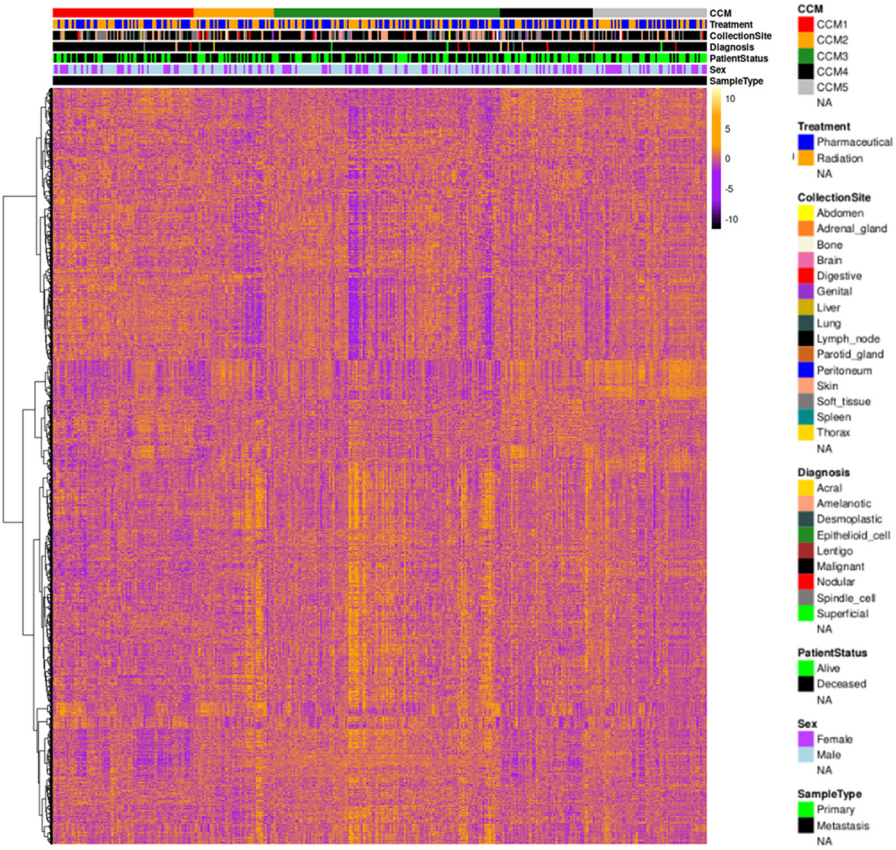


Figure 77. Consensus clustering of melanoma metastatic tumors in TCGA-SKCM.

Heatmap of row-scaled expression for the 5000 most differential protein-coding genes following consensus clustering of TCGA-SKCM metastatic tumor tissues (clusters named CCM1-CCM5).

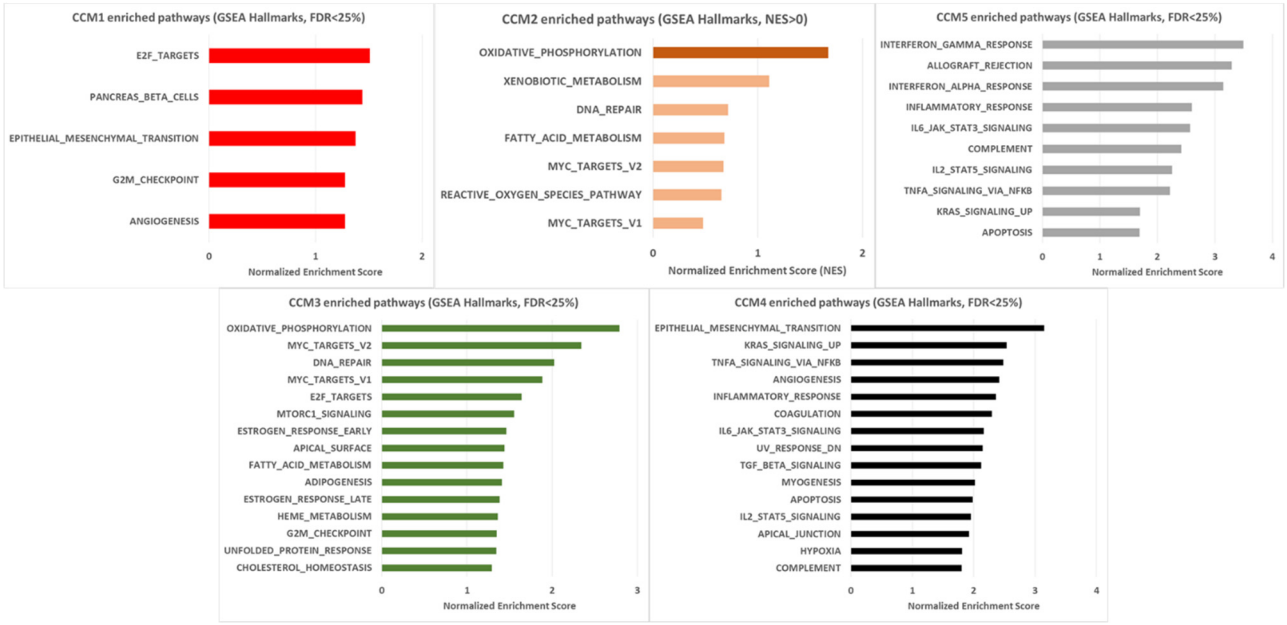


Figure 78. Enriched pathways in CCM clusters.

Chart of enriched human hallmarks following GSEA analysis of each CCM cluster against all other samples.

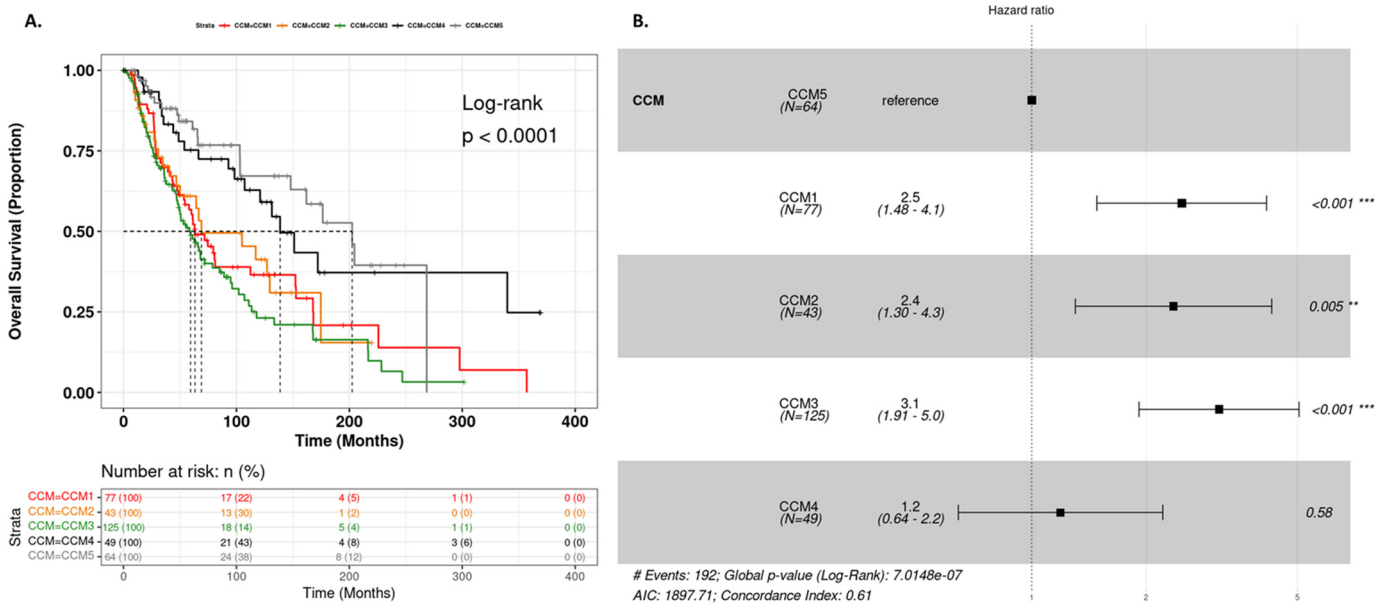


Figure 79. Patient survival in TCGA-SKCM according to CCM clusters.

Kaplan-Meier overall survival curve of TCGA-SKCM patients grouped by CCM clusters (**A**), forest-plot showing result of univariate COX-regression analysis (**B**).

4. SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance.

Bujamin H. Vokshi[1]*, Guillaume Davidson[1]*, Nassim Tawanaie Pour Sedehi[1], Alexandra Helleux[1], Marc Rippinger[1], Alexandre R. Haller[1], Justine Gantzer[1,2], Jonathan Thouvenin[1,2], Philippe Baltzinger[1], Rachida Bouarich[7], Valeria Manriquez[7], Sakina Zaidi[7], Priya Rao[8], Pavlos Msaouel[3], Xiaoping Su[4], Hervé Lang[5], Thibault Tricard[5], Véronique Lindner[6], Didier Surdez[8,9], Jean-Emmanuel Kurtz[2], Franck Bourdeaut[7], Nizar M. Tannir[3], Irwin Davidson[1,10]¤ and Gabriel G. Malouf[1,2,10]¤.

1 Department of Cancer and Functional Genomics, Institute of Genetics and Molecular and Cellular Biology, CNRS/INSERM/UNISTRA, 67400 Illkirch, France.

2 Department of Medical Oncology, Institut de Cancérologie Strasbourg Europe, 67200 Strasbourg, France.

3 Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

4 Department of Bioinformatics and Computational Biology, Division of Quantitative Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

5 Department of Urology, CHRU Strasbourg, Strasbourg University, 67000, Strasbourg.

6 Department of Pathology, CHRU Strasbourg, Strasbourg University, 67200, Strasbourg.

7 INSERM U830, Équipe Labellisée LNCC, Diversity and Plasticity of Childhood Tumors Lab, Institut Curie Research Centre, 75005 Paris, France.

8 Balgrist University Hospital, University of Zurich, Zurich, Switzerland

9 INSERM, U830, Pediatric Translational Research, PSL Research University, SIREDO Oncology Center, Institut Curie, Paris, France.

10 'Équipe Labellisée' Ligue National contre le Cancer.

*Equal contribution as co-first authors

¤ Equal contribution as co-senior authors

To whom correspondence should be addressed

E-mail : irwin@igbmc.fr, maloufg@igbmc.fr

Running Title: Role of SMARCB1 in renal medullary carcinoma

Key words: Ferroptosis / MYC / Renal medullary carcinoma / SMARCB1 / TFCP2L1

The authors declare no potential conflicts of interest.

Published in Nature Communications, 26 May 2023

A. Abstract.

Renal medullary carcinoma (RMC) is an aggressive tumour driven by bi-allelic loss of SMARCB1 and tightly associated with sickle cell trait. However, the cell-of-origin and oncogenic mechanism remain poorly understood. Using single-cell sequencing of human RMC, we defined transformation of thick ascending limb (TAL) cells into an epithelial-mesenchymal gradient of RMC cells associated with loss of renal epithelial transcription factors TFCP2L1, HOXB9 and MITF and gain of MYC and NFE2L2-associated oncogenic and ferroptosis resistance programs. We describe the molecular basis for this transcriptional switch that is reversed by SMARCB1 re-expression repressing the oncogenic and ferroptosis resistance programs leading to ferroptotic cell death. Ferroptosis resistance links TAL cell survival with the high extracellular medullar iron concentrations associated with sickle cell trait, an environment propitious to the mutagenic events associated with RMC development. This unique environment may explain why RMC is the only SMARCB1-deficient tumour arising from epithelial cells, differentiating RMC from rhabdoid tumours arising from neural crest cells.

B. Introduction.

First described in 1995[1], renal medullary carcinoma (RMC) is a lethal malignant neoplasm arising from the kidney medulla region. Despite its relative rarity, RMC is the third most common renal cancer among young adults[2]. It typically afflicts male patients of African descent with sickle cell trait at a median age of 28 years, yet the association is still poorly understood[3,4]. RMC is highly aggressive with most patients presenting metastatic disease at the time of diagnosis and less than 5% survive longer than 36 months[5,6]. In addition, RMC tumours are resistant to targeted therapies used for other renal cancers and the best available cytotoxic chemotherapy regimens produce a brief objective response in less than 30% of cases[7,8]. Alternative treatments such as anti-angiogenics, EZH2 inhibitors and immunotherapy have been tested with varying success[6]. RMC tumour tissue resembles a high-grade carcinoma exhibiting reticular or cribriform patterns and usually stain positive for VIM, MUC1, pankeratins, PAX8, HIF1α and VEGF[8,9]. RMC are also characterized by a strong desmoplasia, a prominent inflammatory infiltrate as well as the frequent presence of sickled red blood cells[10,11].

The hallmark of RMC is loss of SMARCB1 expression[12], a core subunit of the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodelling complex. Several mechanisms lead to SMARCB1 loss in RMC including deletions, point mutations, inactivating translocations and loss-of-heterozygosity[6]. SMARCB1 loss is also the hallmark of malignant rhabdoid tumours (RTs), atypical teratoid/rhabdoid tumours (ATRTs) and epithelioid sarcomas (ESs). The majority of RTs and RMCs share common features such as their renal location and low mutation burden[6]. We recently characterized the molecular characteristics of RMC identifying frequent chromosome 8q gain associated with a copy-number gain of MYC[6]. SMARCB1 loss activates the MYC pathway resulting in increased DNA replication stress and DNA damage response. RMC are thought to arise from the distal region of the nephron, however evidence is limited to correlation inference using bulk RNA-seq data from 8 nephron

biopsies with identified renal cell populations[6,13]. Thus, despite the above pathology and molecular characterization, the cell of RMC origin is as yet not fully defined and the molecular mechanisms involved in oncogenic transformation associated with SMARCB1 loss remain poorly characterized.

To address these issues, we integrated data from single-cell (sc)RNA sequencing of human tumours, multi-region RNA sequencing, bulk transcriptomic data from 2 RMC cohorts, and SMARCB1 gain of function experiments in cellular models. This comprehensive approach revealed how the thick ascending limb (TAL) cells are transformed into RMC through a transcriptional switch involving loss of renal master regulator TFCP2L1 and activation of a MYC and NFE2L2-associated transformation and ferroptosis resistance programs.

C. Results.

a. RMC ontogeny and molecular characterization of tumour cell states.

To characterize the molecular features and ontogeny of RMC, we performed scRNA-seq on a post-treatment primary nephrectomy from an RMC patient with lung metastases at diagnosis. The patient showed complete response following 6 cycles of Methotrexate, Vinblastine, Doxorubicin, Cisplatin (MVAC) treatment. A total of 996 cells from the residual tumour site and 1722 cells from normal adjacent renal tissue (NAT) were aggregated and analysed. Seurat UMAP clustering revealed 14 distinct populations amongst which were 7 renal epithelial clusters and 7 stromal and immune clusters (Fig. 1a-b). Epithelial clusters comprised 6 groups of cells from the proximal and distal tubules and 1 group of collecting duct cells each expressing specific markers (Fig. 1c). Amongst these, we identified thick ascending limb (TAL) cells with expression of *SLC12A1*, *EPCAM*, *CDH1* and keratin 7 (*KRT7*), consistent with previous renal scRNA-seq datasets[14–16].

After merging cancer and NAT samples, we identified populations enriched in the tumour sample comprising TAMs (tumour-associated macrophages) and 2 clusters of cells harbouring an epithelial mesenchymal transition (EMT) signature that we identified to be the RMC tumour and CAF (cancer-associated fibroblast) cells (Fig. 1b). All three clusters expressed specific markers (*LYZ*, *MMP7* and *POSTN*, respectively) with cytokeratin expression in RMC cells (Fig. 1c). Further analyses of RMC and CAFs showed that each expressed overlapping as well as distinct sets of EMT markers (Fig. S1a and Dataset S1a).

The UMAP plot revealed that RMC cells were located close to the TAL population, consistent with a putative cell of origin located in the distal part of the nephron. We interrogated all renal epithelial populations for shared transcriptional signatures with RMC cells and found the best correlation with TAL cells of the kidney medulla (Fig. 1d). Differential gene expression analysis of a pseudo-bulk reconstitution of the RMC versus the CAF populations identified about 150 signature genes for RMC and 50 genes for CAF (Fig. 1e). RMC cells showed a specific oncogenic program, but retained many genes associated with TAL and more broadly epithelial identities (Fig. 1e). RMC and CAF cells did however commonly express EMT genes such as *VIM* and *FN1*, in contrast to TAL cells (Fig. 1c, 1e). Altogether, these

observations identified TAL cells to be the normal renal population most related to RMC and hence the likely cell-of-origin.

To investigate intra-tumoural heterogeneity, we re-clustered the RMC cells identifying distinct RMC1 and RMC2 subpopulations (Fig. 1f). Gene Set Enrichment Analysis (GSEA) revealed that RMC1 were enriched in oxidative phosphorylation (OXPHOS), whereas RMC2 were enriched in EMT, interferon gamma, inflammatory response and hypoxia (Fig. 1g). Correlation of the RMC1 and RMC2 specific signatures to those of normal tubules revealed that RMC1 partly retained a TAL signature that was reduced in RMC2 (Fig. 1h). These observations were independently confirmed by SWNE trajectory analysis that traced transformation of TAL cells to RMC2 via the RMC1 population with some cells retaining a more epithelial identity (Fig. 1i). This was further supported by separation of RMC cells into a 'stressed' epithelial-like phenotype with higher levels of cytokines (*IL8*, *LCN2*), keratins and epithelial markers such as *CDH1*, *CLDN1* and into RMC2 cells with higher expression of mesenchymal markers such as *SFRP2*, *CDH2* and *FN1*. Thus, this RMC tumour comprised epithelial-like RMC1 cells and mesenchymal-like RMC2 cells (Fig. 1j).

We next analysed a naive RMC sample from a primary nephrectomy of a 16-year-old patient with regional lymph node and adrenal gland metastases (pT4N1M1) at presentation capturing a total of 3372 cells. Following surgery, the patient showed rapid progressive disease under adjuvant MVAC regimen. The patient was also primary resistant to durvalumab-tremelimumab immunotherapy and EZH2 inhibitor Tazemetostat leading to death within one year of diagnosis. Among 3372 captured cells, a large group of RMC cells was identified along with TAMs and other CD45-expressing immune cells (Natural killers, neutrophils and T-cells), *POSTN*-expressing CAFs, and an unexpected population of tumour-associated TAL2/3 cells (Fig. 1k-l). Both the RMC and TAL cells, that segregated into two closely located groups on the UMAP plot, expressed *EPCAM* as well as a cytokeratin signature (Fig. 1l-m). The TAL3 population could be distinguished from TAL2 cells by the lowered expression of the *SLC12A1*, *HOXB9* and *PAX8* renal identity markers (Fig. 2a and Dataset S1b). The RMC3 and RMC4 populations were highly similar with the smaller RMC4 cluster displaying an additional G2/M phase cell cycle signature designating them as mitotic RMC3 cells (Fig. 2a). The SWNE trajectory representation of the TAL and RMC populations illustrated the progressive loss of TAL identity markers from the most differentiated TAL2 to TAL3 with some TAL3 cells closely related to the RMC group that retained an epithelial-like signature (Fig. 2b).

Aggregation of the batch corrected data from the two tumours (Fig. 2c), highlighted differences between the TAL and RMC subpopulations illustrated using a collection of epithelial, mesenchymal, endoplasmic reticulum (ER)-related stress genes (Fig. 2d, Fig. S1b). The RMC3/4 cells from the naive tumour had a marked epithelial character compared to intermediate RMC1 cells from the treated tumour, whereas RMC2 cells had the most mesenchymal phenotype (Fig. 2d). GSVA analyses revealed enrichment for cell cycle in RMC4 cells, OXPHOS and apical junction in RMC3 cells, and EMT and interferon gamma response in RMC1 and RMC2 cells (Fig. 2e). SWNE trajectory analyses highlighted the gradient of epithelial to mesenchymal phenotypes of the different populations (Fig. 2f).

We further performed multi-region tumour RNA-seq on a cohort of four patients, for which single region transcriptome sequencing was previously reported along with that of 7 additional cases (6, designated as the MDACC cohort; and Dataset S2). Overall, we generated

an additional 25 bulk RNA-seq from multiple regions of these primary tumours and the corresponding regional lymph nodes as well as 3 NATs and analysed intra- and inter-tumour heterogeneity using CIBERSORTx deconvolution to infer their RMC1-3 composition (Fig. S1c-d). For clarity, we did not include the cycling RMC4 signature. Primary tumour sections showed varying composition, some more enriched in the epithelial-like signature, others with epithelial-like and intermediate signatures, and a third group with all 3 signatures. In contrast, the lymph node metastases sections were strongly enriched in the mesenchymal-like signature. These data unravelled intra-tumour heterogeneity in RMC and the importance of tumour cells with a mesenchymal signature to metastatic progression.

We used SCENIC regulon analyses software to identify transcriptional regulatory networks underlying the above signatures[17]. Comparison of the TAL and RMC populations from the treated tumour revealed a transcriptional switch from high HOXB9 and TFCP2L1 activity in TAL1 cells, to high MYC, HIF1A, YY1 and NFE2L2 activity in RMC cells (Fig. 2g). These data were consistent with the known role of MYC in RMC transformation, whereas TFCP2L1 is a previously described determinant of the distal portion of the nephron[18]. Top TAL regulons were progressively lost upon transformation into RMC1 and RMC2 populations exemplified by TFCP2L1, PPARGC1A, perhaps contributing to the OXPHOS signature[19], and HOXB9, whereas others like SOX9 were maintained (Fig. 2h).

Comparable observations were made between the TAL and RMC populations of the naive tumour with loss of TFCP2L1 activity and gain of MYC and NFE2L2/3 activity (Fig. 2i). Interestingly, while TAL2/3 cells displayed TFCP2L1 activity they also showed a stress signature with prominent activity of ATF4, XBP1 and HIF2A. Moreover, they further showed YY1 and MYC activity, hallmarks of RMC cells. TAL1 cells were derived from NAT, whereas TAL2/3 cells were tightly associated with the RMC cells in the tumour sample and showed a stressed pre-tumoural phenotype with activation of several RMC regulons. Each RMC population displayed a characteristic regulon activity such as cell cycle (BRCA1, E2F4/6) in RMC4 cells[20,21], epithelial-like (OVOL2, ELF3) in RMC3 cells[22,23] and mesenchymal-like (HES1, FOSL2) in RMC2[24,25]. Notably, activity of the PAX8 renal identity marker was strongly reduced in the RMC1 and RMC2 populations compared to RMC3 (Fig. S1e).

The role of TFCP2L1 in driving expression of epithelial genes was reinforced by analyses of the Cancer Cell Line Encyclopedia (CCLE) showing positive correlation between *TFCP2L1* (and also *OVOL2*) and *EPCAM* (Fig. S1f). Similarly, *TFCP2L1* correlated with epithelia markers and anti-correlated with mesenchymal markers (Fig. S1g). In the TCGA chromophobe renal cell carcinoma dataset, originating also from distal tubules, *TFCP2L1* and *MITF* expression correlated with that of *CDH1* (Fig. S1h).

The above data defined an EMT gradient of RMC cells defined by distinct transcriptional signatures also found in patient tumour samples. NAT-derived TAL1 cells were further distinguished from tumour-associated TAL2/3 cells that displayed a stressed, pre-tumoral phenotype in their transcriptional signatures and regulon activities.

b. Tumour cell state of a patient derived RMC xenograft.

We further analysed a patient derived xenograft (IC-PDX-132) from an RMC tumour treated with 6 cycles of cisplatin, gemcitabine and bevacizumab that had undergone 4 passages of subcutaneous injections on immunocompromised mice. Around 10,000 cells were captured and the sequences aligned to a human-mouse hybrid genome. A large group of human RMC tumour cells were identified with high expression of *EPCAM* and the bulk RMC signature as well as a group of murine cells corresponding to CAFs and pericytes, TAMs and monocytes, and a smaller number of other immune cells (Fig. S2a-c and Dataset S1c). A third group that we tagged 'LQ' (low-quality) comprised cells with high levels of mitochondrial genes and potential doublets that were removed from the subsequent analyses.

Re-clustering the RMC cells revealed 4 subpopulations together with some mouse cells of undefined identity that were not further considered (Fig. S2d). The RMC8 cluster showed a strong cell cycle signature and regulon activity designating them as mitotic RMC cells, whereas RMC6 cells displayed high hypoxia and stress-associated regulons such as ATF4 and DDIT3 (Fig S2d-f) 26. RMC5 and RMC7 on the other hand corresponded to epithelial-like and intermediate state cells respectively analogous to the RMC3 and RMC1 cells in the human tumours (Fig. S2e). No distinct highly mesenchymal population was observed, although the mitotic RMC8 cells showed the most dedifferentiated phenotype and highest expression of *FN1* and *CD44*. SCENIC analyses of these populations identified the key MYC, YY1, and NFE2L2 regulons in the RMC cells as seen above in primary human tumours (Fig S2f).

These analyses revealed that the RMC PDX comprised principally epithelial-like, intermediate and mitotic RMC cells as well as a subpopulation of hypoxic cells consistent with the idea that angiogenesis could not fully irrigate the rapidly proliferating tumour.

c. Characterization of the RMC microenvironment.

In addition to TAL and RMC cells, scRNA-seq revealed prominent CAF and TAM populations in the RMC tumour microenvironment (TME). Analyses of CAFs from both tumours revealed two populations with either a myofibroblast myCAF signature (CAF1) predominant in the treated tumour or an inflamed iCAF signature (CAF2) in the naive tumour (Fig. S3a). Renal CAFs may arise from pericyte-like mesangial cells[27]. SWNE analyses incorporating NAT-derived mesangial (MES) cells supported the idea they gave rise to the two CAF populations.

Analyses of the TAM population identified TAM1 cells displaying a pro-inflammatory M1 signature (Fig. S3b). In contrast, TAM2 and TAM3 displayed an anti-inflammatory M2 signature with high expression of known M2 markers *IL10* and *MAF*[28] that was strongest in TAM3. SWNE trajectory analysis further confirmed the idea that the TAM2 signature represented an intermediate state between the most polarized TAM1 and TAM3 states.

We then applied the CAF and TAM signatures to the bulk-RNA-seq data from the patient tumour sections as described above. CAF2 cells were detected in all primary and metastases sections, whereas CAF1 were not present in all primary sections and lowly

represented in metastases sections (Fig. S3a). Likewise, the TAM2 and TAM3 signatures were detected in a subset of primary and metastases sections, whereas the TAM1 signature was poorly represented in the majority of primary tumour sections, but was highly enriched in the lymph node metastases sections (Fig. S3b).

These analyses showed that the naive tumour and untreated primary patient sections displayed a pro-tumoural, immunosuppressive microenvironment with predominantly iCAFs and M2-type TAMs. However, the MVAC-treated microenvironment was characterized by M1-type TAMs and myCAFs.

d. Cultured RMC cells recapitulate the EMT gradient.

To better define the mechanism by which SMARCB1 loss drives transition from the TFCP2L1-TAL epithelial program to the MYC-driven oncogenic program, we analysed RMC2C and RMC219 cells[6,29]. RMC219 cells displayed a regular rounded morphology similar to primary RPTEC renal epithelial cells (Fig. 3a). RMC2C cells were larger with a more mesenchymal morphology and were much more invasive than the RMC219 cells (Fig. 3a-b). Similarly, flow cytometry indicated that RMC219 cells were EPCAM high, whereas RMC2C cells were CD44 high (Fig. 3c), a marker of RCC aggressiveness[30]. A similar analysis of the UOK360 and UOK353[31] lines by flow cytometry revealed intermediate phenotypes. UOK360 displayed higher EPCAM and lower CD44 than UOK353 and more resembled RMC219 cells (Fig. S4a-b). Note however that UOK360 expressed both CD44 and EPCAM discriminating them for the most epithelial RMC219 cells. UOK353 on the other hand, had lower EPCAM, but CD44 levels closer to the RMC2C cells. Moreover, we observed a progressive increase in invasive capacity along the EMT gradient from RMC219-UOK360-UOK353-RMC2C (Fig. S4c). Cultured RMC cells therefore formed an EMT gradient as observed in the scRNA-seq data on the human tumours.

We used RNA-seq to characterize the most epithelial and mesenchymal RMC219 and RMC2C lines identifying an extensive set of differentially expressed genes with preferential expression of epithelial markers in RMC219 cells and mesenchymal markers in RMC2C cells (Fig. 3d). GSEA revealed enrichment of EMT, inflammatory response and hypoxia in RMC2C cells, as seen in the RMC2 tumour population, and enrichment of cell cycle and DNA repair in RMC219 cells (Fig. 3e). Moreover, while OXPHOS was enriched in RMC219, glycolysis was enriched in RMC2C suggesting a metabolic switch upon EMT. *MITF* and *POU3F3*, previously reported determinants of nephron morphogenesis and TAL cell differentiation[32,33], were preferentially expressed in RMC219 cells, whereas EMT-transcription factors like *TWIST1*/*2* and *SNAI2* were preferentially expressed in RMC2C cells (Fig. 3f). Immunoblot analyses confirmed higher expression of VIM, and SNAI2 in RMC2C and higher expression of CDH1 and MITF in RMC219 cells (Fig. 3g). Both cell lines however showed expression of NFE2L2 and MYC and lacked SMARCB1. These cell lines therefore reproduced epithelial-like and mesenchymal-like phenotypes analogous to those observed in human tumours.

e. SMARCB1 re-expression in RMC cells represses the oncogenic program.

We analysed expression of SWI/SNF subunits in RMC2C cells compared to other SMARCB1-deficient cell lines and HEK293T kidney cells. As expected SMARCB1 was absent from all tumour lines (Fig. S5a-b). The catalytic ATPase subunit SMARCA2 (BRM) was absent in all lines except VA-ES-BJ (epithelioid sarcoma), while SMARCA4 (BRG1) was detected in all lines except CHLA-06-ATRT (rhabdoid tumour). RMC2C cells showed the most important changes in SWI/SNF composition with absence of SMARCD3, ARID2 and lowest expression of DPF3, PBRM1, BRD7 and ARID1A. Although the bulk patient RNA-seq data also comprised signal from CAF and TAM cells, RMC-specific reductions in SMARCA2, and DPF3 expression could still be observed (Fig. S5c).

We engineered RMC2C and RMC219 cells to re-express SMARCB1, or mCherry as control, in a doxycycline (Dox)-dependent manner. SMARCB1 was maximally expressed in both RMC cell lines already 12 hours after Dox addition (Fig. 3h). SMARCB1 expression in RMC2C cells was comparable to that seen in HEKT cells, seen in almost all cells of the population and was integrated into SWI/SNF and co-precipitated with BRG1 (Fig. S6a-c]. The renewed presence of SMARCB1 induced rapid re-expression of SMARCA2, but slower re-expression of DPF3 (Fig 3h). Similarly, the TAL-associated TFCP2L1, MITF and CDH1 were also induced, whereas MYC, NFE2L2 and EMT markers VIM and FN1 were down-regulated. Each line showed a similar response, but with faster kinetics in the epithelial RMC219 cells where the oncogenic program was more rapidly repressed and the epithelial program faster induced than in RM2C2 cells. SMARCB1 re-expression therefore reversed key transcriptional changes observed during TAL to RMC transformation.

To globally assess gene expression upon SMARCB1 re-expression, we performed RNA-seq in each cell line 12 and 48 hours after Dox-treatment. In RMC2C cells, a rapid transcriptional response was seen with 938 down-regulated and 1364 up-regulated genes after 12 hours compared to RMC219 cells where only 12 genes were up-regulated over the same period (Fig. 4a, Dataset S3). After 48 hours, a larger number of up and down-regulated genes were observed in both cell lines (Fig. 4b and Fig. S6d). Despite the differences in kinetics and numbers of affected genes, GSEA analyses revealed that in both lines, genes down-regulated were involved in oncogenic functions such as cell cycle and proliferation, designated by the GSEA terms MYC or E2F-targets in agreement with accumulation of G1/S phase RMC2C cells 12 and 48 hours after Dox treatment (Fig. S6e). Up-regulated genes were designated by epithelial-like program terms such as cell adhesion, apical junction and apical surface (Fig. 4c). Comparison with bulk-RNA-seq from RMC patients relative to their NAT from both MDACC and Institut Curie cohorts showed the opposite profile with genes up-regulated in the SMARCB1-deficient tumours enriched in proliferation, cell cycle and JAK-STAT3 pathway, whereas those down-regulated associated with apical surface (Fig. 4d). Similarly, while OXPHOS was increased upon SMARCB1 expression in cell lines, it was reduced in RMC tumours. RMC cell lines hence reproduce phenotypes and transcriptional signatures seen in RMC tumours whose key features were reversed by SMARCB1 re-expression.

f. SMARCB1 re-expression in RMC cells induces ferroptotic cell death.

SMARCB1 re-expression induced cell death with a 10-20-fold increase in the number of Annexin V-expressing cells (Fig. 4e). RMC219 cells responded rapidly with many dead cells detected by 24 hours after Dox addition, whereas death of RMC2C cells was evident at 48 hours, but required a longer time to reach higher levels (Fig. 4e). To understand the mechanism of cell death, we examined the gene expression changes and noted that Heme metabolism was amongst the pathways strongly up-regulated upon SMARCB1 re-expression and down-regulated in RMC patients (Fig 4d). The Heme metabolism GSEA term covers iron homeostasis, response to reactive oxygen species and ferroptosis (Fig. 4d, right panel). Following SMARCB1 re-expression, key anti-ferroptosis genes such as *NFE2L2*, *NUPR1* and their target *GPX4*, a well-characterized inhibitor of lipid peroxidation[34] were down-regulated in both lines (Fig. 5a and Fig 3h). On the other hand, Transferrin (*TF*) and transferrin receptor (*TFRC*) regulating iron uptake were both rapidly induced in RMC219 and RMC2C cells (Fig. 4a-b, 5a and Fig. S6c). Following these acute events, at 48 hours we observed increased expression of a subset of genes involved in lipid peroxidation namely *DPP4*, *LOX*, *LPCAT* paralogs and *ACSL4* (Fig. 5a and Fig 3h). These data suggested that SMARCB1 re-expression induced an acute increase in iron uptake followed by increased lipid peroxidation and ferroptosis.

Complementary observations were made from our scRNA-seq dataset where SCENIC showed that RMC tumour cells were characterized by the activation of the NFE2L2/3 regulon a major regulator of ferroptosis (Figs. 2d, 2i)[35,36]. Consequently, expression of *NFE2L2*, *GPX4* and other anti-ferroptosis genes was upregulated in RMC cells from the MVAC-treated tumour compared to TAL cells, whereas many pro-ferroptosis genes were higher expressed in TAL cells (Fig 5b). Similarly, in the naive tumour, *GPX4* and anti-ferroptosis genes were upregulated in RMC compared to TAL cells (Fig. 5b). RMC tumours further showed staining with 4-Hydroxynonenal (4-HNE) antibodies compared to the surrounding stromal cells confirming their propensity to undergo lipid peroxidation (Fig. S7a). Moreover, in agreement with their pre-tumoural phenotype, the RMC-associated TAL3 cells showed up-regulated expression of anti-ferroptosis genes and down-regulated expression of the pro-ferroptosis genes compared to the TAL2 cells. Activation of the MYC and NFE2L2/3 regulons in these cells was therefore accompanied by activation of the ferroptosis resistance program.

We next assessed if SMARCB1 re-expression and increased expression of the lipid peroxidation genes translated into an elevation of lipid ROS assessed using BODIPY-C11-based flow cytometry (Fig. 5c). SMARCB1 re-expression induced a strong increase of lipid ROS in both lines not seen in mCherry control lines. High lipid ROS was associated with increased AnnexinV-positive cells. Importantly, the increase in lipid ROS and in Annexin-V positive cells were both impaired by ferrostatin-1, a selective ferroptosis inhibitor (Fig. 5c). In contrast, SMARCB1 expression did not induce the activated Caspase 3 apoptosis marker unlike Campothecin treatment. To further confirm ferroptotic cell death, we treated RMC cells with the GPX4 inhibitor RSL3. The RMC cells had IC50 values 2-4 times lower than other RT cell lines and more than 100-fold lower than control RPTEC or HEKT cells (Fig. 5d). We additionally assessed the ability of the pan-caspase inhibitor zVAD-fmk or the necroptosis inhibitor necrostatin-1 (nec1) to inhibit SMARCB1 or RSL3-induced cell death. Reduced cell viability after Dox-induced SMARCB1 expression was rescued by ferrostatin-1 and by nec1,

consistent with the previously reported ability of higher concentrations of nec1 to rescue ferroptosis in other tumour cell lines[37,38], but not by zVAD-fmk (Fig. S7b). Similarly, cell viability in presence of RSL3 was also rescued by ferrostatin-1 and Nec1, but not zVAD-fmk (Fig. S7b). Flow cytometry confirmed that RSL3-induced cell death was rescued by high but not low concentrations of Nec1 (Fig. S7c). As ferrostatin does not rescue other forms of death 34, these data support the observation that SMARCB1 expression induced ferroptotic cell death. Moreover, further evidence for ferroptosis came from immunofluorescence (Fig. S6c) showing not only that TFRC was rapidly induced by SMARCB1 re-expression, but also that while it was located in the cytoplasm in most RMC2C cells at 24 hours, there were already some small rounded dying cells where TFRC was relocated to the plasma membrane, hallmarks of ferroptosis[39]. TFRC located to the plasma membrane in almost all RMC219 cells at the same stage consistent with the observation that these cells undergo very rapid ferroptosis. These results confirmed that RMC cells were highly sensitive to GPX4 inhibition and that cell death was due to ferroptosis.

IFNg, secreted by the immune microenvironment in tumours in situ, induces tumour cell dedifferentiation and ferroptotic cell death in melanoma[40,41]. IFNg treatment of RMC219 and RMC2C resulted in durable expression of PDL1, STAT1 and IRF1 and of mesenchymal markers JUN and ZEB1, induced in RMC219 cells and up-regulated in RMC2C cells (Fig. 5e-f). In contrast, NFE2L2 expression was reduced. IFNg treatment induced death of RMC2C cells between 48 and 72 hours, whereas death of RMC219 cells required 72 hours (Fig 5g). Importantly, treatment with ferrostatin 1 diminished the IFNg-induced cell death showing that it involved ferroptosis (Fig. 5h), while as control no induced cell death was seen with HEK293T.

These results revealed that TAL cells were characterized by a ferroptosis sensitivity program that was progressively replaced in pre-tumoural TAL3 cells, in the RMC tumour populations and in RMC cell lines by a NFE2L2 and GPX4-high ferroptosis resistance program. This process was reversed by SMARCB1 re-expression that down-regulated NFE2L2 and GPX4 or by IFNg treatment leading to cell death by ferroptosis.

g. SMARCB1 re-expression promotes genomic SWI/SNF re-localization to enhancers with TFCP2L1 motifs.

To investigate the consequences of SMARCB1 re-expression on SWI/SNF localization and the epigenome of RMC2C cells, we performed BRG1 and H3K27ac ChIP-seq 48 hours after Dox treatment of SMARCB1 or control mCherry expressing cells. RMC219 cells could not be used due to the rapid cell death upon SMARCB1 expression. SMARCB1 re-expression increased the overall number of H3K27ac peaks, but had little impact on their relative genomic distribution with similar fractions of sites at transcription start sites (TSS) and other genomic regions (Fig. S8a-b). However, comparison of read density at more than 46000 non-redundant H3K27ac sites revealed a gain of sites located distal to the TSS following SMARCB1 re-expression (cluster G2, Fig. S8c), whereas only a minor change was seen at the TSS. A fraction of gained peaks were extended regions reminiscent of super-enhancers (SE) known to regulate genes involved in critical aspects of lineage identity or oncogenic transformation[42,43]. While a large number of H3K27ac-marked SEs and their associated genes

were shared between the mCherry and SMARCB1 expressing cells, 240 SE-associated genes were specific to the mCherry line and associated with a variety of functions notably DNA repair and cell cycle (Fig. S8f). More strikingly, 330 SE-associated genes specific to SMARCB1 expressing cells were associated with kidney epithelium development and differentiation as well as cell polarity and junction (Fig. S8g).

SMARCB1 re-expression also modified BRG1 genomic occupancy with a loss mainly at the TSS (H4, Fig. S8d), but a gain at distal sites (H8, Fig. S8d). Integration of BRG1 and H3K27ac read density profiles at more than 40,000 non-redundant co-occupied sites identified those with concomitant gain of H3K27ac and BRG1 following SMARCB1 re-expression (A3, Fig. 6a) predominantly located distal to the TSS (C2). In contrast, cluster A2 defined sites with reduced BRG1 predominantly located at the TSS (A2/B1) with a smaller set at distal sites (C1). Correlation with RNA-seq data indicated that genes associated with cluster A3/C2 showed increased expression following SMARCB1 re-expression (Fig. 6a). RSAT analyses revealed a strong enrichment for TFCP2L1, HOXB9, and MITF binding motifs at the distal gained A2/C3 sites (Fig. 6b). Moreover, ontology analyses of the nearest genes to the A3/C3 sites showed enrichment in differentiation, cell adhesion and kidney epithelium development (Fig. 6c). Enhanced BRG1 recruitment and H3K27ac modification was exemplified by the CDH1 and TJP2 loci where SMARCB1 re-expression led to increased H3K27ac at several putative upstream and intronic enhancers where a strong BRG1 recruitment was also seen (Fig. S8h). An analogous profile was observed at the MITF locus with de novo recruitment of BRG1 and H3K27ac modification at a set of putative enhancer elements located between the promoters of the MITF-A and B isoforms and downstream of the MITF-B TSS (Fig. S8i). SMARCB1 re-expression therefore led not only to re-expression of TFCP2L1 and MITF, but also re-localization of BRG1 to putative H3K27ac marked distal enhancers and super-enhancers associated with the epithelial gene expression program.

We additionally performed Cut&Tag experiments to profile BRG1 and SMARCB1 genomic localization 24 hours after Dox treatment. While no SMARCB1 signal was seen in control mCherry cells, strong SMARCB1 occupancy was seen following its Dox-induced expression (Fig. S9a). At a subset of sites, low BRG1 binding and H3K27ac was seen in absence of SMARCB1 (cluster A1), whereas at the remainder BRG1 and H3K27ac were seen only in presence of SMARCB1 (cluster A2). When SMARCB1-occupancy was examined at the 10 983 distal sites observed at 48 hours (Fig. 6a), de novo recruitment of SMARCB1, BRG1 and marking by H3K27ac was observed at 24 hours (Fig. S9b, clusters B2 and B3). Moreover, in accordance with the strong enrichment for TFCP2L1 binding motifs at these sites (Fig. 6b), TFCP2L1 co-precipitated with BRG1 from extracts of Dox-treated cells (Fig. S9c). Together these results showed that upon its re-expression, SMARCB1 integrated the SWI/SNF complex that interacted with TFCP2L1 and was rapidly recruited to the H3K27ac-marked regulatory elements associated with epithelial genes.

As mentioned above, TFRC was rapidly induced 12 hours after SMARCB1 re-expression. The *TFRC* promoter was strongly marked by H2K27ac in both the mCherry control and 24 hours after SMARCB1 re-expression (Fig. S9d). Moreover, BRG1 and SMARCB1 were recruited at 24 hours. TFRC therefore behaved as an ‘immediate-early’ gene whose promoter was pre-marked with H3K27ac, but whose activation was associated with rapid BRG1 and SMARCB1 recruitment. This contrasts with epithelial program genes whose

activation was slower and where both BRG1/SMARCB1 recruitment and H3K27ac modification occurred de novo.

h. SMARCB1 re-expression remodels MYC genomic binding.

It has been reported that SMARCB1 interacts directly with MYC to antagonize its DNA binding and genomic occupancy in RT cells[44,45]. To address this in RMC cells, we performed MYC ChIP-seq in SMARCB1-expressing and mCherry control cells 48 hours after Dox addition. We identified 54,786 non-redundant MYC sites, a much larger number than previously observed 44. All MYC-bound sites in G401 RT cells, that were predominantly located close to the TSS, were occupied also in RMC2C cells (Fig. S10a-b). For example, MYC sites commonly bound in G401, RMC2C2 and in the Hela ENCODE data sets were observed at the *NCL* and *CDK4* loci (Fig. S10b).

In keeping with reduced MYC expression, around 50% fewer peaks were observed in SMARCB1 expressing cells where its occupancy was remodelled with a relative re-localization to the TSS that increased from 24% to 41% of the detected peaks (Fig. S8a-b). Read density profiles at the non-redundant MYC sites identified those with gained (I2/I8, Fig. S8e) or diminished (I3/I9) occupancy located at both TSS proximal and distal regions. Notably, integration with BRG1 and H3K27ac datasets revealed that MYC occupancy was increased at TSS proximal sites marked by H3K27ac, but characterized by diminished BRG1 occupancy (D1/E1, Fig.6d). In contrast, a large set of distal located sites were lost upon SMARCB1 re-expression (D4/F3) with a smaller number showing increased occupancy (D3/F2). Global analyses confirmed that BRG1 flanking a subset of MYC bound sites in the mCherry control cells was diminished following SMARCB1 re-expression, whereas H3K27ac was unchanged (Fig. S8c). RSAT analysis of the top 1000 MYC peaks confirmed a strong enrichment of the cognate E-box motif (Fig. 6e). MYC-binding motifs were also strongly enriched at the D1-D4 sub-clusters, together with MAZ at D1 sites, whose activity was associated with TAL transformation (Figs. 2g-i and S10d).

As shown above, the term ‘MYC targets’ was a prominent hallmark of genes down-regulated by SMARCB1 re-expression. We determined the % of genes in the GSEA hallmark gene sets overlapping with those associated with each MYC sub-cluster. Genes associated with D1 sites were strongly enriched in MYC targets, mitotic spindle, mTOR, E2F. DNA repair and G2M hallmark signatures (Fig. S11a). Genes associated with D4 also displayed a similar, yet lower, enrichment in many of these pathways. Correspondingly, genes associated with D1 and D4 showed global down-regulation (Fig. S11b), whereas those associated with D2 and D3 showed up-regulated expression. Thus, many genes associated with oncogenic transformation and down-regulated by SMARCB1 re-expression were associated with a gain of promoter-proximal MYC, but strongly reduced BRG1 binding.

A similar analysis of BRG1 sub-clusters, showed genes associated with A2 were strongly enriched in the above oncogenic-associated hallmarks (Fig. S11c). In contrast, A3 sites with strongly gained BRG1 binding were enriched in genes associated with apical junction/surface and kidney morphogenesis hallmarks, consistent with re-activation of an epithelium program. We used ROSE to identify MYC-H3K27ac-marked or BRG1-H3K27ac-

marked SEs in control and SMARCB1-expressing cells (Fig. S11d-e). The ontology of the SE-associated genes was consistent with a switch from MYC/BRG1 driving proliferation and oncogenesis in absence of SMARCB1 to TFCP2L1/BRG1 driving an epithelium program in presence of SMARCB1.

To better understand the paradoxical observation that MYC binding increases at down-regulated oncogenic genes, we looked more closely at the large set of diminished D4 sites associated with similar ontology terms to D1. Re-clustering of D4 identified a small number (J1, Fig. S11f) of promoter-proximal sites associated with H3K27ac and a large majority of distal sites (J2, Fig. S11f). Strikingly, a large number of genes were commonly associated with both clusters (Fig. S11g). Hence many genes of the oncogenic program had both promoter-proximal and distal MYC sites showing increased and decreased occupancy, respectively. Importantly, the D4 sites were enriched in binding motifs for HIF1A and SNAI1 in agreement with coordinate activation of MYC, HIF1A and EMT programs in RMC. Loss of MYC at the D4 sites upon SMARCB1 expression was therefore consistent with their role in driving transformation.

Overall, these results showed that SMARCB1 re-expression did not repress MYC genomic occupancy, but rather remodelled its binding profile in a manner suggesting that altered enhancer-promoter communications and loss of promoter-proximal BRG1 binding underlie reduced expression of the proliferation/oncogenic program.

D. Discussion.

a. Oncogenic transformation of TAL cells into RMC.

Here we integrate transcriptomic data from RMC patients with gain and loss of SMARCB1 function in cell-based models to decipher the mechanism of a transcriptional switch driving oncogenic transformation and ferroptosis resistance of TAL epithelial cells.

ScRNA-seq analyses of RMC cells compared to NAT identified TAL cells as RMC cell-of-origin. TAL cells were marked by strong activity of TFCP2L1, HOXB9 and MITF transcription factors associated with the epithelial expression program. TAL transformation was characterized by loss of expression and activity of these factors, but gain of MYC and NFE2L2 that drive proliferation and ferroptosis resistance. Further evidence for this series of oncogenic events came from the fortuitous capture of tumour-associated TAL2/3 cells that displayed a pre-transformed state retaining TFCP2L1 activity, while at the same time showing MYC and YY1 activity accompanied by a hypoxia and stress signature.

TAL transformation generated an epithelial-mesenchymal gradient of RMC tumour cells that was reproduced by RMC219, UOK360, UOK353 and RMC2C cell lines. Mesenchymal-like cells were observed in the treated tumour and the mesenchymal transcriptional signature was present in primary tumours from naive patients and was predominant in the lymph nodes. Thus, de-differentiation into this mesenchymal state is not specific to drug-treated tumours, but appears to be an intrinsic feature of RMC tumours that likely contributes to their metastatic spread

SMARCB1 re-expression in RMC2C cells provided experimental mechanistic support for the above model of TAL-RMC transformation. SMARCB1 expression reactivated *TFCP2L1*, *HOXB9* and *MITF* expression and promoted BRG1 re-localization to enhancers and super-enhancers driving expression of an epithelial expression program that were de novo marked by H3K27ac and enriched in binding motifs for these factors (Fig. 6f). The lack of ChIP-grade TFCP2L1 and MITF antibodies did not allow us to directly confirm their presence at these enhancers. However, we previously showed that MITF interacts with SWI/SNF and actively recruits BRG1 to melanoma-cell promoters and enhancers[46] and here we showed that TFCP2L1 also co-precipitated with SWI/SNF. In contrast, SMARCB1 re-expression led to reduced levels of MYC and NFE2L2. Genomic profiling revealed a remodelling of MYC genomic binding with sites showing both gained or reduced occupancy. Paradoxically, while MYC binding increased at the proximal promoters of genes involved in oncogenesis, it was lost at sites distal to these genes. Although there are clear limitations in assigning distal binding sites to regulation of a given gene, a large set of genes showed increased MYC binding at the promoter and diminished binding at distal sites suggesting the importance of enhancer-promoter communication in their activation. More importantly however, BRG1 occupancy was strongly reduced at these promoters showing that MYC cooperated with SWI/SNF lacking SMARCB1 to activate the oncogenic program and that BRG1 eviction and not MYC loss reversed the oncogenic process.

Integrating patient and in cellulo-derived data converged to show that pre-tumoral TAL2/3 cells displayed a hypoxia/stress state activating MYC and NFE2L2 to drive ferroptosis resistance allowing survival under conditions favourable to SMARCB1 loss (Fig. 6f). Subsequently, SMARCB1 loss led to BRG1 recruitment at promoters of MYC occupied oncogenic genes and inhibition of the *TFCP2L1*/*HOXB9*/*MITF*-driven TAL epithelial program. In RMC cells, SWI/SNF lacking SMARCB1 cooperates with MYC to drive the oncogenic program, whereas SMARCB1-containing SWI/SNF is evicted from MYC-driven oncogenic promoters and re-located to enhancers driving the TAL epithelial program.

b. Distinct cells-of-origin and oncogenic mechanisms in RMC and RT.

The above observations highlight major differences with previous studies on RT cells. In G401 RT cells, SMARCB1 antagonized MYC DNA binding and chromatin occupancy[44]. In contrast, in RMC cells, antagonism translated not as a loss of MYC binding, but eviction of SMARCB1-containing SWI/SNF from MYC occupied promoters and reduced oncogenic gene expression. It has been reported that BRD9-containing non-canonical (nc)BAF plays a critical role in driving the oncogenic state in SMARCB1-deficient RT[47,48]. NcBAF strongly colocalized with CTCF although other enriched transcription factor motifs were also identified. However, MYC was not amongst the strongly enriched motifs in either study. This leads to the paradoxical observation that despite the essential role of BRD9/ncBAF in driving the transformed state, it is not enriched at MYC-bound sites, whereas depletion studies in RT cells[44] and our current data revealed MYC as the essential oncogenic driver. Given the strong association of ncBAF with CTCF and not MYC, it is unlikely that the BRG1 seen at the MYC promoters in RMC cells corresponds to ncBAF. Moreover, CTCF motifs were not enriched at

the promoter sites where BRG1 was evicted, but were present at distal sites, where no BRG1 was detected. Thus, the role of ncBAF in RMC remains to be determined.

In RT cells, SMARCB1 re-expression led to SWI/SNF re-localization to what have been described as lineage-specific enhancers[49,50]. However, the transcription factor motifs at SMARCB1-bound enhancers were not always informative as to the nature of the cell of origin. The lack of a clearly defined cell(s) of origin, and their intrinsic biology has hampered a detailed understanding of the transformation process. Mechanistic studies were often limited to SMARCB1 re-expression in RT cell lines with little supporting patient data. An exception is atypical teratoid RT (AT/RT) where the epigenetic profiles of the AT/RT tumours were compared with other types of brain tumours or normal brain[51]. Nevertheless, the validity of these comparisons is limited since more recent data provide strong evidence that RT arise following arrest of neural crest cell differentiation into mesenchyme, in particular Schwann cells[52]. Many of the above limitations have been overcome in our study, where patient-derived and functional in cellulo data converged to define the transcriptional program of the TAL cell of origin and to decipher the mechanistic details of a reversible transcriptional switch driving their transformation into epithelial- and mesenchymal-type RMC states. We thus highlight the fundamental difference between RMC arising from mesoderm-derived differentiated epithelial TAL cells and RT derived from differentiating neural crest cells.

c. A link between RMC ferroptosis and sickle cell trait.

A key finding of our study is activation of a ferroptosis resistance pathway in RMC cells. Analyses of gene expression signatures in scRNA-seq, patient cohort RNA-seq and the RMC cell lines defined how the ferroptosis sensitivity signature in TAL cells is replaced by a ferroptosis resistance signature in RMC cells. This process is reversed in RMC cells upon SMARCB1 re-expression leading to their ferroptotic cell death unlike other RT cells where SMARCB1 re-expression leads to cell cycle arrest or apoptosis[49,50,53]. Indeed, RMC cells are more sensitive to GPX4 inhibition than RT lines. Ferroptosis is therefore a specific vulnerability of RMC tumours.

The above observations link the RMC oncogenic process with sickle cell trait. The kidney medulla is amongst the most hypoxic micro-environments in the organism[54]. Due to its central role in urine concentration, the loop of Henle is characterized by increasing osmolarity and hypoxia that are highest in the TAL region. Msaouel et al. proposed a model where the high interstitial NaCl concentration induces DNA double strand breaks (DSB), whereas microcirculatory ischemia induced by red blood cell (RBC) sickling reduces this osmolarity reactivating DSB repair in a chronic hypoxic environment by NEHJ favoring translocations and deletions, particularly in fragile regions such as chromosome 22q where the *SMARCB1* locus is located[55].

Our observations enrich this model with iron release by RBC sickling favouring ferroptosis of TAL cells and their renewal to maintain the homeostasis of the epithelium[56,57]. Early initiation of ferroptosis resistance observed in the pre-tumoural TAL cells would thus promote their survival under the high NaCl and hypoxic conditions driving error-prone DSB repair. The increased extracellular iron concentration due to the fragility of the sickled RBCs

acts as a selective pressure for survival of ferroptosis resistant TAL cells in an environment propitious to the mutagenic events associated with RMC development. This unique set of circumstances may explain why RMC is the only SMARCB1-deficient tumour arising from epithelial cells, compared to RTs arising from a developmental block of neural crest differentiation.

E. Methods.

Tumour Samples

The two RMC samples subjected to scRNAseq were collected from Strasbourg University Hospital and Curie Institute, according to institutional guidelines. Sample collection for further research analysis was approved ethical Committees of participating institutions and all patients provided an informed written consent for the use of material for further research. Regarding bulk RNAseq, beside the RNAseq of 11 patients recently reported 6, we generated an additional dataset of multi-region RNAseq of a cohort of 4 RMC patients, including multiple sections and lymph nodes metastasis (Dataset S2).

Human single-cell sample preparation and RNA-seq

Following the treated tumour resection, samples from the tumour and adjacent non-malignant normal adjacent tissue were each conserved at 4°C in 1mL of MACS Tissue Storage Solution (Miltenyi Biotech). Single cell suspensions were prepared using gentleMACSTM dissociator and human tumour dissociation kit (Miltenyi Biotech) following manufacturer's instructions. Samples were applied to a MACS SmartStrainer 70µm (Miltenyi Biotech) placed on a 15mL Falcon tube and 10mL DMEM were used to wash C tube and SmartStrainer. Following centrifugation at 300g and 4°C for 10min, cells were sorted using CD45 (TIL) Microbeads (Miltenyi Biotech). CD45+ and CD45- fractions were centrifuged (300g, 10min, 4°C) and dead cells were removed using Dead cell removal kit (Miltenyi Biotech). CD45- and CD45+ were mixed in 1 to 4 ratios. Cell viability and concentration were assessed before 3'-mRNA single-cell libraries were prepared using the Chromium (10x Genomics) following the manufacturer's instructions. Libraries were sequenced 2x100bp on HiSeq4000 sequencer.

Folowing resection of the naive tumour, the sample was cut in small pieces then dissociated 30 min at 37°C in CO2-independent medium (Gibco) + 0,4 g/l of human albumin (Vialebex) with Liberase TL (Roche) 150 ug/ml and DNase 1 (Sigma) 150 ug/ml. Dissociated cells were then filtered with a 40 mm cell strainer, then washed and resuspended with CO2-independent medium + 0,4 g/l of human albumin. A fraction of the cell suspension was used to enrich tumor cells using Tumor isolation kit (Miltenyi Biotech, cat#130-108-339). Cells were then resuspended at 800 cells/ul in PBS + BSA 0,04%. Tissues were processed within 1 hour after tumor resection, and sorted cells were loaded in a 10x Chromium instrument within 6 hours.

Patient-derived xenograft sample preparation

Renal medullary carcinoma (RMC) patient derived xenograft (IC-pPDX-132) was established from a resected RMC tumour treated with 6 cycles of cisplatin, gemcitabine and bevacizumab. The undissociated tumor was engrafted in the subscapular fat pad of NSG

(NOD.Cg-Prkdcscid IL2rgtm1Wjl/SzJ) mice. A PDX tumor fragment was then serially transplanted using the same procedure into Swiss Nude (Crl:NU(Ico)-Foxn1nu) mice until passage 4 which was used for the single cell RNA-seq experiments. Animal care and use for this study were performed in accordance with the recommendations of the European Community (2010/63/UE) for the care and use of laboratory animals. The housing conditions were specific pathogen free (SPF) for all models. Experimental procedures were specifically approved by the ethics committee of the Institut Curie CEEA-IC #118 (Authorization APAFIS#11206-2017090816044613-v2 given by National Authority) in compliance with the international guidelines. The establishment of PDX received approval by the Institut Curie institutional review board OBS170323 CPP ref 3272; n de dossier 2015- A00464-45). Written institutional informed consent was obtained from the patient.

scRNA-seq analysis of human primary RMC tumours

After sequencing, raw reads were processed using CellRanger (v 3.1) to align on the hg19 human genome, remove unexpressed genes and quantify barcodes and UMIs. Data were then analysed in R (v4.0.2). For the treated tumour, tumour and NAT samples were aggregated with the cellranger 'aggr' command. The resulting aggregation was analysed with Seurat v3.2.0 following the recommended workflow. Cells were filtered for feature count ranging from 120 to 2000 and percentage of mitochondrial reads <15%. Counts were normalized with the "LogNormalize" method and data scaled to remove unwanted sources of variation (UMI count and mitochondrial reads). The number of principal components was determined from the Jackstraw plots. Clustering was performed on variable features using the 25 most significant principal components and a resolution of 1.15. For the naive tumour, the same Seurat pipeline was performed using feature counts from 200 to 6000, mitochondrial read fraction <20% and a resolution of 1.0 using the 20 most significant principal component for the clustering. Aggregate analyses of tumours 1 and 2 was performed by merging the two R objects and using the Seurat sctransform with batch correction function to normalize and scale data reducing the impact of technical factors.

scRNA-seq analysis of patient-derived RMC xenograft

For the IC-pPDX-132 sample raw reads were aligned on an hg19-mm10 hybrid genome. Cells were filtered based on feature counts ranging from 200 to 7000 and global clustering performed with a resolution of 0.3 using the 20 most significant components. Human and Mouse cells were re-clustered separately by first filtering cells with mitochondrial read fraction >20% and then using a resolution of 0.4 with 25 principal components.

Functional analysis using scRNA-seq data

Regulome analyses of active transcription factors were performed using the SCENIC v1.1.2.2 package[17]. Transcription factor activities were visualized on the UMAP using AUCell or as heatmaps using the R-package 'pheatmap'. RMC correlations with the different renal tubule clusters were computed by Clustifyr v1.0.0[58] using cluster marker signatures for RMC (*TIMP1, FN1, CTHRC1, DCBLD2, COL1A2, COL1A1, ARL4C, COL6A2, LGALS1, CD44, VIM, CLU, MMP7, SERPINA1, WFDC2, SFRP2, MUC1, KRT18, KRT7, EPCAM, CDH1, CLDN4, CLDN10, DEFB1*), RMC1 (*WFDC2, FXYD2, SLPI, CLDN4, KRT7, KLF6, GSTP1, EEF1A1, CLDN3, TM4SF1*) and RMC2 (*FN1, COL1A2, COL1A1, TIMP1, CD44, CTHRC1, RARRES3, BGN, TFPI2, COL6A2*). Trajectory analyses were plotted and visualized using

Similarity Weighted Nonnegative Embedding (SWNE)[59]. Gene set variation analysis were performed using the r-package GSVA[60].

For the "bulk RMC signature", the upregulated genes from the differential analysis of the MDACC RMC cohort (11 tumours versus 6 NAT) were selected using log2FC > 2 and FDR < 0.01 6. For all signatures, gene sets were retrieved from either Hallmarks MSigDB or KEGG pathways. Gene signatures were computed and visualized on UMAPs using the R package VISION (https://github.com/YosefLab/VISION).

Anti-4 Hydroxynonenal staining of RMC tumours.

Sections from 2 independent RMC tumours and as control a colorectal cancer were fixed in 10% neutral-buffered formalin, paraffin embedded, sectioned, and stained with hematoxylin and eosin. 4-µm tissue sections were processed on VENTANA-Benchmark-XT, with incubation at room temperature in an antigen retrieval process (EDTA citrate buffer, pH 8,3, CC1 buffer, 8 min), then incubated with 4HNE (Anti-4 Hydroxynonenal antibody, mouse monoclonal, clone HNEJ-2, Abcam; dilution: 1/5000 during 32 min), revealed with 'Ultra View' Universal DAB Detection kit and counterstained with Hematoxylin solution (Ventana Roche Systems).

Cell culture, establishment of RMC lines stably expressing SMARCB1

RMC219 cells were grown in HAM-F12/D-MEM (1:1) medium supplemented with 10% foetal calf serum (FCS), Glutamine 2mM, AANE and PS. UOK360 and UOK353 cells were grown in D-MEM medium supplemented with 10% foetal calf serum (FCS) and Glutamine 2mM. RMC2C cells were grown in MEM medium with 10% FCS, AANE, 50ng/mL EGF and PS. RMC cells infected with lentiviral constructs were grown in respective media replacing normal FCS with tetracyclin-free FCS (Dutscher) and supplemented with G418 (300ug/mL). SMARCB1 expression was induced by treatment with either DMSO or 2µM of doxycycline.

Lentiviral pInducer20 vector was obtained from Addgene and the cDNA of either SMARCB1 or mCherry was cloned into the vector by Gateway. We then used pInducer20-mCherry or -SMARCB1 containing lentiviruses to infect 1x106 RMC2C or RMC219 cells. Cells were selected using 500ug/mL G418 for a week and then maintained under these conditions.

In vitro treatments

For ferroptosis, cells were either treated with DMSO or 2uM doxycycline alone or co-treated with 2uM doxycycline and 1uM ferrostatin-1 (SelleckChem, #S7243), zVAD-fmk (MedChemExpress, #HY-16658B) or necrostatin-1 (MedChemExpress, #HY-15760) for the indicated times. For the Caspase-3 assays, cells were either treated with 5uM camptothecin (SelleckChem, #S1288) for 4hr, DMSO or 2uM doxycycline for the indicated times. For the IFNg experiments, cells were either treated with DMSO or 10ng/mL of IFNg (Peprotech, 300-02).

Cell death, caspase-3 and lipid peroxidation analyses by flow cytometry

Cells were harvested at the indicated times and co-stained with Annexin-V-FITC and propidium iodide following manufacturer instructions (BioLegend, #640914). To assess active

Caspase-3, cells were fixed and permeabilized before incubation with the FITC-conjugated caspase-3 antibody following manufacturer's instructions for subsequent flow cytometry analysis (Abcam, #65613). To assess membrane lipid perodixation, cells were stained using 10uM of Bodipy 581/591 C11 (ThermoFisher, #D3861) following manufacturer's instructions. To assess senescence, cells were treated with 100nM bafilomycin A1 (Sigma, #19-148) for 1hr followed by 2mM C12FDG (Invitrogen, #D2893) for 2hrs before being washed and harvested for flow cytometry analyses. All assays were analysed on a FACS Analyzer Fortessa (BD Biosciences) and data were analysed using Flowjo v6.8.

Immunofluorescence.

Cells grown on glass slides in 24-well plates, were fixed with 4% paraformaldehyde for 15 minutes. After two washes with PBS buffer they were permeabilized in PBS+triton X-100 0,1% for 5 minutes and blocked with PBS+10% FCS inactivated for 20 minutes. Primary antibodies were incubated overnight at 4°C and after three washes with PBS+Triton 0,1%, cells were stained for 1 hour at room temperature with AlexaFluor-488 conjugated secondary antibodies (Life technologies) diluted 1/500 in PBS+10% FCS. After three washes with PBS+Triton 0,1%, cells were stained with DAPI (final concentration 1 ug/ml) and mounted on microscopy slides. Images were captured with a confocal (Leica DMI6000) microscope. The references of all antibodies are available in Table 1.

Cell viability assay by fluorescence screening

5 x 103 of indicated cell types were seeded on 96-well plates in four technical replicates on day 1. The next day, cells were treated either with DMSO control or with an increasing concentration of RSL3 (SelleckChem, 8155) ranging from 0 to 10μM. At day 3, cells were washed with PBS and stained using PrestoBlue (Invitrogen, A13261) according to manufacturer instructions before fluorescence was quantified on a multi-modal spectrometer (Berthold Mithras, LB940). IC50 values were calculated using the fraction of DMSO control.

Immunostaining quantification by flow cytometry

Wildtype RMC219 and RMC2C cells were harvested and 1 x 106 cells were resuspended in buffer A (PBS 1X, EDTA 2mM, inactivated FCS 1%) and 5uL of Human TruStain FcX (Biolegend, 422301) was added for 10 min at room temperature. Following blocking, cells were stained for 1hr with 5μL of conjugated EPCAM-FITC (Biolegend, 324203, dilution 1:100) and conjugated CD44-PE (Biolegend, 103023, dilution 1:100). Following two PBS washes, cells were resuspended in buffer A before flow cytometry on a FACS analyzer Fortessa (BD Biosciences) and analysis using Flowjo v6.8.

Boyden Chamber Invasion assays

Before seeding, 100ul of diluted Matrigel (1:20, 356234, Corning) was added in each insert (24-well 8um inserts, Corning) and left to dry for 2hrs at 37°C before being washed twice with PBS. Subsequently, RMC cells were harvested and 2 x 105 cells and seeded in the Boyden chambers in corresponding media without serum. 24hrs later, migrated cells were fixed using PFA 4% for 10 min before being stained using Crystal violet for 10 min. Excess stain was washed 3 times in PBS before images were captured on phase contrast microscope. Quantification of migrated cells was done by resuspension of staining using 100mM acetic acid

for 15min before absorbance was measured on a BioTek microplate reader (using Gen5 software).

RNA preparation and quantitative PCR

RNA isolation was performed according to standard procedures (Macherey Nagel RNA Plus kit). RT-qPCR was carried out with SYBR Green I (Roche) and SuperScript IV Reverse Transcriptase (Invitrogen) and monitored using a LightCycler 480 (Roche). The mean of ACTB, TBP, RPL13A and GAPDH gene expressions was used to normalize the results. Primer sequences for each cDNA were designed using Primer3 Software and are available in Table 2.

Public data correlation analysis using TGCA and CCLE database

Spearman correlation for all selected genes were retrieved from co-expression studies using the Cancer Cell Line Encyclopaedia (Broad, 2019) and the TCGA chromophobe renal cell carcinoma (KICH) databases. All transcription factors were extracted using the "Full Human TFs" list from 61. Scatter plots were made using Prism5. For the correlation with TFCP2L1, the epithelial and mesenchymal genes were retrieved from 23.

Bulk RNA sequencing

RMC cell lines were analysed by RNA-seq under the different indicated conditions. After sequencing raw reads were pre-processed in order to remove adapter and low-quality sequences (Phred quality score below 20) using cutadapt version 1.10. and reads shorter than 40 bases were discarded. Reads were maping to rRNA sequences using bowtie version 2.2.8, were also removed. Reads were mapped onto the hg19 assembly of Homo sapiens genome using STAR version 2.5.3a. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.6.1p1, with annotations from Ensembl version 75 and "union" mode. Only non-ambiguously assigned reads were retained for further analyses. Read counts were normalized across samples with the median-of-ratios method. Comparisons of interest were performed using the Wald test for differential expression and implemented in the Bioconductor package DESeq2 version 1.16.1. Genes with high Cook's distance were filtered out and independent filtering based on the mean of normalized counts was performed. P-values were adjusted for multiple testing using the Benjamini and Hochberg method. Deregulated genes were defined as genes with log2(foldchange) >1 or <-1 and adjusted p-value <0.05.

Analysis of bulk RNA-seq of patient samples

For RMC cohorts, raw counts were retrieved in excel format and normalized first by sequencing depth using DESeq2 sizefactors and then divided by median of gene length. Samples were clustered using the hclust function with "ward.D2" linkage function and visualized as heatmaps using pheatmap package v1.0.12. The deconvolution of immune and stromal cells was done using MCP-counter v1.2.0[62]. Sample compositions were also estimated by deconvolution from our single-cell data using the CIBERSORTx algorithm[63]. Volcano plots were generated with ggplot2 v3.3.2. Gene set enrichment analyses were done with the GSEA software v3.0 using the hallmark gene sets of Molecular Signature Database v6.2. Gene Ontology analysis was done using DAVID (http://david.abcc.ncifcrf.gov/). Gene list intersections and Venn diagrams were performed by Venny.

Protein extraction and Western blotting

Whole cell extracts were prepared by the standard freeze-thaw technique using LSDB 500 buffer (500 mM KCl, 25 mM Tris at pH 7.9, 10% glycerol (v/v), 0.05% NP-40 (v/v), 16mM DTT, and protease inhibitor cocktail). Cell lysates were subjected to SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and proteins were transferred onto a nitrocellulose membrane. Membranes were incubated with primary antibodies in 5% dry fat milk and 0.01% Tween-20 overnight at 4 °C. The membrane was then incubated with HRP-conjugated secondary antibody (Jackson ImmunoResearch) for 1h at room temperature, and visualized using the ECL detection system (GE Healthcare). The references of all antibodies are available in Table 1. All unprocessed data are available in the Source Data file.

Chromatin immunoprecipitation and sequencing (ChIP-seq)

BRG1 ChIP experiments were performed on native MNase-digested chromatin. Between 10 to 20 × 108 freshly harvested RMC2C cells bearing either SMARCB1 or mCHERRY and treated 2uM doxycycline for 48hrs were resuspended in 1.5 ml ice-cold hypotonic buffer (0.3M Sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris–HCl pH 7.5, 0.5 mM DTT, 0.1 mM PMSF, PIC) and cytoplasmic fraction was released by incubation with 1.5 ml of lysis-buffer (0.3M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris–HCl pH 7.5, 0.5 mM DTT, 0.1 mM PMSF, PIC, 0.5% (vol/vol) IGEPAL CA-630) for 10 min on ice. The suspension was layered onto a sucrose cushion (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris–HCl [pH 7.5], 0.5 mM DTT, 0.1 mM PMSF, PIC) and centrifuged for 30 min 4°C at 4700 rpm in a swing rotor. The nuclear pellet was resuspended in digestion buffer (0.32Msucrose, 50 mM Tris–HCl [pH 7.5], 4 mM MgCl2, 1 mM CaCl2, 0.1 mM PMSF) and incubated with 10ul of Micrococcal Nuclease (NEB) for 7 min at 37˚C. The reaction was stopped by addition of 20ul of EDTA 0,5M and suspension chilled on ice for 10 min. The suspension was cleared by centrifugation at 10,000 rpm (4˚C) for 10 min and supernatant (chromatin) was used for further purposes. Chromatin was digested to around 80% of mono-nucleosomes as judged by extraction of the DNA and agarose gel electrophoresis. H3K27ac and MYC ChIP experiments were performed on 0.4% PFA-fixed chromatin isolated from RMC2C cells bearing either SMARCB1 or mCHERRY and treated 2uM doxycycline for 48hrs according to standard protocols as previously described 64. ChIP-seq libraries were prepared using MicroPlex Library Preparation kit v2 and sequenced on the Illumina Hi-seq 4000 as single-end 50-base reads[65]. Sequenced reads were mapped to the Homo sapiens genome assembly hg19 using Bowtie with the following arguments: -m 1 --strata --best -y -S -l 40 -p 2. Cut&Tag was performed using the Active Motif CUT&Tag-IT kit following the manufacturer's instructions. The references of all antibodies are available in Table 1.

ChIP-seq analysis

After sequencing, peak detection was performed using the MACS software (Zhang et al., 2008). Peaks were annotated with Homer using the GTF from ENSEMBL v75. Global clustering analysis and quantitative comparisons were performed using seqMINER[66]. Super-enhancers were called with the python package Ranking Of Super Enhancers (ROSE) (https://github.com/stjude/ROSE).

De novo motif discovery on FASTA sequences corresponding to windowed peaks was performed using MEME suite (meme-suite.org). Motif correlation matrix was calculated with in-house algorithms using JASPAR database as described in 67. Motif discovery from ChIP-seq peaks was performed using the RSAT peak-motifs algorithm (http://rsat.sb-roscoff.fr/peak-motifs_form.cgi).

Motif analysis Searching of known TF motifs from the Jaspar 2014 motif database at BRG1-bound sites was made using FIMO 68 within regions of 200 bp around peak summits, FIMO results were then processed by a custom Perl script which computed the frequency of occurrence of each motif. To assess the enrichment of motifs within the regions of interest, the same analysis was done 100 times on randomly selected regions of the same length as the BRG1 bound regions and the results used to compute an expected distribution of motif occurrence. The significance of the motif occurrence at the BRG1-occupied regions was estimated through the computation of a Z-score (z) with $z = (x - \mu)/\sigma$, where: − x is the observed value (number of motif occurrence), − μ is the mean of the number of occurrences (computed on randomly selected data), − σ is the standard deviation of the number of occurrences of motifs (computed on randomly selected data). The source code is accessible at https:// github.com/slegras/motif-search-significance.git.

Statistics and reproducibility

All experiments were performed in three independent biological replicates, unless stated otherwise in the figure legends. All tests used for statistical significance were calculated using Prism5 and are directly indicated in the figures. Note that the following significance values have been used throughout the manuscript: ****, $p < 0.0001$; ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; ns, $p > 0,05$.

Data availability

Source data for this paper are available from the authors upon reasonable request. All sequencing data reported here have been submitted to the GEO database under accession number GSE181001.

Acknowledgements

We thank all the staff of the IGBMC common facilities in particular Betty Heller and Patricia Wagner from Cell Culture, Claudine Ebel and Muriel Philipps from Flow Cytometry and Dr. Paola Rossolillo and Karim Essabri of the molecular biology facility. We also would like to thank the Institut Curie facilities and in particular Dr Pascale Philippe-Chomette, Pr Michel Peuchmaur, Dr Yves Allory and Dr Pascale Maille for proving the primary specimen for the PDX experiments.

This work was supported by institutional grants from the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Université de Strasbourg, the Association pour la Recherche contre le Cancer (CR, contract number PJA 20181208268), the Ligue Nationale contre le Cancer, the Institut National du Cancer (PLBIO-2021-079), the ANR-10-LABX-0030-INRT French state fund through the Agence Nationale de la Recherche under the frame programme Investissements d'Avenir labelled ANR-10-IDEX-0002-02.

The IGBMC high throughput sequencing facility is a member of the "France Génomique" consortium (ANR10-INBS-09-08). FB was supported by grants from the INCA for the scRNAseq program (INCA PRTK19-2020-036) and the Saint Bladrick Foundation. ID is an 'équipe labellisée' of the Ligue Nationale contre le Cancer. BHV was supported by fellowships from the ANR, the Ligue Nationale contre le Cancer.

Conceptualization, BHV, ID, GGM; Methodology, BHV, GD, AH, MR, ARH, JG, JT, PB, PM, XS, DS, RB, JEK, FB, NMT, ID, GGM; Formal Analysis, BHV, GD, ID, GGM. Investigation, BHV, GD, AH, MR, ARH, JG, JT, PB, PM, XS, HL, TT, VL, DS, RB, JEK, FB, NMT, ID, GGM. Bioinformatics: GD, BHV, XS; Writing – Original Draft, BHV, ID, GGM; Writing – Review and Editing, BHV, GD, ID, GGM. Investigation, BHV, GD, AH, MR, ARH, JG, JT, PB, PM, XS, HL, TT, VL, DS, RB, JEK, FB, NMT, ID, GGM; Resources, PM, XS, HL, TT, VL, DS, RB, FB, NMT, ID and GGM; Supervision: ID, GGM

F. Bibliography

1. Davis, C. J., Mostofi, F. K. & Sesterhenn, I. A. Renal medullary carcinoma. The seventh sickle cell nephropathy. Am. J. Surg. Pathol. 19, 1–11 (1995).

2. Cajaiba, M. M. et al. The classification of pediatric and young adult renal cell carcinomas registered on the children's oncology group (COG) protocol AREN03B2 after focused genetic testing. Cancer 124, 3381–3389 (2018).

3. Alvarez, O., Rodriguez, M. M., Jordan, L. & Sarnaik, S. Renal medullary carcinoma and sickle cell trait: A systematic review. Pediatr. Blood Cancer 62, 1694–1699 (2015).

4. Msaouel, P., Carugo, A. & Genovese, G. Targeting proteostasis and autophagy in SMARCB1-deficient malignancies: where next? Oncotarget 10, 3979–3981 (2019).

5. Carugo, A. et al. p53 Is a Master Regulator of Proteostasis in SMARCB1-Deficient Malignant Rhabdoid Tumors. Cancer Cell 35, 204-220.e9 (2019).

6. Msaouel, P. et al. Comprehensive Molecular Characterization Identifies Distinct Genomic and Immune Hallmarks of Renal Medullary Carcinoma. Cancer Cell 37, 720-734.e13 (2020).

7. Msaouel, P., Walker, C. L., Genovese, G. & Tannir, N. M. Molecular hallmarks of renal medullary carcinoma: more to c-MYC than meets the eye. Mol. Cell. Oncol. 7, 1777060 (2020).

8. Swartz, M. A. et al. Renal medullary carcinoma: clinical, pathologic, immunohistochemical, and genetic analysis with pathogenetic implications. Urology 60, 1083–1089 (2002).

9. Gupta, R. et al. Carcinoma of the collecting ducts of Bellini and renal medullary carcinoma: clinicopathologic analysis of 52 cases of rare aggressive subtypes of renal cell carcinoma with a focus on their interrelationship. Am. J. Surg. Pathol. 36, 1265–1278 (2012).

10. Beckermann, K. E. et al. Renal Medullary Carcinoma: Establishing Standards in Practice. J. Oncol. Pract. 13, 414–421 (2017).

11. Dimashkieh, H., Choe, J. & Mutema, G. Renal medullary carcinoma: a report of 2 cases and review of the literature. Arch. Pathol. Lab. Med. 127, e135-138 (2003).

12. Elliott, A. & Bruner, E. Renal Medullary Carcinoma. Arch. Pathol. Lab. Med. 143, 1556–1561 (2019).

13. Cheval, L., Pierrat, F., Rajerison, R., Piquemal, D. & Doucet, A. Of mice and men: divergence of gene expression patterns in kidney. PloS One 7, e46876 (2012).

14. Lake, B. B. et al. A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. Nat. Commun. 10, 2832 (2019).

15. Muto, Y. et al. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. Nat. Commun. 12, 2190 (2021).

16. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science 361, 594–599 (2018).

17. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat. Protoc. 15, 2247–2276 (2020).

18. Werth, M. et al. Transcription factor TFCP2L1 patterns cells in the mouse kidney collecting ducts. eLife 6, e24265 (2017).

19. LeBleu, V. S. et al. PGC-1α mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. Nat. Cell Biol. 16, 992–1003, 1–15 (2014).

20. Dimova, D. K. & Dyson, N. J. The E2F transcriptional network: old acquaintances with new faces. Oncogene 24, 2810–2826 (2005).

21. Mullan, P. B., Quinn, J. E. & Harkin, D. P. The role of BRCA1 in transcriptional regulation and cell cycle control. Oncogene 25, 5854–5863 (2006).

22. Sengez, B. et al. The Transcription Factor Elf3 Is Essential for a Successful Mesenchymal to Epithelial Transition. Cells 8, E858 (2019).

23. Watanabe, K. et al. OVOL2 induces mesenchymal-to-epithelial transition in fibroblasts and enhances cell-state reprogramming towards epithelial lineages. Sci. Rep. 9, 6490 (2019).

24. Liu, Z.-H., Dai, X.-M. & Du, B. Hes1: a key role in stemness, metastasis and multidrug resistance. Cancer Biol. Ther. 16, 353–359 (2015).

25. Yin, J. et al. HGF/MET Regulated Epithelial-Mesenchymal Transitions And Metastasis By FOSL2 In Non-Small Cell Lung Cancer. OncoTargets Ther. 12, 9227–9237 (2019).

26. Senft, D. & Ronai, Z. A. UPR, autophagy, and mitochondria crosstalk underlies the ER stress response. Trends Biochem. Sci. 40, 141–148 (2015).

27. LeBleu, V. S. & Kalluri, R. A peek into cancer-associated fibroblasts: origins, functions and translational impact. Dis. Model. Mech. 11, dmm029447 (2018).

28. Conejo-Garcia, J. R. & Rodriguez, P. C. c-Maf: a bad influence in the education of macrophages. J. Clin. Invest. 130, 1629–1631 (2020).

29. Dong, Y. et al. Tumor Xenografts of Human Clear Cell Renal Cell Carcinoma But Not Corresponding Cell Lines Recapitulate Clinical Response to Sunitinib: Feasibility of Using Biopsy Samples. Eur. Urol. Focus 3, 590–598 (2017).

30. Zanjani, L. S. et al. Increased expression of CD44 is associated with more aggressive behavior in clear cell renal cell carcinoma. Biomark. Med. 12, 45–61 (2018).

31. Wei, D. et al. Novel renal medullary carcinoma cell lines, UOK353 and UOK360, provide preclinical tools to identify new therapeutic treatments. Genes. Chromosomes Cancer 59, 472–483 (2020).

32. Nakai, S. et al. Crucial roles of Brn1 in distal tubule formation and function in mouse kidney. Dev. Camb. Engl. 130, 4751–4759 (2003).

33. Phelep, A. et al. MITF - A controls branching morphogenesis and nephron endowment. PLoS Genet. 13, e1007093 (2017).

34. Dixon, S. J. et al. Ferroptosis: an iron-dependent form of nonapoptotic cell death. Cell 149, 1060–1072 (2012).

35. Chen, X., Kang, R., Kroemer, G. & Tang, D. Broadening horizons: the role of ferroptosis in cancer. Nat. Rev. Clin. Oncol. 18, 280–296 (2021).

36. Dai, C. et al. Transcription factors in ferroptotic cell death. Cancer Gene Ther. 27, 645–656 (2020).

37. Yuk, H., Abdullah, M., Kim, D.-H., Lee, H. & Lee, S.-J. Necrostatin-1 Prevents Ferroptosis in a RIPK1- and IDO-Independent Manner in Hepatocellular Carcinoma. Antioxidants 10, 1347 (2021).

38. Stockwell, B. R. Ferroptosis turns 10: Emerging mechanisms, physiological functions, and therapeutic applications. Cell 185, 2401–2421 (2022).

39. Feng, H. et al. Transferrin Receptor Is a Specific Ferroptosis Marker. Cell Rep. 30, 3411-3423.e7 (2020).

40. Tsoi, J. et al. Multi-stage Differentiation Defines Melanoma Subtypes with Differential Vulnerability to Drug-Induced Iron-Dependent Oxidative Stress. Cancer Cell 33, 890-904.e5 (2018).

41. Wang, W. et al. CD8+ T cells regulate tumour ferroptosis during cancer immunotherapy. Nature 569, 270–274 (2019).

42. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. Cell 155, 934–947 (2013).

43. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319 (2013).

44. Weissmiller, A. M. et al. Inhibition of MYC by the SMARCB1 tumor suppressor. Nat. Commun. 10, 2014 (2019).

45. Woodley, C. M. et al. Multiple interactions of the oncoprotein transcription factor MYC with the SWI/SNF chromatin remodeler. Oncogene 40, 3593–3609 (2021).

46. Laurette, P. et al. Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. eLife 4, (2015).

47. Michel, B. C. et al. A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. Nat. Cell Biol. 20, 1410–1420 (2018).

48. Wang, X. et al. BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. Nat. Commun. 10, 1881 (2019).

49. Nakayama, R. T. et al. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. Nat. Genet. 49, 1613–1623 (2017).

50. Wang, X. et al. SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. Nat. Genet. 49, 289–295 (2017).

51. Erkek, S. et al. Comprehensive Analysis of Chromatin States in Atypical Teratoid/Rhabdoid Tumor Identifies Diverging Roles for SWI/SNF and Polycomb in Gene Regulation. Cancer Cell 35, 95-110.e8 (2019).

52. Custers, L. et al. Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. Nat. Commun. 12, 1407 (2021).

53. Betz, B. L., Strobeck, M. W., Reisman, D. N., Knudsen, E. S. & Weissman, B. E. Re-expression of hSNF5/INI1/BAF47 in pediatric tumor cells leads to G1 arrest associated with induction of p16ink4a and activation of RB. Oncogene 21, 5193–5203 (2002).

54. Evans, R. G., Smith, D. W., Lee, C.-J., Ngo, J. P. & Gardiner, B. S. What Makes the Kidney Susceptible to Hypoxia? Anat. Rec. Hoboken NJ 2007 303, 2544–2552 (2020).

55. Msaouel, P., Tannir, N. M. & Walker, C. L. A Model Linking Sickle Cell Hemoglobinopathies and SMARCB1 Loss in Renal Medullary Carcinoma. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 24, 2044–2049 (2018).

56. Humphreys, B. D. et al. Intrinsic epithelial cells repair the kidney after injury. Cell Stem Cell 2, 284–291 (2008).

57. Scindia PhD, Y., Leeds Md, J. & Swaminathan Md, S. Iron Homeostasis in Healthy Kidney and its Role in Acute Kidney Injury. Semin. Nephrol. 39, 76–84 (2019).

58. Fu, R. et al. clustifyr: an R package for automated single-cell RNA sequencing cluster classification. F1000Research 9, 223 (2020).

59. Wu, Y., Tamayo, P. & Zhang, K. Visualizing and Interpreting Single-Cell Gene Expression Datasets with Similarity Weighted Nonnegative Embedding. Cell Syst. 7, 656-666.e4 (2018).

60. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14, 7 (2013).

61. Lambert, S. A. et al. The Human Transcription Factors. Cell 172, 650–665 (2018).

62. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 17, 218 (2016).

63. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. 37, 773–782 (2019).

64. Strub, T. et al. Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. Oncogene 30, 2319–2332 (2011).

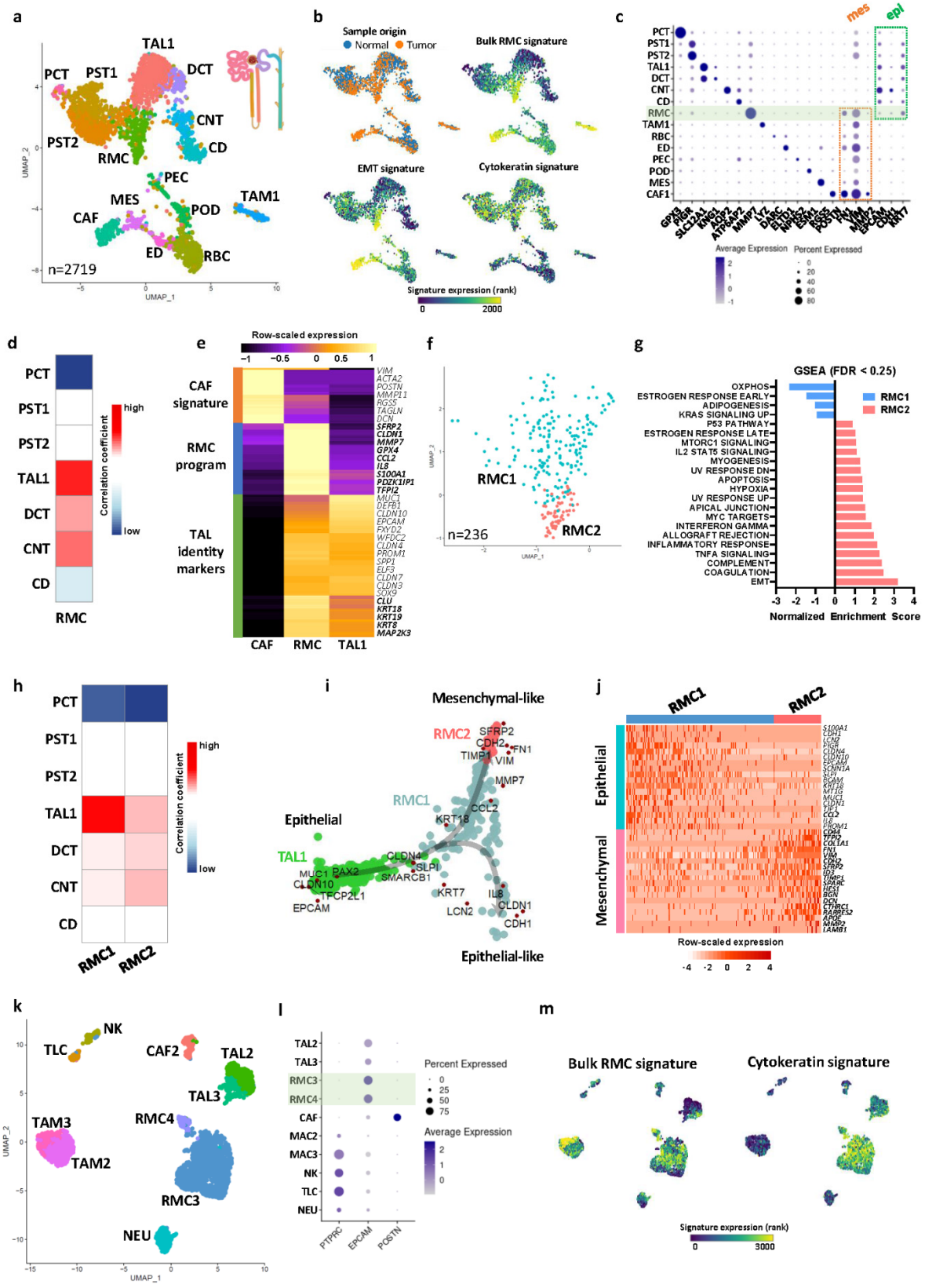
65. Laurette, P. et al. Chromatin remodellers Brg1 and Bptf are required for normal gene expression and progression of oncogenic Braf-driven mouse melanoma. Cell Death Differ. 27, 29–43 (2020).

66. Ye, T. et al. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. 39, e35 (2011).

67. Joshi, S. et al. TEAD transcription factors are required for normal primary myoblast differentiation in vitro and muscle regeneration in vivo. PLoS Genet. 13, e1006600 (2017).

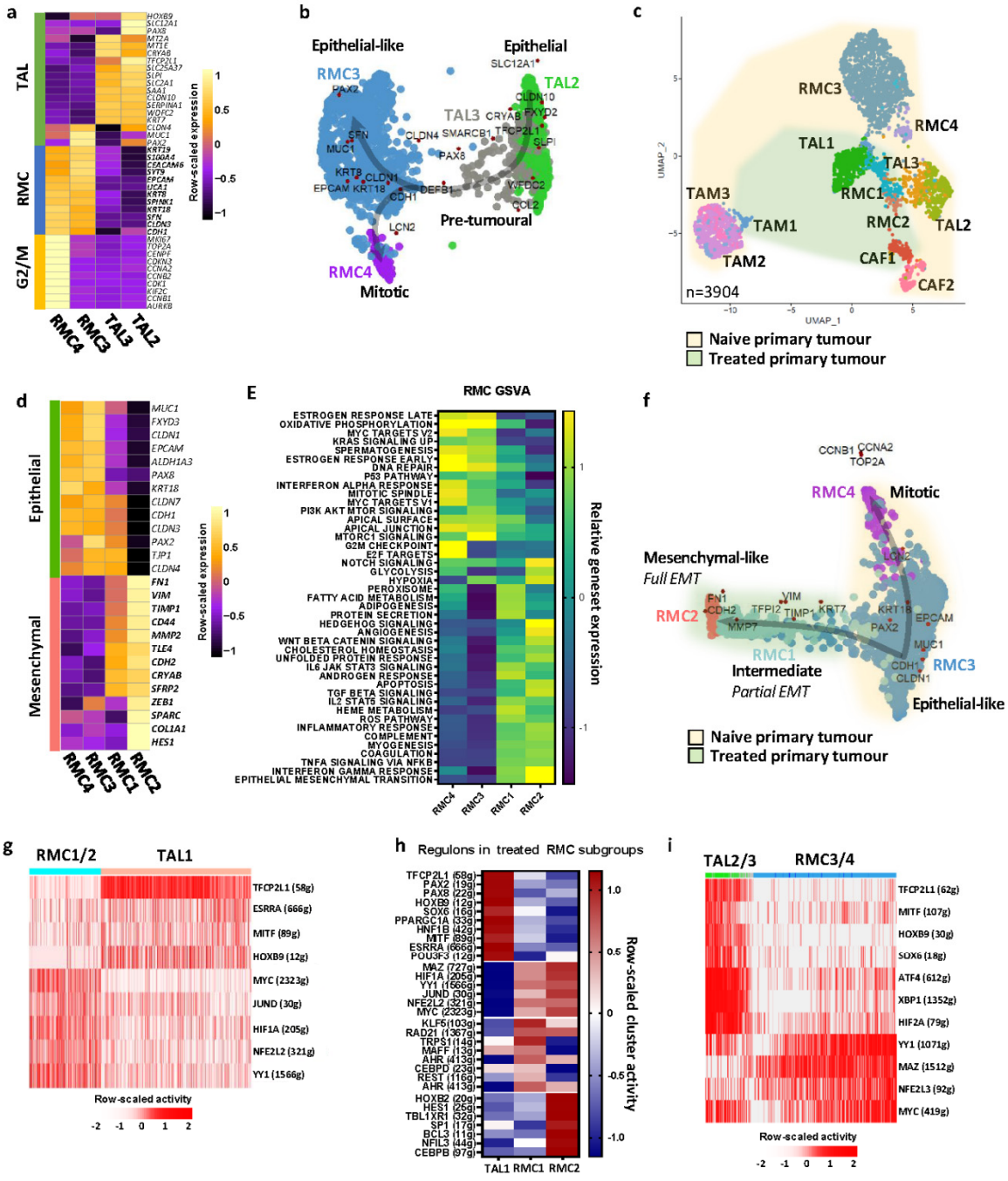
68. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. Bioinforma. Oxf. Engl. 27, 1017–1018 (2011).

G. Main figures



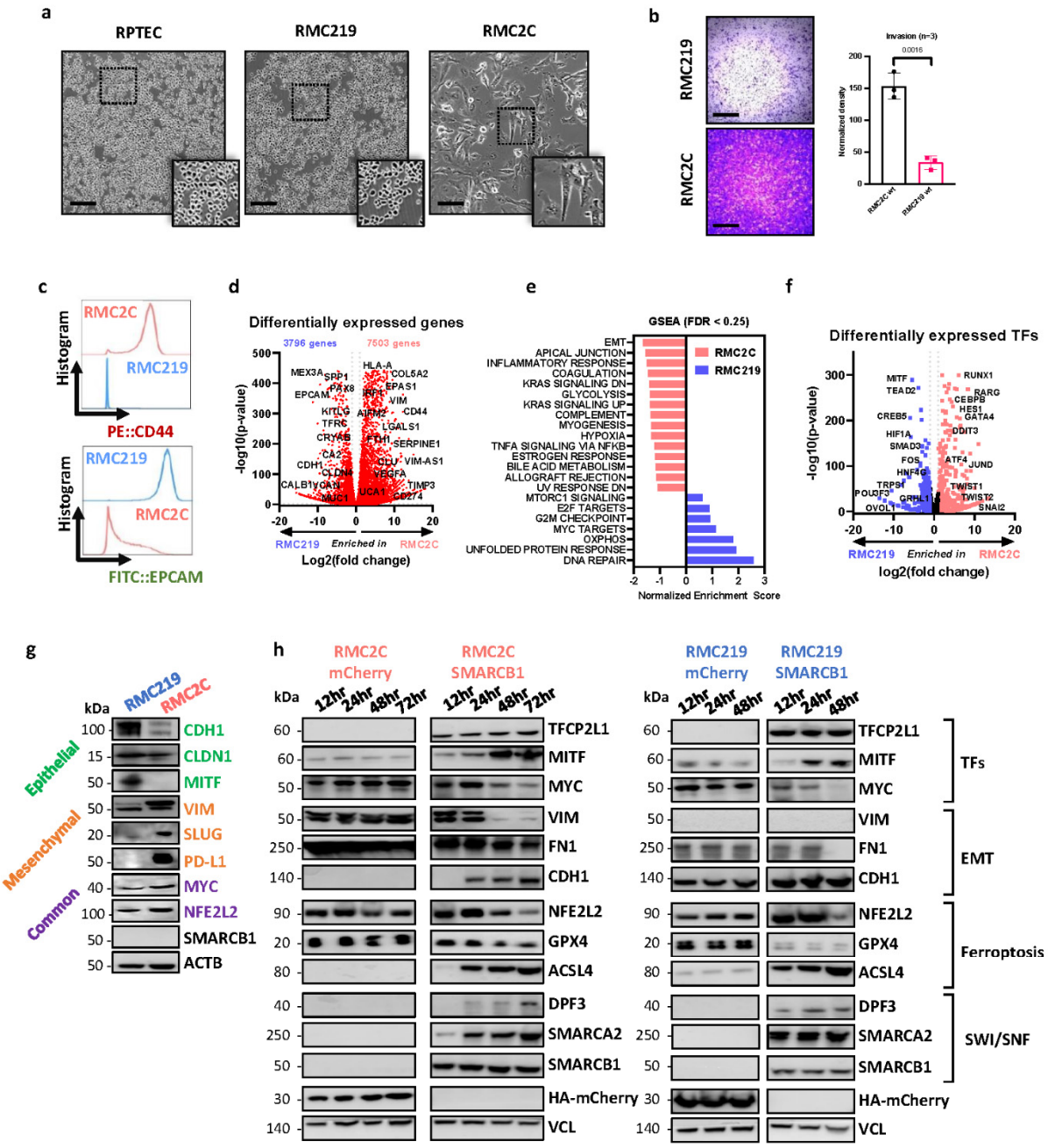
Vokshi et al., Fig. 1

Figure 1. Single-cell RNA sequencing of treated (A-J) and (K-M) naive RMC tumours. a. UMAP plot of the aggregated treated tumour and normal adjacent tissue (NAT) representing the clusters identified by Seurat using a resolution of 1.12. *Abbreviations:* PCT: proximal convoluted tubule cells; PST1/2: proximal straight tubule cells 1 and 2; RMC: Renal medullary carcinoma cells; TAL1: thick ascending tubule cells of Henle's loop; DCT: distal convoluted tubule cells; CNT: connecting tubule cells; CD: collecting duct cells; CAF: cancer-associated fibroblasts; MES: mesangial cells; ED: endothelial cells; RBC: red blood cells; PEC: parietal epithelial cells; POD: podocytes; TAM1: tumour-associated macrophages. **b.** UMAP projection of sample origin or selected gene signatures. **c.** Dot-plots representing gene markers of each identified cluster in the RMC treated sample. Rectangles regroup clusters according to either mesenchymal or epithelial markers. **d.** Clustifyr correlation between RMC cells and renal epithelial tubules transcriptomes. **e.** Pseudo-bulk heatmap of 100 RMC-specific and 50 CAF-specific genes using CAF1, RMC and TAL1 clusters as a matrix. **f.** UMAP representing RMC subclusters as identified by Seurat using a resolution of 1. **g.** GSEA showing enriched "Hallmark gene sets" in RMC1 relative to RMC2 cell clusters. **h.** Clustifyr correlation between RMC subclusters and renal epithelial tubules transcriptomes**. i.** SWNE trajectory analysis of the treated RMC clusters using a set of selected markers per cluster and assuming TAL1 cells as origin. **j.** Heatmap representation of a set of selected EMT genes in the 2 RMC subclusters. **k**. UMAP plot of the naive tumour cell clusters as identified by Seurat. RMC3/4: Renal medullary carcinoma cells; TAL2/3: thick ascending tubule cells of Henle's loop; NEU: neutrophils; CAF2 : cancer-associated fibroblasts; NK : natural killers; TLC : T-lymphocyte cells; TAM2/3 : tumour-associated macrophages. **l.** Dot-plots of selected gene markers of immune, epithelial and CAF cells. **m.** UMAP projection of the bulk RMC and cytokeratin signatures.



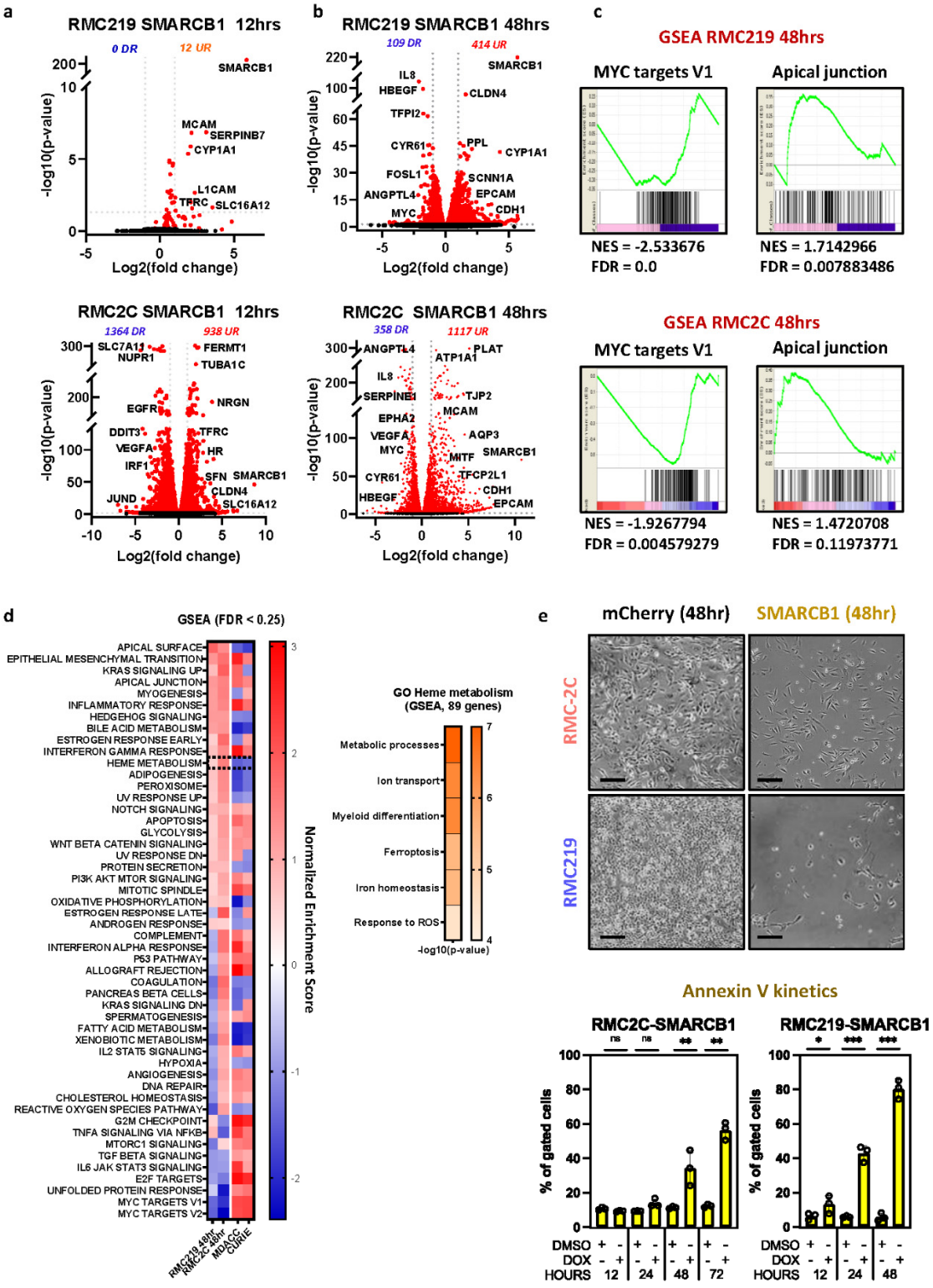
Vokshi et al., Fig. 2

Figure 2. Intra-tumoural heterogeneity of RMC. a. Pseudo-bulk heatmap of expression of top markers of RMC and TAL clusters. **b.** SWNE trajectory analysis of the naive RMC clusters using a set of selected markers per cluster and assuming TAL2 cells as origin. **c.** UMAP representing the normalized merge of selected TAL, RMC, CAF and TAM clusters from the treated tumour (green hue) and the naive tumour (yellow hue). **d.** Pseudo-bulk heatmap showing a set of known EMT markers in all RMC clusters. Note that as RMC4 were cycling RMC3 cells, they were omitted from the analysis to avoid redundancy. **e.** GSVA analysis showing ontologies of indicated RMC clusters. **f.** SWNE trajectory analysis of normalized merged RMC clusters from treated and naive tumours using a set of differentially expressed EMT markers. **g.** SCENIC analysis of the treated tumour showing top regulons of RMC1/2 and TAL1 cells. Note that in brackets are indicated the number of genes (g) per selected regulon. **h.** SCENIC analysis of the treated tumour indicating activities of TAL regulons and RMC1- or RMC2-specific regulons. **i.** SCENIC analysis of the naive tumour revealing top regulons of RMC3/4 and TAL2/3 cells.



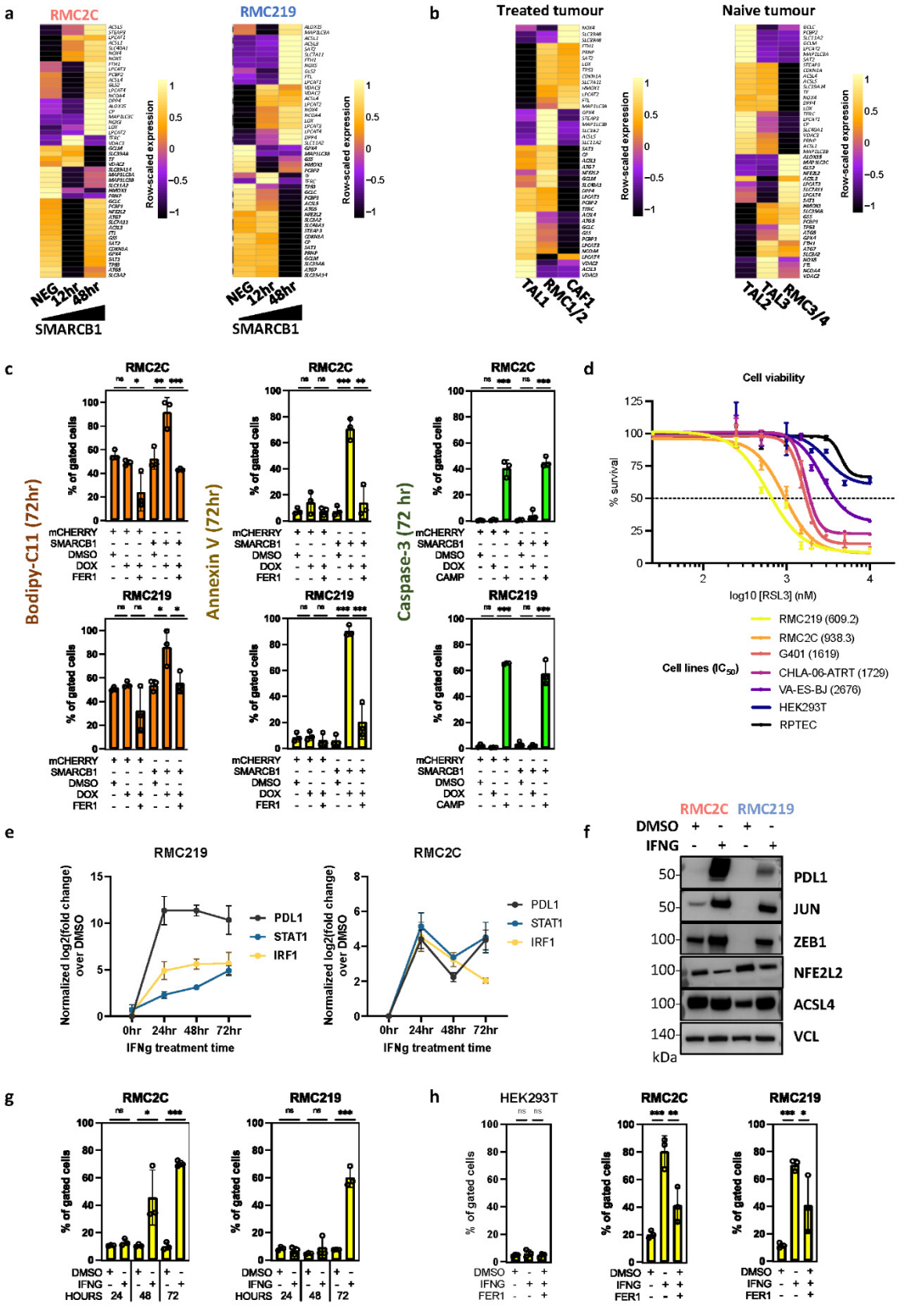
Vokshi et al., Fig. 3

Figure 3. Cultured RMC cells recapitulate EMT cell states. a. Phase-contrast microscopy at 20X magnification of normal kidney (RPTEC) and tumour cells (RMC219 and RMC2C) showing distinct morphologies of RMC lines. Scale bar: 250 µm. n=3 independent biological replicates. **b.** Brightfield microscopy at 4X magnification of Boyden chamber matrigel assays using RMC lines (left) and absolute quantification using absorbance of resuspended crystal violet (right). Scale bar: 1000µm. Biological triplicates are plotted as means ± SD and one-sided unpaired t-test analyses were calculated by Prism 5. **c.** Flow cytometry of membrane protein expression of EPCAM and CD44 in RMC lines. **d.** Volcano plot depicting differentially expressed genes using normalized bulk RNA-seq of RMC lines. P-values were derived using the Wald test and adjusted using Benjamini-Hochberg FDR correction. **e.** GSEA using the Hallmarks genesets showing pathways enrichment in respective RMC lines. Note that only pathways with FDR < 0.25 are shown. **f.** Volcano plot of differentially expressed 1681 FANTOM5-defined TFs using normalized bulk RNA-seq of RMC lines. P-values were derived using the Wald test and adjusted using Benjamini-Hochberg FDR correction. **g.** Immunoblots detecting the indicated proteins. n=3 independent biological replicates. **h.** Immunoblots showing expression of selected proteins upon re-expression of SMARCB1 in RMC2C (left) and RMC219 (right). n=3 independent biological replicates.



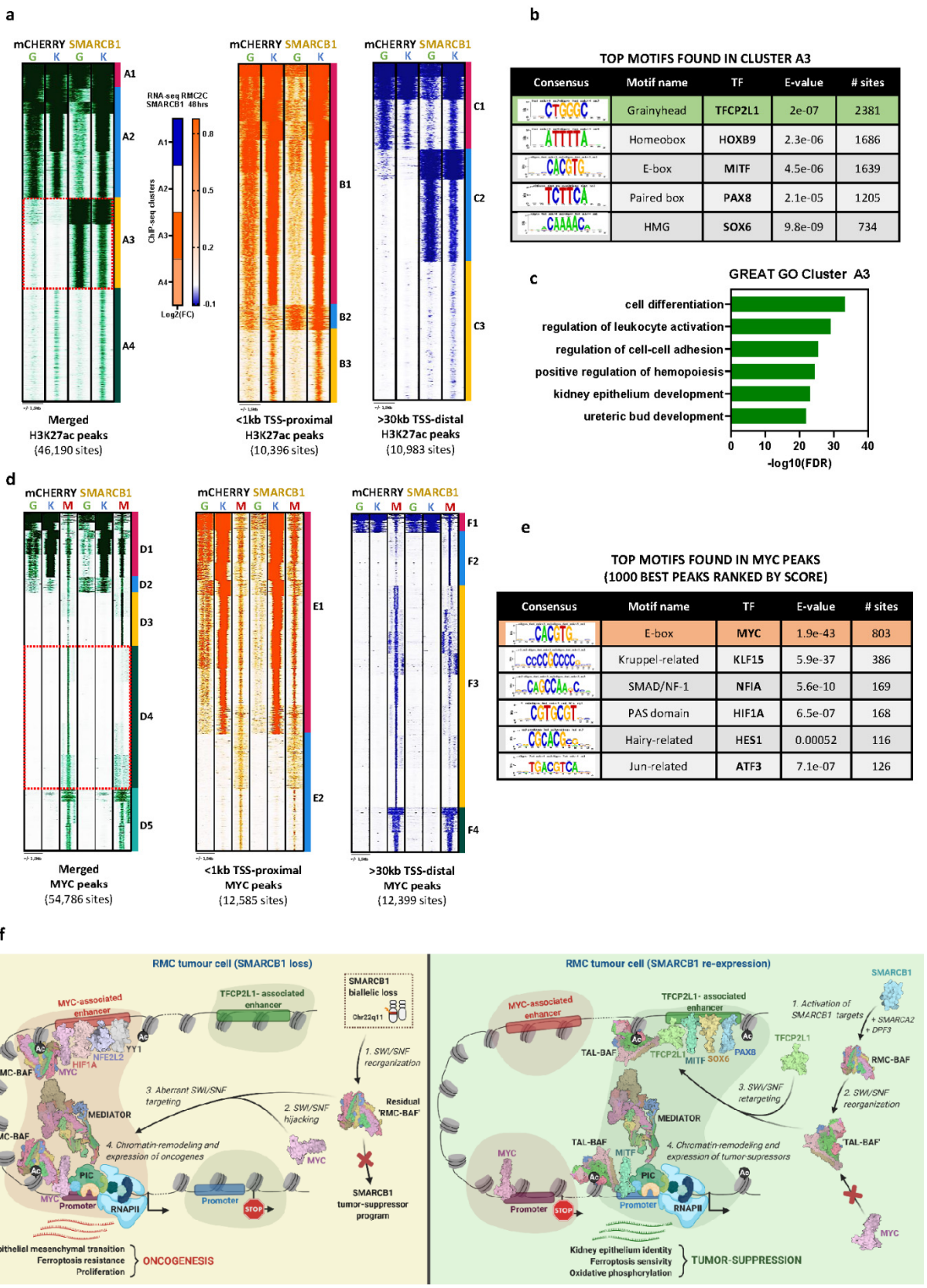
Vokshi et al., Fig. 4

Figure 4. Tumour-suppressor function of SMARCB1. a. Volcano plot revealing up- and down-regulated genes at 12 hrs after SMARCB1 re-expression in RMC lines. P-values were derived using the Wald test and adjusted using Benjamini-Hochberg FDR correction. **b.** Volcano plot revealing up- and down-regulated genes at 48hrs after SMARCB1 re-expression in RMC lines. P-values were derived using the Wald test and adjusted using Benjamini-Hochberg FDR correction. **c.** GSEA showing top up- and down-regulated pathways upon SMARCB1 re-expression (48 hrs) with similar ontologies observed in both lines. **d.** Integrative heatmap showing GSEA Hallmarks enrichments (left panel) in SMARCB1 re-expressing RMC lines and 2 cohorts of RMC primary tumours (MDACC: n=11; Curie: n=5) and Metascape ontology analysis of genes constituting the GSEA "Heme metabolism" term (right panel). FDR values were derived by GSEA using permutation and Benjamini-Hochberg correction. **e.** Phase-contrast microscopy at 10X magnification of RMC lines 48 hrs after re-expression of either SMARCB1 or mCHERRY control. Scale bar: 500µm. (upper panel) Quantification of cell death in RMC lines at selected time-points upon SMARCB1 re-expression, as assessed by flow cytometry (lower panel). Note that the % of cells staining positive for either ANXA5 or propidium iodide were tagged as 'dead'. The remaining unstained cells were tagged 'viable'. Biological triplicates are plotted as means ± SD and one-sided unpaired t-test analyses were performed by Prism 5 by comparing matched time-points. P-values: ns= p>0,05; *= p<0,05; **= p<0,01; ***= p<0,001 and ****=p<0,0001.



Vokshi et al., Fig. 5

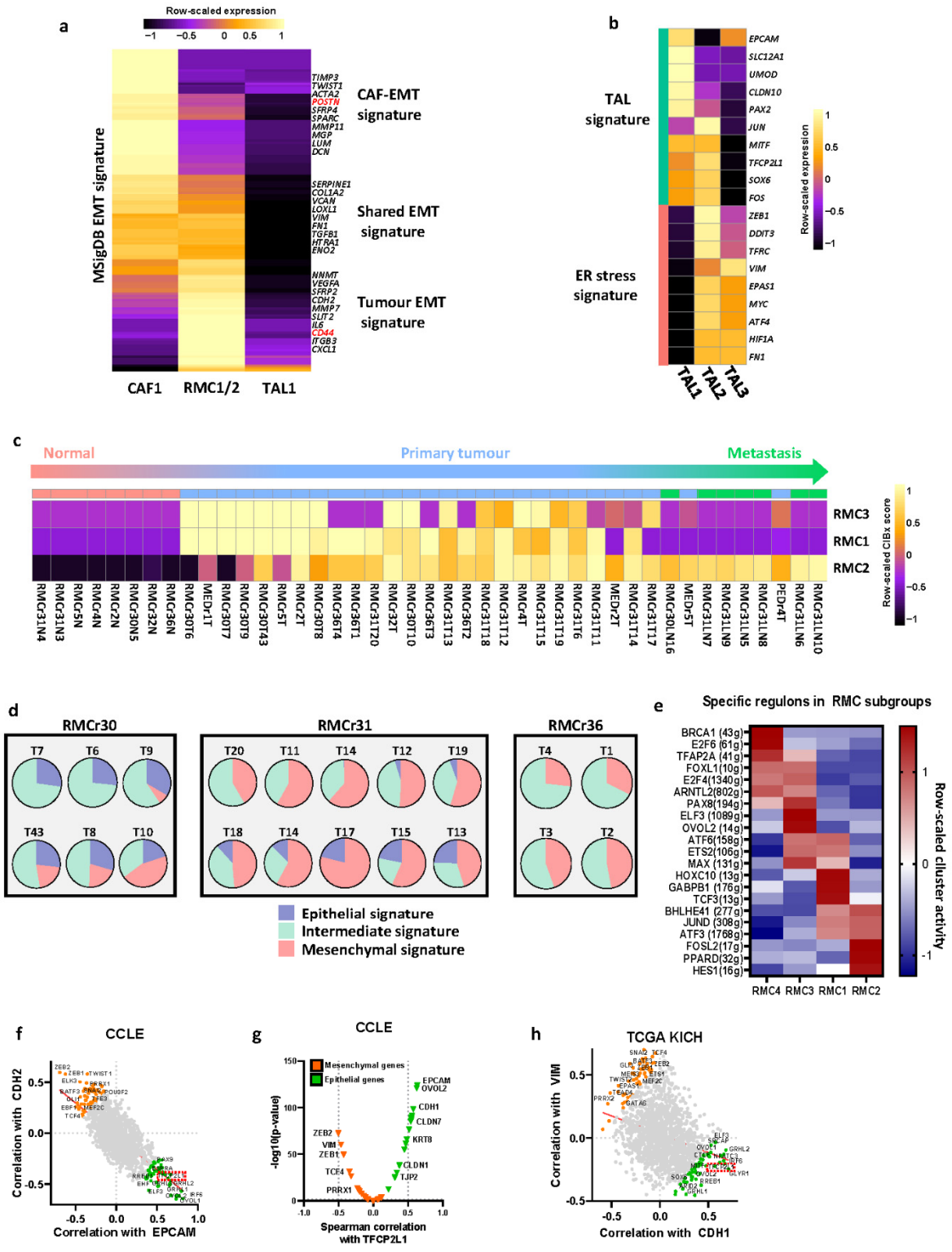
Figure 5. SMARCB1 regulates ferroptosis. a. Heatmap showing the KEGG ferroptosis gene signature in SMARCB1 re-expressing RMC2C (left) and RMC219 (right) cells. **b.** Heatmap showing expression of the ferroptosis gene signature in RMC and TAL clusters. **c.** Flow cytometry quantification of Bodipy-C11, ANXA5 and cleaved CASP3 at 72hrs in SMARCB1 or mCHERRY expressing cells and using either Ferrostatin-1 (Fer1) or camptothecin (CAMP) as controls. Biological triplicates are plotted as means ± SD and one-sided unpaired t-test analyses were performed by Prism 5, ns= p>0,05; *= p<0,05; **= p<0,01; ***= p<0,001 and ****=p<0,0001. **d.** Cell viability (IC50) upon increasing concentrations of RSL3, a class II ferroptosis inducer. Biological triplicates are plotted as means ± SEM. **e.** Gene expression changes of known IFNg downstream targets upon treatment of RMC lines with 10ng/mL recombinant human IFNg. Biological triplicates are plotted as means ± SEM. **f.** Immunoblots showing expression of selected EMT and ferroptosis markers in RMC lines treated either with IFNg or DMSO vehicle control. n=3 independent biological replicates. **g.** Cell death quantified by flow cytometry using annexin-V in RMC lines. Biological triplicates are plotted as means ± SD and one-sided unpaired t-test analyses were performed by Prism 5, ns= p>0,05; *= p<0,05; **= p<0,01; ***= p<0,001 and ****=p<0,0001. **h.** Flow cytometry-based quantification of cell death at 72hrs upon treatment with IFNg alone, IFNg with Fer1 or DMSO in RMC lines and normal kidney cells as control. Represented values are the mean of three biological replicates and unpaired t-test analyses were performed with Prism5 by comparing conditions to matched DMSO. P-values: ns= p>0,05; *= p<0,05; **= p<0,01; ***= p<0,001 and ****=p<0,0001.



Vokshi et al., Fig. 6

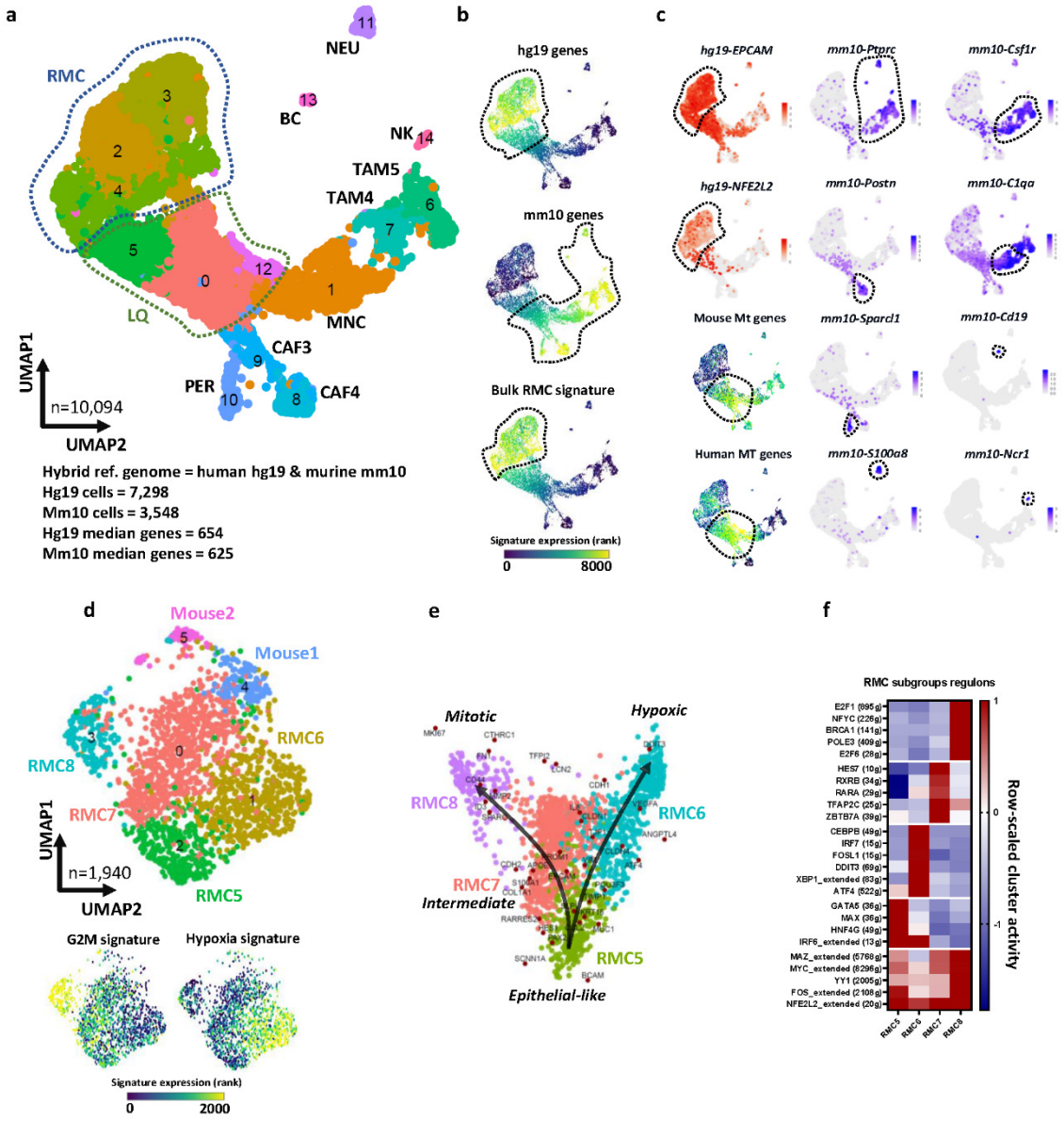
Figure 6. SMARCB1 retargets SWI/SNF complexes to enhancers bearing TFCP2L1 motifs. a. Read density maps showing genome localization BRG1 (G) and H3K27ac (K) in RMC2C cells expressing either SMARCB1 or mCHERRY using as a reference all merged H3K27ac sites (1st panel), all TSS-proximal H3K27ac sites (3rd panel) and all TSS-distal H3K27ac sites (4th panel). Expression changes for genes associated with BRG1/H3K27ac- clusters following SMARCB1 re-expression are shown in the 2nd panel. **b.** RSAT-based motif enrichment analysis using A3 sites ranked by number of sites. **c.** Ontology analysis of genes associated with A3 as annotated by GREAT. **d.** Read density maps showing genome localization of BRG1 (G), H3K27ac (K) and MYC (M) in RMC2C cells expressing either SMARCB1 or mCHERRY using as a reference all merged MYC sites (1st panel), all TSS-proximal MYC sites (2nd panel) and all TSS-distal MYC sites (3rd panel). **e.** RSAT-based motif enrichment analysis using one thousand best MYC peaks ranked (by peak score). **f.** Working model of the oncogenic and SMARCB1 tumour-suppressor events in RMC. Created with BioRender.com with publication and licensing rights (LV22PUT4WB).

H. Supplementary figures



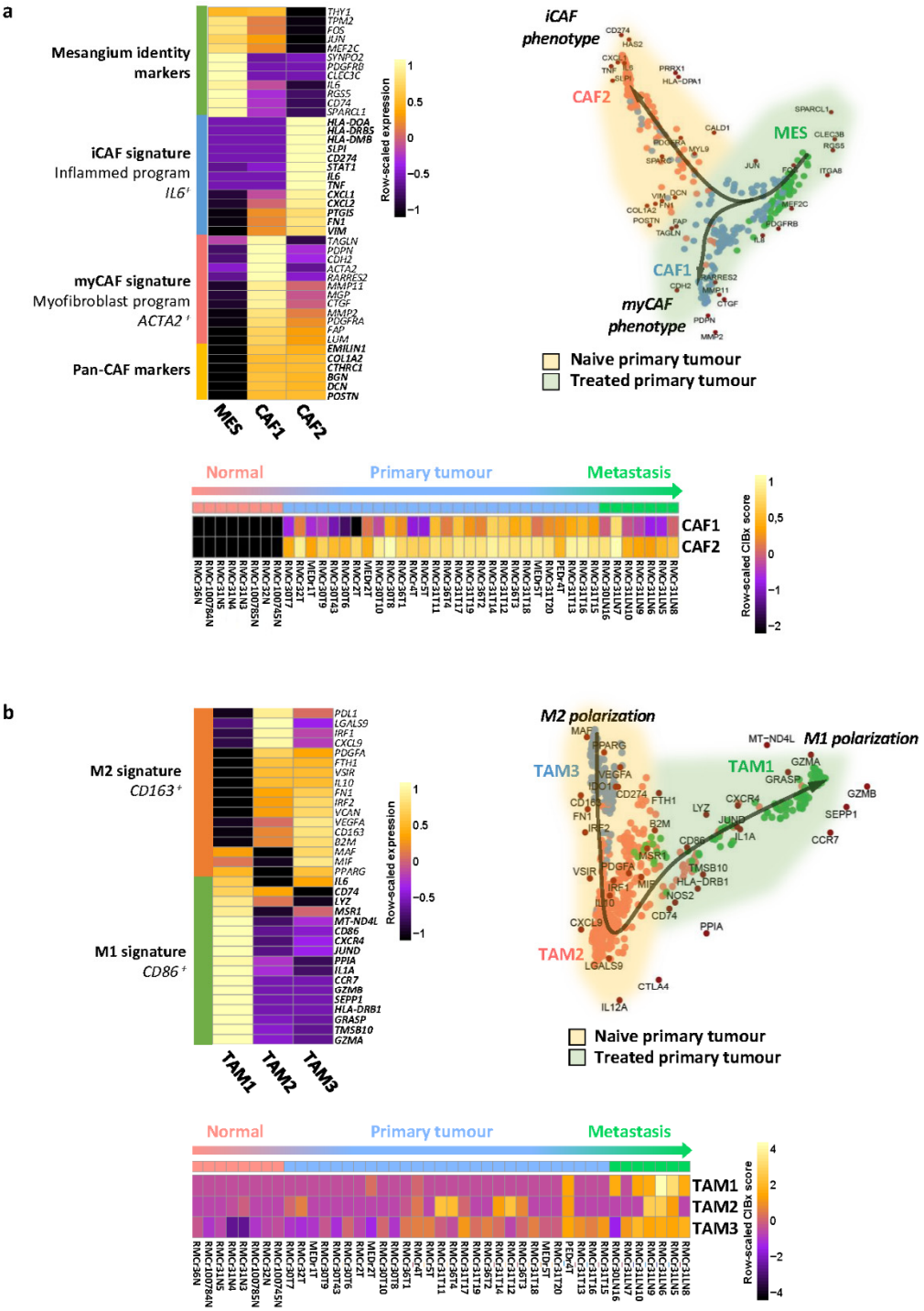
Vokshi et al., Suppl. Fig. 1

Supplementary Figure 1. a. Pseudo-bulk heatmap showing expression of the MSigDB Hallmark EMT signature (183 genes) in RMC1/2 and CAF1 cells from the treated tumour. **b.** Pseudo-bulk heatmap showing heterogeneous expression of selected TAL identity markers, mesenchymal and ER stress genes in all TAL clusters. **c.** Deconvolution of RMC specific signatures as calculated by CIBERSORTx on bulk RNA-seq from sections of RMC primary tumours, lymph node metastasis and normal adjacent tissues. **d.** Pie charts representing intratumoural heterogeneity of RMC signatures using multi-region RNA sequencing of primary RMC tumours (n=3). Note that relative proportions (in %) were inferred by CIBERSORTx using our scRNA-seq normalized merge**. e.** SCENIC analysis of normalized merge of treated and naive RMC samples revealing specific regulons of all RMC clusters. **f.** Pearson correlation analysis of 1683 human transcription factors with selected genes in CCLE database. **g.** Pearson correlation analysis of TFCP2L1 expression with a set of epithelial and mesenchymal genes in CCLE database. P-values were calculated using a two-sided t-test corrected with Benjamin-Hochberg FDR adjustment. **h.** Pearson correlation analysis of 1683 human transcription factors with selected genes in TGCA KICH RCC samples. Note the positive correlation of TFCP2L1 expression level with epithelial markers along with other Grainyhead family members, OVOL1/2 and MITF.



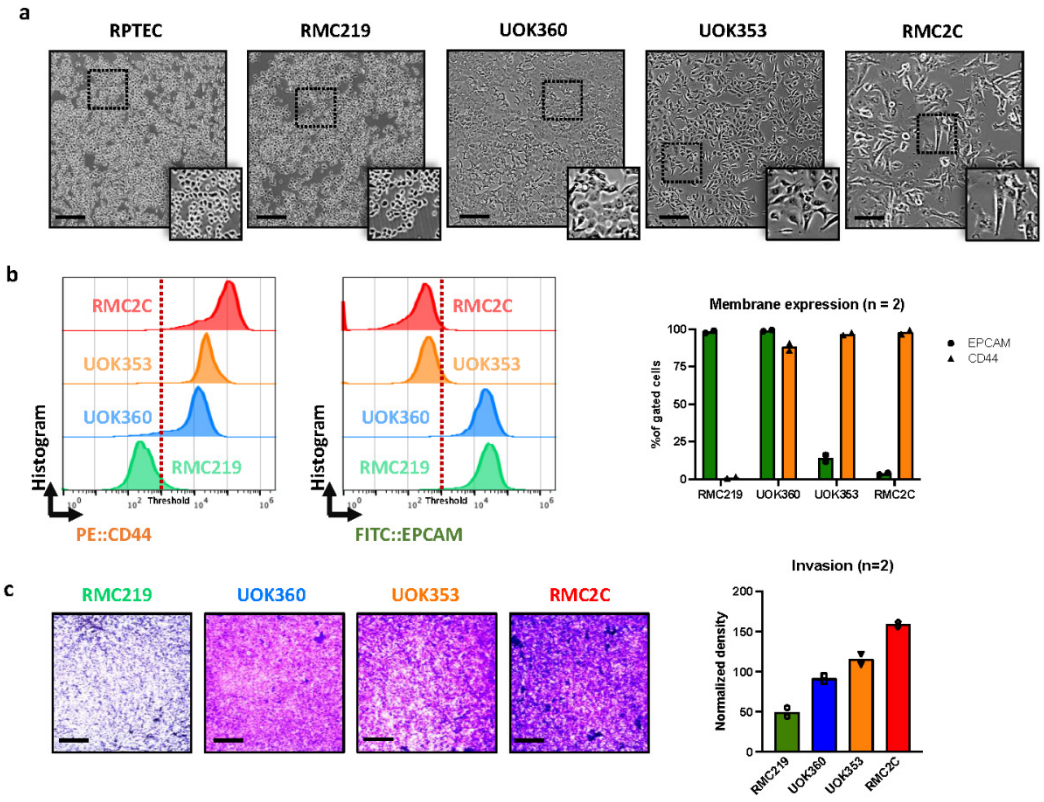
Vokshi et al., Suppl. Fig. 2

Supplementary Figure 2. a. UMAP plot of scRNA-seq from the RMC PDX (IC-pPDX-132) representing cell clusters as calculated by Seurat at a resolution of 0.3. Clusters were identified using hallmark genes shown in Fig. 3b-c. *Abbreviations:* RMC: Renal medullary carcinoma cells; LQ: low quality cells; PER: pericytes; CAF3/4: cancer-associated fibroblasts; MNC: monocytes; TAM4/5: tumour-associated macrophages; NK: natural killers; BC: B-cells; NEU: neutrophils. **b.** UMAP projection of selected gene signatures. Human and murine signatures were established using differential gene nomenclature. **c.** UMAP projection of marker genes and mitochondrial gene signatures. Human and murine mitochondrial gene signatures were established using differential gene nomenclature. **d**. UMAP representing PDX RMC subclusters as identified by Seurat using a resolution of 1 (upper panel), Average expression of selected MSigDB gene signatures (lower panel). **e.** SWNE trajectory analysis of RMC cells using markers of each cluster. **f.** SCENIC analysis of RMC subclusters.



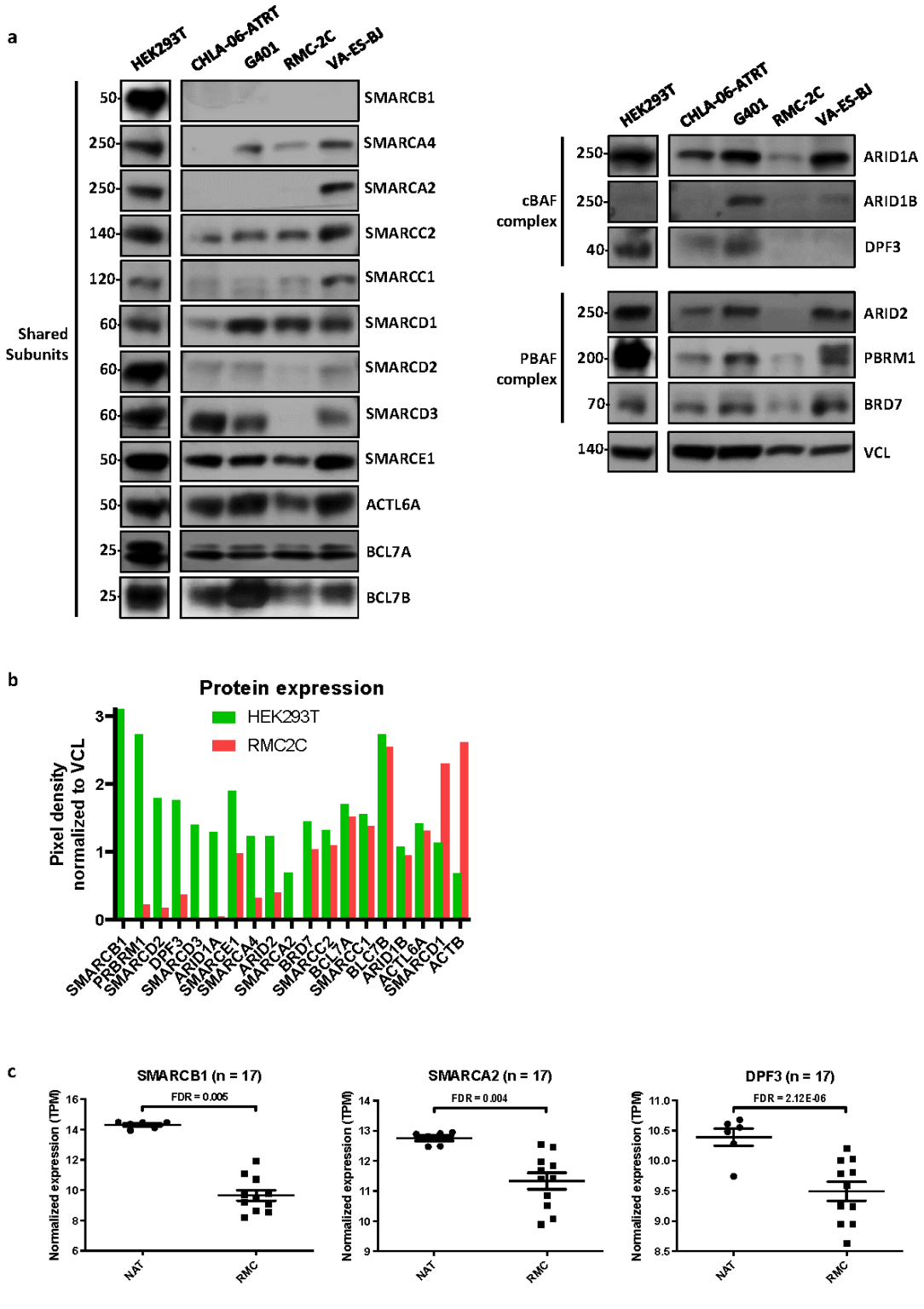
Vokshi et al., Suppl. Fig. 3

Supplementary Figure 3. a. Pseudo-bulk heatmap of iCAF and myCAF signature genes as well as MES and CAF identity markers in CAF clusters and their putative MES cell-of-origin. SWNE trajectory analysis using a set of selected CAF and MES markers revealing distinct CAF phenotypes. Arrow indicates the putative trajectory of CAF activation from MES cells. Deconvolution of CAF specific signatures on bulk RMC RNA-seq. **b.** Pseudo-bulk heatmap of macrophages M1 and M2 gene signatures in TAM clusters from the treated and naive tumours. SWNE trajectory analysis of TAM clusters using a set of selected M1/M2 polarization markers revealing distinct TAM phenotypes. Arrow indicates the putative trajectory in the treated RMC sample. Deconvolution of TAM specific signatures on bulk RNA-seq from sections of RMC.



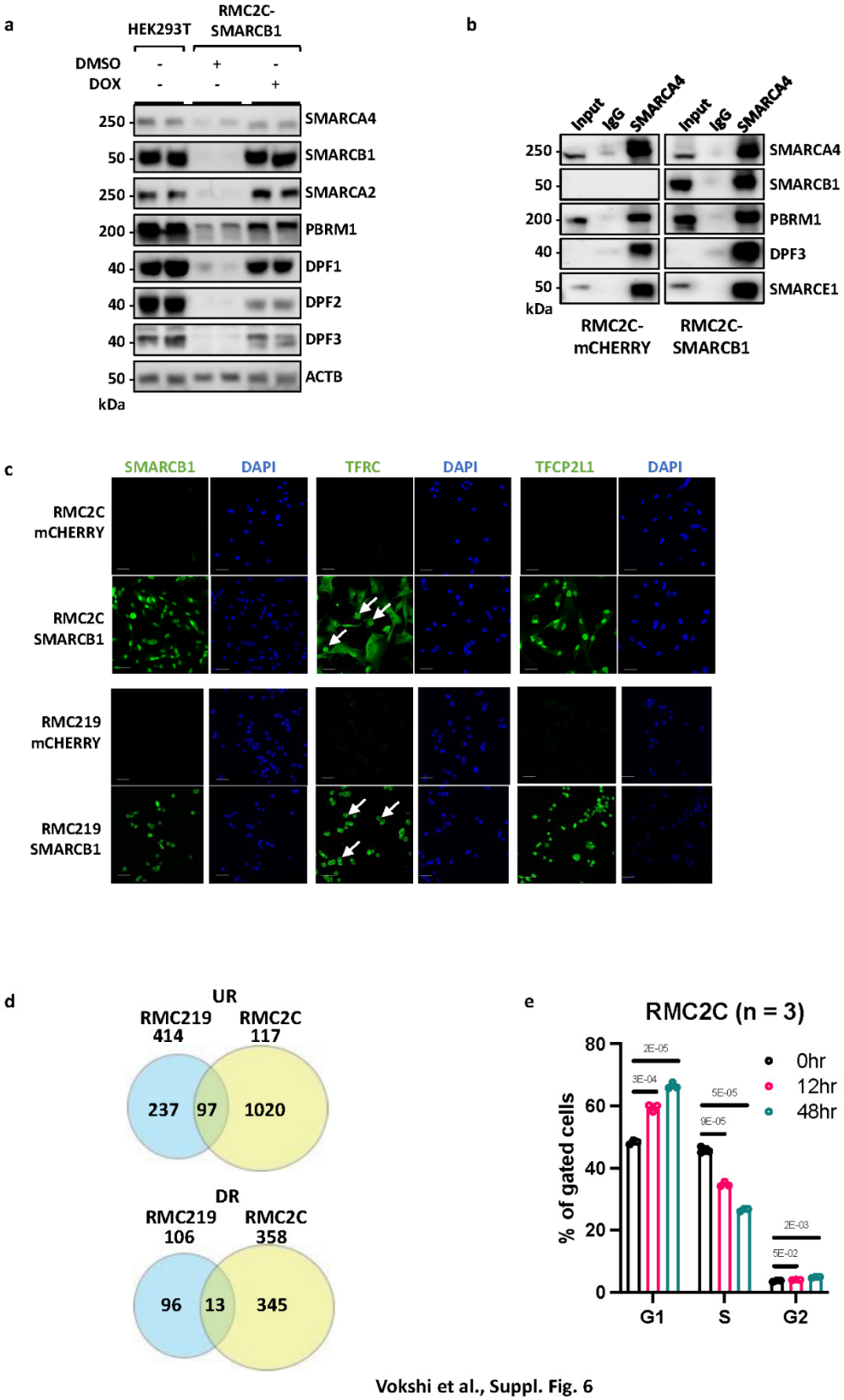
Vokshi et al., Suppl. Fig.4

Supplementary Figure 4. a. Phase-contrast microscopy at 20X magnification of normal kidney (RPTEC) and the indicated tumour cells. Scale bars: 250µm. n=3 independent biological replicates. **b.** Flow cytometry of membrane protein expression of EPCAM and CD44 in RMC lines. n=2 independent biological replicates. **c.** Brightfield microscopy at 4X magnification of Boyden chamber matrigel assays using RMC lines (left) and absolute quantification using absorbance of resuspended crystal violet (right). n=2 independent biological replicates. Scale bars: 1000µm.



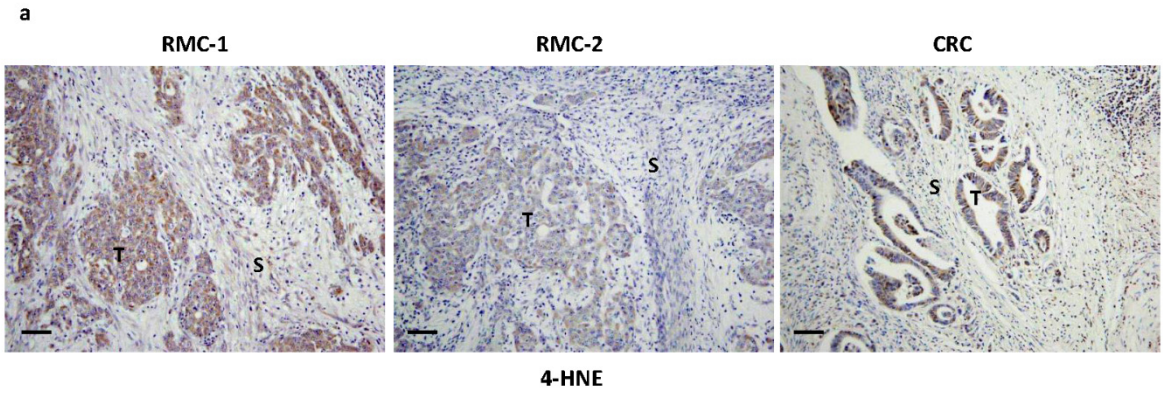
Vokshi et al., Suppl. Fig. 5

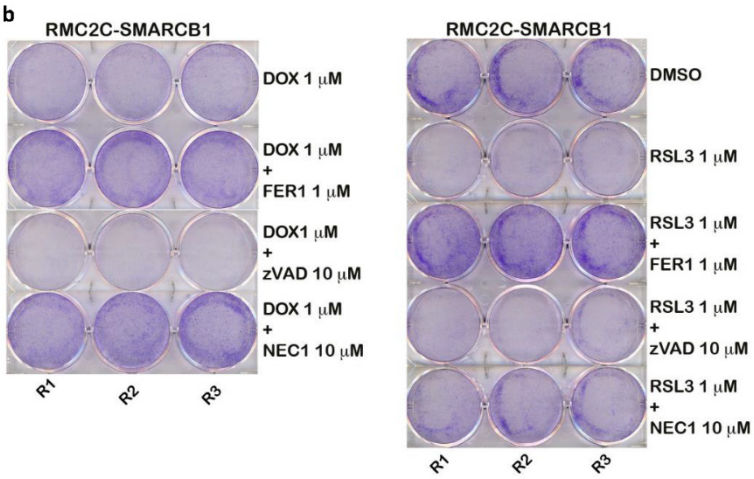
Supplementary Figure 5. a. Immunoblots revealing expression of SWI/SNF subunits in RMC2C cells, 3 additional SMARCB1-deficient lines and HEK293T cells. HEK293T: immortalized human embryonic kidney cells; CHLA-06-ATRT: atypical teratoid/rhabdoid tumour cell line; G401: malignant rhabdoid tumour cell line; VA-ES-BJ: epithelioid sarcoma cell line. Loading normalisation: VCL. **b**. Quantification of expression of the indicated subunits based on scanning of the immunoblots in panel a and corrected for VCL levels. **c.** Scatter plots showing the expression of a selection of SWI/SNF genes in RMC and normal adjacent tissue (NAT). n=17 biological samples: 11 tumor and 6 normal adjacent tissues. Data are shown as means ± SEM and p-values were derived using the Wald test and adjusted with the Benjamini-Hochberg FDR procedure.

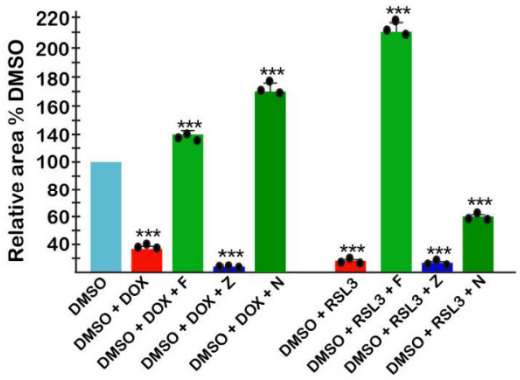


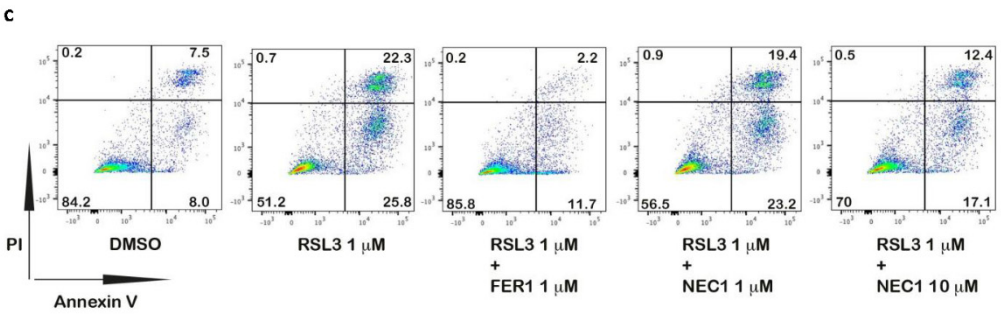
Vokshi et al., Suppl. Fig. 6

Supplementary Figure 6. a. Immunoblots showing re-expression of SMARCB1 and expression of selected SWI/SNF subunits in RMC2C cells with or without Dox compared with HEK293T cells. n=3 independent biological replicates. **b**. Co-precipitation of re-expressed SMARCB1 with SWI/SNF subunits in RMC2C cells. n=3 independent biological replicates. **c**. Immunostaining of RM2C2 and RMC219 cells 24 hours after Dox treatment with the indicated antibodies. Arrows indicate cells where TFRC is localized at the plasma membrane. Captured on a confocal microscope at 40X magnification, scale bars: 100µm. n=3 independent biological replicates. **d**. Venn diagrams showing overlap between genes up and down regulated in RMC219 and RMC2C cells 48 hours after Dox treatment. **e.** Cell cycle analyses of RMC2C cells at the indicated times after Dox treatment illustrating accumulation of cells at G1/S and G2/M and reduction in S-phase cells. n= 3 independent biological replicates. Data are shown as means ± SEM and p-values were calculated with Prism5 using two-sided unpaired Student t-tests.



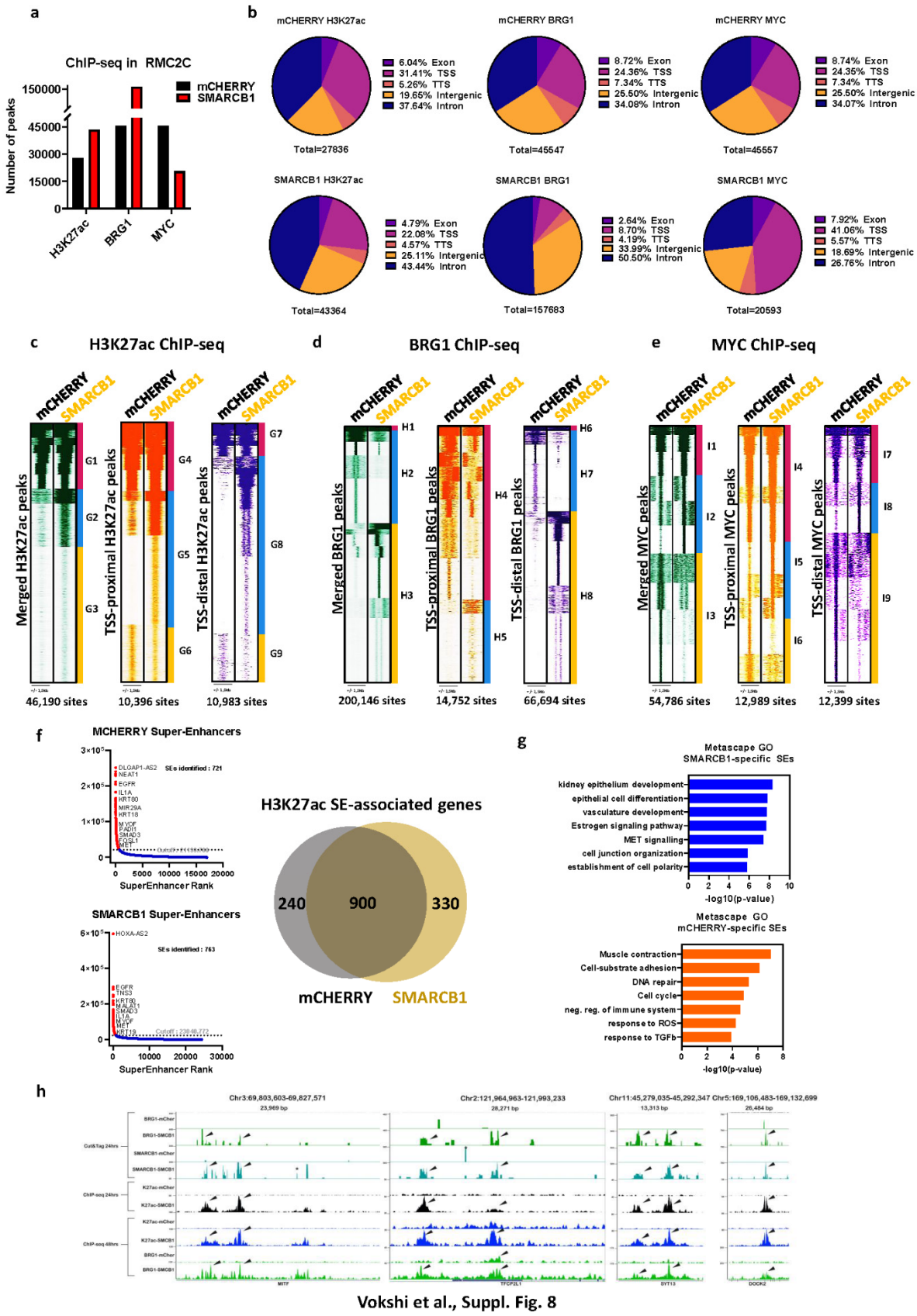






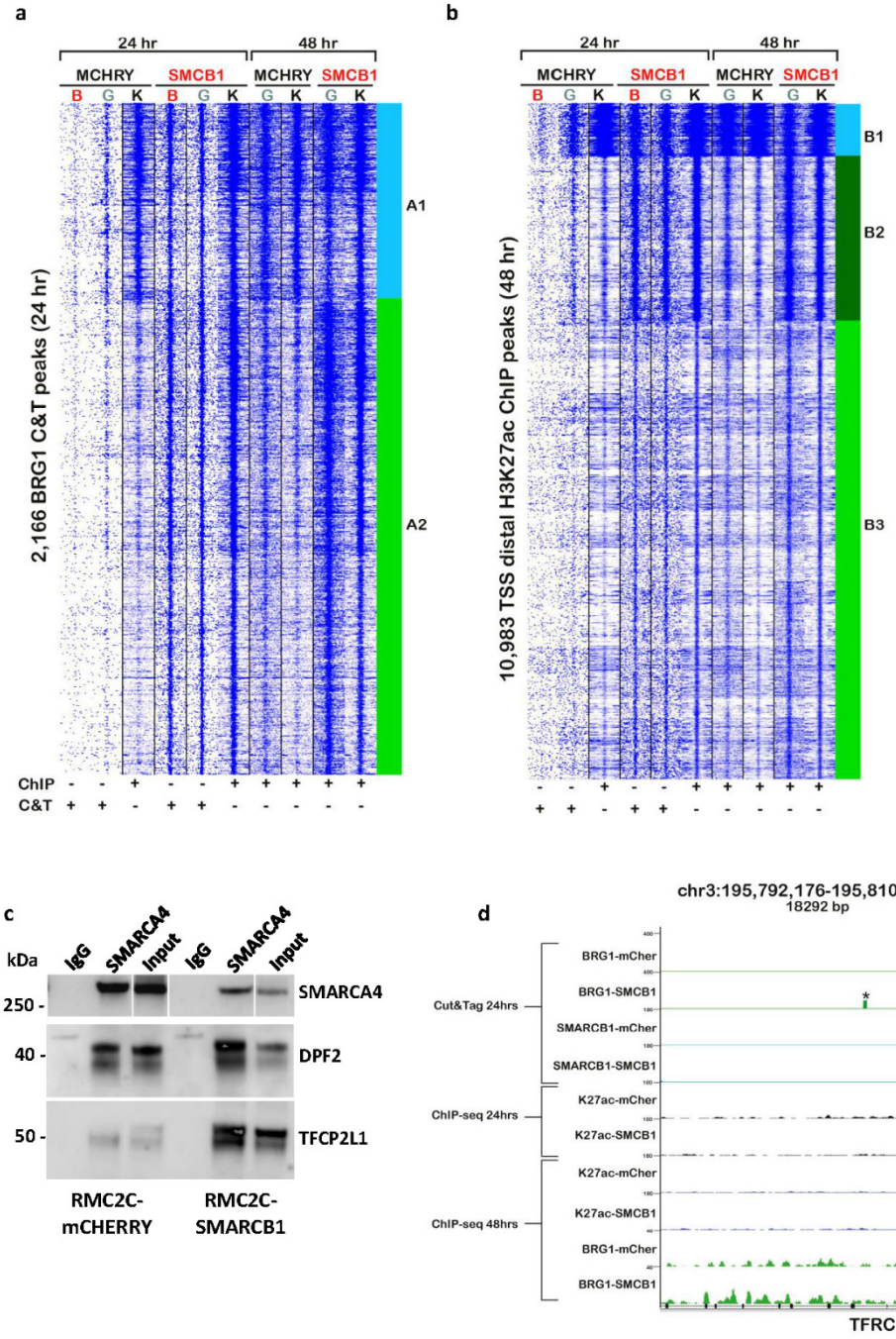
Vokshi et al., Suppl. Fig. 7

Supplementary Figure 7. a. Immunohistochemistry revealing Anti-4 Hydroxynonenal staining of sections from 2 independent RMC tumours and a colorectal tumour as positive control. T = tumour; S = stroma; all images 20X magnification. Scale bars: 100µm. **b.** Cell viability assays in presence of Dox or RSL3 and the indicated inhibitors. 1.5 x10^3 cells were plated and treated as indicated for 72 hours before staining with Crystal Violet and quantification by calculating the area fraction covered by using the FIJI software. The same threshold value was applied to every well and the "relative area" expressed as a fraction of the DMSO control set to 100%. n=3 independent biological replicates. R1-R3 indicate replicates, data are shown as means ± SEM and p-values were calculated using a two-way Anova test, *** indicates p<0.001. **c**. Flow cytometry of RSL3-treated cells in presence of the indicated inhibitors.



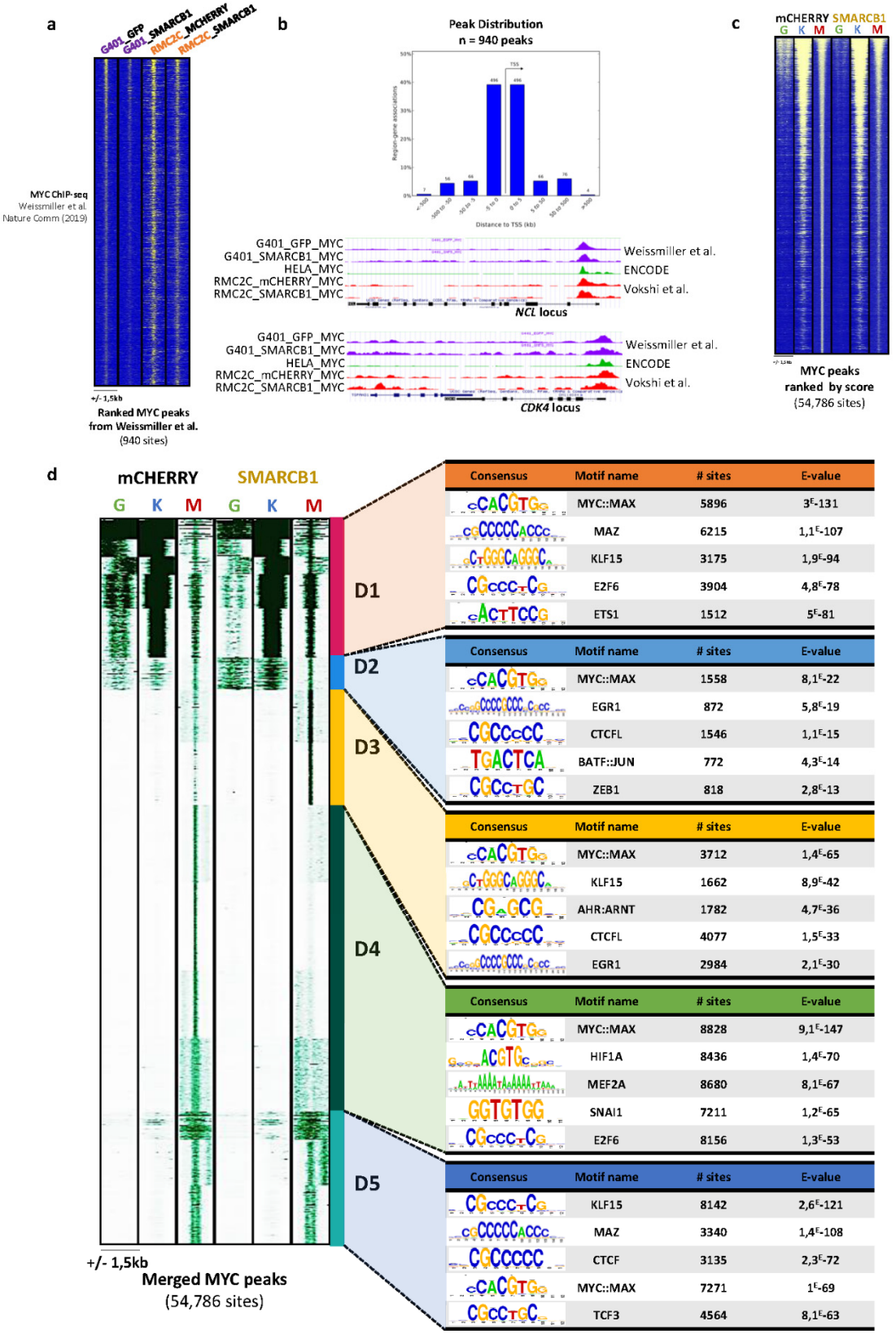
Vokshi et al., Suppl. Fig. 8

Supplementary Figure 8. a. Number of peaks of H3K27ac, BRG1 and MYC in SMARCB1 or mCHERRY-expressing cells as quantified by the MACS algorithm. **b.** Pie charts showing the relative distribution of H3K27ac, BRG1 and MYC peaks on defined genome elements. **c-e**. Read density maps of H3K27ac (C), BRG1 (D), MYC (E) peaks in SMARCB1- or mCHERRY-expressing cells using either all merged, TSS-proximal or TSS-distal sites as a reference. **f.** ROSE identification of H3K27ac Super-Enhancers (SE) in RMC2C cells expressing SMARCB1 or mCHERRY (left), and Venn diagram of shared and specific SE-associated genes (right). **g.** Ontology enrichment analysis of mCHERRY- and SMARCB1-specific SE-associated genes. P-values were calculated by GREAT using a binomial test corrected with Benjamin-Hochberg FDR adjustment. **h**. UCSC genome track snapshots showing the SMARCB1, BRG1 and H3K27ac signals at regulatory elements of selected relevant genes.



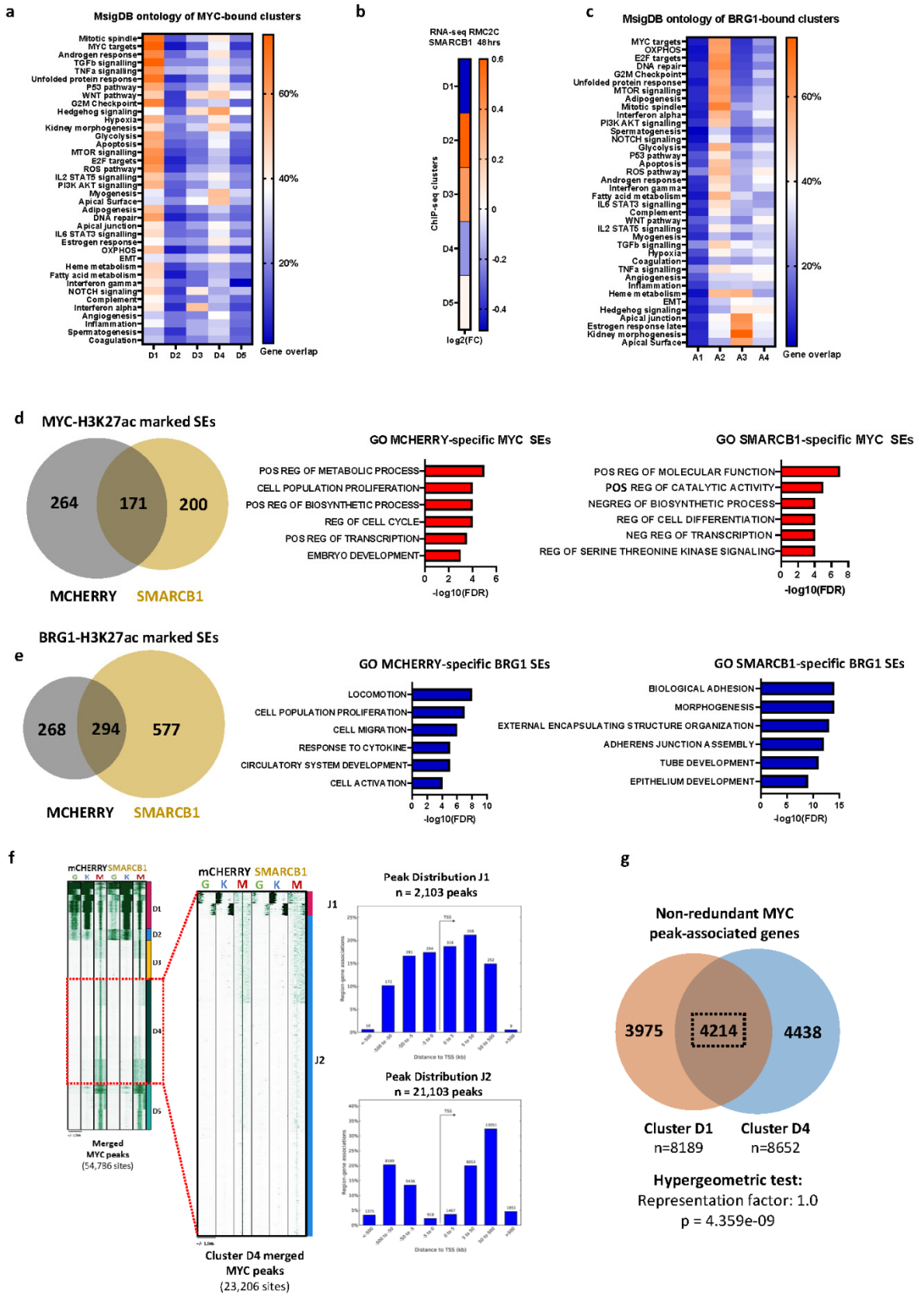
Vokshi et al., Suppl. Fig. 9

Supplementary Figure 9. **a.** Read density heat map of Cut&Tag and H3K27ac ChIP-seq 24 hours after Dox treatment of mCherry (MCHRY) or SMARCB1 (SMCB1) expressing cells. **b**. Read density heat map of Cut&Tag and H3K27ac ChIP-seq 24 hours after Dox treatment at the distal located BRG1 and H3K27ac marked sites 48 hours after Dox treatment. B= SMARCB1; G = BRG1; K = H3K27ac. **c**. Immunoblot showing co-precipitation of TFCP2L1 along with SWI/SNF subunits BRG1 and DPF2. n=3 independent biological replicates. **d**. UCSC genome track snapshots showing the SMARCB1, BRG1 and H3K27ac signals at the TFRC promoter region.



Vokshi et al., Suppl. Fig. 10

Supplementary Figure 10. **a.** Tornado read density maps comparing MYC occupancy in G401 and RMC2C cells. **b.** Distribution of G401 MYC peaks with respect to the TSS. Examples of MYC binding at the *NCL* and *CDK4* loci. **c.** Tornado read density maps showing BRG1, H3K27ac and MYC sites ranked decreasingly by MYC peak score. **d.** RSAT-based motif enrichment analysis at each of the MYC sub-clusters.



Vokshi et al., Suppl. Fig. 11

Supplementary Figure 11. **a.** Percentage of genes associated with MYC clusters as defined by HOMER in the indicated GSEA Hallmark Genesets. **b.** Changes in expression of genes associated with each MYC cluster upon SMARCB1 expression. **c.** Percentage of genes associated with BRG1 clusters as defined by HOMER in the indicated GSEA Hallmark Genesets. **d-e.** Venn diagram of SE-associated genes defined by MYC peak score (D) or BRG1 peak score (E) revealing common and specific SEs in SMARCB1- and mCHERRY-expressing RMC2C cells (left) and associated ontology analysis of SMARCB1- and mCHERRY-specific SE-associated genes (right). **f.** Read density maps showing sub-clustering of MYC D4 sites, with their peak distribution as calculated by GREAT (middle), and the associated ontology analysis of associated genes. **g.** Venn diagram of non-redundant MYC-bound genes found in clusters D1 and D4. P-values were calculated using a two-tailed hypergeometric test.

5. Material and methods

The methods for the main analyses are described in their respective papers (II.1 and II.4 sections), in this section I will describe methods used for the follow-up analyses presented in II.2 and II.3.

Analysis of tumor composition by stage

To analyze tumor composition by stage in the TCGA-KIRC, I used CIBERSORTx inferred absolute score for each population and applied the same methodology as for tumor grade presented in the II.1 section.

Analysis of ligand-receptor interactions

Ligand-receptor interactions for ccRCC.mes/CD8.eff and myCAF/CD8.eff were computed by CellPhoneDB and significant interaction couples extracted in a similar way as for ccRCC.mes/myCAF interactions presented in the II.1 section.

Survival analysis

To build the MES.score, I summed ccRCC.mes, ccRCC.inf and myCAF absolute scores inferred by CIBERSORTx. Analyses of survival are done using R survival and survminer packages as described in the II.1 section. Forest-plots for COX multiple-regression and univariate analyses were generated using the ggforest() function.

Generation of the myCAF signatures

To build the myCAF signatures, I first performed DESeq2 differential gene expression analysis of myCAF-low versus myCAF-high patients in the BIONIKK dataset. I crossed genes enriched in myCAF-high samples (log2FC>1, adj-pval<5%) with myCAF markers from single-cell data and selected the top 5 and top 10 most enriched genes for the LE5 and LE10 signatures. I then computed the value of these signatures for each sample by using the geometric mean of *CTSK*, *COL12A1*, *ASPN*, *POSTN* and *FAP* (LE5) or *CTSK*, *COL12A1*, *ASPN*, *POSTN*, *FAP*, *COL3A1*, *SERPINF1*, *PDGFRA*, *LUM* and *INHBA* (LE10).

Clustering of the BIONIKK cohort

To assign the BIONIKK samples into immune-high and immune-low groups, I performed deconvolution analysis using the MCP-counter v1.2.0 R-package. I performed hierarchical clustering based on the MCP-counter population scores using the hclust() function with "ward.D2" linkage.

To cluster the BIONIKK samples based on lincRNAs, I first retrieved all lincRNAs based on their biotype annotations and filtered the matrix to keep only genes with an average of 1 normalized count per sample (rowSums>100). I ranked the most variable genes using the mad() function and searched for an optimal number of genes to use. For this, I tested multiple hierarchical clustering by using the 50 most variable genes then incrementing this number by 50 until the genes included did not show enough differential expression patterns across groups which happened when using the top 450 genes. I used the 400 most differential lincRNAs for the final hierarchical clustering with hclust() and "ward.D2" linkage to define the LINC1-LINC5 clusters. I computed marker genes for these groups by performing DESeq2 differential expression analyses of each group against all others and selecting up-regulated genes (log2FC>1, adj-pval<5%). GSEA analysis are performed using the JAVA GSEA software with the pre-ranked algorithm using DESeq2-computed fold-change. Ontology analysis are performed using metascape[255] and DAVID[256].

