Dissertation
submitted to the

Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany

for the degree of
Doctor of Natural Sciences

Presented by

Yasamin Dabiri, Pharm.D.
born in: Shiraz, Iran

Oral examination: 2nd of July 2020

Small Molecule Approaches for Targeting Signaling Networks in Cancer and Cellular Reprogramming

Referees:

Prof. Dr. Stefan Wölfl
Prof. Dr. Steven Dooley

To my parents:
Nasrin and Reza

ACKNOWLEDGEMENTS

Putting my sense of appreciation into words was one of the most difficult tasks while writing this thesis. However, I wish to express my whole-hearted gratitude to all the people whose love, care, and support made the PhD experience possible and turned this stage of my life into a beautiful unforgettable memory.

Prof. Dr. Stefan Wölfl, my "Doktorvater", to whom I will be forever indebted for giving me the invaluable opportunity to join his research group and later to start a PhD. Thank you very much for all the extremely kind, enlightened, and empowering guidance and support during my studies and for creating a liberal intellectual atmosphere in the group, which brought lots of motivation and encouragement to the complicated scientific matters.

Prof. Dr. Steven Dooley, my second supervisor and a member of my thesis advisory committee (TAC), for his clever contribution to my PhD project, as well as the valuable suggestions and criticism of my TAC reports and presentations. Thank you very much for the continuing intellectual and experimental support you and your group, in particular Dr. Anne Dropmann, have offered during our ongoing collaborations, as well as your help in applying for scholarships and jobs.

Prof. Dr. Martina Muckenthaler, a member of my TAC, for all the constructive discussions in the meetings and her useful comments on my annual reports and presentations. Thank you very much for providing me the opportunity to perform some experiments in your group with the help of Silvia Colucci; I am grateful for the precious technical experience.

Prof. Dr. Ralf Bartenschlager and Prof. Dr. Gert Fricker, for agreeing to be a member of my oral-examination committee.

Dr. Xinlai Cheng, my mentor in Prof. Wölfl's group, for his close and responsible supervision of my PhD thesis. My deepest appreciation for believing in me and providing the opportunity of getting involved in several diverse projects, all of those have left me with remarkable scientific experiences. I wish you success in the academic path and look forward to continuing working on the ongoing projects together.

The graduate academy of Heidelberg University, for awarding me the full-term individual doctoral fellowship within the Landesgraduiertenförderung (LGF) program, which funded me throughout my studies. I wish to particularly thank the administrative director, Dr. Helke Hillebrand for all her efforts to secure the LGF grants in 2018, and also the contact persons in the administrative office, Ms. Birgit Bell and Ms. Elif Avcu for all their fast, kind, and helpful responses.

The Heidelberg biosciences international graduate school (HBIGS), for giving me a chance to get to know many brilliant fellow PhD students and to participate in lots of courses as well as the school retreat, all of which have taught me the principles of various aspects of science and carrier. I would like to pay special regards to the program coordinator, Dr. Rolf Lutz, the administration, Ms. Martina Galvan, and the carrier development advisor, Ms. Sandra Martini, whose efforts and kind support made all of this possible.

My fantastic peers at the Wölfl's lab, many of those have long left the group, yet are always on my mind and in my heart. Thank you for making me feel welcome and comfortable when I was new in the lab, for kindly helping me with the technical and intellectual matters, that seemed to be very complicated in the beginning, and for creating a warm environment, which quite often made me forget that I am far away from home;

Dr. Mohamed Abu el Maaty; you filled the space of many people for me so generously; my whole-hearted appreciation to you for being an outstanding colleague in the lab whom I have always looked up to and for being my life best friend...you made me realize that despite all the life uncertainties, how far one can go if she receives genuine spiritual support and mutual understanding. For that, I am forever indebted.

Dr. Suzan Can, for always offering me her kindness and understanding since I joined the lab as well as her scientific assistance whenever it was needed. I am also very thankful for your help with the German translation of the thesis summary.

Dr. Ali Ghanem, for being a great lab-mate and a lovely friend. I am appreciative for all our analytical discussions on science, culture, and psychology (with a focus on weird behavior!), despite some of them turning into conflicts! Never lose your impressive level of general information, passion, and sense of humor, no matter what!

Dr. Jannick Theobald, for keeping the lab sprit cool and entertaining (indeed in collaboration with Mohamed and sometimes at Ali's expense!). Thank you for being a lovely friend, an amazing travel companion, and a fantastic co-teacher (!) who made Mohamed's repetitive introductions on TXNIP and vitamin D (!) during the cell culture and Omics practical courses bearable. I will always recall my PhD experience with the friendly atmosphere I had in the office with you, Ali, and Mohamed, as well as our scientific teamwork, which is partly manifested in our publications.

Dr. Jee Young Kim, I wish to show my gratitude to you for kindly supporting me in the very first project I was involved in and for your patience in teaching me some of the basic molecular biology techniques.

Dr. Shahrouz Ghafoory, my deepest gratitude for helping me to get in contact with Prof. Wölfl when I was applying to the group, which paved the way for me to start my PhD.

Rodrigo Gama-Brambila, for being an amazing colleague and a sweet friend of mine. Thank you very much for all the friendship, kindness, and compassion you have offered during your stay in the lab and later out of the lab (!). Never stop being the resilient individual I know, who will not give up on enjoying life despite all the hardships.

Biljana Blagojevic, my kind regards for our lovely (and sometimes quite girly!) conversations during my never-ending western blot imaging (!), and on our way to the German course. It was also a pleasure supervising the literature seminars and the Omics practical course with you. Fadi Almouhanna, my colleague and neighbor, for his technical assistance in the lab and for kindly taking care of my plants when I was traveling! Anastasia Zuieva, a lovely lab-mate, thank you for offering help whenever it was needed, I wish you a successful PhD experience. Sawsan Saleh, for being welcoming and helpful when I was new. Finally, I would like to thank Saskia Schmitteckert, the technical assistance of the Wölfl's lab, for all her efforts in organizing and maintaining the lab tools and materials, and for supervising the trainees, whose performance in the lab has had a significant impact on the quality of my work.

I wish to express my deepest appreciation towards the brilliant students I had the honor to supervise, in particular the very talented Alice Schmid, Hoi Yin (Helen) Chan, and Ian Dirk Fichtner, whose efforts are reflected in my thesis as well as in our publications.

Prof. Dr. Fanak Fahimi and Prof. Dr. Mohammad Reza Hashemian, my Pharm.D. thesis advisors, for patiently teaching me the basics of scientific research.

I have always felt blessed for being surrounded by certain people as family and friends throughout the years; I am indebted to Farhad Azarnoosh, one of the greatest, most intelligent, and kindest guys I have ever known. Thank you for your life-changing advice on considering Heidelberg University among the many applications I was sending around to peruse my dream of doing a PhD abroad. My sincere appreciation for your love and care towards me and my family ever since I can remember. Dr. Farshid Sisani, I feel privileged for getting to know a person like you in Heidelberg at the exact time of my need. Thank you for always enlightening me with your wisdom and precious experiences, and for making me feel I could always count on your support. Amir Saadat Fard, whose friendship has had a great impact on the person I am. I also wish to express my kindest regards to the invaluable friendships I developed during my stay in Heidelberg; Dr. Fereydoon Taheri, Babak Loghmani, and Farshad Shobeiry for being such kind and reliable friends. Aria Ahadzada (very soon to be Dr. med.), for our sophisticated discussions (!) over late-night walks after doing sports and our tea parties!

In the end, I wish to express my infinite appreciation towards my family; my sister, Dr. Maryam Dabiri, a brilliant architect and a genuine artist, my deepest gratitude to you for always being a true older sister and for your life-long love and support. I also wish to show my warmest appreciation for generously having your creative touch on my thesis, posters, and presentations, and for the uniquely beautiful experience of the journal cover. My loving gratitude extends to my younger sister, Golnar Dabiri and my brother, Keivan Dabiri, whose love has sustained me throughout life. I would like to finally thank my parents; my mother, Prof. Nasrin Arshadi, who set an exceptional example of a devoted mom, an ambitious individual, a creative psychologist, and a successful academic. My father, Mr. Abdolreza Dabiri, a truly smart engineer, a great manager, and one of the humblest, most honest, and most content souls I know. I am forever indebted to both of you for raising me to value education, knowledge, and intellectuality above many things in life, for teaching me to always stay positive, kind, truthful, and giving no matter how good or bad life is...I believe those are my most precious belongings, and I will always try to protect them at the core of my being, as I did during this stage of my life.

LIST OF PUBLICATIONS

Peer-Reviewed Journals

(1) **Dabiri Y.**, Gama-Brambila R.A., Taskova K., Herold K., Reuter S., Adjaye J., Utikal J., Mrowka R., Wang J., Andrade-Navarro M.A., Cheng X., Imidazopyridines as potent KDM5 demethylase inhibitors promoting reprogramming efficiency of human iPSCs. iScience, 2019. *Article selected as the journal cover for volume 13 (March 2019 issue): https://www.cell.com/iscience/issue?pii=S2589-0042(19)X0003-6*

(2) **Dabiri Y.**, Abu el Maaty M.A., Chan H.Y., Wölker J., Ott I., Wölfl S., Cheng X., p53-dependent anti-proliferative and pro-apoptotic effects of a gold(I) *N*-heterocyclic carbene (NHC) complex in colorectal cancer cells. Front Oncol, 2019.

(3) **Dabiri Y**., Schmid A., Theobald J., Blagojevic B., Wojciech C., Ott I., Wölfl S., Cheng X., A ruthenium(II) *N*-heterocyclic carbene (NHC) complex with naphthalimide ligand triggers apoptosis in colorectal cancer cells via activating the ROS-p38 MAPK pathway. Int J Mol Sci, 2018.

(4) Abu el Maaty M.A., **Dabiri Y.**, Almouhanna F., Blagojevic B., Theobald J., Büttner M., Wölfl S., Activation of pro-survival metabolic networks by 1,25(OH)2D3 does not hamper the sensitivity of breast cancer cells to chemotherapeutics. Cancer Metab, 2018.

(5) Wölker J., Terenzi A., Cheng X., Hager L., **Dabiri Y.**, Prochnow P., Bandow J.E., Wölfl S., Keppler B.K., Ott I., Fluorescent organometallic rhodium(I) and ruthenium(II) metallodrugs with 4-ethylthio-1,8-naphthalimide ligands: anti-proliferative effects, cellular uptake, and DNA interaction. Eur J Med Chem, 2018.

(6) **Dabiri Y.**, Kalman S., Gürth C.M., Kim J.Y., Mayer V., Cheng X., The essential role of TAp73 in bortezomib-induced apoptosis in p53-deficient colorectal cancer cells. Sci Rep, 2017.

(7) Abu el Maaty M.A., Strassburger W., Qaiser T., **Dabiri Y.**, Wölfl S. Differences in p53 status significantly influence the cellular response and cell survival to 1,25-dihydroxyvitamin D3-metformin co treatment in colorectal cancer cells. Mol Carcinog, 2017.

Book Chapters

(1) **Dabiri Y.**, Song G., Cheng X. “Indirubins as multitarget anti-tumor agents” In Herbal medicine: back to the future, Edition: 2019, Publisher: Bentham Science, Editors: F. Murad, Rahman A., Bian K.

TABLE OF CONTENTS

LIST OF FIGURES (CHAPTER 1)

LIST OF FIGURES (CHAPTER 2)

LIST OF TABLES

LIST OF ABBREVIATIONS

2DG	2-deoxyglucose
3IO	Indirubin-3'-monoxime
6BIO	6-bromoindirubin-3'-monoxime
ALK	Activin receptor like kinase
AML	Acute myelogenous leukemia
AMPK	AMP-activated protein kinase
AP	Alkaline phosphatase
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
C/EBPβ	CCAAT/enhancer binding protein β
CAFs	Cancer-associated fibroblasts
CAMKK	Calcium/calmodulin-dependent kinase kinase
CDI	coefficient of drug interaction
CDK	Cyclin-dependent kinases
CHX	Cycloheximide
ciPSCs	Chemically-induced pluripotent stem cells
CK	Casein kinase
CML	Chronic myeloid leukemia
Co-Smad	Common Smad
COMPASS	Complex proteins-associated with SET1
CR	Complete hematologic response
CTEN	C-terminal tension like protein
CTGF	Connective tissue growth factor
D-PBS	Dulbecco's Phosphate-Buffered Saline
DAPK	Death-associated protein kinase
DARPP-32	Dopamine- and cyclic adenosine 3′-5′-monophosphate-regulated phosphoprotein 32
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNMT	DNA methyl transferase
DPC4	Deleted in pancreatic carcinoma locus 4
DTT	1,4-Dithioththeriotol
DUBs	Deubiquitinating enzymes
ECATs	Embryonic stem cell-associated transcripts
ECL	Enhanced chemoluminescence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid

EGFR	Epidermal growth factor receptor
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid
EMT	Epithelial-to-mesenchymal transition
EndMT	Endothelial-to-mesenchymal transition
EPAC	Exchange protein directly activated by cAMP
EPCAM	Epithelial cell adhesion marker
Erk	Extracellular signal-regulated kinase
ESCs	Embryonic stem cells
Fbxo15	F-box only protein 15
FCS	Fetal calf serum
FGFR	Fibroblast growth factor receptor
GDF	Growth and differentiation factors
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GSK3	Glycogen kinase 3
HDAC	Histine deacetylases
HEPES	4-(2-hydroxyethyl)-1- piperazineethanesulfonic acid
HFs	Human foreskin fibroblasts
HIF1	Hypoxia-inducible factor
HMGA2	High mobility group A2
HMT	Histone methyl trasnferase
hPSCs	Human pluripotent stem cells
HUVECs	Human umbilical vein endothelial cells
I-Smad	Inhibitory Smad
IDs	Inhibitors of cell differentiation
IGF	Insulin-like growth factor
IGF1R	Insulin-like growth factor 1 receptor
IKK	IκB kinase
IL	Interleukin
InsR	Insulin receptor
iPSCs	Induced pluripotent stem cells
IRD	Indirubin derivatives
JAK	Janus kinase
JNK	c-Jun amino terminal kinase
KSR	Knockout serum replacement supplement
LIF	Leukemia inhibitory factor
LKB1	Liver kinase B1
LSD1	Lysine-specific demethylase 1
MAPK	Mitogen-activated protein kinase