RNA extraction and quantitative PCR in 501M and MM099 cells

Wet-lab experiment performed by Alexandra Helleux. Total mRNA isolation was performed using Trizol and isopropanol precipitation. RNA was then treated with DNAseI following the TurboDNAse kit instructions (ThermoFisher) and 1µg of RNA was reversed transcribed using SuperScript IV (ThermoFisher) following manufacturer instructions. qPCR was performed using SYBR Green I (Roche) and monitored by a LightCycler 480 (Roche). Target genes expression was normalized using HMBS and GAPDH as reference genes. Primers for RT-qPCR were designed using the OligoAnalyzerTMTool from IDT and their sequences are as follows:

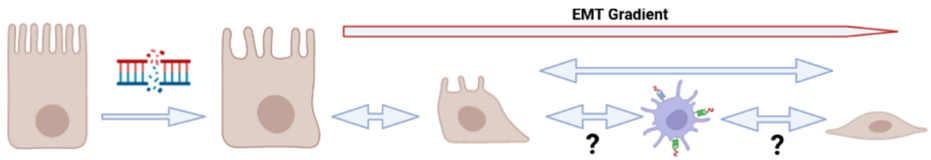
- forward(*GAPDH*) 5'-ACAACTTTGGTATCGTGGAAGG
- reverse(*GAPDH*) 5'-GCCATCACGCCACAGTTTC
- forward(*HMBS*) 5'-GGCAATGCGGCTGCAA
- reverse(*HMBS*) 5'-GGGTACCCACGCGAATCAC
- forward(*LINC01615*) 5'-ACACTGGTTGAGCAGACAGC
- reverse(*LINC01615*) 5'-CTGCCCCTCCTACCCCAG
- forward(*FN1*) 5'-TACAATGTGGGTCCCTCTGTC
- reverse(*FN1*) 5'-CTCTTGGTTGCCCTTTATGG

- forward(*TGFBI*) 5'-TGTAACTGTGAACTGTGCCCG
- reverse(*TGFBI*) 5'-CATCGTGTTGAGCCCTGATGC
- forward(*CD44*) 5'CGAATCCTGAAGACATCTACCC
- reverse(*CD44*) 5'-TGAAGTGCTGCTCCTTTCAC

Clustering of the TCGA-SKCM cohort

For analysis of TCGA-SKCM, I retrieved raw-counts and separated primary tumors and distant metastatis samples. The raw-counts matrices were normalized by sequencing depth using DESeq2 size-factors and then gene-counts were divided by median transcript length. I performed consensus clustering in R using the ConsensusClusterPlus v3.17 package[257] following standard procedure. In short, I filtered the matrices to keep only coding genes based on their biotype annotation, I selected the 5000 most variable genes with the mad() function and median centered the matrices with sweep(), apply() and median() functions before performing consensus clustering with ConsensusClusterPlus() using base parameters. I chose the appropriate number of clusters based on the curve of cumulative distribution function in order to define 4 clusters for primary tumors (CCP1-CCP4) and 5 clusters for distant metastasis samples (CCM1-CCM5). I used DESeq2 and GSEA as described before to characterize these clusters.

Chapter III: Concluding discussion



	Proximal Tubules	ccRCC (epithelial)	ccRCC (intermediate)	ccRCC (inflamed)	ccRCC (mesenchymal)
pVHL expression	Expressed	Lost	Lost	Lost	Lost
TP53 activity	High	None	None	None	None
FOXP4 activity	High	None	None	None	None
PPARG activity	High	None	None	None	None
Gluconeogenesis signature	High	High	High	High	High
OxPhos signature	High	Medium	Low	Very Low	Very Low
PAX8 expression	High	Medium	Low	Low	Very Low
Angiogenesis (VEGF) signature	None	High	High	High	High
Hypoxia response signature	None	Medium	High	High	High
MYC activity	None	High	High	High	High
CEBPD activity	None	High	High	High	High
ATF4 activity	None	High	High	High	High
AP-1 (JUNB/FOSB) expression	Low	Low	High	Medium	Low
EMT signature	None	Low	Medium	High	High
BHLHE41 activity	None	Low	High	High	High
POU (BRN1/OCT4) activity	None	Low	High	High	High
CEBPB activity	None	None	Medium	High	High
SNAI (SNAIL/SLUG) activity	None	None	None	Low	High
ZEB1 activity	None	None	None	None	High
ZEB2 expression	None	None	Medium	Medium	High
SOX4 activity	None	None	None	Low	High
SOX13 activity	None	None	None	None	High
SOX9 expression	High	Low	Low	None	Medium
PD-L1 correlation	None	None	None	None	High
IRF1 activity	Low	None	Low	High	None
IRF9 activity	None	None	Low	High	High
NFKB2 activity	Low	None	Low	High	Low
TNF-α signaling signature	Low	Low	High	High	Low

Figure 80. Summary of ccRCC transformation and dedifferentiation.

Table showing proximal tubules and ccRCC characteristics, features are colored by phenotype: epithelial (green), ccRCC (magenta), intermediate/mesenchymal (gray) and inflamed (purple). Transcription factor activity is inferred by SCENIC analysis. Gene signatures are defined from GSEA MSigDB for OxPhos, hypoxia response (which has major overlaps with EMT and glycolysis hallmarks) and TNFA signaling. Angiogenesis signature represents *VEGFA*, *VEGFB* and *VEGFC* expression. Gluconeogenesis level is inferred from VISION using the KEGG pathway signature. Illustration created with BioRender.

Chapter III: Concluding discussion

The main goal of my thesis was to characterize the ccRCC heterogeneity and TME landscapes in order to better understand mechanisms of malignant transformation, progression and resistance to treatment. This could be achieved by bioinformatics analyses integrating sequencing data from bulk and single-cell RNA-seq as well as spatial transcriptomics. In this section, I will discuss the main findings of this study and put them in perspective with the current knowledge of the field.

A. Malignant transformation of proximal tubules into ccRCC

Proximal tubules give rise to ccRCC following a combination of mutations always including loss of pVHL. Still, the question remains of why the cell of origin for this cancer is specifically a PT and not another cell type of the kidney nephron and what are the exact mechanisms leading to malignant transformation? My results comparing proximal tubule and ccRCC expression programs can help to form hypotheses (Figure 80).

The main hypothesis is that the pseudo-hypoxic state generated by HIF activation gives the pre-malignant cell a selective advantage. My analysis shows progressive loss of the OxPhos nuclear gene expression program along dedifferentiation of ccRCC, this program is inversely correlated with the EMT and hypoxia response programs. Despite HIF activation normally repressing OxPhos, ccRCC.epi shows intermediate levels of the OxPhos signature co-existing with the glycolysis signature indicating that both these pathways might be used by ccRCC cells in this state. The HIF- induced aerobic glycolysis produces the Warburg effect and the altered lipid metabolism provides components for genesis of new membranes. These two aspects support a high proliferation rate and neighboring non-transformed cells could be weakened by uptake of nutrients or release of byproducts from this new metabolism. These metabolic-related advantages are only part of the mechanism and do not explain the PT cell origin.

One possible explanation for this is that the location of the PT cell of origin is more important that the cell-type itself. We saw in RMC that the transformation of TAL cells occurs in specific conditions created by the hypoxic medulla and sickle-cell trait. Similarly, in ccRCC, there might exist a specific hotspot located alongside the PT network and close to the glomerulus where pseudo-hypoxic metabolism confers an advantage leading to oncogenesis.

Another hypothesis is that the type of cell is important which means some answers might lie in the characteristics of the PT cells and their transcriptional program. My analysis identifies loss of activity for kidney and proximal tubule epithelial regulators PAX8, PPARG, FOXP4 and TP53. These results indicate that loss of a tumor-suppressing TP53-driven program might be a requirement for ccRCC oncogenesis. As TP53 itself is not so often mutated in ccRCC, its activity is most likely hindered by loss of PBRM1[203] or SETD2[209] in $TP53^{wt}$ tumors. Escape from terminal differentiation seems like a pre-requisite for ccRCC oncogenesis as well. One previously introduced study identified a transcriptional hub governed by PAX8 and PBAF maintaining kidney cell identity and repressing a c-MYC oncogenic program[204] which

resembles our RMC model where TFCP2L1 and SMARCB1 also repress a c-MYC oncogenic program. In line with this ccRCC model, I found loss of renal transcription factor activity and activation of c-MYC in ccRCC cells. I also identified additional potential oncogenic factors such as CCAAT enhancer-binding protein (C/EBP) delta (*CEBPD*) and activating transcription factor 4 (*ATF4*). CEBPD expression has been found tied to c-MYC in urothelial carcinoma. In this setting it was shown to promote angiogenesis via VEGFA, enforce an aerobic glycolytic program and likely sustained proliferation through MAPK and mTOR signaling[258]. ATF4 is a stress response regulator that promotes adaptation to low nutrient conditions by dimerizing with ATF, C/EBP and AP-1 proteins. Although ATF4 can induce apoptosis, cancer cells were found to exploit its protection against oxidative stress and protein homeostasis functions to support their rapid proliferation. ATF4 intersects with mTOR and regulates a balance between autophagy and protein synthesis[259]. Also, some AP-1 subunits (JUN/FOS) were expressed in PSTs and could participate in ccRCC development as it has oncogenic properties[260]. On the other hand, PAX8 expression is progressively lost in ccRCC along with loss of FOXP4 activity, a factor that was previously found to be expressed in kidney tubules[261, 262] and might be part of the same transcriptional hub. In a similar manner, PPARG activity is lost in ccRCC, this factor has been shown to be expressed in proximal tubules and blocks EMT induced during renal fibrosis[263] suggesting it may be a key factor in maintaining kidney proximal identity. I observed loss of epithelial adhesion molecules and gain of CDH2 already in ccRCC.epi. This is indicating that an early-EMT event might be necessary for oncogenesis and it can only be enabled by loss of activity of kidney epithelial transcription factors. The capacity for early dedifferentiation could be one characteristic specific of the proximal epithelium that makes it prone to oncogenesis as it is observed during acute kidney injury and renal fibrosis[264]. Additionally, PTs are highly metabolic and fragile cells and, compared to other renal tubules, they could be more easily overtaken by a more "selfishly" energy-efficient and resilient transformed cell.

It is possible that not only one single hypothesis is true but that combination of multiple elements discussed above lead to ccRCC oncogenesis and some these hypotheses could apply to other RCC types deriving from proximal epithelial cells.

B. Dedifferentiation of ccRCC and disease progression

Cellular plasticity confers cancer cells the adaptability necessary for invasion, metastasis and resistance to treatment. My analyses uncovered the epithelial-mesenchymal plasticity of ccRCC cells but questions remain as to how and why cancer cells dedifferentiate and migrate to form distant metastasis?

a. Transcriptional control of EMT in ccRCC

I could profile four ccRCC phenotypes (epithelial-like, intermediate, inflamed and mesenchymal-like) fitting in an EMT gradient and identified several candidate factors with specific expression patterns potentially explaining how cancer cell adopt early and late

mesenchymal phenotypes (Figure 80). Similar cancer cell phenotypes were reported in a recent single-cell study of ccRCC tumors[265].

One of the early expressed factor is the stress-response regulator AP-1 of which some specific subunits (FOSB/JUNB) are found transiently up-regulated in the intermediate state, it could be part of an early invasion program as described in one ccRCC study[236]. BHLHE41 (DEC2) is a HIF target[266] that showed an activity in the intermediate state that was conserved in the mesenchymal state, one study showed its knock-down reduces invasion of ccRCC cell lines but the mechanism remain elusive[267].

I found the two POU-family members OCT4 and BRN1 to be expressed in the intermediate and mesenchymal states. OCT4 is a HIF-2α target that could be an early driver of dedifferentiation in ccRCC. BRN1 is an epithelial distal tubule regulator that is also involved in kidney development[157]. This factor was lost in RMC that originate from distal cells but in ccRCC its role might be rather similar to that of BRN2 in melanoma[68] and potentiate dedifferentiation as it was not found to be expressed in PTs. The POU family synergizes with the SOX factors, SOX4 and SOX13 were found active in ccRCC.mes along with moderate expression of SOX9. I found PST cells to express SOX9 as it is implicated in tissue repair by PTs following injury but its expression greatly diminishes in epithelial-like and intermediate ccRCC before increasing again in the mesenchymal-like state. This is perhaps indicating that SOX9 could play a role in ccRCC similar to the one it has in melanoma[268].

Another transcription factor following a gradual activation pattern across dedifferentiated states is CEBPB. This factor has multiple functions and was found to be implicated in invasion of ccRCC cells via an IL-1β/CEBPB/MMP axis[265] although it seems to have an opposite role in breast cancer depending on the isoform[269].

Lastly, I detected expression for well-known EMT regulators. ZEB1 activity was specific to ccRCC.mes and ZEB2 was found expressed in ccRCC.int/ccRCC.inf then greatly enriched in ccRCC.mes showing both ZEB factors to be involved in EMT of ccRCC cells. Similarly, SNAIL and SLUG showed activity and expression specific to ccRCC.mes indicating that the SNAI family is a regulator of full-EMT ccRCC cells. As HIF-1α is highly active in these cells, it might facilitate the action of these transcription factors.

b. Induction of EMT in ccRCC and metastatic spread

A great mystery is why some cancer cells fully dedifferentiate and migrate to form distant metastasis while others seem to keep epithelial and intermediate phenotypes. While some mutations are linked with increased disease aggressiveness, they are only a partial explanation and more of a predisposition as genetic mutations are unlikely to be reversible and distant metastatic cancer cells show a similar genetic background to that of the primary tumor[236,270]. I looked at the variety of different signals that could induce EMT in ccRCC to understand how interactions with the TME can influence invasion and metastasis of ccRCC cells.

One of the potential main drivers of ccRCC dedifferentiation is their interaction with fibroblasts. I found myCAF and ccRCC.mes to be highly correlated and spatial transcriptomics

revealed they co-localize at the tumor/NAT interface. Ligand-receptors interactions analyses identified several pathways by which myCAFs can induce EMT in ccRCC cells. One of the central pathways seems to be through secretion of TGF-β but other key axes were GAS6/AXL and JAG1/NOTCH. Interactions went both ways and TGF-β, IL-6, PKD1 and GAS6 expressed by ccRCC.mes can potentiate EMT of other ccRCC cells as well as induce the myCAF phenotype in resident fibroblastic cells. This poses a chicken and egg problem and it's difficult to know who comes first. Some receptor-ligand couples like GAS6/AXL are expressed on ccRCC.int and mesangial cells so it is possible that ccRCC cells from the primary tumor dedifferentiate when in contact with normal interstitial fibroblasts or pericytes then in turn converts them into activated CAFs. I found ccRCC.mes and myCAFs to be significantly enriched at metastatic sites, this might indicate that the migrating cell that establishes the metastatic niche had the ccRCC.mes phenotype and needs to convert the local fibroblasts into activated myCAFs for the metastasis to be viable (Figure 81). It is also a possibility that ccRCC.mes and myCAF migrate together but the later are untransformed and should not be able to repopulate a new site. We know tumor metastatic localizations are not random[271] and ccRCC has preferential metastasis sites for lung and liver[272]. One of the main reasons for this might be the fibroblastic environment and signaling pathways as both of these organs are also prone to tissue fibrosis[273].

Other potential EMT drivers in ccRCC are the multiple cytokines produced during inflammatory responses. The inflamed ccRCC cells had strong upregulation of TNF-α and interferon (α/γ) response genes with upregulation of MHC-II. These cells were poorly differentiated, even less so than ccRCC.int, but it was not possible to infer whether ccRCC.inf was a mandatory step in the EMT gradient and this phenotype was quite rare and never very abundant in the deconvolution analysis. Intriguingly, MHC-II expression should be detrimental to the tumor[274] yet ccRCC.inf is associated with worse patient prognosis. The undifferentiated nature of these cells points to a role of cytokines for EMT induction. TNF-α response was transiently expressed in ccRCC.int then lost in ccRCC.mes indicating that TNF-α which is produced by myeloid and lymphoid cells might be one of the early dedifferentiation signals. Interferon regulatory factor 1 (IRF1) was slightly active in ccRCC.int and IRF9 was highly active in ccRCC.mes pointing to type-I and type-II interferons as probable EMT-inducing signals. Interferon-γ is produced by cytotoxic lymphoid cells while interferon-α is produced by dendritic cells and is associated with metastasis of breast cancer[275]. Additionally, one study established IL-1β to be correlated with EMT of ccRCC cells, this factor is produced by TAMs establishing a role similar to CAFs for these cells[265].

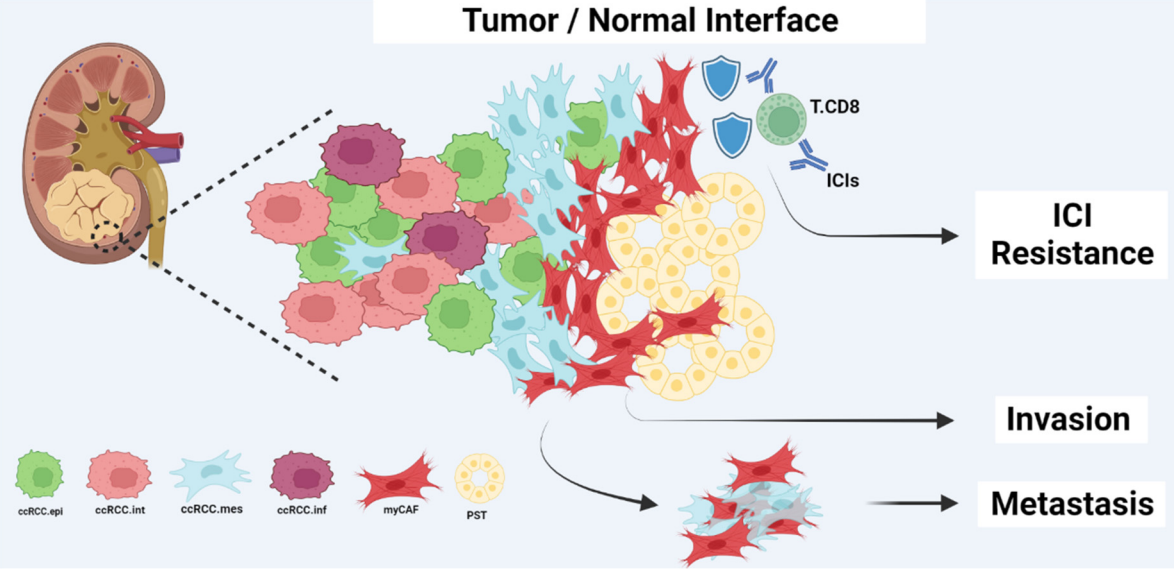


Figure 81. Cooperation of ccRCC.mes and myCAFs for invasion and ICI resistance.

Illustration made by Alexandra Helleux with BioRender.

c. Impact of EMT on tumor heterogeneity during disease progression

I analyzed cellular-composition changes across different tumor grades and stages. This result found that early grade/stage tumors had a more prominent epithelial component along with more vascularization displayed by greater proportions of endothelial and mesangial cells. On the other hand, higher grade/stage tumors were found enriched in mesenchymal populations (ccRCC.mes, ccRCC.inf and myCAFs) alongside enrichment of immune cells with diverse phenotypes including exhausted T-cells and TAMs as reported in the literature[230].

i. EMT and vascularization

The difference in vascularization can be surprising as undifferentiated ccRCC cells express higher levels of *VEGFA* but this difference was also observed in spatial-transcriptomics where spots from the lower-grade showed higher endothelial signatures compared with spots from the higher-grade area of the same tumor sample. One possible explanation is that endothelial cells could become less responsive to VEGF after prolonged exposure or angiogenesis could be attenuated by other signals leading to lesser endothelial vascularization of advanced ccRCC tumors. This phenomenon could be compensated by vasculogenic mimicry, this property of undifferentiated cancer cells to form their own blood vessels was observed in multiple cancers including ccRCC[276,277]. This could also have implications for treatment resistance as drugs are less likely to be properly distributed inside these tumors and mesenchymal cancer cells are known to resist treatment by mechanisms such as membrane transporters that pump drugs out of the cell[278]. This seems to be in line with observations that the "pro-angiogenic" ccRCC2 group in the BIONIKK classification responds well to TKIs compared to less differentiated groups[242]. My analyses indicate that the increased expression

observed at the bulk-RNA level of this angiogenic signature in ccRCC2 samples seems due to an increase in endothelial and mesangial populations that moderately express VEGF and not to a change of VEGF expression in cancer cells.

ii. EMT and immune evasion

Advanced tumors were progressively enriched in immune cells together with undifferentiated cancer cells further pointing to inflammatory signals as promoters of EMT but also indicating these cancer cells might be well suited at eluding the immune system. To study this question, I looked into immune-escape mechanisms that could be employed by ccRCC cells. CAFs and TAMs already have mechanisms described in the literature and ligand-receptor analyses of ccRCC cells pointed towards MIF, SPP1 and FAM3C as important immunosuppressive ligands that could be targeted in this cancer type (Figure 82). These analyses indicate that epithelial-like and intermediate ccRCC cells evade the immune system via expression of SPP1 that dampens cytotoxicity[279] and via recruitment and polarization of TAMs through the MIF axis. On the other hand, ccRCC.mes seem to be able to directly stimulate PD-1 via FAM3C or PD-L1 which is positively correlated with ccRCC.mes proportion and these characteristics are shared with myCAFs. We observe that ccRCC employs TAMs and CAFs to create an immunosuppressive environment but autonomous stimulation of PD-1 by mesenchymal-like ccRCC could be of particular importance. It has been suspected that the immune system dictates tumor development and metastasis in part through clonal selection[280,281] so this mechanism of immune evasion by undifferentiated ccRCC cells could be a prerequisite for disease progression and even become a dependency for certain tumors. Such a dependency could be one of the reasons for the better clinical response rate observed in tumors with sarcomatoid components treated with ICIs[232].

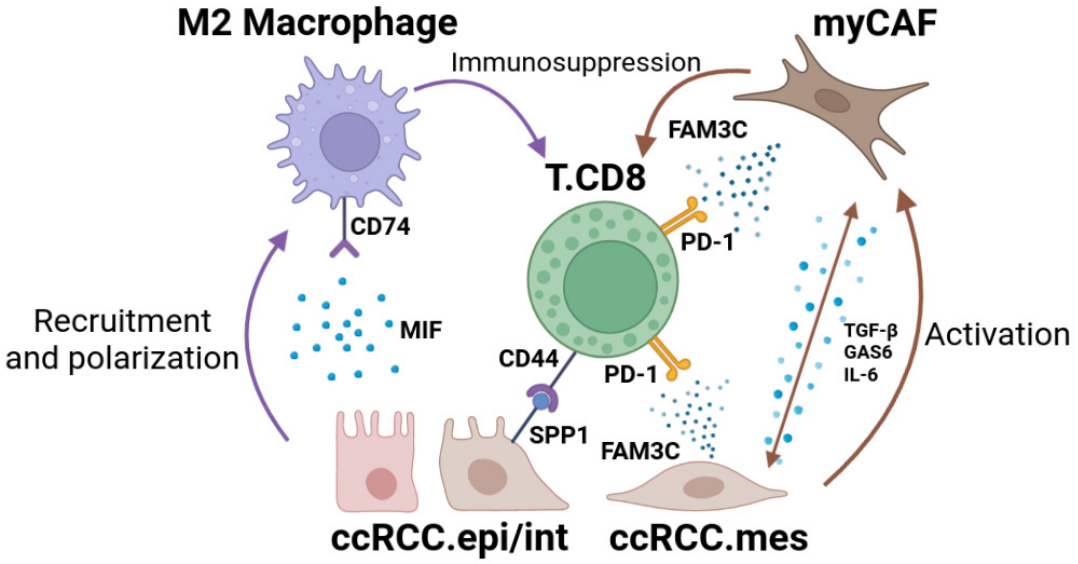


Figure 82. Immune evasion in ccRCC.

Illustration created with BioRender.

C. Using myCAFs as biomarkers for immunotherapy resistance

Immunotherapy has improved prognosis for patients presenting metastatic ccRCC but a majority of patients still do not respond to these treatments so it is critical to identify mechanisms of resistance and biomarkers for this resistance are needed to select patients appropriately[223]. One of the key results of my thesis work is the association of the proportion of myCAFs with poor survival in patients treated by ICIs in the BIONIKK cohort. This relationship appeared to be non-linear as it was dependent on a threshold to stratify patients into myCAF-high and myCAF-low. There are several ways myCAFs may modulate the immune response. As resistance to ICIs required a high amount of myCAF infiltration it is possible that they oppose a physical barrier to the treatment or to re-activated immune cells but the exact mechanism remain elusive. Irrespective of the mechanisms involved, targeting myCAFs could potentiate the effects of immunotherapy in ccRCC. As myCAFs also form a capsule that contains the tumor, targeting the myCAF component would also require a strong anti-cancer agent to avoid increasing the dissemination of cancer cells as observed in mice[111]. Such an approach showed success in pancreatic cancer mouse models by combining PIN-1 and PD-L1 inhibition[282]. Interestingly, recently tested combinations of ICIs with TKIs showed improvement of patient survival[216]. These TKIs target RTKs involved in myCAF communication axes such as PDGFR, EGFR and AXL indicating that maybe such combination could work better on myCAF-high patients than ICB alone.

I explored biomarkers linked to myCAFs that could predict response to ICIs. Using myCAF marker genes that were highly expressed in myCAF-high BIONIKK samples, I could derive two small 5 and 10 genes signatures that were significantly associated with overall survival of ICI-treated patients. These two signatures could be measured by RTqPCR in order to select patients before treatment. By using unsupervised clustering of the BIONIKK cohort according to lincRNAs, one cluster (LINC5) showed significant enrichment of myCAF markers and correlated with poor prognosis. This shows that epigenetic profiling can reflect tumor composition with enough accuracy to be used for prognosis and other epigenetic markers such as other categories of lncRNAs or DNA methylation could maybe achieve the same result. I identified the LINC5 marker SFTA1P as a lncRNA expressed in mesenchymal cells that correlates with poor survival in ICI-treated patients. This gene has been shown to be regulated by the mesenchymal and muscle regulator TEAD family of transcription factors[283] and could represent a biomarker for myCAF presence that predicts response to ICI.

D. Using LINC01615 as a marker of the mesenchymal state in cancer cells

My analysis of LINC5 markers uncovered two lincRNAs with expression specific to ccRCC cells in single-cell data. LINC01638 was only expressed in intermediate and fully undifferentiated ccRCC clusters and was found related to the Wnt/β-catenin[284] and TGF- β[285] signaling pathways. This gene could serve as a broad ccRCC marker and constitute a potential therapeutic target linked with dedifferentiation of cancer cells.

LIN01615 expression was very specific to ccRCC.mes in single-cell data. This gene is detectable at the bulk-RNA level and was much enriched in ccRCC tumor samples compared

to NATs. I observed a strong association between the proportion of undifferentiated ccRCC cells (ccRCC.mes, ccRCC.inf) and myCAF with poor patient survival. Combining information of these 3 populations into one score (MES.score) was able to efficiently predict overall survival though it showed a tendency to override tumor grade in multiple regression analyses indicating collinearity between these two variables. As MES.score was inferred from CIBERSORTx analyses, it is not easily measured and histological assessment of tumor grade might be more suited for practical use. On the other hand, LINC01615 expression was an even stronger predictor of patient survival (HR=8.24, p=0.004) and measurement of this gene could be used in clinical practice as an indicator of ccRCC.mes presence. I explored potential regulators of LINC01615 in ccRCC by looking at co-expressed genes and predicted binding motifs present in its promoter. These analyses pointed to BRN2, AP-1, SOX, MYC and CEBPB as potentially regulating LINC01615 expression since these factors showed either expression or activity in the ccRCC.mes cluster.

I explored if LINC01615 could also play a role in melanoma. I found RNA-seq expression of this gene to be associated with undifferentiated cellular models and a validation RTqPCR experiment showed it to be detectable and greatly enriched in MM099 compared to melanocytic 501M cells. I checked in-vivo expression patterns using the TCGA-SKCM cohort which confirmed association of LINC01615 with mesenchymal tumors both in primary and metastatic samples. In melanoma, it is not as straightforward as in ccRCC to use LINC01615 expression for patient prognosis as distant metastasis samples with a melanocytic signature displayed a worse survival. However, melanoma undifferentiated states are associated with minimal residual disease[286] and suspected to be involved in resistance to ICB[287] so measuring LINC01615 expression can have useful applications for this cancer type.

LINC01615 has been reported in hepatocellular carcinoma[288], head and neck squamous cell carcinoma[289] and stomach adenocarcinoma[290]. In these settings, the gene was associated with poor prognosis and metastasis which was supported by a study linking LINC01615 to EMT[291]. Together with my findings in ccRCC that were backed by one study[292] and my results in melanoma, LINC01615 appears to be an important EMT-related lincRNA with pan-cancer applications.

E. Parallels between pathogenesis of RMC and ccRCC

Results in RMC indicated that loss of SMARCB1 affects TFCP2L1 activity and allows oncogenic activation of c-MYC. As discussed previously, it is interesting to note that ccRCC which is frequently mutated in SWI/SNF subunit PBRM1 showed activation of c-MYC in my analyses. Together, these observations highlight the roles of SWI/SNF in maintenance of PT/TAL epithelial identity and the roles of c-MYC activation in the transformed state that could perhaps be defining a more general model of RCC oncogenesis. My analysis of RMC also found abundant CAF and TAM populations forming an immunosuppressive TME. The presence of these two cell types could be a recurrent factor for disease progression of all RCCs. Based on my results in ccRCC, the abundance of myCAFs in RMC sequencing data together with the known desmoplastic nature of these tumors indicate that they should be poorly responsive to ICIs. The first study (NCT03274258) evaluating nivolumab+ipilimumab in RMC is no longer recruiting patients so depending on the results from a second clinical trial

(NCT05347212) that is evaluating nivolumab+relatimab (anti-LAG3), combination of ICIs with drugs that also target the immunosuppressive component of the TME could be considered.

F. Future perspectives

Results from my thesis open up different prospects. I highlighted the importance of epithelial-mesenchymal plasticity in ccRCC but mechanistic studies and use of the latest sequencing technologies such as 10x Xenium and nanoString GeoMx Digital Spatial Profiler could help to gain valuable insights on this phenomenon. Many ccRCC cellular models exist, their phenotypes should be elucidated so they can be used to study the functions of epithelial and mesenchymal transcription factors. It is also possible to influence ccRCC cell lines differentiation level by interferon-γ or TGF-β treatments in order to conduct these studies. Forced expression of kidney epithelial regulators such as PAX8 and PPARG in these models could shed some light on the oncogenic process. Roles of mesenchymal factors such as those from the POU/SOX/ZEB/SNAI families should be studied in ccRCC to understand the dedifferentiation process and discover ways to prevent metastatic spread. RCC tumors areas with sarcomatoid and rhabdoid phenotypes should be sequenced to understand the particularities of their dedifferentiation. The interactions between ccRCC/myCAF could be studied using co-culture or by exposing primary fibroblasts to supernatants of ccRCC cells to induce their activation. Then, selective inhibition of different receptors could pinpoint the crucial interaction pathways involved. The reverse experiments, exposing ccRCC to CAF supernatant or co-cultures could be performed to study CAF-induced dedifferentiation of ccRCC. All studies on dedifferentiation in ccRCC could have applications relevant for kidney fibrosis as well.

Considering the high resistance rate to immunotherapy, selecting patients appropriately and circumventing this resistance is crucial. The relationship between myCAF levels and ICI resistance should be validated in bigger cohorts than the BIONIKK one, also this relationship should be studied for newer ICI/TKI combinations. The 5 and 10 genes myCAF signatures and SFTA1P expression should be validated and tested by RTqPCR to potentially be used for patient selection. Other epigenetic classification methods that can reflect tumor composition such as DNA methylation should be explored as well. If drug combinations that target CAF and cancer cells show efficacy in other cancers then they should be considered for ccRCC.

Lastly, LINC01615 should be functionally studied in ccRCC and melanoma. Impairing its activity could hinder EMT of cancer cells and this could apply to any cancer expressing this gene. Additionally, its expression could be measured in clinical practice to predict survival or for other uses depending on what we will discover in the future about the roles of mesenchymal cancer cells.

Bibliography

Chapter IV: Bibliography

1. References

1. Mattick, J. and P. Amaral, *RNA, the Epicenter of Genetic Information: A new understanding of molecular biology*. 2022: CRC Press.
2. Randolph-Quinney, P.S., et al., *Osteogenic tumour in Australopithecus sediba: Earliest hominin evidence for neoplastic disease.* South African Journal of Science, 2016. **112**(7/8): p. 7.
3. Ekhtiari, S., et al., *First case of osteosarcoma in a dinosaur: a multimodal diagnosis.* Lancet Oncol, 2020. **21**(8): p. 1021-1022.
4. Hassanpour, S.H. and M. Dehghani, *Review of cancer from perspective of molecular.* Journal of Cancer Research and Practice, 2017. **4**(4): p. 127-129.
5. Wapner, J., *The Philadelphia chromosome: a mutant gene and the quest to cure cancer at the genetic level*. 2013, The Experiment, LLC.
6. Hanahan, D., *Hallmarks of Cancer: New Dimensions.* Cancer Discov, 2022. **12**(1): p. 31-46.
7. Williams, E.D., et al., *Controversies around epithelial-mesenchymal plasticity in cancer metastasis.* Nat Rev Cancer, 2019. **19**(12): p. 716-732.
8. Noubissi Nzeteu, G.A., et al., *Role of Epithelial-to-Mesenchymal Transition for the Generation of Circulating Tumors Cells and Cancer Cell Dissemination.* Cancers (Basel), 2022. **14**(22).
9. Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition.* J Clin Invest, 2009. **119**(6): p. 1420-8.
10. Muhr, J., *Embryology, Gastrulation*. 2023: StatPearls Publishing.
11. Henderson, N.C., F. Rieder, and T.A. Wynn, *Fibrosis: from mechanisms to medicines.* Nature, 2020. **587**(7835): p. 555-566.
12. Kalluri, R. and E.G. Neilson, *Epithelial-mesenchymal transition and its implications for fibrosis.* J Clin Invest, 2003. **112**(12): p. 1776-84.
13. Sistigu, A., et al., *Deciphering the loop of epithelial-mesenchymal transition, inflammatory cytokines and cancer immunoediting.* Cytokine Growth Factor Rev, 2017. **36**: p. 67-77.
14. Sinha, D., et al., *Emerging Concepts of Hybrid Epithelial-to-Mesenchymal Transition in Cancer Progression.* Biomolecules, 2020. **10**(11).
15. Meigs, T.E., et al., *Galpha12 and Galpha13 negatively regulate the adhesive functions of cadherin.* J Biol Chem, 2002. **277**(27): p. 24594-600.
16. Lamouille, S., J. Xu, and R. Derynck, *Molecular mechanisms of epithelial-mesenchymal transition.* Nat Rev Mol Cell Biol, 2014. **15**(3): p. 178-96.
17. Toivola, D.M., et al., *Cellular integrity plus: organelle-related and protein-targeting functions of intermediate filaments.* Trends Cell Biol, 2005. **15**(11): p. 608-17.
18. Theveneau, E. and R. Mayor, *Cadherins in collective cell migration of mesenchymal cells.* Curr Opin Cell Biol, 2012. **24**(5): p. 677-84.
19. Hansen, S.M., V. Berezin, and E. Bock, *Signaling mechanisms of neurite outgrowth induced by the cell adhesion molecules NCAM and N-cadherin.* Cell Mol Life Sci, 2008. **65**(23): p. 3809-21.
20. Hassn Mesrati, M., et al., *CD44: A Multifunctional Mediator of Cancer Progression.* Biomolecules, 2021. **11**(12).

21. Mendez, M.G., S. Kojima, and R.D. Goldman, *Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition.* FASEB J, 2010. **24**(6): p. 1838-51.
22. Nistico, P., M.J. Bissell, and D.C. Radisky, *Epithelial-mesenchymal transition: general principles and pathological relevance with special emphasis on the role of matrix metalloproteinases.* Cold Spring Harb Perspect Biol, 2012. **4**(2).
23. Liu, J., *Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target.* Discovery Medicine, 2018. **25**(139): p. 211-223.
24. Feng, J. and L. Tang, *SPARC in Tumor Pathophysiology and as a Potential Therapeutic Target.* Curr Pharm Des, 2014. **20**(39): p. 6182-90.
25. Rostagno, A., *Further characterization of the NH2-terminal fibrin-binding site on fibronectin.* The Journal of Biological Chemistry, 1994. **269**(50): p. 31938-31945.
26. Qureshi, I.A., S. Gokhan, and M.F. Mehler, *REST and CoREST are transcriptional and epigenetic regulators of seminal neural fate decisions.* Cell Cycle, 2010. **9**(22): p. 4477-86.
27. Lin, Y., et al., *The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1.* EMBO J, 2010. **29**(11): p. 1803-16.
28. Tran, H.D., et al., *Transient SNAIL1 expression is necessary for metastatic competence in breast cancer.* Cancer Res, 2014. **74**(21): p. 6330-40.
29. Yang, F., et al., *SET8 promotes epithelial-mesenchymal transition and confers TWIST dual transcriptional activities.* EMBO J, 2012. **31**(1): p. 110-23.
30. Cakouros, D., et al., *Twist-1 induces Ezh2 recruitment regulating histone methylation along the Ink4A/Arf locus in mesenchymal stem cells.* Mol Cell Biol, 2012. **32**(8): p. 1433-41.
31. Yang, M.H., et al., *Bmi1 is essential in Twist1-induced epithelial-mesenchymal transition.* Nat Cell Biol, 2010. **12**(10): p. 982-92.
32. Lu, W. and Y. Kang, *Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis.* Dev Cell, 2019. **49**(3): p. 361-374.
33. Krebs, A.M., et al., *The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer.* Nat Cell Biol, 2017. **19**(5): p. 518-529.
34. Caramel, J., et al., *A switch in the expression of embryonic EMT-inducers drives the development of malignant melanoma.* Cancer Cell, 2013. **24**(4): p. 466-80.
35. Joshi, S., et al., *TEAD transcription factors are required for normal primary myoblast differentiation in vitro and muscle regeneration in vivo.* PLoS Genet, 2017. **13**(2): p. e1006600.
36. Huh, H.D., et al., *Regulation of TEAD Transcription Factors in Cancer Biology.* Cells, 2019. **8**(6).
37. Lee, K.W., et al., *PRRX1 is a master transcription factor of stromal fibroblasts for myofibroblastic lineage progression.* Nat Commun, 2022. **13**(1): p. 2793.
38. Du, W., et al., *The Regulatory Role of PRRX1 in Cancer Epithelial-Mesenchymal Transition.* Onco Targets Ther, 2021. **14**: p. 4223-4229.
39. Mehta, G.A., P. Khanna, and M.L. Gatza, *Emerging Role of SOX Proteins in Breast Cancer Development and Maintenance.* J Mammary Gland Biol Neoplasia, 2019. **24**(3): p. 213-230.
40. Martinelli, P., et al., *GATA6 regulates EMT and tumour dissemination, and is a marker of response to adjuvant chemotherapy in pancreatic cancer.* Gut, 2017. **66**(9): p. 1665-1676.
41. Katoh, M., et al., *Cancer genetics and genomics of human FOX family genes.* Cancer Lett, 2013. **328**(2): p. 198-206.

42. Hao, Y., D. Baker, and P. Ten Dijke, *TGF-beta-Mediated Epithelial-Mesenchymal Transition and Cancer Metastasis.* Int J Mol Sci, 2019. **20**(11).
43. Liu, R.Y., et al., *JAK/STAT3 signaling is required for TGF-beta-induced epithelial-mesenchymal transition in lung cancer cells.* Int J Oncol, 2014. **44**(5): p. 1643-51.
44. Al-Ismaeel, Q., et al., *ZEB1 and IL-6/11-STAT3 signalling cooperate to define invasive potential of pancreatic cancer cells via differential regulation of the expression of S100 proteins.* Br J Cancer, 2019. **121**(1): p. 65-75.
45. Wu, Y., et al., *Stabilization of snail by NF-kappaB is required for inflammation-induced cell migration and invasion.* Cancer Cell, 2009. **15**(5): p. 416-28.
46. Lv, N., et al., *Inflammatory mediators, tumor necrosis factor-alpha and interferon-gamma, induce EMT in human PTC cell lines.* Oncol Lett, 2015. **10**(4): p. 2591-2597.
47. Paszek, M.J., et al., *Tensional homeostasis and the malignant phenotype.* Cancer Cell, 2005. **8**(3): p. 241-54.
48. Wei, S.C., et al., *Matrix stiffness drives epithelial-mesenchymal transition and tumour metastasis through a TWIST1-G3BP2 mechanotransduction pathway.* Nat Cell Biol, 2015. **17**(5): p. 678-88.
49. Griggs, L.A., et al., *Fibronectin fibrils regulate TGF-beta1-induced Epithelial-Mesenchymal Transition.* Matrix Biol, 2017. **60-61**: p. 157-175.
50. Radisky, D.C., et al., *Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability.* Nature, 2005. **436**(7047): p. 123-7.
51. Li, J., et al., *Chemoresistance to doxorubicin induces epithelial-mesenchymal transition via upregulation of transforming growth factor beta signaling in HCT116 colon cancer cells.* Mol Med Rep, 2015. **12**(1): p. 192-8.
52. Luond, F., et al., *Distinct contributions of partial and full EMT to breast cancer malignancy.* Dev Cell, 2021. **56**(23): p. 3203-3221 e11.
53. Jemal, A., et al., *Recent trends in cutaneous melanoma incidence and death rates in the United States, 1992-2006.* J Am Acad Dermatol, 2011. **65**(5 Suppl 1): p. S17-25 e1-3.
54. Eddy, K., R. Shah, and S. Chen, *Decoding Melanoma Development and Progression: Identification of Therapeutic Vulnerabilities.* Front Oncol, 2020. **10**: p. 626129.
55. Rastrelli, M., et al., *Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification.* In Vivo, 2014. **28**(6): p. 1005-11.
56. Cancer Genome Atlas, N., *Genomic Classification of Cutaneous Melanoma.* Cell, 2015. **161**(7): p. 1681-96.
57. Burotto, M., et al., *The MAPK pathway across different malignancies: a new perspective.* Cancer, 2014. **120**(22): p. 3446-56.
58. Viros, A., et al., *Ultraviolet radiation accelerates BRAF-driven melanomagenesis by targeting TP53.* Nature, 2014. **511**(7510): p. 478-482.
59. Wennerberg, K., K.L. Rossman, and C.J. Der, *The Ras superfamily at a glance.* J Cell Sci, 2005. **118**(Pt 5): p. 843-6.
60. Williams, E.A., et al., *Melanoma with in-frame deletion of MAP2K1: a distinct molecular subtype of cutaneous melanoma mutually exclusive from BRAF, NRAS, and NF1 mutations.* Mod Pathol, 2020. **33**(12): p. 2397-2406.
61. Rambow, F., J.C. Marine, and C.R. Goding, *Melanoma plasticity and phenotypic diversity: therapeutic barriers and opportunities.* Genes Dev, 2019. **33**(19-20): p. 1295-1318.
62. Pedri, D., et al., *Epithelial-to-mesenchymal-like transition events in melanoma.* FEBS J, 2022. **289**(5): p. 1352-1368.
63. Goding, C.R. and H. Arnheiter, *MITF-the first 25 years.* Genes Dev, 2019. **33**(15-16): p. 983-1007.

64. Ennen, M., et al., *Single-cell gene expression signatures reveal melanoma cell heterogeneity.* Oncogene, 2015. **34**(25): p. 3251-63.
65. Ennen, M., et al., *MITF-High and MITF-Low Cells and a Novel Subpopulation Expressing Genes of Both Cell States Contribute to Intra- and Intertumoral Heterogeneity of Primary Melanoma.* Clin Cancer Res, 2017. **23**(22): p. 7097-7107.
66. Karras, P., et al., *A cellular hierarchy in melanoma uncouples growth and metastasis.* Nature, 2022. **610**(7930): p. 190-198.
67. Shirley, S.H., et al., *Slug expression during melanoma progression.* Am J Pathol, 2012. **180**(6): p. 2479-89.
68. Fane, M.E., et al., *BRN2, a POUerful driver of melanoma phenotype switching and metastasis.* Pigment Cell Melanoma Res, 2019. **32**(1): p. 9-24.
69. Zhu, C., Y. Wei, and X. Wei, *AXL receptor tyrosine kinase as a promising anti-cancer approach: functions, molecular mechanisms and clinical applications.* Mol Cancer, 2019. **18**(1): p. 153.
70. Wouters, J., et al., *Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma.* Nat Cell Biol, 2020. **22**(8): p. 986-998.
71. Maniotis, A.J., et al., *Vascular channel formation by human melanoma cells in vivo and in vitro: vasculogenic mimicry.* Am J Pathol, 1999. **155**(3): p. 739-52.
72. Li, X., et al., *Disseminated Melanoma Cells Transdifferentiate into Endothelial Cells in Intravascular Niches at Metastatic Sites.* Cell Rep, 2020. **31**(11): p. 107765.
73. Frankish, A., et al., *Gencode 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D916-D923.
74. Ponting, C.P., P.L. Oliver, and W. Reik, *Evolution and functions of long noncoding RNAs.* Cell, 2009. **136**(4): p. 629-41.
75. Schorderet, P. and D. Duboule, *Structural and functional differences in the long non-coding RNA hotair in mouse and human.* PLoS Genet, 2011. **7**(5): p. e1002071.
76. Wu, Y., et al., *Long noncoding RNA HOTAIR involvement in cancer.* Tumour Biol, 2014. **35**(10): p. 9531-8.
77. Cifuentes-Rojas, C., et al., *Regulatory interactions between RNA and polycomb repressive complex 2.* Mol Cell, 2014. **55**(2): p. 171-85.
78. Peng, J.C., et al., *Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells.* Cell, 2009. **139**(7): p. 1290-302.
79. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes.* Science, 2010. **329**(5992): p. 689-93.
80. Rajagopal, T., et al., *HOTAIR LncRNA: A novel oncogenic propellant in human cancer.* Clin Chim Acta, 2020. **503**: p. 1-18.
81. Gambi, G., et al., *The LncRNA LENOX Interacts with RAP2C to Regulate Metabolism and Promote Resistance to MAPK Inhibition in Melanoma.* Cancer Res, 2022. **82**(24): p. 4555-4570.
82. Zander, R., et al., *CD4(+) T Cell Help Is Required for the Formation of a Cytolytic CD8(+) T Cell Subset that Protects against Chronic Infection and Cancer.* Immunity, 2019. **51**(6): p. 1028-1042 e4.
83. Dolina, J.S., et al., *CD8(+) T Cell Exhaustion in Cancer.* Front Immunol, 2021. **12**: p. 715234.
84. Neumann, H., et al., *Cytotoxic T lymphocytes in autoimmune and degenerative CNS diseases.* Trends Neurosci, 2002. **25**(6): p. 313-9.
85. Zhou, Z., et al., *Granzyme A from cytotoxic lymphocytes cleaves GSDMB to trigger pyroptosis in target cells.* Science, 2020. **368**(6494).
86. Martinez-Lostao, L., A. Anel, and J. Pardo, *How Do Cytotoxic Lymphocytes Kill Cancer Cells?* Clin Cancer Res, 2015. **21**(22): p. 5047-56.

87. Wang, W., et al., *CD8(+) T cells regulate tumour ferroptosis during cancer immunotherapy.* Nature, 2019. **569**(7755): p. 270-274.
88. Weigelin, B., et al., *Cytotoxic T cells are able to efficiently eliminate cancer cells by additive cytotoxicity.* Nat Commun, 2021. **12**(1): p. 5217.
89. Weulersse, M., et al., *Eomes-Dependent Loss of the Co-activating Receptor CD226 Restrains CD8(+) T Cell Anti-tumor Functions and Limits the Efficacy of Cancer Immunotherapy.* Immunity, 2020. **53**(4): p. 824-839 e10.
90. Kumar, V. and M.E. McNerney, *A new self: MHC-class-I-independent natural-killer-cell self-tolerance.* Nat Rev Immunol, 2005. **5**(5): p. 363-74.
91. Wu, S.Y., et al., *Natural killer cells in cancer biology and therapy.* Mol Cancer, 2020. **19**(1): p. 120.
92. Oh, D.Y. and L. Fong, *Cytotoxic CD4(+) T cells in cancer: Expanding the immune effector toolbox.* Immunity, 2021. **54**(12): p. 2701-2711.
93. Saravia, J., N.M. Chapman, and H. Chi, *Helper T cell differentiation.* Cell Mol Immunol, 2019. **16**(7): p. 634-643.
94. Speiser, D.E., et al., *CD4(+) T cells in cancer.* Nat Cancer, 2023. **4**(3): p. 317-329.
95. Siska, P.J. and J.C. Rathmell, *T cell metabolic fitness in antitumor immunity.* Trends Immunol, 2015. **36**(4): p. 257-64.
96. Wynn, T.A., A. Chawla, and J.W. Pollard, *Macrophage biology in development, homeostasis and disease.* Nature, 2013. **496**(7446): p. 445-55.
97. Boutilier, A.J. and S.F. Elsawa, *Macrophage Polarization States in the Tumor Microenvironment.* Int J Mol Sci, 2021. **22**(13).
98. Sica, A. and A. Mantovani, *Macrophage plasticity and polarization: in vivo veritas.* J Clin Invest, 2012. **122**(3): p. 787-95.
99. DeNardo, D.G. and B. Ruffell, *Macrophages as regulators of tumour immunity and immunotherapy.* Nat Rev Immunol, 2019. **19**(6): p. 369-382.
100. Lee, Y.S. and K.J. Radford, *The role of dendritic cells in cancer.* Int Rev Cell Mol Biol, 2019. **348**: p. 123-178.
101. Engelhard, V., et al., *B cells and cancer.* Cancer Cell, 2021. **39**(10): p. 1293-1296.
102. Hedrick, C.C. and I. Malanchi, *Neutrophils in cancer: heterogeneous and multifaceted.* Nat Rev Immunol, 2022. **22**(3): p. 173-187.
103. Aponte-Lopez, A. and S. Munoz-Cruz, *Mast Cells in the Tumor Microenvironment.* Adv Exp Med Biol, 2020. **1273**: p. 159-173.
104. Koliaraki, V., et al., *The mesenchymal context in inflammation, immunity and cancer.* Nat Immunol, 2020. **21**(9): p. 974-982.
105. Geng, X., et al., *Cancer-Associated Fibroblast (CAF) Heterogeneity and Targeting Therapy of CAFs in Pancreatic Cancer.* Front Cell Dev Biol, 2021. **9**: p. 655152.
106. Galbo, P.M., Jr., X. Zang, and D. Zheng, *Molecular Features of Cancer-associated Fibroblast Subtypes and their Implication on Cancer Pathogenesis, Prognosis, and Immunotherapy Resistance.* Clin Cancer Res, 2021. **27**(9): p. 2636-2647.
107. Kennel, K.B., et al., *Cancer-Associated Fibroblasts in Inflammation and Antitumor Immunity.* Clin Cancer Res, 2023. **29**(6): p. 1009-1016.
108. Biffi, G., et al., *IL1-Induced JAK/STAT Signaling Is Antagonized by TGFbeta to Shape CAF Heterogeneity in Pancreatic Ductal Adenocarcinoma.* Cancer Discov, 2019. **9**(2): p. 282-301.
109. Ohlund, D., et al., *Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer.* J Exp Med, 2017. **214**(3): p. 579-596.
110. Hamson, E.J., et al., *Understanding fibroblast activation protein (FAP): substrates, activities, expression and targeting for cancer therapy.* Proteomics Clin Appl, 2014. **8**(5-6): p. 454-63.

111. Rhim, A.D., et al., *Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma.* Cancer Cell, 2014. **25**(6): p. 735-47.
112. Calon, A., et al., *Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation.* Cancer Cell, 2012. **22**(5): p. 571-84.
113. Padua, D., et al., *TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4.* Cell, 2008. **133**(1): p. 66-77.
114. Erdogan, B., et al., *Cancer-associated fibroblasts promote directional cancer cell migration by aligning fibronectin.* J Cell Biol, 2017. **216**(11): p. 3799-3816.
115. Kieffer, Y., et al., *Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer.* Cancer Discov, 2020. **10**(9): p. 1330-1351.
116. Chen, L., et al., *IL-6 influences the polarization of macrophages and the formation and growth of colorectal tumor.* Oncotarget, 2018. **9**(25): p. 17443-17454.
117. Elyada, E., et al., *Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts.* Cancer Discov, 2019. **9**(8): p. 1102-1123.
118. Rodriguez, A.B., et al., *Immune mechanisms orchestrate tertiary lymphoid structures in tumors via cancer-associated fibroblasts.* Cell Rep, 2021. **36**(3): p. 109422.
119. Huang, H., et al., *Mesothelial cell-derived antigen-presenting cancer-associated fibroblasts induce expansion of regulatory T cells in pancreatic cancer.* Cancer Cell, 2022. **40**(6): p. 656-673 e7.
120. De Palma, M., D. Biziato, and T.V. Petrova, *Microenvironmental regulation of tumour angiogenesis.* Nat Rev Cancer, 2017. **17**(8): p. 457-474.
121. Blancas, A.A., et al., *Specialized tip/stalk-like and phalanx-like endothelial cells from embryonic stem cells.* Stem Cells Dev, 2013. **22**(9): p. 1398-407.
122. Hida, K., et al., *Contribution of Tumor Endothelial Cells in Cancer Progression.* Int J Mol Sci, 2018. **19**(5).
123. Sobierajska, K., et al., *Endothelial Cells in the Tumor Microenvironment.* Adv Exp Med Biol, 2020. **1234**: p. 71-86.
124. Armulik, A., G. Genove, and C. Betsholtz, *Pericytes: developmental, physiological, and pathological perspectives, problems, and promises.* Dev Cell, 2011. **21**(2): p. 193-215.
125. Birbrair, A., et al., *Type-2 pericytes participate in normal and tumoral angiogenesis.* Am J Physiol Cell Physiol, 2014. **307**(1): p. C25-38.
126. Dasgupta, S., et al., *RGS5-TGFbeta-Smad2/3 axis switches pro- to anti-apoptotic signaling in tumor-residing pericytes, assisting tumor growth.* Cell Death Differ, 2021. **28**(11): p. 3052-3076.
127. Delon, C., et al., *Differences in cancer incidence by broad ethnic group in England, 2013-2017.* Br J Cancer, 2022. **126**(12): p. 1765-1773.
128. Gray, R.E., *Renal Cell Carcinoma: Diagnosis and Management.* American Family Physician, 2019. **99**(3): p. 179-184.
129. Rizzo, M., et al., *Clinico-pathological implications of the 2022 WHO Renal Cell Carcinoma classification.* Cancer Treat Rev, 2023. **116**: p. 102558.
130. Rini, B.I., S.C. Campbell, and B. Escudier, *Renal cell carcinoma.* Lancet, 2009. **373**(9669): p. 1119-32.
131. Ricketts, C.J., et al., *The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma.* Cell Rep, 2018. **23**(1): p. 313-326 e5.
132. Lobo, J., et al., *WHO 2022 landscape of papillary and chromophobe renal cell carcinoma.* Histopathology, 2022. **81**(4): p. 426-438.

133. Amin, M.B., et al., *Chromophobe renal cell carcinoma: histomorphologic characteristics and evaluation of conventional pathologic prognostic parameters in 145 cases.* Am J Surg Pathol, 2008. **32**(12): p. 1822-34.
134. Lin, R., et al., *Sarcomatoid renal cell tumor harboring a novel BYSL::TFEB fusion with concurrent TFEB amplification.* Genes Chromosomes Cancer, 2023. **62**(6): p. 353-360.
135. Rizzo, M., et al., *MiT translocation renal cell carcinoma: A review of the literature from molecular characterization to clinical management.* Biochim Biophys Acta Rev Cancer, 2022. **1877**(6): p. 188823.
136. Lindner, A.K., et al., *Targeting strategies in the treatment of fumarate hydratase deficient renal cell carcinoma.* Front Oncol, 2022. **12**: p. 906014.
137. Aggarwal, R.K., et al., *Functional succinate dehydrogenase deficiency is a common adverse feature of clear cell renal cancer.* Proc Natl Acad Sci U S A, 2021. **118**(39).
138. Ames, M.K., C.E. Atkins, and B. Pitt, *The renin-angiotensin-aldosterone system and its suppression.* J Vet Intern Med, 2019. **33**(2): p. 363-382.
139. Spivak, J.L., *Erythropoietin.* Blood Rev, 1989. **3**(2): p. 130-5.
140. Bertram, J.F., et al., *Human nephron number: implications for health and disease.* Pediatr Nephrol, 2011. **26**(9): p. 1529-33.
141. Fransen, M.F.J., et al., *Bioprinting of kidney in vitro models: cells, biomaterials, and manufacturing techniques.* Essays Biochem, 2021. **65**(3): p. 587-602.
142. Grahammer, F., C. Schell, and T.B. Huber, *The podocyte slit diaphragm--from a thin grey line to a complex signalling hub.* Nat Rev Nephrol, 2013. **9**(10): p. 587-98.
143. Nakhoul, N. and V. Batuman, *Role of proximal tubules in the pathogenesis of kidney disease.* Contrib Nephrol, 2011. **169**: p. 37-50.
144. Ishihara, H., et al., *Genetic and epigenetic profiling indicates the proximal tubule origin of renal cancers in end-stage renal disease.* Cancer Sci, 2020. **111**(11): p. 4276-4287.
145. Medendorp, K., et al., *Molecular mechanisms underlying the MiT translocation subgroup of renal cell carcinomas.* Cytogenet Genome Res, 2007. **118**(2-4): p. 157-65.
146. Mount, D.B., *Thick ascending limb of the loop of Henle.* Clin J Am Soc Nephrol, 2014. **9**(11): p. 1974-86.
147. Cabanillas, G., et al., *"Collecting duct carcinoma of the kidney: diagnosis and implications for management".* Urol Oncol, 2022. **40**(12): p. 525-536.
148. Chen, L., P.J. Higgins, and W. Zhang, *Development and Diseases of the Collecting Duct System.* Results Probl Cell Differ, 2017. **60**: p. 165-203.
149. Gargiuli, C., et al., *Integrative Transcriptomic Analysis Reveals Distinctive Molecular Traits and Novel Subtypes of Collecting Duct Carcinoma.* Cancers (Basel), 2021. **13**(12).
150. Deng, F. and T. Radswiki, *Renal medullary carcinoma.* 2011.
151. Khoshdel Rad, N., N. Aghdami, and R. Moghadasali, *Cellular and Molecular Mechanisms of Kidney Development: From the Embryo to the Kidney Organoid.* Front Cell Dev Biol, 2020. **8**: p. 183.
152. McMahon, A.P., *Development of the Mammalian Kidney.* Curr Top Dev Biol, 2016. **117**: p. 31-64.
153. Lang, D., et al., *PAX genes: roles in development, pathophysiology, and cancer.* Biochem Pharmacol, 2007. **73**(1): p. 1-14.
154. Bouchard, M., et al., *Nephric lineage specification by Pax2 and Pax8.* Genes Dev, 2002. **16**(22): p. 2958-70.

155. Laszczyk, A.M., et al., *Pax2 and Pax8 Proteins Regulate Urea Transporters and Aquaporins to Control Urine Concentration in the Adult Kidney.* J Am Soc Nephrol, 2020. **31**(6): p. 1212-1225.
156. Barr, M.L., et al., *PAX-8 expression in renal tumours and distant sites: a useful marker of primary and metastatic renal cell carcinoma?* J Clin Pathol, 2015. **68**(1): p. 12-7.
157. Nakai, S., et al., *Crucial roles of Brn1 in distal tubule formation and function in mouse kidney.* Development, 2003. **130**(19): p. 4751-9.
158. Schmid, M., et al., *Uromodulin facilitates neutrophil migration across renal epithelial monolayers.* Cell Physiol Biochem, 2010. **26**(3): p. 311-8.
159. Kuhlbrodt, K., et al., *Cooperative function of POU proteins and SOX proteins in glial cells.* J Biol Chem, 1998. **273**(26): p. 16050-7.
160. Kang, H.M., et al., *Sox9-Positive Progenitor Cells Play a Key Role in Renal Tubule Epithelial Regeneration in Mice.* Cell Rep, 2016. **14**(4): p. 861-871.
161. Kha, M., et al., *The injury-induced transcription factor SOX9 alters the expression of LBR, HMGA2, and HIPK3 in the human kidney.* Am J Physiol Renal Physiol, 2023. **324**(1): p. F75-F90.
162. Kabaria, R., Z. Klaassen, and M.K. Terris, *Renal cell carcinoma: links and risks.* Int J Nephrol Renovasc Dis, 2016. **9**: p. 45-52.
163. Russo, P., *End stage and chronic kidney disease: associations with renal cancer.* Front Oncol, 2012. **2**: p. 28.
164. Scelo, G., et al., *Genome-wide association study identifies multiple risk loci for renal cell carcinoma.* Nat Commun, 2017. **8**: p. 15724.
165. Pavlovich, C.P., et al., *Renal tumors in the Birt-Hogg-Dube syndrome.* Am J Surg Pathol, 2002. **26**(12): p. 1542-52.
166. Jinzaki, M., et al., *Renal angiomyolipoma: a radiological classification and update on recent developments in diagnosis and management.* Abdom Imaging, 2014. **39**(3): p. 588-604.
167. Mittal, M.K. and B. Sureka, *Solid renal masses in adults.* Indian J Radiol Imaging, 2016. **26**(4): p. 429-442.
168. Arora, R.D., *Renal Clear Cell Cancer*, in *StatPearls*. 2023, StatPearls Publishing: Treasure Island (FL).
169. Wu, H., et al., *mTOR Activation Initiates Renal Cell Carcinoma Development by Coordinating ERK and p38MAPK.* Cancer Res, 2021. **81**(12): p. 3174-3186.
170. Espana-Agusti, J., et al., *Loss of PBRM1 rescues VHL dependent replication stress to promote renal carcinogenesis.* Nat Commun, 2017. **8**(1): p. 2026.
171. Le, V.H. and J.J. Hsieh, *Genomics and genetics of clear cell renal cell carcinoma: a mini-review.* Journal of Translational Genetics and Genomics, 2018.
172. Mendoza-Alvarez, A., et al., *Whole-Exome Sequencing Identifies Somatic Mutations Associated With Mortality in Metastatic Clear Cell Kidney Carcinoma.* Front Genet, 2019. **10**: p. 439.
173. Turajlic, S., et al., *Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal.* Cell, 2018. **173**(3): p. 595-610 e11.
174. Jonasch, E., C.L. Walker, and W.K. Rathmell, *Clear cell renal cell carcinoma ontogeny and mechanisms of lethality.* Nat Rev Nephrol, 2021. **17**(4): p. 245-261.
175. Zhang, J. and Q. Zhang, *VHL and Hypoxia Signaling: Beyond HIF in Cancer.* Biomedicines, 2018. **6**(1).
176. Park, K.C., D.C. Lee, and Y.I. Yeom, *NDRG3-mediated lactate signaling in hypoxia.* BMB Rep, 2015. **48**(6): p. 301-2.

177. Manning, B.D. and A. Toker, *AKT/PKB Signaling: Navigating the Network.* Cell, 2017. **169**(3): p. 381-405.
178. Chae, Y.C., et al., *Mitochondrial Akt Regulation of Hypoxic Tumor Reprogramming.* Cancer Cell, 2016. **30**(2): p. 257-272.
179. Janku, F., T.A. Yap, and F. Meric-Bernstam, *Targeting the PI3K pathway in cancer: are we making headway?* Nat Rev Clin Oncol, 2018. **15**(5): p. 273-291.
180. Yang, H., et al., *pVHL acts as an adaptor to promote the inhibitory phosphorylation of the NF-kappaB agonist Card9 by CK2.* Mol Cell, 2007. **28**(1): p. 15-27.
181. Zhang, J., et al., *VHL substrate transcription factor ZHX2 as an oncogenic driver in clear cell renal cell carcinoma.* Science, 2018. **361**(6399): p. 290-295.
182. Qi, H., *The von Hippel-Lindau tumor suppressor protein sensitizes renal cell carcinoma cells to tumor necrosis factor-induced cytotoxicity by suppressing the nuclear factor-kappaB-dependent antiapoptotic pathway.* Cancer Research, 2003. **63**(21): p. 7076-7080.
183. Lolkema, M.P., et al., *Tumor suppression by the von Hippel-Lindau protein requires phosphorylation of the acidic domain.* J Biol Chem, 2005. **280**(23): p. 22205-11.
184. Hasanov, E., et al., *Ubiquitination and regulation of AURKA identifies a hypoxia-independent E3 ligase activity of VHL.* Oncogene, 2017. **36**(24): p. 3450-3463.
185. Thoma, C.R., et al., *VHL loss causes spindle misorientation and chromosome instability.* Nat Cell Biol, 2009. **11**(8): p. 994-1001.
186. Ogawa, K., et al., *Development of radiolabeled compounds for molecular imaging and imaging-based therapy.* ScientificWorldJournal, 2015. **2015**: p. 365418.
187. Kaelin, W.G., Jr. and P.J. Ratcliffe, *Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway.* Mol Cell, 2008. **30**(4): p. 393-402.
188. Zimna, A. and M. Kurpisz, *Hypoxia-Inducible Factor-1 in Physiological and Pathophysiological Angiogenesis: Applications and Therapies.* Biomed Res Int, 2015. **2015**: p. 549412.
189. Mertz, K.D., et al., *Automated immunofluorescence analysis defines microvessel area as a prognostic parameter in clear cell renal cell cancer.* Hum Pathol, 2007. **38**(10): p. 1454-62.
190. Luo, W., et al., *Pyruvate kinase M2 is a PHD3-stimulated coactivator for hypoxia-inducible factor 1.* Cell, 2011. **145**(5): p. 732-44.
191. Kierans, S.J. and C.T. Taylor, *Regulation of glycolysis by the hypoxia-inducible factor (HIF): implications for cellular physiology.* J Physiol, 2021. **599**(1): p. 23-37.
192. Liberti, M.V. and J.W. Locasale, *The Warburg Effect: How Does it Benefit Cancer Cells?* Trends Biochem Sci, 2016. **41**(3): p. 211-218.
193. Tam, S.Y., V.W.C. Wu, and H.K.W. Law, *Hypoxia-Induced Epithelial-Mesenchymal Transition in Cancers: HIF-1alpha and Beyond.* Front Oncol, 2020. **10**: p. 486.
194. Hapke, R.Y. and S.M. Haake, *Hypoxia-induced epithelial to mesenchymal transition in cancer.* Cancer Lett, 2020. **487**: p. 10-20.
195. Davis, L., et al., *Targeting HIF-2alpha in the Tumor Microenvironment: Redefining the Role of HIF-2alpha for Solid Cancer Therapy.* Cancers (Basel), 2022. **14**(5).
196. Mylonis, I., G. Simos, and E. Paraskeva, *Hypoxia-Inducible Factors and the Regulation of Lipid Metabolism.* Cells, 2019. **8**(3).
197. Nagao, K., et al., *Fatty acid binding protein 7 may be a marker and therapeutic targets in clear cell renal cell carcinoma.* BMC Cancer, 2018. **18**(1): p. 1114.
198. Tostain, J., et al., *Carbonic anhydrase 9 in clear cell renal cell carcinoma: a marker for diagnosis, prognosis and treatment.* Eur J Cancer, 2010. **46**(18): p. 3141-8.

199. Kubala, J.M., et al., *NDUFA4L2 reduces mitochondrial respiration resulting in defective lysosomal trafficking in clear cell renal cell carcinoma.* Cancer Biol Ther, 2023. **24**(1): p. 2170669.
200. Verine, J., et al., *Determination of angptl4 mRNA as a diagnostic marker of primary and metastatic clear cell renal-cell carcinoma.* PLoS One, 2010. **5**(4): p. e10421.
201. Kadoch, C., et al., *Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy.* Nat Genet, 2013. **45**(6): p. 592-601.
202. Wanior, M., et al., *Exploiting vulnerabilities of SWI/SNF chromatin remodelling complexes for cancer therapy.* Oncogene, 2021. **40**(21): p. 3637-3654.
203. Cai, W., et al., *PBRM1 acts as a p53 lysine-acetylation reader to suppress renal tumor growth.* Nat Commun, 2019. **10**(1): p. 5800.
204. Gu, X., et al., *PBRM1 loss in kidney cancer unbalances the proximal tubule master transcription factor hub to repress proximal tubule differentiation.* Cell Rep, 2021. **36**(12): p. 109747.
205. Dey, A., et al., *Loss of the tumor suppressor BAP1 causes myeloid transformation.* Science, 2012. **337**(6101): p. 1541-6.
206. Endoh, M., et al., *Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity.* PLoS Genet, 2012. **8**(7): p. e1002774.
207. Peng, J., et al., *Stabilization of MCRS1 by BAP1 prevents chromosome instability in renal cell carcinoma.* Cancer Lett, 2015. **369**(1): p. 167-74.
208. Wang, S.S., et al., *Bap1 is essential for kidney function and cooperates with Vhl in renal tumorigenesis.* Proc Natl Acad Sci U S A, 2014. **111**(46): p. 16538-43.
209. Fahey, C.C. and I.J. Davis, *SETting the Stage for Cancer Development: SETD2 and the Consequences of Lost Methylation.* Cold Spring Harb Perspect Med, 2017. **7**(5).
210. Doll, K.M., A. Rademaker, and J.A. Sosa, *Practical Guide to Surgical Data Sets: Surveillance, Epidemiology, and End Results (SEER) Database.* JAMA Surg, 2018. **153**(6): p. 588-589.
211. Bahadoram, S., *Renal cell carcinoma: an overview of the epidemiology, diagnosis, and treatment.* Giornale Italiano Di Nefrologia: Organo Ufficiale Della Societa Italiana Di Nefrologia, 2022. **39**(3): p. 2022-vol3.
212. Yang, J., K. Wang, and Z. Yang, *Treatment strategies for clear cell renal cell carcinoma: Past, present and future.* Front Oncol, 2023. **13**: p. 1133832.
213. Huang, L., S. Jiang, and Y. Shi, *Tyrosine kinase inhibitors for solid tumors in the past 20 years (2001-2020).* J Hematol Oncol, 2020. **13**(1): p. 143.
214. Hsieh, J.J., et al., *Renal cell carcinoma.* Nat Rev Dis Primers, 2017. **3**: p. 17009.
215. Kato, Y., et al., *Lenvatinib plus anti-PD-1 antibody combination treatment activates CD8+ T cells through reduction of tumor-associated macrophage and activation of the interferon pathway.* PLoS One, 2019. **14**(2): p. e0212513.
216. Liu, Y.F., et al., *Immune checkpoint inhibitor-based therapy for advanced clear cell renal cell carcinoma: A narrative review.* Int Immunopharmacol, 2022. **110**: p. 108900.
217. Motzer, R.J., et al., *Nivolumab versus everolimus in patients with advanced renal cell carcinoma: Updated results with long-term follow-up of the randomized, open-label, phase 3 CheckMate 025 trial.* Cancer, 2020. **126**(18): p. 4156-4167.
218. Albiges, L., et al., *Nivolumab plus ipilimumab versus sunitinib for first-line treatment of advanced renal cell carcinoma: extended 4-year follow-up of the phase III CheckMate 214 trial.* ESMO Open, 2020. **5**(6): p. e001079.

219. Motzer, R.J., et al., *Avelumab plus axitinib versus sunitinib in advanced renal cell carcinoma: biomarker analysis of the phase 3 JAVELIN Renal 101 trial.* Nat Med, 2020. **26**(11): p. 1733-1741.
220. Bagaev, A., et al., *Conserved pan-cancer microenvironment subtypes predict response to immunotherapy.* Cancer Cell, 2021. **39**(6): p. 845-865 e7.
221. Barrett, R.L. and E. Pure, *Cancer-associated fibroblasts and their influence on tumor immunity and immunotherapy.* Elife, 2020. **9**.
222. Mariathasan, S., et al., *TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells.* Nature, 2018. **554**(7693): p. 544-548.
223. Moreira, M., et al., *Resistance to cancer immunotherapy in metastatic renal cell carcinoma.* Cancer Drug Resist, 2020. **3**(3): p. 454-471.
224. Smaldone, M.C., et al., *Clinical characteristics associated with treatment type for localized renal tumors: implications for practice pattern assessment.* Urology, 2013. **81**(2): p. 269-75.
225. Ljungberg, B., et al., *European Association of Urology Guidelines on Renal Cell Carcinoma: The 2022 Update.* Eur Urol, 2022. **82**(4): p. 399-410.
226. Becht, E., et al., *Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression.* Genome Biol, 2016. **17**(1): p. 218.
227. Chevrier, S., et al., *An Immune Atlas of Clear Cell Renal Cell Carcinoma.* Cell, 2017. **169**(4): p. 736-749 e18.
228. Krishna, C., et al., *Single-cell sequencing links multiregional immune landscapes and tissue-resident T cells in ccRCC to tumor topology and therapy efficacy.* Cancer Cell, 2021. **39**(5): p. 662-677 e6.
229. Borcherding, N., et al., *Mapping the immune environment in clear cell renal carcinoma by single-cell genomics.* Commun Biol, 2021. **4**(1): p. 122.
230. Braun, D.A., et al., *Progressive immune dysfunction with advancing disease stage in renal cell carcinoma.* Cancer Cell, 2021. **39**(5): p. 632-648 e8.
231. Meylan, M., et al., *Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer.* Immunity, 2022. **55**(3): p. 527-541 e5.
232. Debien, V., et al., *Sarcomatoid Dedifferentiation in Renal Cell Carcinoma: From Novel Molecular Insights to New Clinical Opportunities.* Cancers (Basel), 2019. **12**(1).
233. Hahn, A.W., et al., *The significance of sarcomatoid and rhabdoid dedifferentiation in renal cell carcinoma.* Cancer Treat Res Commun, 2022. **33**: p. 100640.
234. Chen, Q., et al., *Deep Learning-Based Classification of Epithelial-Mesenchymal Transition for Predicting Response to Therapy in Clear Cell Renal Cell Carcinoma.* Front Oncol, 2021. **11**: p. 782515.
235. Lyu, C., et al., *Identification of EZH2 as Cancer Stem Cell Marker in Clear Cell Renal Cell Carcinoma and the Anti-Tumor Effect of Epigallocatechin-3-Gallate (EGCG).* Cancers (Basel), 2022. **14**(17).
236. Kim, K., et al., *Determinants of renal cell carcinoma invasion and metastatic competence.* Nat Commun, 2021. **12**(1): p. 5760.
237. Feng, D., et al., *SMOC2 promotes an epithelial-mesenchymal transition and a pro-metastatic phenotype in epithelial cells of renal cell carcinoma origin.* Cell Death Dis, 2022. **13**(7): p. 639.
238. Rabjerg, M., et al., *Comparing World Health Organization/International Society of Urological Pathology Grading and Fuhrman Grading with the Prognostic Value of Nuclear Area in Patients with Renal Cell Carcinoma.* Uro, 2021. **1**(1): p. 2-13.

239. Offermann, A., C. Kuempers, and S. Perner, *Histological (Sub)Classifications and Their Prognostic Impact in Renal Cell Carcinoma.* 2019: p. 537-553.
240. Brannon, A.R., et al., *Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns.* Genes Cancer, 2010. **1**(2): p. 152-163.
241. Cancer Genome Atlas Research, N., *Comprehensive molecular characterization of clear cell renal cell carcinoma.* Nature, 2013. **499**(7456): p. 43-9.
242. Beuselinck, B., et al., *Molecular subtypes of clear cell renal cell carcinoma are associated with sunitinib response in the metastatic setting.* Clin Cancer Res, 2015. **21**(6): p. 1329-39.
243. Su, Y. and A.L. Hong, *Recent Advances in Renal Medullary Carcinoma.* Int J Mol Sci, 2022. **23**(13).
244. Caramel, J., F. Quignon, and O. Delattre, *RhoA-dependent regulation of cell migration by the tumor suppressor hSNF5/INI1.* Cancer Res, 2008. **68**(15): p. 6154-61.
245. Wilson, B.G. and C.W. Roberts, *SWI/SNF nucleosome remodellers and cancer.* Nat Rev Cancer, 2011. **11**(7): p. 481-92.
246. Wang, X., et al., *SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation.* Nat Genet, 2017. **49**(2): p. 289-295.
247. Msaouel, P., et al., *Molecular hallmarks of renal medullary carcinoma: more to c-MYC than meets the eye.* Mol Cell Oncol, 2020. **7**(5): p. 1777060.
248. Msaouel, P., N.M. Tannir, and C.L. Walker, *A Model Linking Sickle Cell Hemoglobinopathies and SMARCB1 Loss in Renal Medullary Carcinoma.* Clin Cancer Res, 2018. **24**(9): p. 2044-2049.
249. Figueiredo, C.R., et al., *Blockade of MIF-CD74 Signalling on Macrophages and Dendritic Cells Restores the Antitumour Immune Response Against Metastatic Melanoma.* Front Immunol, 2018. **9**: p. 1132.
250. Cheng, M., et al., *Immunosuppressive role of SPP1-CD44 in the tumor microenvironment of intrahepatic cholangiocarcinoma assessed by single-cell RNA sequencing.* J Cancer Res Clin Oncol, 2022.
251. Campbell, J.J., et al., *The chemokine receptor CCR4 in vascular recognition by cutaneous but not intestinal memory T cells.* Nature, 1999. **400**(6746): p. 776-80.
252. Xu, W., Z. Yang, and N. Lu, *A new role for the PI3K/Akt signaling pathway in the epithelial-mesenchymal transition.* Cell Adh Migr, 2015. **9**(4): p. 317-24.
253. Shukla, S., et al., *Activation of PI3K-Akt signaling pathway promotes prostate cancer cell invasion.* Int J Cancer, 2007. **121**(7): p. 1424-32.
254. Tripathi, K. and M. Garg, *Mechanistic regulation of epithelial-to-mesenchymal transition through RAS signaling pathway and therapeutic implications in human cancer.* J Cell Commun Signal, 2018. **12**(3): p. 513-527.
255. Zhou, Y., et al., *Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.* Nat Commun, 2019. **10**(1): p. 1523.
256. Sherman, B.T., et al., *DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update).* Nucleic Acids Res, 2022. **50**(W1): p. W216-W221.
257. Wilkerson, M.D. and D.N. Hayes, *ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking.* Bioinformatics, 2010. **26**(12): p. 1572-3.
258. Chan, T.C., Y.L. Shiue, and C.F. Li, *The biological impacts of CEBPD on urothelial carcinoma development and progression.* Front Oncol, 2023. **13**: p. 1123776.

259. Wortel, I.M.N., et al., *Surviving Stress: Modulation of ATF4-Mediated Stress Responses in Normal and Malignant Cells.* Trends Endocrinol Metab, 2017. **28**(11): p. 794-806.
260. Verde, P., et al., *Deciphering AP-1 function in tumorigenesis: fra-ternizing on target promoters.* Cell Cycle, 2007. **6**(21): p. 2633-9.
261. Teufel, A., et al., *FoxP4, a novel forkhead transcription factor.* Biochim Biophys Acta, 2003. **1627**(2-3): p. 147-52.
262. Nowik, M., et al., *Genome-wide gene expression profiling reveals renal genes regulated during metabolic acidosis.* Physiol Genomics, 2008. **32**(3): p. 322-34.
263. Zhao, M., et al., *Renal tubular epithelium-targeted peroxisome proliferator-activated receptor-gamma maintains the epithelial phenotype and antagonizes renal fibrogenesis.* Oncotarget, 2016. **7**(40): p. 64690-64701.
264. Gewin, L.S., *Renal fibrosis: Primacy of the proximal tubule.* Matrix Biol, 2018. **68-69**: p. 248-262.
265. Li, R., et al., *Mapping single-cell transcriptomes in the intra-tumoral and associated territories of kidney cancer.* Cancer Cell, 2022. **40**(12): p. 1583-1599 e10.
266. Hu, T., et al., *DEC2 expression is positively correlated with HIF-1 activation and the invasiveness of human osteosarcomas.* J Exp Clin Cancer Res, 2015. **34**(1): p. 22.
267. Shen, Z., et al., *Overexpression of BHLHE41, correlated with DNA hypomethylation in 3'UTR region, promotes the growth of human clear cell renal cell carcinoma.* Oncol Rep, 2019. **41**(4): p. 2137-2147.
268. Yang, X., et al., *SOX9 is a dose-dependent metastatic fate determinant in melanoma.* J Exp Clin Cancer Res, 2019. **38**(1): p. 17.
269. Matherne, M.G., et al., *Emerging functions of C/EBPbeta in breast cancer.* Front Oncol, 2023. **13**: p. 1111522.
270. Simeonov, K.P., et al., *Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states.* Cancer Cell, 2021. **39**(8): p. 1150-1162 e9.
271. Ben-Baruch, A., *Site-specific metastasis formation: chemokines as regulators of tumor cell adhesion, motility and invasion.* Cell Adh Migr, 2009. **3**(4): p. 328-33.
272. Xue, J., et al., *Patterns of distant metastases in patients with clear cell renal cell carcinoma--A population-based analysis.* Cancer Med, 2021. **10**(1): p. 173-187.
273. Zeisberg, M. and R. Kalluri, *Cellular mechanisms of tissue fibrosis. 1. Common and organ-specific mechanisms associated with tissue fibrosis.* Am J Physiol Cell Physiol, 2013. **304**(3): p. C216-25.
274. Axelrod, M.L., et al., *Biological Consequences of MHC-II Expression by Tumor Cells in Cancer.* Clin Cancer Res, 2019. **25**(8): p. 2392-2402.
275. Tsuno, T., et al., *IRF9 is a key factor for eliciting the antiproliferative activity of IFN-alpha.* J Immunother, 2009. **32**(8): p. 803-16.
276. Hernandez de la Cruz, O.N., et al., *Regulation Networks Driving Vasculogenic Mimicry in Solid Tumors.* Front Oncol, 2019. **9**: p. 1419.
277. Wu, Y., et al., *A novel definition of microvessel density in renal cell carcinoma: Angiogenesis plus vasculogenic mimicry.* Oncol Lett, 2020. **20**(5): p. 192.
278. Lyu, C., et al., *A Pan-Cancer Landscape of ABCG2 across Human Cancers: Friend or Foe?* Int J Mol Sci, 2022. **23**(24).
279. Klement, J.D., et al., *Osteopontin Blockade Immunotherapy Increases Cytotoxic T Lymphocyte Lytic Activity and Suppresses Colon Tumor Progression.* Cancers (Basel), 2021. **13**(5).
280. Gonzalez, H., C. Hagerling, and Z. Werb, *Roles of the immune system in cancer: from tumor initiation to metastatic progression.* Genes Dev, 2018. **32**(19-20): p. 1267-1284.

281. Davidson, G., et al., *Dynamic Evolution of Clonal Composition and Neoantigen Landscape in Recurrent Metastatic Melanoma with a Rare Combination of Driver Mutations.* J Invest Dermatol, 2019. **139**(8): p. 1769-1778 e2.
282. Koikawa, K., et al., *Targeting Pin1 renders pancreatic cancer eradicable by synergizing with immunochemotherapy.* Cell, 2021. **184**(18): p. 4753-4771 e27.
283. Zhu, B., et al., *LncRNA SFTA1P mediates positive feedback regulation of the Hippo-YAP/TAZ signaling pathway in non-small cell lung cancer.* Cell Death Discov, 2021. **7**(1): p. 369.
284. Lv, P. and Y. Xue, *ETS like-1 protein ELK1-induced lncRNA LINC01638 accelerates the progression of papillary thyroid cancer by regulating Axin2 through Wnt/beta-catenin signaling pathway.* Bioengineered, 2021. **12**(1): p. 3873-3885.
285. Lu, H., et al., *Downregulation of LINC01638 lncRNA inhibits migration and invasion of pancreatic ductal adenocarcinoma cells by reducing TGF-beta signaling.* Mol Med Rep, 2019. **20**(5): p. 4533-4539.
286. Rambow, F., et al., *Toward Minimal Residual Disease-Directed Therapy in Melanoma.* Cell, 2018. **174**(4): p. 843-855 e19.
287. Pozniak, J. and F. Rambow, *A TCF4/BRD4-dependent regulatory network confers cross-resistance to targeted and immune checkpoint therapy in melanoma.* 2022.
288. Ji, D., et al., *Identification of LINC01615 as potential metastasis-related long noncoding RNA in hepatocellular carcinoma.* J Cell Physiol, 2019. **234**(8): p. 12964-12970.
289. Hu, Y., et al., *Screening key lncRNAs with diagnostic and prognostic value for head and neck squamous cell carcinoma based on machine learning and mRNA-lncRNA co-expression network analysis.* Cancer Biomark, 2020. **27**(2): p. 195-206.
290. Zhang, S., et al., *Establishment and Validation of a Ferroptosis-Related Long Non-Coding RNA Signature for Predicting the Prognosis of Stomach Adenocarcinoma.* Front Genet, 2022. **13**: p. 818306.
291. Kumar, P., et al., *An epithelial-mesenchymal plasticity signature identifies two novel LncRNAs with the opposite regulation.* Front Cell Dev Biol, 2022. **10**: p. 885785.
292. Zhou, Z., et al., *Identification and Validation of a Ferroptosis-Related Long Non-Coding RNA (FRlncRNA) Signature to Predict Survival Outcomes and the Immune Microenvironment in Patients With Clear Cell Renal Cell Carcinoma.* Front Genet, 2022. **13**: p. 787884.