MEFs	Mouse embryonic fibroblasts
MET	Mesenchymal-to-epithelial transition
MH1	Mad homology 1
MIP	Macrophage inflammatory protein
miPSCs	Mouse induced pluripotent stem cells
MMPs	Metalloproteinases
mTOR	Mechanistic target of rapamycin
MTT dye	Thiazolyl blue tetrazolium bromide dye
NEDD4L	Neural precursor cell expressed developmentally downregulated protein 4-like
NF-κB	Nuclear factor-κB
NK cells	Natural killer cells
NRTK	Non-receptor tyrosine kinase
O4Is	OCT4-inducing compounds
OSKM	OCT4, SOX2, KLF4, MYC
OTUs	Ovarian tumor proteases
PAINS	Pan-assay interference compounds
PDA	Pancreatic ductal adenocarcinoma
PDGFB	Growth factor beta polypeptide
PDGFR	Platelet-derived growth factor receptor
PDK1	Pyruvate dehydrogenase kinase 1
PECAM1	Platelet endothelial cell adhesion molecule
PFA	Paraformaldehyde
PI3K	phosphatidylinositol-3 kinase
PKA	Protein kinase A
PMSF	Phenylmethane sulfonyl fluoride
Polyvinylidene fluoride	PVDF
PR	Partial hematologic response
PSCs	Pluripotent stem cells
PTK	Protein tyrosine kinases
R-Smad	Receptor-regulated Smads
Rb	Retinoblastoma protein
ROCKi	Rho-associated kinase inhibitor
RTK	Receptor tyrosine kinase
SALL4	Sal-like protein 4
SAR	Structure-activity relationship
SBE	Smad binding element
SCNT	Somatic cell nuclear transfer
SDS	Sodium dodecyl sulfate

SFK	Src family kinase
SMEDDS	Self-micro-emulsifying drug delivery system
Smurf	Smad-specific E3 ubiquitin protein ligase
SNEDDS	self-nanoemulsifying drug delivery system
SSEA1	Stage-specific embryonic antigen 1
STAT3	signal transducer and activator of transcription 3
TAK1	Transforming growth factor β-activated kinase 1
TCM	Traditional Chinese medicine
Tetramethylethylenediamine	TEMED
TGFβ	Transforming growth factor β
TGFβR	Transforming growth factorβ receptor
TK	Tyrosine kinase
TKI	Tyrosine kinase inhibitor
TKL	Tyrosine kinase-like
TRAF6	Tumor necrosis factor receptor-associated factor 6
TRIMM33	Tripartite motif containing 33
UCHs	Ubiquitin C-terminal hydrolyses
UPP	Ubiquitin-proteasome pathway
USPs	Ubiquitin-specific proteases
UTF1	Undifferentiated embryonic cell transcription factor 1
VEGFR	Vascular endothelial growth factor receptor
VHL protein	von Hippe-Lindau protein
WDR5	WD repeat protein 5
YAP	Yes-associated protein

SUMMARY

Aberrations in signaling networks have long been studied for their contribution to tissue degeneration and cancer. Chemical modulation of such pathways provides an excellent opportunity to fine-tune their activity, owing to the rapid and dose-dependent effects of small molecules. The results of the presented thesis demonstrate the application of small molecules in: (i) studying one of the key pathways in cancer and development–TGFβ/Smad signaling; and (ii) modulation of cellular fate towards stemness via targeting self-renewal- and pluripotency-associated signaling molecules.

In the first chapter, I describe the use of newly-synthesized molecules derived from the indirubin family of natural compounds in the regulation of R-Smad signaling, with a focus on TGFβ-related Smad2/3. Using a large screen encompassing cell lines of various cancerous- and non-cancerous tissue origins, I show that indirubin derivatives (IRDs) collectively, and the IRD E738 analogue specifically, interfere with TGFβ/Smad signaling via a profound reduction in steady-state R-Smad levels. My analyses illustrate that this effect is not only due to reduced R-Smad gene transcription by E738, but also a result of protein degradation, demonstrating that the IRD modulates the TGFβ/Smad pathway through different mechanisms. Further investigation by transient over-expression of wild-type-, truncated-, and phosphorylation-defective mutant forms of Smad2 and Smad3 reveals the significance of the "linker domain", in particular its phosphorylation status, in the small molecule's regulation of basal Smad stability. Given the role of Smad linker phosphorylation–referred to as "phospho-Smad signaling"–in the malignant switch of TGFβ in some tumor types, the potential connection between E738's regulation of steady-state Smad2/3 stability and the phospho-Smad pathway is explored. Using patient-derived cholangiocarcinoma (CCA) cell lines, I show that IRD E738 inhibits oncogenic phospho-Smad isoforms (linker region sites), while maintaining those associated with a cytostatic phenotype (C-terminally phosphorylated sites). This effect reflects the molecule's inhibitory activity on kinases (e.g., GSK3 and CDKs) involved in the phosphorylation of Smad2/3 linker region.

The results of the second chapter outline a combined chemical and genetic approach for enhancing the generation and maintenance of human iPSCs. Using a cell-based high-throughput screen, a series of OCT4-inducing compounds (O4Is) are identified, including imidazopyridine analogues, with the lead compound termed O4I3, as well as 4-(tert-Butyl)-N-(2,3-dimethylphenyl) thiazol-2-amine, herein O4I4. The small molecules either alone or in combination are shown to activate pluripotency-associated signaling, and to increase the reprogramming efficiency of human fibroblasts into iPSCs when combined with ectopic expression of the master pluripotency transcription factors: OCT4, SOX2, KLF4, and MYC (known as "OSKM"). Transcriptomic analyses (DNA microarray/RNA-sequencing) and ATAC-sequencing revealed previously unrecognized, targetable molecular events in the path towards pluripotency. Indeed, O4I3 lifts an epigenetic barrier of reprogramming through inhibiting the histone demethylase enzyme, KDM5A, thus allowing the enrichment of H3K4Me3 at the *OCT4* promoter. Additionally, applying a combination of O4I3 and O4I4 to OSKM-based reprogramming reveals the involvement of bone morphogenetic protein (BMP)/Smad signaling upstream of high mobility group A1 (HMGA1) in O4I3/4-mediated induction of endogenous OCT4, which in turn initiates the reprogramming process.

Altogether, the results of both chapters demonstrate the potential use of new classes of small molecules in targeting signaling networks associated with cancer and somatic cell reprogramming.

ZUSAMMENFASSUNG

Schon lange werden Veränderungen in Signalnetzwerken hinsichtlich ihres Beitrags zu Gewebsdegeneration und Krebs untersucht. Dank der schnellen und dosisabhängigen Wirkung kleiner Moleküle (Small Molecules) liefert eine chemische Modulation solcher Signalwege eine hervorragende Möglichkeit deren Aktivität genau einzustellen. Die Ergebnisse der vorliegenden Arbeit zeigen die Anwendung kleiner Moleküle: (i) bei der Untersuchung eines Schlüssel-Signalwegs bei Krebs und der Entwicklung - der TGFβ/Smad-Signalweg; und (ii) bei der Modulation des Zellschicksals: Ausbildung von Stammzelleigenschaften durch Angriff auf Signalmoleküle, die eine Rolle bei Selbsterneuerung und Pluripotenz spielen.

Im ersten Teil beschreibe ich die Anwendung neu-synthetisierter Moleküle, die sich von der Naturstoff-Familie der Indirubine ableiten, bei der Regulation der R-Smad-Signalisierung mit Fokus auf TGFβ-assoziiertes Smad2/3. Mittels eines umfangreichen Screenings, welches verschiedene aus Tumor- und Nicht-Tumor-Gewebe stammende Zelllinien umfasst, zeige ich, dass Indirubin-Derivate (IRDs) insgesamt, und besonders das IRD-E738-Analogon, den TGFβ/Smad-Signalweg über eine erhebliche Reduktion des Steady-State-Levels von R-Smad beeinflussen. Meine Untersuchungen verdeutlichen, dass dieser Effekt nicht nur auf eine verringerte R-Smad-Gentranskription durch E738 zurückzuführen ist, sondern auch Folge von Proteinabbau ist. Dies zeigt, dass IRDs den TGFβ/Smad-Signalweg über verschiedene Mechanismen modulieren. Weitere Untersuchungen mittels transienter Überexpression von Wildtyp-Smad2 und -Smad3 sowie der durch Mutation entweder verkürzten oder nicht-phosphorylierbaren Formen offenbaren die Bedeutung der „*Linker-Domäne*“, insbesondere deren Phosphorylierungsstatus, in der Regulation der basalen Smad-Stabilität durch kleine Moleküle. Angesichts der Rolle, welche die Linker-Phosphorylierung - bezeichnet als „*Phospho-Smad-Signalisierung*“ - beim Umschalten der TGFβ-Funktion in die maligne Form in einigen Tumorarten spielt, wird die mögliche Verbindung zwischen der Regulation der Steady-State-Stabilität von Smad2/3 durch E738 und dem Phospho-Smad-Signalweg untersucht. Durch Verwendung von Cholangiokarzinom-(CAA-)Zelllinien, die von Patienten stammen, zeige ich, dass das IRD E738 onkogene Phospho-Smad-Isoformen (Phosphorylierung in der Linker-Region) hemmt, während es jene erhält, die zum zytostatischen Phänotyp (Phosphorylierung im c-terminalen Bereich) gehören. Diese Wirkung spiegelt die inhibitorische Aktivität des Moleküls auf Kinasen (z. B. GSK3 und CDKs) wider, welche an der Phosphorylierung der Linker-Region von R-Smad2/3 beteiligt sind.

Die Ergebnisse des zweiten Teils geben einen Überblick über eine gleichzeitig chemische und genetische Vorgehensweise, die dazu dient, die Bildung und Aufrechterhaltung humaner iPSCs zu steigern. Mittels eines zellbasierten Hochdurchsatz-Screenings konnten eine Reihe OCT4-induzierender Verbindungen (O4Is) identifiziert werden. Dazu gehören Imidazopyridin-Analoga, darunter die Leitsubstanz O4I3, sowie 4-(tert-Butyl)-N-(2,3-dimethylphenyl)thiazol-2-amin, bezeichnet als O4I4. Die kleinen Moleküle Aktivieren entweder alleine oder in Kombination pluripotenz-assoziierte Signalwege und erhöhen die Effizienz bei der Reprogrammierung humaner Fibroblasten zu iPSCs, wenn sie mit einer ektopischen Expression der Haupt-Pluripotenz-TranskriptionsfAktoren *OCT4*, *SOX2*, *KLF4* und *MYC* (bekannt als „*OSKM*“) kombiniert werden. Transkriptomanalysen (DNA-Microarray/RNA-Sequenzierung) und ATAC-Sequenzierung offenbarten bisher unerkannte, als Angriffsziel geeignete molekulare Ereignisse auf dem Weg zur Pluripotenz. So beseitigt O4I3 eine epigenetische Reprogrammierungsbarriere durch Hemmung der Histon-Demethylase KDM5A, so dass die Anreicherung von H3K4Me3 am *OCT4*-Promotor ermöglicht wird. Zusätzlich wird durch die Anwendung einer Kombination von O4I3 und O4I4 bei der OSKM-basierten Reprogrammierung eine Beteiligung des BMP/Smad-Signalwegs, welcher HMGA1 (high mobility group A1) vorgeschaltet ist, bei der O4I3/4-vermittelten Induktion von endogenem OCT4 gezeigt, was wiederum den Reprogrammierungsprozess initiiert.

Insgesamt zeigen die Ergebnisse beider Teile eine mögliche Verwendung von neuen Klassen kleiner Moleküle für den Angriff auf Signalnetzwerke auf, die in Verbindung mit Krebs und der Reprogrammierung somatischer Zellen stehen.

CHAPTER [1]

Small Molecule Modulation of TGFβ/Smad Signaling

1. INTRODUCTION

1.1. TGFβ/BMP signaling

1.1.1. TGFβ/BMP signaling–the core pathway

Identification of the key players of transforming growth factor β (TGFβ)/bone morphogenetic protein (BMP) signaling dates back to more than a decade ago. With increasing number of studies, the concept of the signaling pathway and its medical relevance has been enriched throughout the years, but also has raised new questions on the topic, in particular the impact of the cellular context on the responses to TGFβ/BMP signaling.

The TGFβ superfamily of ligands includes more than 30 related cytokines with a key role in cellular development and tissue homeostasis. This consists of 3 TGFβ isoforms, 4 activin proteins, nodal, 10 BMPs, and 11 growth and differentiation factors (GDFs), all sharing a so-called "cysteine knot" in their mature form ***(recently reviewed in*** [1]).

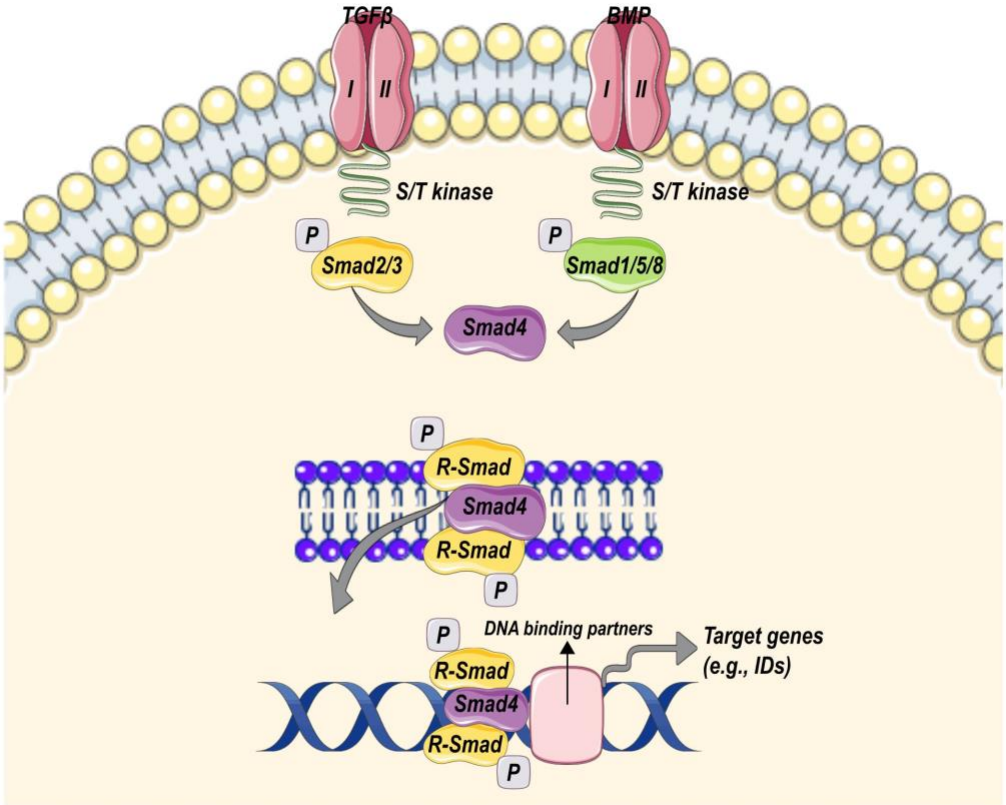


Figure 1. Schematic illustration of the principles of TGFβ/BMP signaling. The simplest way for understanding the basics of the pathway is to divide it into the following sequential steps; (i) receptor-mediated C-terminal phosphorylation of the regulatory Smads (R-Smads), Smad2/3 and Smad1/5/8 in case of TGFβ- and BMP receptors, respectively, (ii) complex formation with the common Smad, Smad4,

and (iii) translocation of the trimeric structure to the nucleus where it directly acts as a transcription factor in cooperation with other DNA binding partners.

The signaling pathway is triggered when the ligands bind to preformed dimeric receptor structures, referred to as activin receptor like kinases (ALKs). Upon ligand binding, type I and type II dimers bridge in order to form the heterotetrameric active receptor complex in which the constitutively active type II receptor phosphorylates and activates type I receptor at its intracellular serine/threonine (S/T) kinase domain. Activated type I receptor then activates receptor-regulated Smads (R-Smads) by means of phosphorylating two seine residues at their C-termini, which enables heteromerization with the common Smad (Co-Smad), Smad4, as well as with other R-Smads (***basics of TGFβ signaling is reviewed in*** [2]). The trimeric structure containing 2 R-Smads and 1 Smad4 molecule is thought to be the principle functional unit of the TGFβ/BMP pathway, which shuttles to the nucleus where together with other DNA binding partners regulates the transcription of various target genes involved in cellular proliferation, differentiation, death, and migration (**Figure 1**).

1.1.2. TGFβ/BMP signaling–the key players (Smads)

Depending on the ligand initiating the pathway, R-Smads fall into 2 groups, which have been traditionally divided into TGFβ/activin/nodal-activated Smad2/3 and BMP/GDF/AMH-activated Smad1/5/8. However, this classification does not precisely reflect the reality. The pathway is split only downstream of type I receptors meaning that ALK4/5/7 specifically phosphorylate Smad2/3 and ALK1/2/3/6 are specifically responsible for phosphorylation of Smad1/5/8 [2].

Aside from R-Smads and Co-Smads, inhibitory Smads (I-Smads) i.e., Smad6 and Smad7, form a distinct subfamily of Smad proteins, known to terminate the signaling through distinct and diverse mechanisms. The latter include: (i) Competitive inhibition of R-Smad binding; I-Smads have been initially found to interact with TGFβ/BMP type I receptors, prompting the idea of a direct inhibitory mechanism, (ii) recruitment of E3 ligases to the active receptor complex, which leads to receptor ubiquitination and subsequent proteasomal degradation, (iii) interacting with protein phosphatase 1 (PP1), which targets them to active receptors, leading to receptor inactivation through dephosphorylation, and (iv) binding to Smad binding elements (SBEs) and inactivating Smad-dependent promoter activation, implicating a nuclear role for I-Smads [2].

R-Smads and Smad4 share two highly conserved domains known as Mad homology 1 (MH1) and MH2 at their amino- and carboxy-terminal sequences, respectively. The MH1 domain contains nuclear localization signals as well as a β-hairpin structure, which is critical for DNA binding ability. The MH2 domain, on the other hand, has a series of hydrophobic surface patches, enabling the protein for various interactions with activated type I receptors, Smad4, and DNA binding partners, to name a few. MH1 and MH2 domains are coupled via a less conserved proline-rich linker region [3]. Although R-Smad activation is initiated by receptor-mediated phosphorylation of the SXS motif in the MH2 domain, it is

increasingly apparent that the linker region plays a critical role in Smad activity and stability, through integration with signaling molecules stemming from other pathways [4].

1.1.3. TGFβ/BMP signaling–noncanonical (non-Smad) pathways

Beyond the Smad-dependent, canonical TGFβ signaling, the activated receptor complexes also signal through other transducers, for example, the mitogen-activated protein kinases (MAPKs), including the extracellular signal-regulated kinases (Erks), c-Jun amino terminal kinase (JNK), p38 MAPK, as well as the IκB kinase (IKK), phosphatidylinositol-3 kinase (PI3K)/Akt, and Rho family GTPases, which in the context of TGFβ family signaling are collectively referred to as non-Smad (noncanonical) signaling pathways (***thoroughly reviewed by Zhang et al. in*** [5]). ***These receptor-activated non-Smad transducers exert their effects either as isolated pathways or hand in hand with Smads, as they converge onto Smads through post-translational modifications in the linker region.***

1.1.3.1. Non-Smad pathways–activation of Erk MAPK

The role of Erk MAPKs in TGFβ signaling has been frequently reported, even before the discovery of Smads. The primary evidence came from the studies showing a rapid activation of Ras (as early as 5-10 min comparable to the kinetics shown by mitogenic growth factors) upon TGFβ stimulation in normal epithelial- as well as colon cancer cells [6-8]. Mechanistically, upon ligand binding the constitutively active type II receptor phosphorylates type I receptor at its S/T residues and induce tyrosine phosphorylation of both type I and II, as well as ShcA. These phosphorylated tyrosines recruit Grb2/Sos, which then activates Erk1/2 through the Ras-Raf-MEK1/2 axis [9]. This leads to the phosphorylation of several target genes, among others, members of the AP-1 family, including p53 and other transcription factors, which interact and cooperate with Smads to regulate gene expression [10]. A key biological function of Erk in the context of TGFβ signaling is the epithelial-to-mesenchymal transition (EMT) [11, 12], a normal event in embryonic development, which is pathologically associated with tumor invasion and fibrosis [13]. Similar to TGFβ, BMPs can also induce EMT, cell migration, and invasion in an Erk-dependent manner. In this regard, Erk activity has been shown to contribute to BMP7-induced morphologic conversion in prostate cancer cells [14], as well as BMP2-mediated motility and invasiveness in lung [15] and gastric cancer cells [16]. Additionally, Erk activity has been shown to be essential in BMP-induced cell differentiation through the regulation of BMP target genes, such as *alkaline phosphatase* (*AP*), *collagen I*, *fibronectin*, *osteopontin*, *osteocalcin*, and *Runx2* [17, 18].

Importantly, Erk modulates the activity of R-Smads including Smad1, Smad2, and Smad3 by direct phosphorylation [19-21]. Although in cell culture systems, Erk-mediated phosphorylation of Smads is normally known to inhibit Smad transcriptional activity [21], some studies have reported the opposite [22], which may be in part due to a different cell type and/or the examined response. The former regulation has been used to explain how oncogenic Ras overrides TGFβ-induced anti-mitogenic responses in cancer cells [20]. Oncogenically-activated Ras acting via Erk, phosphorylates Smad2 and Smad3 at specific residues in the linker region, which are distinct from those of the TGFβ receptor I

(TGFβRI). This, in turn, inhibits Smad2/3 nuclear accumulation and transcriptional activity. Accordingly, mutation of these phosphorylation sites in Smad3 produces a Ras-resistant form, which can recover TGFβ's growth inhibitory effects in Ras-hyperactive cancer cells [20].

1.1.3.2. Non-Smad pathways–activation of TGFβ-activated kinase 1 (TAK1) and downstream JNK- and p38 MAPKs, as well as IKK

In response to cytokines, p38 and JNK MAPKs are activated through the action of MAPK kinases (MKKs), MKK4 and MKK3/6 [23]. Like cytokines, TGFβ can induce the activation of JNK via MKK4 and p38 via MKK3/6 [5]. In contrast to the activation of R-Smads and Erk, which occurs through sequential phosphorylation, regulation of this branch of non-Smad signaling relies on a more elaborate mechanism, involving the ubiquitin-proteasome pathway. Ligand-bound TGFβ receptors interact with the RING-domain E3 ligase, tumor necrosis factor receptor-associated factor 6 (TRAF6) to induce K63-linked polyubiquitination of itself; polyubiquitinated TRAF6 then associates with and activates TAK1 (also known as MAPKKK7), and finally its downstream kinases, JNK and p38 [24, 25]. In addition, other MAPKs such as MEKK1, MEKK4, MLK2, MLK3, and ASK1 have been also reported in the TGFβ-activated JNK and p38 signaling [5]. Although being independent from the activation of R-Smads, regulation of JNK/p38 MAPKs by TGFβ receptors involves the inhibitory Smads, Smad6 and Smad7. Smad6 inhibits TGFβ-activated MAPKs by recruiting the deubiquitinating enzyme A20 to TRAF6 [26], whereas Smad7 promotes TGFβ-activated MAPKs, as it interacts with Smurf2 and routes TGFβ receptors to caveolin-1-containing lipid rafts wherein TRAF6 is localized [27].

One of the important outcomes of JNK- and p38 MAPK activation is apoptosis; this is particularly relevant to BMP's roles in development, as well as TGFβ's roles in tumor suppression [5]. In light of this, JNK- and p38 MAPK have been shown to cooperate with Smad signaling in inducing apoptosis, as Smad3 is known to play an essential role in TGFβ-mediated pro-apoptotic effects (***examples are***; [28, 29]. In addition to apoptosis, TGFβ-activated JNK- and p38 MAPKs also contribute to the induction of EMT [5]. In this regard, blocking p38 MAPK activity using small molecule inhibitors or by expressing a dominant-negative MKK3 mutant [30, 31] as well as silencing TRAF6 expression [25] has been shown to block TGFβ-induced EMT and cancer cell invasion. Additionally, TGFβRI can recruit TRAF4, aside from TRAF6, and TAK1 for the Smad-independent activation of JNK- and p38 MAPK, which appears to be an obligatory step in cellular migration and metastatic dissemination in response to TGFβ [32].

TGFβ- and BMP pathways may crosstalk with nuclear factor-κB (NF-κB) signaling mainly because of TAK1's role in activating IKKs, which in turn activates NF-κB [33]. The cellular cues that lead to either the pro-apoptotic response of TGFβ or the pro-survival effects of NF-κB, both through activating TAK1, are yet to be fully determined. Similar to Erk1/2, JNK- and p38 MAPKs directly regulate R-Smad activity via phosphorylating the protein at specific residues in the linker region, which overlap with those phosphorylated by Erk MAPK.

1.1.4. TGFβ/BMP signaling–regulatory mechanisms

As abovementioned, the critical regulatory steps in R-Smad activity are mediated by C-tail phosphorylation and subsequent nucleoplasmic shuttling (**also see Figure 1**). However, activated R-Smads are prone to a variety of post-translational modifications (e.g., phosphorylation), which enable the peak transcriptional action, on the one hand, and recognition by the ubiquitin-proteasome pathway (UPP), hence destruction, on the other hand (***summarized in*** [34]). This concept although paradoxical presents an opportunity to reveal how the cellular responses to TGFβ and BMP signals are coupled to the turnover of their central mediators, Smads, and more importantly how R-Smads get to act before their disposal.