2. My publications and oral communications

Publications:

- Berico, P., Cigrang, M., **Davidson, G.**, Braun, C., Sandoz, J., Legras, S., Vokshi, B.H., Slovic, N., Peyresaubes, F., Gene Robles, C.M., Egly, J., Compe, E., Davidson, I., Coin, F., 2021. CDK7 and MITF repress a transcription program involved in survival and drug tolerance in melanoma. EMBO Reports 22. https://doi.org/10.15252/embr.202051683
- Coassolo, S., **Davidson, G.**, Negroni, L., Gambi, G., Daujat, S., Romier, C., Davidson, I., 2021. Citrullination of pyruvate kinase M2 by PADI1 and PADI3 regulates glycolysis and cancer cell proliferation. Nat Commun 12, 1718. https://doi.org/10.1038/s41467-021-21960-4
- **Davidson, G.**, Coassolo, S., Kieny, A., Ennen, M., Pencreach, E., Malouf, G.G., Lipsker, D., Davidson, I., 2019. Dynamic Evolution of Clonal Composition and Neoantigen Landscape in Recurrent Metastatic Melanoma with a Rare Combination of Driver Mutations. Journal of Investigative Dermatology 139, 1769-1778.e2. https://doi.org/10.1016/j.jid.2019.01.027
- **Davidson, G.***, Helleux, A.*, Vano, Y.A., Lindner, V., Fattori, A., Cerciat, M., Elaidi, R.Thierry., Verkarre, V., Sun, C.-M., Chevreau, C., Bennamoun, M., Lang, H., Tricard, T., Fridman, W.Herman., Sautes-Fridman, C., Su, X., Plassard, D., Keime, C., Thibault-Carpentier, C., Barthélémy, P., Oudard, S.Marie., Davidson, I., Malouf, G.G., 2023. Mesenchymal-like tumor cells and myofibroblastic cancer-associated fibroblasts are associated with progression and immunotherapy response of clear-cell renal cell carcinoma. Cancer Research CAN-22-3034. https://doi.org/10.1158/0008-5472.CAN-22-3034
- Debien, V., **Davidson, G.**, Baltzinger, P., Kurtz, J.-E., Séverac, F., Imperiale, A., Pessaux, P., Addeo, P., Bachellier, P., Su, X., Davidson, I., Chenard, M.-P., Goichot, B., Malouf, G.G., 2021. Involvement of Neutrophils in Metastatic Evolution of Pancreatic Neuroendocrine Tumors. Cancers 13, 2771. https://doi.org/10.3390/cancers13112771
- Ennen, M., Keime, C., Gambi, G., Kieny, A., Coassolo, S., Thibault-Carpentier, C., Margerin-Schaller, F., **Davidson, G.**, Vagne, C., Lipsker, D., Davidson, I., 2017. MITF -High and MITF -Low Cells and a Novel Subpopulation Expressing Genes of Both Cell States Contribute to Intra- and Intertumoral Heterogeneity of Primary Melanoma. Clinical Cancer Research 23, 7097–7107. https://doi.org/10.1158/1078-0432.CCR-17-0010
- Gambi, G., Mengus, G., **Davidson, G.**, Demesmaeker, E., Cuomo, A., Bonaldi, T., Katopodi, V., Malouf, G.G., Leucci, E., Davidson, I., 2022. The LncRNA LENOX Interacts with RAP2C to Regulate Metabolism and Promote Resistance to MAPK Inhibition in Melanoma. Cancer Research 82, 4555–4570. https://doi.org/10.1158/0008-5472.CAN-22-0959
- Gantzer, J**., Davidson, G.**, Vokshi, B., Weingertner, N., Bougoüin, A., Moreira, M., Lindner, V., Lacroix, G., Mascaux, C., Chenard, M.-P., Bertucci, F., Davidson, I., Kurtz, J.-E., Sautès-Fridman, C., Fridman, W.H., Malouf, G.G., 2022. Immune-Desert Tumor Microenvironment in Thoracic SMARCA4-Deficient Undifferentiated Tumors with Limited Efficacy of Immune Checkpoint Inhibitors. The Oncologist 27, 501–511. https://doi.org/10.1093/oncolo/oyac040
- Joshi, S., **Davidson, G.**, Le Gras, S., Watanabe, S., Braun, T., Mengus, G., Davidson, I., 2017. TEAD transcription factors are required for normal primary myoblast differentiation in vitro and muscle regeneration in vivo. PLOS Genetics 13, e1006600. https://doi.org/10.1371/journal.pgen.1006600
- Kleiber, T.*, **Davidson, G.***, Mengus, G., Martianov, I., Davidson, I., 2021. Single cell transcriptomics reveal trans-differentiation of pancreatic beta cells following inactivation of the TFIID subunit Taf4. Cell Death Dis 12, 790. https://doi.org/10.1038/s41419-021-04067-y

- Laurette, P., Coassolo, S., **Davidson, G.**, Michel, I., Gambi, G., Yao, W., Sohier, P., Li, M., Mengus, G., Larue, L., Davidson, I., 2020. Chromatin remodellers Brg1 and Bptf are required for normal gene expression and progression of oncogenic Braf-driven mouse melanoma. Cell Death Differ 27, 29–43. https://doi.org/10.1038/s41418-019-0333-6
- Mancuso, P., Tricarico, R., Bhattacharjee, V., Cosentino, L., Kadariya, Y., Jelinek, J., Nicolas, E., Einarson, M., Beeharry, N., Devarajan, K., Katz, R.A., Dorjsuren, D.G., Sun, H., Simeonov, A., Giordano, A., Testa, J.R., **Davidson, G.**, Davidson, I., Larue, L., Sobol, R.W., Yen, T.J., Bellacosa, A., 2019. Thymine DNA glycosylase as a novel target for melanoma. Oncogene 38, 3710–3728. https://doi.org/10.1038/s41388-018-0640-2
- Martianov, I., Velt, A., **Davidson, G.**, Choukrallah, M.-A., Davidson, I., 2016. TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. Scientific Reports 6. https://doi.org/10.1038/srep32069
- Säisä-Borreill, S., **Davidson, G.**, Kleiber, T., Thevenot, A., Martin, E., Mondot, S., Blottière, H., Helleux, A., Mengus, G., Plateroti, M., Duluc, I., Davidson, I., Freund, J.-N., 2023. General transcription factor TAF4 antagonizes epigenetic silencing by Polycomb to maintain intestine stem cell functions. Cell Death Differ 30, 839–853. https://doi.org/10.1038/s41418-022-01109-6
- Vokshi, B.H.*****, **Davidson, G.***, Tawanaie Pour Sedehi, N., Helleux, A., Rippinger, M., Haller, A.R., Gantzer, J., Thouvenin, J., Baltzinger, P., Bouarich, R., Manriquez, V., Zaidi, S., Rao, P., Msaouel, P., Su, X., Lang, H., Tricard, T., Lindner, V., Surdez, D., Kurtz, J.-E., Bourdeaut, F., Tannir, N.M., Davidson, I., Malouf, G.G., 2023. SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance. Nat Commun 14, 3034. https://doi.org/10.1038/s41467-023-38472-y
- Yao, W., German, B., Chraa, D., Braud, A., Hugel, C., Meyer, P., **Davidson, G.**, Laurette, P., Mengus, G., Flatter, E., Marschall, P., Segaud, J., Guivarch, M., Hener, P., Birling, M.-C., Lipsker, D., Davidson, I., Li, M., 2022. Keratinocyte-derived cytokine TSLP promotes growth and metastasis of melanoma by regulating the tumor-associated immune microenvironment. JCI Insight 7, e161438. https://doi.org/10.1172/jci.insight.161438

*** Equal contribution**

Posters:

- Alexandra Helleux, Guillaume Davidson, Justine Gantzer, Agathe Bourgamyer, Irwin Davidson, Gabriel Malouf. **Single-cell sequencing reveals cell ontogeny and tumor heterogeneity of a perivascular epithelioid cell neoplasm**. VIB conference: Tumor Heterogeneity, Plasticity and Therapy (virtual edition), presented online on the 5th of May 2021 together with Alexandra Helleux.
- Guillaume Davidson, Bujamin Vokshi, Alexandra Helleux, Marc Rippinger, Alexandre Haller, Justine Gantzer, Philippe Baltzinger, Jonathan Thouvenin, Rachida Bouarich-Bourimi, Valentina Manriquez, Sakina Zaidi, Priya Rao, Pavlos Msaouel, Xiaoping Su, Hervé Lang, Thibault Tricard, Véronique Lindner, Didier Surdez, Jean-Emmanuel Kurtz, Franck Bourdeaut, Nizar M. Tannir, Gabriel Malouf, Irwin Davidson. **Renal medullary carcinoma tumor heterogeneity, microenvironment and novel therapeutic vulnerabilities**. International Cancer Research Symposium, presented in Lyon on the 2nd of November 2022.

Annexes

Chapter V: Annexes

List of annexes:

- Involvement of Neutrophils in Metastatic Evolution of Pancreatic Neuroendocrine Tumors (2021-06-02, Cancers)
- CDK7 and MITF repress a transcription program involved in survival and drug tolerance in melanoma (2021-07-23, EMBO reports)
- Immune-Desert Tumor Microenvironment in Thoracic SMARCA4-Deficient Undifferentiated Tumors with Limited Efficacy of Immune Checkpoint Inhibitors (2022-06-08, The Oncologist)
- The LncRNA LENOX Interacts with RAP2C to Regulate Metabolism and Promote Resistance to MAPK Inhibition in Melanoma (2022-12-16, Cancer Research)
- General transcription factor TAF4 antagonizes epigenetic silencing by Polycomb to maintain intestine stem cell functions (2023-01-13, Cell Death & Differentiation)

cancers

MDPI

Article

Involvement of Neutrophils in Metastatic Evolution of Pancreatic Neuroendocrine Tumors

Véronique Debien [1,2,*,†], Guillaume Davidson [2,†], Philippe Baltzinger [2,3], Jean-Emmanuel Kurtz [1], François Séverac [4], Alessio Imperiale [5], Patrick Pessaux [6], Pietro Addeo [7], Philippe Bachellier [7], Xiaoping Su [8], Irwin Davidson [2], Marie-Pierrette Chenard [2,9], Bernard Goichot [3] and Gabriel G. Malouf [1,2,*]

1 Department of Oncology, Institut de Cancérologie Strasbourg Europe, ICANS, 67200 Strasbourg, France; je.kurtz@icans.eu
2 Department of Cancer and Functional Genomics, Institute of Genetics and Molecular and Cellular Biology, CNRS/INSERM/UNISTRA, 67400 Illkirch, France; davidsoy@igbmc.fr (G.D.); baltzinp@igbmc.fr (P.B.); irwin@igbmc.fr (I.D.); marie-pierrette.chenard@chru-strasbourg.fr (M.-P.C.)
3 Internal Medicine Department, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France; bernard.goichot@chru-strasbourg.fr
4 Department of Public Health and Epidemiology, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France; francois.severac@chru-strasbourg.fr
5 Department of Nuclear Medicine, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France; a.imperiale@icans.eu
6 Department of Surgery, Nouvel Hôpital Civil, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France; patrick.pessaux@chru-strasbourg.fr
7 Department of Hepato-pancreatic Surgery, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France; pietrofrancesco.addeo@chru-strasbourg.fr (P.A.); philippe.bachellier@chru-strasbourg.fr (P.B.)
8 Department of Bioinformatics and Computational Biology, MD Anderson Cancer Center, Houston, TX 77230-1402, USA; XSu1@mdanderson.org
9 Department of Pathology, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France
* Correspondence: veronique.debien@bordet.be (V.D.); maloufg@igbmc.fr (G.G.M.)
† Equal contributions.

check for updates

Citation: Debien, V.; Davidson, G.; Baltzinger, P.; Kurtz, J.-E.; Séverac, F.; Imperiale, A.; Pessaux, P.; Addeo, P.; Bachellier, P.; Su, X.; et al. Involvement of Neutrophils in Metastatic Evolution of Pancreatic Neuroendocrine Tumors. *Cancers* **2021**, *13*, 2771. https://doi.org/10.3390/cancers13112771

Academic Editor: David Wong

Received: 26 April 2021
Accepted: 28 May 2021
Published: 2 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Simple Summary: The neutrophil-to-lymphocytes (NLR) reflects the systemic inflammation. Based on a cohort of 144 patients treated for localized or metastatic well-differentiated pancreatic neuroendocrine tumors (pNET), we identified the NLR $\geq$ 4 to be associated with worse overall survival. Using MCP-Counter on a publicly available pNET RNA-sequencing dataset, we inferred the tumor microenvironment composition of 83 primary pNET and 30 liver metastasis. The neutrophils scores were statistically higher in liver metastasis relative to primary pancreatic tumors ($p = 0.005$). Gene set enrichment analysis further revealed activation of complement pathway signature in liver metastasis. Through combination of the neutrophil and complement pathway genes, we found that pNET can be classified into two clusters: Neu-Comp1 and Neu-Comp2. Notably, the Neu-Compt1 cluster was enriched in neutrophils and complement pathway signatures and was associated with liver metastasis. These data offer new insights into the role of inflammatory factors in the metastatic progression of the pNET.

Abstract: Well-differentiated pancreatic neuroendocrine tumors (pNET) have an unpredictable natural history. The identification of both blood and tumor immune features associated with patients' outcomes remains limited. Herein, we evaluated the best prognostic value of the neutrophils-to-lymphocyte ratio (NLR) in a cohort of 144 pNETs. The NLR $\geq$ 4 was associated with worse overall survival in both univariate analysis (HR = 3.53, CI95% = 1.50–8.31, $p = 0.004$) and multivariate analysis (HR = 2.57, CI95% = 1.061–6.216, $p = 0.036$). The presence of synchronous liver metastasis was identified as a prognostic factor in multivariate analysis (HR = 3.35, CI95% = 1.411–7.973, $p = 0.006$). Interestingly, the absolute tumor-associated neutrophils count was higher in liver metastasis as compared to their paired primary tumor ($p = 0.048$). Deconvolution of immune cells from the transcriptome of 83 primary tumors and 30 liver metastases reveals enrichment for neutrophils in metastasis relative to primary tumors ($p = 0.005$), and this was associated with upregulation of the

complement pathway (NES = 1.84, $p < 0.0001$). Combining neutrophils signature and complement pathway genes, unsupervised clustering identified two pNETs subgroups, namely Neu-Comp1 and Neu-Comp2. Characterized by neutrophils infiltration and activation of the complement pathway, Neu-Comp1 was highly enriched for metastatic liver samples as compared to Neu-Comp2 ($p < 0.0001$). These data suggest the possible link between liver metastasis, complement pathway activation, and neutrophils infiltration in well-differentiated pNET and open avenues for targeting complement pathways in these tumors.

Keywords: pancreatic neuroendocrine tumors; neutrophils-to-lymphocyte ratio; tumor microenvironment; neutrophils; complement; innate immunity; transcriptome

1. Introduction

Pancreatic neuroendocrine tumors (pNET) are the second most frequent tumor arising in the pancreas after adenocarcinoma, accounting for 1–3% of all pancreatic tumors [1]. Their incidence has been increasing over the last few years [2]. According to the Surveillance, Epidemiology, and End Results (SEER) registry, the incidence has reached 0.8 new cases per 100,000 persons per year in 2012 as compared to <0.1/100,000 per year in 1973 [2]. pNET is a heterogeneous disease in which the grade is the most important prognostic factor [3]. The three-trial grading system is based on the evaluation of tumors' proliferative potential estimated by the percentage of Ki67 positive cells or mitotic count [4]. Grade 3 carcinomas are known to have the worst outcome with five-year overall survival (OS) estimated at 13%, while having a five-year OS for grade 1 (G1) and grade 2 (G2) tumors are 80% and 67%, respectively [5]. However, among well-differentiated grade G1 and G2 tumors, there is a true tumor heterogeneity leading to a distinct natural history. Among known prognostic factors for pNET, Ki67 expression (which is in part related to tumor grade) and stage at diagnosis (lymph node involvement, as well as the burden of liver metastases) are the most frequently used [3,6]. Carcinological surgery is the cornerstone of therapy for localized disease (although the wait-and-see strategy is an option for small G1 pNET) [3]. However, surgery can also be curative in some patients with liver metastases [7]. Outcomes in localized, well-differentiated G1 and G2 pNET are generally favorable, although the recurrence rate varies between 12–25% in the literature [8,9]. Therefore, the identification of prognostic biomarkers is an unmet need in this population [10]. Among the explored biomarkers, systemic inflammation has been recognized as and represents a hallmark of cancer [11].

Various biological parameters reflect systemic inflammation such as elevated blood neutrophil-to-lymphocytes (NLR) ratio, which fosters tumor proliferation and metastasis via inhibition of apoptosis, promotion of angiogenesis, and DNA damage [11–13]. Thus, the tumor-associated neutrophils (TANs) were shown to interact with tumor cells. Depending on various extracellular stimulations (e.g., IFNg, TGF-B), TAN may present an "immunosuppressive switch" from antitumor N1 phenotype to pro-tumoral N2 phenotype [14]. Moreover, TAN activity is dependent on the tumor type and location within the tumor (intratumoral versus stromal) [15].

The NLR has been already described as a prognostic factor in different types of tumors [16,17]. High NLR was related to patients' poor overall survival among various metastatic tumors [11,12]. For example, in colorectal cancer (surgically treated localized disease, as well as in the metastatic setting), NLR > 5 was associated with a worse outcome [16]. Several studies described NLR as a prognostic factor of relapse and survival in all-grade resected pNET patients [18–23]. However, different cut-offs have been identified among the reported cohorts. Notably, beyond its simple calculation, NLR might be a surrogate for the immune tumor microenvironment (TME), whereas the presence of tumor-associated immune cells is generally assessed by immunohistochemistry (IHC) [24].

With the advent of immune checkpoint inhibitors, efforts have been made to perform the immune classification of various cancers types in different organs [25]. Those efforts revealed striking associations between somatic mutations and TME composition and response to immune checkpoint inhibitors [26]. However, to the best of our knowledge, no comprehensive subtypes immune TME profiling has been reported to date.

In the present work, we analyzed the association between NLR, clinicopathological tumor features, and patients' outcomes in a cohort of well-differentiated pNET identifying higher NLR and TANs as features of liver metastasis. Furthermore, we have inferred the distribution of immune cells from the transcriptome of 83 primary pNET and 30 liver metastasis revealing striking associations between neutrophils enrichment, complement activation, and liver metastasis.

2. Materials and Methods

2.1. Patient Population

We recorded pNET patients' data from five oncology centers in the area of Alsace, France. Patients were identified from the tumor board database between 1 January 2008 and 1 January 2019, and variables of interest were extracted from medical files. This study was approved by the Institutional Ethics Committee (no. 7435) of Strasbourg University Hospital and conducted in accordance with the principles outlined in the Declaration of Helsinki. Inclusion criteria were as follows: confirmed diagnosis of well-differentiated G1 and G2 pNET (all stages) reviewed by pathologists from national expert board TENPATH, adult patients, availability of white blood count analysis before treatment (surgery or medical treatment). Recorded data included: gender, age, tumor stage, pathology findings (grade, TNM staging, functional or not), prior medical history including a context of multiple endocrine neoplasia (MEN1), blood count characteristics, treatment characteristics, and survival. The definition of OS is the time between the diagnosis until death from any causes or the last day of follow-up. TNM stage was determined from the American Joint Committee on Cancer (AJCC) 2017 classification (seventh edition) [27]. The tumor grade and differentiation were defined according to 2017 WHO classification of pNET [28].

2.2. NLR Calculation

The lymphocyte and neutrophil counts were obtained from the white blood cells count (WBC) performed as close as possible (less than 3 months) to diagnosis or prior to the surgery for patients undergoing surgery. NLR was calculated by dividing the neutrophils absolute count by lymphocyte absolute count. The predictive value for OS of NLR was defined using the receiver operating characteristic (ROC) curve.

2.3. Immunohistochemistry Neutrophils and Lymphocyte Assessment

Three paired tissue samples from primary and liver metastasis were available for immunohistochemistry staining for the evaluation of tumor-associated neutrophils and lymphocytes. Neutrophils staining with the CD66b antibody (BioLegend®, cat. no. 305102) was performed according to the manufacturer's instructions. For this purpose, two FFPE tissue sections from primary pNETs and their corresponding liver metastasis were prepared, washed in pH = 7.4 PBS, and stained with 10µg/mL of diluted CD66b antibody. In addition, tumor-associated lymphocytes were stained with the CD3 antibody (ThemoFicher®, cat. no. RM-9107-S, clone SP7) according to the manufacturer's instructions. The density of staining was evaluated by counting the stained cells by two independent investigators (M.P.C. and V.D.) in one field x400, with an average of 10 fields. Acute appendicitis was used as a positive control for all stains.

2.4. RNA Sequencing Data and Bioinformatic Analysis

Raw data for RNA sequencing (RNA-seq) of 83 primary pNETs and 30 liver metastases were downloaded from a publicly available dataset (GEO: GSE98894) [29]. Raw reads were aligned using STAR v2.5.3a with the "–quantMode Transcriptome SAM" argument

and by providing the GFF file from ENSEMBL v75; the gene expression level was then calculated using RSEM v1.3.3. To infer the distribution of immune populations, we used the MCP-Counter v1.2.0. package for R software [30]. The TME deconvolution tool allows for estimating in an abundance of 10 cell populations including eight immune and stromal cells' population (neutrophils, myeloid dendritic cells, monocytic cells, B lineage, NK cells, cytotoxic lymphocytes, CD8 T cells, T cells), and two stromal cells' populations (endothelial cells and fibroblasts), based on their expression score. To identify subgroups, unsupervised clustering was performed according to the Z-score and visualized as a heatmap with the 'pheatmap' package v1.0.12. The number of clusters was chosen empirically following the obtained dendrograms. The MCP-counter scores for immune cells were compared between identified clusters. To estimate the tumor-associated neutrophils score, we used a mean value of the neutrophil score (defined by expression of *CXCR1*, *CXCR2*, *FCGR3B* genes). The complement pathway was defined by the expression of *SERPINC1*, *C4BPB*, *PLG*, *APOC1*, *C3*, *APOA4CP*, *F2*, *TFPI2*, and *ITIH* genes.

To assess the differential gene expression between primary pNETs and liver metastasis, we used the Wald test for differential expression proposed by Love et al. and implemented in the Bioconductor package DESeq2 version 1.16.1 [31]. Genes with a high Cook's distance were filtered out and independent filtering based on the mean of normalized counts was performed. *p*-values were adjusted for multiple testing using the Benjamini and Hochberg method [32]. Gene Set Enrichment Analysis (GSEA) was done using the GSEA software v4.0.3 with the pre-ranked algorithm on log2 (fold-changes) estimated by DESeq2, using the human hallmark gene sets from Molecular Signatures Database (MSigD) v7.1 [33,34]. Gene sets with a False Discovery Rate (FDR) < 0.05 were considered as significantly differentially expressed.

2.5. Statistical Analysis

The receiver operating characteristics (ROC) curve and its area under the curve (AUC) were used to obtain the best cut-off value for NLR based on the overall survival population. Survival outcomes were determined by the Kaplan–Meier method and compared by the log-rank test. Univariate and multivariate analyses were performed using a logistic regression model. The Akaike Information Criterion was used to keep variables associated with survival in multivariate analysis. The Cox proportional model was used for OS univariate and multivariate analyses.

The unpaired Student's *t*-test and Chi^2 test were performed to identify differences between groups and associations between NLR and categorical variables. The paired Student's *t*-test was performed to evaluate the statistical difference in paired samples. The Kruskal–Wallis *t*-test was performed to compare three and more groups among each other. The analyses were performed with R Studio Version 1.1.463 and GraphPad Prism version 5.0 a. The results were considered statistically significant if $p < 0.05$.

3. Results

3.1. Patients' Characteristics

Overall, 187 patients were identified from the tumor board database. Out of those, 24 and 19 cases were excluded because of missing data and G3 grade, respectively. Hence, 144 well-differentiated G1 and G2 pNET patients' data were available for analysis. Patients' demographics and characteristics are displayed in Table 1. Median patient age was 56 years (range: 20–81 years), with a slight male predominance (57%). Among the total population, 80 patients (55.6%) had the symptomatic disease at diagnosis, with 28 (19.4%) of them displaying functional tumors. Only 10 (6.9%) patients had known MEN1 syndrome. At diagnosis, synchronous metastases were present in 42 (29.2%) of patients, and 50 (34.7%) patients had lymph node involvement. A total of 129 (89.6%) patients underwent surgical treatment, with R0 and R1 resection in 112 and 17 cases, respectively.

Table 1. Baseline clinicopathological characterization of the current cohort.

Variable	*n* = 144
Median age, years (range)	56 (20–81)
Gender, male	82 (57%)
BMI, kg/m^2	25,7 (16.5–46.3)
Missing data	19
Symptoms at diagnosis, Yes	80 (55.6%)
Functional tumor	28 (19.4%)
MEN1	10 (6.9%)
Size, median in mm, range	25 (5–120)
Missing data	9
Metastasis at diagnosis	42 (29.2%)
Ki67, median %	3 (1–20)
Grade 2	69 (48%)
Grade 1	75 (52%)
Size, T from AJCC 2017	
T1	58 (40.3%)
T2	33 (22.9%)
T3	48 (33.3%)
T4	4 (2.8%)
Lymph node status, N1	50 (34.7%)
Missing data	9
Surgery	129 (89.6%)
Surgical margins, R1	17 (11.8%)
Median NLR	2,31 (0.99–14.05)
Median neutrophils count,/mm^3	4245 (1370–14470)
Median lymphocytes count,/mm^3	1695 (420–4040)

NLR, neutrophil-to-lymphocyte ratio.

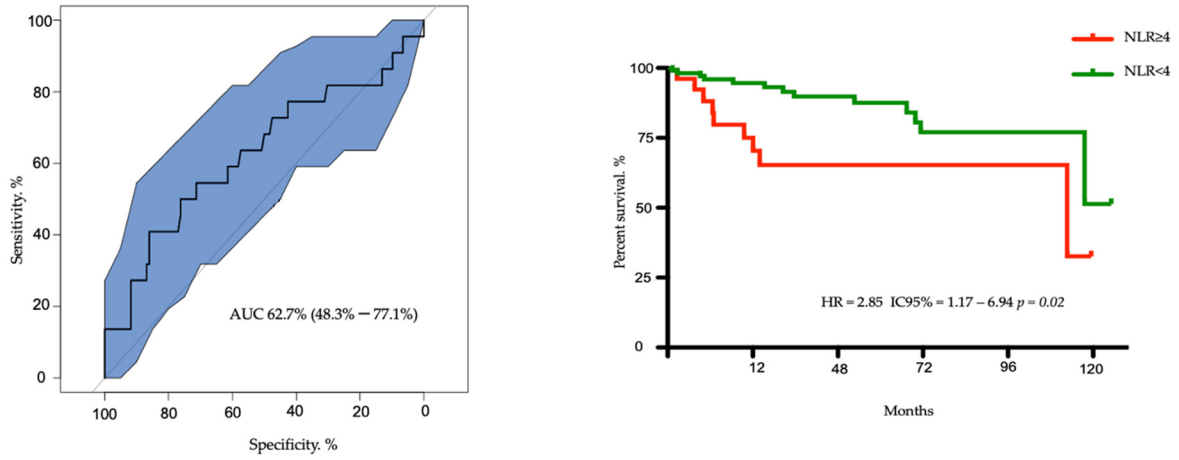
The association between NLR and patient's characteristics is reported in Table S1.

3.2. Survival of Patients According to NLR Ratio

According to the AUC of 0.627, the best NLR cut-off was 4, with a sensitivity of 41% and a specificity of 86% (Figure 1a). Twenty-seven patients had an NLR $\geq$ 4. At the last time of follow-up, 13 patients died in the NLR < 4 subgroup as compared to 9 patients in the NLR $\geq$ 4 subgroup. With a median follow-up of 27 months, median OS was 113 months for patients with NLR $\geq$ 4 versus not reached (NR) for the subgroup of patients with NLR < 4 (HR = 2.850, CI 95% = 1.170–6.94, p = 0.02) (Figure 1b). The two-year OS rates were 74% and 96% in the NLR $\geq$ 4 and <4 subgroups, respectively.

3.3. Univariate and Multivariate Analysis for Overall Survival

In univariate analysis, the presence of metastasis (p = 0.006), lymph node involvement (p = 0.01) and NLR $\geq$ 4 (p = 0.004) were significantly associated with OS. Neither the continuous Ki67 value (p = 0.41) nor tumor T stage (p = 0.72) were identified as prognostic factors for OS (Table 2). In multivariate analysis, NLR $\geq$ 4 (HR = 2.57 CI = 1.061–6.216, p = 0.0036) and presence of synchronous liver metastasis (HR = 3.354 CI = 1.411–7.973, p < 0.006) were associated with poor OS (Table 2).



(**a**) (**b**)

Figure 1. Determination of best cut off of neutrophil-to-lymphocyte ratio (NLR) and analysis of overall survival. (**a**) Receiver operating characteristic (ROC) curve for NLR to determine the best cut-off value: 4; AUC = 62.7) with a sensitivity of 41% and specificity of 86%. (**b**) Kaplan–Meier curves for patients overall survival according to NLR.

Table 2. Univariate and multivariate analysis of variables for overall survival.

***n* = 144**		**Univariate**			**Multivariate**		
Variable	***n***	**HR**	**CI 95%**	***p* Value**	**HR**	**CI 95%**	***p* Value**
Age, >50 years	98	1.88	0.693–5.106	0.22			
Sex, male	82	2.46	0.908–6.676	0.08			
Ki67, continuous value	144	1.03	0.958–1.111	0.41			
Tumor size, T3-T4	52	1.17	0.494–2.783	0.72			
Lymph node involvement	50	3.27	1.315–8.117	0.01			
Metastasis	42	3.32	1.417–7.766	0.006	3.35	1.1411–7.973	0.006
NLR ≥ 4	27	3.53	1.502–8.313	0.004	2.57	1.061–6.216	0.036

NLR, neutrophil-to-lymphocyte ratio.

3.4. Association between NLR and Other Clinicopathological Features

We then investigated the association between NLR ≥ 4 and other clinicopathological factors, such as age, gender, body mass index (BMI) <25 or ≥25kg/m^2, presence of symptoms, Ki67 percentage, tumor size, tumor stage, lymph nodes, and distant metastasis (Table S1). Only the presence of synchronous metastasis was associated with increased NLR (HR = 2.32, CI = 0.98–5.51, $p = 0.05$). Notably, NLR was higher in metastatic as compared to localized pNET ($p = 0.007$, Figure 2a). This difference was associated with higher neutrophil counts and lower lymphocyte counts in metastatic relative to localized pNET subgroups ($p = 0.03$ and $p = 0.045$, respectively) (Figure 2b,c).

3.5. Evaluation of Tumors Associated Neutrophils in Liver Metastasis and Matched Primary pNETs

Out of our whole data set population, TANs were assessed by IHC in the matched primary and liver metastasis of three patients, for which material and informed written consents were available (Figure 2e). Interestingly, the absolute count of TANs was two-fold higher in the metastatic group (median range 2.5–10 neutrophils per field) compared to primary (median range 0.5–5.8 neutrophils per field) ($p = 0.048$) (Figure 2d), although the absolute level was low. The median value of CD3 lymphocytes per field in primary tumors was 7.9 and 5.5 in metastatic samples (Table S2)

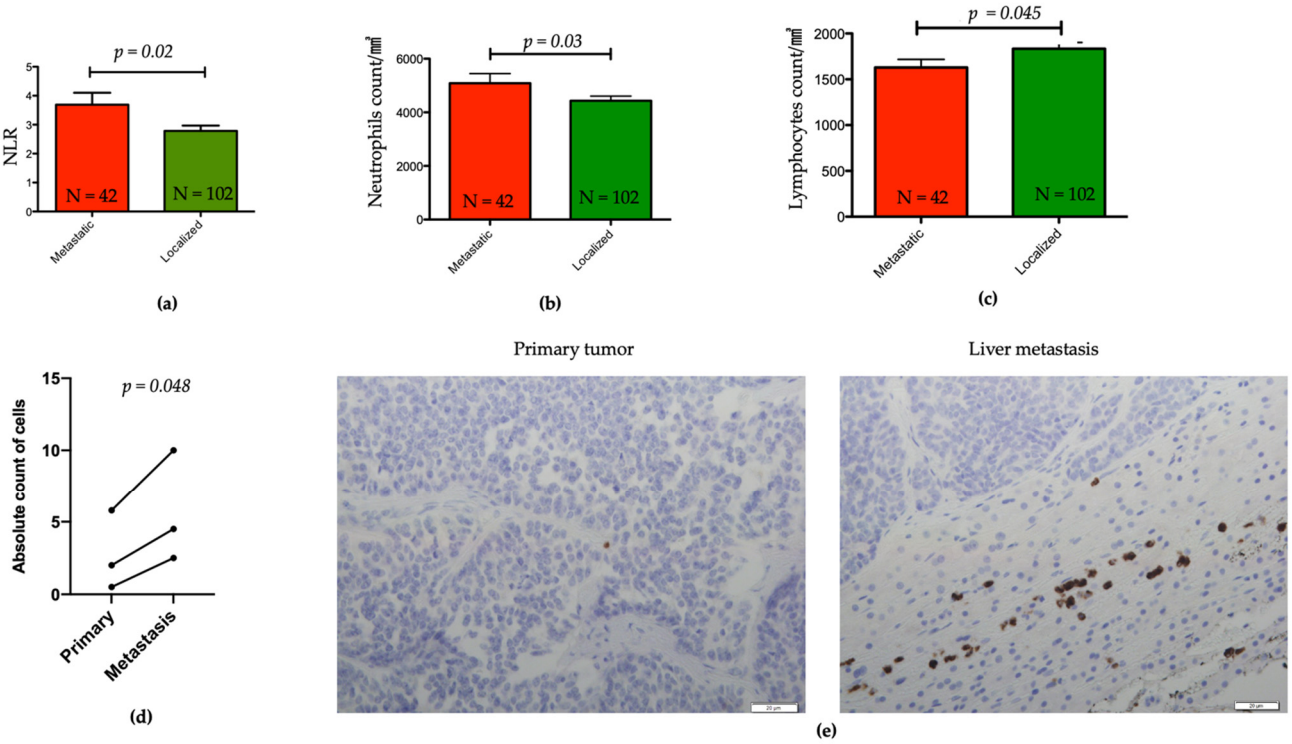


Figure 2. Distribution of patients' blood neutrophils and lymphocytes counts in pNET. (**a**) Bar plot representing NLR from peripherical blood of patients with metastatic versus localized pNET; Student's *t*-test: $p = 0.02$. (**b**,**c**) Bar plot of neutrophils and lymphocytes counts in patients with metastatic versus localized pNET; Student's *t*-test: $p = 0.03$ and $p = 0.045$, respectively. (**d**) Dot plots representing the tumor-associated neutrophils counts in paired metastatic and in primary pNETs. (**e**) Representation of IHC CD66b staining of tumor-associated neutrophils in a primary pancreatic neuroendocrine tumor (left) and its corresponding liver metastasis (right), x400.

3.6. Landscape of the Microenvironment Phenotypes in pNET

To assess the tumor microenvironment (TME) composition in a large collection of 113 pNETs and explore the putative association with metastatic versus localized samples, we inferred the distribution of six immune populations (T cells, CD8 T cells, cytotoxic lymphocytes, B lineage, monocytic lineage, and neutrophils). Unsupervised hierarchical clustering using immune cell scores revealed three heterogeneous clusters: cluster 1 ($n = 7$; 6.2%), the "neutrophils-enriched", with high enrichment for neutrophils ($p < 0.0001$); cluster 2 ($n = 44$; 38.9%), the "immune-desert", with low immune cell infiltration; and cluster 3 ($n = 62$; 54.9%), the "immune-rich" cluster, with high T cells ($p = 0.001$) and cytotoxic lymphocytes ($p < 0.0001$) as compared to the other clusters (Figures 3a and S1). Notably, the neutrophils-enriched cluster was tightly associated with metastatic samples ($n = 5/7$; 71.4%) relative to the remaining C2 ($n = 12/44$; 27.3%) and C3 ($n = 13/62$; 21%) clusters ($p = 0.02$).

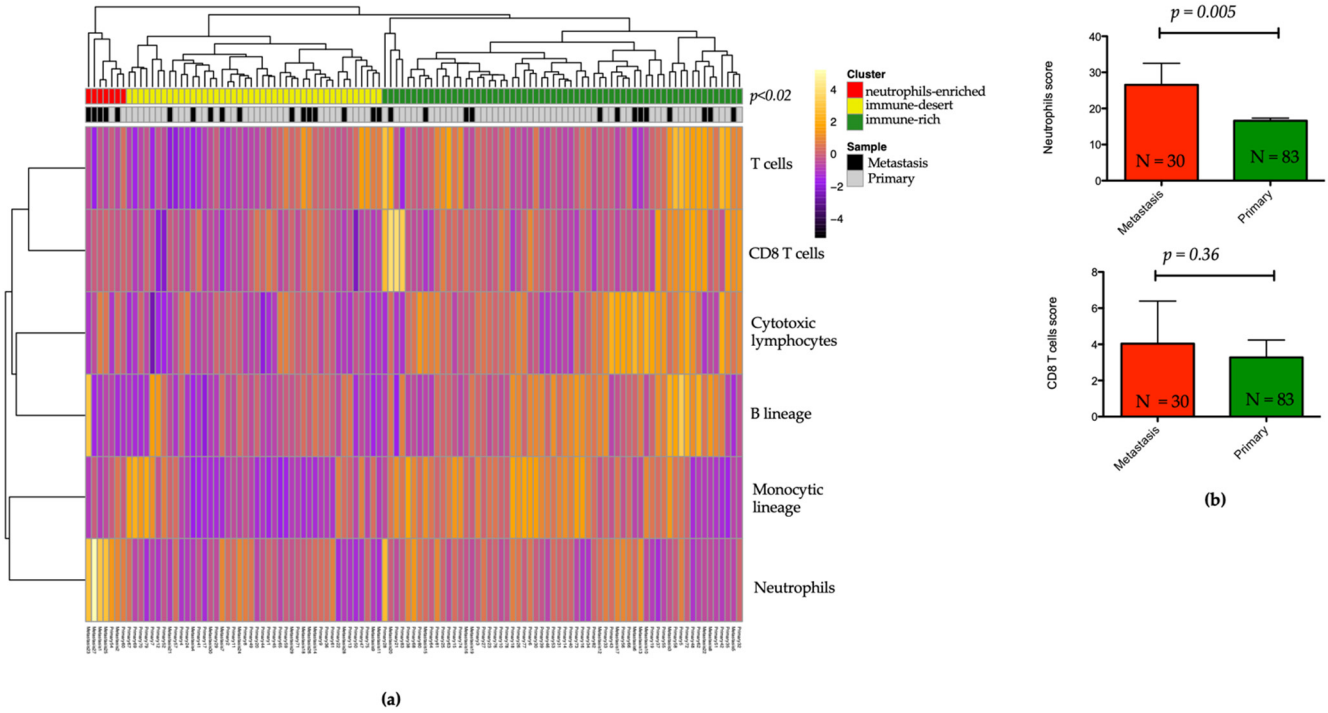


Figure 3. MCP-counter analysis of 113 pancreatic neuroendocrine tumors and distribution of immune cell scores in metastatic and primary tumors. (**a**) Heatmap showing unsupervised hierarchical clustering of immune cells in three clusters. (**b**) Box plot of neutrophils and CD8 lymphocytes score in liver metastasis versus primary pNET tumors.

3.7. Association between Neutrophils Infiltration, Complement Pathway Activation, and Metastatic Tumor Status

To further analyze if any specific molecular features are defining metastatic versus primary pNETs particularly regarding an immunosuppressive myeloid environment possibly linked with higher blood NLR, we compared the distribution of immune cells' scores between primary and metastatic samples. Only the neutrophil score was higher in liver metastasis versus primary pNET samples ($p = 0.005$, Figure 3b). Conversely, no statistically significant difference was observed for CD8 lymphocytes ($p = 0.36$, Figure 3b). We further investigated differentially expressed genes between primary and metastatic tumors; overall, 1041 genes were overexpressed (FC ≥ 2; $p < 0.05$), and 341 genes were downregulated (FC ≤ -2; $p < 0.05$). Gene set enrichment analysis using the Hallmark set identified 25 gene sets with significant enrichment in metastatic relative to primary pNETs (FDR < 0.05) and only two gene sets downregulated (FDR < 0.05). Most upregulated gene sets included E2F targets, xenobiotic metabolism, fatty acid metabolism, G2M checkpoints, and hypoxia along with complement pathway (FDR < 0.05, $p < 0.0001$) (Figure 4).

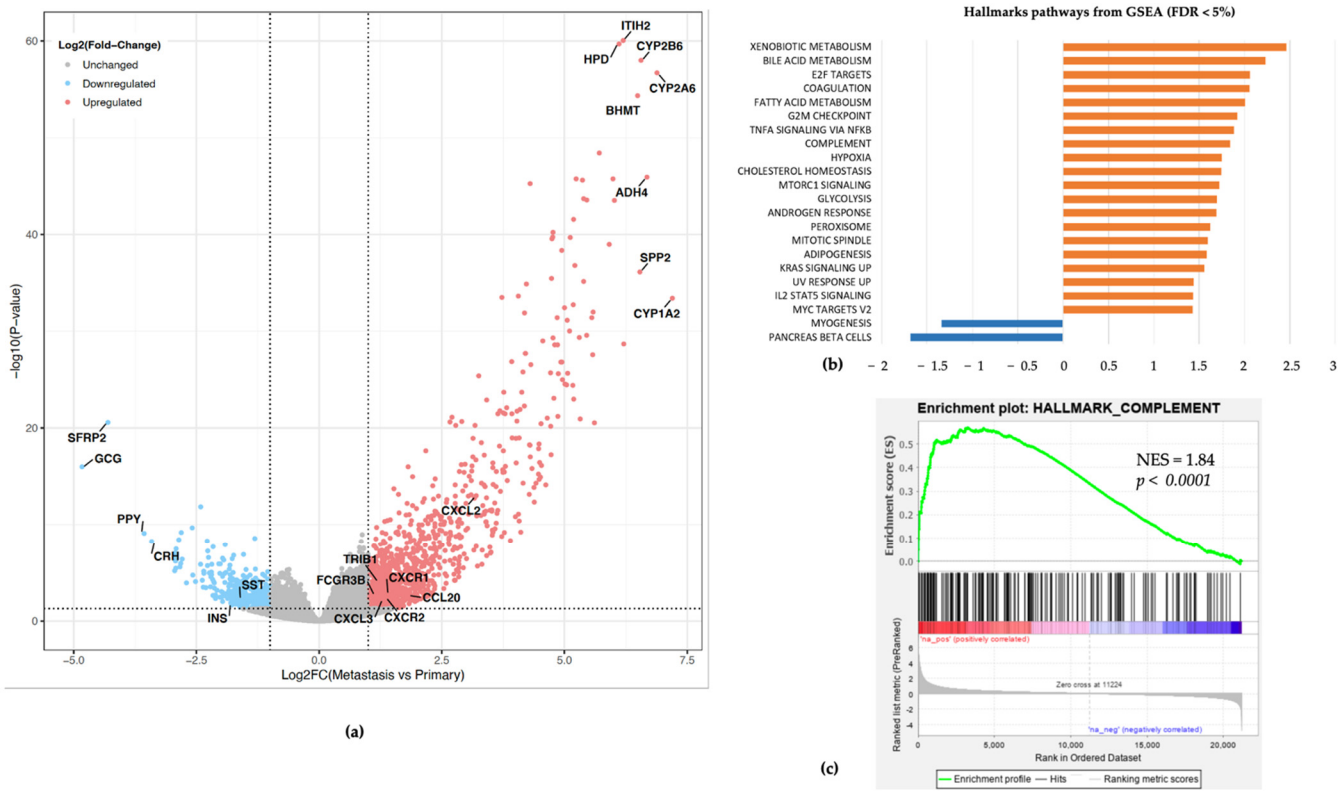


Figure 4. (**a**) Volcano plot of differential gene expression in liver metastasis versus primary pNET tumors. Each point represents a gene. Red represents upregulated genes, while blue is the downregulated one. (**b**) Bar plot representing differentially up- (orange) and down- (blue) regulated Hallmark pathways according to Gene Set Enrichment Analysis (GSEA analysis). NES: Normalized Enrichment Score. (**c**) Enrichment for complement pathway gene set in metastatic liver metastasis versus primary pNET tumors.

3.8. Subtypes Classification of pNETs Using Neutrophils and Complement Pathway Signature

Given the potential link between neutrophils and the complement pathway, we performed hierarchical clustering combining gene signatures of neutrophils and the top 10 expressed genes from the complement GSEA Hallmark gene sets. Unsupervised clustering identified two clusters. The first one (Neu-Comp1) (n = 19; 16.8%) was enriched for neutrophils and complement pathway as compared to the second one (Neu-Comp2) (n = 94; 83.2%). In addition, the Neu-Comp1 cluster was highly enriched for metastatic samples (n = 15; 78.9%) as compared to the Neu-Comp2 cluster (n = 19; 20.2%) ($p < 0.0001$, Figure 5).

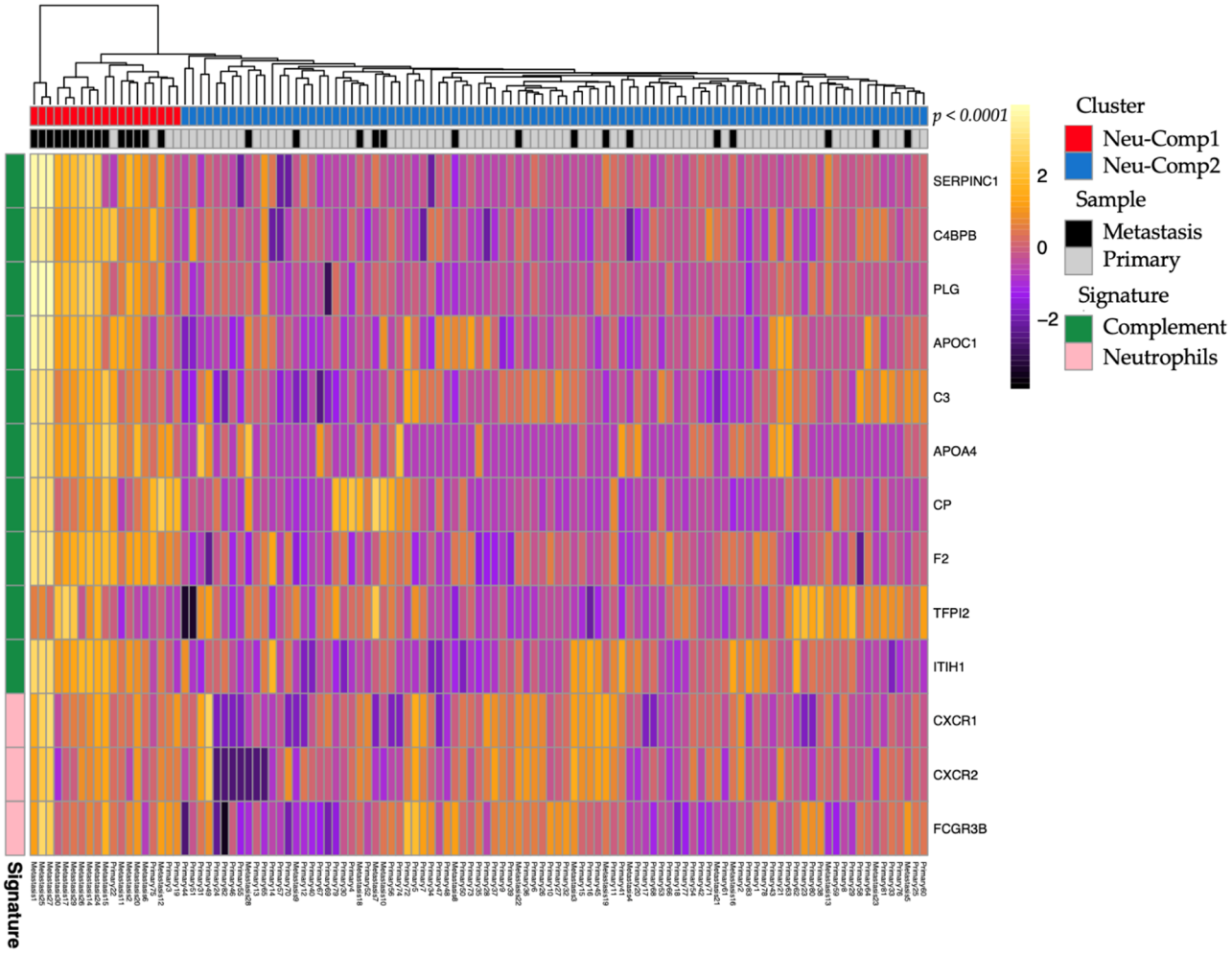


Figure 5. Unsupervised clustering of pNETs according to neutrophils and complement pathway gene expression signatures. Note that metastatic pNET cluster together as compared to localized pNETs ($p < 0.0001$).

4. Discussion

In this study, we identified NLR ≥4 as an independent biomarker for overall survival in well-differentiated pancreatic neuroendocrine tumors and found that it was associated with metastatic disease. We supposed that tumor-associated neutrophils could reflect the difference between primary and metastatic tumors. Moreover, our study investigating the immune TME in pNETs using transcriptome deconvolution, to our knowledge, the first of its kind, identifying TAN enrichment in liver metastasis relative to primary pNET. In addition, we unraveled an association between complement pathway activation and TAN enrichment that suggests the importance of the innate immune system in driving pNET metastasis.

Prognostic factors such as liver metastasis, tumor size, lymph node involvement, WHO grade classification, Ki67, or presence of symptoms have been previously described as prognostic biomarkers of recurrence-free survival [28,35–37]. In addition, scoring systems predicting integrating several clinicopathological parameters have been proposed by Genç et al. [6]. Recently, a high-risk, well-differentiated pNETs score was defined when two out of three of the following variables were present: tumor size > 20 mm, lymph node metastasis, and Ki67 > 5% or mitotic count > 2 [38]. Herein, we believe that NLR ≥ 4 might be added to those factors. To our knowledge, our study is the first to show elevated NLR in metastatic patients as compared to those with localized disease. Interestingly, the NLR ≥ 4 allowed us to identify a subgroup with a higher death risk in the first two years after

diagnosis. To explain such heterogeneity, future investigations are needed to define the genetic and epigenetic molecular underpinnings of these tumors.

Few previous studies explored the role of NLR in well-differentiated pancreatic neuroendocrine tumors in European patients [6,39]. Other published series involved mainly resected patients in the Asian population [18–22,40,41]. Recently, a pooled analyses from RADIANT-3 and RADIANT-4 identified NLR < 2.58 to be associated with longer PFS in all subgroups, including pancreatic neuroendocrine tumors (n = 396, HR = 0.53 CI 95% 0.39–0.70) [42]. All patients enrolled in these trials had metastatic disease, and the majority of them have been already exposed to systemic treatments, including chemotherapy, which could have affected NLR. As NLR is an accessible biomarker of tumor-associated inflammation, we hypothesized that high NLR might be a surrogate marker of tumor microenvironment composition. To date, the correlation between circulating neutrophils and tumor-infiltrating neutrophils is inconsistent among solid tumors. For instance, TANs in pancreatic cancers were shown to be increased in patients with high NLR, although the correlation was not statistically significant [43]. At the functional level, neutrophils are involved in the anti-tumor activity (N1), as well as in the promotion of tumorigenesis (N2); thus, under the pressure of various cytokines they might participate in either angiogenesis and/or metastasis development [13,44,45].

To the best of our knowledge, the difference in TME composition between primary and metastatic pNET is poorly understood. Herein, by using the transcriptome deconvolution for the TME description, we have shown that neutrophils expression scores were higher in liver metastasis relative to primary pNETs, consistent with our IHC staining for available matched primary and liver metastasis. These data are keeping with results showing that higher infiltration of intratumoral neutrophils in localized well-differentiated pNETs has been associated with poor outcomes [46]. Thus, TANs might have an N2 pro-tumoral phenotype driving tumor aggressiveness. Further studies are needed to analyze the distribution and features of these cells using single-cell transcriptome sequencing.

To date, the role of immune cells in pNET has been investigated in several studies using IHC [24,47–51]. A higher level of tumor-infiltrating macrophages (TAMs) was shown to be associated with higher NLR. Both parameters have been statistically correlated with poor recurrence-free survival [22]. In another study, a high level of peritumoral TAMs was associated with lower disease-free survival [50].

Another interesting topic that is important to discuss in our study is the association between neutrophils infiltration with complement activation. Complement is a key factor in tissue inflammation, allowing cancer progression through the release of complement component 5a (C5a). Neutrophil stimulation by cytokines have been shown to activate the alternative complement pathway and release of C5 fragments, which further foster neutrophil proinflammatory responses [52,53]. This mechanism, possibly important for effective immune response, may play a key role in pNETs and highlight potential therapeutic targets to invigorate efficient immune response. Recently, Yang et al. reported compelling results about the potential involvement of the complement C1q activation in liver metastasis of patients with pancreatic adenocarcinomas [54]. Moreover, they showed that C1q is mainly expressed at tumor stroma rather in tumor cells and is involved in complement cascade. Mechanistic experiments further demonstrated that C1q would promote invasion and metastasis. These data are reminiscent with our observations in pNET, suggesting a role of complement pathway activation in hepatic metastasis along with high neutrophils infiltration.

Our study has several weaknesses. Firstly, the analysis of NLR was done retrospectively. The size of our cohort may present another limitation due to the rarity of the disease. However, to our knowledge, our study is one of the largest cohorts in the European population encompassing a large number of cases [20]. Secondly, the elevation of neutrophils or the decrease of lymphocytes may be the consequence of various physiological situations, like infection, or a result of systemic treatments like steroids [55]. Another significant limitation is the size of our setting for TAN evaluation. We only could perform IHC for a handful

of paired samples. This limitation is mainly due to the rarity of the pathology, available tumor material, and retrospective nature of the study. Finally, another limitation is the lack of clinical annotation associated with the retrieved RNA-seq data that we analyzed.

Nonetheless, our study has several strengths. Firstly, it is a multicentric cohort on a period of more than 10 years with centralized cases reviewed by expert pathologists. Secondly, we have managed to establish that NLR ≥ 4, a value found in other solid tumors, is a prognostic tool for overall survival that is accessible for other clinicians and useful in practice. Thirdly, RNA-sequencing mining allowed us to deeply investigate the involvement of innate immunity in pNET. Our data suggest that the difference between localized and metastatic diseases may be related to the tumor microenvironment reflected by variance in NLR. Neutrophils infiltration in liver metastasis in our training and validation dataset strongly suggests that neutrophils may be involved in the development of metastasis, as it has been already reported for colorectal or breast cancer [56–58]. Finally, through the combination of complement pathways with neutrophils signature, we described two pNETs clusters separating metastatic from localized tumors. Altogether, we suggest that activation of the complement pathway may attract neutrophils, promoting not only the inflammation induced by cancer cells, but also their metastatic potential.

5. Conclusions

In summary, our work highlights the importance of tumor-related systemic inflammation biomarkers NLR and TNA as prognostic markers of metastasis in pNETs. Furthermore, this finding indicates the importance of complement activation along with neutrophils infiltration in metastatic pNETS, suggesting that targeting a complement pathway might open avenues for focusing on metastatic pNETs.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/cancers13112771/s1, Figure S1: Bar plots of immune cells score in the three immune clusters: neutrophils-enriched (red), immune-desert (yellow), immune-rich (green). (a) T cells. (b) CD8. (c) Cytotoxic lymphocytes. (d) B lineage. (e) Monocytic lineage. (f) Neutrophils. Table S1: Association between neutrophil-to-lymphocyte ratio and clinical characteristics Table S2: Number of tumor-associated lymphocytes CD3 and neutrophils (CD66b) on three paired hepatic metastasis and primary well-differentiated pancreatic neuroendocrine tumor.

Author Contributions: Conceptualization, V.D. and G.G.M.; Formal analysis, V.D., G.D., and F.S.; Investigation, V.D., G.D., and M.-P.C; Methodology, V.D. and G.G.M.; Resources, P.B. (Philippe Bachellier), J.-E.K., A.I., P.P., P.A., P.B. (Philippe Baltzinger), X.S., I.D., M.-P.C., and B.G.; Software, G.D.; Supervision, G.M.; Writing—original draft, V.D. and G.D.; Writing—review and editing, J.-E.K., G.G.M. All authors have read and agreed to the published version of the manuscript.

Funding: The team has received funding from Ligue Nationale Contre le Cancer.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Strasbourg University Hospital (no. 7435) and approved on 26 June 2019.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study part based on clinical cohort.

Data Availability Statement: The RNA-sequencing dataset is publicly available on GEO: GSE98894. Data is available upon reasonable request.

Acknowledgments: We thank all patients and their families. We also thank Hôpitaux Universitaires de Strasbourg, technicians from Center of Biological Resources (CRB). The authors thank H Fridman and C Sautès-Fridman for helpful discussions.

Conflicts of Interest: The authors declare no conflict of interest related to this work. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Bosman, F.T.; Carneiro, F.; Hruban, R.H.; Theise, N.D. *WHO Classification of Tumours of the Digestive System*; WHO: Geneva, Switzerland, 2010.
2. Dasari, A.; Shen, C.; Halperin, D.; Zhao, B.; Zhou, S.; Xu, Y.; Shih, T.; Yao, J.C. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients with Neuroendocrine Tumours in the United States. *JAMA Oncol.* **2017**, *3*, 1335. [CrossRef] [PubMed]
3. Falconi, M.; Eriksson, B.; Kaltsas, G.; Bartsch, D.K.; Capdevila, J.; Caplin, M.; Kos-Kudla, B.; Kwekkeboom, D.; Rindi, G.; Klöppel, G.; et al. ENETS Consensus Guidelines Update for the Management of Patients with Functional Pancreatic Neuroendocrine Tumours and Non-Functional Pancreatic Neuroendocrine Tumours. *Neuroendocrinology* **2016**, *103*, 153–171. [CrossRef] [PubMed]
4. Klöppel, G.; Couvelard, A.; Hruban, R.; Klimstra, D.S.; Komminoth, P.; Osamura, R.Y.; Perren, A.; Rindi, G. Neoplasms of the Neuroendocrine Pancreas: Introduction. In *WHO Classification of Tumours of Endocrine Organs*; Lloyd, R.V., Osamura, R.Y., Klöppel, G., Rosai, J., Eds.; IARC: Lyon, France, 2017; Volume 10, ISBN 9789283244936.
5. Genc, C.G.; Klümpen, H.J.; van Oijen, M.G.H.; van Eijck, C.H.J.; Nieveen van Dijkum, E.J.M. A Nationwide Population-Based Study on the Survival of Patients with Pancreatic Neuroendocrine Tumours in The Netherlands. *World J. Surg.* **2018**, *42*, 490–497. [CrossRef] [PubMed]
6. Genç, C.G.; Jilesen, A.P.; Partelli, S.; Falconi, M.; Muffatti, F.; van Kemenade, F.J.; van Eeden, S.; Verheij, J.; van Dieren, S.; van Eijck, C.H.J.; et al. A New Scoring System to Predict Recurrent Disease in Grade 1 and 2 Nonfunctional Pancreatic Neuroendocrine Tumours. *Ann. Surg.* **2018**, *267*, 1148–1154. [CrossRef]
7. Pavel, M.; O'Toole, D.; Costa, F.; Capdevila, J.; Gross, D.; Kianmanesh, R.; Krenning, E.; Knigge, U.; Salazar, R.; Pape, U.-F.; et al. ENETS Consensus Guidelines Update for the Management of Distant Metastatic Disease of Intestinal, Pancreatic, Bronchial Neuroendocrine Neoplasms (NEN) and NEN of Unknown Primary Site. *Neuroendocrinology* **2016**, *103*, 172–185. [CrossRef]
8. Gao, H.; Liu, L.; Wang, W.; Xu, H.; Jin, K.; Wu, C.; Qi, Z.; Zhang, S.; Liu, C.; Xu, J.; et al. Novel Recurrence Risk Stratification of Resected Pancreatic Neuroendocrine Tumour. *Cancer Lett.* **2018**, *412*, 188–193. [CrossRef]
9. Marchegiani, G.; Landoni, L.; Andrianello, S.; Masini, G.; Cingarlini, S.; D'Onofrio, M.; De Robertis, R.; Davì, M.; Capelli, P.; Manfrin, E.; et al. Patterns of Recurrence after Resection for Pancreatic Neuroendocrine Tumours: Who, When, and Where? *Neuroendocrinology* **2018**, *108*, 161–171. [CrossRef]
10. Jensen, R.T.; Bodei, L.; Capdevila, J.; Couvelard, A.; Falconi, M.; Glasberg, S.; Kloppel, G.; Lamberts, S.; Peeters, M.; Rindi, G.; et al. Unmet Needs in Functional and Nonfunctional Pancreatic Neuroendocrine Neoplasms. *Neuroendocrinology* **2019**, *108*, 26–36. [CrossRef]
11. Hanahan, D.; Weinberg, R.A. Hallmarks of Cancer: The next Generation. *Cell* **2011**, *144*, 646–674. [CrossRef] [PubMed]
12. Coussens, L.M.; Werb, Z. Inflammation and Cancer. *Nature* **2002**, *420*, 860–867. [CrossRef]
13. Liang, W.; Ferrara, N. The Complex Role of Neutrophils in Tumour Angiogenesis and Metastasis. *Cancer Immunol. Res.* **2016**, *4*, 83–91. [CrossRef] [PubMed]
14. Granot, Z.; Fridlender, Z.G. Plasticity beyond Cancer Cells and the "Immunosuppressive Switch". *Cancer Res.* **2015**, *75*, 4441–4445. [CrossRef] [PubMed]
15. Shaul, M.E.; Fridlender, Z.G. Tumour-Associated Neutrophils in Patients with Cancer. *Nat. Rev. Clin. Oncol.* **2019**, *16*, 601–620. [CrossRef]
16. Haram, A.; Boland, M.R.; Kelly, M.E.; Bolger, J.C.; Waldron, R.M.; Kerin, M.J. The Prognostic Value of Neutrophil-to-Lymphocyte Ratio in Colorectal Cancer: A Systematic Review. *J. Surg. Oncol.* **2017**, *115*, 470–479. [CrossRef] [PubMed]
17. Templeton, A.J.; McNamara, M.G.; Šeruga, B.; Vera-Badillo, F.E.; Aneja, P.; Ocaña, A.; Leibowitz-Amit, R.; Sonpavde, G.; Knox, J.J.; Tran, B.; et al. Prognostic Role of Neutrophil-to-Lymphocyte Ratio in Solid Tumours: A Systematic Review and Meta-Analysis. *JNCI J. Natl. Cancer Inst.* **2014**, *106*, dju124. [CrossRef] [PubMed]
18. Luo, G.; Liu, C.; Cheng, H.; Jin, K.; Guo, M.; Lu, Y.; Long, J.; Xu, J.; Ni, Q.; Chen, J.; et al. Neutrophil-Lymphocyte Ratio Predicts Survival in Pancreatic Neuroendocrine Tumours. *Oncol. Lett.* **2017**, *13*, 2454–2458. [CrossRef]
19. Arima, K.; Okabe, H.; Hashimoto, D.; Chikamoto, A.; Nitta, H.; Higashi, T.; Kaida, T.; Yamamura, K.; Kitano, Y.; Komohara, Y.; et al. Neutrophil-to-Lymphocyte Ratio Predicts Metachronous Liver Metastasis of Pancreatic Neuroendocrine Tumours. *Int. J. Clin. Oncol.* **2017**, *22*, 734–739. [CrossRef]
20. Panni, R.Z.; Lopez-Aguiar, A.G.; Liu, J.; Poultsides, G.A.; Rocha, F.G.; Hawkins, W.G.; Strasberg, S.M.; Trikalinos, N.A.; Maithel, S.; Fields, R.C.; et al. Association of Preoperative Monocyte-to-lymphocyte and Neutrophil-to-lymphocyte Ratio with Recurrence-free and Overall Survival after Resection of Pancreatic Neuroendocrine Tumours (US-NETSG). *J. Surg. Oncol.* **2019**, *22*, 734–739. [CrossRef]
21. Tong, Z.; Liu, L.; Zheng, Y.; Jiang, W.; Zhao, P.; Fang, W.; Wang, W. Predictive Value of Preoperative Peripheral Blood Neutrophil/Lymphocyte Ratio for Lymph Node Metastasis in Patients of Resectable Pancreatic Neuroendocrine Tumours: A Nomogram-Based Study. *World J. Surg. Oncol.* **2017**, *15*, 1–9. [CrossRef]
22. Harimoto, N.; Hoshino, K.; Muranushi, R.; Hagiwara, K.; Yamanaka, T.; Ishii, N.; Tsukagoshi, M.; Igarashi, T.; Tanaka, H.; Watanabe, A.; et al. Prognostic Significance of Neutrophil-Lymphocyte Ratio in Resectable Pancreatic Neuroendocrine Tumours with Special Reference to Tumour-Associated Macrophages. *Pancreantology* **2019**, *19*, 897–902. [CrossRef]
23. Gaitanidis, A.; Patel, D.; Nilubol, N.; Tirosh, A.; Sadowski, S.; Kebebew, E. Markers of Systemic Inflammatory Response Are Prognostic Factors in Patients with Pancreatic Neuroendocrine Tumours (PNETs): A Prospective Analysis. *Ann. Surg. Oncol.* **2018**, *25*, 122–130. [CrossRef] [PubMed]

24. Milione, M.; Miceli, R.; Barretta, F.; Pellegrinelli, A.; Spaggiari, P.; Tagliabue, G.; Centonze, G.; Paolino, C.; Mangogna, A.; Kankava, K.; et al. Microenvironment and Tumour Inflammatory Features Improve Prognostic Prediction in Gastro-Entero-Pancreatic Neuroendocrine Neoplasms. *J. Pathol. Clin. Res.* **2019**, *5*, 217–226. [CrossRef] [PubMed]
25. Bruni, D.; Angell, H.K.; Galon, J. The Immune Contexture and Immunoscore in Cancer Prognosis and Therapeutic Efficacy. *Nat. Rev. Cancer* **2020**, *20*, 662–680. [CrossRef]
26. Petitprez, F.; de Reyniès, A.; Keung, E.Z.; Chen, T.W.-W.; Sun, C.-M.; Calderaro, J.; Jeng, Y.-M.; Hsiao, L.-P.; Lacroix, L.; Bougoüin, A.; et al. B Cells Are Associated with Survival and Immunotherapy Response in Sarcoma. *Nature* **2020**, *577*, 556–560. [CrossRef]
27. Sobin, L.H.; Gospodarowicz, M.K.; Wittekind, C. *TNM Classification of Malignant Tumours*; John Wiley & Sons: Hoboken, NJ, USA, 2011; ISBN 9781444358964.
28. Inzani, F.; Petrone, G.; Rindi, G. The New World Health Organization Classification for Pancreatic Neuroendocrine Neoplasia. *Endocrinol. Metab. Clin. North Am.* **2018**, *47*, 463–470. [CrossRef]
29. Alvarez, M.J.; Subramaniam, P.S.; Tang, L.H.; Grunn, A.; Aburi, M.; Rieckhof, G.; Komissarova, E.V.; Hagan, E.A.; Bodei, L.; Clemons, P.A.; et al. A Precision Oncology Approach to the Pharmacological Targeting of Mechanistic Dependencies in Neuroendocrine Tumours. *Nat. Genet.* **2018**, *50*, 979. [CrossRef] [PubMed]
30. Becht, E.; Giraldo, N.A.; Lacroix, L.; Buttard, B.; Elarouci, N.; Petitprez, F.; Selves, J.; Laurent-Puig, P.; Sautès-Fridman, C.; Fridman, W.H.; et al. Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol.* **2016**, *17*, 1–20. [CrossRef]
31. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef]
32. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57*, 289–300. [CrossRef]
33. 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.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef]
34. Mootha, V.K.; Lindgren, C.M.; Eriksson, K.-F.; Subramanian, A.; Sihag, S.; Lehar, J.; Puigserver, P.; Carlsson, E.; Ridderstråle, M.; Laurila, E.; et al. PGC-1alpha-Responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetes. *Nat. Genet.* **2003**, *34*, 267–273. [CrossRef]
35. Panzuto, F.; Boninsegna, L.; Fazio, N.; Campana, D.; Pia Brizzi, M.; Capurso, G.; Scarpa, A.; De Braud, F.; Dogliotti, L.; Tomassetti, P.; et al. Metastatic and Locally Advanced Pancreatic Endocrine Carcinomas: Analysis of Factors Associated with Disease Progression. *J. Clin. Oncol.* **2011**, *29*, 2372–2377. [CrossRef]
36. Scarpa, A.; Mantovani, W.; Capelli, P.; Beghelli, S.; Boninsegna, L.; Bettini, R.; Panzuto, F.; Pederzoli, P.; delle Fave, G.; Falconi, M. Pancreatic Endocrine Tumours: Improved TNM Staging and Histopathological Grading Permit a Clinically Efficient Prognostic Stratification of Patients. *Mod. Pathol.* **2010**, *23*, 824–833. [CrossRef] [PubMed]
37. Boninsegna, L.; Panzuto, F.; Partelli, S.; Capelli, P.; Fave, G.D.; Bettini, R.; Pederzoli, P.; Scarpa, A.; Falconi, M. Malignant Pancreatic Neuroendocrine Tumour: Lymph Node Ratio and Ki67 Are Predictors of Recurrence after Curative Resections. *Eur. J. Cancer* **2012**, *48*, 1608–1615. [CrossRef] [PubMed]
38. Ausania, F.; Senra Del Rio, P.; Gomez-Bravo, M.A.; Martin-Perez, E.; Pérez-Daga, J.A.; Dorcaratto, D.; González-Nicolás, T.; Sanchez-Cabus, S.; Tardio-Baiges, A. Can We Predict Recurrence in WHO G1-G2 Pancreatic Neuroendocrine Neoplasms? Results from a Multi-Institutional Spanish Study. *Pancreatology* **2019**, *19*, 367–371. [CrossRef] [PubMed]
39. Carmona-Bayonas, A.; Jiménez-Fonseca, P.; Lamarca, Á.; Barriuso, J.; Castaño, Á.; Benavent, M.; Alonso, V.; Riesco-Martínez, M.D.C.; Alonso-Gordoa, T.; Custodio, A.; et al. Prediction of Progression-Free Survival in Patients With Advanced, Well-Differentiated, Neuroendocrine Tumours Being Treated With a Somatostatin Analog: The GETNE-TRASGU Study. *J. Clin. Oncol.* **2019**, *37*, 2571–2580. [CrossRef]
40. Zhou, Y.; Li, D.; Lin, Y.; Yu, M.; Lu, X.; Jian, Z.; Na, N.; Hou, B. Pretreatment Hematologic Markers as Prognostic Predictors of Gastroenteropancreatic Neuroendocrine Tumours: A Systematic Review and Meta-Analysis. *OncoTargets Ther.* **2018**, *11*, 2489–2496. [CrossRef] [PubMed]
41. Salman, T.; Kazaz, S.N.; Varol, U.; Oflazoglu, U.; Unek, I.T.; Kucukzeybek, Y.; Alacacioglu, A.; Atag, E.; Semiz, H.S.; Cengiz, H.; et al. Prognostic Value of the Pretreatment Neutrophil-to-Lymphocyte Ratio and Platelet-to-Lymphocyte Ratio for Patients with Neuroendocrine Tumours: An Izmir Oncology Group Study. *Chemotherapy* **2016**, *61*, 281–286. [CrossRef]
42. Chan, D.L.; Yao, J.C.; Carnaghi, C.; Buzzoni, R.; Herbst, F.; Ridolfi, A.; Strosberg, J.; Kulke, M.H.; Pavel, M.; Singh, S. Markers of Systemic Inflammation in Neuroendocrine Tumours: A Pooled Analysis of the RADIANT-3 and RADIANT-4 Studies. *Pancreas* **2021**, *50*, 130–137. [CrossRef]
43. Takakura, K.; Ito, Z.; Suka, M.; Kanai, T.; Matsumoto, Y.; Odahara, S.; Matsudaira, H.; Haruki, K.; Fujiwara, Y.; Saito, R.; et al. Comprehensive Assessment of the Prognosis of Pancreatic Cancer: Peripheral Blood Neutrophil-Lymphocyte Ratio and Immunohistochemical Analyses of the Tumour Site. *Scand. J. Gastroenterol.* **2016**, *51*, 610–617. [CrossRef]
44. Coffelt, S.B.; Kersten, K.; Doornebal, C.W.; Weiden, J.; Vrijland, K.; Hau, C.-S.; Verstegen, N.J.M.; Ciampricotti, M.; Hawinkels, L.J.A.C.; Jonkers, J.; et al. IL-17-Producing Γδ T Cells and Neutrophils Conspire to Promote Breast Cancer Metastasis. *Nature* **2015**, *522*, 345–348. [CrossRef]

45. Masucci, M.T.; Minopoli, M.; Carriero, M.V. Tumor Associated Neutrophils. Their Role in Tumorigenesis, Metastasis, Prognosis and Therapy. *Front. Oncol.* **2019**, *9*, 1146. [CrossRef] [PubMed]
46. Zhang, W.-H.; Wang, W.-Q.; Gao, H.-L.; Xu, S.-S.; Li, S.; Li, T.-J.; Han, X.; Xu, H.-X.; Li, H.; Jiang, W.; et al. Tumour-Infiltrating Neutrophils Predict Poor Survival of Non-Functional Pancreatic Neuroendocrine Tumour. *J. Clin. Endocrinol. Metab.* **2020**, *105*. [CrossRef]
47. da Silva, A.; Bowden, M.; Zhang, S.; Masugi, Y.; Thorner, A.R.; Herbert, Z.T.; Zhou, C.W.; Brais, L.; Chan, J.A.; Hodi, F.S.; et al. Characterisation of the Neuroendocrine Tumour Immune Microenvironment. *Pancreas* **2018**, *47*, 1123–1129. [CrossRef]
48. de Reuver, P.R.; Mehta, S.; Gill, P.; Andrici, J.; D'Urso, L.; Clarkson, A.; Mittal, A.; Hugh, T.J.; Samra, J.S.; Gill, A.J. Immunoregulatory Forkhead Box Protein P3-Positive Lymphocytes Are Associated with Overall Survival in Patients with Pancreatic Neuroendocrine Tumours. *J. Am. Coll. Surg.* **2016**, *222*, 281–287. [CrossRef] [PubMed]
49. Katz, S.C.; Donkor, C.; Glasgow, K.; Pillarisetty, V.G.; Gönen, M.; Espat, N.J.; Klimstra, D.S.; D'Angelica, M.I.; Allen, P.J.; Jarnagin, W.; et al. T Cell Infiltrate and Outcome Following Resection of Intermediate-Grade Primary Neuroendocrine Tumours and Liver Metastases. *HPB* **2010**, *12*, 674–683. [CrossRef] [PubMed]
50. Cai, L.; Michelakos, T.; Deshpande, V.; Arora, K.S.; Yamada, T.; Ting, D.T.; Taylor, M.S.; Castillo, C.F.-D.; Warshaw, A.L.; Lillemoe, K.D.; et al. Role of Tumour-Associated Macrophages in the Clinical Course of Pancreatic Neuroendocrine Tumours (PanNETs). *Clin. Cancer Res.* **2019**, *25*, 2644–2655. [CrossRef]
51. Takahashi, D.; Kojima, M.; Suzuki, T.; Sugimoto, M.; Kobayashi, S.; Takahashi, S.; Konishi, M.; Gotohda, N.; Ikeda, M.; Nakatsura, T.; et al. Profiling the Tumour Immune Microenvironment in Pancreatic Neuroendocrine Neoplasms with Multispectral Imaging Indicates Distinct Subpopulation Characteristics Concordant with WHO 2017 Classification. *Sci. Rep.* **2018**, *8*, 1–10. [CrossRef]
52. Camous, L.; Roumenina, L.; Bigot, S.; Brachemi, S.; Frémeaux-Bacchi, V.; Lesavre, P.; Halbwachs-Mecarelli, L. Complement Alternative Pathway Acts as a Positive Feedback Amplification of Neutrophil Activation. *Blood* **2011**, *117*, 1340–1349. [CrossRef]
53. Roumenina, L.T.; Daugan, M.V.; Noé, R.; Petitprez, F.; Vano, Y.A.; Sanchez-Salas, R.; Becht, E.; Meilleroux, J.; Clec'h, B.L.; Giraldo, N.A.; et al. Tumour Cells Hijack Macrophage-Produced Complement C1q to Promote Tumour Growth. *Cancer Immunol. Res.* **2019**, *7*, 1091–1105. [CrossRef]
54. Yang, J.; Lin, P.; Yang, M.; Liu, W.; Fu, X.; Liu, D.; Tao, L.; Huo, Y.; Zhang, J.; Hua, R.; et al. Integrated Genomic and Transcriptomic Analysis Reveals Unique Characteristics of Hepatic Metastases and Pro-Metastatic Role of Complement C1q in Pancreatic Ductal Adenocarcinoma. *Genome Biol.* **2021**, *22*, 4. [CrossRef] [PubMed]
55. Mehra, N.; Sharp, A.; Lorente, D.; Dolling, D.; Sumanasuriya, S.; Johnson, B.; Dearnaley, D.; Parker, C.; de Bono, J. Neutrophil to Lymphocyte Ratio in Castration-Resistant Prostate Cancer Patients Treated With Daily Oral Corticosteroids. *Clin. Genitourin. Cancer* **2017**, *15*, 678–684.e1. [CrossRef] [PubMed]
56. Hsu, B.E.; Tabariès, S.; Johnson, R.M.; Andrzejewski, S.; Senecal, J.; Lehuédé, C.; Annis, M.G.; Ma, E.H.; Völs, S.; Ramsay, L.; et al. Immature Low-Density Neutrophils Exhibit Metabolic Flexibility That Facilitates Breast Cancer Liver Metastasis. *Cell Rep.* **2019**, *27*, 3902–3915.e6. [CrossRef] [PubMed]
57. Gordon-Weeks, A.N.; Lim, S.Y.; Yuzhalin, A.E.; Jones, K.; Markelc, B.; Kim, K.J.; Buzzelli, J.N.; Fokas, E.; Cao, Y.; Smart, S.; et al. Neutrophils Promote Hepatic Metastasis Growth through Fibroblast Growth Factor 2-Dependent Angiogenesis in Mice. *Hepatol. Baltim. Md* **2017**, *65*, 1920–1935. [CrossRef] [PubMed]
58. Mizuno, R.; Kawada, K.; Itatani, Y.; Ogawa, R.; Kiyasu, Y.; Sakai, Y. The Role of Tumour-Associated Neutrophils in Colorectal Cancer. *Int. J. Mol. Sci.* **2019**, *20*, 529. [CrossRef] [PubMed]

Article

SOURCE DATA | TRANSPARENT PROCESS

EMBO reports

CDK7 and MITF repress a transcription program involved in survival and drug tolerance in melanoma

Pietro Berico[1,2,3,4], Max Cigrang[1,2,3,4], Guillaume Davidson[1,2,3,4], Cathy Braun[1,2,3,4], Jeremy Sandoz[1,2,3,4], Stephanie Legras[1,2,3,4], Bujamin Hektor Vokshi[1,2,3,4], Nevena Slovic[1,2,3,4], François Peyresaubes[1,2,3,4], Carlos Mario Gene Robles[1,2,3,4], Jean-Marc Egly[1,2,3,4], Emmanuel Compe[1,2,3,4], Irwin Davidson[1,2,3,4,*] & Frederic Coin[1,2,3,4,**]

Abstract

Melanoma cell phenotype switching between differentiated melanocytic and undifferentiated mesenchymal-like states drives metastasis and drug resistance. CDK7 is the serine/threonine kinase of the basal transcription factor TFIIH. We show that dedifferentiation of melanocytic-type melanoma cells into mesenchymal-like cells and acquisition of tolerance to targeted therapies is achieved through chronic inhibition of CDK7. In addition to emergence of a mesenchymal-type signature, we identify a GATA6-dependent gene expression program comprising genes such as AMIGO2 or ABCG2 involved in melanoma survival or targeted drug tolerance, respectively. Mechanistically, we show that CDK7 drives expression of the melanocyte lineage transcription factor MITF that in turn binds to an intronic region of GATA6 to repress its expression in melanocytic-type cells. We show that GATA6 expression is activated in MITF-low melanoma cells of patient-derived xenografts. Taken together, our data show how the poorly characterized repressive function of MITF in melanoma participates in a molecular cascade regulating activation of a transcriptional program involved in survival and drug resistance in melanoma.

Keywords CDK7; TFIIH; MITF; melanoma; GATA6

Subject Categories Cancer; Chromatin, Transcription & Genomics; Signal Transduction

DOI 10.15252/embr.202051683 | Received 7 September 2020 | Revised 18 June 2021 | Accepted 25 June 2021 | Published online 23 July 2021

EMBO Reports (2021) 22: e51683

Introduction

Malignant melanoma is responsible for 70% of skin cancer deaths in Western countries (Eggermont *et al*, 2014). Somatic gain-of-function mutations in the proto-oncogene kinase *BRAF* are the commonest mutations (60%) with the T → A transversion underlying $BRAF^{V600E}$ comprising the majority of *BRAF* mutations (Brose *et al*, 2002; Davies *et al*, 2002). As an alternative to *BRAF* mutations, human melanomas frequently (35%) carry *NRAS* or *NF1* mutations, while the remainder (5%) shows no mutations of these three genes (Triple-Wt) (Hodis *et al*, 2012).

Melanoma is notorious for its heterogeneity based on co-existing melanoma cell phenotypes. *In vitro*, transcriptomic analysis of melanoma cells has established two main and distinct signatures defined as either melanocytic-type (proliferative) or mesenchymal-like (invasive) melanoma cell states (Carreira *et al*, 2006; Widmer *et al*, 2012; Verfaillie *et al*, 2015). At the transcriptional level, the differentiated melanocytic-type melanoma cells display high levels of lineage-specific transcription factors, including the *SRY-box 10* (*SOX10*) and the *MIcrophthalmia-associated Transcription Factor* (*MITF*) that drive expression of melanocyte lineage genes. Undifferentiated mesenchymal-like melanoma cells express low levels of *MITF* and *SOX10,* and their gene expression signature, including markers like the *AXL receptor tyrosine kinase* (*AXL*) and *SOX9,* is driven by AP1-TEAD factors (Verfaillie *et al*, 2015; Minnoye *et al*, 2020). The discovery of cells with intermediate signatures (Ennen *et al*, 2017; Wouters *et al*, 2020) supports the initial concept of phenotypic plasticity driving melanoma progression through conversion from one phenotype into another in response to external cues (Hoek *et al*, 2008; Ennen *et al*, 2017; Rambow *et al*, 2018).

Treatment options for patients with metastatic melanoma include combination therapies with inhibitors targeting the BRAF (i.e.,

1 Institut de Génétique et de Biologie Moléculaire et Cellulaire, Equipe Labélisée Ligue contre le Cancer, Strasbourg, France
2 Centre National de la Recherche Scientifique, UMR7104, Illkirch, France
3 Institut National de la Santé et de la Recherche Médicale, Illkirch, France
4 Université de Strasbourg, Illkirch, France
*Corresponding author. Tel: +33 3 88653445; E-mail: irwin@igbmc.fr
**Corresponding author. Tel: +33 3 88653270; E-mails: fredr@igbmc.fr

vemurafenib and dabrafenib) and MEK (i.e., trametinib) kinases (BRAFi and MEKi, respectively), whose efficiency is limited by development of resistance and subsequent progression (Menzies & Long, 2014). It is well established that tolerance to targeted therapies can involve various phenotype changes, including epithelial–mesenchymal transition(-like) (EMT) from a melanocytic to a mesenchymal state (Kemper *et al*, 2014; Arozarena & Wellbrock, 2019; Rambow *et al*, 2019). Therefore, understanding the molecular details of phenotypic plasticity and transcriptional reprograming of melanoma cells is crucial for the development of future therapeutic approaches.

Among the protein complexes essential for gene expression in eukaryotes, the basal transcription factor TFIIH is unique due to its various enzymatic activities, including helicase, translocase, and kinase functions (Villicana *et al*, 2014; Berico & Coin, 2018). The CDK7 subunit of TFIIH is a kinase that phosphorylates transcription factors, including the largest subunit of RNA polymerase II, to promote gene expression (Eick & Geyer, 2013; Compe & Egly, 2016; Fisher, 2019). Surprisingly, CDK7 kinase activity inhibition (CDK7i) elicits dramatic responses in various cancers (Cao & Shilatifard, 2014; Christensen *et al*, 2014; Kwiatkowski *et al*, 2014) probably due to the contribution of the TFIIH kinase in super-enhancer (SE)-linked oncogene transcription (Chipumuro *et al*, 2014). SEs are broad genomic regions that drive transcription of cell identity genes in normal tissue or oncogenes in cancer (Hnisz *et al*, 2013). SEs are enriched in specific transcription factors such as CDK7, Mediator, the BET family of bromodomain protein 4 (BRD4), or chromatin marks such as H3K27 acetylation (H3K27ac) (Whyte *et al*, 2013). Besides CDK7i, inhibition of BRD4 (BETi) with the small molecule JQ1 causes loss of expression for many SE-associated genes in cancer cells (Loven *et al*, 2013).

Here, we show that resistance to CDK7i correlated with melanoma cell dedifferentiation and acquisition of tolerance to BRAF and MEK inhibitors. Besides the mesenchymal-like signature, we observed the emergence of a transcription program comprising genes involved in melanoma survival and drug tolerance under the control of the *GATA-binding factor 6 transcription factor* (GATA6). CDK7 prevents the emergence of the GATA6-dependent transcription program in differentiated melanoma cells by promoting the SE-dependent expression of MITF that binds to an intronic regulatory sequence of the GATA6 locus to silence its expression. In agreement with findings in cell cultures, we observed that diminished MITF expression during human melanoma progression and phenotype switching promotes the progressive activation of GATA6 in patient-derived xenografts. We determined that GATA6 emerges in the MITF-low cells of the PDX showing invasive or interferon γ (IFNγ)-active phenotypes.

Results

Melanoma cultures exhibit distinct sensitivity to CDK7i

We explored the sensitivity of melanoma cells to CDK7i using cells with the two main phenotypes and most common driver mutations. The melanocytic-type patient-derived MM011 ($NRAS^{Q61K}$), MM074 ($BRAF^{V600E}$), MM117 (Triple-wt) cell cultures and the melanoma 501mel cell line ($BRAF^{V600E}$) exhibited moderate to high expression of the lineage-specific transcription factors MITF and SOX10 together with low to undetectable levels of SOX9 and c-JUN (Widmer *et al*, 2012; Verfaillie *et al*, 2015) (Fig 1A). In contrast, patient-derived MM029 ($BRAF^{V600K}$), MM047 ($NRAS^{Q61R}$), and MM099 ($BRAF^{V600E}$) cell cultures showed a mesenchymal-like phenotype characterized by low to undetectable levels of MITF and SOX10 coupled to high levels of SOX9 and c-JUN (Widmer *et al*, 2012; Verfaillie *et al*, 2015; Wouters *et al*, 2020). We observed that all melanocytic-type cells together with the MM047 mesenchymal-like cells were sensitive to low concentrations of THZ1, the first-in-class selective and covalent inhibitor of CDK7 (Kwiatkowski *et al*, 2014) (Fig 1B). In marked contrast, the MM099 and MM029

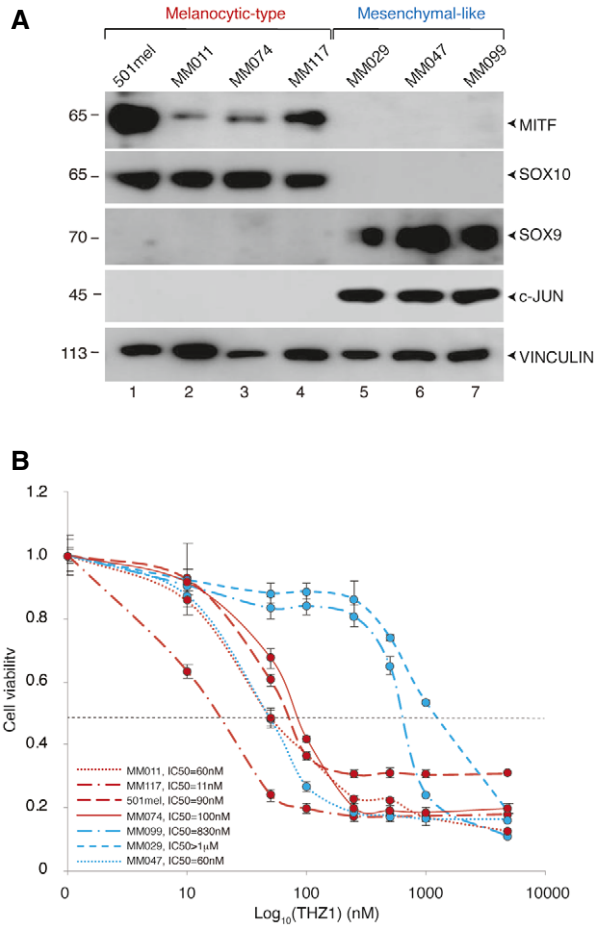


Figure 1. Melanoma cells show differential sensitivity to CDK7i.

A Protein lysates from the melanocytic-like melanoma cells 501mel, MM011, MM074, and MM117 or the mesenchymal-like melanoma cells MM029, MM047, and MM099 were immuno-blotted for proteins as indicated. Molecular mass of the proteins is indicated (kDa).

B Melanoma cells were treated with increasing concentrations of THZ1 as indicated for 72 h. Mean growth is shown relative to vehicle (DMSO)-treated cells. IC50 for each cell line is indicated. Melanocytic-type (MITF-High, proliferative) melanoma cells are shown in red, while mesenchymal-like (MITF-low, invasive) melanoma cells are shown in blue.

Data information: In (B), data are presented as mean values + standard deviation (SD) for three replicates (n = 3).
Source data are available online for this figure.

mesenchymal-like cells were tolerant to CDK7i, even at high concentrations of the drug. These observations demonstrated that melanocytic-type melanoma cells were highly sensitive to CDK7i, regardless of their driver mutation, while some mesenchymal-like melanoma cells were tolerant to the drug.

CDK7i promotes dedifferentiation of melanoma cells

To investigate the role of CDK7 in melanoma cells, we generated several THZ1-resistant (CDK7i) or vemurafenib-resistant (BRAFi) cell lines *ex vivo (*MM074$^{CDK7i\text{-}R}$, MM074$^{BRAFi\text{-}R}$, and MM047$^{CDK7i\text{-}R}$) (Fig EV1A–C). Establishment of CDK7i resistance decreased sensitivity of the MM074$^{CDK7i\text{-}R}$ to BRAFi (vemurafenib) and MEKi (trametinib) (Fig EV1B and D), while the BRAFi-resistant MM074$^{BRAFi\text{-}R}$ remained sensitive to both CDK7i and MEKi (Fig EV1A and D). In agreement with the involvement of CDK7 in gene expression, global transcription activity of MM047 and MM074 was strongly impacted by CDK7i treatment, in contrast to MM047$^{CDK7i\text{-}R}$ and MM074$^{CDK7i\text{-}R}$ where global transcription was not inhibited (Fig EV1E).

RNA-seq revealed a pronounced modification of the transcriptional programs of MM074$^{CDK7i\text{-}R}$ and MM074$^{BRAFi\text{-}R}$ compared to the parental MM074, but a less pronounced modification of the MM047$^{CDK7i\text{-}R}$ compared to MM047 (Fig 2A). More than 6,000 genes were deregulated in MM074$^{CDK7i\text{-}R}$ compared to MM074 and 1,000 genes in MM047$^{CDK7i\text{-}R}$ compared to MM047 (Fig 2B). Despite the fact that the parental cells were of different phenotypes, 261 genes were commonly up-regulated in the two CDK7i-resistant cell cultures (Fig 2B and Dataset EV1). We hereafter defined these genes as the "CDK7i-resistant signature" (K7iRS). As shown by Gene Ontology (GO) analysis, these genes were involved in epithelial cell differentiation or in the transport of small molecules (Appendix Fig S1).