1.1.4.1. Controlling TGFβ/BMP signaling by phosphorylation: Smad action turnover switch

A series of phosphorylation steps drives the process of Smad signaling, starting with receptor S/T kinases. The receptor-induced R-Smad phosphorylation is reversed by the action of tail phosphatases (e.g., PPM1A), limiting the general pool of activated R-Smads (***reviewed in*** [35]). R-Smad molecules engaged in transcription of the target genes rapidly undergo a second step of phosphorylation events; these are mediated by nuclear cyclin-dependent kinases (CDKs) 8 and CDK9, priming R-Smads for further phosphorylation by glycogen synthase kinase 3 (GSK3). CDK8/9 are components of transcriptional mediator and elongation complexes and GSK3 is a Wnt- and phosphoinositide 3 kinase (PI3K) downstream molecule. These phosphorylations–clustered in the interdomain linker region–not only promote R-Smad transcriptional action through recruiting DNA binding factors like Yes-associated protein (YAP, in case of Smad1/5) and Pin1 (in case of Smad2/3), but also add binding sites for the E3 ubiquitin ligases, Smad-specific E3 ubiquitin protein ligase (Smurf) 1/2 and neural precursor cell expressed developmentally downregulated protein 4-like (NEDD4L), targeting R-Smads for polyubiquitination and proteasome-mediated degradation (**summarized in Figure 2**). Degradation of agonist-activated R-Smads occurs simultaneous to dephosphorylation of the linker by small C-terminal domain (SCD) phosphatases, allowing repeated rounds of signaling. If R-Smads are not dephosphorylated, they will be further phosphorylated by GSK3, which marks them for proteasomal degradation, depleting the pool of TGFβ/BMP signal transducers. In this manner, the delivery of TGFβ and BMP signals is coupled to the turnover of R-Smad messenger molecules.

In a different context, mitogens and stress signals acting through MAPKs and the cell cycle acting through cell division-related CDKs (e.g., CDK4) phosphorylate R-Smads at sites similar to those phosphorylated by CDK8/9. MAPK- and/or CDK-primed linker modifications can be followed by GSK3-mediated phosphorylation (**Figure 2**). This not only marks R-Smads for degradation as aforementioned, but also creates entry points for cooperation between the TGFβ/BMP pathway with other pathways including MAPK- and Wnt signaling [19, 36-38].

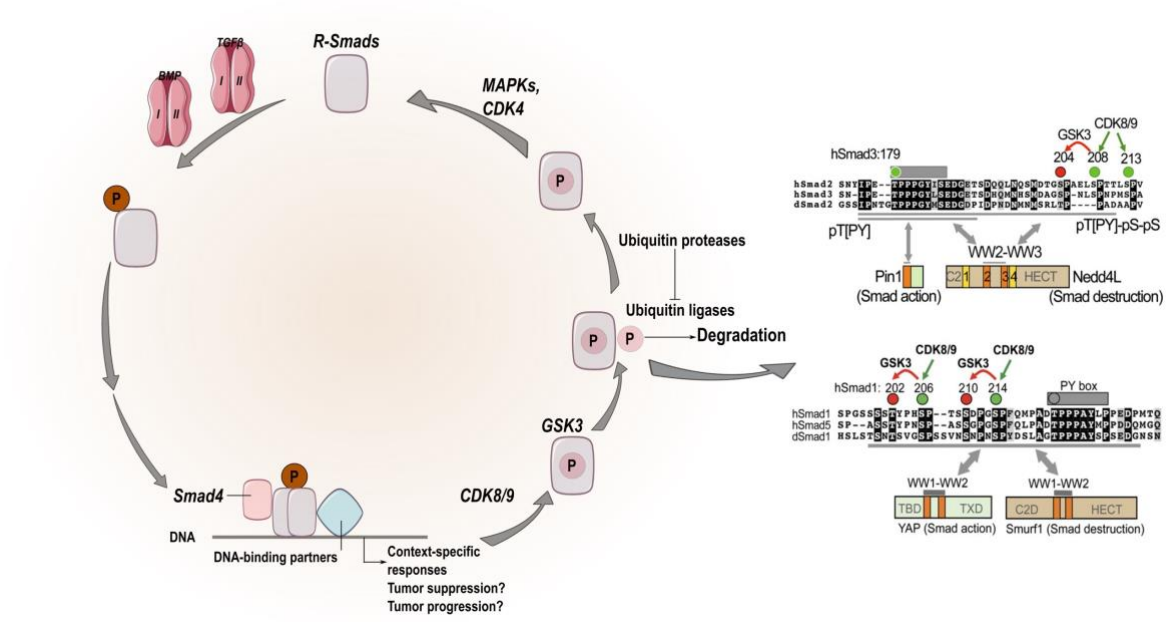


Figure 2. Schematic illustration of *"R-Smad signaling cycle"* shows the connection between R-Smad transcriptional activity and R-Smad turnover, which is mediated via sequential phosphorylation steps, as indicated. On the right, sequence alignments of the linker domain of human (h) Smad2/3 and Smad1/5 as well as those of drosophila (d) are shown with the phosphorylation residues and the involved kinases highlighted. The responsible ubiquitin E3 ligases and the regions mediating the binding of the ligases to R-Smads are indicated. R-Smad sequence alignment illustration is taken from [39].

Binding of the R-Smad partner transcription factors, like YAP and Pin1, as well as E3 ubiquitin ligases (Smurf1/2 and NEDD4L) involves WW domains on such proteins; these are 38- to 40 amino acid sequences–characterized by a pair of highly conserved tryptophan residues–that typically interact with proline-rich regions (PPXY where X is either serine or threonine or *"PY box"*) or in case of Pin1 with phospho-SP motifs (**Figure 2**). The PY box is located next to the phosphorylated regions in the linker domain of Smad [35].

Taken together, the Smad action turnover switch is controlled by a specific phospho-serine code, which is written by nuclear- and cytoplasmic kinases, CDK8/9, MAPKs, and GSK3. The tandem WW domains in R-Smad binding partners are the readers of this code that dictate Smad transcriptional action, on the one hand, and its elimination, on the other hand. In the BMP pathway, CDK8/9 phosphorylate Smad1/5, creating docking sites for interacting with WW domain of YAP. Subsequent phosphorylation by GSK3 switches off YAP binding and turns on binding to WW domains of Smurf1. In a similar way, TGFβ-activated Smad2/3 are first phosphorylated by CDK8/9, which adds binding sites for Pin1 and GSK3 in order to enhance NEDD4L binding [39]***.***

1.1.4.2. Controlling TGFβ/BMP signaling-basics of ubiquitination/deubiquitination and its role in the pathway

As previously described, Smurf1/2 and other HECT domain E3 ligases carry WW domains, which recognize and bind to PY-rich motifs in the linker region of R-Smads and I-Smads, targeting either their own degradation or that of the binding partners. If bound to I-Smads, Smurfs are recruited to the active receptor complex, promoting its destruction through the UPP, as aforementioned. TGFβ or BMP-activated R-Smads, exploit Smurfs for the elimination of the transcriptional corepressor, SnoN (also referred to as Ski-like), which further enhances the transcriptional function of Smad complexes (***reviewed in*** [40]). Additionally, Smurf1 interacts with and causes the degradation of receptor-activated R-Smads of the BMP pathway–Smad1/5–as well as the membrane localization of the inhibitory Smad, Smad7, and hence influences the output of TGFβ signaling [41, 42]. R-Smads of the TGFβ pathway–Smad2/3–are known to be targeted by Smurf2 [43] and the other member of the HECT domain E3 ligases, NEDD4L [44]. Importantly, Smurf2 has been shown to bind to Smad3 without ubiquitinating the protein [43], whereas NEDD4L represents the major ubiquitin ligase responsible for degradation of both Smad2 and Smad3. The latter is known to also cause the degradation of the common mediator, Smad4, thereby affecting both TGFβ- and BMP responses [45]. These examples demonstrate that polyubiquitination is an important regulatory mechanism through which the TGFβ/BMP pathway is controlled; it has to be noted that these ubiquitin ligases not only target R-Smads but also its negative regulators (e.g., Smad7 and SnoN), as stated above, therefore they are able to both enhance and terminate the signaling.

In addition to the HECT domain enzymes, members of the RING-finger family of E3 ligases are also able to mediate the proteasomal degradation of Smads and/or their partners. Binding of TGFβ-induced R-Smads to the E3 ligase Arkadia (also known as RNF111) causes the degradation of SnoN and Ski, and hence enhances Smad transcriptional responses [46]. Another example of the positive regulation of TGFβ signaling by ubiquitination is the interaction of Smad3 and anaphase-promoting complex (APC), which results in the ubiquitin-dependent degradation of the corepressor, SnoN [47]. The E3 ligase, SCF has been shown to mediate the TGFβ-dependent degradation of Smad3 [48] and Smad4 [49], thus inhibiting the TGFβ responses. In the BMP pathway, the U-box-dependent E3 ligase, STUB1 (also known as CHIP) promotes the degradation of Smad1/5, thereby terminating Smad signaling [50].

Polyubiquitination has been long studied in the regulation of TGFβ/BMP signaling and is mostly coupled to proteasomal degradation of activated R-Smads, although in many cases–as the above examples show–appears to be a positive regulator of the TGFβ and BMP signal transducers. In this regard, monoubiquitination represents an equally important mechanism in controlling both Smad4 and R-Smads' functions, without promoting their degradation [40]. Monoubiquitination came to our understanding of TGFβ/BMP signaling after unraveling the important role of deubiquitinases (DUBs; also known as deubiquitinating enzymes) in the regulation of the pathway; this suggests the existence

of parallel and continuous rounds of mono-, poly-, and de-ubiquitination in the fine tuning of TGFβ/BMP signal transduction.

Ubiquitination–derived from the word "ubiquitin", a 76 amino acid protein–involves consecutive action of three catalytic enzymes: (i) a ubiquitin-activating enzyme (E1); (ii) a ubiquitin-conjugating enzyme (E2); and (iii) a ubiquitin ligase (E3). Initially, E1 attaches the C-terminal glycine residue of ubiquitin to a cysteine residue within its active site via a thioester bond. The activated ubiquitin is then transferred to the cysteine residue of the second enzyme involved, E2. Ubiquitin-loaded E2 is then paired with specific E3 ligases, which catalyze the ligation of ubiquitin to a lysine residue of the target protein, that can be ubiquitin itself (***reviewed in*** [51]). Compared with other post-translational modifications, ubiquitination presents a particularly diverse process; the attachment of ubiquitin to target proteins manifests as monoubiquitination (if ubiquitin is linked as a monomer to a single lysine residue), multiubiquitination (if ubiquitin is linked as a monomer to multiple lysine residues) and polyubiquitination (if several ubiquitin molecules are added serially to the same lysine residue), with each of these types having distinct cellular functions [51]. Monoubiquitination is normally involved in protein transport, DNA repair, and endocytosis, whereas multiubiquitination often acts on the latter [51]. These examples demonstrate that monoubiquitination is mostly a regulatory mechanism of protein function rather than a signal which dictates protein degradation. Polyubiquitination–normally associated to protein destruction through the 26S proteasome–further adds to the diversity of ubiquitination, as it is subclassified depending on any of the seven lysine residues of ubiquitin (K6, K11, K27, K29, K33, K48, and K63) attached to the first ubiquitin molecule. Importantly, a single ubiquitin chain may contain mixed linkages and/or multiple branches. Additionally, ubiquitin can be conjugated with ubiquitin-like modifiers (e.g., SUMO) or small molecules (e.g., phosphate) [51]. These various ways by which ubiquitin chains are assembled provide countless means in regulating biological processes. In light of this, K48-linked polyubiquitination is thought to be majorly involved in proteasome-mediated degradation of the target proteins, while K63-linked polyubiquitin chains are known to rather influence cellular functions such as endocytosis and DNA repair [51]. In this respect, K63-linked polyubiquitination and monoubiquitination are comparable to protein phosphorylation, meaning that they can modulate protein activity, its subcellular localization, as well as its interaction with other proteins without the need for changing the total amount of protein levels. It is, however, worth mentioning the emerging evidence, showing that monoubiquitin(s) of substrates could be marked for protein degradation either through the UPP or autophagy [51].

With respect to the TGFβ- and BMP pathways, monoubiquitination has been shown to modulate Smad activity without promoting its degradation (***reviewed in*** [52]). An early example of such regulatory mechanism is the RING ubiquitin ligase, TRIMM33 (tripartite motif containing 33; also referred to as TIF1γ or ectodermin), which monoubiquitinates Smad4 in its MH2 domain (K519), and hence inhibits both TGFβ- and BMP signals presumably by promoting Smad4 nuclear exclusion as well as disrupting its complex assembly with R-Smads [53]. TRIMM33 also mediates some Smad4-independent

responses, as it competes with Smad4 for binding to C-terminally phosphorylated Smad2/3. By doing so, TRIM33 forms a complex with activated R-Smads, binds to the promoter of nodal target genes (marked with H3K9 trimethylation and H3K18 acetylation), and displaces the chromatin-compacting factor chromobox homologue 3 (CBX3) [54]. During differentiation of embryonic stem cells, the latter process is a prerequisite step for the transcriptional activation of nodal target genes by the Smad complex [54]. It remains to be investigated, however, how general the involvement of TRIM33 is in the regulation of Smad signaling, and whether it activates or represses the target genes. Another example of Smad activity modulation by mono- and multiubiquitination is the inhibition of Smad3 signaling by the HECT-domain E3 ligase, Smurf2. The use of mice strains harboring a target-disrupted Smurf2 allele has shown enhanced TGFβ responses in the absence of the E3 ligase. However, this effect was found to be independent of Smurf2-mediated polyubiquitination, and to be rather associated with Smad3 multiple monoubiquitination at lysine residues in the MH2 domain (K333 and K378), which blocks the formation of Smad3 homo- and heteromeric complexes [55].