We next clustered melanoma cells based on the expression of a hundred genes corresponding to previously described signatures of melanocytic *vs* mesenchymal transcriptional cell states (Widmer *et al*, 2012). In agreement with the literature (Verfaillie *et al*, 2015; Wouters *et al*, 2020), 501mel and MM074 showed a melanocytic-type transcriptional signature (Fig 2C, lanes 1–2), while the MM047, MM099, and MM029 cells showed a mesenchymal-like signature (lanes 5–7). Surprisingly, chronic exposure of MM074 to CDK7i induced the emergence of a stable mesenchymal-like signature (compare lane 2 with 4) correlating with increased invasion capacity (Fig EV1F). In apparent contrast with MM074$^{CDK7i\text{-}R}$, the melanocytic-type signature of MM074 persisted in MM074$^{BRAFi\text{-}R}$ where we further observed a significant increase in the expression of a set of *bone fide* pigmentation genes (Fig 2C, compare lane 2 with 3). RT–qPCR confirmed the increased expression of genes involved in pigmentation such as *MLANA* in MM074$^{BRAFi\text{-}R}$ (Fig EV1G), which correlated with higher cellular pigmentation (Fig EV1H). In agreement with mRNA, we observed that the MM074$^{BRAFi\text{-}R}$ exhibited significantly higher amounts of the melanocyte lineage-specific proteins MITF and TFAP2A compared to MM074 (Fig 2D, compare lane 1 with 2). In contrast, MM074$^{CDK7i\text{-}R}$ showed a dramatic decrease of these proteins together with the emergence of SOX9 (compare lane 1 with 3).

Altogether, these data showed that MM074 melanocytic-type cells chronically exposed to CDK7i dedifferentiated to adopt a mesenchymal state, whereas those exposed to BRAFi acquired a highly pigmented hyper-differentiated cell state. Furthermore, both mesenchymal-like and melanocytic-type melanoma cells chronically exposed to CDK7i displayed common altered expression of 261 genes corresponding to the K7iRS.

A GATA6-dependent transcription program in CDK7i-resistant melanoma cells

We compared the MM074$^{CDK7i\text{-}R}$ and MM047$^{CDK7i\text{-}R}$ gene expression programs to potentially identify a signature involved in drug tolerance that emerges as melanocytic-type cells undergo a phenotype switch and that is shared with the drug-resistant mesenchymal cells. This comparison focusing on genes commonly regulated during drug tolerance bypassing the much larger number of genes characterizing the phenotype switch *per se* identified the K7iRS genes. Merging these genes with a list of annotated transcription factors identified 16 common up-regulated transcription factors (TFs) in MM074$^{CDK7i\text{-}R}$ and MM047$^{CDK7i\text{-}R}$ (Fig EV2A). Analysis of their expression in RNA-seq data from melanoma cells showed that only 4 were significantly more expressed in the CDK7i-resistant MM029, MM099, MM074$^{CDK7i\text{-}R}$, and MM047$^{CDK7i\text{-}R}$ cells, compared to CDK7i-sensitive cells (Fig EV2B). Of these, only GATA6 was significantly overexpressed in primary melanoma *vs* nevi (Fig EV2C). We confirmed by RT–qPCR and immuno-blot higher levels of *GATA6* mRNA and GATA6 protein, respectively, in the CDK7i-insensitive cells (Fig 3A and B). We also noted that

Figure 2. Exposure to CDK7i induces melanoma dedifferentiation.

A Volcano plots were used to demonstrate differentially expressed genes as determined by RNA-seq in either MM047$^{CDK7i\text{-}R}$ *vs* MM047 (top), MM074$^{CDK7i\text{-}R}$ *vs* MM074 (middle), or MM074$^{BRAFi\text{-}R}$ *vs* MM074 (bottom). Red dots show significantly over-represented (top) or under-represented (bottom) RNAs in drug-resistant cells compared to parental cells. All data were evaluated with the DESeq2 R package. The value for a given gene is the normalized gene expression value relative to the mean of all samples belonging to the same condition.

B Proportional Venn diagrams indicating the number of up-regulated (top) and down-regulated (bottom) genes in MM047$^{CDK7i\text{-}R}$ and MM074$^{CDK7i\text{-}R}$ compared to the parental MM047 and MM074, respectively. The number of genes overlapping between the datasets is indicated. 261 genes were found up-regulated and 241 down-regulated in MM047$^{CDK7i\text{-}R}$ and MM074$^{CDK7i\text{-}R}$. Hypergeometric *P*-value is indicated.

C Genes characterizing the melanocytic-type and mesenchymal-like transcription signatures (Widmer *et al*, 2012) have been plotted on a heatmap and are shown in relation to their expression in different melanoma cells. RPKM values are represented as *z*-score. The group of genes related to pigmentation has been highlighted in red. The color key shows the $\log_2$ expression values. Yellow color stands for high expression and dark violet for low expression.

D Protein lysates from MM074, MM074$^{BRAFi\text{-}R}$, or MM074$^{CDK7i\text{-}R}$ were immuno-blotted for indicated proteins. Molecular sizes of the proteins are indicated (kDa). The numbers below the gel lanes represent relative protein level, which was determined from the band intensity using ImageJ software and normalized to each relative vinculin control.

Source data are available online for this figure.

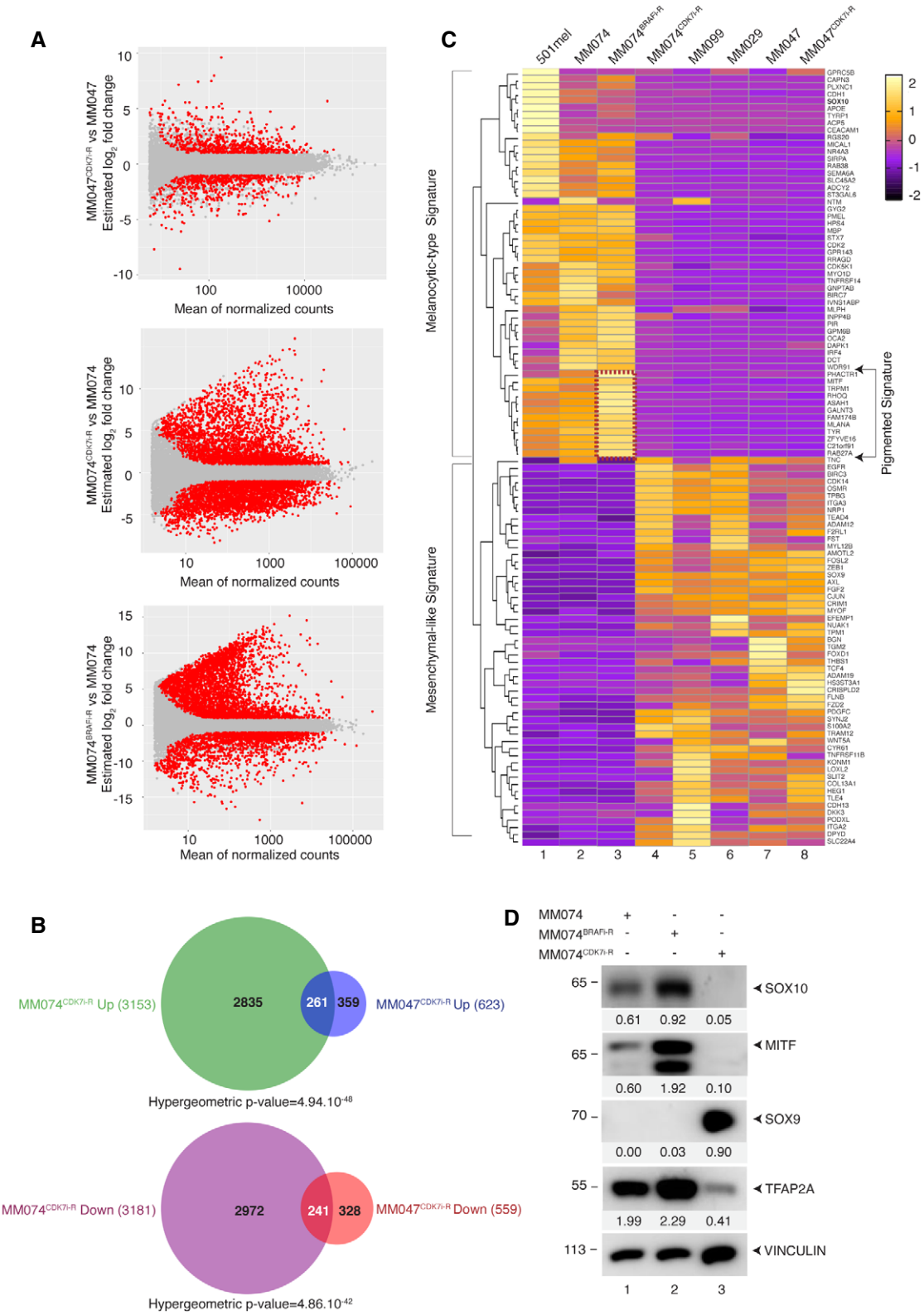


Figure 2.

GATA6 protein in MM047-sensitive cells was much lower than in the tolerant MM029 and MM099 cells and was potently induced in MM047$^{CDK7i\text{-}R}$ (Fig 3B).

We then analyzed the transcriptomic profiles of the CDK7i-insensitive mesenchymal-like MM099 cells in which GATA6 was depleted using siRNA and observed a significant down-regulation of

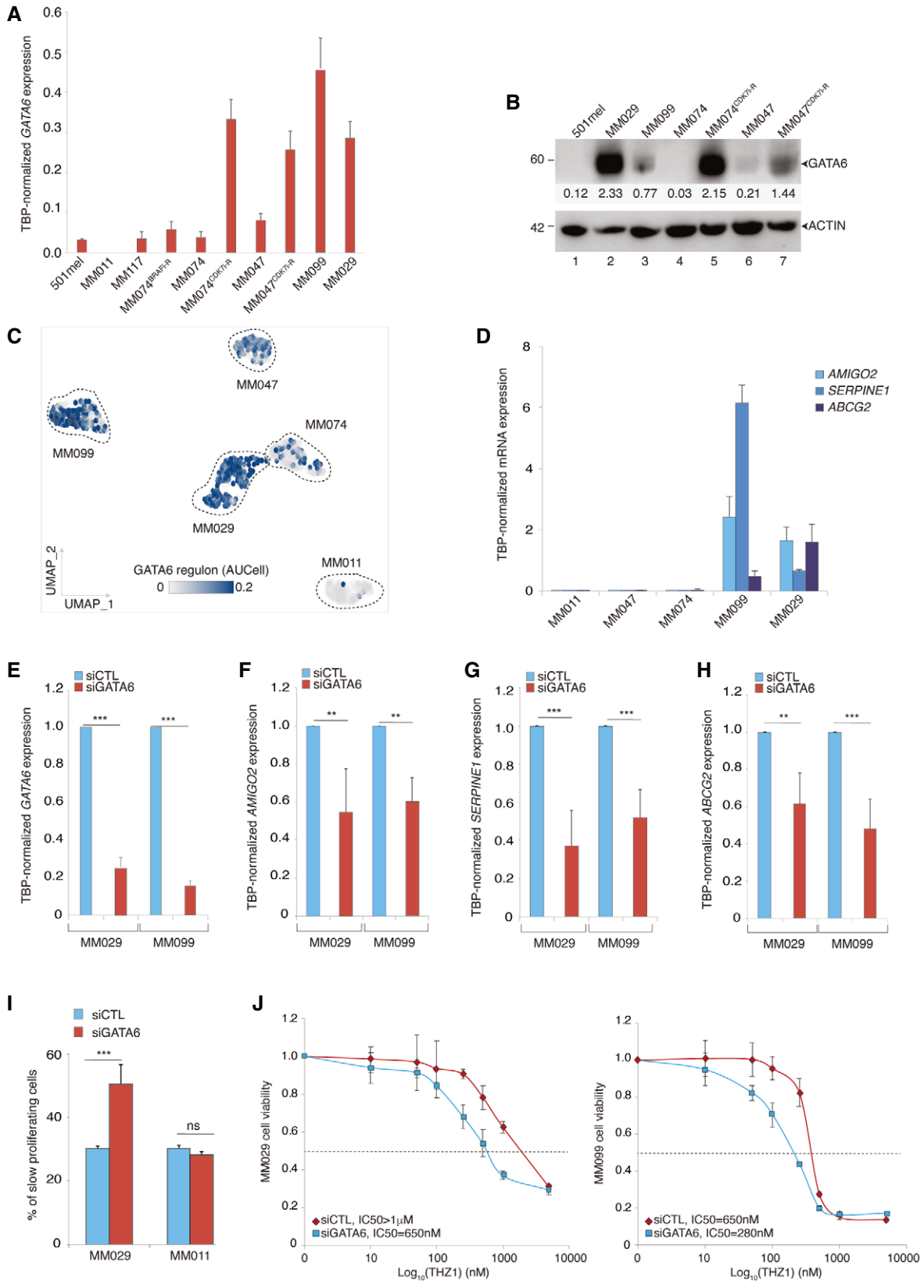


Figure 3.

Figure 3. GATA6 and its regulon are expressed in CDK7i-tolerant melanoma cells.

A qRT–PCR analysis showing average *TBP*-normalized expression of *GATA6* in the indicated cells.

B Protein lysates from the indicated cells were immuno-blotted for the indicated proteins. Molecular sizes of the proteins are indicated in kDa. The numbers below the gel lanes represent relative protein level, which was determined from the band intensity using ImageJ software and normalized relative to each relative actin control.

C UMAP (Uniform Manifold Approximation and Projection) dimension reduction representative of GATA6 regulon expression in MM011, MM029, MM047, MM074, and MM099 from Wouters dataset (Wouters *et al*, 2020). UMAP is colored according to the continuous GATA6 AUCell values (from 0 to 0.2).

D qRT–PCR analysis showing average *TBP*-normalized expression of *AMIGO2, SERPINE1, and ABCG2* in the indicated cells.

E–H qRT–PCR analysis showing average *TBP*-normalized expression of *GATA6* (E), *AMIGO2* (F), *SERPINE1* (G), and *ABCG2* (H) in the indicated cells treated with either siCTL or si*GATA6* for 72 h.

I MM011 and MM029 were treated with either siCTL or si*GATA6* for 72 h. Cell proliferation was analyzed using CellTrace staining and flow cytometry in the indicated cell lines, and the % of slow proliferating cells is shown for each condition.

J MM029 (left) and MM099 (right) were pre-treated with either siCTL or si*GATA6* for 48 h and treated with increasing concentrations of THZ1 for 72 h. Mean growth is shown relative to vehicle (DMSO)-treated cells.

Data information: In (A, D-J), data are presented as mean values + SD for six replicates ($n = 6$). The *P*-value (Student's *t*-test) is indicated, ** < 0.01, *** < 0.005, and ns, non-significant.
Source data are available online for this figure.

86 genes following GATA6 silencing (defined below as "GATA6 regulon") (Dataset EV2). We next examined expression of the GATA6 regulon in single-cell transcriptomic data recently obtained from MM011, MM029, MM047, MM074, and MM099 (Wouters *et al*, 2020). The GATA6 regulon was more enriched in CDK7i-insensitive MM099 and MM029 cells compared to the others (Fig 3C). Within the GATA6 regulon, we identified genes whose function was previously defined as important for melanoma such as the *Adhesion Molecule with Ig like dOmain 2* (*AMIGO2*) (Fontanals-Cirera *et al*, 2017) and the *SERPIN family E member* 1 (*SERPINE1*) (Klein *et al*, 2012) together with genes contributing to multidrug resistance in cancer cells such as the efflux pump *ATP Binding Cassette Subfamily G member 2* (*ABCG2*) (Robey *et al*, 2018). RT–qPCR showed that expression of these genes was significantly higher in the CDK7i-resistant MM099 and MM029 cells compared to MM011, MM047, and MM074 (Fig 3D). As in MM099, GATA6 depletion in MM029 decreased expression of *AMIGO2*, *SERPINE1,* and *ABCG2* (Fig 3E–H). Furthermore, GATA6 depletion diminished MM029 proliferation compared to MM011 (Figs 3I and EV2D) and sensitized MM029 and MM099 to CDK7i (Figs 3J and EV2D). We also tried to overexpress GATA6 in melanocytic-like cells; however, its expression was toxic in these cells, leading to cell cycle arrest that compromised the isolation of stably expressing clones. Therefore, we overexpressed GATA6 in the mesenchymal-like MM047 cells and obtained stable expression of GATA6 (Fig EV2E). Ectopic expression of GATA6 induced expression of *ABCG2, AMIGO2,* and *SERPINE1* (Fig EV2E and F) and increased resistance to CDK7i (Fig EV2G). These data suggested that GATA6 coordinated the expression of a set of genes specifically expressed in drug-tolerant mesenchymal-like melanoma cells and required for proliferation/survival and drug resistance.

ABCG2 is involved in tolerance to CDK7i and BRAFi in melanoma cells

The above data suggest that up-regulation of *ABCG2* expression by GATA6 in mesenchymal-like melanoma cells may promote CDK7i resistance. RNA-seq data from melanoma tumors and *in situ* mRNA hybridization of melanoma tumor sections demonstrated higher expression of *ABCG2* in cutaneous metastatic melanoma compared to primary tumors (Fig EV3A and Appendix Fig S2). Three ABC transporters (*ABCG2, ABCB1,* and *ABCC3*) were up-regulated in $MM047^{CDK7i\text{-}R}$ and/or $MM074^{CDK7i\text{-}R}$ (Fig EV3B and C), but only *ABCG2* was overexpressed in the CDK7i-insensitive MM099 and MM029 (Fig 4A and B). Depletion of ABCG2 using siRNA (Fig EV3D) significantly sensitized MM099 and MM029 to CDK7i (Fig 4C and D). Interestingly, depletion of ABCG2 also sensitized MM099 cells to BRAFi (Fig 4E), showing the potential pleiotropic impact of this efflux pump on drug resistance. Consistently, decrease of ABCG2 in MM029 cells did not impact their sensitivity to BRAFi since they harbored the vemurafenib-resistant $BRAF^{V600K}$ mutation (Fig 4F). Taken together, these data suggested that the ABC transporter ABCG2 played a significant role in tolerance to CDK7i and BRAFi in melanoma cells.

CDK7 regulates expression of MITF and SOX10

We investigated the regulation of GATA6 regulon that was repressed in melanocytic melanoma cells and activated by chronic exposure to CDK7i. Previous work suggested that CDK7 occupied SEs regulating MITF and SOX10 expression in melanoma cells (Eliades *et al*, 2018), but the presence of CDK7 at MITF/SOX10-associated super-enhancers was not observed so far. We performed ChIP-seq chromatin profiling of CDK7 using 501mel where the CDK7 locus was tagged with a Biotin-3xFlag tag by CRISPR/Cas9 genome editing ($501mel^{BIO\text{-}FLAG:CDK7}$) (Appendix Fig S3). FLAG ChIP-seq identified numerous CDK7-binding sites throughout the *MITF* locus and of its transcriptional activator *SOX10* (Fig 5A and B). CDK7 occupancy co-localized with H3K27ac, binding of MITF and/or of SOX10, BRG1, or H2AZ, all characterizing SE elements. A short 24 h CDK7i treatment impaired *MITF* and *SOX10* expression in 501mel, whereas exposure to BETi JQ1 had no effect (Fontanals-Cirera *et al*, 2017) (Fig 5C and D and Appendix Fig S4). Interestingly, decrease of *MITF* and *SOX10* following CDK7i occurred in parallel with increased expression of *GATA6* (Fig 5E). Moreover, expression of *CDK7, MITF,* and *SOX10* anti-correlated with that of *GATA6* in published RNA-seq data from human patient cutaneous melanoma (SKCM from TCGA) (Appendix Fig S5A).

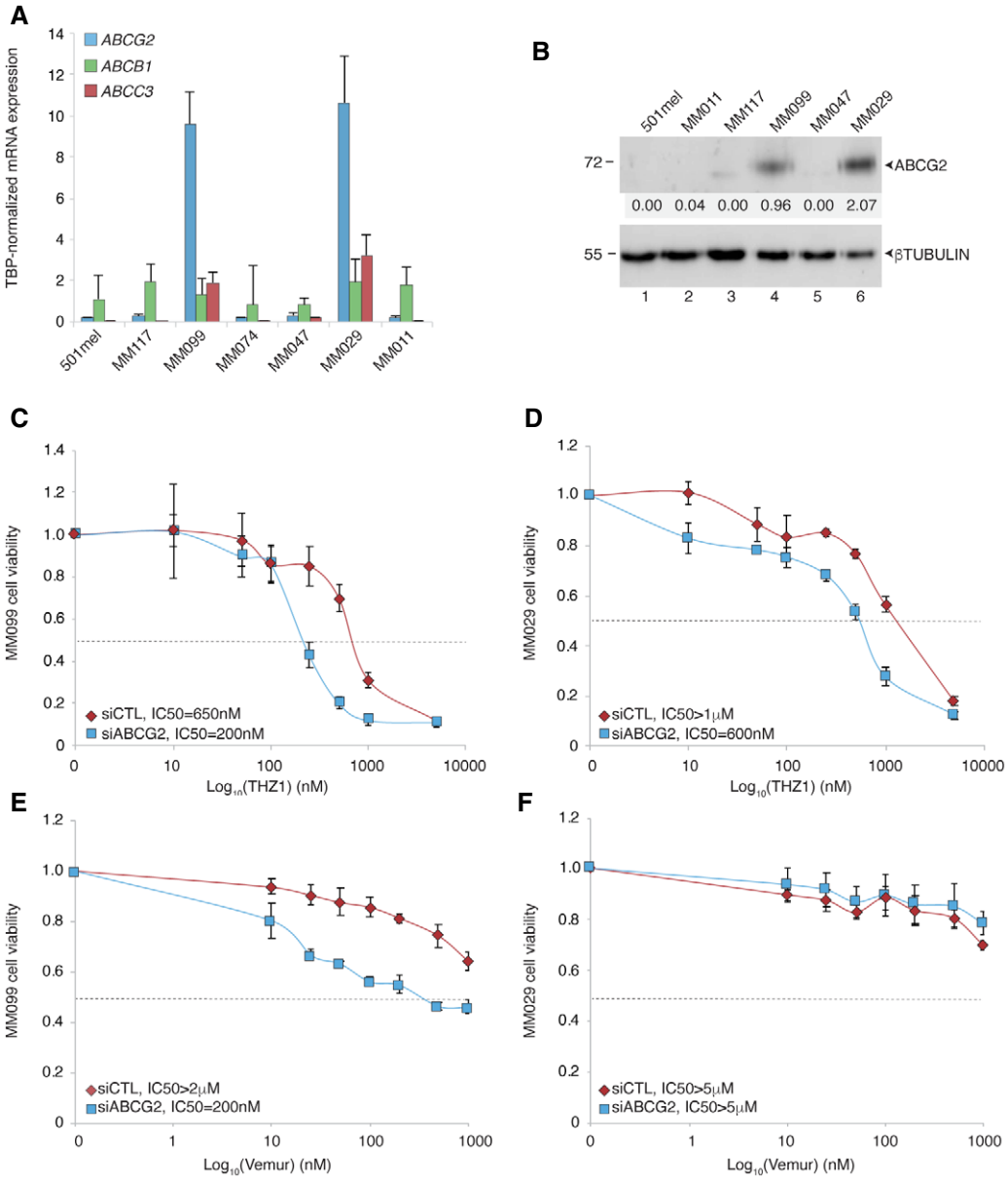


Figure 4. ABCG2 is involved in multidrug tolerance in melanoma cells.

A qRT–PCR analysis showing average *TBP*-normalized expression of *ABCB1*, *ABCC3*, and *ABCG2* in the indicated cells.

B Protein lysates from the indicated cells were immuno-blotted for the indicated proteins. Molecular masses of the proteins are indicated in kDa. The numbers below the gel lanes represent relative protein level, which was determined from the band intensity using ImageJ software and normalized to each relative β-tubulin control.

C, D MM099 (C) and MM029 (D) were pre-treated with either siCTL or si*ABCG2* as indicated and treated with increasing concentrations of THZ1 for 72 h. Mean growth is shown relative to vehicle (DMSO)-treated cells.

E, F MM099 (E) and MM029 (F) were pre-treated with either siCTL or si*ABCG2* as indicated and treated with increasing doses of vemurafenib for 72 h. Mean growth is shown relative to vehicle (DMSO)-treated cells.

Data information: In (A), data are presented as mean values + SD for six replicates (n = 6). In (C-F), data are presented as mean values + SD for three replicates (n = 3). IC50 for each cell line is indicated.
Source data are available online for this figure.

SOX10 silencing induces release of GATA6 regulon expression

The above data suggested that decreased MITF and/or SOX10 expression may induce *GATA6* expression. To test this, we depleted SOX10 with siRNA in 501mel cells and observed a significant decrease of *MITF* and induction of *GATA6* expression (Fig 6A and Appendix Fig S5B). In agreement, bioinformatic analyses of published scRNA-seq performed at different times after SOX10 depletion in melanocytic-like

MM074 cells (Wouters *et al*, 2020) showed concomitant activation of the GATA6 regulon (Fig 6B and C and Appendix Fig S5C). We observed that progressive *SOX10* and *MITF* down-regulation (Fig 6D and E) correlated with concomitant up-regulation of *GATA6*, *ABCG2*, *SERPINE1*, and *AMIGO2* (Fig 6F–I). Altogether, these data showed an antagonism between MITF/SOX10 and GATA6 regulon in melanoma.

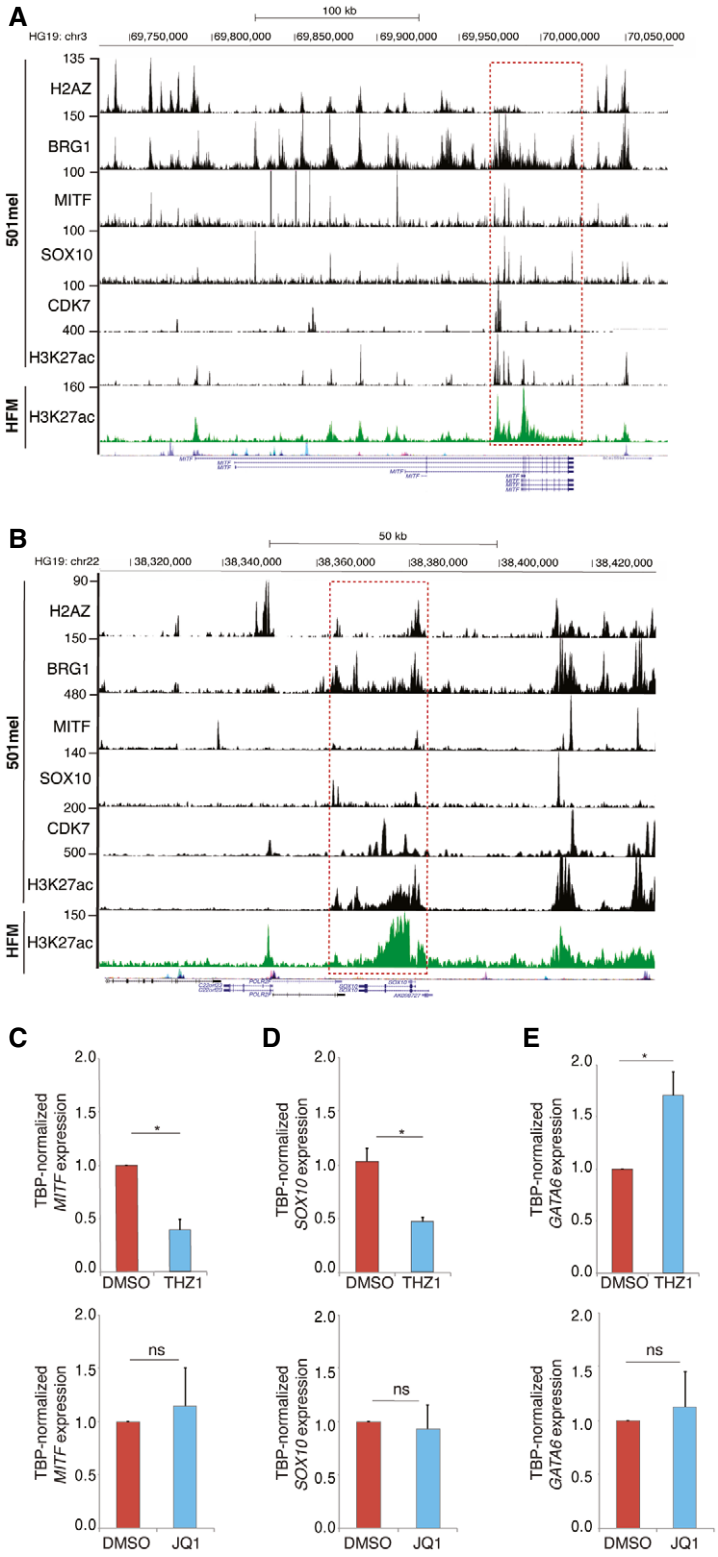


Figure 5.

Figure 5. CDK7i induced inhibition of MITF and SOX10 and release of GATA6 expression.

A, B Gene track of CDK7 occupancy at *MITF* (A) or *SOX10* (B) loci in 501mel$^{BIO\text{-}FLAG:CDK7}$ cell line. Gene tracks of H2A.Z, BRG1, MITF, SOX10, and H3K27ac (GSE94488 and GSE61967) at the same loci in parental 501mel are indicated. SE is denoted by a red opened square. H3K27ac deposition is also shown in Hair Follicle Melanocytes (HFM) (GSE94488).

C–E qRT–PCR analysis showing average *TBP*-normalized fold expression of *MITF* (C), *SOX10* (D), and *GATA6* (E) in 501mel treated with either DMSO/THZ1 (50 nM) (upper) or DMSO/JQ1 (10 μM) (lower) for 24 h.

Data information: In (C-E), data are presented as mean values + SD for six replicates ($n = 6$). The *P*-value (Student's *t*-test) is indicated, * < 0.05.

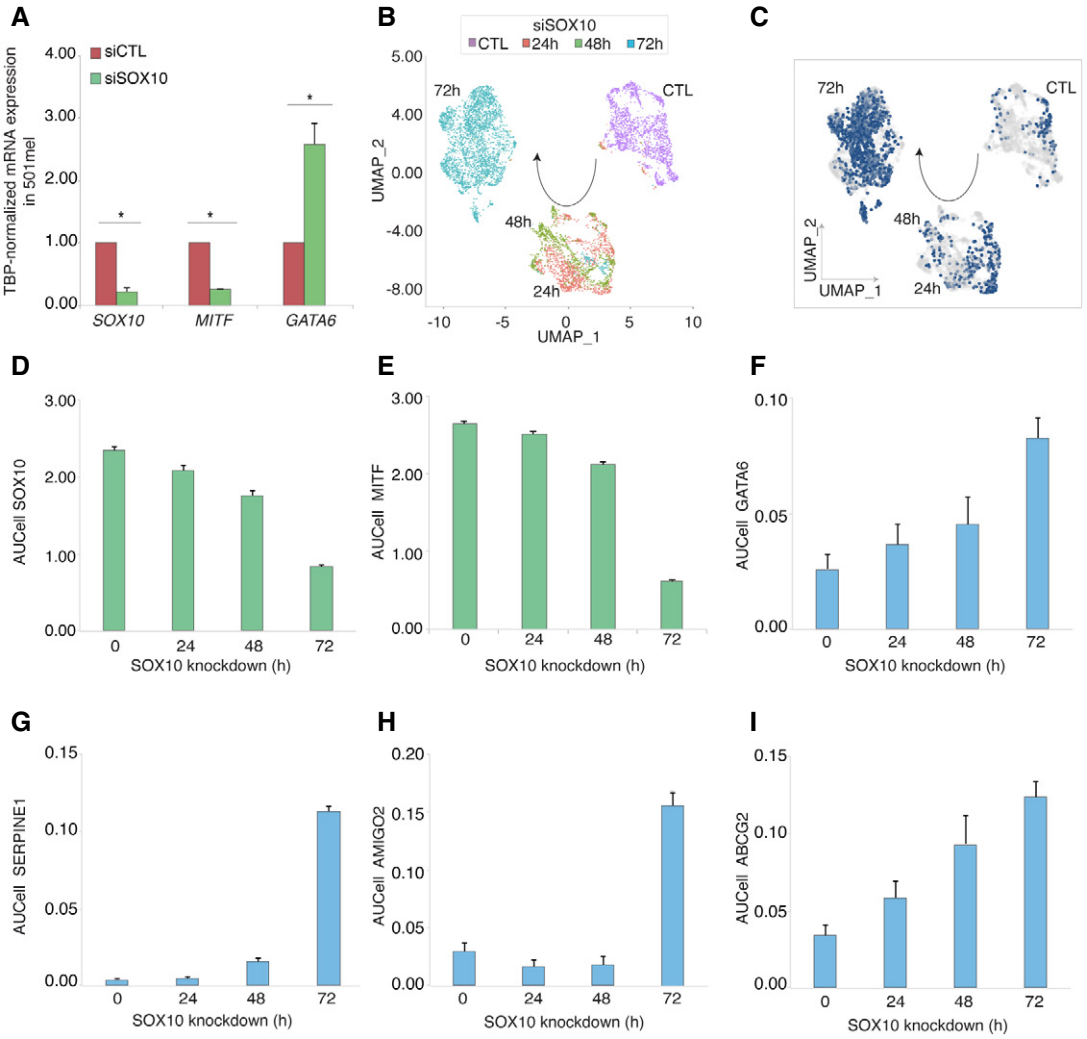


Figure 6. Loss of SOX10 and MITF releases GATA6 expression.

A qRT–PCR analysis showing average *TBP*-normalized fold expression of *SOX10*, *MITF*, and *GATA6* in 501mel treated with either siCTL or si*SOX10* for 48 h.

B Seurat UMAP of MM074 treated with siCTL or siSOX10 (24, 48, and 72 h post-treatment). The arrow indicated the trajectory from control to 72 h post-siSOX10 transfection.

C UMAP of AUCell GATA6 regulon in MM074 shows that GATA6 regulon is up-regulated along the trajectory from siCTL to 72 h post-siSOX10 treatment (GSE116237) (Wouters *et al*, 2020). The arrow indicated the trajectory from siCTL to 72 h post-siSOX10 transfection. We considered cell with GATA6 regulon activity of AUCell > 0.15 as active (see Appendix Fig S5).

D–I Graphs showing the average expression of the *SOX10* (D), *MITF* (E), *GATA6* (F), *SERPINE1* (G), *AMIGO2* (H), and *ABCG2* (I) per individual melanoma cell measured by AUCell on MM074 at different time points post-transfection of si*SOX10* (GSE116237) (Wouters *et al*, 2020).

Data information: In (A), data are presented as mean values + SD for six replicates ($n = 6$). The *P*-value (Student's *t*-test) is indicated, * < 0.05. In (D-I), data are presented as mean values + standard error of the mean (SEM) for six replicates ($n = 6$).

MITF drives direct transcriptional repression of GATA6

The above data suggested a direct mechanistic link between SOX10 and/or MITF and the repression of *GATA6* in melanoma cells. ChIP-seq did not reveal SOX10 binding at the *GATA6* locus in 501Mel cells; however, a prominent MITF-binding site was observed in an intronic region of the *GATA6* gene body (hereafter called "intGATA6r", for intronic GATA6 locus region) containing potential

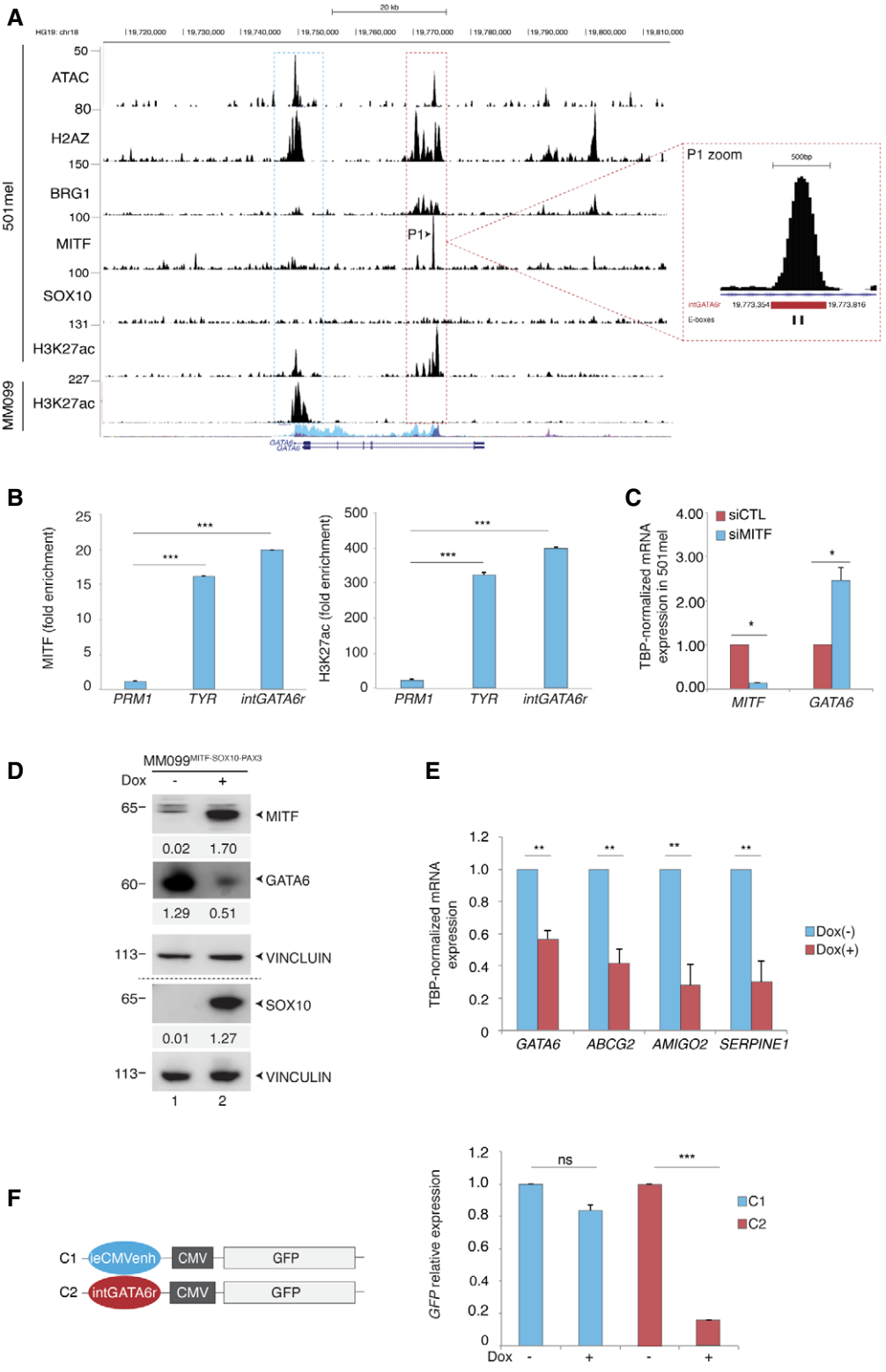


Figure 7.

Figure 7. MITF binds and represses the GATA6 locus.

A ChIP-seq track of 3HA-MITF signal occupancy showing a significant MITF-binding peak (P1) in the GATA6 gene body in 501mel (GSE61967). Additional tracks indicate potential regulatory regions highlighted by ATAC-seq and H3K27ac, BRG1, SOX10, and H2A.Z deposition (GSE94488 and GSE61967). H3K27ac deposition is also shown in MM099 at the *GATA6* locus. The scale bar indicates the size of the genomic region in kilobases (Kb). A magnification of the P1 region for MITF occupancy is shown in which the "intGATA6r" region is indicated in red and the two E-boxes in black.

B ChIP qPCR experiment monitoring the fold enrichment (compare to control IgG) of MITF protein and H3K27ac mark at the "intGATA6r" region. *Proteamine 1* (*PRM1)* and Tyrosinase (*TYR)* regulatory regions were used as negative and positive controls, respectively (Laurette *et al*, 2015).

C qRT–PCR analysis showing average *TBP*-normalized fold expression of *MITF* and *GATA6* in 501mel treated with either siCTL or si*MITF* for 48 h.

D MM099$^{MITF\text{-}SOX10\text{-}PAX3}$ expressing inducible *MITF-SOX10-PAX3* genes was treated or not with doxycycline (1 μg/ml) for 24 h, and protein lysates were immuno-blotted for the indicated protein. The numbers below the gel lanes represent relative protein level, which was determined from the band intensity using ImageJ software and normalized relative to vinculin control.

E qRT–PCR analysis showing average *TBP*-normalized fold expression of *GATA6*, *ABCG2*, *AMIGO2*, or *SERPINE1* in MM099$^{MITF\text{-}SOX10\text{-}PAX3}$ treated or not with doxycycline (1 μg/ml) for 24 h.

F Left panel: Schematic representation of pCDNA-ieCMVenh-CMV-GFP (C1) or pCDNA-intGATA6r-CMV-GFP (C2) reporter vectors. The ieCMVenh sequence in C1 was replaced by the "intGATA6r" sequence to generate C2. Right panel: qRT–PCR analysis showing average *TBP*-normalized fold expression of *GFP* in MM099$^{MITF\text{-}SOX10\text{-}PAX3}$ transfected with C1 or C2 vectors for 48 h before treatment or not with doxycycline (1 μg/ml) for 24 h.

Data information: In (B, C, E), data are presented as mean values + SD for three biological triplicates. The *P*-value (Student's *t*-test) is indicated, * < 0.05, ** < 0.01, and *** < 0.001. In (F), data are presented as mean values + SD for three technical replicates ($n = 6$). The *P*-value (Student's *t*-test) is indicated, *** < 0.005 and ns, non-significant (> 0.05).
Source data are available online for this figure.

MITF-binding sites (E-box motifs) (Fig 7A) (Laurette *et al*, 2015). In addition, "IntGATA6r" was enriched in H3K27ac, BRG1, and H2AZ, marks of enhancer elements. Interestingly, in MM099 where MITF is not expressed, intronic H3K27ac was lost, but rather replaced by strong H3K27ac labeling at the *GATA6* promoter, correlating with its high expression in these cells. ChIP-qPCR confirmed enrichment of MITF and H3K27ac at the "intGATA6r" region in 501mel (Fig 7B). In agreement with a role for MITF in *GATA6* repression, siMITF silencing in 501mel induced *GATA6* expression (Fig 7C and Appendix Fig S5B).

To determine whether MITF was able to transcriptionally repress *GATA6*, we generated MM099$^{MITF\text{-}SOX10\text{-}PAX3}$ in which MITF, SOX10, and PAX3 expression could be induced by doxycycline (Dox) treatment (Fig 7D). We co-expressed MITF, SOX10, and PAX3 as we observed that the presence of SOX10 and PAX3 stabilized MITF in these cells. Following induction of MITF-SOX10-PAX3, *GATA6* mRNA expression was repressed and level of GATA6 protein decreased (Fig 7D and E). Consequently, the expression of the GATA6 regulon genes *ABCG2*, *AMIGO2,* and *SERPINE1* was inhibited (Fig 7E).

To establish its repressive role, the 500bp "intGATA6r" sequence was inserted upstream of the CMV promoter of the pcDNA-CMV vector to replace the immediate early CMV enhancer ("ieCMVenh") in the context of a GFP reporter vector (Fig 7F, left panel). The reporter construct was transiently transfected into MM099$^{MITF\text{-}SOX10\text{-}PAX3}$ with or without Dox-induced MITF-SOX10-PAX3 expression. While ieCMVenh-dependent GFP expression was barely affected by MITF-SOX10-PAX3 expression, the presence of the "intGATA6r" element upstream of the promoter strongly impacted expression of the GFP compared to cells that did not express MITF-SOX10-PAX3 (Fig 7F, right panel). Altogether, these data strongly suggested that MITF transcriptionally repressed *GATA6* by binding to a negative regulatory sequence located in an intronic region of *GATA6*.

GATA6 is expressed in MITF-low cells of human melanoma

Our *in vitro* data suggested that GATA6 and its regulon may be expressed in MITF-low melanoma cells in human tumors. To test this hypothesis, we first performed an immunohistological (IHC) examination of human tumor samples. Because MITF antibodies are poorly efficient in IHC, we rather detected its transcriptional activator SOX10. While GATA6 was not observed in nevi and primary melanomas that showed high SOX10 expression (Fig 8A, panels a-d), it was highly expressed in a subpopulation of cells in cutaneous metastases that did not express SOX10 (Fig 8A, panels e-f). In line with the above data, analyses of public DNA microarray (Xu *et al*, 2008) or RNA-seq data (TCGA) consistently revealed higher expression of *GATA6* in metastatic melanoma compared with primary melanoma (Fig EV4A).

To further define which melanoma cell subtypes express *GATA6* and its regulon, we re-analyzed scRNA-seq data from a PDX tumor before and after BRAFi (dabrafenib) and MEKi (trametinib) combination treatment (Rambow *et al*, 2018). An unsupervised gene clustering analysis that included more cells than in the original published analyses detected 9 different cell subpopulation clusters (Fig 8B). GO analysis attributed the four previously well characterized phenotypes to clusters 1, 5, 7, and 8, specifically starved-like melanoma cells (SMC), pigmented, invasive and neural crest-stem cells (NCSC) characterized by many of the previously described genes of each signature (Fig 8C). We attributed two additional phenotypes to clusters 4 and 6 that we defined as "Mitotic" due to the high expression of late S-phase and G2 M phase genes and "IFN-active" (previously designated as Immune (Rambow *et al*, 2018)) due to the enriched expression of interferon response genes. Cluster 0 corresponded to MITF-intermediate proliferative cells, while no specific ontology could be assigned to clusters 2 and 3 that were characterized by high expression of mitochondrial or pseudogenes, respectively, and were excluded from subsequent analyses.

The frequency of cells of each phenotype was then analyzed at the different phases defined by Rambow before and after MAPKi exposure (T0 is the drug-naïve phase, phases 1 and 2 are the minimal residual disease phase (MRD), and phase 3 is the development of drug resistance (Rambow *et al*, 2018)). As previously described, an increase in SMC, pigmented and NCSC at minimal residual disease (MRD) phases 1 and/or 2 was observed (Fig EV4B), while the frequency of mitotic cells was strongly reduced in phases 1 and

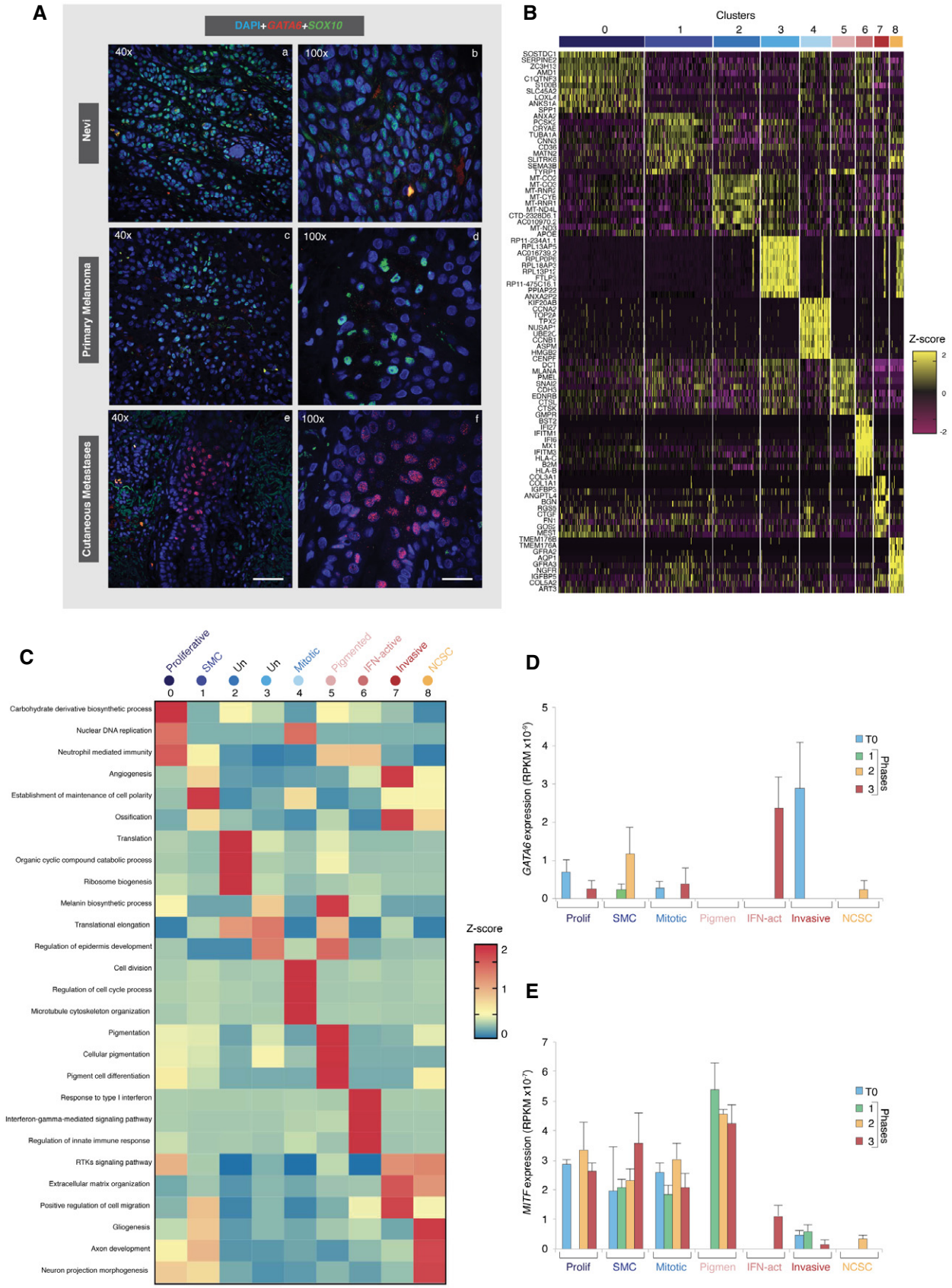


Figure 8.

Figure 8. GATA6 is expressed in MITF-low melanoma cells *in vivo*.

A Tumor sections were immuno-labeled (IHC) with anti-GATA6 (red) and anti-SOX10 (green) antibodies, and images were captured by confocal microscopy at the indicated magnification. We analyzed six tumor sections of metastases and observed significant GATA6 expression in only one of them. Scale bar 250 μm for 40× and 100 μm for 100×.

B Seurat cluster heatmap was generated from published scRNA-seq performed on PDX tumor (n = 674 cells) (GSE116237) (Rambow *et al*, 2018). The heatmap shows 9 different clusters into which the cells can be divided according to the expression of different referenced genes (Z-score). The top 10 genes are indicated in the left for each cluster.

C GO was used to analyze the genes characterizing each cluster identified above. The average *P*-value was retrieved for each cluster taking the 3 best GO per cluster, and then, z-score ((*P*-value of each biological process-average of *P*-value of each biological process)/standard deviation) was calculated. Clusters 2 and 3 were undefined (un).

D, E Graphs showing the average expression of *GATA6* (D) and *MITF* (E) (RPKM) for each phenotype cluster in T0 (drug naïve) (blue), phases 1–2 (MDR) (green and yellow), and phase 3 (drug resistance) (red).

Data information: In (D, E), data are presented as mean values + SEM (n = 6,574 cells from 5 PDX).

2 but increased in the drug-resistance phase 3. The IFN-active cells were present uniquely in phase 3. MITF-intermediate proliferative cells were prevalent at the drug-naïve phase T0, but declined strongly in phase 1 before becoming more numerous in phases 2 and 3.

Analyses of GATA6 expression in the seven cell types during the defined phases indicated that GATA6 significantly emerged in cells displaying "Invasive" phenotype at T0, persisted at low levels in "SMC" cells during the MRD phase before re-emerging in cells with "IFN-active" phenotype during the drug-resistant phase 3 (Fig 8D). In contrast, MITF was expressed in cells with "SMC" or "Mitotic" phenotypes in T0, but showed essentially no expression in "Invasive" or "NCSC" cells and was strongest expressed in "Pigmented" cells with lower expression in the "IFN-active" cells at the drug-resistance phases (Fig 8E). The expression of GATA6 in "IFN-active" melanoma cells in the PDX prompted us to treat MM074 cells with the pro-inflammatory cytokine IFNγ. This treatment decreased expression of MITF and up-regulated expression of GATA6, c-JUN (Riesenberg *et al*, 2015), and the positive control PD-L1 at both mRNA (Fig EV4C) and protein levels (Fig EV4D). These data indicated an anti-correlation between MITF and GATA6 in cells from PDX tumors together with the emergence of GATA6 in MITF-low cells of the drug-resistance phase. This anti-correlation was recapitulated in cells treated with the pro-inflammatory cytokine IFNγ.

Discussion

In this work, we have shown that CDK7i sensitivity of melanoma cells was independent of driver mutation status, but strongly influenced by their phenotype. MITF-high melanocytic-type melanoma cells were highly sensitive to CDK7i, while MITF-low mesenchymal-like melanoma cells were largely insensitive. As shown before (Ennen *et al*, 2017; Wouters *et al*, 2020), mesenchymal-like MM099, MM029, and MM047 showed similar, but not identical signatures. Our current data show that MM047 differs from the MM099 and MM029 cells in its resistance to CDK7i. As each are primary cultures from different patients with a different natural history of the disease, differences between lines are only to be expected. Through the establishment of CDK7i-resistant cells from two different parental phenotypes, we defined a set of 261 genes reflecting the adaptation of melanoma cells to the exposure to CDK7i. Among these, we identified a network governed by GATA6 and containing genes such as *AMIGO2*, involved in melanoma cell survival. GATA6 and its regulon were not only expressed in melanoma cells chronically exposed to CDK7i *in vitro*, but also more broadly in melanoma cells showing low expression of the lineage-specific markers SOX10 and MITF in tumors. We further observed that depletion of SOX10 or MITF proteins also activated GATA6-dependent genes, suggesting that the decommission of the CDK7-dependent SEs regulating MITF and SOX10 expression following CDK7i exposure (Eliades *et al*, 2018) is a key step in their activation. Our results also established that CDK7i more strongly inhibited *MITF* and *SOX10* expression in melanoma cells compared to BETi (Fontanals-Cirera *et al*, 2017), despite the presence of BRD4 at their corresponding SEs (Eliades *et al*, 2018), that may just be a collateral non-functional recruitment associated with strong enrichment of coactivators at SEs. Further evidence for the critical role of MITF/SOX10 in GATA6 repression comes from their ectopic expression in mesenchymal-like cells that inhibited *GATA6* expression. We further identified a short regulatory sequence in a GATA6 intron that is bound by MITF and conferred MITF-driven transcriptional repression in a heterologous setting, a recognized criterion for *bone fide* repressor elements.

Consistent with our analysis showing GATA6 expression in both "invasive" and "IFN-active" PDX melanoma cells, IFNγ treatment of melanocytic cells repressed MITF/SOX10 leading to the concomitant activation of *GATA6* (Son *et al*, 2014). Since MITF has been shown to participate in stabilization of CDK7 in melanocyte-type melanoma cells (Seoane *et al*, 2019; Louphrasitthiphol *et al*, 2020), our data suggest a negative feedback loop where the progressive loss of MITF during melanoma progression and inflammation triggers decreased CDK7 protein levels that in turn promotes lower MITF expression leading to de-repression of GATA6 expression in MITF/CDK7-low melanoma cells (Fig EV5). In line with this model, a negative correlation between CDK7 and GATA6 was observed in human SKCM.

GATA6 is expressed in various normal tissues derived from the mesoderm and endoderm (Almalki & Agrawal, 2016). An oncogenic role for GATA6 has been proposed in various cancers including pancreatic cancer where its knockdown reduced cell proliferation and cell cycle progression (Sun & Yan, 2020). We show that the decrease of GATA6 impaired proliferation of MITF-low mesenchymal-like melanoma cells. Since *GATA6* is expressed in normal adult tissues, it is unlikely that its targeting would lead to efficient therapy. However, identification of its downstream regulon genes may help to identify molecular targets in mesenchymal-like melanoma cells that could be exploited therapeutically to prevent

acquisition of metastatic and drug resistance potential. One of the GATA6 regulon genes, *AMIGO2,* has already been identified as targetable for metastatic melanoma (Fontanals-Cirera *et al*, 2017). We observed that the GATA6-dependent multidrug transporter *ABCG2* is, at least in part, responsible for cross-resistance to targeted therapies in mesenchymal-like cells and is significantly overexpressed in metastatic melanoma tumors compared with primary tumors, suggesting that it may mediate ubiquitous cross-resistance to targeted therapies clinically.

Our results also clearly established a role for CDK7 in transcriptional reprograming of melanoma cells. MITF-high melanoma cells exposed to CDK7i progressively lost melanocytic-type markers and acquired those of the undifferentiated mesenchymal-like state. In acquired CDK7i-resistant melanoma cells, we detected both a mesenchymal-like transcriptional signature and the acquisition of programs responsible for invasion. In apparent contrast, we observed that the acquired resistance of MITF-high melanoma cells to BRAFi was not accompanied by a loss of lineage-specific markers. In our hands, and as previously observed (Haq *et al*, 2013) (Smith *et al*, 2016), chronic exposure of melanocytic-type melanoma cells to escalating doses of BRAFi switched them to a highly pigmented state, which is likely a consequence of the increased MITF expression that we observed in these cells (Khaled *et al*, 2010).

Finally, an increasing number of studies identified CDK7 as a therapeutic target in various cancers (Fisher, 2019). However, the phenotype reprograming observed during prolonged exposure of melanoma cells to CDK7i illustrates the potential danger of targeting this kinase in cancers where EMT plays an important role in therapeutic resistance and metastasis, an issue that has not been fully investigated so far. Future studies should therefore take into consideration the potential of CDK7i treatment to promote emergence of mesenchymal-like cells and therapeutic resistance.

Materials and Methods

A full list of reagents including antibodies, commercial kits, and oligonucleotides is supplied in Appendix Table S1.

Patients

Gene expressions in tumors and nevi were retrieved from several previously published datasets (including TCGA) indicated in the figure legends.

Cell culture and treatment

Cells were grown in 5% CO_2 at 37°C in HAM-F10 (Gibco, Invitrogen) supplemented with 10% FCS and penicillin–streptomycin. Melanoma cell line 501mel was grown in 5% CO_2 at 37°C in RPMI w/o HEPES (Gibco, Invitrogen) supplemented with 10% FCS and gentamycin.

Cells were transfected with Lipofectamine RNAiMAX following the manufacturer's instructions with 25 nM of siRNA ON-TARGETplus SMARTPool (Horizon Discovery), and cells were harvested 48 and/or 72 h after transfection. All cell lines used were mycoplasm negative.

MM099$^{MITF\text{-}SOX10\text{-}PAX3}$, MM047^{GFP}, and MM047^{GATA6} cells were generated as followed. Lentiviral vectors pTET-SMP encoding human untagged MITF, SOX10, and PAX3 proteins, and pLenti-EF1a-GFP and pLenti-EF1-3xFLAG-GATA6 encoding for GFP and GATA6 proteins, respectively, were transduced in either MM099 or MM047 in the presence of polybrene, and cells were selected with 3 μg/ml of puromycin. Conditional expression of pTET vector was carried out by adding 1 μg/ml of doxycycline in the medium for at least 24 h.

Generation of CDK7i- and BRAFi-resistant cells

To generate CDK7i- or BRAFi-resistant cells, we chronically exposed the MM074 (BRAF^{V600E}) melanocytic-type cells to escalating doses of THZ1 or vemurafenib over several weeks. These treatments were carried out until the cells proliferated in drug concentrations equal to at least 5 times the original IC50 values, allowing us to generate stable MM074$^{CDK7i\text{-}R}$ and MM074$^{BRAFi\text{-}R}$ lines, respectively. In parallel, the MM047 (NRAS^{Q61R}) mesenchymal-like cells were chronically exposed to THZ1 following the same protocol to generate stable MM047$^{CDK7i\text{-}R}$. Once established, the resistance was permanent and drugs can be removed without affecting cell phenotype.

CRISPR/Cas9 editing of 501mel$^{BIO\text{-}FLAG:CDK7}$

A 501mel were co-transfected with vector px738 (encoding Cas9-HF-GFP and two guide RNAs targeting CDK7 locus) and construct m599 linear DNA fragment carrying homology regions to CDK7 locus and puromycin-P2A-BIO-FLAG-CDK7N-termsequence (Appendix Table S1) with transfection reagent FuGENE6. Twenty-four h later, single cells GFP positive were sorted in P96-well plates in the presence of puromycin (3 μg/ml) with cell sorter. Single clones were let grown and selected for 4–6 weeks and surviving ones screened for gene editing through PCR using Phusion High-Fidelity TAQ Polymerase using different combination of primers (F1, F5, R3, R4, R5, see Appendix Table S1). PCR-positive clones were finally further amplified to perform Western blot and Co-IPs validation.

Cell proliferation assay

To measure proliferation, cells were incubated first with CellTrace Violet according to the manufacturer's instructions. Cell proliferation was detected on a BD LSRFortessa™ Flow Cytometer. Data were analyzed with FlowJo software. To define slow proliferating cells, we proceeded as follows: We considered that slow proliferating cells represented the 30% of cells with the highest concentration of BV421 in the siCTL treatment. We then calculated the % of cells that had a concentration greater than or equal to this value after treatment with siRNA.

Reporter assay

The intGATA6r element was isolated by genomic PCR using Phusion High-Fidelity TAQ polymerase (Thermo Fisher) with specific primers (Appendix Table S1). To allow the cloning within pCDNA-GFP vector, the first PCR product was further amplified by PCR with primers carrying MluI and SmaI restriction sites at the 5' and 3', respectively (MluI_F and SmaI_R primers in Appendix Table S1).

The immediate early CMV enhancer (ieEnh) in the pCDNA-GFP vector (pCDNA-ieEnh-CMV-GFP) was then replaced with the intGATA6r element (pCDNA-intGATA6r-CMV-GFP).

MITF-SOX10-PAX3 expression was induced in MM099$^{MITF\text{-}SOX10\text{-}PAX3}$ cells with doxycycline for 48 h, and cells were subsequently transfected with pCDNA-ieEnh-CMV-GFP or pCDNA-intGATA6r-CMV-GFP vectors for 24 h with FuGENE6 following the manufacturer's instructions. RNA was then collected for qPCR, and GFP protein signal was detected on cytofluorometer. FACS data were analyzed with FlowJo software.

Histology

Human tissue sections were de-paraffinized and dehydrated with Histosol and dilutions of ethanol (100, 90, 70, and 30%) and then rehydrated with demineralized water. Subsequently, sections were boiled in sodium citrate buffer (0.1 M citric acid, 0.1 M sodium citrate) for 15 min to unmask antigens. Alternatively, cells were grown on glass slides and fixed with 4% formaldehyde. Both tissues and cells were permeabilized with PBS and 0.1% Triton X-100. Blocking was done with 10% fetal bovine serum before incubation with primary antibodies.

In situ hybridization of *ABCG2* mRNA was performed using the RNAscope assay according to the manufacturer's instructions (ACDBio). Cells and tissue sections were counterstained with DAPI and visualized using confocal microscope Spinning disk Leica CSU W1. Probes' sequences were not provided by the manufacture.

EU incorporation assay

RNA labeling by EU incorporation was performed with Click-iT RNA Imaging kit following the manufacturer's protocol. EU signal intensity was quantified using imaging system.

Cell survival assay

Normal or transfected cells were seeded at 5,000 cells/well in a 96-well plate and treated with increasing concentrations of THZ1, vemurafenib, or trametinib. After 72 h of incubation, cells were treated with PrestoBlue reagent according to the manufacturer's instructions. The absorbance per well was measured with a microplate reader. The data were then analyzed using Prism8.

RT–qPCR

Total RNA was isolated from cells using a GenElute Mammalian Total RNA Miniprep kit (Sigma) and reverse-transcribed with SuperScript IV reverse transcriptase (Invitrogen). The quantitative PCR was done using LightCycler. The primer sequences for the different genes used in qPCR are indicated in Appendix Table S1. The mRNA expression of the various analyzed genes represents the ratio between values obtained from treated and untreated cells normalized with the housekeeping genes mRNA.

ChIP

Cells were grown on 15-cm plates and, once reached 80% of confluence, were fixed with PBS + 0.4% formaldehyde solution for 10 min. Fixation reaction was stopped with 2 M Glycin pH 8. Cells were then pelleted and suspended in lysis buffer (EDTA 10 mM, Tris–HCl pH8 50 mM, SDS 1%) and sonicated with Covaris E220 AFA power 200 Hz 6 cycles 200 s to get a DNA fragmentation between 500 and 200 bp. Chromatin was then diluted in 60 µg aliquots with 8 volumes of ChIP dilution buffer (Tris–HCl pH8 16.7 mM, EDTA 1.2 mM, NaCl 167 mM, Triton X-100 1.1%, SDS 0.01%). The immuno-precipitations were done as follows. 1–5 µg of antibody was incubated overnight with chromatin, and the complex antibody–chromatin was then captured with G protein sepharose beads (Sigma-Aldrich) 1 h at 4°C. Beads were washed 2 times with low salt buffer (Tris–HCl pH8 20 mM, EDTA 2 mM, NaCl 150 mM, Triton X-100 1%, SDS 0.1%), high salt buffer (Tris–HCl pH8 20 mM, EDTA 2 mM, NaCl 500 mM, Triton X-100 1%, SDS 0.1%), LiCl buffer (Tris–HCl pH8 500 mM, EDTA 1 mM, Na deoxycholate 1%, NP40 1%, LiCl 0.25 M), and TE buffer (Tris–HCl pH8 10 mM, EDTA 1 mM), and DNA was eluted 30 min at room temperature with Elution buffer ($NaHCO_3$ 0.1 M, SDS 1%). DNA was finally purified through phenol–chloroform, re-suspended in 100 µl of water, and analyzed by qPCR using a set of primers indicated in Appendix Table S1.

RNA-seq

RNA-seq was performed as previously described (Laurette *et al*, 2019). Reads were preprocessed in order to remove adapter and low-quality sequences (Phred quality score below 20). After this preprocessing, reads shorter than 40 bases were discarded for further analysis. These preprocessing steps were performed using cutadapt version 1.10. Reads were mapped to rRNA sequences using bowtie version 2.2.8, and reads mapping to rRNA sequences were removed for further analysis. Reads were mapped onto the hg19 assembly of Homo sapiens genome using STAR version 2.5.3a. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.6.1p1, with annotations from Ensembl version 75 and "union" mode. Only non-ambiguously assigned reads have been retained for further analyses. Read counts have been normalized across samples with the median-of-ratios method (Anders & Huber, 2010). Differential gene expression analysis was performed using the methodology implemented in the Bioconductor package DESeq2 version 1.16.1 (Love *et al*, 2014). *P*-values were adjusted for multiple testing by the method proposed by Benjamini and Hochberg (Benjamini & Hochberg, 1995). Deregulated genes were defined as genes with log2(foldchange) > 1 or < -1 and adjusted *P*-value < 0.05. Heatmaps were generated with the R package pheatmap v1.0.12.

ChIP-seq

Purified DNA fragments for ChIP-seq were prepared by using the ChIP-IT High Sensitivity Kit (Active Motif) and the related antibodies. ChIP-seq was performed on an Illumina sequencer as single-end 50 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA 1.17.20 and CASAVA 1.8.2. Reads were mapped onto the hg19 assembly of the human genome. Peak detection was performed using MACS (https://github.com/macs3-project/MACS) under settings where the input fraction was used as negative control. Peaks detected were annotated using

HOMER (http://biowhat.ucsd.edu/homer/ngs/annotation.html) as well as TSS protein enrichment comparison. Quantitative comparison of RNA Pol II gene body enrichment was performed using seqMINER (http://bips.u-strasbg.fr/seqminer/). As reference coordinates, we used the MACS-determined peaks or the annotated TSS/TTS of human genes as defined by RefSeq database. Sequence enrichment was performed using RSAT (http://rsat.sb-roscoff.fr) with MACS-determined peaks as reference.

Analysis of scRNA-seq data from short-cultured melanoma cells

After being downloaded, raw reads from scRNA-seq from MM011, MM029, MM047, MM074, and MM099 (Wouters *et al*, 2020) were processed using CellRanger (v 3.1) to align on the hg19 human genome, remove unexpressed genes, and quantify barcodes and UMIs. Data were then analyzed in R (v4.0.2) with Seurat v3.2.0 following the recommended workflow. Cells were filtered for feature count ranging from 120 to 2,000 and percentage of mitochondrial reads < 15%. Counts were normalized with the "LogNormalize" method and data scaled to remove unwanted sources of variation (UMI count and mitochondrial reads). The number of principal components to use was determined from the Jackstraw plots. Clustering was performed on variable features using the 25 most significant principal components and a resolution of 1.15. Regulome analyses of active transcription factors were performed using the SCENIC v1.1.2.2 package. Transcription factor activities were visualized on the UMAP using AUCell or as heatmaps using the R package pheatmap. Trajectory on the UMAP projection was resolved by monocle3 v0.2.0.

Analysis of scRNA-seq from PDX

Expression matrix with row reads counts for the single-cell experiment was retrieved from GEO (GSE116237). Then, data were normalized and clustered using the Seurat software package version 3.1.4 (Butler *et al*, 2018) in R version 3.6.1. Data were filtered, and only genes detected in at least 3 cells and cells with at least 350 detected genes were kept for further analysis. Expression of 26,661 transcripts in 674 cells was quantified. To cluster cells, read counts were normalized using the method "LogNormalize" of the Seurat function NormalizeData. It divides gene expression counts by the total expression, multiplies this by a scale factor (10,000 was used), and log-transforms the result. Then, 2,000 variable features were selected with the variance stabilizing transformation method using the Seurat function FindVariableGenes with default parameters. Integrated expression matrices were scaled (linear transformation) followed by principal component analysis (PCA) for linear dimensional reduction. The first 20 principal components (PCs) were used to cluster the cells with a resolution of 0.5 and as input to tSNE to visualize the dataset in two dimensions. The Bioconductor package AUCell v 1.6.1 (Aibar *et al*, 2017) was used to assess whether some cells from the Rambow dataset were enriched in gene sets of interest. AltAnalyze was used for the supervised clustering of TCGA samples (Olsson *et al*, 2016).

Gene ontology

Gene ontology was performed using Metascape software developed by (Zhou *et al*, 2019).

Quantification and statistical analysis

Statistical details of experimental can be found in figure legends or in the methods details. Hypergeometric distribution tests for the Venn diagrams were performed using: https://systems.crump.ucla.edu/hypergeometric/.

Data availability

The datasets produced in this study are available in the following databases:

Access numbers for data generated in this paper are as follows:
ChIP-Seq data CDK7: Gene expression Omnibus GSE158118.
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158118
RNA-seq data CDK7i cells: Gene expression Omnibus GSE 158119.
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158119
RNA-seq data: Gene expression Omnibus GSE164431.
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164431.

Expanded View for this article is available online.

Acknowledgments
We thank the IGBMC antibody and cell culture facilities, Prof D. Lipsker, and the staff of the Strasbourg University Hospital dermatology clinic for tumor sections and Dr Ghanem Ghanem and J-C Marine for providing us the MM-series melanoma cultures. This study was supported by the Institut National Du Cancer (INCa) (2015-9378 and 2017-11537), the Ligue contre le Cancer (Equipe labélisée 2019, FC and équipe labélisée 2018, ID) and the ANR (START-2021), the ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02. Sequencing was performed by the IGBMC GenomEast platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009). PB is supported by the Ligue contre le Cancer. M.C is supported by the Fondation pour la Recherche Médicale.

Author contributions
PB, ID, and FC conceived the study. PB, ID, FC, and EC analyzed the data. PB generated the resistant cells and performed the RNA-seq and the IH staining. SL, GD, BV, and PB, performed the bioinformatics analyses. PB, MC, and CMGR performed the survival assays. EC, FP, PB, and MC performed the RT–qPCR and the WB. JS performed the EU staining. NS performed the IFNγ experiments. CB provided a valuable technical assistance. JME provided valuable materials. FC and ID wrote the manuscript with input from EC.

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

References

Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J *et al* (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 14: 1083 – 1086

Alekseev S, Nagy Z, Sandoz J, Weiss A, Egly JM, Le May N, Coin F (2017) Transcription without XPB establishes a unified helicase-independent

mechanism of promoter opening in eukaryotic gene expression. *Mol Cell* 65: 504 – 514

Almalki SG, Agrawal DK (2016) Key transcription factors in the differentiation of mesenchymal stem cells. *Differentiation* 92: 41 – 51

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106

Arozarena I, Wellbrock C (2019) Phenotype plasticity as enabler of melanoma progression and therapy resistance. *Nat Rev Cancer* 19: 377 – 391

Badal B, Solovyov A, Di Cecilia S, Chan JM, Chang LW, Iqbal R, Aydin IT, Rajan GS, Chen C, Abbate F *et al* (2017) Transcriptional dissection of melanoma identifies a high-risk subtype underlying TP53 family genes and epigenome deregulation. *JCI Insight* 2: e92102

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57: 289 – 300

Berico P, Coin F (2018) Is TFIIH the new Achilles heel of cancer cells? *Transcription* 9: 47 – 51

Brose MS, Volpe P, Feldman M, Kumar M, Rishi I, Gerrero R, Einhorn E, Herlyn M, Minna J, Nicholson A *et al* (2002) BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Res* 62: 6997 – 7000

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36: 411 – 420

Cao K, Shilatifard A (2014) Inhibit globally, act locally: CDK7 inhibitors in cancer therapy. *Cancer Cell* 26: 158 – 159

Carreira S, Goodall J, Denat L, Rodriguez M, Nuciforo P, Hoek KS, Testori A, Larue L, Goding CR (2006) Mitf regulation of Dia1 controls melanoma proliferation and invasiveness. *Genes Dev* 20: 3426 – 3439

Chipumuro E, Marco E, Christensen C, Kwiatkowski N, Zhang T, Hatheway C, Abraham B, Sharma B, Yeung C, Altabef A *et al* (2014) CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* 159: 1126 – 1139

Christensen C, Kwiatkowski N, Abraham B, Carretero J, Al-Shahrour F, Zhang T, Chipumuro E, Herter-Sprie G, Akbay E, Altabef A *et al* (2014) Targeting transcriptional addictions in small cell lung cancer with a covalent CDK7 inhibitor. *Cancer Cell* 26: 909 – 922

Compe E, Egly JM (2016) Nucleotide excision repair and transcriptional regulation: TFIIH and beyond. *Annu Rev Biochem* 85: 265 – 290

Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W *et al* (2002) Mutations of the BRAF gene in human cancer. *Nature* 417: 949 – 954

Eggermont AM, Spatz A, Robert C (2014) Cutaneous melanoma. *Lancet* 383: 816 – 827

Eick D, Geyer M (2013) The RNA Polymerase II Carboxy-Terminal Domain (CTD) code. *Chem Rev* 113: 8456 – 8490

Eliades P, Abraham BJ, Ji Z, Miller DM, Christensen CL, Kwiatkowski N, Kumar R, Njauw CN, Taylor M, Miao B *et al* (2018) High MITF expression is associated with super-enhancers and suppressed by CDK7 inhibition in melanoma. *J Invest Dermatol* 138: 1582 – 1590

Ennen M, Keime C, Gambi G, Kieny A, Coassolo S, Thibault-Carpentier C, Margerin-Schaller F, Davidson G, Vagne C, Lipsker D *et al* (2017) MITF-high and MITF-low cells and a novel subpopulation expressing genes of both cell states contribute to intra- and intertumoral heterogeneity of primary melanoma. *Clin Cancer Res* 23: 7097 – 7107

Fisher RP (2019) Cdk7: a kinase at the core of transcription and in the crosshairs of cancer drug discovery. *Transcription* 10: 47 – 56

Fontanals-Cirera B, Hasson D, Vardabasso C, Di Micco R, Agrawal P, Chowdhury A, Gantz M, de Pablos-Aragoneses A, Morgenstern A, Wu P *et al* (2017) Harnessing BET inhibitor sensitivity reveals AMIGO2 as a melanoma survival gene. *Mol Cell* 68: 731 – 744.e9

Haq R, Shoag J, Andreu-Perez P, Yokoyama S, Edelman H, Rowe GC, Frederick DT, Hurley AD, Nellore A, Kung AL *et al* (2013) Oncogenic BRAF regulates oxidative metabolism via PGC1alpha and MITF. *Cancer Cell* 23: 302 – 315

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA (2013) Super-enhancers in the control of cell identity and disease. *Cell* 155: 934 – 947

Hodis E, Watson I, Kryukov G, Arold S, Imielinski M, Theurillat J-P, Nickerson E, Auclair D, Li L, Place C *et al* (2012) A landscape of driver mutations in melanoma. *Cell* 150: 251 – 263

Hoek KS, Eichhoff OM, Schlegel NC, Dobbeling U, Kobert N, Schaerer L, Hemmi S, Dummer R (2008) *In vivo* switching of human melanoma cells between proliferative and invasive states. *Cancer Res* 68: 650 – 656

Kemper K, de Goeje PL, Peeper DS, van Amerongen R (2014) Phenotype switching: tumor cell plasticity as a resistance mechanism and target for therapy. *Cancer Res* 74: 5937 – 5941

Khaled M, Levy C, Fisher DE (2010) Control of melanocyte differentiation by a MITF-PDE4D3 homeostatic circuit. *Genes Dev* 24: 2276 – 2281

Klein RM, Bernstein D, Higgins SP, Higgins CE, Higgins PJ (2012) SERPINE1 expression discriminates site-specific metastasis in human melanoma. *Exp Dermatol* 21: 551 – 554

Kwiatkowski N, Zhang T, Rahl PB, Abraham BJ, Reddy J, Ficarro SB, Dastur A, Amzallag A, Ramaswamy S, Tesar B *et al* (2014) Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* 511: 616 – 620

Laurette P, Coassolo S, Davidson G, Michel I, Gambi G, Yao W, Sohier P, Li M, Mengus G, Larue L *et al* (2019) Chromatin remodellers Brg1 and Bptf are required for normal gene expression and progression of oncogenic Braf-driven mouse melanoma. *Cell Death Differ* 27: 29 – 43

Laurette P, Strub T, Koludrovic D, Keime C, Le Gras S, Seberg H, Van Otterloo E, Imrichova H, Siddaway R, Aerts S *et al* (2015) Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. *Elife* 4: e06857

Louphrasitthiphol P, Siddaway R, Loffreda A, Pogenberg V, Friedrichsen H, Schepsky A, Zeng Z, Lu M, Strub T, Freter R *et al* (2020) Tuning transcription factor availability through acetylation-mediated genomic redistribution. *Mol Cell* 79: 472 – 487.e10

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

Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153: 320 – 334

Menzies AM, Long GV (2014) Systemic treatment for BRAF-mutant melanoma: where do we go next? *Lancet Oncol* 15: e371 – e381

Minnoye L, Taskiran II, Mauduit D, Fazio M, Van Aerschot L, Hulselmans G, Christiaens V, Makhzami S, Seltenhammer M, Karras P *et al* (2020) Cross-species analysis of enhancer logic using deep learning. *Genome Res* 30: 1815 – 1834

Olsson A, Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, Singh H, Grimes HL (2016) Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* 537: 698 – 702

Rambow F, Marine JC, Goding CR (2019) Melanoma plasticity and phenotypic diversity: therapeutic barriers and opportunities. *Genes Dev* 33: 1295 – 1318

Rambow F, Rogiers A, Marin-Bejar O, Aibar S, Femel J, Dewaele M, Karras P, Brown D, Chang YH, Debiec-Rychter M *et al* (2018) Toward minimal residual disease-directed therapy in melanoma. *Cell* 174: 843 – 855.e19

Riesenberg S, Groetchen A, Siddaway R, Bald T, Reinhardt J, Smorra D, Kohlmeyer J, Renn M, Phung B, Aymans P *et al* (2015) MITF and c-Jun antagonism interconnects melanoma dedifferentiation with pro-inflammatory cytokine responsiveness and myeloid cell recruitment. *Nat Commun* 6: 8755

Robey RW, Pluchino KM, Hall MD, Fojo AT, Bates SE, Gottesman MM (2018) Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nat Rev Cancer* 18: 452 – 464

Seoane M, Buhs S, Iglesias P, Strauss J, Puller A-C, Müller J, Gerull H, Feldhaus S, Alawi M, Brandner JM *et al* (2019) Lineage-specific control of TFIIH by MITF determines transcriptional homeostasis and DNA repair. *Oncogene* 38: 3616 – 3635

Smith MP, Brunton H, Rowling EJ, Ferguson J, Arozarena I, Miskolczi Z, Lee JL, Girotti MR, Marais R, Levesque MP *et al* (2016) Inhibiting drivers of non-mutational drug tolerance is a salvage strategy for targeted melanoma therapy. *Cancer Cell* 29: 270 – 284

Son J, Kim M, Jou I, Park KC, Kang HY (2014) IFN- γ inhibits basal and α -MSH-induced melanogenesis. *Pigment Cell Melanoma Res* 27: 201 – 208

Sun Z, Yan B (2020) Multiple roles and regulatory mechanisms of the transcription factor GATA6 in human cancers. *Clin Genet* 97: 64 – 72

Verfaillie A, Imrichova H, Atak ZK, Dewaele M, Rambow F, Hulselmans G, Christiaens V, Svetlichnyy D, Luciani F, Van den Mooter L *et al* (2015) Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat Commun* 6: 6683

Villicana C, Cruz G, Zurita M (2014) The basal transcription machinery as a target for cancer therapy. *Cancer Cell Int* 14: 18

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153: 307 – 319

Widmer DS, Cheng PF, Eichhoff OM, Belloni BC, Zipser MC, Schlegel NC, Javelaud D, Mauviel A, Dummer R, Hoek KS (2012) Systematic classification of melanoma cells by phenotype-specific gene expression mapping. *Pigment Cell Melanoma Res* 25: 343 – 353

Wouters J, Kalender-Atak Z, Minnoye L, Spanier KI, De Waegeneer M, Bravo González-Blas C, Mauduit D, Davie K, Hulselmans G, Najem A *et al* (2020) Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma. *Nat Cell Biol* 22: 986 – 998

Xu L, Shen SS, Hoshida Y, Subramanian A, Ross K, Brunet JP, Wagner SN, Ramaswamy S, Mesirov JP, Hynes RO (2008) Gene expression changes in an animal melanoma model correlate with aggressiveness of human melanoma metastases. *Mol Cancer Res* 6: 760 – 769

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10: 1523

The Oncologist, 2022, **27**, 501–511
https://doi.org/10.1093/oncolo/oyac040
Advance access publication 12 March 2022
Original Article

OXFORD

Immune-Desert Tumor Microenvironment in Thoracic SMARCA4-Deficient Undifferentiated Tumors with Limited Efficacy of Immune Checkpoint Inhibitors

Justine Gantzer[1,2,3,*], Guillaume Davidson[3], Bujamin Vokshi[3], Noëlle Weingertner[2,4], Antoine Bougoüin[5], Marco Moreira[5], Véronique Lindner[2,4], Guillaume Lacroix[5], Céline Mascaux[6,7], Marie-Pierre Chenard[2,4], François Bertucci[8], Irwin Davidson[3], Jean-Emmanuel Kurtz[1,2], Catherine Sautès-Fridman[5], Wolf H. Fridman[5,‡], Gabriel G. Malouf[1,2,3,*,‡]

[1]Department of Medical Oncology, Strasbourg-Europe Cancer Institute (ICANS), Strasbourg, France
[2]Fédération de Médecine Translationnelle (FMTS), Strasbourg, France
[3]Department of Cancer and Functional Genomics, INSERM UMR_S1258, Institute of Genetics and of Molecular and Cellular Biology, Illkirch, France
[4]Department of Pathology, University Hospital, Strasbourg, France
[5]Centre de recherche des Cordeliers, INSERM, Université de Paris, Sorbonne Université, Team 13- Complement, Inflammation and Cancer, Équipe labellisée Ligue contre le cancer, Paris, France
[6]Department of Pneumology, University Hospital, Strasbourg, France
[7]University of Strasbourg, Inserm UMR_S 1113, IRFAC, Laboratory Streinth (STress REsponse and INnovative THerapy against cancer), Strasbourg, France
[8]Department of Medical Oncology, Cancer Research Center of Marseille (CRCM), INSERM U1068, CNRS UMR7258, Institut Paoli Calmettes, Aix-Marseille University, Marseille, France

*Corresponding author: Justine Gantzer, Department of Medical Oncology, Strasbourg-Europe Cancer Institute (ICANS), 17 rue Albert Calmette, 67033 Strasbourg, France. Tel: +33 3 68 76 72 25; Email: j.gantzer@icans.eu; or, Gabriel G. Malouf, Department of Medical Oncology, Strasbourg-Europe Cancer Institute (ICANS), 17 rue Albert Calmette, 67033 Strasbourg, France. Tel: +33 3 68 76 72 17; Email: g.malouf@icans.eu
‡Co-senior authors.

Abstract

Background: Thoracic SMARCA4-deficient undifferentiated tumors (SMARCA4-UT) are aggressive neoplasms. Data linking BAF alterations with tumor microenvironment (TME) and efficacy of immune checkpoint inhibitors (ICI) are contradictory. The TME of SMARCA4-UT and their response to ICI are unknown.

Materials and Methods: Patients diagnosed with SMARCA4-UT in our institution were included. Immunostainings for tertiary lymphoid structures (TLS), immune cell markers, and checkpoints were assessed. Validation was performed using an independent transcriptome dataset including SMARCA4-UT, non–small cell lung cancers (NSCLC) with/without *SMARCA4* mutations, and unclassified thoracic sarcomas (UTS). CXCL9 and PD-L1 expressions were assessed in NSCLC and thoracic fibroblast cell lines, with/without *SMARCA4* knockdown, treated with/without interferon gamma.

Results: Nine patients were identified. All samples but one showed no TLS, consistent with an immune desert TME phenotype. Four patients received ICI as part of their treatment, but the only one who responded, had a tumor with a TLS and immune-rich TME. Unsupervised clustering of the validation cohort using immune cell scores identified 2 clusters associated with cell ontogeny and immunity (cluster 1 enriched for NSCLC independently of *SMARCA4* status (n = 9/10; P = .001); cluster 2 enriched for SMARCA4-UT (n = 11/12; P = .005) and UTS (n = 5/5; P = .0005). *SMARCA4* loss-of-function experiments revealed interferon-induced upregulation of CXCL9 and PD-L1 expression in the NSCLC cell line with no effect on the thoracic fibroblast cell line.

Conclusion: SMARCA4-UT mainly have an immune desert TME with limited efficacy to ICI. TME of SMARCA4-driven tumors varies according to the cell of origin questioning the interplay between BAF alterations, cell ontogeny and immunity.

Key words: SMARCA4-deficiency; thoracic tumors; sarcomas; immunotherapy; immune infiltrate; tumor microenvironment.

Implications for Practice

Thoracic SMARCA4-deficient undifferentiated tumors harbor mostly an immune desert TME but, as in soft-tissue sarcomas, immune-rich tumors characterized by TLS do exist and respond to ICI. The TME of SMARCA4-driven tumors varies according to the cell of origin highlighting the need to explore the interplay between alterations of BAF complexes, cell ontogeny and immunity.

Received: 20 September 2021; Accepted: 14 January 2022.

Background

Thoracic SMARCA4-deficient undifferentiated tumors (SMARCA4-UT) are a rare type of neoplasm, characterized by inactivating *SMARCA4* mutations leading to protein loss.[1] SMARCA4-encoded protein, BRG1 is one of the ATPase subunits part of the BRG1/BRM-associated factors (BAF) chromatin remodeling complex, also known as the mammalian SWItch/sucrose Non-Fermenting (mSWI/SNF) ATP-dependent chromatin remodeling complex.[2] The BAF complexes, composed of multiple subunits, regulate transcription and recent studies revealed their critical roles as tumor suppressors.[3-5] Gene profiling analyses revealed that SMARCA4-UT are related to other BAF complex-deficient tumors (ie, *SMARCB1*-inactivated malignant rhabdoid tumors (MRT) and *SMARCA4*-mutated-small cell carcinomas of the ovary, hypercalcemic type (SCCOHT)), but differ from SMARCA4-deficient non–small cell lung cancers (NSCLC) and not otherwise specified (NOS) NSCLC.[1] Histologically SMARCA4-UT are poorly differentiated tumors with rhabdoid or epithelioid features and harbor a specific "immunohistochemical signature". Indeed, in the fifth edition of World Health Organization (WHO) classification of Thoracic Tumors, the community of pathologists recognized this newly described entity but changed its name from SMARCA4-deficient thoracic sarcomas to SMARCA4-UT.[6] Clinically, the majority of SMARCA4-UT shared an aggressive clinical course with a median overall survival (OS) between 4 and 7 months.[1,7-11] Mostly male adults around 45 years old with a heavy smoking history are affected. Primary tumor location is thoracic with a median tumor size of 10 cm.[7]

Currently, there is no approved treatment for SMARCA4-UT. The first-line setting is mainly anthracycline-based chemotherapy identical to what is done for soft-tissue sarcomas (STS), but with very limited efficacy. Based on encouraging preclinical data, immune checkpoint inhibitors (ICI), especially anti-programmed cell death protein 1 (PD1), have been tested in STS patients, but showed limited efficacy with observed objective response rates always below 20%, although exceptional responders were identified.[12-15] This limited efficacy has been explained by the analysis of tumor microenvironment (TME) composition, based on the study of the immune populations from the transcriptome of 213 STS.[16] Notably, only 2 out of the 5 immune subtypes described harbored an immune-high phenotype despite a low tumor mutational burden (TMB). Within these subtypes, the one with a high response rate to ICI and an improved survival was characterized by the presence of tertiary lymphoid structures (TLS) in tumors. TLS are ectopic lymphoid formations structured with a T-cell zone with mature dendritic cells, a germinal center with proliferating B cells and are the lymphoid organs closest to the tumor generating an adaptative immune response.[17-19] Other biomarkers of response to ICI, also described in STS, include higher densities of cytotoxic tumor-infiltrating T cells, activated T cells, and an increased percentage of tumor-associated macrophages expressing PD1 ligand 1 (PD-L1).[20]

Previous data from the literature regarding the link between alterations of BAF complexes and their correlation with immune infiltrate are contradictory.[21] Indeed, while some studies have reported that BAF deficiency itself may enhance tumor cell susceptibility to immune control, others suggested that it may lead to impaired interferon (IFN) signaling leading to a non-immunogenic phenotype.[22-25] Focusing on SMARCA4-deficient tumors, the first report of SMARCA4-deficiency TME was described in SCCOHT and unraveled an immune-active TME despite low TMB, with significant levels of T-cell infiltration and PD-L1 expression.[26] Interestingly, among 4 patients treated with ICI, 1 patient had a sustained partial response for 6 months, and the 3 remaining patients remained disease-free for 1.5 years or more.[27] Other promising results on the efficacy of ICI were published on different SMARCA4-deficient tumors,[28,29] while others suggested the opposite.[30] Recently, results of a phase II study assessing the efficacy of pembrolizumab in patients with rare sarcomas described 3 out of the 11 patients (27%) with SMARCA4-deficient MRT responding to ICI,[15] which was consistent with several previously published case reports responding to ICI combined or not to chemotherapy.[31-34] However, to our knowledge, the immune landscape of SMARCA4-UT remains unknown and there is no series describing patient's response to ICI.

Herein, we report the first comprehensive analysis on SMARCA4-UT immune infiltrate, based on a retrospective cohort of 9 cases from a single institution and validate our findings in an independent cohort.[1] We also describe the response of 4 patients to ICI and performed a temporal comprehensive genomic as well as immune tumor profiling of an exceptional responder with complete and lasting response to ICI. Finally, CXCL9 chemokine and PD-L1 expressions were assessed in NSCLC and thoracic fibroblast cell lines, treated with/without interferon gamma (IFNG).

Materials and Methods

Patients and Samples

Patients of Strasbourg University hospital were identified prospectively during the period of 2016 to 2019. The main clinicopathological data and outcomes were recorded. Sample collection for further research analysis was approved by an Ethical Committee ("Comité de Protection des Personnes Est IV", Strasbourg, France) and the study was performed according to the Declaration of Helsinki.

Immunohistochemistry and Immunofluorescence

Immunostainings carried out on diagnostic purposes were performed by the local expert sarcoma pathologist, according to routine practice (Supplementary Table S1). Immunostainings for immune cell markers (CD3, CD8, CD20, and CD68) and immune checkpoints (PD1, PD-L1, and TIM3) were performed and evaluated according to the methodology detailed in the Supplementary Appendix and in the list of antibodies (Supplementary Table S2).

Tumors were considered TLS positive when a CD3 aggregate was found juxtaposing a CD20 aggregate, as previously described.[19] TLS with surface above 60 000 μm^2, containing at least 700 cells were considered as mature when a network of follicular dendritic cells was detected in B cell follicle on haematoxylin and eosin slides.[35,36] In lymph node metastases, TLS were only taken into account if the sample was invaded by more than 85% of tumor cells and without any residual lymph node tissue. PD-L1 was quantified using a score from 0 to 2 for both immune and tumor cells; 0 when no staining or less than 5% of positive cells, 1 between 5% and 50%, and 2 over 50% of positive cells.

For 7 patients, studied samples were either from the primary tumor (n = 2) or from distant metastasis (n = 5); for 2

patients, samples were available in both the primary tumor and matched distant metastasis.

Comprehensive Genomic Profiling

A FoundationOne Heme test was performed for 2 distinct samples in patient no. 3. This available commercial genomic profiling test from Foundation Medicine sequences DNA and RNA to detect cancer-related genomic alterations in more than 400 genes and around 250 gene fusions. Microsatellite status and TMB were also assessed.

Gene Expression Analysis

Publicly available fastq files of gene expression profiling from SMARCA4-UT (n = 12), SMARCA4-deficient NSCLC (n = 4), NOS NSCLC (n = 10), and unclassified thoracic sarcoma (UTS; n = 5) were downloaded from Sequence Read Archive accession SRP052896.[1] Briefly, raw reads were aligned using STAR v2.5.3a with the "--quantMode TranscriptomeSAM" argument and by providing the GFF file from ENSEMBL v75, gene expression level was then calculated using RSEM v1.3.3.[37,38] Gene expression data were processed using R v3.6.3. The TME composition of each sample was inferred using the microenvironment cell populations (MCP)-counter v1.2.0,[39] providing abundance score for 8 immune populations and 2 stromal populations (fibroblast and endothelial cells) based on analysis of specific transcriptomic markers expressed by each cell population.[16] We removed the fibroblast signature since sarcomas highly expressed this signature because of their mesenchymal origin, in order to better see differences in the other signatures. Samples were clustered based on their MCP-counter analysis using an unsupervised hierarchical clustering on the metagene Z-score for the 8 immune populations, with complete linkage using the hclust function and visualized as a heatmap with the pheatmap package v1.0.12.

Gene signatures for the functional orientation were computed as the geometric mean expression of the following genes: immunosuppression (*CXCL12*, *TGFB1*, *TGFB3*, and *LGALS1*), T-cell activation (*CXCL9*, *CXCL10*, *CXCL16*, *IFNG*, and *IL15*), T-cell survival (*CD70* and *CD27*), regulatory T cells (*FOXP3* and *TNFRSF18*), major histocompatibility complex class I (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-E*, *HLA-F*, *HLA-G*, and *B2M*), myeloid cell chemotaxis (*CCL2*), tertiary lymphoid structures (*CXCL13*). Expression of genes related to immune checkpoint were also computed (*PD1*, *PD-L1*, *PD-L2*, *CTLA4*, *TIM3*, and *LAG3*).

In vitro Analyses

Two commercially available cell lines were purchased from ATCC: the CCD_19Lu derived from human thoracic fibroblasts and the Calu-1 derived from NSCLC pleural metastasis. CCD_19Lu cells were grown in EAGLE medium supplemented with 10% FCS and gentamicine, and Calu-1 in MEM alpha medium with ribonucleosides and deoxyribonucleosides supplemented with 10% FCS and gentamicine. In the setting of this study, each cell line was transfected transiently with siRNAs either targeting SMARCA4 or non-targeting control for 6 hours, followed by either addition of DMSO as a control or human recombinant IFNG (20 ng/mL, Peprotech 500-P32). Cells were collected after 24 hours for RNA and protein extractions.

RNA isolation was performed according to standard procedure (NucleoSpin RNA Plus Kit). RT-qPCR was carried out with SYBR Green I (Roche) and SuperScript IV Reverse Transcriptase (Invitrogen) and monitored using a LightCycler 480 (Roche). The mean of *ACTB* gene expression levels was used to normalize the results. Primer sequences for each cDNA were designed using Primer3 Software and are available upon request.

Whole cell extracts were prepared by the standard freeze-thaw technique using LSDB 500 buffer (500 mM KCl, 25 mM Tris at pH 7.9, 10% glycerol (v/v), 0.05% NP-40 (v/v), 16 mM DTT, and protease inhibitor cocktail). Cell lysates were subjected to SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and proteins were transferred onto a nitrocellulose membrane. Membranes were incubated with primary antibodies in 5% dry fat milk and 0.01% Tween-20 overnight at 4 °C. The membrane was then incubated with HRP-conjugated secondary antibody (Jackson ImmunoResearch) for 1 hour at room temperature, and visualized using the ECL detection system (GE Healthcare). The antibodies used are as follows: SMARCA4 (ab110641, Abcam), VCL (V4505, Sigma-Aldrich), and PD-L1 (13684, Cell Signaling).

Statistical Analysis

Analysis of expression profiles was performed using Fischer's exact test for categorical variable using Prism v7.0. Overall survival was analyzed with Kaplan-Meier estimates. All tests were 2-sided with a significative P-value defined <.05.

Results

Population Characteristics

We studied the immune landscape of 11 SMARCA4-UT samples (total of 9 patients), including 4 primary thoracic tumors (2 biopsies and 2 surgical specimens) and 7 distant metastases: bone (n = 1), lymph node (n = 4), brain (n = 1), and jaw (n = 1). For 2 patients (no. 3 and no. 5), tumor specimens were available from paired primary and distant metastasis, and at different timepoints for patient no. 3. Patient and sample characteristics are described in Table 1. The median age was 60 years (range, 39-73 years) and most of them were male (67%) with a poor performance status. Primary tumor locations were well balanced with 5 and 4 tumors located in the mediastinum and lungs, respectively. Median tumor size was 8.5 cm (range, 4.2-16.9). All patients displayed metastatic disease at diagnosis with 78% and 45% of them with lymph node and bone metastasis, respectively. As first-line treatment, 5 patients (55%) received chemotherapy and 2 patients received ICI (one received nivolumab alone and the other one the association nivolumab-ipilimumab), whereas 2 received only best supportive care. There was no statistical difference between median progression-free survival of patients treated by chemotherapy (2 months) versus ICI (0.5 months) (P = .12). The median OS was 1.8 months (range, 0.2-not reached) (Fig. 1). Eight patients died of the evolution of their disease and only one is still alive without evidence of disease.

Characterization of SMARCA4-UT Immune Infiltrate

Pathological features were faithful to the known morphological description and "immunohistochemical signature" (Fig. 2a).[8] Analyzed samples to assess the tumor immune infiltrate by immunostainings were those available at the diagnosis, prioritizing the primary tumor sample when possible (Fig. 2b). Eight over the 9 patients had an immune-low

Table 1. Clinical characteristics of a cohort of patients with a thoracic SMARCA4-deficient undifferentiated tumor.

Pt	Age (years)	Sex	Primary tumor site (size)	Comorbidity	Presenting symptoms (OMS)	Specimen type (specimen site)	Sync. mets	Treatment	Outcome	Time to death (months)
1	63	M	Lung (7.5 cm)	Smoking	DGC, pain from bone mets (2)	Core biopsy (bone mets)	Liver, AG, bone	CT no response on 2 regimens, TT no response	DOD	7
2	73	F	Mediastinum (10 cm)	Diabetes AVC	DGC, chest pain, dysphagia (3)	Core biopsy (primary mediastinal tumor)	Lung, bone, LN	BSC	DOD	1
3	39	M	Lung (NA)	Smoking	DGC, jaw mass (2)	Surgical specimen (jaw mets) + surgical specimen (primary lung tumor)	Jaw	Surgery of jaw mets, CT no response on 2 regimens, ICI response, surgical resection of residual lung mass, adjuvant ICI	Alive	NA
4	54	M	Lung (NA)	Smoking	Neurologic disorders (2)	Surgical specimen (brain mets)	Brain	CT no response	DOD	1
5	40	M	Mediastinum (4.8 cm)	Smoking	Dyspnea, chest pain (2)	Core biopsy (LN mets)+ surgical specimen (primary mediastinal tumor)	LN	ICI no response	DOD	1.2
6	56	M	Mediastinum (9.5 cm)	Smoking, CLL	DGC, dyspnea (2)	Core biopsy (primary mediastinal tumor)	Lung, LN	CT no response	DOD	1.5
7	60	F	Mediastinum (16.9 cm)	Smoking	DGC, dyspnea (3)	Core biopsy (LN mets)	Bone, LN, AG	BSC	DOD	0.2
8	66	F	Mediastinum (NA)	Smoking	DGC, cervical mass (3)	Surgical specimen (LN mets)	Bone, LN, AG	ICI no response	DOD	2.2
9	69	M	Lung (4.2 cm)	Smoking, Lung cancer	Hemoptysis (2)	Fine needle aspiration (LN mets)	LN	CT no response, ICI no response	DOD	6.5

Abbreviations: AG, adrenal glands; BSC, best supportive care; CLL, chronic lymphocytic leukemia; CT, chemotherapy; DGC, deterioration of general condition; DOD: dead of disease; F, female; ICI, immune checkpoint inhibitor; L, line; LN: lymph nodes; M, male; mets: metastasis; NA, not available; Sync., synchronous; TT: targeted therapy; yrs, years.

profile tumor with low CD3-positive T-cells density [median at 33.3 per mm²; range (1.1-187.3)], low CD8-positive T-cells density [median at 15.6 per mm²; range (0.6-142.7)], and low CD20-positive B-cells density [median at 0.3 per mm²; range (0-10.9)]. In these 8 tumors, CD68-positive macrophage-cells were the most represented cell type [median at 427.4 per mm²; range (140.4-1014.3)] and 5 out of them displayed TIM3 positivity in tumor cells (>100 cells per mm²). These 8 tumors were also lowly infiltrated by efficient T cells, defined by CD8-positive PD1-negative cells [median at 15.3 per mm²; range (0.1-137.7) and none contained any TLS; only one tumor had a weak PD-L1 expression on tumor cells (no. 7) (Fig. 2c, d). Conversely, one tumor (no. 3) was characterized by high immune cell densities (CD3+ at 654.2 per mm²; CD8+ at 244.4 per mm², almost all PD1 negative; CD20+ at 37.3 per mm²), presence of TLS aggregate, and weak PD-L1 expression on both tumor cells and macrophages.

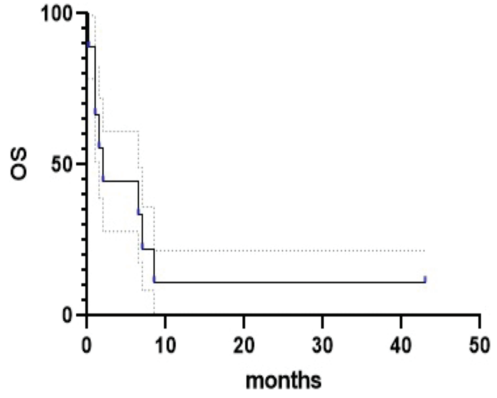


Figure 1. Kaplan-Meier overall survival (OS) curves for all 9 patients.

Analysis of the Immune Infiltrate in an Independent Cohort

To validate our observation, we analyzed an independent transcriptome dataset.[1] Unsupervised hierarchical clustering using immune scores for different immune cell types identified 2 clusters namely C1 and C2; while C1 was characterized by immune rich TME, C2 was immune desert (Fig. 3a). Notably, C1 was enriched for NSCLC independently from *SMARCA4* status (n = 9/10; 90%) (P = .001), while C2 was enriched for SMARCA4-UT (n = 11/12; 91.7%) (P = .005) and UTS (n = 5/5; 100%) (P = .0005). Similarly, gene signatures associated with immunosuppression, T-cell activation and survival, regulatory T cells, major histocompatibility complex class I, and myeloid cell chemotaxis were downregulated in SMARCA4-UT and UTS relative to NSCLC (Fig. 3b). Similarly, we observed lower lymphoid-structures-associated B-cell-specific chemokine *CXCL13* in C2-enriched sarcomas cluster. Finally, we observed lower expression of immune checkpoints genes in the majority of C2-enriched sarcomas cluster relative to C1-NSCLC enriched cluster (Fig. 3c).

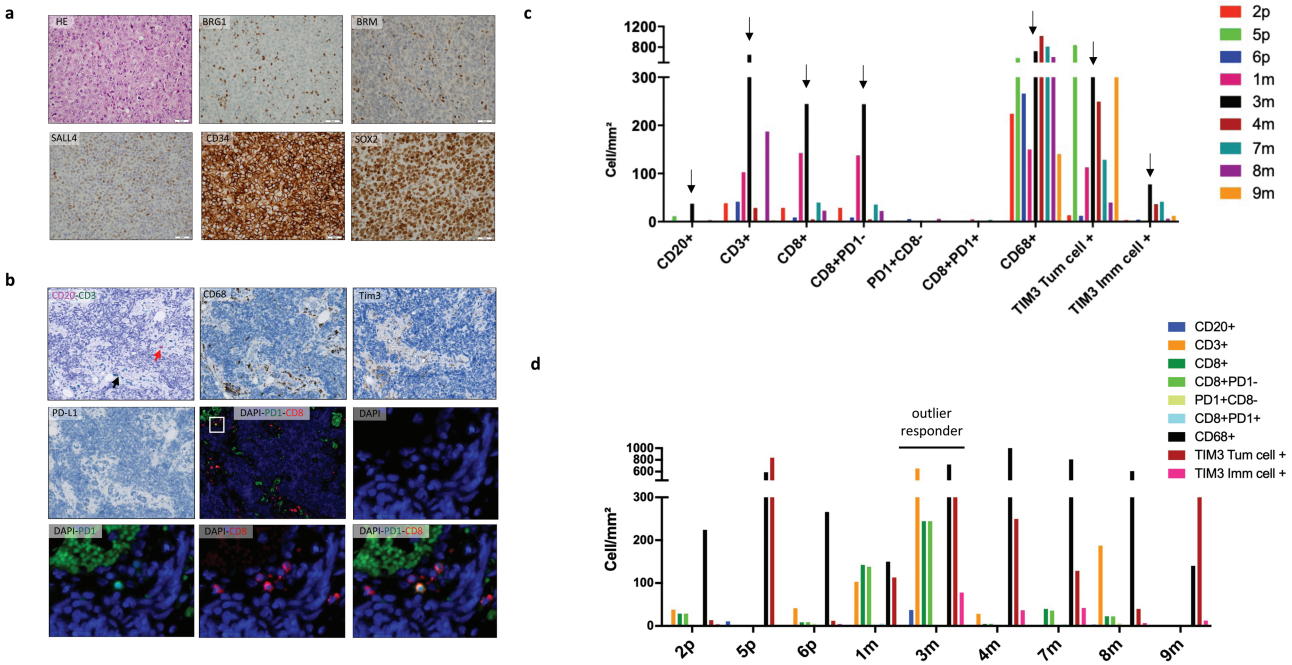


Figure 2. Pathologic features and immune infiltrate of SMARCA4-UT. (**a**) Hematoxylin and eosin staining showing tumor cells with epithelioid/rhabdoid features (×40). Immunohistochemical profile (×40): BRG1 and BRM, both diffusely lost in tumor cells with internal control in inflammatory cells; SALL4, focally expressed in tumor cells; CD34 and SOX2, diffusely expressed in tumor cells. (**b**) Low immune infiltrate on immunohistochemical profile: very few CD3 and CD8-positive T cells, whereas the CD68-positive macrophage cells are the most represented cell type in the TME with some TIM3-positive but PD-L1-negative tumor cells. Immunofluorescent assays: global visualization for CD8 and PD1 staining with autofluorescence from red blood cells in green, with a zoom in the very few double-positive cells in the white square. (**c**) Histogram representing immune cells densities (cells/mm²) shows low densities of immune infiltrate on primary tumor (p) or metastasis (m) by immune cells type (arrow indicates individual cell density for sample no. 3m from patient with the outlier response to ICI). (**d**) Histogram representing immune cells densities (cells/mm²) in tumor cores by specimen: primary tumors on the left and metastasis on the right. Abbreviations: SMARCA4-UT, SMARCA4-deficient thoracic sarcoma; TME, tumor microenvironment.

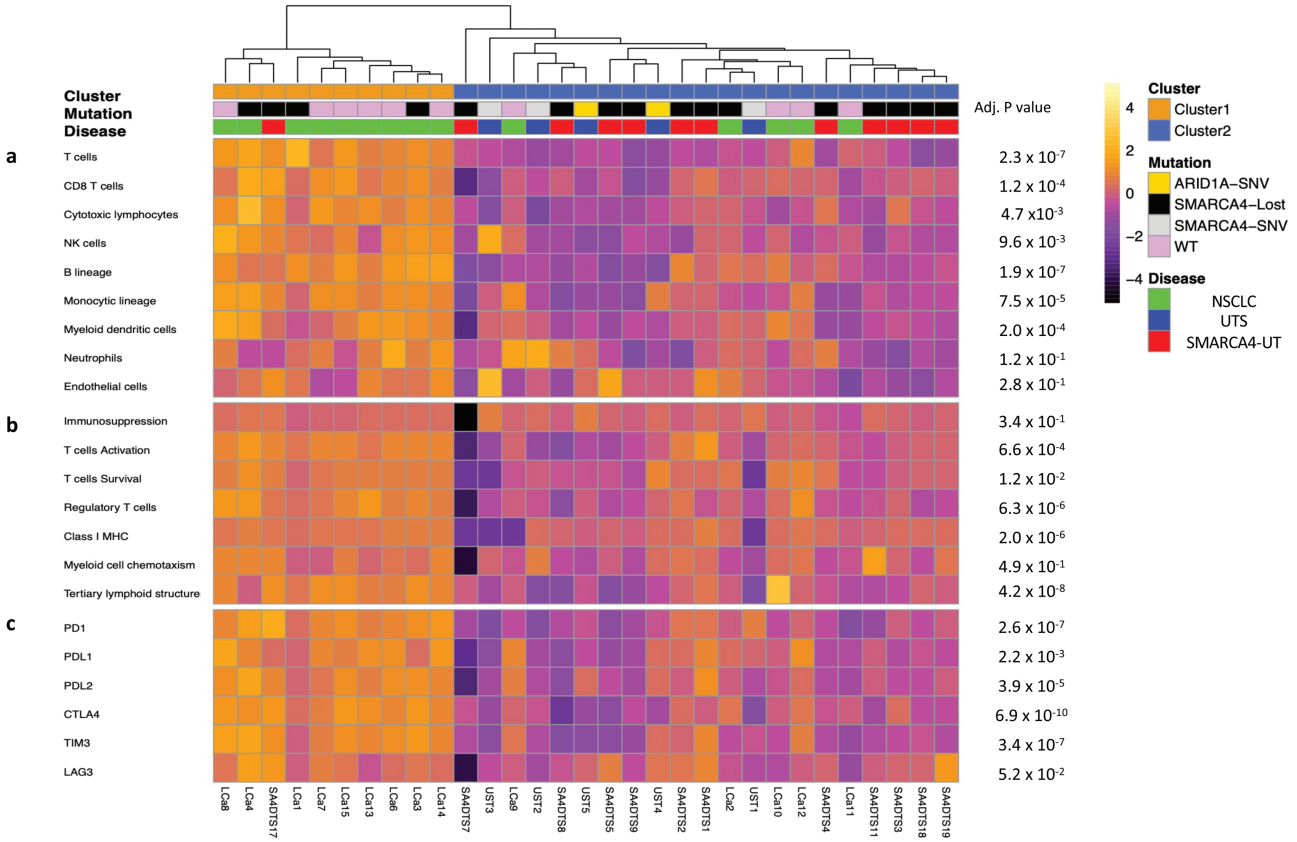


Figure 3. SMARCA4-UT exhibits strongly different TMEs than NOS NSCLC regardless of *SMARCA4* mutation status. This figure refers to the Le Loarer et al cohort (n = 31). (**a**) Unsupervised clustering of samples was performed based on the metagene Z-score for the included populations of MCP-counter. (**b**) Expression of gene signatures related to the functional orientation of the immune TME. (**c**) Expression of genes related to immune checkpoints. Abbreviations: MCP, microenvironment cell population; NK cells, natural killer cells; NOS NSCLC, not other specified non–small cell lung cancer; SMARCA4-UT, SMARCA4-deficient thoracic sarcoma; TME, tumor microenvironment; UTS: undifferentiated thoracic sarcoma.

Patient's Response to Immune Checkpoint Inhibitors and Immune Infiltration Correlate

Among our patients, 4 received ICI as one of their anticancer treatment. Their characteristics and outcomes are detailed in Fig. 4. Two of them (no. 5 and no. 8) died quickly within 10 weeks after the diagnosis and had ICI as their only treatment (Fig. 4a). The third one (no. 9) died within 6 months and received chemotherapy as first line and ICI as second line. Those 3 patients had immune-low tumor profile (Fig. 4b). For the patient no. 5 matched distant metastasis was also available and, in contrast with the primary tumor, weak PD-L1 expression (Fig. 4c) and mature TLS were found in the lymph node metastasis (Fig. 4d).

The only alive patient of the cohort was patient no. 3, a 40-year-old man who was referred for the onset of a rapidly growing mass of the lower jawbone in the setting of weight loss (Fig. 5a, b). After initial surgery of the jaw metastasis and the failure of 2 lines of chemotherapy, the patient showed rapid and partial long-term response following combination of ipilimumab and nivolumab, followed by nivolumab maintenance and surgical resection of lung residual tumor. Resection was complete leading to a complete remission status, and nivolumab was kept as maintenance treatment for almost 2 years after the initial diagnosis. Nowadays, while being on follow-up, patient is fully active and able to carry on all pre-disease performance without restriction. Regarding his tumor's immune profile, the first tumor specimen presented a TLS aggregate even although it was not a mature one, whereas no TLS nor aggregate were observed in the second tumor specimen after several months on ICI treatment (Fig. 5c-e). Whereas CD3-positive cells density was similar in both specimens, CD20-positive cells density was higher in the specimen collected before any systemic anticancer treatment than in the residual one collected after the different treatments. We thus investigated double-positive (CD8+PD1+) cells, which showed low density in the jaw specimen, but had a tenfold increase in the residual lung tumor. In addition, PD1-positive but CD8-negative cells density was also significantly higher in the residual lung tumor compared with the jaw lesion. Puzzlingly, PD-L1 intensity staining was higher in the residual tumor. Overall, the immune infiltrate in the sample collected after ICI showed clear signs of exhaustion.

As this case was an outlier, we performed a comprehensive genomic profiling on the 2 samples collected before and after ICI. On the first sample, *SMARCA4* Q183* mutation was found along with *HRAS* (G13V), *TP53* (G245V), and *BLM* (D107fs*3) mutations. Tumor was not considered as microsatellite instable and TMB profile was considered as high (29 mutations per megabase). On the second sample, no additional somatic mutation was observed. However, we observed *JAK1* and *JUN* amplification, suggesting a role of the copy number variation in genes related to IFNG pathway in resistance and immune exhaustion.

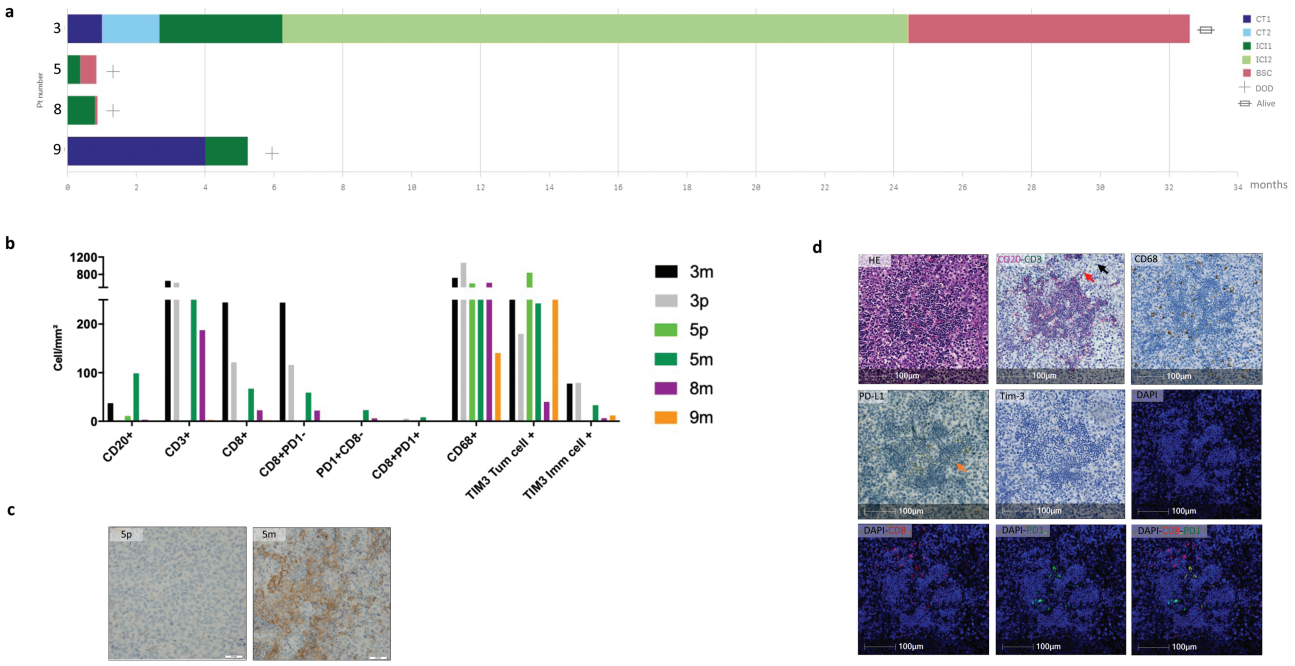


Figure 4. Sub-population treated by immune checkpoint inhibitor (n = 4). (**a**) Swimming plot of the received treatment: patient no. 3: L1: doxorubicin/ifosfamide; L2: actinomycin D/dacarbazine/cisplatin; L3: ipilimumab/nivolumab; L4 nivolumab; patient no. 5: L1: ipilimumab/nivolumab; patient no. 8: L1: nivolumab; patient no. 9: L1: carboplatin/paclitaxel; L2 nivolumab. (**b**) Immune cells densities (cells/mm^2) in tumor cores by specimen. (**c**) PD-L1 staining on primary tumor (p) and metastasis (m) of patient no. 5. (**d**) Representation of mature TLS in the lymph node metastasis of patient no. 5 with: hematoxylin and eosin stain; IHC CD20 pink stain for B cells (red arrow) and CD3 green for T cells (black arrow); IHC CD68-positive cells for macrophages; very few PD-L1-positive cells (orange arrow); absence of TIM3-positive cells; immunofluorescent with CD8 (red) and PD1 (green) stain. Abbreviations: BSC, best supportive care; CT, chemotherapy; DOD: dead of the disease; ICI, immune checkpoint inhibitor; IHC: immunohistochemical.

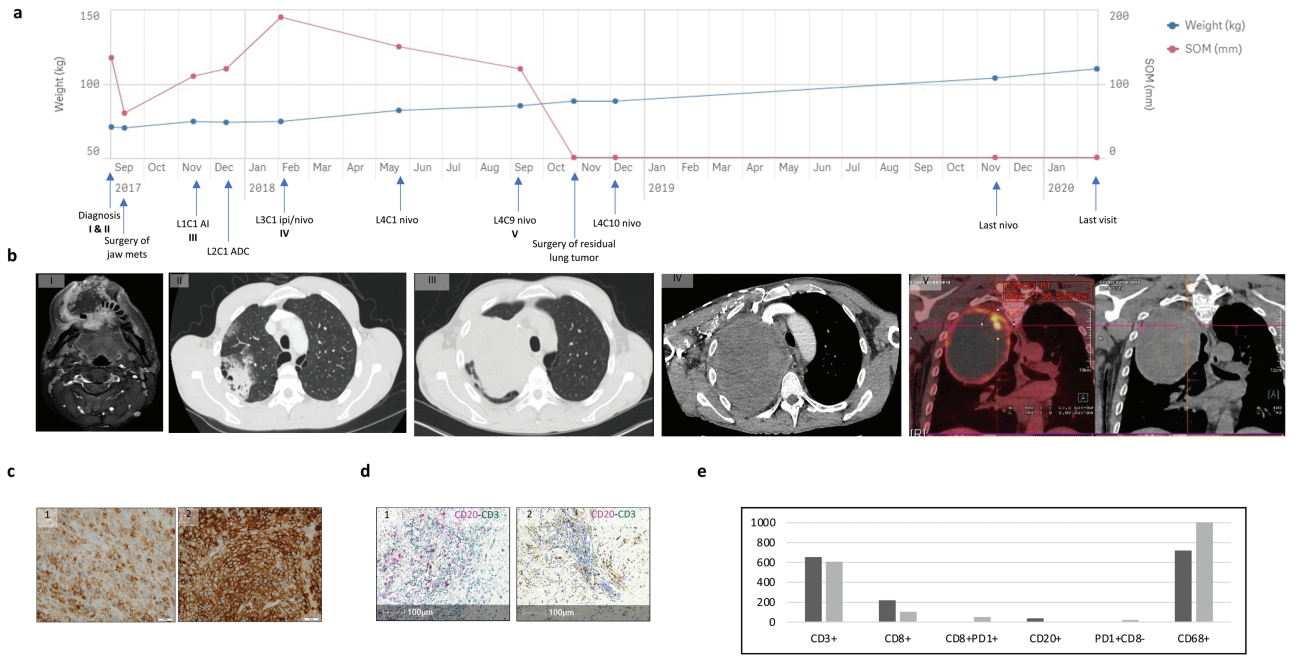


Figure 5. Focus on patient no. 3 with the outlier response to ICI. (**a**) Timeline of his medical history reporting the evolution of his weight and the SOM on computed tomography (CT) scans. (**b**) Images features of the evolution of the main target lesions: I: Cervical MRI-enhanced gadolinium T1-weighted axial sections of the bulky jaw metastasis; II: Thoracic CT-scan axial section showing the initial lesion in the right upper pulmonary lobe in September 2017; III: massive progression of the lesion at the beginning of the first line of chemotherapy in November 2017; IV: progressive lesion at the beginning of the ICI combination with a superior vena cava syndrome in February 2018; V: positron emission tomography-scan frontal sections showing partial response of the lesion after 5 cycles of combination of ipilimumab with nivolumab and 10 cycles of maintenance with nivolumab in September 2018, with a hypermetabolism at the periphery of the right apical solid-cystic pulmonary mass predominantly apical (SUV max 17.6). (**c**) PD-L1 IHC (×40) and (**d**) TLS presence before (1) and after ICI (2). Before ICI, one aggregate CD20 (IHC pink stain)-CD3 (IHC green stain)-positive, no longer seen after ICI (in yellow-brown, hemosiderin deposit). (**e**) Comparison of cells densities (cells/mm^2) between tissues before (dark gray) and after (light gray) ICI. Abbreviations: ADC, actinomycin D/dacarbazine/cisplatin; AI, doxorubicin/ifosfamide; C, cycle; CT, computed tomography; IHC: immunohistochemical; ICI, immune checkpoint inhibitor; L, line; MRI, magnetic resonance imaging; SOM, sum of diameters of target lesions.

In Vitro Analyses of knockdown of SMARCA4 in Cell Lines

To investigate whether TME changes observed are linked to cell ontogeny, we checked in vitro if SMARCA4 loss of function could induce gene expression changes of CXCL9 chemokine and PD-L1 in NSCLC and thoracic fibroblast cell lines, treated with/without IFNG. Strikingly, we observed a significant increase of CXCL9 expression levels in NSCLC cell line following SMARCA4 knockdown with ($P < .05$) and without ($P = .05$) IFNG stimulations, while no effect was observed in the lung fibroblast cell line (Fig. 6a, b). In addition, SMARCA4 knockdown induced an upregulation of PD-L1 expression in NSCLC cell line, but only with a significant increase in the cells without IFNG stimulation ($P < .001$), with no effect in the lung fibroblast cell line (Fig. 6c).

Discussion

To our knowledge, our study is the first one to report a comprehensive analysis of the immune infiltrate in SMARCA4-UT, by analyzing the immune cell infiltrates, the presence of TLS and PD-L1 expression in tumor and macrophage cells. Our study showed that SMARCA4-UT are mainly immune-desert tumors, differing from NSCLC but similar to most STS subtypes. This might explain limited efficacy of ICI in this setting, although we did observe one responder in our cohort.

Regardless of the tumor types, immune cell infiltrates, TLS, PD-L1 expression in tumor cells or immune infiltrating cells, TMB, IFN signature and DNA mismatch-repair deficiency status have been shown to represent potential biomarkers of response to ICI in several cancer subtypes.[40-44] In our cohort, SMARCA4-UT without TLS-positive tumor did not respond to ICI and had a very poor OS. Among them, one patient died of local progression of his thoracic TLS-negative tumor in less than 2 months despite mature TLS in his lymph node metastasis. The outcome of this patient raises questions about the heterogeneity of the TLS status between the different tumor sites and the relevance of the evaluation of predictive factors of response to treatment at a metastatic site, while the local progression of the primary tumor is the main prognosis factor. The only surviving patient is the one who responded dramatically to ICI and had a TLS aggregate in his primary tumor. Sample from the surgery of the residual tumor after ICI showed the loss of TLS, reinforcing the correlation between TLS and response to ICI. This loss was never previously reported as a secondary resistance mechanism but might explain the clone resistant's immune infiltrate. With this loss, immune cells could not be educated any more to recognize new tumor antigens which could thereby lead to their exhaustion. Our observation is consistent with the previous report that STS showing a class E immune-high subgroup characterized by the presence of TLS have a high response rate to ICI regardless of their histologic subtypes.[16] The results of PEMBROSARC, the first clinical trial investigating the efficacy of ICI in TLS-positive STS, were recently presented.[45] Briefly, TLS-positive STS were found in 48 of the 240 patients screened and among them 35 were included and 30 were eligible for efficacy. While no patient had a complete response in the TLS-positive STS, 8 patients had a partial response and 5 a stable disease. With a 6-month non-progression rate of 40% in TLS-positive STS while it was only of 4.2% in all comers, this study corroborated that TLS status is an efficient approach to tailor ICI in STS patients.

In our cohort, in addition to having TLS, the only high immune infiltrate tumor also had a high TMB which might be another explanation to the outlier response to ICI. Similarly

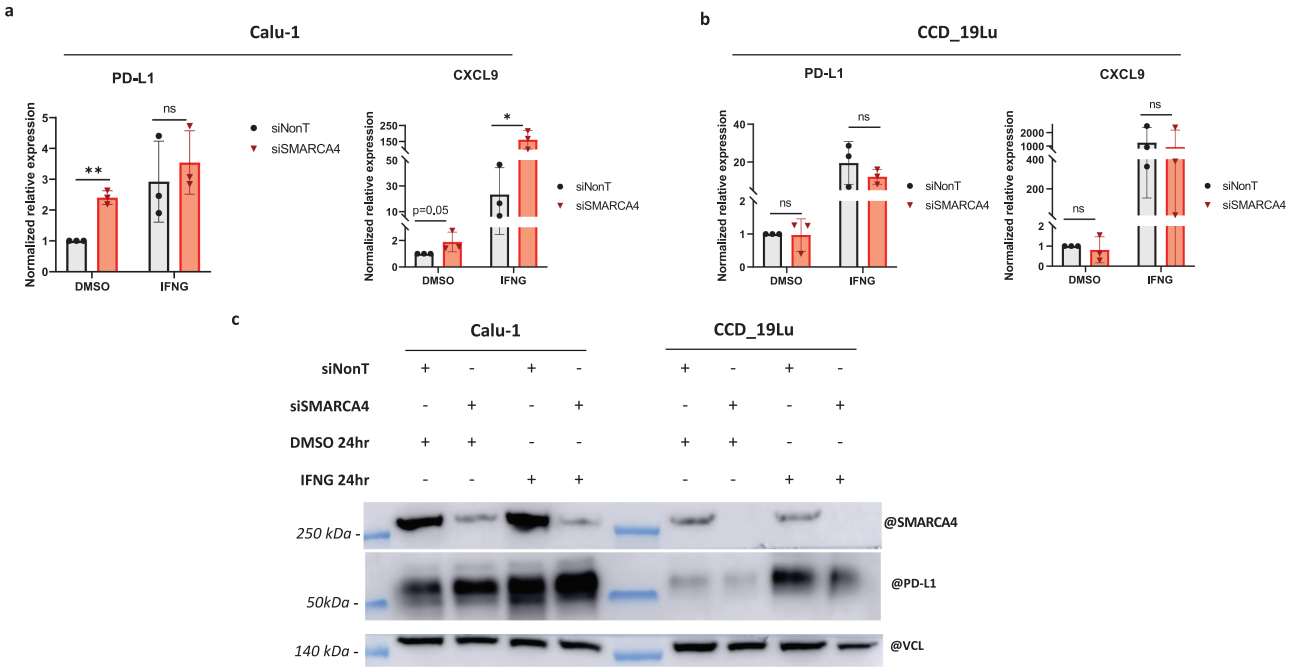


Figure 6. In vitro analyses on the impact of SMARCA4 loss on the response to interferon gamma treatment in 2 different cell lines (Calu-1: human lung carcinoma; CCD_19Lu: human lung fibroblasts). PD-L1 and CXCL9 chemokine relative expression in Calu-1 cell line (**a**) and in CCD_19Lu cell line (**b**). (**c**) PD-L1 protein expression according to *SMARCA4* deficiency status and interferon treatment in the 2 cell lines. Abbreviations: DMSO, dimethylsulfoxid; IFNG, interferon gamma; ns, no significate; VCL, vinculin; $*P < .05$,$**P < .001$.

to the change in the TLS status between samples, a *JAK1* amplification appeared on the lung residual tumor. Intriguingly, *JAK1* loss-of-function mutations or deletions have been previously described as a mechanism of ICI resistance. *JAK1/2* loss of function leads to loss of IFNG signaling, preventing PD-L1 expression and thus making ICI futile.[46] In our case, we could hypothesize that the *JAK1* amplification might be responsible for the increased PD-L1 expression and contributed to immune cells exhaustion.

Interestingly, despite previous promising results in different cancer cohorts suggesting that BAF-deficient tumors might be more sensitive to ICI with improved clinical outcomes,[26-29,47-49] our conflicting results among others[22-24,30] enforced the hypothesis that wider genomic contexts than just BAF deficiency and cell's ontogeny account for the efficacy of ICI. Indeed, in our cohort, efficacy of ICI was limited and restricted to a tumor with specific immune TME, in contrast to previous published data.[31,32,50] The immune infiltrate does not seem to be dependent on SMARCA4-deficiency itself. Indeed, in vitro analyses reinforced the hypothesis than cell ontogeny may play a central role on the effect of *SMARCA4* loss of function. Calu-1 cell line was chosen as a NSCLC cell line, whereas the human thoracic fibroblasts cell line CCD_19Lu had to replace the non-existing SMARCA4-UT cell line. These experiments tend to show that SMARCA4 loss upregulated PD-L1 expression in the NSCLC cell line, whereas there was no difference of its expression in thoracic fibroblasts. Even after a strong induction of the IFN pathway through IFNG treatment, PD-L1 expression tends to only be slightly downregulated by SMARCA4 loss in fibroblasts. CXCL9 expression was also assessed since CXLC9 has been described as one of the chemokine reflecting the most the IFN pathway activation and has also been suggested as a predictor of ICI response.[51] Similarly to PD-L1 expression, we did find an upregulated CXCL9 expression in the NSCLC cell line with SMARCA4 knockdown with/without IFNG stimulation, whereas there was no difference seen in the thoracic fibroblast cell line. Though the underlying mechanisms are not yet understood, SMARCA4-dependent activation of IFN signaling did differ from one cell line to the other. Moreover, in our work, immune deconvolution using an independent transcriptome dataset demonstrated a distinct immune profile between NOS NSCLC and sarcomas regardless of the *SMARCA4* mutational status. Another example is SCCOHT, previously described as immune-high tumors as opposed to SMARCA4-UT.[26,27] Like other mutations or deficiencies, tumor cell of origin may change the driver mutation pathogenesis and clinical outcomes.

As already mentioned, in the latest edition of WHO classification of thoracic tumors, the pathology community chose to not use anymore the term "sarcoma" to described this entity, instead the term SMARCA4-UT was preferred to better reflect their possible epithelial histogenesis.[52] In a useful way our study highlighted SMARCA4-UT similarities with other subtypes of thoracic sarcomas, but also their distinct immune profile from SMARCA4-deficient NSCLC. Therefore, these 2 tumor subtypes appear to have different characteristics, that should encourage specific research into the clinical implications of this distinction.

It is important to mention herein that our study has several limitations including the low number of cases explored, mainly due to the extreme rarity of the disease. Among the potential biomarkers of response to ICI, immune infiltrates, TLS, and PD-L1 expression were studied for all samples. However, the genomic comprehensive analysis, allowing to assess TMB and DNA mismatch-repair deficiency status, was only performed on the samples of the responder.

Conclusion

Overall, our data demonstrate that SMARCA4-UT are mainly immune desert tumors similarly to most STS subtypes with limited efficacy to ICI. The detailed analysis of the tumor landscape of the long-term responder highlights the relevance of TLS as predictive markers to ICI efficacy in SMARCA4-UT as in STS. Finally, our data suggest that TME of SMARCA4-driven tumors varies according to the cell of origin. Further studies are needed to understand the interplay between alterations of BAF complexes, cell ontogeny and immunity.

Acknowledgments

We would like to thank the patients for giving consent for publication.

Funding

None declared.

Conflict of Interest

Celine Mascaux: Roche, AstraZeneca, Kephren, Bristol-Myers Squibb, Pfizer, Sanofi, MSD, Takeda, Amgen (ET, H, SAB), European Patent Application EP19305434.3, European Patent Application EP19305535.7 (OI), **Gabriel Malouf**: BMS, MSD, Ipsen, Pfizer (C/A). The other authors indicated no financial relationships.

(C/A) consulting/advisory relationship; (RF) research funding; (E) employment; (ET) expert testimony; (H) honoraria received; (OI) ownership interests; (IP) intellectual property rights/inventor/patent holder; (SAB) scientific advisory board.

Author Contributions

Conception/design: J.G., N.W., F.B., J-E.K., C.S-F., W.H.F., G.G.M. Provision of study material/patients: J.G., B.V., N.W., V.L., C.M., M-P.C., J-E.K., G.G.M. Collection and/or assembly of data: J.G., G.D., B.V., N.W., A.B., M.M., G.L., C.S-F., W.H.F., G.G.M. Data analysis and interpretation: J.G.,G.D., B.V., N.W., A.B., I.D., J-E.K., C.S-F., W.H.F., G.G.M. Performed the experiments and analyzed the data; I.D., J.-E.K., C.S.-F., W.H.F., G.G.M. Supervised the experiments; N.W., V.L., F.B., M.-P.C. Provided resources: J.G., G.D., B.V., N.W., C.M., F.B., J.-E.K., C.S.-F., W.H.F., G.G.M. Manuscript writing: All authors.

Data Availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Supplementary Material

Supplementary material is available at *The Oncologist* online.

References

1. Le Loarer F, Watson S, Pierron G et al. SMARCA4 inactivation defines a group of undifferentiated thoracic malignancies transcriptionally related to BAF-deficient sarcomas. *Nat Genet*. 2015;47:1200-1205.
2. Mashtalir N, D'Avino AR, Michel BC et al. Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell*. 2018;175:1272-1288.e20.
3. Kadoch C, Hargreaves DC, Hodges C et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet*. 2013;45:592-601.
4. Shain AH, Pollack JR. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS One*. 2013;8:e55119.
5. Mittal P, Roberts CWM. The SWI/SNF complex in cancer - biology, biomarkers and therapy. *Nat Rev Clin Oncol*. 2020;17:435-448.
6. Tsao M. PL01.05 The new WHO classification of lung tumors. *J Thorac Oncol*. 2021;16:S63.
7. Crombé A, Alberti N, Villard N et al. Imaging features of SMARCA4-deficient thoracic sarcomas: a multi-centric study of 21 patients. *Eur Radiol*. 2019. https://doi.org/10.1007/s00330-019-06017-x.
8. Perret R, Chalabreysse L, Watson S et al. SMARCA4-deficient thoracic sarcomas: clinicopathologic study of 30 cases with an emphasis on their nosology and differential diagnoses. *Am J Surg Pathol*. 2019;43:455-465.
9. Sauter JL, Graham RP, Larsen BT et al. SMARCA4-deficient thoracic sarcoma: a distinctive clinicopathological entity with undifferentiated rhabdoid morphology and aggressive behavior. *Mod Pathol Off J U S Can Acad Pathol Inc*. 2017;30:1422-1432.
10. Yoshida A, Kobayashi E, Kubo T et al. Clinicopathological and molecular characterization of SMARCA4-deficient thoracic sarcomas with comparison to potentially related entities. *Mod Pathol Off J U S Can Acad Pathol Inc*. 2017;30:797-809.
11. Takeda M, Tani Y, Saijo N et al. Cytopathological features of SMARCA4-deficient thoracic sarcoma: report of 2 cases and review of the literature. *Int J Surg Pathol*. 2020;28:109-114.
12. Groisberg R, Hong DS, Behrang A et al. Characteristics and outcomes of patients with advanced sarcoma enrolled in early phase immunotherapy trials. *J Immunother Cancer*. 2017;5:100.
13. Tawbi HA, Burgess M, Bolejack V et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol*. 2017. https://doi.org/10.1016/S1470-2045(17)30624-1.
14. Toulmonde M, Penel N, Adam J et al. Use of PD-1 Targeting, macrophage infiltration, and ido pathway activation in sarcomas: A phase 2 clinical trial. *JAMA Oncol*. 2017. https://doi.org/10.1001/jamaoncol.2017.1617.
15. Blay J-Y, Penel N, Ray-Coquard IL et al. High clinical activity of pembrolizumab in chordoma, alveolar soft part sarcoma (ASPS) and other rare sarcoma histotypes: The French AcSé pembrolizumab study from Unicancer. *J Clin Oncol*. 2021;39:11520-11520.
16. Petitprez F, de Reyniès A, Keung EZ et al. B cells are associated with survival and immunotherapy response in sarcoma. *Nature*. 2020. https://doi.org/10.1038/s41586-019-1906-8.
17. Dieu-Nosjean M-C, Goc J, Giraldo NA et al. Tertiary lymphoid structures in cancer and beyond. *Trends Immunol*. 2014;35:571-580.
18. Engelhard VH, Rodriguez AB, Mauldin IS et al. Immune cell infiltration and tertiary lymphoid structures as determinants of antitumor immunity. *J Immunol Baltim MD 1950*. 2018;200:432-442.
19. Sautès-Fridman C, Petitprez F, Calderaro J et al. Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat Rev Cancer*. 2019;19:307-325.
20. Keung EZ, Burgess M, Salazar R et al. Correlative analyses of the SARC028 trial reveal an association between sarcoma-associated immune infiltrate and response to pembrolizumab. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2020. https://doi.org/10.1158/1078-0432.CCR-19-1824.
21. Zhou M, Yuan J, Deng Y et al. Emerging role of SWI/SNF complex deficiency as a target of immune checkpoint blockade in human cancers. *Oncogenesis* 2021;10:3.
22. Abou Alaiwi S, Nassar AH, Xie W et al. Mammalian SWI/SNF complex genomic alterations and immune checkpoint blockade in solid tumors. *Cancer Immunol Res*. 2020. https://doi.org/10.1158/2326-6066.CIR-19-0866.
23. Li J, Wang W, Zhang Y et al. Epigenetic driver mutations in ARID1A shape cancer immune phenotype and immunotherapy. *J Clin Invest*. 2020;130:2712-2726.
24. Liu X-D, Kong W, Peterson CB et al. PBRM1 loss defines a nonimmunogenic tumor phenotype associated with checkpoint inhibitor resistance in renal carcinoma. *Nat Commun*. 2020;11:2135.
25. Leruste A, Chauvin C, Pouponnot C et al. Immune responses in genomically simple SWI/SNF-deficient cancers. *Cancer*. 2021;127:172-180.
26. Lu B, Shi H. An in-depth look at small cell carcinoma of the ovary, hypercalcemic type (SCCOHT): clinical implications from recent molecular findings. *J Cancer*. 2019;10:223-237.
27. Jelinic P, Ricca J, Van Oudenhove E et al. Immune-active microenvironment in small cell carcinoma of the ovary, hypercalcemic type: rationale for immune checkpoint blockade. *J Natl Cancer Inst*. 2018;110:787-790.
28. Hanna GJ, Lizotte P, Cavanaugh M et al. Frameshift events predict anti-PD-1/L1 response in head and neck cancer. *JCI Insight*. 2018;3. https://doi.org/10.1172/jci.insight.98811.
29. Schoenfeld AJ, Bandlamudi C, Lavery JA et al. the genomic landscape of SMARCA4 alterations and associations with outcomes in patients with lung cancer. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2020;26:5701-5708.
30. Marinelli D, Mazzotta M, Scalera S et al. KEAP1-driven co-mutations in lung adenocarcinoma unresponsive to immunotherapy despite high tumor mutational burden. *Ann Oncol Off J Eur Soc Med Oncol*. 2020;31:1746-1754.
31. Henon C, Blay J-Y, Massard C et al. Long lasting major response to pembrolizumab in a thoracic malignant rhabdoid-like SMARCA4-deficient tumor. *Ann Oncol Off J Eur Soc Med Oncol*. 2019;30:1401-1403.
32. Takada K, Sugita S, Murase K et al. Exceptionally rapid response to pembrolizumab in a SMARCA4-deficient thoracic sarcoma overexpressing PD-L1: A case report. *Thorac Cancer*. 2019;10:2312-2315.
33. Anžič N, Krasniqi F, Eberhardt A-L et al. Ipilimumab and pembrolizumab mixed response in a 41-year-old patient with SMARCA4-deficient thoracic sarcoma: an interdisciplinary case study. *Case Rep Oncol*. 2021;14:706-715.
34. Kawachi H, Kunimasa K, Kukita Y et al. Atezolizumab with bevacizumab, paclitaxel and carboplatin was effective for patients with SMARCA4-deficient thoracic sarcoma. *Immunotherapy*. 2021;13:799-806.
35. Calderaro J, Petitprez F, Becht E et al. Intra-tumoral tertiary lymphoid structures are associated with a low risk of early recurrence of hepatocellular carcinoma. *J Hepatol*. 2019;70:58-65.
36. Vanhersecke L, Brunet M, Guégan J-P et al. Mature tertiary lymphoid structures predict immune checkpoint inhibitor efficacy in solid tumors independently of PD-L1 expression. *Nat Cancer*. 2021;2:794-802.
37. Dobin A, Davis CA, Schlesinger F et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma Oxf Engl*. 2013;29:15-21.
38. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
39. Becht E, Giraldo NA, Lacroix L et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* .2016;17:218.
40. Duffy MJ, Crown J. Biomarkers for predicting response to immunotherapy with immune checkpoint inhibitors in cancer patients. *Clin Chem*. 2019;65:1228-1238.

41. Lapuente-Santana Ó, Eduati F. Toward systems biomarkers of response to immune checkpoint blockers. *Front Oncol*. 2020;10:1027.
42. Walk EE, Yohe SL, Beckman A et al. The cancer immunotherapy biomarker testing landscape. *Arch Pathol Lab Med*. 2020;144:706-724.
43. Jardim DL, Goodman A, de Melo Gagliato D et al. The challenges of tumor mutational burden as an immunotherapy biomarker. *Cancer Cell*. 2021;39:154-173.
44. Paijens ST, Vledder A, de Bruyn M et al. Tumor-infiltrating lymphocytes in the immunotherapy era. *Cell Mol Immunol*. 2021;18:842-859.
45. Italiano A, Bessede A, Bompas E et al. PD1 inhibition in soft-tissue sarcomas with tertiary lymphoid structures: a multicenter phase II trial. *J Clin Oncol*. 2021;39:11507-11507.
46. Shin DS, Zaretsky JM, Escuin-Ordinas H et al. Primary resistance to PD-1 blockade mediated by JAK1/2 mutations. *Cancer Discov*. 2017;7:188-201.
47. Miao D, Margolis CA, Gao W et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science*. 2018;359:801-806.
48. Braun DA, Ishii Y, Walsh AM et al. Clinical validation of PBRM1 alterations as a marker of immune checkpoint inhibitor response in renal cell carcinoma. *JAMA Oncol*. 2019. https://doi.org/10.1001/jamaoncol.2019.3158.
49. Li L, Li M, Jiang Z et al. ARID1A mutations are associated with increased immune activity in gastrointestinal cancer. *Cells*. 2019;8. https://doi.org/10.3390/cells8070678.
50. ESMO. Pembrolizumab Delivers Clinical Benefit in Selected Histotypes of Rare Sarcoma. Available at https://www.esmo.org/oncology-news/pembrolizumab-delivers-clinical-benefit-in-selected-histotypes-of-rare-sarcoma Accessed March 1, 2021.
51. Litchfield K, Reading JL, Puttick C et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell*. 2021;184:596-614.e14.
52. Rekhtman N, Montecalvo J, Chang JC et al. SMARCA4-deficient thoracic sarcomatoid tumors represent primarily smoking-related undifferentiated carcinomas rather than primary thoracic sarcomas. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2020;15:231-247.

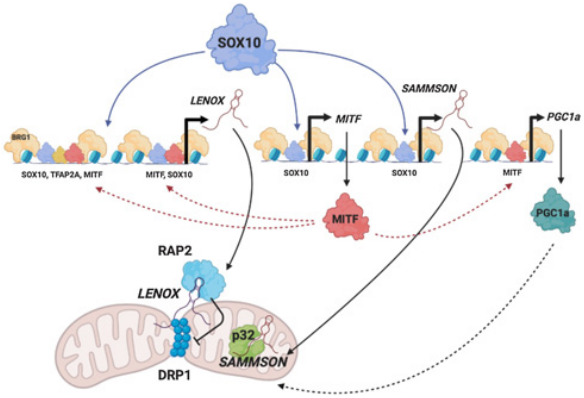
The LncRNA LENOX Interacts with RAP2C to Regulate Metabolism and Promote Resistance to MAPK Inhibition in Melanoma

Giovanni Gambi[1,2,3,4], Gabrielle Mengus[1,2,3,4], Guillaume Davidson[1,2,3,4], Ewout Demesmaeker[5], Alessandro Cuomo[6], Tiziana Bonaldi[6], Vicky Katopodi[5], Gabriel G. Malouf[1,2,3,4], Eleonora Leucci[5], and Irwin Davidson[1,2,3,4,7]

ABSTRACT

Tumor heterogeneity is a key feature of melanomas that hinders development of effective treatments. Aiming to overcome this, we identified LINC00518 (LENOX; lincRNA-enhancer of oxidative phosphorylation) as a melanoma-specific lncRNA expressed in all known melanoma cell states and essential for melanoma survival *in vitro* and *in vivo*. Mechanistically, LENOX promoted association of the RAP2C GTPase with mitochondrial fission regulator DRP1, increasing DRP1 S637 phosphorylation, mitochondrial fusion, and oxidative phosphorylation. LENOX expression was upregulated following treatment with MAPK inhibitors, facilitating a metabolic switch from glycolysis to oxidative phosphorylation and conferring resistance to MAPK inhibition. Consequently, combined silencing of LENOX and RAP2C synergized with MAPK inhibitors to eradicate melanoma cells. Melanomas are thus addicted to the lncRNA LENOX, which acts to optimize mitochondrial function during melanoma development and progression.

Significance: The lncRNA LENOX is a novel regulator of melanoma metabolism, which can be targeted in conjunction with MAPK inhibitors to eradicate melanoma cells.



SOX10 orchestrates mitochondrial function and metabolism in melanoma.

Introduction

Intratumor heterogeneity is a major determinant of therapeutic resistance. Melanoma tumors are notoriously heterogeneous comprising cell populations with distinct properties and gene expression signatures (1–5). Rare vulnerabilities common to most melanoma cell states have been identified and successfully exploited to overcome therapy resistance, such as inhibition of mitochondrial translation (6).

Transcription factor SOX10 is an essential determinant of heterogeneity in melanoma. Melanocytic state cells express microphthalmia-associated transcription factor (MITF) and SOX10, whereas neural crest-like cells express SOX10, but not MITF (4, 7). In melanocytic cells, MITF and SOX10 bind together at *cis*-regulatory elements to promote proliferation, survival, and cell metabolism (8–11). Targeted therapies can induce the undifferentiated state expressing neither MITF nor SOX10 (4, 12–14). Upon MAP kinase inhibition, the neural crest and undifferentiated states play key roles in minimal residual disease and the emergence of drug-resistant populations (15, 16).

Long noncoding (Lnc) RNAs are regulators of virtually every process in the cell (17) in particular adaptive processes involved in tumor progression and therapy resistance (18–20). The melanoma-specific lncRNA SAMMSON (Linc01212) is a SOX10-regulated gene essential for melanoma cell proliferation and survival (21, 22). SAMMSON inhibition severely affects mitochondrial function, inducing proteotoxic stress, resulting in cell death *in vitro* and tumor regression in preclinical settings when combined with MAPK inhibition (21).

Here we characterize the melanoma-specific lncRNA LENOX (LINC00518) expressed in all known melanoma states and essential for proliferation and survival. LENOX interacts with the RAP2C small GTPase promoting its interaction with DRP1 and impairing mitochondrial fission through enhanced DRP1 S637 phosphorylation. The resulting increase in mitochondrial elongation and optimal oxidative phosphorylation promotes melanoma cell survival and resistance to MAP kinase inhibitors.

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France. [2]Centre National de la Recherche Scientifique, UMR7104, Illkirch, France. [3]Institut National de la Santé et de la Recherche Médicale, U1258, Illkirch, France. [4]Université de Strasbourg, Illkirch, France. [5]Laboratory for RNA Cancer Biology, KU Leuven, Leuven, Belgium. [6]Nuclear Proteomics Institute to Study Gene Expression, Milano, Italy. [7]Equipe Labélisée Ligue contre le Cancer.

Corresponding Authors: Irwin Davidson, Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire, 1 Rue Laurent Fries, Illkirch, 67404, France. E-mail: irwin@igbmc.fr; and Eleonora Leucci, Laboratory for RNA Cancer Biology, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. E-mail: eleonora.leucci@kuleuven.be

Cancer Res 2022;82:4555–70

doi: 10.1158/0008-5472.CAN-22-0959

Materials and Methods

Analysis of RNA sequencing data

Single-cell RNA sequencing (RNA-seq) data from MEL006 patient-derived xenograft (PDX) tumors (GSE116237) were processed using Seurat and genes were divided into protein-coding and noncoding based on their biotype (Ensembl 104). Mean expression level in each melanoma population was calculated excluding cells in which the gene was not captured. SD across the melanoma populations was divided by mean expression to calculate the coefficient of variation and rank genes. The 50 protein-coding and noncoding genes with lowest coefficient of variation were identified and their expression was measured across normal tissues and melanoma samples in the genotype-tissue expression (GTEx) and The Cancer Genome Atlas (TCGA) databases, respectively. All other used RNA sequencing datasets are indicated in the text.

Cell culture and GapmeR transfections

Melanoma cell lines Sk-mel-25, Sk-mel-25R, Sk-mel-28, and 501Mel were grown in RPMI1640 medium supplemented with 10% fetal calf serum (FCS) and gentamycin; IGR-37 and IGR-39 in RPMI1640 medium supplemented with 15% FCS and gentamycin. MM011, MM117, MM047, and MM099 were grown in HAM-F10 medium supplemented with 10% FCS, 5.2 mmol/L glutamax, 25 mmol/L Hepes, and penicillin/streptomycin (7.5 μg/mL). M229, M229R, M249, and M249R were grown in DMEM medium supplemented with glucose (4.5 g/L), 5% FCS, and penicillin/streptomycin (7.5 μg/mL). A375 cells were grown in DMEM medium supplemented with glucose (4.5 g/L), 10% FCS, and gentamycin. HEK293T cells were grown in DMEM medium supplemented with glucose (1 g/L), 10% FCS, and penicillin/streptomycin (7.5 ug/mL). To assess cell growth and viability cells were stained with Trypan Blue (Invitrogen). Vemurafenib (PLX4032), trametinib (GSK1120212), and dabrafenib (GSK2118436) were purchased from Selleckchem. Sk-mel-25, Sk-mel-28, A375, and 501Mel were obtained from ATCC, all other cell lines were gifts from collaborators. All cell lines were regularly tested using the Venor GeM Mycoplasma Detection Kit, and used at less than 10 passages.

GapmeR and siRNA were transfected using Lipofectamine RNAiMAX (Invitrogen) with 20 nmol/L of GapmeR (Qiagen) or siRNA (Thermo Fisher Scientific). GapmeRs and siRNAs sequences are listed in Supplementary Table S1. For combination, GapmeR experiments cells were transfected with 15 nmol/L of LENOX GAP#2 and/or 5 nmol/L SAMMSON Gapmer. For vemurafenib/trametinib+dabrafenib-GapmeR cotreatment, cells were cultured for 3 days in presence or absence (DMSO only) of vemurafenib (1 μmol/L), transfected with 15 nmol/L of control GapmeR, LENOX GAP#2, siRAP2C or siTFAP2A, then cultured for additional 3 days before harvesting. Colony-forming ability was assessed by plating 500 cells/9.6 cm^2, for 10 days and fixing in formalin and staining with 0.05% Crystal Violet solution (Sigma Aldrich).

Melanosphere formation assay

501Mel cells were plated in 10 cm petri dishes without any coating in KO DMEM medium supplemented with 25% KSR, AANE, 2.5 mmol/L Glutamax, 125 μg/mL penicillin/streptomycin, and 50 mmol/L β-mercaptoethanol. Every 3 days pictures of 10 different areas uniformly distributed across the petri were taken by light microscopy. Quantification was performed by ImageJ to calculate the mean and standard deviation for each sample.

Plasmid cloning and lentiviral transduction

GFP-RAP2B was a gift from Philip Stork (Addgene plasmid #118321; http://n2t.net/addgene:118321; RRID:Addgene 118321), whereas pLJC2-RAP2A-3xFLAG was a gift from David Sabatini (Addgene plasmid #87974; http://n2t.net/addgene:87974; RRID: Addgene_87974). LENOX isoforms and RAP2C cDNAs were synthesized by Genscript under the control of a CMV promoter. These plasmids were used for transient transfection of HEKT using PEI (Polysciences, ref. 23966). LENOX and RAP2 cDNAs were cloned into the pCW57-GFP-P2A-MCS vector (a gift from Adam Karpf; Addgene plasmid #71783; http://n2t.net/addgene:71783; RRID: Addgene_71783). A LENOX shRNA (shLENOX) or a scrambled control (shCTRL) were cloned in LT3GEPIR (a gift from Johannes Zuber; Addgene plasmid #111177; http://n2t.net/addgene:111177; RRID:Addgene_111177). Lentiviral particles were produced in HEK293T cells, purified by ultracentrifugation, and resuspended in PBS. After titration, melanoma cells were infected at MOI of 1 and selected by puromycin addition to the media (1 μg/mL).

CRISPR interference

501Mel cells were cotransfected with plasmid expressing dead Cas9 protein fused to the Kruppel-associated box (KRAB) domain-containing KAP1 (dCas9-KAP1) and the red fluorescent protein mScarlet (pX-dCas9-KRAB-Scarlet), together with a plasmid expressing GFP and three single guide RNAs targeting the transcription start site of LENOX (pcDNA3-sgRNA-GFP) or a control plasmid expressing GFP only (pCMV-GFP). Double Scarlet-GFP positive cells were sorted 24 hours after cotransfection, stained with Cell Trace Violet and cultured for additional 96 hours.

A375 xenograft model and bioluminescent imaging

Swiss nude mice were purchased from Charles River Laboratories (France) and housed under specific pathogen-free conditions. Animal care, use, and experimental procedures were conducted in accordance with recommendations of the European Community (86/609/EEC) and Union (2010/63/UE) and the French National Committee (87/848). The ethics committee of IGBCM in compliance with institutional guidelines approved animal care and use (APAFIS#31519–2021051708529028). A375 cells expressing the Dox-inducible shLENOX or scrambled control (shCTRL) were transduced with a lentiviral vector containing the Firefly luciferase gene (kind gift of Catherine-Laure Tomasetto, IGBMC). Mice were injected on the rear flank with 5×10^6 cells resuspended in 100 μL of PBS+ Cultrex Basement Membrane Extract (ref. 3432–005–01; R&D Systems). After 4 days, shRNAs and GFP were induced by Dox administration (ref. D9891; Sigma Aldrich) in drinking water supplemented with 5% sucrose. Tumor growth was monitored by caliper measurement every 2 days and volume was calculated with the formula: $4/3 \times \pi \times \text{length}/2 \times \text{width}/2 \times h/2$. After sacrifice, primary tumors were dissected and dissociated as single cells using the MACS Tumor Dissociation Kit (ref. 130–095–929; Milteny Biotech). Cells were stained with Zombie violet (ref. 423113; BioLegend) and with an alexa-647 anti-H2kq antibody (ref. 115106; BioLegend) to identify dead and murine cells, respectively. Bioluminescence imaging was performed at 7 and 14 days after injection. Prior to imaging, mice were injected intraperitoneally with 100 μL of Xenolight D-luciferin potassium salt (15 mg/mL, #122799; Perkin Elmer). Mice were anesthetized with 5% isoflurane and transported in a sterile cage inside the IVIS Spectrum Imager (Perkin Elmer). Bioluminescence acquisition was performed in auto-mode and expressed as radiance (photon/second).

Patient-derived xenografts

The cutaneous melanoma MEL006res is part of Trace collection (https://gbiomed.kuleuven.be/english/research/50488876/54502087/Trace). MEL006res was derived from the $BRAF^{V600E}$ MEL006 PDX originally sensitive to the dabrafenib–trametinib combination (DT) and rendered resistant by continuous DT treatment (23). Written informed consent was obtained from the patient and all procedures involving human samples were approved by the UZ Leuven/KU Leuven Medical Ethical Committee (S63799) and carried out in accordance with the principles of the Declaration of Helsinki and with GDPR regulations. The experiments were approved by the KU Leuven animal ethical committee under ECD P164–2019 and performed in accordance with the internal, national, and European guidelines of Animal Care and Use. Mice were maintained in IVC cages in a semi-pathogen-free facility under standard housing conditions with continuous access to food and water. The KU Leuven animal facilities comply with all appropriate standards [cages, space per animal, temperature (22°C), light, humidity, food, water], and all cages are enriched with nesting materials. Tumor pieces were implanted subcutaneously in the hip of female NMRI nude BomTac:NMRI-Foxn1nu, 10-week-old females (Taconic Biosciences). Mice were engrafted with PDX MEL006res and drug naïve tumors were grown to 100 mm^3 before administration of antisense oligonucleotides (ASO; 15 mg/kg) every second day by subcutaneous injection in the back. A second cohort was treated daily with DT by oral gavage with 30 mg and 0.3 mg/kg DT, respectively, was started when the tumors became palpable. Once tumors reached 200 mm^3, the mice were randomly assigned to the different cohorts and treated with daily with ASO as described above.

Analysis of oxygen consumption rate in living cells

Oxygen consumption rate (OCR) was measured in an XF96 extracellular analyzer (Seahorse Bioscience). 20,000 transfected cells per well were seeded 48 hours prior the experiment. The cells were incubated at 37°C and the medium was changed to XF base medium supplemented with 1 mmol/L pyruvate, 2 mmol/L glutamine, and 10 mmol/L glucose for 1 hour before OCR profiling with the Mitostress Test Kit sequentially exposed to 2 μmol/L oligomycin, 1 μmol/L carbonyl cyanide-p-trifluoromethoxyphenylhydrazone (FCCP), and 0.5 μmol/L rotenone and antimycin A. Cells were washed with PBS, fixed with 3% PFA, permeabilized with 0.2% triton. Nuclei were counterstained with DAPI (1:500) and number of cells per well determined.

Proliferation and viability analyses by flow cytometry

To assess cell viability and proliferation, cells were stained with Cell Trace Violet (Invitrogen) on the day of transfection harvested after 72 hours and stained with Annexin V (BioLegend) and TOPRO-3 (Invitrogen) or the active caspase-3 Kit (BD Biosciences). Cells were analyzed on a LSRII Fortessa (BD Biosciences) and data were analyzed with FlowJo software (TreeStar). To assess cytochrome C content, 501Mel and A375 cells were transfected with the indicated GapmeRs or siRAP2C, permeabilized with digitonin (50 μg/mL) for 5 minutes on ice, fixed in 4% PFA for 20 minutes at room temperature, and incubated in blocking buffer (3% BSA, 0.05% saponin in PBS) for 1 hour. Cells were stained overnight at 4°C with an anti-cytochrome C mAb coupled to Alexa Fluor 647 (BioLegend, #612310). The day after, cells were stained with Zombie Violet and the anti-active caspase-3 antibody for 30 minutes at room temperature and analyzed by flow cytometry as above.

Cell-cycle progression analysis by flow cytometry

Cells were stained using the Click-it Edu Kit (Thermo Fisher Scientific) and TOPRO-3. Briefly, cells were cultured with 10 μmol/L Edu for 1.5 hours, harvested by trypsinization, washed once in 1% BSA-PBS, and fixed in 4% PFA for 15 minutes at room temperature. After a wash in 1% BSA-PBS cells were permeabilized with saponin-permeabilization buffer for 15 minutes and stained with the Click-it reaction cocktail for 30 minutes. Finally, they were washed once in 1% BSA-PBS, resuspended in 500 μL of PBS with 10 nmol/L TOPRO-3, and left for 10 minutes at room temperature.

Intracellular reactive oxygen species analysis by flow cytometry

Cells were stained in adherent conditions with CellRox Deep Red (Thermo Fisher Scientific) at final concentration of 500 nmol/L following manufacturer instructions. After harvesting, cells were stained for active caspase-3 (BD Biosciences) and analyzed on a LSRII Fortessa (BD Biosciences). To induce reactive oxygen species (ROS), cells were treated with THBP (200 μmol/L) for 30 minutes, and to inhibit ROS induction, cells were treated with NAC (1 μmol/L) for 1 hour before THBP administration. To induce apoptosis, cells were treated with staurosporine (500 nmol/L) for 16 hours.

MitoTracker analysis by flow cytometry

Cells were stained with Annexin V and after one wash in Annexin V binding buffer and PBS, diluted in PBS+5% FCS and stained with MitoTracker CMXRos Red (200 nmol/L) for 25 minutes at 37°C. After one wash in PBS, cells were stained with TOPRO-3 for 10 minutes at room temperature. For FCCP-treated samples, cells were incubated with 50 μmol/L FCCP for 30 minutes before MitoTracker staining. Samples were analyzed on a LSRII Fortessa (BD Biosciences).

RNAscope

RNAs for LENOX, MITF, and SOX10 were detected with the RNAscope assay (Advanced Cell Diagnostics, ACD) according to the manufacturer's protocols. Patient sections were deparaffinized, incubated with hydrogen peroxide at room temperature for 10 minutes, boiled with target retrieval reagent for 15 minutes, and then treated with protease plus reagent at 40°C for 30 minutes. Sections were hybridized with Hs-MITF probe (ACD, catalog no. 310951), Hs-SOX10 probe (ACD, catalog no. 484121), at 40°C for 2 hours. Probes for Hs-LENOX and were custom designed by ACD. Hybridization signals were amplified and visualized with RNAscope Multiplex Fluorescent Reagent Kit v2 (ACD, catalog no. 323100). For co-detection of RAP2 and p32 with LENOX, cells were fixed for 30 minutes with formaldehyde 3.7%, washed with PBS and incubated 10 minutes at room temperature with H2O2. After one wash in distilled water, primary antibodies diluted in co-detection diluent (1/100 for RAP2, 1/200 for p32) were added o/n at 4°C. Slides were washed in PBS+tween 0.1% (PBST), fixed in formaldehyde 3.7% for 30 minutes, and washed again in PBST. Slides were treated with protease III and washed with PBS. LENOX hybridization signals were amplified following the Multiplex Fluorescent Kit. Finally, RAP2 and p32 signals were developed by secondary antibodies incubation (diluted 1/2,000 in co-detection diluent), followed by tyramide signal amplification (TSA Plus Kit, NEL760001KT, Perkin Elmer). Images were captured with a confocal (Leica DMI6000) microscope.

Immunofluorescence of fixed cells

Cells grown on glass slides in 24-well plates, were fixed with 4% paraformaldheyde for 15 minutes. After two washes with PBS buffer they were permeabilized in PBS+triton X-100 0.1% for 5 minutes and blocked with PBS+10% FCS inactivated for 20 minutes. Primary antibodies were incubated overnight at 4°C and after three washes with PBS+Triton 0.1%, cells were stained for 1 hour at room

temperature with Alexa Fluor-488 conjugated secondary antibodies (Life technologies) diluted 1/500 in PBS+10% FCS. After three washes with PBS+Triton 0.1%, cells were stained with DAPI (final concentration 1 μg/mL) and mounted on microscopy slides. Anti-TP53BP1 (NB100–304), anti-RAP2 (Santa Cruz Biotechnology, sc-515711), and anti-p32 (Bethyl, A302–863A) were diluted 1/200 in PBS+10% FCS; anti-γH2AX (Ser139, Abcam, ab11174) 1/400. Images were captured with a confocal (Leica DMI6000) microscope. RAP2-p32 colocalization and the number of gH2Ax or TP53BP1 positive nuclear foci were quantified using ImageJ. Mander's correlation coefficient of RAP2 over p32 was calculated using JACoP.

MitoTracker live imaging

Cells were cultured in 4 wells 35 × 10 mm dishes (CellView, Greiner Bio-one), stained for 1 hour with MitoTracker Red CMXRos (125 nmol/L) and Hoechst 33342 (1 μg/mL) and z-stack images acquired on a confocal (Leica DMI6000) microscope in a temperature controlled (37°C) chamber. Mitochondrial shape and network analyses were performed on ImageJ as described previously (24, 25). Briefly, projections of multiple z-stack sections into one image were performed by generating a maximum intensity composite and preprocessed using "subtract background" (radium 1um), "sigma filter plus" (radius 0.1 μm, 2.0 sigma), "enhance local contrast/CLAHE" (block size 64, slope 2.0), "gamma correction" (0.8), and tubeness (sigma 0.361). The adaptive threshold plugin (https://sites.google.com/site/qingzongtseng/adaptivethreshold#use) was used to identify mitochondria and the image post-processed using "despeckle." The resulting binary image was used as the input for the "analyze particles" command, measuring for area, perimeter, and shape descriptors. Form factor (FF) was derived as the inverse of the circularity (26). For network connectivity analysis, the "skeletonize 2D/3D" command was applied to produce a skeleton map and the "analyze skeleton" command calculated the number of branches and branch junctions in the network.

RNA extraction and qPCR

Total mRNA isolation was performed using TRIzol and isopropanol precipitation. Isolation of cytosolic, nuclear soluble, and chromatin-associated RNA was performed as described in ref. 27. Briefly, cells were harvested and washed in PBS buffer, resuspended in 0.15% NP-40 lysis buffer, and centrifuged on a 24% sucrose cushion taking supernatant as the cytosolic fraction. Nuclei were resuspended in 1M urea, 1% NP-40 lysis buffer, and centrifuged to recover the nuclear soluble fraction in the supernatant. The chromatin pellet was finally resuspended in 1 mL of TRIzol, solubilized using a 21-gauge needle, and isolated following manufacturer instructions. Cyosolic and nuclear soluble fractions were cleared by centrifugation and RNA was isolated from 200 μL of each using 1 mL of TRIzol. Total and fractionated RNAs were treated with DNAseI following the TurboDnase Free Kit instructions (Thermo Fisher Scientific) and reversed transcribed using Superscript IV reverse transcriptase (Thermo Fisher Scientific) following manufacturer instructions. qRT-PCR was carried out with SYBR Green I (Roche) and monitored by a LightCycler 480 (Roche). Target gene expression was normalized using TBP, HBMS, GAPDH, ACTB, RPL13A as reference genes. Primers for RT-qPCR were designed using Primer3 and listed in the Supplementary Table S2.

Protein extraction and Western blotting

Whole cell extracts were prepared by freeze–thaw technique using LSDB 500 buffer [500 mmol/L KCl, 25 mmol/L Tris at pH 7.9, 10% glycerol (v/v), 0.05% NP-40 (v/v), 16 mmol/L DTT, and protease inhibitor cocktail]. Lysates were subjected to SDS-PAGE and proteins were transferred onto a nitrocellulose membrane. Membranes were incubated with primary antibodies in TBS+ 5% BSA + 0.01% Tween-20. Overnight at 4°C. The membrane was then incubated with HRP-conjugated secondary antibody (Jackson ImmunoResearch) for 1 hour at room temperature, and visualized using the ECL detection system (GE Healthcare). Antibodies used are listed in Supplementary Table S3.

Mitochondria fractionation

Mitochondria were isolated with the Mitochondria Isolation Kit (Thermo Fisher Scientific) following manufacturer instructions. Harvested cells were washed and pelleted, resuspended in buffer A, and incubated 2 minutes on ice. Buffer B was added for 5 minutes, vortexing every minute, and diluted with buffer C. Nuclei were pelleted 10 minutes at 700 × *g* and supernatant centrifuged for 15 minutes at 3,000 × *g*. Purified mitochondria were washed once in buffer C and used for RNA (TRIzol–isopropanol precipitation) or protein (TBS+CHAPS 2%) extraction.

LENOX pulldown and LC/MS-MS analysis

MM011 cells were grown in 15 cm petri dishes, harvested by trypsinization, washed, pelleted, resuspended in lysis buffer (TrisHCl 20 mmol/L pH8, NaCl 200 mmol/L, $MgCl_2$ 2.5 mmol/L, Triton 0.05%, DEPC water) supplemented with fresh DTT (1 mmol/L), protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific) and RNAsin (Thermo Fisher Scientific) and kept 20 minutes on ice. Membranes were pelleted 3,000 × *g* for 3 minutes at 4°C and supernatant precleared for 1 hour at 4°C with streptavidin-coated sepharose beads. The lysate was incubated 2 hours with streptavidin coated beads and 400 pmol anti-PCA3 or LENOX-specific DNA biotinylated oligonucleotides (listed in Supplementary Table S4). Beads were pelleted for 3 minutes at 3,000 × *g* and washed three times with lysis buffer. After final wash beads were divided for RNA and protein extraction. RNA was purified by TRIzol and isopropanol precipitation, digested with DNAse, reverse transcribed and analyzed by qPCR for LENOX and TINCR. Proteins were eluted by boiling beads in Laemli sample buffer and separated on NuPAGE Novex 4% to 12% gradient gels. Three independent experiments were performed and the entire lane was excised in seven consecutive bands and subjected to "in-gel" digestion. Proteins were reduced in 10 mmol/L DTT for 1 hour at 56ºC and alkylated with 55 mmol/L iodoacetamide for 45 minutes at room temperature. Enzymatic digestion was performed using 12.5 ng/μL trypsin overnight at 37ºC. Tryptic peptides were extracted from the gel with 3% trifluoroacetic acid (TFA) and 30% acetonitrile (ACN). The extracted peptides were concentrated onto homemade StageTips reversed phase microcolumns. Peptides were eluted in 40 μL buffer B [80% ACN, 0.1% formic acid (FA)]. ACN was evaporated using a vacuum concentrator (Speed Vac, Eppendorf) and the volume of the eluates were adjusted to 5 μL with 1% TFA for peptide separation and analysis in a reversed-phase nano-flow liquid chromatographic (nRP-LC) column using an EASY-nLC 1200 (Thermo Fisher Scientific) coupled to an Q-Exactive HF instrument (Thermo Fisher Scientific) through a nano-electrospray ion source (EASY-Spray; Thermo Fisher Scientific). The nRP-LC system was operated in one column set-up with an EasySpray PEPMAP RSLC C18 (Thermo Fisher Scientific) kept at 45°C constant. Solvent A was 0.1% FA and solvent B was 0.1% FA in 80% ACN. Samples were loaded in aqueous 0.1% (FA) solution at constant pressure (980 Bar). Peptides were separated with a gradient of 3% to 30% solvent B over 69 minutes followed by a gradient of 30% to 60% for 5 minutes and 60% to 95% over 5 minutes at a flow rate of

300 nL/min. The Q-Exactive was operated in the data-dependent acquisition (DDA) mode and MS spectra (from m/z 375–1,550) were analyzed in the Orbitrap detector with resolution $R = 60{,}000$ at m/z 200. The fifteen most intense peptide ions were isolated to a target value of 3e6 and fragmented by Higher Energy Collision Dissociation (HCD) with a normalized collision energy (NCE) setting of 28. The maximum allowed ion accumulation times was 80 milliseconds for MS-MS. The dynamic exclusion time was set to 20 seconds.

Post-acquisition MS data analysis for proteomics

Acquired raw data were analyzed with the integrated suite of algorithms MaxQuant, version 1.6.1.1, using the Andromeda search engine. FDR for both peptides and protein identifications was set to a maximum of 0.01. Carbamidomethylation of cysteine was set as a fixed modification. Uniprot Human sequence database was used for peptide identification (74470 Entries). LFQ intensity calculation was enabled requiring a minimum LFQ ratio count equal to two. The match between runs (MBR) feature was selected and a tolerance of 0.7 minutes was specified for the match time window option. The "protein groups" (.txt) output file from MaxQuant was processed by Perseus software for statistics. Briefly, no imputation for missing values was used, and the data were filtered, to have 3 valid values in at least one group. A t Student test was used to compare protein co-enriched in the LENOX- versus the PCA3- pulldown experiments and the threshold settings to select significant enriched proteins were S0 = 0 and FDR = 0.05.

Immunoprecipitation

Cells were grown in 15 cm petri dishes, harvested by scraping, resuspended in lysis buffer (20 mmol/L Tris-HCl pH 8, 200 mmol/L NaCl, 2.5 mmol/L $MgCl_2$, 0.05% Triton, DEPC water) supplemented with DTT (1 mmol/L), protease/phosphatase inhibitor cocktail (Thermo Fisher Scientific) and RNAsin (Thermo Fisher Scientific) and kept on ice for 15 minutes, pipetting every 3 minutes. Membranes were pelleted 10 minutes at 10,000 × g at 4°C and the supernatant precleared 1 hour at 4°C with protein G magnetic beads. Lysate was quantified by Bradford protein quantification assay (Bio-Rad) and incubated overnight at 4°C with indicated antibodies. Protein G magnetic beads were added for 3 hours at 4°C to isolate RNA–protein complexes and washed five times in lysis buffer. After final wash RNA was purified by TRIzol and isopropanol precipitation and proteins eluted by boiling beads in Laemli sample buffer.

Proximity ligation assay

The PLA Duolink FarRed protocol (Sigma Aldrich, ref. DUO92008) was performed following manufacturer's instructions. Primary antibodies recognizing RAP2 (sc-515711) and DRP1 (12957–1-AP) were used at a concentration of 1:200. PLA Probe Anti-Rabbit PLUS and PLA Probe Anti-Mouse MINUS were used.

Statistical analysis

Comparison between experimental groups were mainly performed using one-way ANOVA, Dunnett, or Tukey multiple comparison test, as indicated in legends. Other statistical tests are indicated in the appropriate legends (*, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$).

Resources

All oligonucleotides and antibodies used are listed in Supplementary Tables S1 to S4.

Data availability

The datasets generated during and/or analyzed in this study are available from the corresponding authors E. Leucci or I. Davidson upon request.