The level of ubiquitinated proteins is rapidly readjusted by the action of deubiquitinating enzymes (DUBs), which remove ubiquitin chains from the target proteins (***regulation of TGFβ signaling by deubiquitination has been summarized in two excellent reviews;*** [56, 57]). The catalytic activity of DUBs cleaves the isopeptide bond between the glycine residue of ubiquitin and the lysine on the target protein. Human genome encodes more than 100 DUBs, which are classified in 6 groups based on their catalytic features; (i) ubiquitin-specific proteases (USPs), (ii) ubiquitin C-terminal hydrolyses (UCHs), (iii) ovarian tumor proteases (OTUs), (iv) Josephins, (v) MINDYs (motif-interacting with ubiquitin-containing novel DUB family), and (vi) JAB1/MPN/ MOV34 metalloenzymes (JAMM/MPN+). The latter family are zinc metalloproteases, whereas USPs, UCHs, OTUs, Josephins, and MINDYs contain cysteine peptidase activity. The USP subfamily represents the most abundant group of DUBs (approximately 55%) and has been largely implicated in the regulation of TGFβ and BMP pathways.

Deubiquitinating enzymes act on different components of the TGFβ/BMP pathway; from receptors to Smads. USP family members, USP4, USP11, and USP15 display highly conserved structural domains and protein sequences and have been all mentioned in the deubiquitination of either TGFβ- and/or BMP receptors [56]. USP4 is known to relocate in the plasma membrane upon Akt-mediated phosphorylation where it binds to ALK5, deubiquitinates the protein, and hence protects it from degradation, leading to enhanced TGFβ responses [58]. In this manner, USP4 depletion was shown to inhibit TGFβ-induced epithelial-to-mesenchymal transition (EMT) and Akt-induced breast cancer cell migration [58]. In fact, there is a dynamic balance between polyubiquitination and deubiquitination in controlling the fate of the TGFβR1, ALK5. This is achieved by the action of Smad7, on the one hand, which recruits Smurf1/2 and WWP1 to the receptor, and the deubiquitinase activity of USP4 and USP11, on the other hand [59]. As a proof of concept, in cells overexpressing Smad7, the interaction between ALK5 and USP11 is not promoted upon TGFβ stimulation, suggesting a competition between the two protein in binding to the receptor complex. Similar to USP4, USP11 depletion has been associated with reduced TGFβ-related

transcriptional activity and EMT [59]. USP15 enhances both TGFβ and BMP pathways using a comparable mechanism, which counteracts the process of Smad6/7-Smurf-mediated ALK5 receptor ubiquitination [60]. USP15 amplification has been associated with several cancers, such as glioblastoma, ovarian, as well as breast and has been correlated with high activity of TGFβ signaling [60]. These findings highlight the importance of DUB inhibition, by genetic- as well as chemical means, in the regulation of signal transduction pathways, and hence in controlling diseases (e.g., cancer) which involve such pathways.

A more recent example from the regulation of TGFβ signaling by the USP family of DUBs is USP2a, which has been initially identified as a p53 negative regulator through stabilizing MDM2 [61]. In the TGFβ pathway, USP2a was found to support metastasis by removing K33-linked ubiquitin chains from TGFβRI, leading to enhanced Smad2 recruitment, and thus TGFβ responses [62]. Additionally, TGFβRII phosphorylates USP2a; such phosphorylation promotes the dissociation USP2a and Smad2 from the receptor complex, releasing Smad2 into the cytoplasm [62]. As aforementioned, several USPs (e.g., USP4, USP11, and USP15) are known to deubiquitinate TGFβRI through complex formation with Smad7. This concept of multiple DUBs acting upon the same substrate–TGFβRI–poses interesting questions; do these DUBs work together in controlling TGFβ signaling? or are these enzymes able to act independently and in the absence of one another? The evidence from different lines of studies suggests that both possibilities are likely. On the one hand, USP4 phosphorylation at S445 has been shown to trigger heteromeric complexes with USP11, USP15, and USP19, implicating that these DUBs could act as a complex [58]. On the other hand, mice knock-out experiments demonstrate that upon USP15 depletion, TGFβ signaling is still normally regulated [63], suggesting that in the prolonged absence of one of the enzymes, the others could compensate for its loss.

In addition to receptor deubiquitination, USP15 has been shown to reverse monoubiquitination of Smad3 [64], providing an additional layer of control in the regulation of TGFβ signaling by the ubiquitin system. Consistent with receptor deubiquitination, USP15 was found to reverse monoubiquitination of R-Smads involved in both TGFβ- and BMP pathways [64], whereas its family members–USP4 and USP11–are thought to be limited to TGFβ signaling. Restoring TGFβ signaling through reversible monoubiquitination has been also shown in case USP9x (also known as FAM), which deubiquitinates monoubiquitinated Smad4 (at K519) [65]. Importantly, USP9x knockdown was found to negatively influence the TGFβ-induced p21 transcription, and to be required for TGFβ-mediated growth arrest and cellular migration [65].

Among other members of the USP family of DUBs, CYLD has been reported to inhibit TGFβ signaling through specifically reversing K63-linked ubiquitination of Akt [66] and Smad7 [67]. The former results in inhibition of Akt, subsequent activation of GSK3, which in turn promotes CHIP-mediated degradation of Smad3, thereby inhibiting TGFβ signaling and fibrotic responses [66]. With regards to Smad7, data obtained from CYLD-knockout mice experiments showed that the enzyme targets K63-ubiquitination

sites (K360 and K374) of Smad7, which are important ubiquitination sites required for the activation of transforming growth factor β-activated kinase 1 (TAK1) and p38 MAPK. In light of this, Smad7 knockdown and/or p38 MAPK inhibition in primary T cells was found to impair differentiation of regulatory T cells (Tregs), highlighting the significance of TGFβ signaling in the development of Tregs which is mediated through K63-linked polyubiquitination of Smad7 [67].

Other families of DUBs have been also shown to regulate the TGFβ/BMP pathway both positively and negatively. From the OTU family, OTUB1 is known to impact TGFβ signaling via interaction with active Smad2/3 [68]. What appears to be special with regards to this deubiquitinating enzyme is that unlike other DUBs, it acts only upon TGFβ exposure, reversing the polyubiquitination of C-terminally phosphorylated Smad2/3, and hence enhancing TGFβ-mediated gene transcription and cellular migration [68]. The mode of action of OTUB1 highlights the dynamic interplay between phosphorylation and ubiquitination in controlling TGFβ signaling. Although being known to cleave K48-linked ubiquitin chains, OTUB1 has been shown to stabilize the active Smad2/3 complex in a noncanonical manner and independently of its catalytic activity and through inhibition of ubiquitin transfer from E2 to E3 [56]. TGFβ activity can also be sustained by the UCH family member, UCH37 (also known as UCHL5), which acts through targeting the type I receptor, ALK5. Accordingly, UCH37 knockdown has been shown to decrease TGFβ-induced transcriptional responses and lateral cell migration during the early phase of TGFβ signaling activation [69, 70]. The zinc metalloprotease DUBs, AMSH [71] and AMSH-LP [72] exert a positive feedback on the TGFβ- and BMP signals via sequestering Smad6 and Smad7, respectively.

Several DUBs have been reported to have a role in the modulation of non-Smad TGFβ signaling. Among the USP family members, USP4, which is thought to be a positive regulator of TGFβ signaling as aforementioned, has been reported to suppress the noncanonical TGFβ pathway through deubiquitinating TAK1, thus inhibiting the activation of TNFα-induced NF-κB [73]. The negative regulation of TAK1 could also be achieved by the action of USP18 during the process of T helper 17 (Th17) cell differentiation. In this regard, USP18 knockout mice was found to be defective in the generation of Th17 and resistant to autoimmune encephalomyelitis, majorly through restricting IL2 expression [74]. Focusing on the TRAF6-TAK1-p38 MAPK axis, A20 from the OTU family of DUBs has been reported to inhibit the pathway by deubiquitinating TRAF6. Accordingly, knockdown of either A20 or Smad6, which acts as a transporter of A20 to direct the enzyme to TRAF6, results in increased apoptosis, and maintains the phosphorylation of TAK1 and p38/JNK MAPK. This indicates that the A20-Smad6 axis is essential in the modulation of TGFβ-induced TRAF6-TAK1-p38 MAPK signaling and shows a new role for Smad6–not Smad7–in controlling the noncanonical TGFβ pathway [26]. The abovementioned examples shed light on the involvement of deubiquitinating enzymes not only in the regulation of canonical TGFβ signaling but also in that of the noncanonical pathways.

The regulation of TGFβ/BMP signaling by ubiquitination and deubiquitination has been mostly focused on the receptor-activated R-Smad pool, which is followed by multiple sequential

phosphorylations in the linker region, marking activated R-Smads for degradation (**see Figure 2**). Although this regulatory cycle holds true, the small constant pool of Smurfs and other E3 ligases as well as deubiquitinating enzymes implicates that they may have some effect on steady-state Smad stability (***reviewed in*** [40]). In light of this, Smurf1 has been shown to bind to the PY motifs of Smad1 and Smad5, promoting their degradation [41]. Additionally, Smurf2 is known to control baseline properties of Smad1 and Smad2 through ubiquitination [75]. Although not being a target of Smurf2, basal Smad3 has been reported to be stabilized by a proteasome inhibitor [76]. One of the E3 ligases that can ubiquitinate Smad3 at steady state is CHIP [77], which was also found to contribute to the ubiquitination of nonactivated Smad1 [50]. Similar to the regulation of receptor-activated R-Smad levels, steady-state R-Smad stability appears to be finetuned by a continuous interplay between phosphorylation and ubiquitination. A clear example of such regulatory mechanism is the role of the scaffolding protein, Axin and its associated kinase, GSK3β in basal Smad3 turnover. Mechanistically, Axin facilitates GSK3β-mediated phosphorylation of Smad3 at a conserved threonine residue in the MH1 domain (T66), which triggers Smad3 ubiquitination and degradation in the absence of TGFβ stimulation, and hence desensitizing cells to TGFβ [78]. Other kinases have also been reported to control baseline activity of R-Smads in different contexts. In this regard, activated Erk1 has shown to positively regulate Smad2 signaling by phosphorylating Smad2. Importantly, the Erk-mediated elevation of Smad2 protein levels was found to occur as a result of increased half-life of the protein [21].

1.1.5. TGFβ signaling–tumor suppressor *vs.* tumor promoter

The suitability of the TGFβ pathway as a target in cancer therapy can only be evaluated by a deep understanding of its dual role in tumorigenesis, i.e., how tumor cells in contribution with stroma evolve to evade the tumor-suppressive functions of this signaling, and more dramatically hijack the signal to their advantage, turning it into a tumor-promoting response. As aforementioned, TGFβ- and BMP responses are greatly influenced by the cellular context as well as by integration with other signaling pathways. In cancer, this pleotropic reaction to the pathway, makes TGFβ a powerful tumor suppressor in the context of early stage tumors, but an enhancer of invasion and metastasis in the context of more advanced cancer (***summarized in two excellent reviews; by Massague J. in*** [35], ***by Seoane J. and Gomis R. in*** [79]). This dichotomy although paradoxical, has its own logic. TGFβ's cytostatic effects are important in normal tissue homeostasis. As cells acquire mutations and reach a pre-malignant state, TGFβ triggers apoptotic cell death. In some cancer types, this can select for mutations, which eliminate the TGFβ pathway (e.g., inactivating mutations in *TGFβR2* and *Smad4* in gastrointestinal- and pancreatic tumors). Upon TGFβ signaling elimination, tumor cells can create a TGFβ-rich microenvironment, which favors cancer progression through effects on the stroma. Certain tumor types, such as breast, melanoma, and glioma are even more skillful in using the remaining signaling capacity to their benefit, leading to a diverse set of tumorigenic features ranging from immune evasion to tumor growth, migration, invasion, and metastasis.

1.1.5.1. TGFβ/BMP signaling in cancer–the cytostatic and apoptogenic program

TGFβ's signaling mediators, R-Smads, regulate the expression of hundreds of genes including those that are involved in cell cycle progression (***reviewed in*** [79]). TGFβ inhibits G1 to S phase transition through either the induction of CDK inhibitors or the suppression of *c-Myc* expression. In epithelial cells, TGFβ induces the expression of p15 and p21, thereby inhibiting the formation of cyclin D/CDK4/6 and cyclinE- and A/CDK2 complexes, respectively. Smad co-transcription factors, FoxO and Sp1, associate with the Smad3-Smad4 complex in order to target the promoters of *CDKN2B* and *CDKN1A*, coding for p15 and p21, respectively. *c-Myc*, a key transcriptional modulator of cell growth and division, is downregulated by TGFβ in mammary epithelial cells through the action of Smad3-Smad4 complex in combination with several other transcription factors, such as E2F4, E2F5, and CCAAT/enhancer binding protein β (C/EBPβ).

In addition to the regulation of cell cycle, TGFβ can induce apoptosis via various mechanisms, including an increase in the expression of death-associated protein kinase (DAPK) and GADD45β, the activation of death receptor, FAS, as well as facilitating the binding of Bim to Bcl-2 and Bcl-xL (***reviewed in*** [79]). The pro-apoptotic effects of Bim could be also promoted through Smad3-dependent transcription of the MAPK phosphatase, MKP2. Moreover, TGFβ can block the expression of pro-survival signaling molecules, such as survivin and Akt (***reviewed in*** [79]). As previously described, noncanonical (non-Smad) TGFβ signaling can induce apoptosis through the TRAF6-TAK1-JNK/p38 MPAK axis in some cell types, an effect which can be facilitated by Smad7. Additionally, some of the TGFβ's pro-apoptotic responses could be linked to the tumor-suppressor, p53, involving Smad7, p38 MAPK, and the DNA-damage sensitive kinase, ATM [80]. Interestingly, TGFβ-mediated apoptosis can also occur as a result of oxidative stress through the inhibition of anti-oxidant genes or the induction of NADPH oxidase (Nox) activity; this leads to mitochondria-dependent apoptosis, at least in part due to the regulation of various Bcl-2 family members [81].

TGFβ-mediated tumor suppression does not always occur through direct growth-inhibitory effects on target cells and may be exerted via indirect effects on the tumor microenvironment. TGFβ limits the production of paracrine factors in stromal fibroblasts and inflammatory cells. In this regard, inactivation of TGFβRII in mouse fibroblasts has been shown to result in intraepithelial neoplasia in prostate and invasive squamous cell carcinoma of the forestomach, possibly through the activation of hepatocyte growth factor (HGF) signaling [82].

1.1.5.2. TGFβ/BMP signaling in cancer–from evading TGFβ's tumor suppression to cancer progression

During the tumorigenic path, cancer cells tend to escape the action of tumor suppressors, such as p53 and APC and/or to activate oncogenes for example, *KRAS* and *HER2*. If a stem or progenitor cell with such oncogenic mutations is exposed to TGFβ, it will become liable to undergo apoptosis.

Therefore, this cell must acquire additional alterations which, among other events, disable the tumor-suppressive responses to TGFβ in order to advance on the oncogenic track. One way to tumor evolution are genetic variations in the core components of TGFβ-Smad signal transduction through the accumulation of somatic mutations and epigenetic silencing. A prominent example of such alterations are colorectal cancers, which consistently incur replication errors in the TGFβRII gene, *TGFBR2* (***reviewed in*** [83]). Additionally, in gastrointestinal tumors, methylation status of *TGFBR2* is known to be correlated with its decreased expression levels. In the BMP pathway, mutations in the *BMPR1A*, coding for the BMP type I receptor (ALK3) cause juvenile polyposis, which predisposes patients to intestinal polyposis and cancer (***reviewed in*** [84]).

Mutations of *R-Smads*, in particular *Smad3*, are infrequent in cancer and have been reported only in limited number of cases [85]. For example, *Smad3* loss of expression has been described in gastric cancer and T cell lymphoblastic leukemia, and loss of heterozygosity of the 18q21 chromosomal region containing *Smad2* and *Smad4* has been reported in pancreatic- and colon tumors [86]. Unlike *R-Smads*, *Smad4* (deleted in pancreatic carcinoma locus 4; DPC4) mutations are a common genetic alteration in cancer, and in pancreatic cancer occur as frequent as mutations in the *KRAS*, *TP53*, and *CDKN2A* [83, 87]. *Smad4* mutations are also recurrent in colorectal tumors, and have been observed in esophageal and other cancers, with different incidence rates [88, 89]. The disruption of TGFβ signaling can also occur at the level of singling's negative modulators, including inhibitory Smads, and transcriptional corepressors. In support of this, increased expression of Smad7 and Smad6 has been shown to blunt TGFβ signaling, and hence prevents its anti-tumoral effects in endometrial carcinomas and thyroid follicular cancers (in case of Smad7), as well as pancreatic- and breast tumors (in case of Smad6) [90-93]. Moreover, Smad7 over-expression in immune cells has been linked to chronic inflammation in the colonic mucosa, similar to the role of *Smad4* mutations in the stromal cell proliferation and submucosal invasion [79]. In addition to I-Smads, deletions and amplifications of *SKI* and *SKIL* (the gene coding for SnoN) have been observed in gastrointestinal tumors [94].

Once cancer cells are released from TGFβ-mediated tumor suppression, they will use the remaining ability of the pathway to support tumor cell growth, migration and invasion. Additionally, TGFβ has pleotropic effects on the immune system through inhibiting most components of both adaptive- (e.g., T cells) and innate (e.g., macrophages and natural killer (NK) cells) immune system directly and indirectly [95]. In this context, TGFβ stands out as a physiological immunosuppressor; the immunosuppressive effects of this growth factor allow tumor cells to escape the anti-cancer immune response, and hence TGFβ represents an appealing therapeutic target with regards to cancer immunomodulation. In light of this, the inhibitory effect of TGFβ on NK cells is one way through which cancer cells create a permissive stroma, which promotes tumor progression [79]. As an example, TGFβ regulates the anti-cancer effects of tamoxifen *in vivo* via controlling the population of NK cells; in this regard the tamoxifen-resistant breast cancer cell line, LCC2 is sensitized to tamoxifen treatment upon TGFβ inhibition [96]. Moreover, this cytokine can inhibit tumor cell death via regulating gene expression in monocytes and macrophages

[79]. In monocytes, TGFβ induces the production of pro-inflammatory molecules, including interleukin 1 and 6 (IL1 and IL6), while blocking oxygen-free radical production [97]. In macrophages, TGFβ suppresses the expression of chemokines, including macrophage inflammatory protein 1α and 2 (MIP-1α and MIP-2) and the chemokine CXCL1 factor, the cytokine granulocyte-macrophage colony-stimulating factor (GM-CSF) and interleukins IL-1β, IL-8, and IL-10, altogether preventing early premature immune response [98].

Similar to TGFβ's anti-cancer responses, pro-cancer effects of this cytokine rely on the tumor microenvironment; in this context the tumor alters the host immune response to TGFβ in order to favor tumor initiation and progression. Tumor microenvironment consists of various cell types, including fibroblasts, endothelial- and immune cells, as well as extracellular matrix proteins [99]. TGFβ is produced in these tumor-adjacent cells and activates stromal signals; the outcome of these signals depends on whether the epithelial tumor cell retains a functional TGFβ signaling or not, as described above. For example, in certain tumor types (e.g., colorectal- pancreatic- and prostate cancers) increased levels of TGFβ are associated with poor prognosis and relapse due to the activated TGFβ signaling in tumor-adjacent cells, not in the malignant epithelial tissue [79]. Things are more devious in breast cancer, gliomas, and melanomas, as they have an intact Smad machinery, thus they are able to respond to stroma-derived TGFβ, resulting in Smad-dependent gene transcription, which in such context is profitable for metastasis [35]. Examples of this pro-metastatic program include the Smad-dependent expression of angiopoietin-like 4, which promotes extravasation of circulating tumor cells in breast cancer, induction of parathyroid hormone-like hormone, IL11, connective tissue growth factor (CTGF), and NOTCH ligand, JAGGED1, which allows cancer cells for osteolytic metastasis in breast carcinoma and melanoma (***reviewed in*** [35]).

During EMT, as the abbreviation describes, cells lose their epithelial features and acquire mesenchymal properties, including down-regulation of adherent junctions and associated proteins (e.g., E-cadherin), elevated matrix metalloproteinases (MMPs), as well as an architectural rearrangement of the cytoskeleton, which altogether result in cell motility and invasiveness (***reviewed in*** [100]). EMT has a central role in various events during embryonic development in which cells originate far from their final destination, such as gastrulation and formation of neural crest and structures of the heart. In cancer, the EMT program helps tumor cells to disseminate and release to the surrounding environment. In addition, EMT has been associated with chemo-resistance in breast- and pancreatic cancer. Importantly, tumor cells often return to an epithelial phenotype through a reversed phenomenon, referred to as mesenchymal-to-epithelial transition (MET); MET-related transcription factors (e.g., Prrx1) uncouple stemness and EMT, and are required by tumor cells to colonize at distant sites and form metastases [100]. TGFβ regulates a set of transcription factors, including Snail (Snail1), Slug (Snail2), TWIST, Par6, and the zinc finger E-box binding homeobox proteins, ZEB1 and ZEB2 [101]. Moreover, in certain contexts, Smad-mediated complexes may have an indirect role in controlling the expression of Snail1/2 via high mobility group A2 (HMGA2). Through its role in inducing EMT, TGFβ can support the generation

of fibroblasts from epithelial cells, as well as from endothelial cells via a closely-related process, referred to as endothelial-to-mesenchymal transition (EndMT) [79]. EndMT has been reported to be an essential source of cancer-associated fibroblasts (CAFs) in a pancreatic mouse model, and myofibroblast accumulation has been repeatedly described in the pathogenesis of solid tumors [102, 103]. Although TGFβ-induced EMT is generally thought to be pro-tumorigenic, in TGFβ-sensitive pancreatic ductal adenocarcinoma (PDA) cells, EMT can become pro-apoptotic through the action of Sox4. In these cells, Smad4 is necessary for EMT induction but not for the expression of Sox4, therefore the Smad4 status of the cell defines whether TGFβ-induced EMT favors tumor progression or a pro-apoptotic response [104]. Taking all these findings into consideration illustrates that there is a strong interplay between TGFβ and EMT induction, and that the overall response is highly subject to cell type- and context-specific factors.

One of the biological functions of TGFβ is enhancing the tumor initiating qualities of cancer stem cells (CSCs), which becomes particularly significant in the process of metastatic colonization when cells are prone to severe conditions, and eventually only a few can succeed in forming secondary tumors at distant regions. If these cells do not have tumor-initiating properties, metastasis may not cause clinical symptoms (***the role of CSCs in metastasis has been summarized by Trumpp A. et al. in*** [105]). In light of this, TGFβ has shown to promote the CSC potential in glioblastoma, and to cooperate with Wnt signaling in order to activate mesenchymal status of CSC in association with EMT [106]. TGFβ supports a $CD44^{high}/Id1^{high}$ glioma-initiating cell population, which is responsible for tumor initiation, relapse, and chemo-resistance [106]. Additionally, TGFβ helps the self-renewal traits of glioblastoma stem cells through the Smad-dependent induction of *Sox2*, *growth factor beta polypeptide* (*PDGFB*), and *leukemia inhibitory factor* (*LIF*) [107, 108]. LIF is a cytokine with an essential role in embryonic stem cells; it can be secreted by cancer cells in some glioblastoma to promote self-renewal of cancer-initiating cells, and hence tumor relapse [108]. TGFβ can also induce the expression of the CSC marker, CD133 in hepatic epithelial cells, which is associated with a more aggressive and resistant phenotype [109]. In chronic myeloid leukemia (CML), TGFβ has a critical role in maintaining CML stem cells through the TGF-β–FOXO–BCL6 axis in cooperation with nutrient signaling [110]. As described above, TGFβ's effects on EMT and tumor-initiating properties occur concomitantly under certain circumstances, and both events (EMT- and stem cell-like markers) are observed in circulating tumor cells from patients with metastasis. An example is the induction of stem cell traits by enforced expression of EMT-related transcription factors, such as TWIST in breast cancer cells [111]. In this context, the role of other TGFβ family members has to be carefully taken into consideration, as they can suppress self-renewal and enhance cancer cell differentiation. For example, lung-derived BMP ligands have been shown to impose metastatic dormancy in breast cancer cells, and that the BMP antagonist, Coco supports metastatic reactivation [112].

1.2. Indirubin and indirubin derivatives (IRDs)

The successful isolation and identification of Qinghaosu (青蒿素) as an anti-malarial agent from more than 500 recipes of traditional Chinese medicine (TCM) has promised clinical application of TCM-derived active ingredients against a variety of diseases, ranging from infections to metabolic disorders and cancer. Indirubins represent the active ingredient of a Chinese herbal prescription named Dangui Luhui Wan, which has been traditionally used in TCM for the treatment of various chronic diseases. Its anti-tumor effect was first reported in 1970s in a clinical trial involving 22 patients with CML. Mechanistically, the cytotoxic effects of indirubins have been mainly attributed to their inhibitory function against a number of membrane-bound- nuclear- and cytoplasmic protein kinases, such as CDKs, GSK3, as well as receptor tyrosine kinases (RTKs). Recent discoveries are increasingly adding new molecular targets to indirubins' biological profile. However, the clinical application of indirubin and IRDs is largely hampered due to the limited pharmacokinetic properties. Therefore, a lot of effort has been directed to the chemical modification of indirubin's backbone, leading to the synthesis of a number of different derivatives with new biological activities, which influence various cellular signaling networks associated with cancer and other diseases (***we have recently discussed the impact of indirubins on cancer-associated signaling networks in*** [113]).

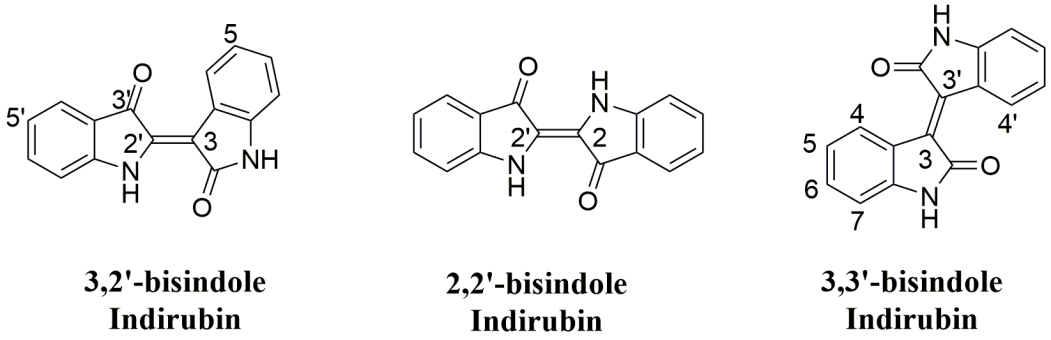


Figure 3. Chemical structures of indirubin, indigo (2,2'-bisindole indirubin), and isoindigo (3,3'-bisindole indirubin).

1.2.1. Indirubins–towards better bioavailibilty: *chemical modification and pharamceutical formulation*

The synthesis of indirubin was first described by Baeyer in 1881 using isatin and indoxyl in MeOH under basic condition. Since the indoxyl moiety is not stable and spontaneously converts to indigo upon exposure to air, Russell and Kaupp replaced this chemical with the more stable indoxyl acetate in 1969. This method under mild basic condition is suitable for the synthesis of most indirubin derivatives. However, 7-aza indoxyl acetate is not stable and degrades under this condition. The synthesis of aza-indirubins is therefore preferentially conducted in acetic acid with 10% concentrated HCl, as developed and reported by Martinet and Dornier (***reviewed in*** [114]).