Results

LENOX is expressed in all known melanoma states and is associated with melanoma progression

To identify melanoma-specific genes common to all cellular states, we re-analyzed singe cell (sc)RNA-seq of a melanoma PDX undergoing targeted MAP kinase inhibition (MAPKi) therapy (4). We selected the 50 mRNAs and lncRNAs with the lowest coefficient of variation across all melanoma cell states (**Fig. 1A**). Protein coding genes associated with housekeeping functions were expressed in the different melanoma states and across normal tissues. However, LINC00518 (hereafter LENOX) was highly expressed in human melanoma and across melanoma cell states, but poorly in normal tissues (**Fig. 1A** and **B**).

In normal tissues LENOX expression was restricted to sun-exposed and nonexposed skin and testis (Supplementary Fig. S1A), but it was strongly and specifically expressed in cutaneous and uveal melanoma (Supplementary Fig. S1B). LENOX expression in MITF-expressing melanocytes was confirmed by RNAscope on sections from normal skin (**Fig. 1C**) and in MITF-expressing cells of primary melanoma (**Fig. 1C**). LENOX was expressed in melanoma cells, but not in infiltrating immune or stromal cells (Supplementary Fig. S1C). LENOX expression was upregulated in melanoma compared with nevi (Supplementary Fig. 1D; refs. 28, 29) and increased with Clark score (Supplementary Fig. S1E) and increased copy number due to its coamplification with TFAP2A in around 6.5% to 8% of melanomas (Supplementary Fig. S1F). High LENOX expression was associated with decreased survival of patients with melanoma (Supplementary Fig. S1G). RT-qPCR on a collection of melanocytic and undifferentiated lines confirmed that LENOX was consistently expressed irrespective of cell state and driver mutation (**Fig. 1D**).

The LENOX promoter comprises a MLT1B ERVL-MaLR retroviral long terminal repeat conferring primate specificity analogous to SAMMSON containing the LTR1 ERV1 element (30). Juxtaposition of the MLT1B element to a SINE element (MIRb) created a SOX10 binding site and prominent SOX10 binding was observed with that of MITF, BRG1, and the H3K27ac mark (Supplementary Fig. S1H). Accordingly, LENOX was regulated by both MITF and SOX10 in 501Mel cells (Supplementary Fig. S1I; ref. 10]. However, LENOX expression across cell states suggested additional modes of regulation. Public ChIP-seq data from melanocytic and undifferentiated melanoma cells identified several potential enhancer elements (EN 1–6) in the TFAP2A/LENOX locus (Supplementary Fig. S1H; refs. 10, 31, 32). For example, EN1 was marked by H3K27ac, H3K4me1, TFAP2A, and FOSL2, whereas the LENOX promoter showed SOX10 binding and H3K4me3. EN1 and EN5 showed binding of TFAP2A that is expressed in drug-tolerant states (**Fig. 1B**). siTFAP2A silencing diminished LENOX expression in 501Mel and MM047 lines, but had little effect on SAMMSON (Supplementary Fig. S1J). Furthermore, LENOX and TFAP2A expression were positively correlated in TCGA and in melanoma cell lines (Supplementary Fig. S1K; refs. 2, 33). In contrast, LENOX silencing (see below) did not affect TFAP2A expression (Supplementary Fig. S1L). Combinatorial regulation by MITF, SOX10, and TFAP2A may account for LENOX expression across melanoma cell states in accordance with the observation that LENOX and TFAP2A are part of a gene signature discriminating melanomas from

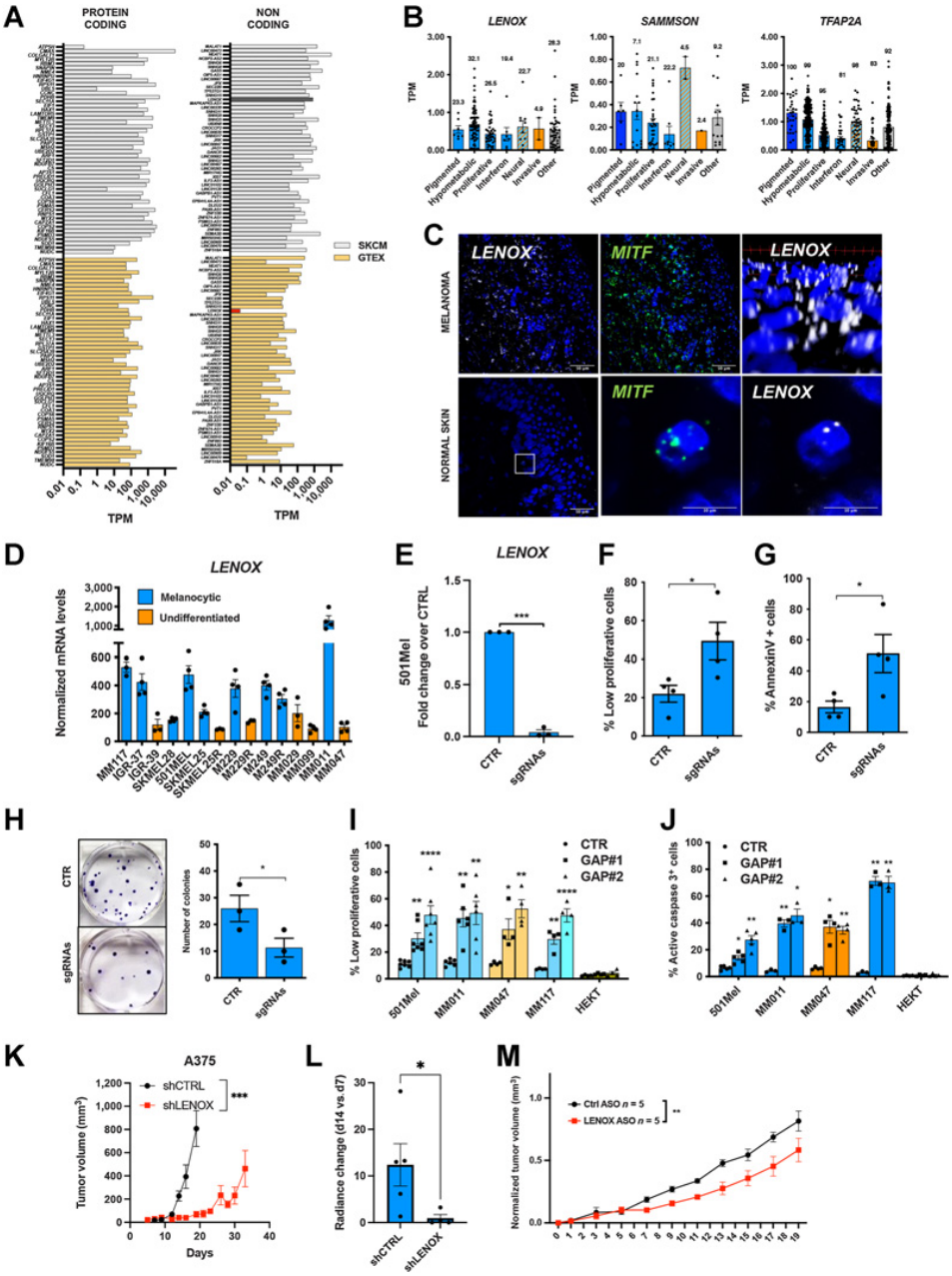


Figure 1.

LENOX is expressed in normal skin melanocytes and melanoma. **A,** Top 50 protein-coding and noncoding transcripts based on expression across melanoma states (GSE116237) and in GTEx and SKCM TCGA. LENOX is highlighted by dark bars. **B,** Expression of LENOX, SAMMSON, and TFAP2A in melanoma subpopulations. Only cells where the RNAs were captured were included in the analysis. The percentage of cells analyzed for each population is indicated. **C,** RNAscope detection of LENOX and MITF in normal skin and melanoma FFPE sections. **D,** LENOX expression in melanocytic (blue) and undifferentiated (orange) cell lines. **E,** LENOX expression in 501Mel cells expressing the dCas9-KAP1 protein with control or LENOX promoter-targeting sgRNAs. **F–H,** Proliferation, apoptosis, and colony formation following dCas9-KAP1-mediated LENOX silencing compared with negative control by one-way ANOVA (Dunnett test). **I** and **J,** Proliferation and apoptosis of LENOX or control GapmeR-transfected cells compared by one-way ANOVA (Dunnett test). **K,** Tumor growth of A375 cells expressing shCTRL or shLENOX ($N = 5$). **L,** Bioluminescence emission was measured on day 7 and 14 using the IVIS Spectrum Image and compared by Mann–Whitney test. **M,** Growth of MEL006res PDX treated with control or LENOX targeting ASO. Tumor size was measured daily for 20 days. Statistical analysis were calculated using two-way ANOVA. *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

other tumors and positively correlating with disease aggressiveness (designated C6ORF218 in ref. 34).

LENOX is required for melanoma cell proliferation and survival

To silence LENOX, we transfected melanoma cells with a vector encoding the CRISPR/dCAS9-KAP1 fusion protein with sgRNAs specific for the LENOX promoter that reduced its expression by more than 90% compared with control (**Fig. 1E**), resulting in an increase in slow proliferating cells (**Fig. 1F**), in Annexin V–positive apoptotic cells (**Fig. 1G**) and decreased colony formation (**Fig. 1H**).

We designed two independent locked nucleic acid ASOs (LNA-GapmeRs) whose transfection reduced LENOX expression by over 80% compared with a nontargeting control (CTR; Supplementary Fig. S2A). Reduced cell numbers were detected 72 hours after LENOX silencing in melanocytic and undifferentiated lines (Supplementary Fig. S2B), but not in HEK293T cells where LENOX was not expressed. GapmeR-mediated silencing resulted in strongly reduced cell proliferation (**Fig. 1I**) and apoptosis (**Fig. 1J**; Supplementary Fig. S2C) with early and late apoptotic cells observed in flow cytometry using Annexin V labeling (Supplementary Figs. S2D and S2E). We also silenced LENOX using a doxycycline (Dox)-inducible shRNA coupled to GFP for monitoring by flow cytometry (Supplementary Fig. S3A). LENOX was strongly decreased in 501Mel and A375 cells, with reduced proliferation (Supplementary Figs. S3B and S3C).

A375 cells were injected subcutaneously in immunodeficient mice to form tumors and Dox-induced GFP confirmed in tumor cells (Supplementary Fig. S3D). Cells expressing LENOX shRNA showed strongly impaired tumor growth confirmed by bioluminescence measurements (**Fig. 1K** and **L**; Supplementary Figs. S3E and S3F). Tumors were significantly smaller and in two cases regressed. We also targeted LENOX in mice engrafted with the $BRAF^{V600E}$ cutaneous melanoma PDX MEL006res (23). In drug naïve conditions, the LENOX GAP#2 ASO significantly reduced LENOX expression and PDX growth compared with control GAP ASO (**Fig. 1M**; Supplementary Fig. S3G). All three targeting strategies therefore supported the essential role of LENOX in melanocytic, neural crest, and undifferentiated melanoma cells irrespective of driver mutation.

Databases predict several alternatively spliced LENOX isoforms sharing a common last exon with three potential polyadenylation sites mapped by 3′RACE as present in all isoforms (Supplementary Fig. S4A). Exon–exon junctions were confirmed on 501Mel RNA-seq data. Isoforms 1, 3, and 5 were most abundantly expressed in human melanomas and in melanoma cell lines (Supplementary Fig. S4B). We generated 501Mel cells with Dox-inducible expression of isoforms 1, 2 or 3, or GFP as control showing a time-dependent increase of their expression (Supplementary Figs. S4C and S4D). Dox treatment increased proliferation of LENOX overexpressing cells, their colony formation and growth as 3D melanospheres (Supplementary Figs. S4E–S4G). Sustained LENOX expression was confirmed in spheroids after 12 days of culture (Supplementary Fig. S4H). Analogous observations were made in A375 cells (Supplementary Figs. S4I and S4J). Thus, although LENOX silencing compromised melanoma cell proliferation and survival, its ectopic expression promoted growth under 2D and 3D conditions.

LENOX interacts with the RAP2 GTPases

RNAscope showed that LENOX was predominantly expressed in the cytoplasm around the nuclear periphery in melanoma cells *in vivo* (**Fig. 1C**) and in cultured melanoma cells confirmed by RT-qPCR on RNA from different cell compartments (Supplementary Figs. S5A and S5B). Reconstitution of 3D cellular images showed LENOX enrichment in the cytoplasm (**Fig. 1C**; Supplementary Fig. S5A).

To identify LENOX interacting proteins, we performed pulldown from cytoplasmic extracts of MM011 cells using a tiling array of LENOX-complementary biotinylated oligonucleotides, or negative control prostate cancer lincRNA PCA3 oligonucleotides followed by MS. LENOX was enriched using its cognate oligonucleotides, but not those of the PCA3 control, whereas lincRNA TINCR was not enriched under any conditions (**Fig. 2A**). Proteins found uniquely in the LENOX pulldown included several ribosomal proteins or endoplasmic reticulum, Golgi, or mitochondrial proteins (**Fig. 2B**, top). To assess their interaction with LENOX, we immunoprecipitated (IP) RAP2, SURF4, SAR1B, and NDUFA6. However, only the antibody against RAP2 efficiently precipitated its target protein from cell extracts (**Fig. 2B**, bottom; Supplementary Fig. S5C). RAP2 IP fractions showed strong enrichment of LENOX, but not SAMMSON, MALAT, or NEAT1. As additional controls, we performed IP of XRN2 that enriched SAMMSON and NEAT1, but not LENOX (Supplementary Fig. S5C).

The RAP2 small GTPases are encoded by 3 paralogous genes *RAP2A*, *RAP2B*, and *RAP2C* with high amino acid identity, but distinguished by a short hypervariable C-terminal region (**Fig. 2C**). Consequently, the RAP2 antibody recognizes all three paralogs. Each paralog was expressed across melanoma cell states, but also in stromal and immune cells with RAP2A showing the lowest expression in melanoma patients (Supplementary Figs. S5D–S5G).

The 501Mel, A375, and MM047 cells showed variable levels of each paralog (**Fig. 2D**). Silencing of each with specific siRNAs (**Fig. 2E** and **F**) revealed that only silencing of RAP2C induced slow growth and apoptosis thus phenocopying LENOX silencing (**Fig. 2G**; Supplementary Figs. S5H–S5J). We generated cell lines with Dox-inducible expression of FLAG-tagged RAP2A, B, or C (Supplementary Fig. S5K). RAP2C protein accumulated to higher levels than RAP2A or B. FLAG-IP of each recombinant protein or pan-RAP2 IP of endogenous RAP2 from the GFP-control cells showed that LENOX, but not SAMMSON, was enriched with all RAP2 paralogues proportionate to their variable expression levels (Supplementary Fig. S5L). Thus, although LENOX interacted with all RAP2 paralogs, RAP2C was the biologically relevant form required for proliferation and survival.

LENOX and RAP2C promote oxidative phosphorylation

Coupling RNAscope with immune-staining showing overlapping localized cytoplasmic RAP2 and LENOX (**Fig. 3A**). RAP2 and LENOX labeling tightly coincided with the mitochondrial p32 protein, and RAP2-p32 colocalization was further seen in primary melanoma (**Fig. 3B**). Purification of mitochondrial and cytoplasmic fractions showed the presence of 16S mitochondrial ribosomal RNA almost exclusively in the mitochondrial fraction, whereas LENOX and SAMMSON were present in both the cytoplasmic and mitochondrial fractions (**Fig. 3C**; Supplementary Fig. S6A). Mitochondrial purity was confirmed by strong enrichment of p32 and HSP60 in the mitochondrial fraction where RAP2 was also enriched even in HEK293T cells (**Fig. 3C**).

Although RAP2 and p32 colocalized in cells transfected with control GapmeR, LENOX silencing resulted in a 5-fold decrease of RAP2-p32 colocalization with RAP2 signal seen in the surrounding cytoplasm (**Fig. 3A**; Supplementary Fig. S6B). This was not due to decreased RAP2 levels that were unaffected by LENOX silencing (Supplementary Figs. S6C and S6D). ShLENOX knockdown strongly

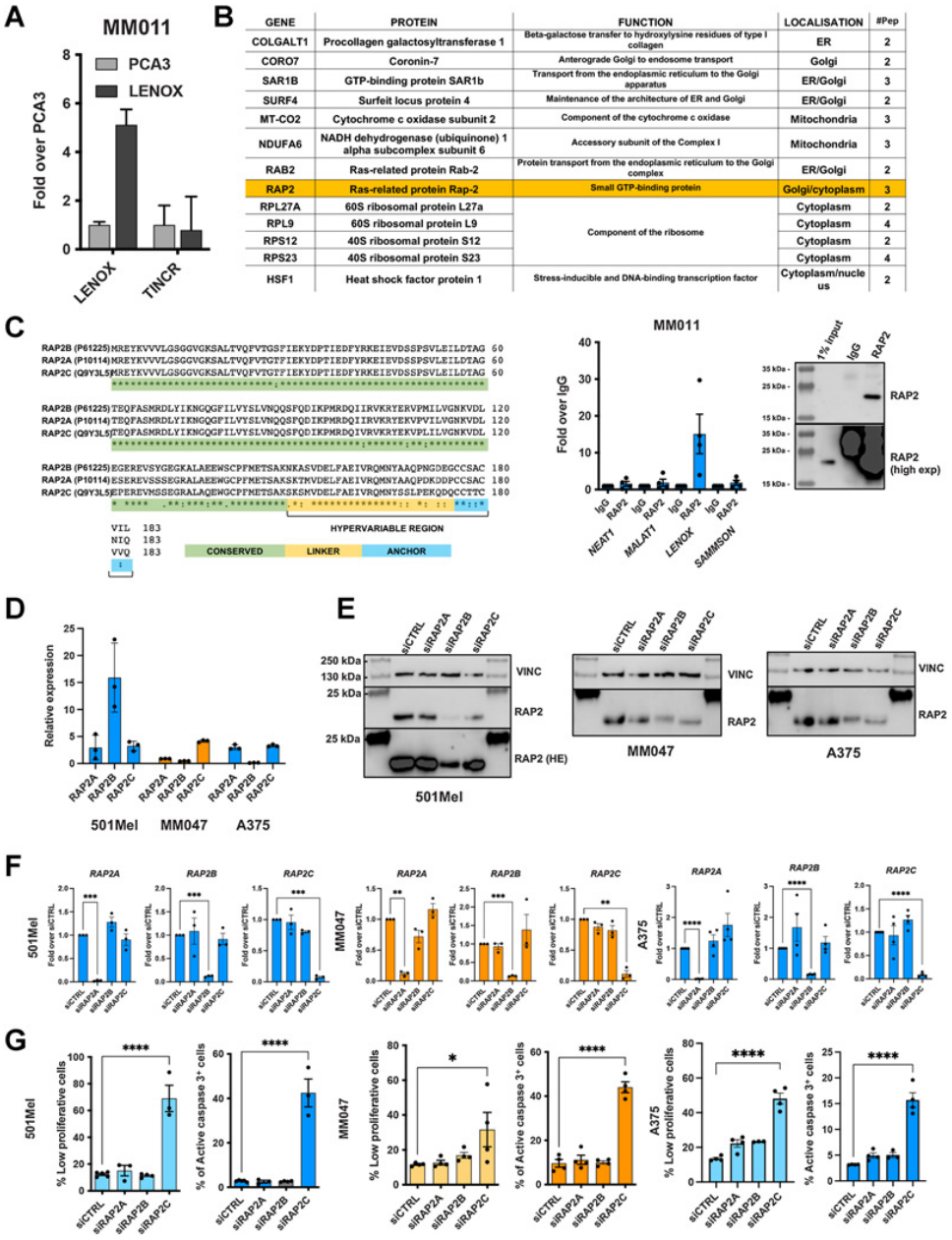


Figure 2.

LENOX interacts with the RAP2 GTPases. **A,** Enrichment of LENOX or TINCR after oligonucleotide-mediated pulldown from MM011 cells. **B,** Top, proteins detected by mass spectrometry only in the LENOX pulldown from MM011 cells. Bottom left, selective enrichment of LENOX following RAP2 IP. Bottom right, IP of RAP2 with 1% input showing short and long exposures. **C,** RAP2A, B, and C protein sequences illustrating amino acid homology. **D,** RAP2A/B/C expression in indicated cells normalized over housekeeping genes. **E** and **F,** RAP2A/B/C expression in the above cells after transfection with indicated siRNAs compared with siCtrl by one-way ANOVA (Dunnett test). VINCULIN was used as a loading control in right panel. **G,** Proliferation and apoptosis following silencing with indicated siRNA and comparison with control by one-way ANOVA (Dunnett test). *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

reduced LENOX, but not SAMMSON expression in both the cytoplasm and mitochondria fractions, reduced RAP2-p32 colocalization and reduced RAP2 levels in mitochondria (Supplementary Figs. S6E–S6G).

We investigated the effect of LENOX and RAP2C silencing on mitochondrial activity by profiling OCR as a measure of oxidative phosphorylation (OXPHOS). LENOX or RAP2C silencing did not affect basal OCR, but potently decreased maximal and reserve capacities

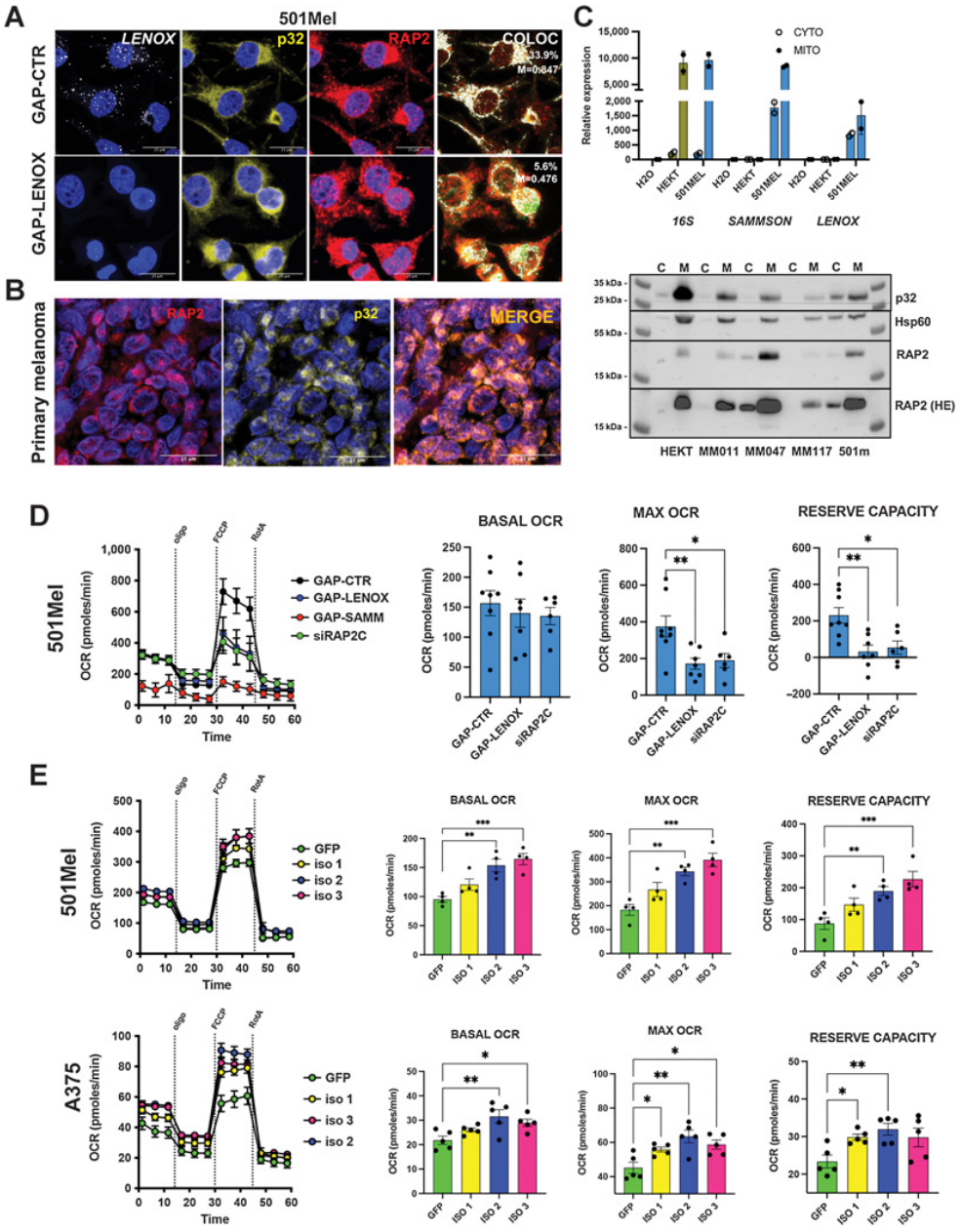


Figure 3.

LENOX and RAP2 modulate oxidative phosphorylation. **A,** Co-imaging of LENOX, p32, and RAP2 in 501Mel cells transfected with the indicated GapmeRs. **B,** RAP2-p32 colocalization in a human melanoma FFPE section. **C,** Levels of 16S rRNA, SAMMSON, and LENOX in mitochondrial and cytoplasmic fractions (top) and levels of RAP2, P32, and HSP60 proteins (bottom) **D**. Mitostress test of 501Mel cells 48 hours after transfection with indicated GapmeRs or siRAP2C. Experimental groups were compared by one-way ANOVA (Dunnett test). **E,** Mito Stress Test on cells with ectopic LENOX isoform expression. Experimental groups were compared as in **D**. *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$.

(**Fig. 3D**). SAMMSON knockdown potently reduced OCR under all conditions. Ectopic LENOX expression increased basal, maximal, and reserve OCR in both 501Mel and A375 cells (**Fig. 3E**).

To link compromised OXPHOS to decreased cell growth and apoptosis, we asked if LENOX or RAP2C silencing and the impaired mitochondrial function and induced ROS. We performed flow cytometry on cells labeled with Annexin V, TOPRO, and MitoTracker Red CMXRos and gated on the nonapoptotic Annexin V/TOPRO-negative cells. Compared with control, LENOX or RAP2C silencing led to the appearance of higher proportions of CMXRos low cells

indicative of reduced mitochondrial polarization that was strongly induced using FCCP as a positive control (Supplementary Fig. S7A).

We next stained LENOX, RAP2C, or control silenced cells with anti-caspase-3 and Cell ROX to distinguish nonapoptotic cells (Q4) from apoptotic cells (Q3) and cells with elevated ROS (Q1; Supplementary Fig. S7B). As positive control, THBP (tert-Butyl hydroperoxide) induced elevated ROS that was suppressed using the ROS scavenger N-acetyl-L-cysteine (NAC), whereas staurosporine induced apoptosis, but not ROS. LENOX or RAP2C silencing induced a large increase in apoptotic cells, but also cells with augmented ROS as well as apoptotic ROS-high cells (Q2). Increased ROS following LENOX or RAP2C silencing was associated with DNA damage observed by increased γH2AX and TP53BP1 foci, not seen following RAP2A or control silencing (Supplementary Figs. S7C and S7D). LENOX or RAP2C silencing also increased the population of cells arrested in G_2–M phase (Supplementary Fig. S7E). We further gated cells negative for caspase-3 and zombie violet to select nonapoptotic cells that were analyzed for cytochrome C content. Compared with controls, LENOX or RAP2 silencing significantly increased the number of cytochrome C low cells, indicating increased cytochrome C leakage may also promote subsequent cell apoptosis (Supplementary Fig. S7F).

LENOX or RAP2C silencing therefore led to mitochondrial depolarization, generating ROS and cytochrome C leakage, inducing DNA damage, cell-cycle block and ultimately apoptosis.

LENOX and RAP2C regulate DRP1 phosphorylation and mitochondrial fission

LENOX or RAP2C silencing did not alter the total number of mitochondria per cell, but led to smaller, shorter, and rounder mitochondria with reduced numbers of branches and junctions (**Fig. 4A**; Supplementary Figs. S8A and S8B). In contrast, ectopic LENOX expression increased mitochondrial size and length, but did not induce a hyperfused phenotype as the numbers of branches and junctions were unaffected (**Fig. 4B**).

Mitochondrial homeostasis is dynamically regulated by a balance between phosphorylation of S616 of the GTPase DRP1 by ERK2 that stimulates mitochondrial fission (35, 36) and phosphorylation at S637 inhibiting DRP1 and promoting mitochondrial fusion (37–39). LENOX or RAP2C silencing reduced DRP1 S637 phosphorylation (**Fig.4C** and **D**), whereas it was increased in cells ectopically expressing LENOX (**Fig. 4E**) consistent with the changes in mitochondrial morphology.

We investigated interactions between DRP1 and RAP2. DRP1 was coprecipitated with RAP2 antibody, but not control IgG, in cells expressing control shRNA, but lost in shLENOX-expressing cells showing their LENOX-dependent interaction (**Fig. 4F**). Accordingly, LENOX, but not SAMMSON, was enriched in both the RAP2 and DRP1 IPs, but not the control IP (**Fig. 4F** and **G**). Similarly, LENOX was enriched in the RAP2 IP in control cells, but not in shLENOX cells, whereas SAMMSON was not enriched under any condition (**Fig. 4F**). In an independent approach, proximity ligation assay with DRP1 and RAP2 antibodies showed their cytoplasmic interaction in control cells that was strongly diminished in the shLENOX cells (**Fig. 4H**). LENOX therefore promotes a RAP2–DRP1 interaction that enhanced S637 phosphorylation, mitochondrial fusion, and optimized OXPHOS.

RAP2 associated with mitochondria in HEK293T cells (**Fig. 3C**) prompting us to investigate RAP2 function in these cells in absence of LENOX. RAP2B and RAP2C knockdown induced the strongest reductions of RAP2 protein levels, however, no effect on proliferation or apoptosis was seen upon knockdown of any RAP2 paralog (Supplementary Figs. S9A and S9B). Accordingly, LENOX-GapmeR or siRAP2C did not induce changes in mitochondrial morphology (Supplementary Fig. S9C). In contrast, ectopic LENOX expression by transient transfection, in particular of isoforms 2 and 3, induced a RAP2-DRP1 interaction not seen in GFP transfected cells (Supplementary Figs. S9D–S9F). Moreover, ectopic LENOX expression led to increased mitochondrial elongation in transfected cells and increased OCR (Supplementary Figs. S9G and S9H).

Hence, despite its mitochondrial association, RAP2 did not regulate OXPHOS and/or cell viability in HEK293T cells. In contrast, ectopic LENOX expression promoted a RAP2-DRP1 interaction, hijacking RAP2 to regulate mitochondrial homeostasis and function.

Cooperativity of LENOX and SAMMSON

Expression of LENOX and SAMMSON across melanoma cell states and their complementary roles on mitochondrial function suggested they may act cooperatively with one another. Compared with individual SAMMSON or LENOX knockdown with suboptimal concentrations of GapmeR for each lincRNA, combinatorial knockdown led to a more potent reduction in 501Mel and MM011 proliferation (**Fig. 5A**) and a strong increase in their apoptosis (**Figs. 5B**). Similarly, combinatorial knockdown in undifferentiated MM047 cells cooperatively induced slow proliferation and induced a potent increase in apoptosis (**Fig. 5A** and **B**). The synergistic effect of combinatorial silencing was evident after 10 days of culture, by which time, essentially all the MM047 cells were eliminated (**Fig. 5C**). LENOX and SAMMSON therefore cooperatively promoted melanoma cell proliferation and survival.

LENOX and RAP2C are required for increased OXPHOS upon BRAF inhibition

As LENOX and RAP2C regulated mitochondrial morphology and OXPHOS under basal growth conditions, we asked if this pathway was also important upon metabolic stress such as suppression of glycolysis by BRAF inhibitors that promote a metabolic switch to OXPHOS (40, 41).

M229 cells treated with BRAF inhibitor vemurafenib (2) showed upregulated LENOX expression after 3 days that returned to basal level at later times. Similarly, TFAP2A, SOX10, and PAX3 were all upregulated during this acute phase, whereas SAMMSON was reduced (Supplementary Fig. S10A). We confirmed upregulated LENOX, TFAP2A, SOX10, and PAX3 expression between 12 and 72 hours in vemurafenib-treated M229, Sk-Mel-25, 501Mel, and A375 cells (Supplementary Fig. S10B). Consistently, LENOX levels in mitochondria increased upon vemurafenib treatment, whereas SAMMSON displayed a progressive reduction (Supplementary Fig. S10C). LENOX, TFAP2A, and SOX10 upregulation was also seen with encorafenib, DT, and confirmed for LENOX in A375 and 501Mel cells (Supplementary Figs. S10D and S10F). Increased LENOX expression after MAPKi was also observed in MEL006 PDX treated with DT and in triple wild-type patients treated with trametinib and durvalumab (Supplementary Figs. S10G and S10H), suggesting that LENOX induction was not restricted to BRAF mutant melanoma.

Upregulation of TFAP2A protein was observed upon vemurafenib, dabrafenib, or trametinib exposure (Supplementary Fig. S10I). Consistently, TFAP2A-bound EN#1 and 5 EN#5 showed increased H3K27ac levels in vemurafenib-treated A375 cells (Supplementary Fig. S10J). The LENOX promoter showed an analogous increase and stronger BRD4 recruitment that was abrogated by concomitant THZ1 treatment. RNA-seq data confirmed that upregulation of LENOX, TFAP2A, and SOX10 was also inhibited by THZ1 (Supplementary

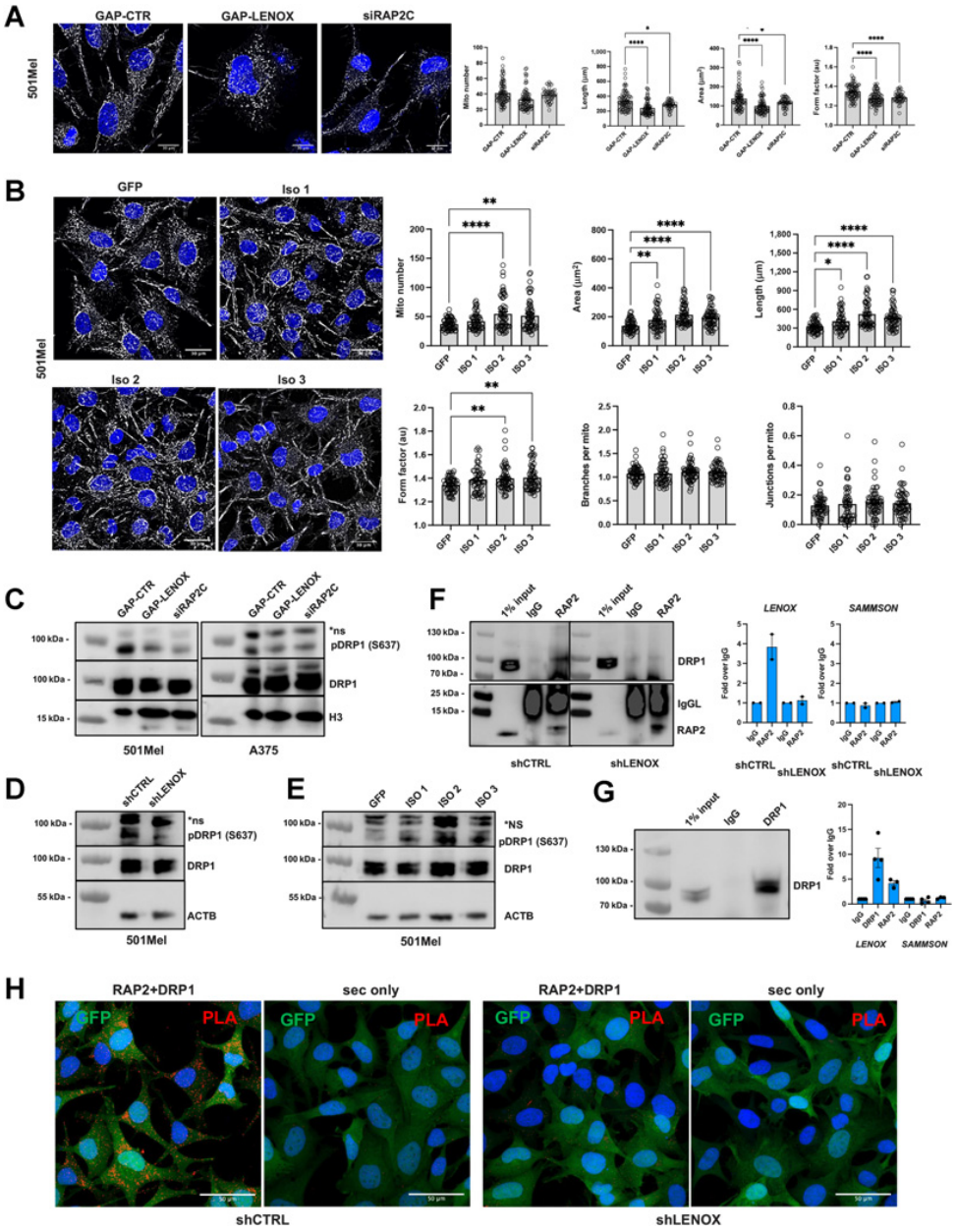


Figure 4.

LENOX promotes RAP2 interaction with DRP1 regulating its phosphorylation on serine 637 and mitochondrial fusion. **A,** MitoTracker CMXROS Red staining of 501Mel cells transfected with the indicated GapmeR or siRAP2C. MitoTracker signal is shown in grayscale and Hoescht in blue. Total mitochondrial count, area, perimeter, and mean form factor were calculated and compared by one-way ANOVA (Dunnett test). **B,** MitoTracker CMXROS Red staining of 501Mel cells with ectopic LENOX isoform expression. Mitochondrial parameters were calculated and compared as above. **C–E,** Phospho-DRP1 S637 and total DRP1 levels in 501Mel and A375 cell extracts, with H3 and ACTB as loading controls. **F,** Left, RAP2 IP from 501Mel cells expressing shRNAs revealed with antibodies to the indicated proteins. Right, LENOX and SAMMSON levels in the indicated IP fractions. **G,** IP of DRP1 from 501Mel cells (left) and LENOX and SAMMSON levels in the IP (right). **H,** PLA-mediated detection RAP2–DRP1 interaction in 501Mel cells. Appropriate negative controls using secondary antibodies only are shown. *, $P < 0.033$; **, $P < 0.0021$; ****, $P < 0.0001$.

Fig. S10K). These results highlight the role of TFAP2A and SOX10 in LENOX induction by MAPKi treatment.

We investigated if vemurafenib-induced LENOX expression was required to stimulate OXPHOS and promote cell survival and hence an increased vulnerability to LENOX silencing. LENOX, RAP2C, or TFAP2A silencing strongly reduced the vemurafenib-mediated increase in basal, reserve, and maximal OCR compared with DMSO control cells showing the essential role of this axis in the adaptive

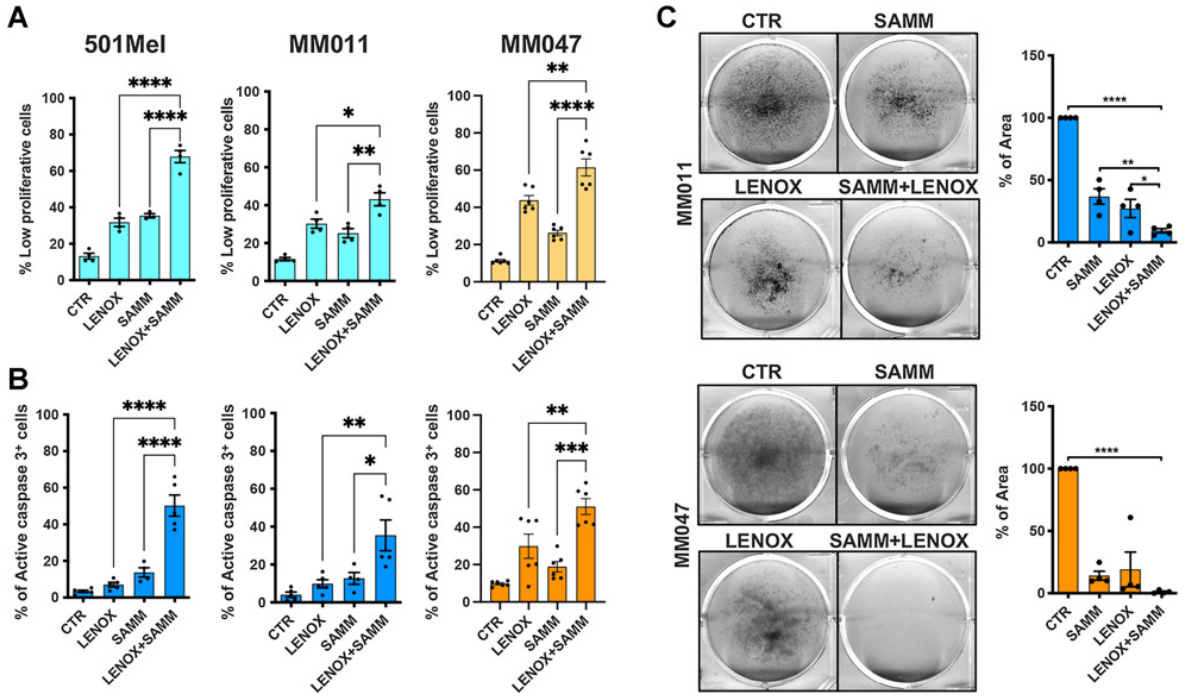


Figure 5.

LENOX and SAMMSON cooperatively promote melanoma cell survival. **A** and **B,** Proliferation and apoptosis after transfection with suboptimal doses of LENOX or SAMMSON GapmeRs as single or pairwise combinations compared between groups by one-way ANOVA. **C,** Crystal violet staining of cells transfected as indicated and cultured for 10 days. Percentages of area occupied in each condition were compared by one-way ANOVA. *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

metabolic switch (**Fig. 6A** and **B**). LENOX knockdown was comparable in DMSO- and vemurafenib-treated cells and its levels were also reduced by siTFAP2A, corroborating its role in LENOX induction (Supplementary Fig. S10L).

In accordance with increased OXPHOS, vemurafenib-treated cells displayed a more complex and elongated mitochondria network compared with DMSO (**Fig. 6C**; Supplementary Figs. S10M and S10N). LENOX or RAP2C silencing in vemurafenib-treated cells restored the rounder and shorter morphology, reduced mitochondrial connectivity and DRP1 S637 phosphorylation (**Fig. 6D**). LENOX and RAP2C therefore promoted mitochondrial fusion required for adaptive increase of OXPHOS upon vemurafenib treatment.

Vemurafenib potently induced arrest of cell proliferation after 3 and 6 days, which was not enhanced by LENOX, RAP2C, or TFAP2A silencing that otherwise strongly reduced proliferation of DMSO control cells (**Fig. 6E**). In contrast, vemurafenib alone did not induce apoptosis during this period (**Fig. 6F**). Importantly, vemurafenib-treated cells displayed increased apoptosis compared DMSO controls after LENOX, RAP2C, or TFAP2A silencing that was particularly evident at 6 days. DT treated cells were also more sensitive to LENOX silencing (**Fig. 6G** and **H**). Thus, MAPKi cooperated with LENOX silencing to induce apoptosis underlining the critical role of the TFAP2A/LENOX /RAP2C axis in the metabolic adaptation to drug treatment.

We tested whether LENOX silencing could inhibit growth of the MEL006res PDX in presence of MAPKi. Mice were engrafted and once palpable, the tumors were treated with DT and the resistant lesions subsequently challenged with either control GAP ASO (DT+GAP-CTR) or LENOX GAP#2 ASO (DT+GAP-LENOX). Compared with DT alone, or DT with control GAP-ASO, tumor progression with LENOX GAP#2 ASO was strongly and significantly reduced, leading to better overall survival (**Fig. 6I** and **J**), revealing that LENOX targeting inhibited MAPKi resistant PDX tumor growth.

We assessed OXPHOS in vemurafenib-treated cells ectopically expressing LENOX. In control cells, OCR levels were stimulated by vemurafenib and further increased by ectopic LENOX expression (Supplementary Fig. S11A). As vemurafenib stimulates endogenous LENOX expression, overall LENOX levels were higher in vemurafenib-treated cells than in the DMSO controls (Supplementary Fig. S11B). Furthermore, inhibition of melanospheres by vemurafenib treatment was rescued by ectopic LENOX expression (Supplementary Fig. S11C). Loss and gain of LENOX underscore its essential role of LENOX in the adaptive response to MAPKi.

As MAPKi treatment inhibited glycolysis and increased sensitivity to LENOX silencing, we asked whether such vulnerability was observed under other conditions where glycolysis was inhibited. Growth under low glucose or in presence of galactose upregulated both TFAP2A and LENOX expression (Supplementary Figs. S11D and S11E), highlighting the role of TFAP2A as a metabolic sensor promoting compensatory gene expression changes such as increased LENOX. Increased DRP1 S637 phosphorylation and OXPHOS were also observed and reduced by LENOX silencing, particularly in cells grown in galactose (Supplementary Figs. S11F and S11H). Thus, inhibition of glycolysis was compensated by a LENOX-dependent increase in OXPHOS increasing their vulnerability to LENOX loss (Supplementary Fig. S11I).

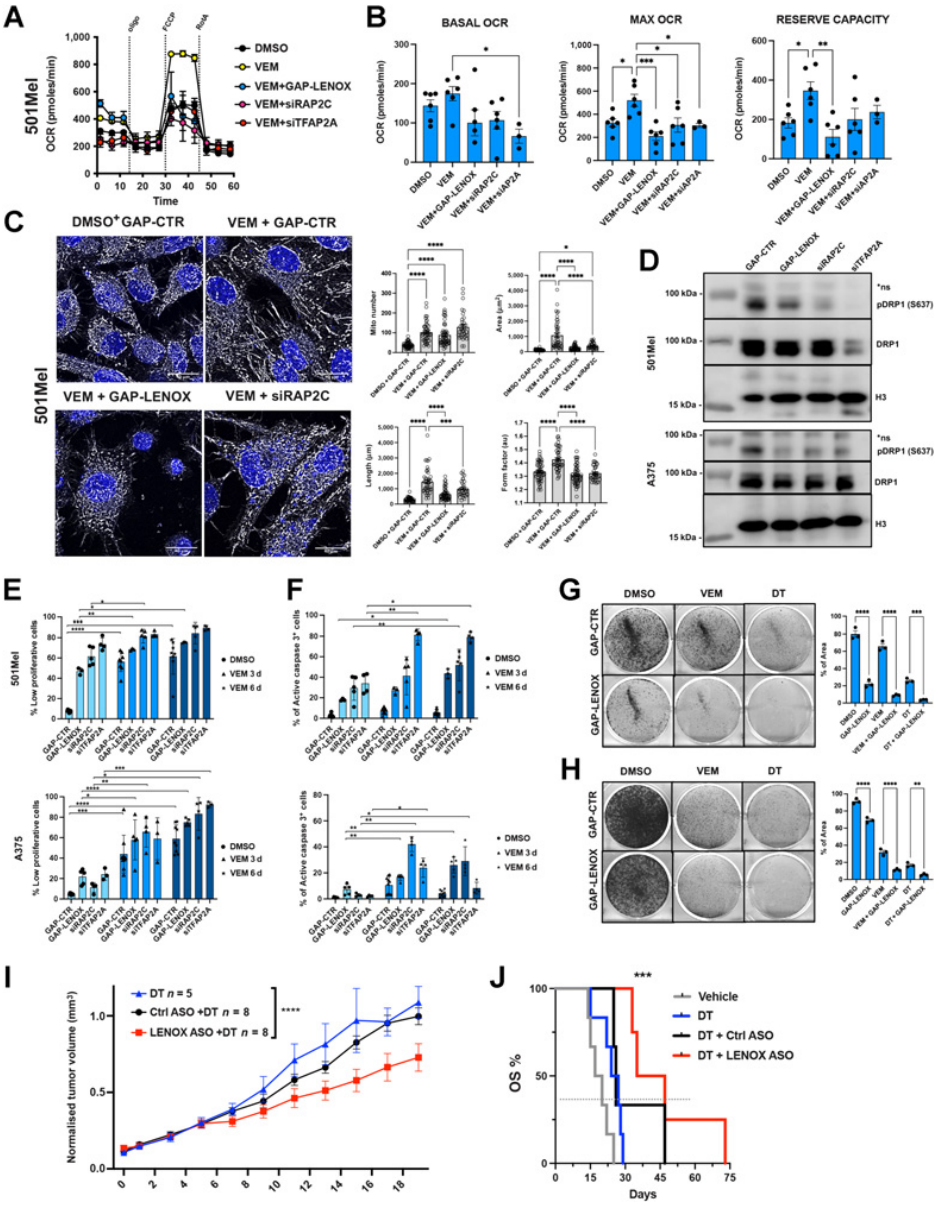


Figure 6.

LENOX and RAP2C promote metabolic switch upon BRAF inhibition. **A** and **B,** Mito Stress Test of 501Mel cells grown for 3 days with DMSO or vemurafenib (1 μmol/L) and transfected as indicated. Experimental groups were compared by one-way ANOVA (Tukey test). **C,** MitoTracker CMXROS Red staining of cells treated with DMSO or vemurafenib, transfected as indicated and analyzed by confocal microscopy as in **Fig. 4A**. **D,** Phospho-DRP1 S637 and total-DRP1 levels in cells treated as above. H3 was used as loading control. **E** and **F,** Proliferation and apoptosis of cells treated as indicated. Experimental groups were compared by one-way ANOVA. **G** and **H,** Crystal violet staining of GapmeR-transfected cells. Percentages of occupied areas were compared by one-way ANOVA. **I,** Growth of MEL006res PDX in mice treated with DT with or without nontargeting ASO (DT+GAP-CTR) or LENOX-targeting ASO (DT+GAP-LENOX). Tumor size was measured daily for 20 days. Statistics were calculated using two-way ANOVA (Šidák multiple comparisons test). **J,** Kaplan–Meier plot showing overall survival (OS) of mice described in **I**. DT, $n = 5$; DT+GAP-CTR, $n = 3$; DT+GAP-LENOX ASO, $n = 4$. Statistical analysis were calculated by log-rank (Mantel–Cox) test. *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

Discussion

Metabolic reprogramming is a hallmark of cancer cells allowing them to adapt to changing and often stressful environments while balancing energy production with the availability of metabolites used to fuel anabolic process required for cell proliferation (42, 43). Emerging evidence suggests that targeting mitochondrial function and ATP generation may represent a common vulnerability across different cell states and tumor types (23). Here we describe how LENOX regulates

mitochondrial dynamics and OXPHOS under basal, MAPKi, and metabolic stress conditions.

We identify the poorly characterized GTPase RAP2C as a binding partner of LENOX mediating its biological effects. RAP2 proteins have been proposed as intermediate activation molecules in the LATS1/2–YAP/TAZ signaling cascade (44). However, in melanoma cells, RAP2C has a distinct function in regulating mitochondrial fusion and metabolism. LENOX or RAP2C silencing was associated with increased fission and reduced maximal and reserve but not basal OXPHOS levels. Gain of LENOX function increased mitochondrial fusion and stimulated basal, maximal, and reserve capacities, further linking mitochondrial fusion status to altered OXPHOS capacity (45).

Mitochondrial homeostasis involves several GTPases (46) including DRP1 that forms oligomeric rings to drive fission through GTP hydrolysis, a process dynamically regulated by its posttranslational modification (46, 47). RAP2C interacted with DRP1 in a LENOX-dependent manner to enhance S637 phosphorylation and mitochondrial fusion and promote survival by limiting ROS production and cytochrome C release. DRP1 inhibition and mitochondrial fusion inhibit apoptosis induced by a variety of stress situations (38, 39, 48, 49). Moreover, fine tuning of DRP1 activity to increase mitochondrial fusion and OXPHOS has further been associated with increased oncogenic transformation (50, 51). LENOX may therefore optimize OXPHOS and survival to favor oncogenic transformation and subsequent progression in stressful cellular environments encountered during primary tumor growth and metastatic dissemination (52). We note also that increased DRP1 S637 phosphorylation by LENOX-RAP2C associated with increased melanoma cell proliferation was distinct from a P53-dependent increase in phospho-S637 observed at the onset of cellular senescence in nonmelanoma lines (53).

TFAP2A acts as a metabolic sensor that promotes OXPHOS and cell survival by upregulating LENOX in response to metabolic stress induced by MAPKi and/or glycolysis inhibition. Consequently, MAPKi or glycolysis inhibited cells showed enhanced vulnerability to LENOX silencing since reducing OXPHOS capacity under conditions where glycolysis was also impaired potently induced cell death. Moreover, ASO-mediated LENOX targeting *in vivo*, inhibited growth of both drug naïve melanoma PDX and MAPKi-resistant PDX suggesting LENOX inhibition as a promising therapeutic strategy for drug-resistant melanoma. Combinatorial targeting of glycolysis and mitochondrial metabolism in tumors has indeed been previously proposed as a therapeutic approach (54, 55).

Intriguingly, SAMMSON and LENOX are located adjacent to the *MITF* and *TFAP2A* loci, respectively, and are coamplified with them in 8% to 10% of melanoma. These lincRNAs may have acquired their melanoma specific functions through genomic association with these highly expressed loci. As each lincRNA regulated complementary pathways converging on mitochondria, their combinatorial silencing potently induced apoptosis, suggesting an effective therapeutic option to target undifferentiated cell states in minimal residual disease and relapse. This cooperativity highlights how critical hallmark functions of MITF, SOX10, and TFAP2A are mediated by their protein-coding targets and cooperatively acting lincRNAs. SOX10 orchestrates multiple aspects of mitochondrial function, directly regulating SAMMSON and activating MITF that drives PPARGC1A and LENOX expression promoting mitochondrial biogenesis and homeostasis (21, 40).

In conclusion, melanomas are addicted to both SAMMSON and LENOX that play distinct and complementary functions to optimize protein translation and mitochondrial function, two critical needs of cancer cells.

Authors' Disclosures

G. Gambi reports personal fees from FRM during the conduct of the study; personal fees from SATT Conectus outside the submitted work; and also has a patent for EP20305074.5 (Antisense oligonucleotide targeting Linc00518 for treating melanoma) pending. V. Katopodi reports grants from FWO during the conduct of the study. E. Leucci reports grants from Belgian Federation for Cancer FAF-F/2018/1184, Melanoma Research Alliance (MRA; https://doi.org/10.48050/pc.gr.80542), and FWO-KOTK grant (#G097620N) during the conduct of the study. I. Davidson reports grants from ITMO Cancer and Ligue National Contre le Cancer during the conduct of the study; in addition, I. Davidson has a patent for Antisense oligonucleotide targeting LINC00518 for treating melanoma (EP20305074.5) pending. No disclosures were reported by the other authors.

Authors' Contributions

G. Gambi: Conceptualization, data curation, formal analysis, investigation, visualization, methodology, writing–original draft. **G. Mengus:** Investigation, methodology. **G. Davidson:** Formal analysis, investigation. **E. Demesmaeker:** Investigation, methodology. **A. Cuomo:** Data curation, formal analysis, investigation, methodology. **T. Bonaldi:** Data curation, formal analysis, methodology. **V. Katopodi:** Investigation, methodology. **G.G. Malouf:** Resources, funding acquisition, methodology. **E. Leucci:** Conceptualization, resources, supervision, funding acquisition, writing–original draft, project administration, writing–review and editing. **I. Davidson:** Conceptualization, supervision, funding acquisition, writing–original draft, project administration, writing–review and editing.

Acknowledgments

The authors thank J-C. Marine and G. Ghanem for the MM-series primary melanoma cells, D. Lipkser and the Dermatology Clinic of Strasbourg University Hospital for patient melanoma sections, the staff of the IGBMC common facilities. This work was supported by grants from the ITMO Cancer, the Ligue Nationale contre le Cancer, the Fondation ARC 2020 PJA3, the ANR-10-LABX-0030 and ANR-10-IDEX-0002–02. Work in the Leucci laboratory was supported by the Melanoma Research Alliance (MRA; https://doi.org/10.48050/pc.gr.80542) grants, the Amanda and Jonathan Eilian Young Investigator Award 2018, the Belgian Federation for Cancer grant FAF-F/2018/1184, and by the FWO-KOTK grant (#G097620N). Trace staff were supported by Stichting Tegen Kanker grant 2016-054. Work in the Bonaldi and Cuomo laboratory was supported by EPIC-XS, project number 823839, funded by the Horizon 2020 program of the European Union. ID is an 'équipe labellisée' of the Ligue Nationale contre le Cancer. G. Gambi was supported by fellowships from the Fondation pour la Recherche Médicale and the SATT-Conectus Alsace. V. Katopodi is a recipient of the FWO PhD fellowship 1S47519N.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Note

Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Received March 21, 2022; revised June 27, 2022; accepted September 21, 2022; published first October 10, 2022.

References

1. Muller J, Krijgsman O, Tsoi J, Robert L, Hugo W, Song C, et al. Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. Nat Commun 2014;5:5712.

2. Tsoi J, Robert L, Paraiso K, Galvan C, Sheu KM, Lay J, et al. Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. Cancer Cell 2018;33:890–904.

3. Ennen M, Keime C, Gambi G, Kieny A, Coassolo S, Thibault-Carpentier C, et al. MITF-high and MITF-low cells and a novel subpopulation expressing genes of both cell states contribute to intra and inter-tumoral heterogeneity of primary melanoma. Clin Cancer Res 2017;23:7097–107.
4. Rambow F, Rogiers A, Marin-Bejar O, Aibar S, Femel J, Dewaele M, et al. Toward minimal residual disease-directed therapy in melanoma. Cell 2018; 174:843–55.
5. Rambow F, Marine JC, Goding CR. Melanoma plasticity and phenotypic diversity: therapeutic barriers and opportunities. Genes Dev 2019;33:1295–318.
6. Vendramin R, Konnova A, Adnane S, Cinque S, Katopodi V, Knezevic Z, et al. Activation of the integrated stress response in drug-tolerant melanoma cells confers vulnerability to mitoribosome-targeting antibiotics. Cancer Biology; 2020 Available from: http://biorxiv.org/lookup/doi/10.1101/2020.06.26.173492.
7. Laurette P, Coassolo S, Davidson G, Michel I, Gambi G, Yao W, et al. Chromatin remodellers Brg1 and Bptf are required for normal gene expression and progression of oncogenic Braf-driven mouse melanoma. Cell Death Differ 2019;27:29–43.
8. Strub T, Giuliano S, Ye T, Bonet C, Keime C, Kobi D, et al. Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. Oncogene 2011;30:2319–32.
9. Vivas-Garcia Y, Falletta P, Liebing J, Louphrasitthiphol P, Feng Y, Chauhan J, et al. Lineage-restricted regulation of SCD and fatty acid saturation by MITF controls melanoma phenotypic plasticity. Mol Cell 2020;77:120–37.
10. Laurette P, Strub T, Koludrovic D, Keime C, Le Gras S, Seberg H, et al. Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. Elife 2015;4:e06857.
11. Goding CR, Arnheiter H. MITF-the first 25 years. Genes Dev 2019;33:983–1007.
12. Riesenberg S, Groetchen A, Siddaway R, Bald T, Reinhardt J, Smorra D, et al. MITF and c-Jun antagonism interconnects melanoma dedifferentiation with pro-inflammatory cytokine responsiveness and myeloid cell recruitment. Nat Commun 2015;6:8755.
13. Bai X, Fisher DE, Flaherty KT. Cell-state dynamics and therapeutic resistance in melanoma from the perspective of MITF and IFNgamma pathways. Nat Rev Clin Oncol 2019;16:549–62.
14. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. Cell. 2018;175:984–97.
15. Marin-Bejar O, Rogiers A, Dewaele M, Femel J, Karras P, Pozniak J, et al. Evolutionary predictability of genetic versus nongenetic resistance to anticancer drugs in melanoma. Cancer Cell 2021;39:1135–49.
16. Boshuizen J, Vredevoogd DW, Krijgsman O, Ligtenberg MA, Blankenstein S, de Bruijn B, et al. Reversal of pre-existing NGFR-driven tumor and immune therapy resistance. Nat Commun 2020;11:3946.
17. Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 2021;22:96–118.
18. Huarte M. The emerging role of lncRNAs in cancer. Nat Med 2015;21:1253–61.
19. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. Cancer Cell 2016;29:452–63.
20. Aprile M, Katopodi V, Leucci E, Costa V. LncRNAs in cancer: from garbage to Junk. Cancers (Basel) 2020;12:3220.
21. Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, et al. Melanoma addiction to the long non-coding RNA SAMMSON. Nature 2016; 531:518–22.
22. Vendramin R, Verheyden Y, Ishikawa H, Goedert L, Nicolas E, Saraf K, et al. SAMMSON fosters cancer cell fitness by concertedly enhancing mitochondrial and cytosolic translation. Nat Struct Mol Biol 2018;25:1035–46.
23. Vendramin R, Katopodi V, Cinque S, Konnova A, Knezevic Z, Adnane S, et al. Activation of the integrated stress response confers vulnerability to mitoribosome-targeting antibiotics in melanoma. J Exp Med 2021;218:e20210571.
24. Nikolaisen J, Nilsson LIH, Pettersen IKN, Willems PHGM, Lorens JB, Koopman WJH, et al. Automated quantification and integrative analysis of 2D and 3D mitochondrial shape and network properties. PLoS One 2014;9:e101365.
25. Chaudhry A, Shi R, Luciani DS. A pipeline for multidimensional confocal analysis of mitochondrial morphology, function, and dynamics in pancreatic β-cells. Am J Physiol Endocrinol Metab 2020;318:E87–101.
26. Merrill RA, Flippo KH, Strack S. Measuring mitochondrial shape with ImageJ. In: Strack S, Usachev YM, editors. Techniques to investigate mitochondrial function in neurons [Internet]. New York, NY: Springer New York; 2017 [cited 2021 Nov 29]. page 31–48. Available from: http://link.springer.com/10.1007/978–1-4939–6890–9_2.
27. Conrad T, Ørom UA. Cellular fractionation and isolation of chromatin-associated RNA. Ørom UA, editor. Enhancer RNAs: methods and protocols. New York, NY: Springer New York; 2017.
28. Badal B, Solovyov A, Di Cecilia S, Chan JM, Chang LW, Iqbal R, et al. Transcriptional dissection of melanoma identifies a high-risk subtype underlying TP53 family genes and epigenome deregulation. JCI Insight 2017;2:e92102.
29. Kunz M, Loffler-Wirth H, Dannemann M, Willscher E, Doose G, Kelso J, et al. RNA-seq analysis identifies different transcriptomic types and developmental trajectories of primary melanomas. Oncogene 2018;37:6136–51.
30. Babaian A, Mager DL. Endogenous retroviral promoter exaptation in human cancer. Mob DNA 2016;7:24.
31. Fontanals-Cirera B, Hasson D, Vardabasso C, Di Micco R, Agrawal P, Chowdhury A, et al. Harnessing BET inhibitor sensitivity reveals AMIGO2 as a melanoma survival gene. Mol Cell 2017;68:731–44.
32. Seberg HE, Van Otterloo E, Loftus SK, Liu H, Bonde G, Sompallae R, et al. TFAP2 paralogs regulate melanocyte differentiation in parallel with MITF. PLoS Genet 2017;13:e1006636.
33. Wouters J, Kalender-Atak Z, Minnoye L, Spanier KI, De Waegeneer M, Bravo González-Blas C, et al. Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma. Nat Cell Biol 2020;22:986–98.
34. Rambow F, Job B, Petit V, Gesbert F, Delmas V, Seberg H, et al. New functional signatures for understanding melanoma biology from tumor cell lineage-specific analysis. Cell Rep 2015;13:840–53.
35. Kashatus JA, Nascimento A, Myers LJ, Sher A, Byrne FL, Hoehn KL, et al. Erk2 phosphorylation of Drp1 promotes mitochondrial fission and MAPK-driven tumor growth. Mol Cell 2015;57:537–51.
36. Serasinghe MN, Wieder SY, Renault TT, Elkholi R, Asciolla JJ, Yao JL, et al. Mitochondrial division is requisite to RAS-induced transformation and targeted by oncogenic MAPK pathway inhibitors. Mol Cell 2015;57:521–36.
37. Chang CR, Blackstone C. Cyclic AMP-dependent protein kinase phosphorylation of Drp1 regulates its GTPase activity and mitochondrial morphology. J Biol Chem 2007;282:21583–7.
38. Cribbs JT, Strack S. Reversible phosphorylation of Drp1 by cyclic AMP-dependent protein kinase and calcineurin regulates mitochondrial fission and cell death. EMBO Rep 2007;8:939–44.
39. Wang Z, Jiang H, Chen S, Du F, Wang X. The Mitochondrial phosphatase PGAM5 functions at the convergence point of multiple necrotic death pathways. Cell 2012;148:228–43.
40. Haq R, Shoag J, Andreu-Perez P, Yokoyama S, Edelman H, Rowe GC, et al. Oncogenic BRAF regulates oxidative metabolism via PGC1alpha and MITF. Cancer Cell 2013;23:302–15.
41. Parmenter TJ, Kleinschmidt M, Kinross KM, Bond ST, Li J, Kaadige MR, et al. Response of BRAF-mutant melanoma to BRAF inhibition is mediated by a network of transcriptional regulators of glycolysis. Cancer Discov 2014;4: 423–33.
42. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144:646–74.
43. Vander Heiden MG, DeBerardinis RJ. Understanding the Intersections between metabolism and cancer biology. Cell 2017;168:657–69.
44. Meng Z, Qiu Y, Lin KC, Kumar A, Placone JK, Fang C, et al. RAP2 mediates mechanoresponses of the hippo pathway. Nature 2018; 560:655–60.
45. Chan DC. Mitochondrial dynamics and its involvement in disease. Annu Rev Pathol 2020;15:235–59.
46. Sabouny R, Shutt TE. Reciprocal regulation of mitochondrial fission and fusion. Trends Biochem Sci 2020;45:564–77.
47. Chang C-R, Blackstone C. Dynamic regulation of mitochondrial fission through modification of the dynamin-related protein Drp1: chang & blackstone. Ann N Y Acad Sci 2010;1201:34–9.
48. Li J, Huang Q, Long X, Guo X, Sun X, Jin X, et al. Mitochondrial elongation-mediated glucose metabolism reprogramming is essential for tumour cell survival during energy stress. Oncogene 2017;36:4901–12.
49. Gomes LC, Di Benedetto G, Scorrano L. During autophagy mitochondria elongate, are spared from degradation and sustain cell viability. Nat Cell Biol 2011;13:589–98.
50. Spurlock B, Parker D, Basu MK, Hjelmeland A, Gc S, Liu S, et al. Fine-tuned repression of Drp1-driven mitochondrial fission primes a "stem/progenitor-like state" to support neoplastic transformation. Elife 2021;10:e68394.
51. Bonnay F, Veloso A, Steinmann V, Köcher T, Abdusselamoglu MD, Bajaj S, et al. Oxidative metabolism drives immortalization of neural stem cells during tumorigenesis. Cell 2020;182:1490–507.

52. Bergers G, Fendt S-M. The metabolism of cancer cells during metastasis. Nat Rev Cancer 2021;21:162–80.
53. Kim YY, Um J, Yoon J, Lee D, Lee YJ, Kim DH, et al. p53 regulates mitochondrial dynamics by inhibiting Drp1 translocation into mitochondria during cellular senescence. FASEB J 2020;34:2451–64.
54. Weinberg SE, Chandel NS. Targeting mitochondria metabolism for cancer therapy. Nat Chem Biol 2015;11:9–15.
55. Martinez-Outschoorn UE, Peiris-Pagés M, Pestell RG, Sotgia F, Lisanti MP. Cancer metabolism: a therapeutic perspective. Nat Rev Clin Oncol 2017;14:11–31.

CDDpress

www.nature.com/cdd

ARTICLE OPEN

Check for updates

General transcription factor TAF4 antagonizes epigenetic silencing by Polycomb to maintain intestine stem cell functions

Susanna Säisä-Borreill[1], Guillaume Davidson[2], Thomas Kleiber[2,3], Andréa Thevenot[1], Elisabeth Martin[1], Stanislas Mondot[4], Hervé Blottière[4], Alexandra Helleux[2], Gabrielle Mengus[2], Michelina Plateroti[1], Isabelle Duluc[1,5], Irwin Davidson[2,5] and Jean-Noel Freund[1,5]✉

Taf4 (TATA-box binding protein-associated factor 4) is a subunit of the general transcription factor TFIID, a component of the RNA polymerase II pre-initiation complex that interacts with tissue-specific transcription factors to regulate gene expression. Properly regulated gene expression is particularly important in the intestinal epithelium that is constantly renewed from stem cells. Tissue-specific inactivation of Taf4 in murine intestinal epithelium during embryogenesis compromised gut morphogenesis and the emergence of adult-type stem cells. In adults, Taf4 loss impacted the stem cell compartment and associated Paneth cells in the stem cell niche, epithelial turnover and differentiation of mature cells, thus exacerbating the response to inflammatory challenge. *Taf4* inactivation ex vivo in enteroids prevented budding formation and maintenance and caused broad chromatin remodeling and a strong reduction in the numbers of stem and progenitor cells with a concomitant increase in an undifferentiated cell population that displayed high activity of the Ezh2 and Suz12 components of Polycomb Repressive Complex 2 (PRC2). Treatment of Taf4-mutant enteroids with a specific Ezh2 inhibitor restored buddings, cell proliferation and the stem/progenitor compartment. *Taf4* loss also led to increased PRC2 activity in cells of adult crypts associated with modification of the immune/inflammatory microenvironment that potentiated *Apc*-driven tumorigenesis. Our results reveal a novel function of Taf4 in antagonizing PRC2-mediated repression of the stem cell gene expression program to assure normal development, homeostasis, and immune-microenvironment of the intestinal epithelium.

Cell Death & Differentiation (2023) 30:839–853; https://doi.org/10.1038/s41418-022-01109-6

INTRODUCTION

The intestinal epithelium is characterized by dynamic cell renewal every 5–6 days fueled by active stem cells (SCs) located in crypts, the Lgr5^{high} Crypt Base Columnar cells (CBCs), that generate committed progenitors and ultimately mature digestive cells of the absorptive and secretory lineages [1]. These complex processes are controlled by complementary signaling pathways of which the main ones are Wnt, Notch and BMP. The last decades identified a whole range of transcription factors among which bHLH, HMG, homeodomain, Zinc-finger, Krüppel-like and other factors that are targeted by and/or cooperate with these pathways to ensure the dynamic homeostasis of the gut epithelium. These factors bind specific *cis*-regulatory DNA sequences and interact with the pre-initiation complex (PIC) of the general transcription machinery to promote gene expression. Their precise contribution to gut development and homeostasis has been deciphered through genetic knockouts. However, beyond these tissue-specifying transcription factors, the role of the general transcription machinery has not been addressed so far in the gut.

Taf4a (hereafter Taf4, for TATA-box binding protein-associated factor 4) and Taf4b are paralogous subunits of the general transcription factor TFIID comprising the TATA-box binding protein (TBP) and 13–14 TBP-associated factors (TAFs), that plays a critical role in PIC formation. While Taf4 is widely expressed, Taf4b displays a cell-restricted expression. We previously inactivated the *Taf4* gene in a variety of somatic murine tissues or during embryogenesis showing how Taf4 controlled gene expression programs in a tissue-specific manner regulating embryonic tissue differentiation [2] as well as the homeostasis of the epidermis [3] and the function of several endoderm-derived cell types such as the activation of post-natal metabolism genes in neonatal hepatocytes and the identity and activity of pancreatic beta cells [4, 5]. In addition to the liver and pancreas, the intestinal epithelium is another endodermal derivative that deserves attention because of its constant and active turnover.

Here, we inactivated *Taf4* in the intestinal epithelium in vivo during gut morphogenesis, adult homeostasis, and ex vivo in enteroid models. *Taf4* knockout compromised gut morphogenesis in embryos and impaired SCs and the dynamic homeostasis of the adult intestine. We reveal a novel function of Taf4 in antagonizing Polycomb Repressive Complex 2 (PRC2) to maintain the SC gene expression program. In absence of Taf4, SCs show increased PRC2 activity that

[1]University of Strasbourg, Inserm, UMR-S1113/IRFAC, FHU ARRIMAGE, FMTS, 67200 Strasbourg, France. [2]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Department of Functional Genomics and Cancer, CNRS/Inserm/University of Strasbourg, 1 Rue Laurent Fries, 67404 Illkirch Cédex, France. [3]Orphazyme, Ole Maaloes 3, 2200 Copenhagen, Denmark. [4]University Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350 Jouy-en-Josas, France. [4]These authors contributed equally: Isabelle Duluc, Irwin Davidson, Jean-Noel Freund. ✉email: jean-noel.freund@inserm.fr

Received: 20 June 2022 Revised: 15 December 2022 Accepted: 21 December 2022
Published online: 13 January 2023

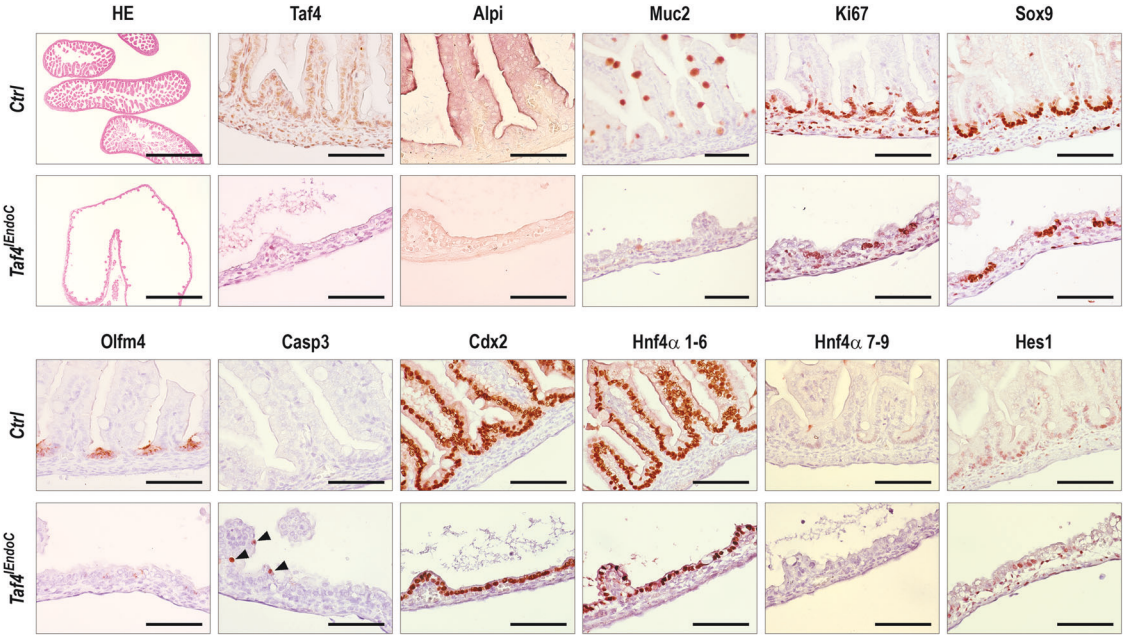


Fig. 1 Morphological alteration resulting from *Taf4* inactivation in the gut endoderm of E18.5 fetuses. Morphology (HE) and immunohistochemical detection of the indicated proteins in E18.5 control $Taf4^{los/lox}$ (Ctrl) and $Taf4^{IEndoC}$ littermates. Bars are 50 µm except for HE and Muc2 where they represent 500 µm and 100 µm, respectively.

correlates with an altered mucosal immune-microenvironment possibly involved in increased *Apc*-driven tumorigenesis.

RESULTS

Taf4 is required for proper development and morphogenesis of the embryonic intestinal endoderm

Taf4 gene inactivation starting around days 10–11 post coitum (dpc) in the presumptive intestinal endoderm of $Taf4^{lox/lox}$::*VilCre* embryos (hereafter $Taf4^{IEndoC}$) led to perinatal lethality. Loss of Taf4 protein was progressive, being mosaic at 14.5 dpc and almost complete at 17.5 dpc. At this stage, the length of the gut was reduced in mutant compared to control fetuses (3.29 vs. 4.52 cm, $p = 0.0007$) with impaired morphogenesis and altered alkaline phosphatase activity and Muc2 expression indicated perturbed cell differentiation, whereas cell proliferation labeled by Ki67 was unchanged. The impact of *Taf4* loss on the dynamic morphogenetic process was enhanced at E18.5 as the mucosa became flat with only few bulged villi (Fig. 1). Alkaline phosphatase (Alpi) activity and mucin Muc2 expression were strongly reduced. Nevertheless, presumptive crypt regions were preserved as evidenced by labeling with Sox9 and Ki67 that further demonstrated the proliferative capacity of these cells. However, Olfm4 marking the emergence of adult-type SCs in the inter-villi regions was barely detectable. By contrast, expression of Cdx2 and Hnf4α 1–6, two important transcription factors for intestinal identity and differentiation, was unaltered, whereas Hes1 and Hnf4α 7–9 were diminished and irregular in the regions presenting the most altered phenotype. In addition, Caspase-3-stained apoptotic cells were observed mainly at the level of remnant villi.

Taf4 regulates the dynamic homeostasis of the adult intestinal epithelium

Taf4 was inactivated in the adult gut epithelium of $Taf4^{lox/lox}$::$VilCreER^{T2}$ (hereafter $Taf4^{IEC}$) and control mice by Tamoxifen injection at the age of 2–3-month-old. *Taf4* inactivation did not alter the general morphology of the intestinal crypt-villus axis (Fig. 2A). However, a significant proportion of mice of the $Taf4^{IEC}$ series (15/28, 54%) had to be sacrificed during the course of the experiment as they reached the ethical limit point of weight lost. At later time points, the remaining 13 mice showed significantly lower body weight compared to controls that gained body weight over the same period (Fig. 2B). Progressively, $Taf4^{IEC}$ mice were less active and suffered from diarrhea. They exhibited a swollen cecum distended by gas and the whole intestine appeared mechanically fragile.

The gut epithelium of Tamoxifen-injected $Taf4^{IEC}$ mice exhibited cellular alterations in both proliferation and differentiation compartments (Fig. 2C). In the villi, enterocytes showed lower Alpi activity but increased Aldolase levels. The number of Muc2-positive goblet cells was reduced and enteroendocrine cells expressed less Chga. Dclk1 labeling of tuft cells was also strongly reduced. In the crypts, Sox9 was unaltered whereas Lyz in Paneth cells and Olfm4 in SCs were reduced, but still expressed. Ephb2, another marker of SCs was also reduced, although less than Olfm4 (Supplementary Fig. 1A). Functionally, the overall reduction of differentiated cells and/or of their features were associated with increased cell proliferation visualized with Ki67, higher phospho-Erk1/2 staining in the proliferative compartment, and accelerated cell turnover assessed by BrdU pulse-chase labeling (Fig. 2D, F). There was also a tendency to increased apoptosis displayed by activated Caspase-3, whereas the expression patterns of transcription factors Cdx2, Hnf4α 1–6 and 7–9, and Hes1 were unaltered (Supplementary Fig. 2).

To strengthen the relationship between the loss of Taf4 protein in SCs and the defective differentiation of mature villi cells, $Taf4^{lox/lox}$::$Lgr5\text{-}GFP\text{-}CreER^{T2}$ mice (hereafter $Taf4^{CBC}$) were generated to selectively inactivate *Taf4* in the CBCs. Adult mice were treated with Tamoxifen and analyzed 6 days later. As expected from the mosaic expression of the $Lgr5\text{-}GFP\text{-}CreER^{T2}$ allele in CBCs, only few crypts exhibited Taf4 loss in SCs whereas it remained present in adjacent Paneth cells due to their slow turnover (Supplementary Fig. 1B). $Taf4^{CBC}$ mice showed villi lined by Taf4-positive cells with few intermingled ribbons of Taf4-negative cells originating from the Taf4-depleted crypts. Taf4-negative cells in the villi exhibited less Alpi activity compared to adjacent Taf4-expressing cells, indicating that Taf4 inactivation in SCs compromised the terminal differentiation of their progeny migrating up the villi (Supplementary Fig. 1C).

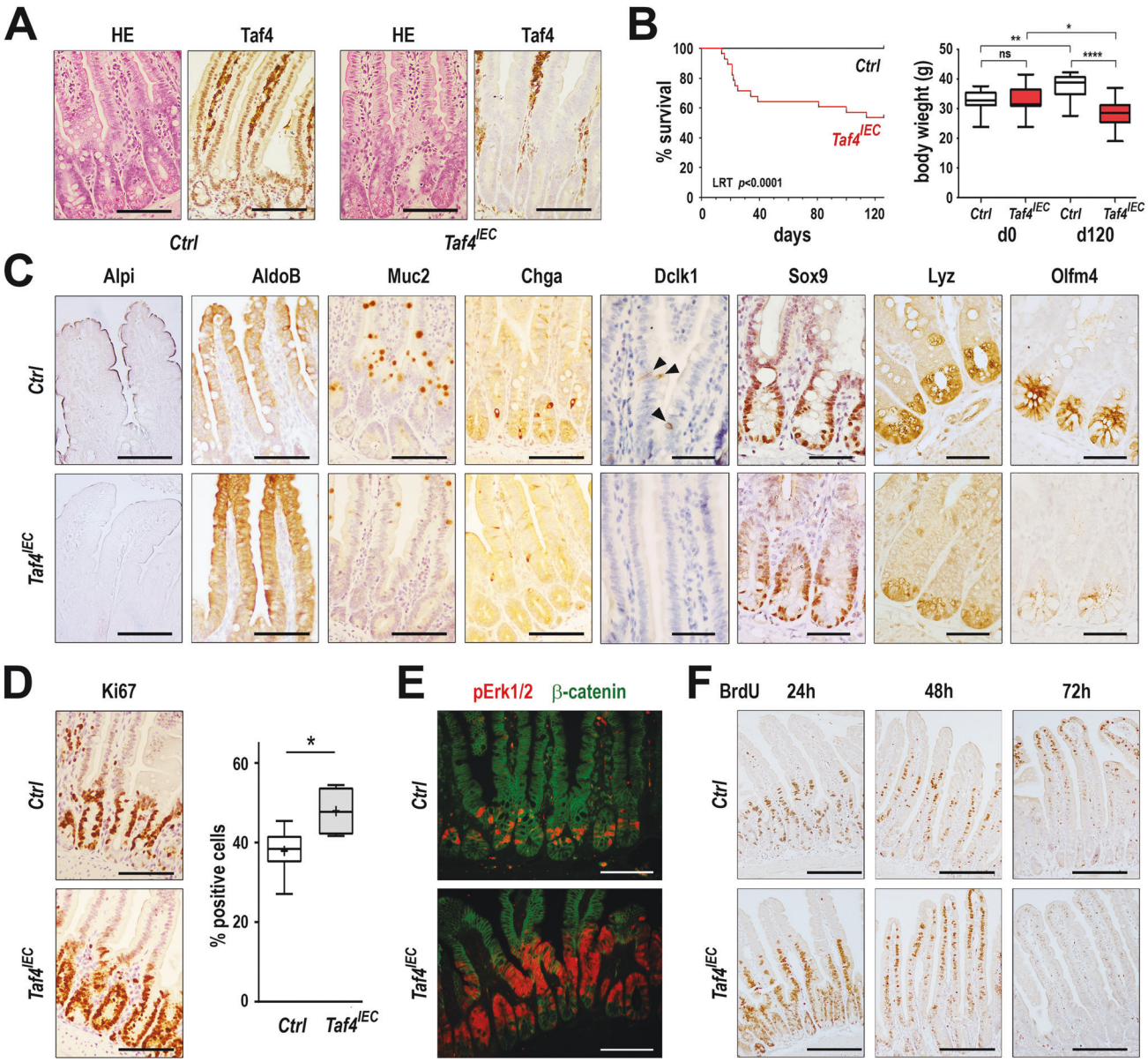


Fig. 2 Homeostasis defects induced by *Taf4* inactivation in the adult gut epithelium. A Histology (HE) and immunodetection of the Taf4 protein in the intestine of adult $Taf4^{IEC}$ and control $Taf4^{lox/lox}$ (Ctrl) mice 10 days after Tamoxifen injection. Bars are 100 µm. **B** Overall survival of $Taf4^{IEC}$ (red boxes) and $Taf4^{lox/lox}$ (Ctrl; white boxes) mice after Tamoxifen injection ($n = 28$ in each group; LRK Logrank test). Body weight of the 13 $Taf4^{IEC}$ mice surviving up to the end of the experiment and of 13 mice of the Ctrl group at day 0 and day 120 after Tamoxifen administration. ns not significant; $*p = 0.01$; $**p < 0.02$; $****p < 0.0001$. Boxes extend from the 25th–75th percentile and whiskers represent mean to max. **C** Immunodetection of the indicated proteins in the ileum of $Taf4^{IEC}$ and control mice 10 days after Tamoxifen injection. Bars are 100 µm for Alpi, Aldob, Muc2 and Chga, and 50 µm for Dclk1, Olfm4, Lyz and Sox9. **D** Ki67 immunostaining and cell counts in the ileum of $Taf4^{\Delta IEC}$ and control mice. Bars are 100 µm. Boxes extend from the 25th–75th percentile and whiskers represent mean to max. $*p < 0.05$. **E** Immunofluorescent staining of phospho-Erk1/2 (red) and β-catenin (green) in $Taf4^{IEC}$ and control mice. Bars are 100 µm. **F** BrdU detection in the jejunal mucosa of $Taf4^{IEC}$ and control mice at the indicated days after the single injection of BrdU. Bars are 200 µm.

Altogether, these results highlight that *Taf4* inactivation perturbs the dynamic homeostasis of the SC niche and transit amplifying cell compartment ultimately leading to impaired differentiation of both absorptive and secretory cell lineages.