Indirubin, indigo, and isoindigo are three isoforms of the small family of bis-indole alkaloids, which share the bis-indole's backbone (**see Figure 3**). Similar to its isoforms, indirubin and many of its derivatives have a poor solubility in water and in major organic solvents because of the stable intra- and inter-molecular hydrogen bonds between 1-NH and 2-CO and between 1'-NH and 2-CO, respectively (**Figure 4**) [114]. Studying the co-crystal structure of indirubin-3'-oxime and indirubin-5-sulfonate in complex with the ATP-binding pocket of CDK2 has unveiled potential positions on the bis-indole moiety of indirubin for chemical modifications, which allow improved pharmacokinetics with minimal negative effects on bioactivity–i.e., using 5- and/or 3'-positions facing the ribose/triphosphate canal of the ATP-binding pocket of the kinase, as well as 5'-position, which is free and open to the solvent (**Figure 5**) [115-117]. The modification at 3'-position is possibly the most successful approach in order to introduce substituents with basic functional groups for improvement of solubility and biological activity. In view of this, we and others have synthesized the water-soluble indirubin derivatives using the 3'-position. In addition, 5- and 5'-positions could also be used for the synthesis of IRDs, however with more complicated protocols [114].

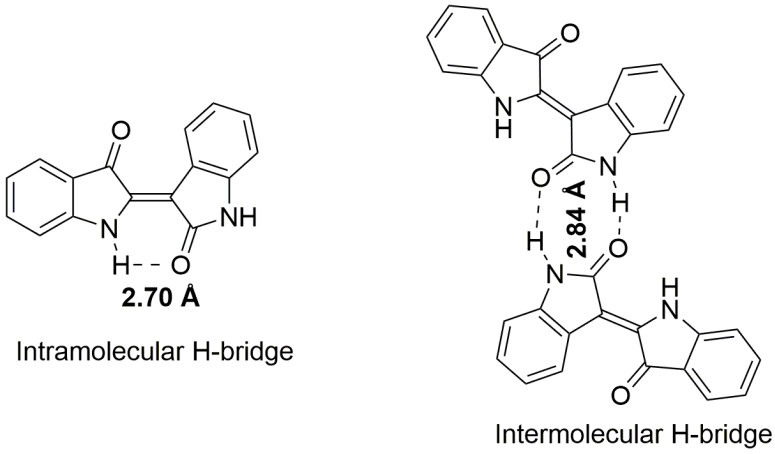


Figure 4. Inter- and intra-molecular hydrogen bonds cause indirubins' poor water solubility.

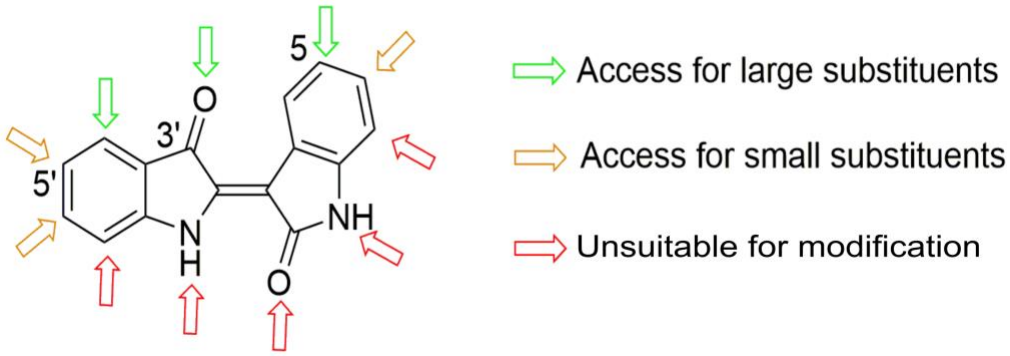


Figure 5. Demonstration of possible positions for chemical modification of the indirubin backbone. A color code has been used to indicate the suitability of each position, with green being representative of full accessibility, orange representative of limited accessibility, and red presenting unsuitable positions.

As mentioned above, the bioavailability of indirubins suffers from the poor water solubility of the molecule; an early finding in support of this idea is the study on the interaction of indirubin and p-glycoprotein, showing the low binding affinity when the pH is in the range of 2 to 11 [118]. Although chemical modifications in the aforementioned positions have led to the synthesis of more water-soluble derivatives, low pharmaceutical activity of those analogues (e.g., inidirubin-5-sulfonate) due to the extra hydrophilic property limits the use of IRDs in pre-clinical and clinical models [118]. Alternatively, pharmaceutical formulation presents an opportunity to enhance solubility and bioactivity. In this regard, both self-micro-emulsifying and self-nanoemulsifying drug delivery systems (SMEDDS and SNEDDS, respectively) have shown promising results [119, 120]. Using SMEDDS, 50 ng/mL indirubin was detected in the plasma after oral administration, whereas a significantly higher plasma concentration (1800 ng/mL or ~10 µM) was achieved using SNEDDS. Given that the reported IC_{50} values of indirubin and indirubin derivatives in cellular assays are normally below 10 µM, SNEDDS provides a suitable approach for the clinical application of indirubins.

1.2.2. Indirubins–molecular mode of action

1.2.2.1. molecular targets–cyclin dependent kinases

Indirubins were primarily found to have a high inhibitory activity towards cyclin-dependent kinases with IC_{50} values down to the nano molar range (5-100 nM). As previously mentioned, CDKs are serine/threonine protein kinases with a major role in the regulation of R-Smad activity and stability. Outside the context of TGFβ/Smad signaling, CDKs together with their binding partners, cyclins, regulate the progression of cell cycle. There are three main transition steps at which the cell cycle regulators including CDKs/cyclins decide on passing of cells through the next phase; G1/S, G2/M and during mitosis (M) phase. The first checkpoint, G1/S is mostly regulated by CDK4,6/cyclin D and CDK2/cyclin E complexes, which are responsible for phosphorylation, and hence inactivation of retinoblastoma protein (Rb). CDK2/cyclin A is essential for the S phase progression and is accompanied by the action of CDK1/cyclin B and CDK2/cyclin B complexes at the G2/M transition, preparing cells for mitotic division [113].

With regards to CDKs, indirubin and its analogues have been shown to bind to and inhibit CDK1/cyclin B, CDK2/cyclin A, CDK2/cyclin E, and CDK5/p25 complexes [113]. Crystal structure studies of CDK2 in complex with indirubin derivatives have revealed that indirubin occupies the ATP binding pocket of the CDK's catalytic subunit via hydrophobic interactions and hydrogen bonds [115]. The more water-soluble derivatives, indirubin-3'-monoxime and indirubin-5-sulfonate have been demonstrated higher affinity for CDKs and potent inhibition of the enzymes. While the sulfonate analogue fails to be up-taken by tumor cells, the oxime analogue (3IO) is cell-permeable and has demonstrated high efficacy in several malignancy models, making it the most widely-studied derivative of indirubins (***reviewed in*** [121]. In this regard, Hoessel *et al.* showed that the kinase inhibitory activity of indirubin-3'-oxime causes reduced phosphorylation of Rb, and subsequently induces G1- and G2/M

cell cycle arrest in the human leukemia cell line, Jurkat [115]. Moreover, the compound has shown anti-proliferative effects in the breast cancerous cell line, MCF-7 by inducing a G2/M cell cycle arrest through inhibiting CDK1 [122]. Furthermore, it was found to increase the expression of the CDK inhibitor, p27, and to reduce that of CDK2 and cyclin E, resulting in a G0/G1 arrest in human neuroblastoma cell lines [123].

1.2.2.2. molecular targets–glycogen synthase kinase 3

The indirubin-mediated CDK inhibition mostly occurs in combination with the inhibition of glycogen synthase kinase 3β. GSK3β is a multifaceted serine/threonine kinase, and a key player of the canonical Wnt signaling pathway, which controls several physiological processes ranging from cell cycle and metabolism to neuroprotection and dorsal-ventral patterning during embryonic development [124]. In addition, GSK3 represents an important enzyme in insulin signaling; over-expression of GSK3β has been reported in type 2 diabetes and GSK3β inhibitors have shown anti-diabetic effects *in vitro* and *in vivo* [125]. Moreover, GSK3β in cooperation with CDK5 mediates hyperphosphorylation of the microtubule binding protein, tau which is a diagnostic marker in certain neurodegenerative diseases, including Alzheimer's disease [121]. In light of this, it has been reported that several indirubin derivatives which were initially introduced as potent CDK inhibitors, block the GSK3β enzyme in an ATP-competitive manner similar to their interaction with CDKs, and by doing so represent the first class of GSK3β pharmacological inhibitors that are effective at low nano molar concentrations (5-50 nM) [113]. Leclerc and colleagues showed that the indirubin-3'-oxime derivative inhibits tau protein aggregation *in vitro* and *in vivo* by the concomitant inhibition of two protein kinases, CDK5/p25 and GSK3β. Additionally, this analogue was found to block the phosphorylation of dopamine- and cyclic adenosine 3′-5′-monophosphate-regulated phosphoprotein 32 (DARPP-32) by CDK5, thus imitating dopamine effects in the stratum [126]. Looking for new scaffolds that are able to selectively target GSK3β, it has been reported by Meijer *et al.* that a simple substitution at position 6 increases the GSK3 specificity over CDKs. In this regard, 6-bromoindirubin derived from "tyrian purple" dye of mollusks and its oxime substituent, 6-bromoindirubin-3'-oxime (6BIO) have been reported as selective small molecule inhibitors of GSK3α/β. The compounds were found to mimic Wnt signaling, and to reduce phosphorylation of β-catenin *in vivo*. Additionally, they showed more than 16-fold higher specificity for CDK2 as compared to CDK5 [127]. With this information in mind, indirubin and its derivatives present promising candidates not only in cancer therapy, but also in the treatment of other diseases in that selective GSK3 inhibition is profitable, such as diabetes and Alzheimer's disease. Importantly, the dual and opposing roles of GSK3β in oncogenesis should be taken into consideration when applying GSK3β inhibitors for cancer treatment. On the one hand, GSK3β inhibits Wnt signaling by phosphorylation and subsequent proteasomal degradation of β-catenin; in this context GSK3β acts as a tumor suppressor. On the other hand, it can switch to a tumor promoter by stimulating the activity of Wnt/β-catenin signaling through phosphorylation of Wnt co-receptor, LRP 5/638 [128]. The latter mechanism might explain the tumor-suppressive effects observed with GSK3β inhibitors, such as indirubins. In addition, this kinase is increasingly identified as a modulator of invasion and metastasis. Williams and colleagues

demonstrated that the GSK3β inhibitory activity of indirubins offers a new therapeutic mechanism through preventing invasion of glioma cells and glioma-initiating cells *in vitro* and in a glioma bearing mouse model [129]. Moreover, indirubins were found to decrease cellular migration of endothelial cells, an essential component of tumor angiogenesis [129]. Mechanistically, GSK3β inhibition causes dephosphorylation, thereby stabilization of β-catenin, leading to reduced migratory phenotypes. Despite the significance of β-catenin in the proposed mechanism, they showed that silencing this protein does not completely rescue the anti-migratory effects of indirubins in glioma [129], implicating additional targets other than GSK3β in mediating the observed effects. In a different study, Braig *et al.* reported the IRD 6BIO significantly diminishes migration of breast cancer cells into the lungs at subtoxic concentrations (1 mg/kg) in an aggressive breast cancer mouse model (4T1). While GSK3β as well as the phosphoinositide-dependent tyrosine kinase 1 (PDK1)12 should be granted for the overall anti-migratory activity of 6BIO in this study, they did not appear to be the major targets. Instead, the inhibition of two other kinases, Src and janus kinase (JAK), followed by decreasing the activity of their substrate, signal transducer and activator of transcription 3 (STAT3) turned out to be the dominant mechanism. STAT3 inhibition caused decreased expression of the cellular migration regulators C-terminal tension like protein (CTEN) and MMP2 [130]. The combined inhibition of multiple kinases by the IRD 6BIO enables the simultaneous targeting of different cellular networks related to cancer, ranging from extensive proliferation of cells to migration, invasion, angiogenesis and metastasis.