Taf4 regulates proliferation and the immune/inflammatory microenvironment in fetal and adult intestine

The effect of *Taf4* inactivation on gene expression was investigated by RNA-seq in the intestine of E17.5 $Taf4^{lEndoC}$ fetus vs. control littermates (Supplementary Table 1), and in the ileum of adult $Taf4^{IEC}$ vs. control $Taf4^{lox/lox}$ mice 10 days after Tamoxifen administration (Supplementary Table 2). Comparing the transcriptional changes resulting from *Taf4* inactivation in E17.5 $Taf4^{lEndoC}$ and adult $Taf4^{IEC}$ mice revealed a higher number of down-regulated than up-regulated genes at both stages (respectively 1160 vs. 173 in fetuses and 769 vs. 284 in adults, ($|\log_2(FC)| > 1$, p value < 0.05) (Fig. 3A). At E17.5, KEGG and GSEA analyses showed that down-regulated genes were enriched in terms associated with metabolism, reflecting the alteration of functional epithelial cell differentiation (Fig. 3B, C). GSEA analyses of up-regulated genes revealed enrichment in terms associated with increased proliferation (Fig. 3C). Loss of *Taf4* in the adult intestine corroborated and extended the immune-histological data, with many genes involved in enterocyte functions down-

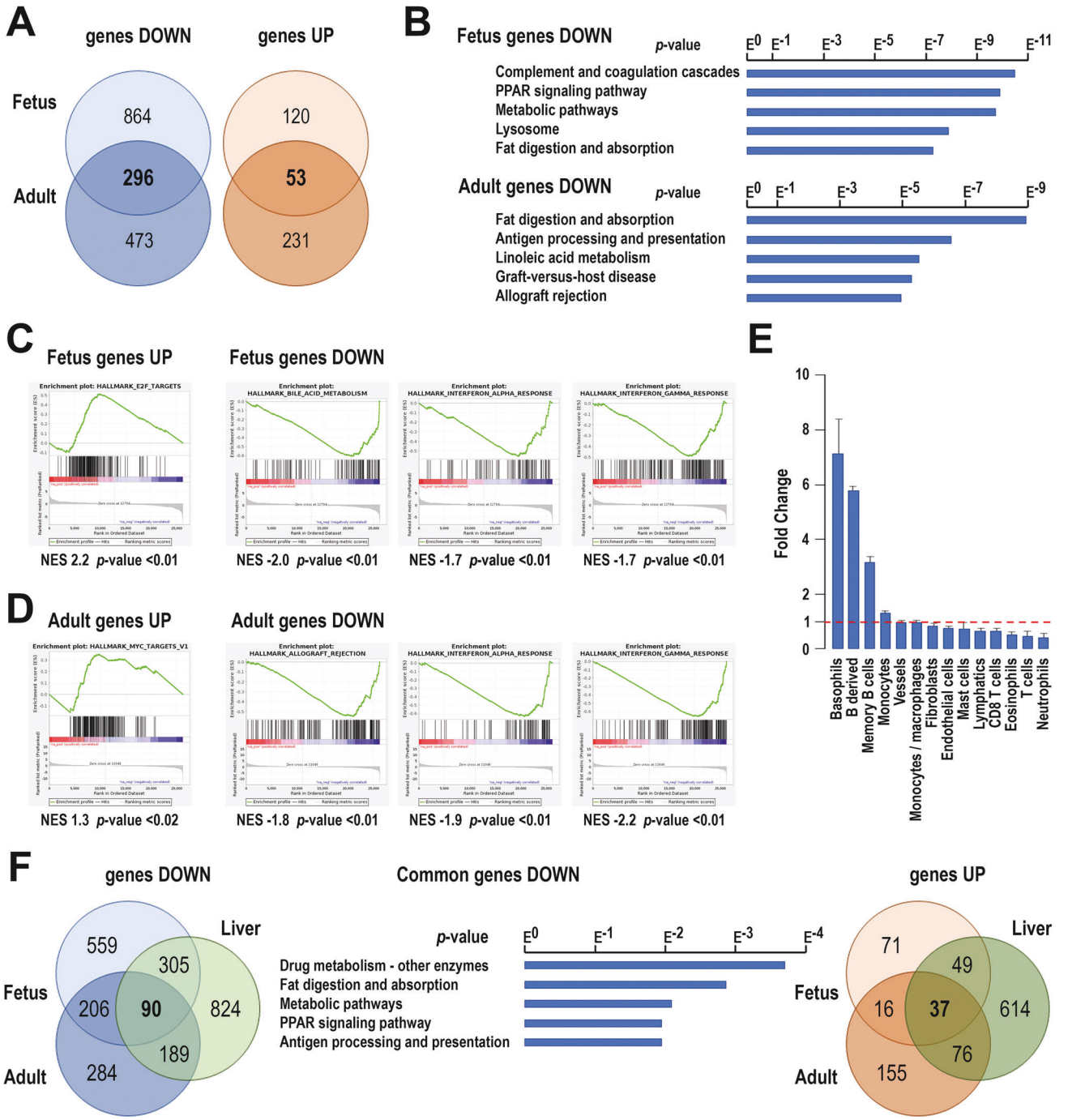


Fig. 3 Gene expression changes after *Taf4* inactivation. A Venn diagrams representing the down- and up-regulated genes in E17.5 *Taf4*^{IEndoC} fetuses and adult *Taf4*^{IEC} mice compared to their respective controls. **B** KEGG ontology enrichment is shown for the downregulated genes in fetuses and adults, ordered according to the *p* value. **C** GSEA analysis of the up-regulated and downregulated genes in fetuses identifying respectively hallmarks for E2F targets and for bile acid metabolism, and IFNα and IFNγ response. **D** GSEA analysis of the up-regulated and downregulated genes in adults identifying respectively hallmarks for Myc targets and for allograft rejection, and IFNα and IFNγ response. **E** Stromal cell population evaluation from RNA-seq data using the MCP method, expressed as the proportion of each cell type in *Taf4*^{IEC} mice compared to controls. **F** Venn diagrams of the down-regulated genes (left) and up-regulated genes (right) enriched in common in the intestine of E17.5 *Taf4*^{IEndoC} fetuses and adult *Taf4*^{IEC} mice, and in the *Taf4*-null liver of 12 days suckling mice. Middle: KEGG ontology enrichment of the 90 genes down-regulated in common in the fetal and adult intestine and in the liver of *Taf4*-deficient mice.

regulated, including those coding for digestive enzymes (Alpi, Anpep, Treh), fatty acid binding proteins (Fabp2, -6, -7), and more than 40 soluble carrier family members among which the transporters/cotransporters of Na/Glucose (Slc5a11/Sglt6, Slc5a12/Smct2), oligopeptides (Slc15a1/Ppept1), short-chain fatty acids (Slc16a3), amino acids (Slc7a8, Slc7a9, Slc7a15, Slc38a3) and folate (Slc46a1, Slc46a3) (Supplementary Table 2). However, the α-glucosidase genes *Sis* and *Mgam* and the aldolase gene *Aldob* were upregulated. Goblet cell mucin genes *Muc4* and *Muc20* were decreased, along with enteroendocrine genes encoding precursors and regulatory peptides (*Chga*, *Chgb*, *Cck*, *Gip*, *Pyy*, *Sct* and *Sst*) as well as *Dclk1* and *Pou2f3* for tuft cells. KEGG and GSEA analysis

confirmed enrichment of the down-regulated genes in different aspects of metabolism (Fig. 3B, D), whereas GSEA analysis of the up-regulated genes showed enrichment in cell proliferation (Fig. 3D). In addition, altered gene expression was observed in the SC niche, as exemplified by reduced expression of CBC signature genes (*Agr3*, *Ciita*, *Esrrg*, *Fras1*, *H2-Eb1*, *Hk2*, *Lect2*, *Olfm4*, *Rdh16*, *Sdsl*, *Sectm1b*, *Tifa*, *Tnfsf10*, *Tns4*, *Vnn1*) [6] as well as Paneth cell genes (*Defa26*, *Dafa-rs1*, *Defb1*, *Defb37*, *Lyz1*, *Mmp7*). Thus, *Taf4* inactivation perturbed the SC compartment and terminal differentiation of all mature epithelial cell types migrating along the villi.

A subset of genes was de-regulated in a similar manner in both the embryonic and adult contexts (Fig. 3A), sharing enrichment of several GSEA and KEGG ontology pathways notably associated with increased proliferation for the up-regulated genes and inflammation/immune functions including allograft rejection and the interferon (IFN) alpha and IFN gamma pathways for the down-regulated genes. Murine Microenvironment Cell Population (MCP [7]) counter analyses of the RNA-seq data further showed a modification of the immune microenvironment with increased number of basophils and B cells and reduced T cells in the *Taf4*-mutant adult intestinal mucosa (Fig. 3E). *Taf4* inactivation thus had cell-autonomous effects on epithelial cells, but also a non-cell-autonomous impact on the mucosal microenvironment.

We previously reported that *Taf4* inactivation in endoderm-derived neonatal hepatocytes also upset metabolic pathways [4]. Among the down-regulated genes in hepatocytes, 395 were shared with the fetal gut and 279 with the adult gut, 90 of them being affected in all 3 contexts (Fig. 3F). KEGG pathway analysis linked these 90 genes to several aspects of metabolism, complement function and chemical carcinogenesis (Fig. 3F).

Taf4 inactivation in the adult gut modifies the microbiota and enhances the inflammatory response

Along with digestive and metabolic functions, the gut epithelium plays an important barrier function at the interface between the luminal microbiota and the stroma. Diarrhea and the swollen cecum after *Taf4* loss suggested changes in the microbiota. Comparison of the bacterial composition within the cecum of Tamoxifen-treated $Taf4^{IEC}$ and control $Taf4^{lox/lox}$ mice 1 month after *Taf4* inactivation revealed reduced α-diversity with fewer operational taxonomic units (OTUs) in mutants than in controls (Supplementary Fig. 3A). The microbiota of $Taf4^{IEC}$ mice showed higher abundance of *Helicobacteriaceae*, *Bifidobacteriaceae* and *Deferribacteraceae* families but reduction of *Porpyromonadaceae*. At genus level, seven genera displayed differential levels: *Bifidobacterium*, *Clostridium XlVb* and *XVIII*, *Helicobacter*, *Mucispirillum* and *Olsenella* were significantly more abundant in mutants, whereas controls harbored more *Marvinbryantia*. OTUs included in *Helicobacter* genera were close to *H. typhlonius* (100%), *H. ganmani* (98%) and an uncultured species (JRPC—100%), and those included into *Bifidobacterium* genera close to *B. pseudolongum* (98%) and *B. animalis* (100%); *M schaedleri* (100%) was the main strain among *Mucispirillum* genera (Supplementary Fig. 3A). Two genera found in higher abundance in the cecum of *Taf4* deficient mice, *Helicobacter* and *Olsenella*, usually colonizers of the upper part of the digestive tract, are considered as detrimental for host health. *Bifidobacterium*, *Mucispirillum* and *Clostridium XlVb* and *XVIII*, also in higher proportion in *Taf4*-deficient mice, are commensals of the gut ecosystem with a metabolism oriented toward SCFA production, whereas *Marvinbryantia*, preferentially found associated with small intestinal microbiota and including species that use cellulose to produce acetate [8], are reduced in *Taf4*-deficient mice. Thus, *Taf4* inactivation in the intestinal epithelium impacts the composition of the gut microbiota by reducing its diversity and favoring the presence of microbes from the upper part of the digestive tract in the distal part, while maintaining SCFA-producing bacteria whose activity should support higher cell proliferation rate of colonocytes. Interestingly, the microbiota changes triggered by *Taf4* invalidation were accompanied by increased paracellular permeability of the epithelium measured in vivo by FITC-Dextran luminal-to-blood transfer, whereas transcellular permeability measured with D-Xylose remained unchanged (Supplementary Fig. 3B). It is worth noting that decreased microbiota diversity is observed in several pathological contexts where the mucosal barrier integrity/functionality is altered [9].

Considering the changes in epithelial barrier activity, microbiota composition and stromal cell composition induced by *Taf4* loss, mice were challenged with a pro-inflammatory stimulation. For this purpose, $Taf4^{IEC}$ and control $Taf4^{lox/lox}$ mice ($n = 5$ in each group) were treated with Tamoxifen and 12 days later were given 2% DSS in drinking water for 5 days before returning to tap water for 3 days. Two $Taf4^{IEC}$ mice exhibited traces of blood in the rectum already before DSS treatment, and 2 mice of this group were euthanized in the 2 days after the end of DSS treatment because they had reached the ethical limit point. Survey of the mice during the course of the experiment revealed a significantly higher clinical score in *Taf4*-deficient mice compared to controls ($p = 0.0009$) (Supplementary Fig. 3C), while the length of the colon was also shorter in mutants at the end of experiment (5.6 cm vs. 8.1 cm, $p = 0.012$). Histological examination showed a nearly normal structure with a regular glandular organization of the distal colon of control mice after the 2% DSS treatment, whereas *Taf4*-inactivated mice exhibited a severely altered colonic mucosa with edema, large areas of immune cell clots, and an irregular single epithelial layer with few remnant glands (Supplementary Fig. 3D). The mutant epithelium showed strongly reduced cell proliferation, high proportion of apoptotic cells labeled with activated Caspase-3 and homogeneous expression of Ephb2 (Supplementary Fig. 3D). These observations demonstrated exacerbated sensitivity of *Taf4*-deficient mice to acute inflammation.

Impact of *Taf4* inactivation on enteroid morphogenesis

To better understand the cell autonomous effects of *Taf4* loss in the intestinal epithelium cells, we developed crypt-derived 3D enteroids and inactivated *Taf4* ex vivo [10]. For this purpose, enteroids were established from the ileum of Tamoxifen-free $Taf4^{IEC}$ and control $Taf4^{lox/lox}$ mice. As expected, enteroids of both genotypes cultured in standard medium generated typical 3D structures with crypt-like buddings lined by a single polarized epithelium. An analogous situation was observed with $Taf4^{lox/lox}$ enteroids treated with 4-Hydroxy-Tamoxifen (4-OHT) during the first 3 days of culture. In contrast, 4-OHT-treated $Taf4^{IEC}$ enteroids resulted in altered 3D structures characterized by a flat cuboidal epithelium, reduced budding outgrowths, and accumulation of cell debris and DNA within the lumen and even outside the spheroids, as illustrated at day 5 of culture (Fig. 4A, B). Ultimately, these disorganized cystic enteroids degenerated, indicating compromised morphogenesis and survival upon Taf4 loss.

The consequence of *Taf4* inactivation on enteroid gene expression was investigated by RNA-seq performed after 3 days of 4-OHT treatment, to capture the most direct effects of *Taf4* loss at a stage when the cellular phenotype was still mild. Compared to control 4-OHT-treated $Taf4^{lox/lox}$ enteroids, treated $Taf4^{IEC}$ enteroids showed 3269 deregulated genes (Supplementary Table 3). KEGG annotation of the 1623 down-regulated genes identified DNA replication/cell cycle and DNA repair as the major pathways affected by *Taf4* loss (Fig. 4C). Specifically, 182 of the 510 genes of the intestinal SC signature [6] were down-regulated, including receptor genes present on CBCs: *Fzd7*, *Fzd2*, *Lgr5*, *Notch1* and *Tnfrsf19*. Genes encoding ligands of these receptors, expressed by Paneth cells in the epithelial SC niche, were also down-regulated: *Wnt3*, *Wnt5a*, *Wnt9b*, *Dll3*, *Jag2*, *Egfl8*. Expression of the *Cdx2* homeobox gene, that dictates intestinal identity in SCs, was perturbed as well as genes encoding differentiation markers of

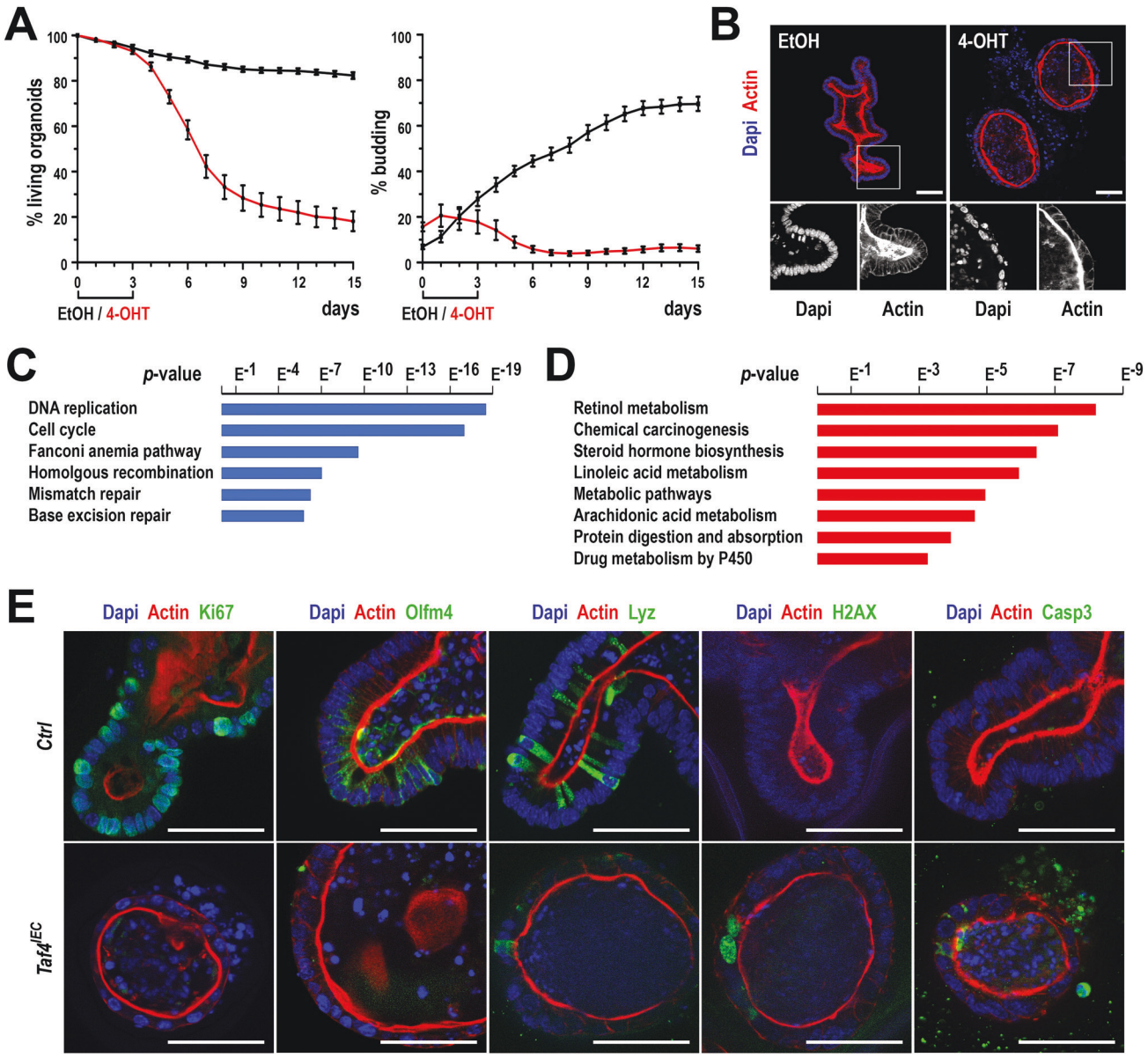


Fig. 4 Effect of *Taf4* inactivation on enteroid morphogenesis. A Survival and budding activity after early *Taf4* gene inactivation. $Taf4^{IEC}$ enteroids were plated and treated 2 h later with 4-OHT in EtOH (red line) or EtOH alone (black line) for 3 days. In total, 50–100 3D structures were counted at each time point. **B** Morphology of the $Taf4^{IEC}$ enteroids at day 5 of culture, that is 2 days after treatment with EtOH (left) or 4-OHT (right). Boxed regions are enlarged below. Nuclei are stained with Dapi (blue) and the actin network with Phalloidin (red). Bars are 50 µm. **C** Top 6 enriched KEGG pathways in downregulated genes in 4-OHT treated $Taf4^{IEC}$ vs. $Taf4^{lox/lox}$ enteroids. **D** Top 8 enriched KEGG pathways in upregulated genes in 4-OHT treated $Taf4^{IEC}$ vs. $Taf4^{lox/lox}$ enteroids. **E** Immunofluorescence detection of the indicated proteins in $Taf4^{lox/lox}$ (Ctrl) and $Taf4^{IEC}$ enteroids treated with 4-OHT at day 5 of culture (2 days after treatment with 4-OHT). Bars are 50 µm.

mature intestinal cell types, among which *Defa17/20/21/24/26* and *Mmp7* for Paneth cells, *Muc4/20* and *Tff2* for goblet cells, *Chga/b* and *Cck* for enteroendocrine cells, and *Aldoa* and *Fabp5* for enterocytes. However, other marker genes were upregulated such as *Cdhr2/5*, *Dpp4*, *Fabp1/2/6*, *Mgam*, *Sis* and *Vill*. KEGG analysis of the upregulated genes identified pathways related to chemical carcinogenesis and metabolic processes (Fig. 4D).

Immunofluorescence at day 5, that is 2 days after the end of 4-OHT treatment, corroborated the transcriptomic results (Fig. 4E). Compared to controls, *Taf4* loss led to a strong reduction of the cell proliferation marker Ki67. The decline in cell proliferation was associated with loss of the SC marker Olfm4 and reduction of Lysozyme in Paneth cells. In addition, considering the number of downregulated genes involved in DNA replication and repair, including *Kdm2b*, *Ppar1* and *Timeless* that signal DNA damage response [11], a strong punctate staining of γH2AX was observed selectively in 4-OHT-treated $Taf4^{IEC}$ enteroids. In line with this, activated Caspase-3-labeled apoptotic cells were detected in $Taf4^{IEC}$ but not in control $Taf4^{lox/lox}$ enteroids.

Thus, *Taf4* loss alters the SC gene expression program and triggers defects in DNA integrity and replication leading to cell apoptosis ultimately eliciting enteroid degeneration and death.

Impact of *Taf4* inactivation on enteroid maintenance

Having demonstrated the important role of Taf4 on enteroid morphogenesis, we asked if it was required for enteroid maintenance, when buddings were already formed with SCs in their niche. For this purpose, enteroids were grown for 5 days in the absence of 4-OHT that was then added to the culture medium for 3 days. Under these conditions, $Taf4^{IEC}$ enteroids progressively lost buddings generated

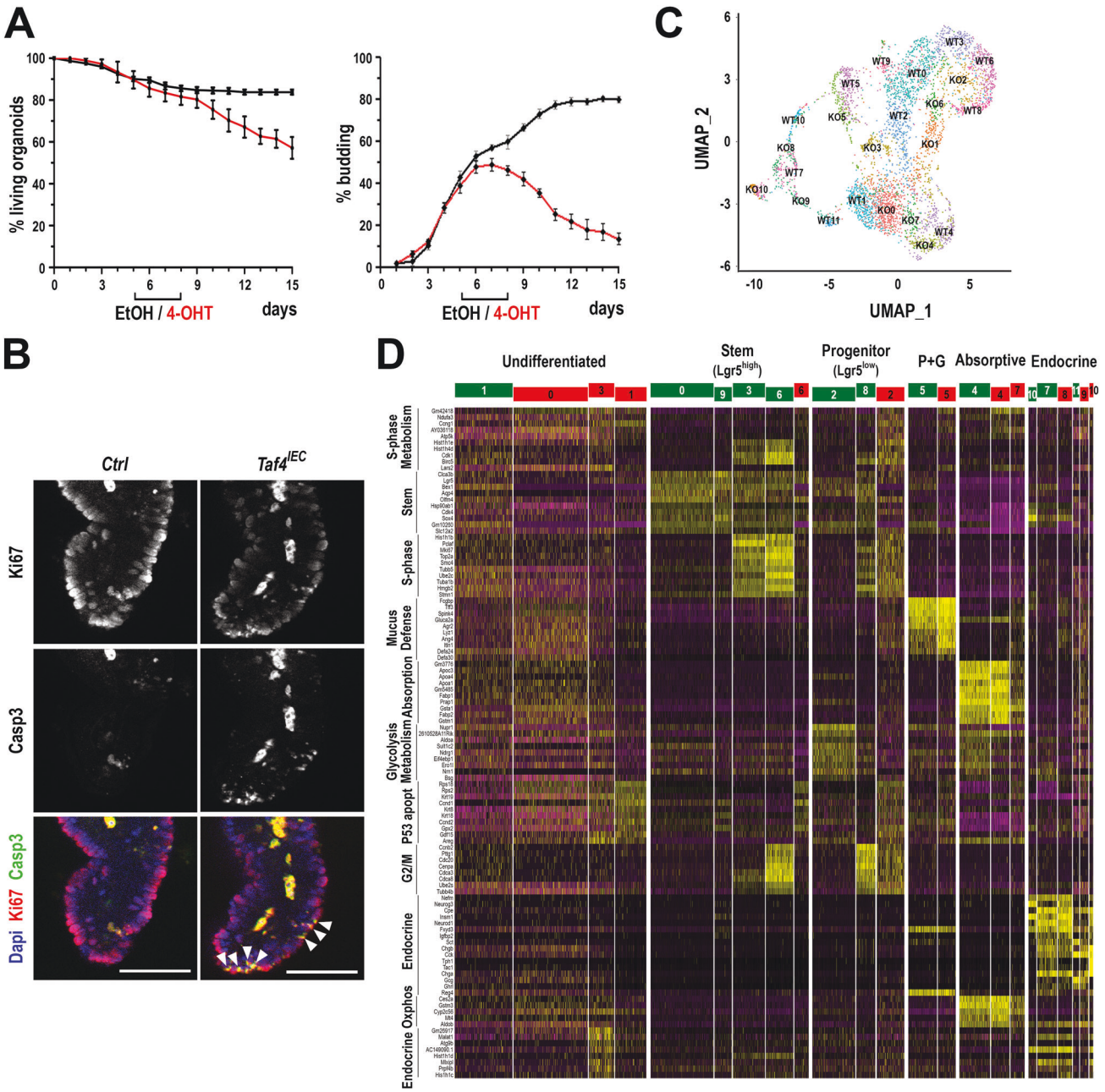


Fig. 5 Effect of *Taf4* inactivation on enteroid homeostasis. A Survival and budding activity after late *Taf4* gene inactivation. $Taf4^{IEC}$ enteroids were plated and treated with 4-OHT (red line) or EtOH (black line) at days 5 to 7 of culture. In total, 50–100 3D structures were counted at each time point. **B** Immunodetection of proliferating cells (Ki67) and apoptotic bodies (activated Caspase-3) at day 8 of culture in 4-OHT-treated $Taf4^{IEC}$ vs. $Taf4^{lox/lox}$ (Ctrl) enteroids. White arrowheads show costaining. Bars are 50 µm. **C** Uniform manifold approximation and projection (UMAP) clustering from scRNA-seq analyses in 4-OHT-treated $Taf4^{IEC}$ and $Taf4^{lox/lox}$ enteroids at day 8 of culture. Cell clusters identified in $Taf4^{lox/lox}$ and $Taf4^{IEC}$ enteroids were labeled WT0 to WT11 and KO0 to K010, respectively. **D** Heatmap of the top markers identified by scRNA-seq in the WT0 to WT11 (green) and KO0 to K010 (red) cell clusters. Absorptive: enterocytes; P + G: Paneth and goblet cells, Endocrine: enteroendocrine cells.

during the first 5 days of culture, stopped growing but remained viable up to day 15, although they could no longer be passaged (Fig. 5A). Immunofluorescence for Ki67 demonstrated reduced cell proliferation in $Taf4^{IEC}$ enteroids (Fig. 5B). Moreover, the presence of apoptotic bodies at the level of Ki67-positive cells was seen by activated Caspase-3 immunodetection, suggesting that proliferative cells enter apoptosis in the absence of Taf4.

To decipher the impact of *Taf4* loss on the different cell populations that make up the enteroid at this stage, we performed single-cell (sc) RNA-seq at day 8 of culture. UMAP representation defined a collection of cell populations from both the 4-OHT-treated control and $Taf4^{IEC}$ enteroids (Fig. 5C) that displayed distinctive expression signatures allowing their identification (Fig. 5D and see Supplementary Datasets 4–8). Based on the level of Lgr5 and Olfm4 expression [12–14], control enteroids exhibited 4 $Lgr5^{high}$ $Olfm4^{high}$ SC clusters, two of them in non-cycling state (WT0, WT9) and two in S and S/G2/M states (WT3, WT6), together with two $Lgr5^{low}$ progenitor clusters, one in S/G2/M state still expressing Olfm4 (WT8), and the other displaying a glycolytic signature without Olfm4 (WT2). Compared to the

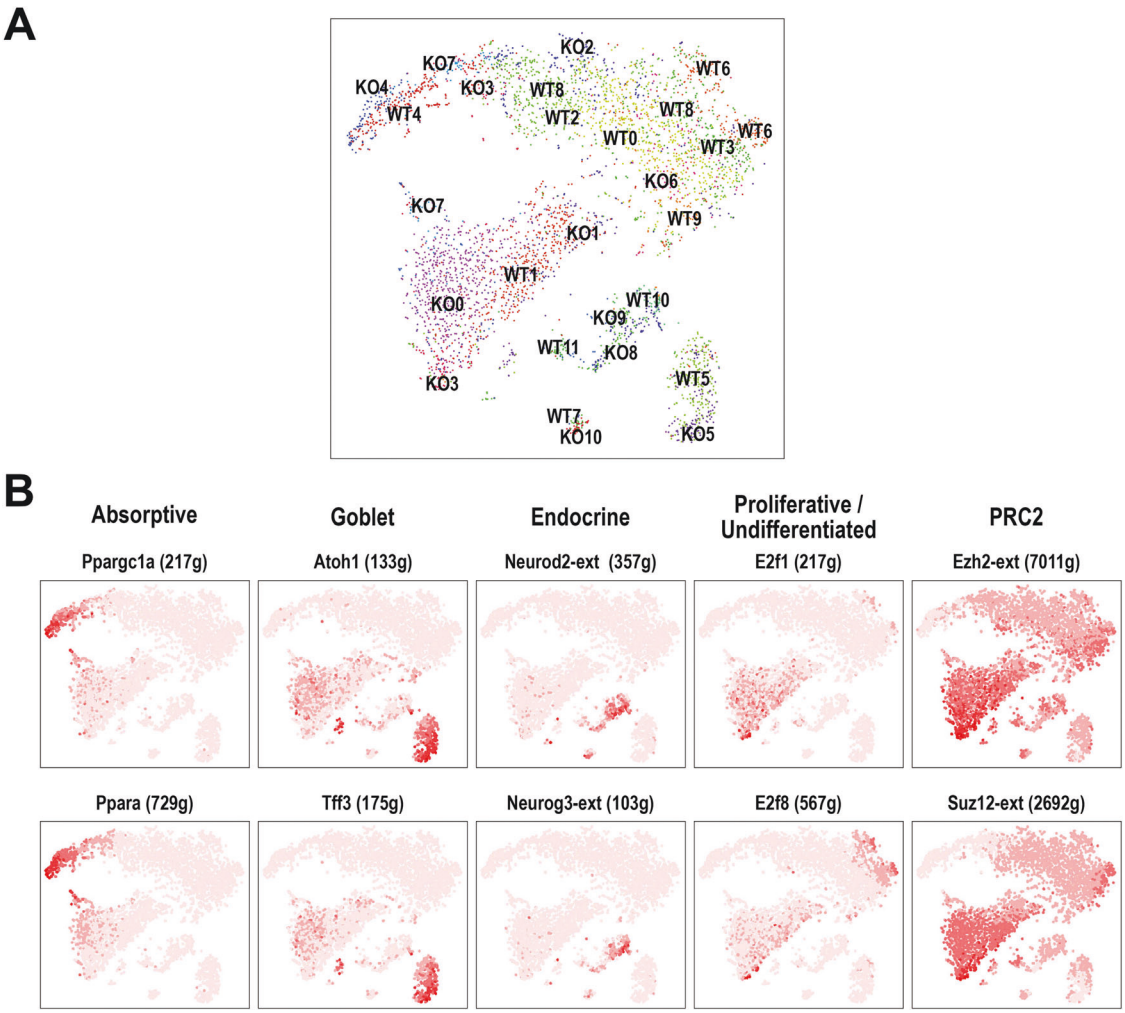


Fig. 6 Regulons changes associated with Taf4 loss. A SCENIC-UMAP of the cell clusters identified by scRNA-seq in 4-OHT-treated *Taf4*$^{lox/lox}$ (WT) and *Taf4*^{IEC} (KO) enteroids. **B** Uncertainty-Aware Face Clustering (AUC) of representative highly active regulons in the indicated cell clusters identified by SCENIC in *Taf4*$^{lox/lox}$ enteroids.

clusters identified in controls, *Taf4*^{IEC} enteroids revealed a strong reduction of Lgr5^{high} SC cells that additionally displayed a p53/apoptotic signature (cluster KO6), and a reduction of Lgr5^{low} progenitors also exhibiting a p53/apoptotic signature (cluster KO2). In addition, *Taf4*-inactivated enteroids showed an increase in cells with no clearly defined identity (undifferentiated; clusters KO0, KO1, KO3 vs. WT1), two of them with a p53/apoptotic signature (KO1, KO3). At this stage of culture, there was no major effect on Paneth and goblet cells (cluster KO5 vs. WT5), whereas a poorly differentiated sub-population of enterocytes emerged (clusters KO4, KO7 vs. WT4). The pattern of enteroendocrine cells was also perturbed, characterized by a reduction of precursor cells and the presence of poorly differentiated cells (clusters KO8, KO9, KO10 vs. WT7, WT10, WT11).

Taf4 loss therefore not only affects enteroid morphogenesis and budding formation, but it also compromises the maintenance of buddings through a strong reduction in stem and progenitor cell populations.

Chromatin conformation changes after *Taf4* inactivation

ATAC-seq performed after early *Taf4* inactivation at day 3 of 4-OHT treatment in *Taf4*^{IEC} vs. *Taf4*$^{lox/lox}$ enteroids provided a broad picture of chromatin remodeling linked to the loss of Taf4 (Supplementary Fig. 4A and Supplementary Table 9). From 118,500 non-redundant peaks present in control and *Taf4*-inactivated enteroids, seq-MINER analysis revealed a cluster of 6874 peaks that were diminished (cluster C7-4) and two clusters of respectively 3798 (cluster 8-1) and 5021 augmented peaks (cluster 8-2) (Supplementary Fig. 4B). DNA-binding motif analysis in these clusters, and in silico footprinting of differentially accessible sites [15] showed a specific enrichment of DNA-binding sites for nuclear receptors (Hnf4α/γ, Pparα/γ, RXRα/γ, Nr2c2, Nr2f6) and Jun- or Fos-containing AP1 dimers in control enteroids and for Zinc finger transcription factors (Gata, Sp2/4, Ctcf) in the Taf4-mutant enteroids (Supplementary Fig. 5C, D). Interestingly, diminished and/or augmented peaks were linked to 585 of the 1623 downregulated genes after Taf4 loss. Gene ontology associated these 585 downregulated genes to cell cycle and division, DNA replication and cell response to DNA damage (Fig. 6E). Among them, 75 were contained in the set of 182 genes of the SC signature downregulated after Taf4 loss, including *Axin2*, *Fzd7*, *Lgr5*, *Notch1* and *Olfm4*.

Rescue of *Taf4*-depleted enteroids by Polycomb inhibition

Taf4 inactivation in enteroids led to loss of the SC and transit amplifying cell compartments with a concomitant increase in undifferentiated cells, accompanied by chromatin remodeling and decreased expression of many SC signature genes consistent with transcriptional reprogramming. The degeneration of *Taf4*-inactivated enteroids indicated the failure of known rescue mechanisms ex vivo. Thus, the mobilization of "+4" or "revival"

cells should be defective despite the increased expression of regenerative markers such as *Tert*, *Kr19* and *Clu* [16–18], as well as the YAP-dependent reprogramming of SCs, even if some genes of the YAP pathway for intestinal repair were modified, including the decrease of *Reg3a*, *Olfm4*, *Wnt3*, *Aqp4, Lgr5* and *Axin2* and the increase of *Areg*, *Il1rn* and *Msln* [19] (Supplementary Table 2). In addition, *Ascl2* expression was markedly reduced in *Taf4*-deficient enteroids preventing *Ascl2*-dependent replenishment of the SC compartment from progenitors [20] (Supplementary Table 2). To address the mechanism underlying enteroid degeneration, we performed single-cell regulatory network inference and clustering (SCENIC) analysis on the control and *Taf4*-inactivated cells (Fig. 6A). Several of the cell populations were strongly marked by activity of factors already shown to be important for their identity and/or specification (Fig. 6B): Atoh1 and Tff3 for Goblet cells, Neurod2 and Neurog3 for enteroendocrine cells, and nuclear receptors such as Ppara and the master regulator of metabolism Ppargc1a for enterocytes in agreement with their high OXPHOS signature. The undifferentiated cells and transit amplifying cells were also marked by activity of E2f factors consistent with their cell cycle signatures. Nevertheless, SCENIC did not identify transcription factor regulons strongly active in the SCs. Instead, we noted that the SCs displayed low Ezh2 and Suz12 activity that was strongly increased in the control and *Taf4*-inactivated undifferentiated cells suggesting that increased PRC2 complex activity played a role in suppressing SC identity. As the undifferentiated cell population increased at the expense of the SC compartment, we thus postulated that Taf4 loss promoted PRC2 activity leading to suppression of SC identity thereby favoring the undifferentiated state.

To test this idea, 4-OHT-treated $Taf4^{IEC}$ and control $Taf4^{lox/lox}$ enteroids were treated with EPZ6438, a selective inhibitor of the Ezh2 histone methyltransferase of PRC2. EPZ6438 was added at the same time as 4-OHT and maintained up to day 15. When *Taf4* was inactivated early after plating, EPZ6438 promoted survival and budding in $Taf4^{IEC}$ enteroids (Fig. 7A, B and Supplementary Fig. 5A). EPZ6438 treatment also rescued budding degeneration by late *Taf4* inactivation at day 5 of culture (Fig. 7C and Supplementary Fig. 5B). Noteworthy, immunofluorescence demonstrated that the buddings restored by EPZ6438 in *Taf4*-inactivated $Taf4^{IEC}$ enteroids treated with 4-OHT were populated with proliferating cells (Ki67), SCs (Olfm4) and Paneth cells (Lyz) (Fig. 7D). This result was further extended by RNA-seq (Supplementary Table 10). Indeed, among the 1623 genes downregulated by *Taf4* inactivation, the expression of 1146 (70.6%) was restored by adding EPZ6438 ($p < 0.05$), of which 949 (58.5%) with a Fold Change > 2. Similarly, among the 1646 genes upregulated after Taf4 loss, 1166 (70.8%) were decreased by EPZ6438, of which 947 (57.5%) with a ($|\log_2(FC)| > 1$. Importantly, expression of 115 of the 182 genes of the SC signature down-regulated by the loss of Taf4 were restored by EPZ6438 (Fig. 7E).

To determine if *Taf4* loss also affected PRC2 function in the developing intestinal epithelium and adult mucosa, we performed immunohistochemistry for the PRC2-dependent histone mark H3K27me3 (Supplementary Fig. 6). Control E18.5 embryos showed low H3K27me3 staining in the intervilli regions and progressively increasing levels along the villi. Similarly, adults exhibited increasing H3K27me3 levels along the villi, with few cells labeling at the bottom of the crypts. In contrast, H3K27me3 staining was strong in all of the cells making up the flat developing epithelium in $Taf4^{IEndoC}$ embryos consistent with the lack of SC emergence. In adult $Taf4^{IEC}$ mice, increased H3K27me3 levels were seen in both the villi and most strikingly in cells at the crypt bottom compared to controls, in line with reduced SC activity. Thus, consistent with the stimulation of PRC2 activity after *Taf4* inactivation in enteroids, *Taf4* gene inactivation during intestinal morphogenesis in embryos and in adult intestinal homeostasis resulted in elevated levels of PRC2-deposited H3K27me3 and impaired SC development or function.

Taf4 inactivation enhances *Apc*-driven tumorigenesis

Since *Taf4* inactivation led to increased PRC2 activity in crypt SCs, that in enteroids resulted in the increased pool of undifferentiated cells, we asked if this alteration may affect intestinal tumorigenesis in the tumor prone model of $Apc^{+/\Delta 14}$ mice. In a cohort of males ($n > 10$ in each group), *Taf4* loss in $Taf4^{IEC}$::$Apc^{+/\Delta 14}$ mice reduced overall survival and aggravated tumor burden compared to $Apc^{+/\Delta 14}$ mice (Fig. 8A, B). $Taf4^{IEC}$::$Apc^{+/\Delta 14}$ tumors were histologically similar to those of $Apc^{+/\Delta 14}$ mice (Fig. 8C), and exhibited a comparable distribution of proliferative cells labeled with Ki67, and a similar intratumor heterogeneity as illustrated by the patterns of Cdx2 and Hnf4α (Fig. 8D). RNA-seq of the tumors showed 699 down-regulated genes and 122 up-regulated genes upon *Taf4* inactivation (Supplementary Table 11). GSEA and KEGG analyses revealed a strong decrease in terms related to allograft-rejection and IFNα, IFNγ, IL2 / STAT5 and IL3 / STAT3 pathways together with increased of Wnt/β-catenin signaling (Fig. 8E, F). The down-regulation of these immune/inflammatory pathways was associated with an altered tumor immune-microenvironment. In *Taf4* depleted tumors, MCP counter showed an increase in Mast cells/Basophils and granulocytes, but a decrease of T cells amongst which CD8+ cytotoxic T lymphocytes (Fig. 8G), supported by reduced levels of their key markers Cd8a, Pdcd1, Tigit, and Lag3 (Supplementary Table 11). The striking reduction of the interferon / inflammation pathways and T cells in tumors mirrored the changes already seen in the normal gut mucosa. Indeed, more than 33% (247/699) of the down-regulated genes in tumors were also down-regulated in the non-tumoral mucosa upon *Taf4* inactivation (compare Supplementary Tables 2 and 11), and these 247 genes were associated with KEGG pathways designating hematopoietic cell lineages, Th1 and Th2 cell differentiation, antigen processing, graft-versus-host disease, and allograft rejection (Fig. 8H). *Taf4* loss both in non-tumoral gut mucosa and in intestinal tumors therefore led to an altered immune environment. In particular, both situations were characterized by reduced gene expression of *Cd7*, a marker of mature CD8+ T cells, of *Cd8a* itself, of the granzymes *Gzma* and *Gzmb* and of *Prf1* for cytotoxic T lymphocytes, together with the repression of the IFN gamma pathway, a major regulator of colon tumor immunity [21]. Moreover, expression of cytokine Cxcl11 that attracts cytotoxic T lymphocytes [22] was reduced, whereas expression of Interleukin-17 receptor (Il17Rb) was increased. The Il17-IL17Rb axis is known to inhibit cytotoxic T lymphocyte recruitment in mouse models of intestinal cancer [23, 24]. Together these results revealed a novel facet of Taf4 function in influencing the immune microenvironment in both normal mucosal tissue and in the tumor context.

DISCUSSION

Taf4 is a critical regulator of development and homeostasis of the intestinal epithelium

This study defines the important role played by the Taf4 component of TFIID in intestinal development and homeostasis. Taf4 loss had not only cell-autonomous effects, but also affected gut permeability, the immune microenvironment and the composition of the luminal microbiota ultimately impacting the response to pro-inflammatory and pro-tumoral stimuli. Taf4 was crucial for the emergence of adult-type intestinal SCs during late embryogenesis and participated in their maintenance throughout adulthood together with differentiation of mature cells. The critical role of Taf4 in SCs was further confirmed in enteroids where its loss depleted the SC and transit amplifying compartments.

Although Taf4 is a subunit of a major component of the general transcription machinery, its loss of function affected only a subset of genes in the gut epithelium, in agreement with previous observations in liver, pancreas and epidermis [3–5]. We previously suggested that these limited effects on gene expression were in part explained by compensation by its paralog Taf4b [2, 25] that integrates and maintains the integrity of TFIID. This mechanism is

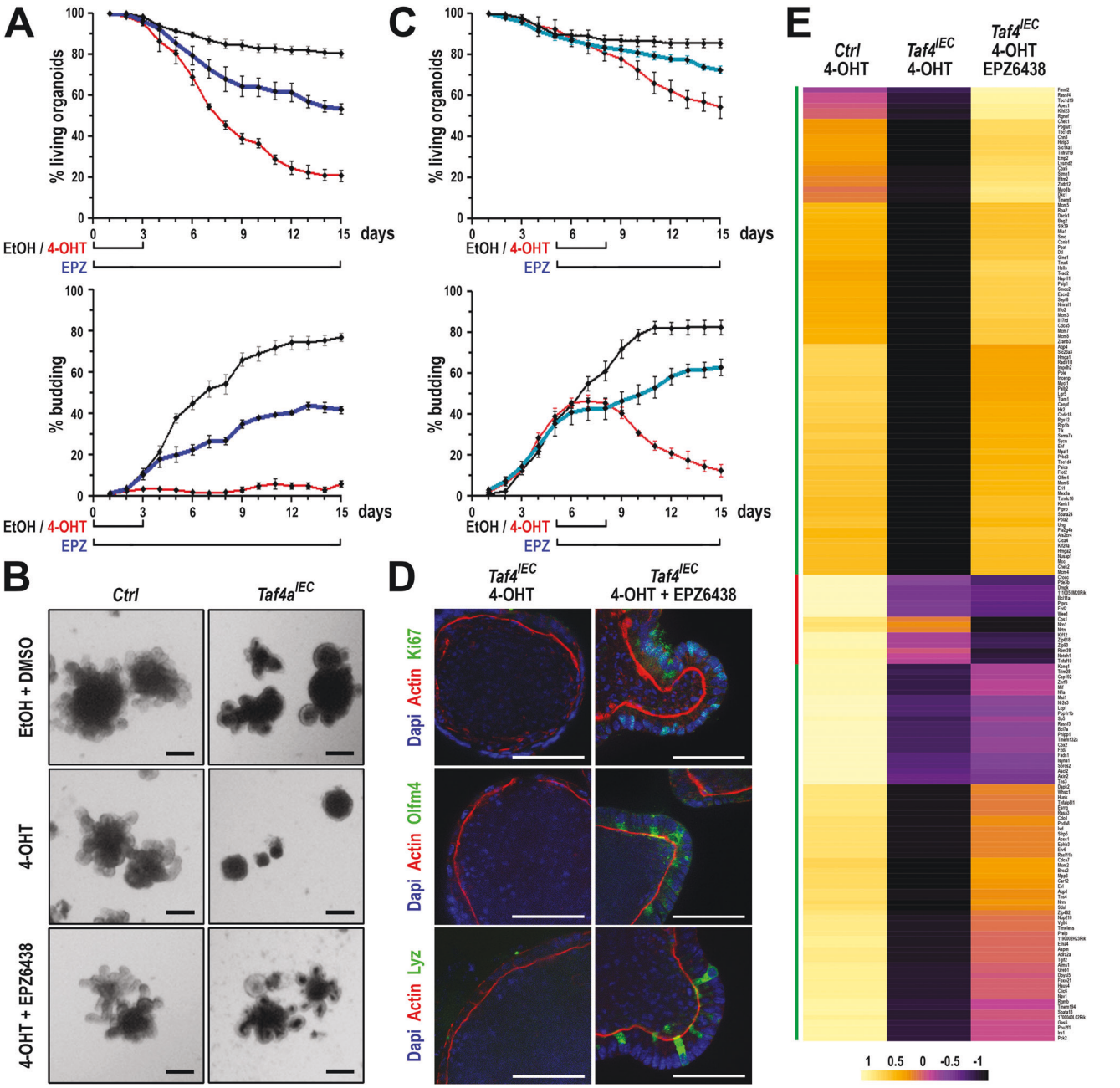


Fig. 7 Rescue of *Taf4* inactivated enteroids by Polycomb complex inhibition. A Survival and budding activity in *Taf4*^{IEC} enteroids treated with EtOH + DMSO (black line), with 4-OHT (2–72 h after plating), or 4-OHT (2–72 h after plating) + EPZ6438 (2 h to 15 days after plating) (red line). In total, 50-100 3D structures were counted at each time point. **B** Morphology at day 15 of culture of *Taf4*$^{lox/lox}$ (ctrl) and *Taf4*^{IEC} enteroids treated with EtOH + DMSO, or 4-OHT (2–72 h after plating), or 4-OHT (2–72 h after plating) + EPZ6438 (2 h to 15 days after plating). Bars are 200 µm. **C** Survival and budding activity in *Taf4*^{IEC} enteroids treated with EtOH + DMSO (black line), with 4-OHT (days 5-8 after plating) (red line), or with 4-OHT (days 5-8 after plating) + EPZ6438 (days 5-15 after plating) (blue line). In total, 50–100 3D structures were counted at each time point. **D** Immunostaining of the indicated proteins in *Taf4*^{IEC} enteroids treated 4-OHT or with 4-OHT + EPZ6438 at day 15 of culture. Bars are 50 µm. **E** Heatmap of the 182 genes of the stem cell signature that are downregulated by Taf4 loss in 4-OHT treated *Taf4*^{IEC} enteroids compared to treated *Taf4*$^{lox/lox}$ control enteroids, and restored by treatment with the Ezh2 inhibitor EPZ6438.

also likely operative in the intestine epithelium since RNA-seq data in both fetal and adult intestine indicated a mild but significant increase of *Taf4b* expression upon *Taf4* loss.

In vivo, the defects observed upon *Taf4* loss demonstrated its involvement in the proper differentiation of mature cells of both absorptive and secretory lineages and in the control of the epithelial turnover and barrier function. Increased paracellular permeability did not reflect expression changes of cell-cell adhesion components, but correlated with a decrease of several integrin genes including *Itgb6* that encodes a subunit of αvβ6 participating in the barrier activity of intestinal epithelial cells [26]. Several of the genes downregulated in the developing and adult gut are shared with *Taf4*-deficient neonatal hepatocytes, including metabolic genes. In addition, ATAC-seq peaks diminished upon *Taf4* loss in intestinal enteroids were enriched in DNA-binding sites for Hnf4, a major regulator of metabolic genes previously shown to interact with Taf4 in hepatocytes [4]. Together, the data obtained in the intestine and liver suggest that Taf4 coordinates

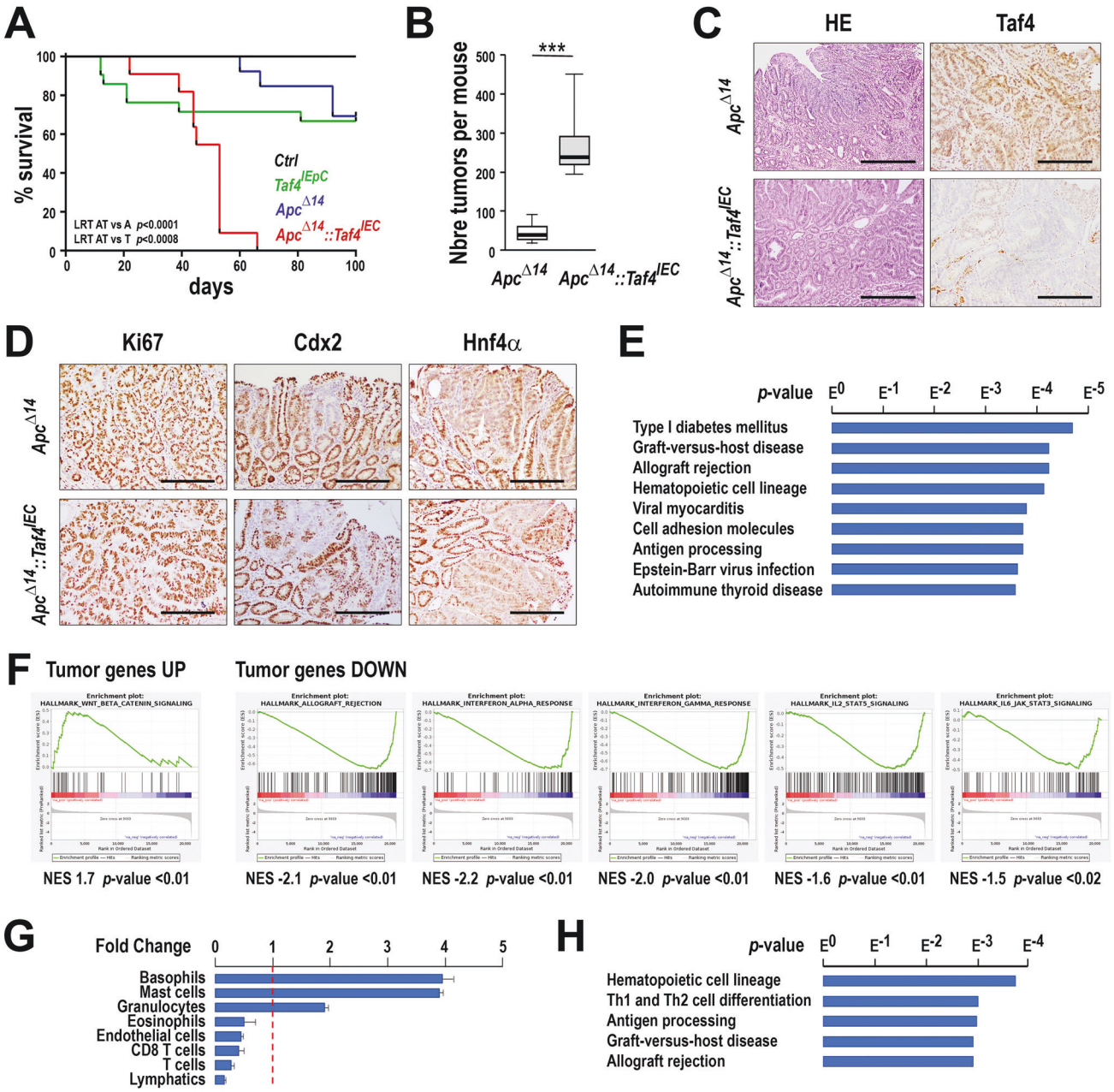


Fig. 8 Impact of *Taf4* inactivation on intestinal tumor development. A Overall survival of $Apc^{+/\Delta 14}$::$Taf4^{IEC}$, $Apc^{+/\Delta 14}$, $Taf4^{IEC}$ and $Taf4^{lox/lox}$ (Ctrl) males after Tamoxifen injection; $n > 10$ for each genotype. LRT AT vs. A: Logrank test between $Apc^{+/\Delta 14}$::$Taf4^{IEC}$ and $Apc^{+/\Delta 14}$ mice; LRT AT vs. T: Logrank test between $Apc^{+/\Delta 14}$::$Taf4^{IEC}$ and $Taf4^{IEC}$ mice. **B** Tumor number in the small intestine of $Apc^{+/\Delta 14}$::$Taf4^{IEC}$ and $Apc^{+/\Delta 14}$ mice. ***$p < 0.0001$. **C** Histology (HE) and immunodetection of Taf4 protein in ileal tumors of $Apc^{+/\Delta 14}$::$Taf4^{IEC}$ and $Apc^{+/\Delta 14}$ mice. Bars are 400 µm for HE and 200 µm for Taf4. **D** Immunohistochemical detection of Ki67, Cdx2 and Hnf4α in $Apc^{+/\Delta 14}$::$Taf4^{IEC}$ and $Apc^{+/\Delta 14}$ tumors. Bars are 200 µm. **E** KEGG ontology enrichment is shown for the 699 downregulated genes in $Apc^{+/\Delta 14}$::$Taf4^{IEC}$ vs. $Apc^{+/\Delta 14}$ mice and ordered according to the *p* value. **F** GSEA analysis of the up-regulated and downregulated genes in $Apc^{+/\Delta 14}$::$Taf4^{IEC}$ vs. $Apc^{+/\Delta 14}$ mice identifying respectively hallmarks for Wnt/β-catenin signaling and for allograft rejection and IFNα, IFNγ, IL2/STAT5 and IL6/STAT3 signaling. **G** Stromal cell population evaluation from RNA-seq data using the MCP method, expressed as the proportion of each cell type in $Apc^{+/\Delta 14}$::$Taf4^{IEC}$ compared to $Apc^{+/\Delta 14}$ mice. **H** KEGG ontology enrichment of the 247 downregulated genes in common between $Apc^{+/\Delta 14}$::$Taf4^{IEC}$ vs. $Apc^{+/\Delta 14}$ mice and $Taf4^{IEC}$ vs. wild type mice.

metabolic pathways by promoting Hnf4 binding and maintaining chromatin accessibility at key metabolic genes.

A novel role of Taf4 in antagonizing PRC2 activity and regulating the immune-microenvironment

A key finding of our study is the novel role for Taf4 in antagonizing the activity of the PRC2 complex to maintain the SC gene expression program and the stem/progenitor cell compartment as highlighted in vivo during embryonic development and in the adult intestine, and ex vivo in enteroids. In the fetus, *Taf4* loss increased PRC2 activity in all cells lining the epithelium consistent with the impaired emergence of adult-type SCs, whereas in adults some features of the already established SCs were perturbed, for example diminished *Olfm4* expression, but a limited SC activity was preserved allowing cell turnover. The difference between fetuses/enteroids and adults suggests that extrinsic signals from

the adult SC niche, likely insufficient during development and in enteroids, may partially compensate for Taf4 deficiency and maintain stemness, and thus that an integrated and fine-tuned epithelial-mesenchymal interaction is necessary to properly switch on the regenerative process when SCs are affected at the adult stage [27]. In line with this, transcriptomic data indicate that the level of *Wnt3*, an important Wnt family ligand for adult SCs [10, 28], was reduced in *Taf4*-deficient fetuses and enteroids, but not in the adult epithelium.

Our results reveal that inhibiting the methyltransferase activity of the Ezh2 component of PRC2 rescued enteroid survival, budding formation and cell proliferation, and restored the SC compartment with active SCs and supporting Paneth cells, all of which were compromised in the absence of Taf4. This was accompanied by the rescued expression of a large proportion of genes deregulated by *Taf4* loss, including genes of the intestinal SC signature. Previous studies showed how SWI/SNF recruitment by transcription factors to both enhancers and proximal promoters upon gene activation antagonizes PRC2 activity [29–31]. Our results extend this model by showing that the Taf4 subunit of TFIID, a key component of the promoter-bound pre-initiation complex and present at enhancers, also antagonizes the ability of PRC2 to supress the expression of the SC gene expression program. Yet, Taf4 is necessary to overcome the repression of this program by PRC2, but not for its expression per se that takes place in absence of Taf4 when Ezh2 is inhibited.

PRC1 and 2 are major regulators of epigenetic silencing in developmental processes and pathologies including cancer and inflammatory diseases [32, 33] and are involved in the control of intestinal crypt homeostasis and regeneration after damage [34, 35]. In enteroids, the SC and progenitor compartments were marked by low activity of Ezh2 and Suz12 in agreement with previous studies showing that PRC2 is essential for SC maintenance and proliferation [34–36]. In addition, differentiation of Paneth cells that provide ligands required for SC functions and emergence in enteroids [37] is also dependent on the level of Ezh2/PRC2 activity [38]. Finely tuned PRC2 activity in SCs and in the niche Paneth cells is therefore essential for proper homeostasis. Our observations in enteroids suggest that SCs exist in equilibrium with the undifferentiated population controlled by a competition between Taf4-driven expression of the SC program and its PRC2-mediated repression. Upon *Taf4* inactivation, this equilibrium is upset with increased PRC2 activity shutting down the SC program and generating undifferentiated cells. We observed a similar effect in the developing epithelium and in adult mucosa where enhanced levels of H3K27me3 were observed in crypt SCs upon *Taf4* inactivation, in line with the idea of increased PRC2 activity impairing SC function in vivo.

An additional consequence of *Taf4* loss in the adult mucosa, in *Apc*-driven tumors and, to a lesser extent, in the embryonic intestine is a modified inflammatory/immune microenvironment. These changes reflected altered immune composition that was most prominently seen in the tumors characterized by higher numbers of mast cells/basophils but reduced numbers of T lymphocytes confirmed by the strong reduction in their key markers and by strongly reduced IFNα/γ signaling. High levels of mast cells have been associated with increased tumor growth in human colon cancer [39], whereas a low rate of cytotoxic T cell recruitment correlates with compromised immunosurveillance and poor prognosis in colon cancer [21, 40]. Hence the increased number of tumors seen upon *Taf4* inactivation likely results from reduced immunosurveillance. Several lines of evidence suggest that the increased PRC2 activity seen upon *Taf4* inactivation may be linked to the altered mucosal and tumor immune microenvironment. Previous studies relate increased Ezh2 expression and PRC2 activity to immunoediting of the tumor environment [41, 42]. Specifically, increased PRC2 activity in colon cancer cells reduced T-cell migration to tumors [43] consistent with our observations. IFNγ signaling and antigen presentation is also consistent with the reduced MHC class II seen in the *Taf4*-null tumors [21, 41]. We therefore propose that increased PRC2 activity in the adult intestinal mucosa alters its immune microenvironment providing an environment that promotes *Apc*-driven tumorigenesis through reduced immune surveillance. Changes in the immune microenvironment, in particular the reduction of IFNα signaling which is protective in acute colitis [44], may also be involved in the hypersensitivity of *Taf4*-deficient mice to DSS-induced inflammatory injury.

Together, our observations describe for the first time a novel function of Taf4 that antagonizes PRC2-mediated repression of the SC gene expression program to assure normal development, homeostasis, and immune microenvironment of the intestine epithelium.

MATERIAL AND METHODS

Mice and treatments

Mice experiments were performed in the certified animal facility (#G/H-67-482-21) according to the protocol approved by the French Ministry of Agriculture under the permit APAFiS #14197. $Taf4^{lox/lox}$ [25], *VilCre* and $VilCreER^{T2}$ [45], $Lgr5\text{-}GFP\text{-}CreER^{T2}$ [12], and $Apc^{\Delta 14/+}$ [46] mice have been described. Animals were genotyped by PCR on tail DNA with the following primers: $Taf4^{lox/lox}$ allele CTAGTTACTGCTCTGCACAAT/GTGCTCCATGACTC TGGCAAG/CAGCCAAAGCTACATAATAAGT; *VilCre* and $VilCreER^{T2}$ alleles CAAGCCTGGCTCGACGGCC/CGCGAACATCTTCAGGTTCT; $Lgr5\text{-}GFP\text{-}CreER^{T2}$ allele CTGCTCTCTGCTCCCAGTCT/ATACCCCATCCCTTTTGAGC/GAACTTCAG GGTCAGCTTGC; Apc^{wt} allele CTGTTCTGCAGTATGTTATCA/CTATGAGTCAAC ACAGGATTA; $Apc^{\Delta 14}$ allele CTGTTCTGCAGTATGTTATCA/TATAAGGGCTAA CAGTCAATA.

For conditional inactivation of the *Taf4* gene, $Taf4^{lox/lox}$::$VilCreER^{T2}$ mice, $Taf4^{lox/lox}$::$Lgr5\text{-}GFP\text{-}CreER^{T2}$ mice or $Taf4^{lox/lox}$::$VilCreER^{T2}$::$Apc^{\Delta 14/+}$ mice aged 2–3 months received intraperitoneal injections of 1.6 mg Tamoxifen (Sigma-Aldrich) in corn oil, once daily for 3 days. Controls, either wild types or $Taf4^{lox/lox}$ or $VilCreER^{T2}$ or $Lgr5\text{-}GFP\text{-}CreER^{T2}$ or $Apc^{\Delta 14/+}$, also received Tamoxifen.

Animals were euthanized when body weight loss reached 20% of initial body weight.

BrdU pulse-chase labeling experiments were performed on $Taf4^{lox/lox}$::$VilCreER^{T2}$ and control $Taf4^{lox/lox}$ mice treated 10 days earlier with Tamoxifen and injected (day 0) with a single dose of 1 mg BrdU (Sigma-Aldrich). Animals were euthanized at day 1, 2 or 3 after BrdU administration.

Paracellular and transcellular permeability was determined by measuring the serum levels of fluorescein isothiocyanate conjugated dextran (FITC-Dextan 4.4 kDa, Sigma-Aldrich) and D-Xylose (Sigma-Aldrich) after gavage. Briefly, 10 days after Tamoxifen treatment, $Taf4^{lox/lox}$::$VilCreER^{T2}$ and control $Taf4^{lox/lox}$ mice were starved during 4 h and then force-fed with 500 mg/kg FITC-Dextran or 2 g/kg D-Xylose in PBS pH 7.4. Blood samples were taken 3 h after gavage by submandibular collection. The serum levels of FITC-Dextran and D-Xylose were respectively measured by direct spectrophotofluometry and with the D-Xylose Kit (Chrondrex, 6601) coupled with spectrophotometry, using the spectrophometer TriStar Multimode reader LB 942 (BERTHOLD Technologies).

For gut inflammation studies, $Taf4^{lox/lox}$::$VilCreER^{T2}$ and control $Taf4^{lox/lox}$ mice 12 days after the Tamoxifen treatment were given 2% Dextran Sulfate Sodium (DSS 36–50 kDa, MP Biomedicals, Illkirch, France) in drinking water for 7 days and then tap water for 3 days. Throughout the experiment, mice were daily injected with 20 mg/kg Lurocaine (Centravet, Lapalisse, France, LUR003). The presence of blood in the stools was analyzed using HemoCARE (Care Diagnostic, Voerde, Germany). The clinical score was determined daily based on body weight loss, stool consistency and blood in the stools, as described [47]. At the end of the experiment, the colon was removed, flushed with PBS, measured, mounted as Swiss Roll, fixed in 4% paraformaldehyde for 4 h and embedded in paraffin. Assessment of inflammation was performed with regard to stiffness, edema, ulcerations and thickness.

Enteroid cultures and treatments

Enteroid cultures were established from ileal crypts of 4-month-old $Taf4^{lox/lox}$::$VilCreER^{T2}$ and control $Taf4^{lox/lox}$ mice, not treated with

Tamoxifen. Ileal fragments were incubated in Gentle Cell Dissociation Reagent (STEMCELL™, 07174) for 15 min, then ~50 crypts were embedded in 20 µl of Matrigel® (Corning®, #356231) in 48-wells plates (Greiner Bio-one, 677180) and grown in 250 µl Mouse IntestiCult™ Organoid Growth Medium (STEMCELL™, 06005). Enteroids were passed every week after mechanical breakage with a 200 µl pipette and dilution at 1:4 for maintenance and RNA extraction or at 1:8 for immunolabelling and kinetics studies. Experiments were performed on enteroids established for at least 5 passages. For early *Taf4* gene inactivation, enteroid fragments were plated for 2 h and then treated for 3 days with 1 µM (Z)-4-hydroxytamoxifen (4-OHT, Sigma-Aldrich, H7904) in EtOH or with EtOH alone for control experiments. During the 3-days treatment, fresh IntestiCult™ medium with 4-OHT was changed every day and then every 2 days without 4-OHT up the end of the experiment. For late *Taf4* gene inactivation, enteroids were grown for 5 days in IntestiCult™ medium, then treated for 3 days with 4-OHT in EtOH or EtOH alone as described above, and then cultured in IntestiCult™ medium without 4-OHT up the end of the experiment. When indicated, EPZ6438 at 1 µM (MedChemExpress, HY-13803) in DMSO was added to the culture medium at the same times as 4-OHT and then changed with fresh IntestiCult™ medium up to the end of the experiment.

Immunostaining of tissue samples and enteroids

Antibodies used for this study are listed in Supplementary Table 12.

Intestinal samples taken from mouse embryos and adult animals were fixed in 4% paraformaldehyde and embedded in paraffin, and then analyzed by immunohistochemistry and immunofluorescence detection as previously described [48]. To determine the proportion of labeled cells, at least three pictures were taken per animal in five animals and cell counting was performed on a minimum of 1000 cells per picture. Pictures were taken either with an Axiophot microscope or an Axio Imager Z2 microscope (Zeiss).

Enteroids grown in IntestiCult™ medium and Matrigel in 8-wells Lab-Tek® Chamber Slide™ (Dominique Dutscher) were directly fixed for 30 min with 4% paraformaldehyde, 60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 2 mM magnesium acetate and permeabilized for 30 min in 0.5% Triton X-100 (Euromedex). After blockade for 2 h at 37 °C in 5% BSA (Euromedex), primary antibodies were added and incubated 2 h at room temperature followed by overnight at 4 °C. Secondary antibodies were incubated for 2 h at 37 °C. Nuclei were stained with 40,6-diamidino-2-phenylindole dihydrochloride (DAPI) and actin was revealed by Phalloidin-TRITC (Sigma-Aldrich, P1951). Samples were mounted in ProLong™ Gold Antifade Mountant (Life Technologie™, P36930) and analyzed with an Axio Zoom.V16 microscope (Zeiss) for stereomicroscopy or with an Axio Imager M2 microscope coupled to a Hamamatsu's camera Orca Flash 4v3 using the ApoTome.2 function (Zeiss) for optical sectioning.

Bacterial 16S RNA analysis

The luminal content was collected in the cecum of 8 5-month-old *Taf4*$^{lox/lox}$*::VilCreER*^{T2} and 8 control *Taf4*$^{lox/lox}$ mice treated with Tam at the age of 3 months. DNA extraction was performed using NucleoSpin® DNA Stool kit (Macherey-Nagel—740472.5) adapted with a mechanical lysis step (Fastprep—6.5 m s^{-1} for 2 min). The V3–V4 region of the 16S rRNA gene was amplified with the primers TACGGRAGGCAGCAG/ATCTTACCAGGGTATCTAATCCT according to GeT-PlaGe platform protocol (INRAE). Sequencing was performed on Illumina MiSeq system using 2*300 bp paired-end mode. For sequence data analysis, remaining adapter/primer sequences were trimmed and reads were checked for quality (≥20) and length (≥200 bp) using Cutadapt [49]. Reads were further corrected for known sequencing errors using SPAdes [50] and then merged using PEAR [51]. OTU were identified using the Vsearch pipeline [52] set up to dereplicate, cluster, chimera check the merged reads. OTU taxonomical classification was performed using classifier from the RDPTools suit [53]. Statistical tests were performed using Wilcoxon Rank Sum Test for group comparison. Multiple tests were corrected using the False Discovery Rate method (*q* value) as required.

Bulk RNA preparation and RNA-seq

RNA was extracted from the small intestine of 3 E17.5 *Taf4*$^{lox/lox}$*::VilCre* and 3 control *Taf4*$^{lox/lox}$ littermates and from 3 adult *Taf4*$^{lox/lox}$*::VilCreER*^{T2} and 3 control *Taf4*$^{lox/lox}$ mice 10 days after Tamoxifen administration. RNA was also extracted from 3 wells of 4-OHT-treated *Taf4*$^{lox/lox}$*::VilCreER*^{T2} and control *Taf4*$^{lox/lox}$ enteroids at day 3 of culture, and from 3 wells of *Taf4*$^{lox/lox}$*::VilCreER*^{T2} enteroids at day 15 of culture after treatment with 4-OHT and EPZ6438. RNA preparation used Tri Reagent (Euromedex) and the quality was analyzed using nanoRNA chips on a Bioanalyser 2100 (Agilent Technologies). Complementary DNA was generated and linearly amplified from 3 ng total RNA using the Ovation RNA-seq V2 system (NuGEN technologies Inc., Leek, The Netherlands), according to the manufacturer's instructions. The amplified cDNA was then purified using Agencourt AMPure XP beads (Beckman-Coulter, Villepinte, France) in a 1.8:1 bead to sample ratio and fragmented by sonication using a Covaris E220 instrument (with duty cycle: 10%, maximum incident power: 175 watts and cycles/burst: 200 for 120 s). The RNA-seq libraries were generated from 100 ng fragmented cDNA using the Ovation Ultralow v2 library system (NuGEN technologies Inc., Leek, The Netherlands) according to the manufacturer's instructions, with 6 PCR cycles for library amplification. The final libraries were verified for quality and quantified using capillary electrophoresis before sequencing on an Illumina Hi-Seq4000.

Reads were preprocessed to remove adapter and low-quality sequences (Phred quality score below 20). After this preprocessing, reads shorter than 40 bases were discarded from further analysis. These preprocessing steps were performed using Cutadapt version 1.10 [49]. Reads were mapped to rRNA sequences using Bowtie version 2.2.8 [54], and reads mapping to rRNA sequences were removed for further analysis. Reads were mapped onto the mm9 assembly of Mus musculus genome using STAR version 2.5.3a [55]. Gene expression quantification was performed from uniquely aligned reads using Htseq-count version 0.6.1p1 [56], with annotations from Ensembl version 67 and "union" mode. Only non-ambiguously assigned reads have been retained for further analyses. Read counts have been normalized across samples with the median-of-ratios method [57], to make these counts comparable between samples. Comparisons of interest were performed using the Wald test for differential expression and implemented in the Bioconductor package DESeq2 version 1.16.1 [58]. Genes with high Cook's distance were filtered out and independent filtering based on the mean of normalized counts was performed. *p* values were adjusted for multiple testing using the Benjamini and Hochberg method [59]. Heatmaps were generated with R-package pheatmap v1.0.12. Deregulated genes were defined as genes with $\log_2$(FoldChange) >1 or < −1 and adjusted *p* value <0.05.

Assay for transposase-accessible chromatin (ATAC-seq)

ATAC-seq was performed at day 3 of culture from 20,000 cells of *Taf4*$^{lox/lox}$*::VilCreER*^{T2} and control *Taf4*$^{lox/lox}$ enteroids treated with 4-OHT. Sequenced reads were mapped to the mouse genome assembly mm9 using Bowtie [54] with the following arguments: "-m 1 -strata -best -y -S -l 40 -p 2".

After sequencing, peak detection was performed using the MACS software [60] v2.1.1.20160309 with arguments "-nomodel -shift −100 -extsize 200 -broad". Peaks were annotated with Homer [61] using the GTF from ENSEMBL v67. Peak intersections were computed using Bedtools [62]. Global Clustering was done using seqMINER [63]. De novo motif discovery was performed using the MEME suite [64]. Footprinting signatures were calculated using TOBIAS v0.5.1 [15], and differential footprinting scores were plotted with R-package ggplot2 [65].

Single-cell RNA-seq (scRNA-seq)

Taf4$^{lox/lox}$*::VilCreER*^{T2} and control *Taf4*$^{lox/lox}$ enteroids treated with 4-OHT from day 5 to 8 of culture were dissociated at the end of the 4-OHT treatment with Accutase (A6964, Sigma) at 27 °C for 5 min and the cells were suspended in culture medium. Cells were then sorted by flow cytometry to select live cells and captured using 10X Genomics Chromium Analyzer. After sequencing, raw reads were processed using CellRanger (v 3.1) to align on the mm10 mouse genome, remove unexpressed genes and quantify barcodes and UMIs. Data were then analyzed in R (v3.6.3) using Seurat v3.1.4 [66]. First cells were filtered, only cells with feature count ranging from 200 to 6000 and with percentage of mitochondrial reads <15% were kept for the analysis. Then counts were normalized with the "LogNormalize" method and scaled to remove unwanted sources of variation. Clustering was performed on variable features using the 20 most significant principal components and a resolution of 0.9. Analysis of regulome was performed using SCENIC v1.1.2.2 [67].

DATA AVAILABILITY

16S rRNA data are publicly available from NCBI SRA under the Bioproject accession number PRJNA842218. RNA-seq, sc-RNA-seq and ATAC-seq data are deposited in the GEO database under the accession number GSE205442.

REFERENCES

1. Gehart H, Clevers H. Tales from the crypt: new insights into intestinal stem cells. Nat Rev Gastroenterol Hepatol. 2019;16:19–34.
2. Langer D, Martianov I, Alpern D, Rhinn M, Keime C, Dollé P, et al. Essential role of the TFIID subunit TAF4 in murine embryogenesis and embryonic stem cell differentiation. Nat Commun. 2016;7:11063.
3. Fadloun A, Kobi D, Pointud J-C, Indra AK, Teletin M, Bole-Feysot C, et al. The TFIID subunit TAF4 regulates keratinocyte proliferation and has cell-autonomous and non-cell-autonomous tumour suppressor activity in mouse epidermis. Development. 2007;134:2947–58.
4. Alpern D, Langer D, Ballester B, Le Gras S, Romier C, Mengus G, et al. TAF4, a subunit of transcription factor II D, directs promoter occupancy of nuclear receptor HNF4A during post-natal hepatocyte differentiation. Elife. 2014;3:e03613.
5. Kleiber T, Davidson G, Mengus G, Martianov I, Davidson I. Single cell transcriptomics reveal trans-differentiation of pancreatic beta cells following inactivation of the TFIID subunit Taf4. Cell Death Dis. 2021;12:790.
6. Muñoz J, Stange DE, Schepers AG, van de Wetering M, Koo B-K, Itzkovitz S, et al. The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. EMBO J. 2012;31:3079–91.
7. Petitprez F, Levy S, Sun C-M, Meylan M, Linhard C, Becht E, et al. The murine microenvironment cell population counter method to estimate abundance of tissue-infiltrating immune and stromal cell populations in murine samples using gene expression. Genome Med. 2020;12:86.
8. Wu M, Li P, Li J, An Y, Wang M, Zhong G. The differences between luminal microbiota and mucosal microbiota in mice. J Microbiol Biotechnol. 2020;30:287–95.
9. Mosca A, Leclerc M, Hugot JP. Gut microbiota diversity and human diseases: should we reintroduce key predators in our ecosystem? Front Microbiol. 2016;7. https://doi.org/10.3389/fmicb.2016.00455.
10. Sprangers J, Zaalberg IC, Maurice MM. Organoid-based modeling of intestinal development, regeneration, and repair. Cell Death Differ. 2021;28:95–107.
11. Rona G, Roberti D, Yin Y, Pagan JK, Homer H, Sassani E, et al. PARP1-dependent recruitment of the FBXL10-RNF68-RNF2 ubiquitin ligase to sites of DNA damage controls H2A.Z loading. Elife. 2018;7:e38771.
12. Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature. 2007;449:1003–7.
13. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature. 2009;459:262–5.
14. Schuijers J, van der Flier LG, van Es J, Clevers H. Robust cre-mediated recombination in small intestinal stem cells utilizing the olfm4 locus. Stem Cell Rep. 2014;3:234–41.
15. Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat Commun. 2020;11:4267.
16. Asfaha S, Hayakawa Y, Muley A, Stokes S, Graham TA, Ericksen RE, et al. Krt19(+)/Lgr5(-) cells are radioresistant cancer-initiating stem cells in the colon and intestine. Cell Stem Cell. 2015;16:627–38.
17. Ayyaz A, Kumar S, Sangiorgi B, Ghoshal B, Gosio J, Ouladan S, et al. Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. Nature. 2019;569:121–5.
18. Montgomery RK, Carlone DL, Richmond CA, Farilla L, Kranendonk MEG, Henderson DE, et al. Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. Proc Natl Acad Sci USA. 2011;108:179–84.
19. Gregorieff A, Liu Y, Inanlou MR, Khomchuk Y, Wrana JL. Yap-dependent reprogramming of Lgr5(+) stem cells drives intestinal regeneration and cancer. Nature. 2015;526:715–8.
20. Murata K, Jadhav U, Madha S, van Es J, Dean J, Cavazza A, et al. Ascl2-dependent cell dedifferentiation drives regeneration of ablated intestinal stem cells. Cell Stem Cell. 2020;26:377–90.e6.
21. Du W, Frankel TL, Green M, Zou W. IFNγ signaling integrity in colorectal cancer immunity and immunotherapy. Cell Mol Immunol. 2022;19:23–32.
22. Gao Q, Wang S, Chen X, Cheng S, Zhang Z, Li F, et al. Cancer-cell-secreted CXCL11 promoted CD8+ T cells infiltration through docetaxel-induced-release of HMGB1 in NSCLC. J Immunother Cancer. 2019;7:42.
23. Chen J, Ye X, Pitmon E, Lu M, Wan J, Jellison ER, et al. IL-17 inhibits CXCL9/10-mediated recruitment of CD8+ cytotoxic T cells and regulatory T cells to colorectal tumors. J Immunother Cancer. 2019;7:324.
24. Wang K, Kim MK, Di Caro G, Wong J, Shalapour S, Wan J, et al. Interleukin-17 receptor a signaling in transformed enterocytes promotes early colorectal tumorigenesis. Immunity. 2014;41:1052–63.
25. Mengus G, Fadloun A, Kobi D, Thibault C, Perletti L, Michel I, et al. TAF4 inactivation in embryonic fibroblasts activates TGF beta signalling and autocrine growth. EMBO J. 2005;24:2753–67.
26. Yu Y, Chen S, Lu G-F, Wu Y, Mo L, Liu Z-Q, et al. Alphavbeta6 is required in maintaining the intestinal epithelial barrier function. Cell Biol Int. 2014;38:777–81.
27. Hageman JH, Heinz MC, Kretzschmar K, van der Vaart J, Clevers H, Snippert HJG. Intestinal regeneration: regulation by the microenvironment. Dev Cell. 2020;54:435–46.
28. Farin HF, Jordens I, Mosa MH, Basak O, Korving J, Tauriello DVF, et al. Visualization of a short-range Wnt gradient in the intestinal stem-cell niche. Nature. 2016;530:340–3.
29. Erkek S, Johann PD, Finetti MA, Drosos Y, Chou H-C, Zapatka M, et al. Comprehensive analysis of chromatin states in atypical teratoid/rhabdoid tumor identifies diverging roles for SWI/SNF and Polycomb in gene regulation. Cancer Cell. 2019;35:95–110.e8.
30. Nakayama RT, Pulice JL, Valencia AM, McBride MJ, McKenzie ZM, Gillespie MA, et al. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. Nat Genet. 2017;49:1613–23.
31. Valencia AM, Kadoch C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. Nat Cell Biol. 2019;21:152–61.
32. Lorzadeh A, Romero-Wolf M, Goel A, Jadhav U. Epigenetic regulation of intestinal stem cells and disease: a balancing act of DNA and histone methylation. Gastroenterology. 2021;160:2267–82.
33. Piunti A, Shilatifard A. The roles of Polycomb repressive complexes in mammalian development and cancer. Nat Rev Mol Cell Biol. 2021;22:326–45.
34. Chiacchiera F, Rossi A, Jammula S, Piunti A, Scelfo A, Ordóñez-Morán P, et al. Polycomb complex PRC1 preserves intestinal stem cell identity by sustaining Wnt/β-catenin transcriptional activity. Cell Stem Cell. 2016;18:91–103.
35. Chiacchiera F, Rossi A, Jammula S, Zanotti M, Pasini D. PRC2 preserves intestinal progenitors and restricts secretory lineage commitment. EMBO J. 2016;35:2301–14.
36. Koppens MAJ, Bounova G, Gargiulo G, Tanger E, Janssen H, Cornelissen-Steijger P, et al. Deletion of Polycomb repressive complex 2 from mouse intestine causes loss of stem cells. Gastroenterology. 2016. https://doi.org/10.1053/j.gastro.2016.06.020.
37. Serra D, Mayr U, Boni A, Lukonin I, Rempfler M, Challet Meylan L, et al. Self-organization and symmetry breaking in intestinal organoid development. Nature. 2019;569:66–72.
38. Nakanishi Y, Reina-Campos M, Nakanishi N, Llado V, Elmen L, Peterson S, et al. Control of paneth cell fate, intestinal inflammation, and tumorigenesis by PKCλ/ι. Cell Rep. 2016;16:3297–310.
39. Yu Y, Blokhuis B, Derks Y, Kumari S, Garssen J, Redegeld F. Human mast cells promote colon cancer growth via bidirectional crosstalk: studies in 2D and 3D coculture models. Oncoimmunology. 2018;7:e1504729.
40. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. 2006;313:1960–4.
41. Sun S, Yu F, Xu D, Zheng H, Li M. EZH2, a prominent orchestrator of genetic and epigenetic regulation of solid tumor microenvironment and immunotherapy. Biochim Biophys Acta Rev Cancer. 2022;1877:188700.
42. Kang N, Eccleston M, Clermont P-L, Latarani M, Male DK, Wang Y, et al. EZH2 inhibition: a promising strategy to prevent cancer immune editing. Epigenomics. 2020;12:1457–76.
43. Nagarsheth N, Peng D, Kryczek I, Wu K, Li W, Zhao E, et al. PRC2 epigenetically silences Th1-type chemokines to suppress effector T-cell trafficking in colon cancer. Cancer Res. 2016;76:275–82.
44. Li J-Y, Xiao J, Gao M, Zhou H-F, Fan H, Sun F, et al. IRF/Type I IFN signaling serves as a valuable therapeutic target in the pathogenesis of inflammatory bowel disease. Int Immunopharmacol. 2021;92:107350.
45. El Marjou F, Janssen KP, Chang BH, Li M, Hindie V, Chan L, et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis. 2004;39:186–93.
46. Colnot S, Niwa-Kawakita M, Hamard G, Godard C, Le Plenier S, Houbron C, et al. Colorectal cancers in a new mouse model of familial adenomatous polyposis: influence of genetic and environmental modifiers. Lab Invest. 2004;84:1619–30.
47. Li Y, Zhang M, Xiao H, Fu H, Ho A, Lin C, et al. Addition of berberine to 5-aminosalicylic acid for treatment of dextran sulfate sodium-induced chronic colitis in C57BL/6 mice. PLoS ONE. 2015;10:e0144101.
48. Balbinot C, Armant O, Elarouci N, Marisa L, Martin E, De Clara E, et al. The Cdx2 homeobox gene suppresses intestinal tumorigenesis through non-cell-autonomous mechanisms. J Exp Med. 2018;215:911–26.
49. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–11.

50. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
51. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30:614–20.
52. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:e2584.
53. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37:D141–5.
54. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.
55. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
56. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.
57. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.
58. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
59. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a pratical and powerful ap- proach to multiple testing. Genome Biol. 1994;57:289–300.
60. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.
61. 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.
62. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
63. Ye T, Krebs AR, Choukrallah M-A, Keime C, Plewniak F, Davidson I, et al. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. 2011;39:e35.
64. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015;43:W39–49.
65. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. Cham: Springer International Publishing: Imprint; 2016. https://doi.org/10.1007/978-3-319-24277-4.
66. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. Cell. 2019;177:1888–902.e21.
67. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020;15:2247–76.

ACKNOWLEDGEMENTS

This work was supported by the Inserm and the ANR grant 15-CE14-0018. SSB and TK were fellows from the Fondation pour la Recherche Médicale, and AT from the Institut National du Cancer (INCa #2019-150). MP is an "équipe labellisée" of the Fondation pour la Recherche Médicale (DEQ20181039598). Work in the Davidson lab was supported by grants from the Ligue Nationale contre le Cancer, the Institut National du Cancer, the ANR-10-LABX-0030-INRT French state fund. ID is an "équipe labellisée" of the Ligue Nationale contre le Cancer. The IGBMC high throughput sequencing facility is a member of the "France Génomique" consortium (ANR10-INBS-09-08).

AUTHOR CONTRIBUTIONS

SSB performed experiments in mice and enteroids, GD performed the bioinformatics analyses, TK performed and analyzed the ATAC-seq experiments, SSB, AT and EM contributed to histological and immunohistological experiments, SM and HB performed microbiota studies, AH performed enteroid dissociation and single-cell capture, GM and MP participated in setting up the project and provided essential material, I Davidson, I Duluc. and JNF conceived the experiments, analyzed the data and wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41418-022-01109-6.

Correspondence and requests for materials should be addressed to Jean-Noel Freund.

Reprints and permission information is available at http://www.nature.com/reprints

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

Guillaume DAVIDSON

Exploring the roles of cell plasticity and tumor heterogeneity in renal cell carcinoma: implications on development, progression and response to immunotherapy.

Résumé

Le cancer est une des principales causes de mortalité dans le monde. Le carcinome renal du rein (RCC) compte parmi les 10 cancers les plus fréquents. Le plus commun, le RCC à cellules claires (cRCC), est traité par inhibiteurs de tyrosine kinases et inhibiteurs de checkpoint immunitaires (ICIs) pour les formes avancées de la maladie mais la plupart de ces patients ne répondent pas aux traitements. J'ai cherché à caractériser la composition tumorale des cRCC par séquençage d'ARN en bulk et en cellule unique ainsi que par transcriptomique spatiale. J'ai découvert quatre phénotypes de cellule de cancer définissant un gradient de transition épithélio-mésenchymateuse (TEM) et une population stromale de myofibroblastes associés au cancer (myCAFs). J'ai trouvé que la présence de cellules de cancer mésenchymateuses (ccRCC.mes) était associée à une mauvaise survie des patients et ces cellules sont enrichies dans les tumeurs avancées et dans les métastases à distance. De plus, la présence de ccRCC.mes corrèle fortement avec celle des myCAFs, ces deux types de cellules sont localisés spatialement de manière très proche à l'interface avec le tissu sain. Aussi, la présence de myCAF est associée à une mauvaise survie des patients traités par ICIs suggérant un rôle de ces cellules dans la résistance à ce traitement. Pour cibler ce dangereux couple ccRCC.mes/myCAF, j'ai identifié de multiples axes de communication et des biomarqueurs tels que STFA1P qui marque la présence de fibroblastes ou LINC01615 qui marque la présence de ccRCC.mes.

Mots-clés : cancer, cRCC, CMR, mélanome, TEM, plasticité, CAFs, lncRNA, scRNA-seq, transcriptomique-spatiale, immunothérapie

Résumé en anglais

Cancer is one the leading causes of death worldwide. Renal cell carcinoma (RCC) is among the ten most frequent cancers. The most common form of RCC, clear-cell RCC (ccRCC), is treated with tyrosine kinase inhibitors and immune checkpoint inhibitors (ICIs) for advanced disease but a large proportion of these patients do not respond to treatments. I sought to characterize ccRCC tumoral composition using single-cell and bulk RNA-sequencing coupled with spatial transcriptomics. I profiled four phenotypes of ccRCC cancer cells defining an epithelial to mesenchymal (EMT) gradient and stromal myofibroblastic cancer-associated fibroblasts (myCAFs). I found that presence of mesenchymal-like cancer cells (ccRCC.mes) was associated with poor patient survival and these cells were enriched in advanced disease tumors and in distant metastases. Furthermore, ccRCC.mes was highly correlated with myCAF presence and both could be found in close proximity to each other at the interface with healthy tissues. Additionally, myCAF presence was associated with poor survival in patients treated with ICIs suggesting a role in resistance to this treatment. To target this dangerous ccRCC.mes/myCAF couple, I identified multiple communication axes and biomarkers such as SFTA1P which marks fibroblast presence or LINC01615 which marks ccRCC.mes presence.

Keywords: cancer, ccRCC, RMC, melanoma, EMT, plasticity, CAFs, lncRNA, scRNA-seq, spatial-transcriptomics, immunotherapy