1.2.2.3. molecular targets–receptor and non-receptor tyrosine kinases

The invention of imatinib–the first FDA-approved tyrosine kinase inhibitor (TKI)–in late 1990s has transformed the approach to the treatment of various cancers by offering a new form of targeted therapy, i.e., selective inhibition of protein tyrosine kinases (PTKs). PTKs represent one of the major enzymes in the process of cellular signal transduction; they possess tyrosine kinase activity, which enables them to catalyze the transfer of phosphate groups from ATP to the tyrosine residues on their substrates (e.g., STATs), in order to phosphorylate them. The phosphorylated substrates then transfer the signal to the nucleus in order to regulate various cellular functions; cell survival- growth- and differentiation, to name a few. Based on their structure, PTKs are classified as either receptor tyrosine kinases (RTKs) or non-receptor tyrosine kinases (NRTKs). The former includes epidermal growth factor receptor (EGFR), platelet-derived growth factor receptor (PDGFR), vascular endothelial growth factor receptor (VEGFR) as well as insulin receptor (InsR) family and so on. Unlike RTKs, NRTKs have no extracellular structure, instead they are usually tethered to the cell membrane or reside in the cytoplasm; the examples include ABL kinase, Janus tyrosine kinase (JAK), and Src family kinases (SFKs). NRTKs have a crucial role in signal transduction pathways initiated by various membrane receptors (e.g., tyrosine kinase receptors, growth factor receptors, and G-protein coupled receptors); additionally, they are involved in the so-called "inter-pathway" crosstalks. The latter is exemplified in the effects of activated STAT3 on Ras- and PI3K/Akt pathways, and in the association of JAK2 with PI3K- and Erk signalings (***reviewed in*** [131]). In light of this, it has been reported by Nam *et al.* that a number of IRDs–designated as E564, E728, and E804–inhibit the constitutively active STAT3 signaling in human breast- and prostate cancer cell

lines. One of the derivatives, indirubin-3'-(2,3 dihyroxypropyl)-oximether, referred to as E804, was found to have a direct interaction with Src at sub-micromolar concentrations (IC_{50} = 0.43 μM), leading to reduced tyrosine phosphorylation of its substrate, STAT3 in both an *in vitro* kinase assay as well as in human breast (MDA-MB-468 and -435) and prostate (DU145) tumor cells. Moreover, two of the STAT3-induced pro-survival proteins, Mcl-1 and survivin, were strongly decreased in response to E804 in breast cancer cells, triggering apoptosis. Importantly, the induction of apoptosis by E804 was not found in cells lacking elevated STAT3 activity, suggesting that the anti-tumoral effects of E804 occur at least in part through targeting Src/STAT3 signaling [132]. Another member of the STAT family with a key role in tumorigenesis, particularly in hematopoietic cancers, is STAT5, which is activated via the action of the non-receptor tyrosine kinases, BCR-ABL and SFKs [133]. With regards to indirubins' effects on the status of the latter kinases, Nam *et al.* have shown in a different study that among a series of derivatives, IRD E804 is the most potent analogue to induce apoptosis in CML through inhibiting SFK/STAT5 signaling, down-stream of BCR-ABL kinase. This inhibition was followed by the small molecule's pro-apoptotic effects against human K-562 CML cells, the imatinib-resistant KCL-22 CML cells, expressing mutant BCR-ABL, as well as the CD34-positive primary CML cells from patients. Moreover, the expression of Mcl-1, known to be upregulated in CML, was decreased upon E804 treatment [134]. It is worth mentioning that the IRD E804 is also a potent CDK inhibitor, a common feature of most indirubin derivatives. Through CDK-inhibition, the compound has shown anti-proliferative effects in different *in vitro* cancerous models, namely human HCT116 colon cancer, human MCF-7 breast cancer, as well as in the human lung carcinoma cell line, LXFL529L [121].

As described earlier, JAK and Src kinases share STAT3 as their substrate. It has been reported that prolonged inhibition of Src triggers a compensatory mechanism in which STAT3 is alternatively phosphorylated and activated through the JAK/STAT3 interaction, mediating cancer cell survival and proliferation [135]. Therefore, dual inhibition of SFKs and JAK may have synergistic effects by blocking STAT3 reactivation, a common drawback of Src-specific inhibitors (e.g., dasatinib). In view of this, the IRD E738, possessing a propanediol group at 3'- and a methoxy group at 5-position has been introduced by Nam and colleagues as the first dual inhibitor of JAKs and SFKs with IC_{50} values between 0.7-74.1 nM and 10.7-263.9 nM, respectively. As compared to E804, E738 showed 40-fold increase in the Src inhibitory activity. Combined inhibition of JAK/Src by IRD E738 caused reduced tyrosine phosphorylation of STAT3, followed by reduced Mcl-1 levels, and apoptosis in human pancreatic cancer cells [136]. Several other studies have identified the JAK/Src/STAT signal transduction as a mode of action of indirubin's bromo-substituents. It has been reported by Liu and colleagues that the compound, 6BIO induces apoptosis in melanoma by selective targeting of JAK/STAT3 signaling. 6BIO, among a series of IRDs, showed the most efficacy in reducing the phosphorylation of JAK and STAT3, and was found to suppress tumor growth in a human melanoma xenograft mouse model [137]. In a different study, the anti-cancer activity of a 7-bromoindirubin derivative (known as MLS-2438) against human melanoma cells was attributed to the inhibition of STAT3 and Akt kinases. The compound also reduced tumor growth *in vivo* with relatively low toxicity [138]. Interestingly, 7BIO, unlike its close isomers 5- and 6-

bromoindirubins, showed insignificant inhibition of the classical targets, CDKs and GSK3β, which account for most of the anti-tumoral effects of indirubins [139]. Instead, 7BIO with a minimal kinase inhibitory activity was able to induce a rapid cell death without triggering the typical markers of apoptosis [139, 140].

In addition to non-receptor PTKs, RTKs could be also targeted by indirubins. An early example is the anti-angiogenic properties of the water-soluble derivative, indirubin-3'-oxime, which were found to occur through inhibiting VEGFR-2 [141]. Besides the oxime derivative, indirubin itself has been shown to inhibit tumor angiogenesis in human prostate cancer cell line, PC-3 via preventing VEGFR2-mediated JAK/STAT3 signaling in endothelial cells [142]. With regards to IRD E804, several molecular targets, including CDKs, Src/STAT3, and BCR-ABL/STAT5 had been described so far. In a different context, Chan *et al.* tested the anti-angiogenic profile of the compound, and showed a potent *in vitro* angio-suppression by reducing the proliferation, migration, and tube formation of human umbilical vein endothelial cells (HUVECs). Additionally, E804 (at a concentration of 4 μM) completely inhibited the outgrowth of endothelial cells from the rat aorta, as determined by the *ex vivo* aortic ring assay, as well as the activity of two members of the MMP family, MMP-2 and MMP-9 [143]. The anti-angiogenic activity of the cell permeable derivative, E804, was further confirmed in a different study where the compound inhibited the angiogenic markers in HUVECs and micro-vessel sprouting from the rat aortic ring. Mechanistically, E804 inhibited the kinase activity VEGFR-2 *in vitro* and in endothelial cells, blocking receptor phosphorylation, and subsequently the activity of its down-stream kinases, Akt and Erk MAPK [144].

Similar to other kinase alterations, fibroblast growth factor receptors (FGFRs)–which also belong to the family of RTKs–are targetable by different approaches, such as antibodies, DNA/RNA aptamers, as well as small molecules (***reviewed in*** [131]). In the context of indirubins as FGFR small molecule inhibitors, Zhen *et al.* have reported that indirubin-3'-oxime strongly and selectively inhibits the activity of this tyrosine kinase receptor at concentrations below those required for CDK2 inhibition (5 μM *vs.* 20 μM, respectively). Additionally, the compound was found to enhance Erk1/2 activity, independently of FGFR1 inhibition, but through p38 MAPK activation. Blocking FGFR1 activation, indirubin-3'-oxime caused reduced proliferation of the FGF1-stimulated NIH/3T3 cell line and the human acute myelogenous leukemia (AML) cells, KG-1a. At a higher concentration (20 μM), the small molecule was able to show similar effects on the FGFR-1-negative leukemia cell line, K-562, which may be through the inhibition of CDK [145]. These findings become particularly significant in the treatment of FGFR1-dependent cancers (e.g., the AML-derived cancer cell line, KG-1a), however, as aforementioned (in case of K-562 cells), the anti-cancer properties of indirubin-3'-oxime is not limited to FGFR-1-sensitive tumor cells due to the pleotropic nature of cellular responses to this derivative as well as other IRDs.

Overactivation of the receptor tyrosine kinase, insulin-like growth factor 1 receptor (IGF1R) and its ligands, IGF1/2 have been associated with various human diseases, including a number of different cancers (e.g., breast- colorectal- and prostate cancers). Like other RTKs, several strategies, including

antibodies, neutralizing ligands, as well as chemical TKIs have been tested for targeting IGF1R signaling, some of which are in late phases of clinical trials (***reviewed in*** [146]). Recently, screening a series of newly synthesized 5'-carboxy substituents of indirubins identified a water-soluble derivative, indirubin-3′-[2-(4-methylpiperazino)ethyl]oxime ether hydrochloride–referred to as 6ha–with a potent inhibitory activity against IGF1R, and only a marginal inhibition of CDKs and GSK3β. Accordingly, 6ha showed cytotoxic effects in the NCI-60 cell line panel at low micro molar concentrations [147].

1.2.2.4. molecular targets–interfering with the TGFβ/BMP pathway

As thoroughly discussed above, the TGFβ/BMP pathway is regulated at multiple levels, in particular, through controlling the duration of activated- and steady-state R-Smad signaling, a process that requires a continuous cooperation between phosphorylation, ubiquitination, and deubiquitination. We have tested the biological activity of indirubins in this context, and showed that among a series of derivatives, IRD E738 has the strongest effect in depleting steady-state R-Smad levels of both major TGFβ-family pathways–Smad2/3 and Smad1/5/8–Detailed analyses of the molecule's effects revealed that this intervention with the signaling happens at clearly separated levels: (i) reduction of total R-Smad pools, (ii) transient stabilization of receptor-activated R-Smads (possibly due to the GSK3 inhibitory activity of E738), and (iii) activation of the noncanonical TGFβ signalings, p38- and JNK MAPKs [148]. The latter effect is consistent with the previously reported induction of MAPKs in response to other derivatives [145]. Importantly, the E738-mediated R-Smad depletion was found to occur through the control of protein stability and was associated with the deubiquitinating enzymes, USP9x and USP34, adding a new target to indirubins' profile; the ubiquitin-proteasome pathway. Inhibition of these enzymes, using either IRD E738 or siRNA caused reduced levels of steady-state Smad2/3 and Smad1/5/8, downstream of the TGFβ- and BMP pathways, respectively [148]. ***Altogether, these findings highlight the regulation of TGFβ/BMP signaling by reversible ubiquitination, which occurs not only through controlling activated R-Smad pools, but also through that of total non-activated R-Smads. Moreover, this regulatory mechanism could be modulated using small molecules, as shown by the action of IRD E738.***

1.2.2.5. molecular targets–other mechanisms

In addition to the aforementioned cellular networks, many more signaling molecules have been shown to be regulated by indirubin analogues, including the cytoplasmic kinases, IKK and casein kinases (CKs). With regards to the former, Sethi *et al.* reported that the indirubin-mediated inhibition of IKK leads to the inactivation of the NF-κB pathway, an effect which most likely has a crucial role in the small molecule's anti-inflammatory effects [149]. With regards to the latter, IRD 7,7'-diazaindirubin–from the family of azaindirubins with a more metabolically stable profile–has been found to selectively inhibit CKs. Interestingly, the anti-proliferative and pro-apoptotic effects of this derivative in MCF-7 breast cancer cells were abrogated upon knocking down *CK2*, implicating the role of this kinase in the molecule's mode of action [150].

In a new context, the bioactivity of indirubins was investigated against the tumor initiating cancer stem cells. So far, several combinations of cell surface markers are known to be selectively expressed in CSCs, such as $CD44^+/CD24^-$, $CD44^+/CD24^+/ESA^+$, and $CD34^+/CD133^+104$ (***recently reviewed in*** [151]). In this respect, meisoindigo, an analogue of isoindigo with a methyl group at 1-position, has been shown to preferentially decrease CSC populations ($CD133^+$ and $CD24^+/CD44^+/ESA^+$ cells) of PDAC [152]. Previous studies had reported on the ability of this derivative in inducing apoptosis, inhibiting cell proliferation and differentiation, as well as angiogenesis in both *in vitro* and *in vivo* leukemia models. Notably, meisoindigo, in contrast to its parental molecule–indirubin–does not interact with protein kinases, suggesting other mechanistic pathways than the conventional targets (e.g., CDKs and GSK3β) underpinning the anti-tumoral effects of isoindigo analogues (***reviewed in*** [113]). For example, we have proposed that meisoindigo triggers AMPK activation through the altered mitochondrial function, followed by oxidative stress-related apoptosis in PDAC cells. Additionally, the upstream kinase of AMPK, liver kinase B1 (LKB1) was inhibited upon meisoindigo treatment, which was accompanied by a decrease in cancer stem cell markers (e.g., CD133 and CD44) [152]. These results provide a deeper understanding into the molecular basis of the cytotoxic- and apoptogenic responses, as well as anti-cancer stem cell properties of the new generation of indirubins, isoindigo derivatives.

In summary, protein kinase inhibition is seemingly the main contributor to the in vitro and in vivo effects of indirubins in the context of cancer and other diseases. However, as described in detail, recent studies have proposed new targets, such as the ubiquitin-proteasome pathway. The bioactivity of indirubins extends even further through the regulation of microRNAs, as recently described for indirubin and indirubin-3'-oxime to regulate the tumor suppressor microRNA, let7 [153]. ***Another example is the decreased global microRNA maturation in response to IRD 6BIO, as well as another small molecule inhibitor of GSK3, CHIR*** [154]. ***These findings provide more interesting aspects to the multi-faceted biological profile of this family of natural compounds.***

1.2.3. Indirubins–clinical perspectives

As briefly mentioned above, the clinical use of indirubin was first tested in 1970-1980s in a number of clinical trials for the treatment of CML patients and showed more than 50% partial or complete remission (***summarized in*** [113]). In comparison to the standard regimen, busulfan, indirubin exhibited similar efficacy, however with a lower rate of complete hematologic response (CR) [155]. Regarding toxicity, indirubin was associated with minor side effects in most cases which were limited to abdominal pain, diarrhea, and nausea, except for three cases who experienced reversible pulmonary hypertension and cardiac insufficiency [156]. Lately, a complete cytogenetic response and survival of 32 years in a CML patient was reported who was treated with a combination of the TKI, imatinib, indirubin, and meisoindigo [157]. Comparing indirubin and meisoindigo, they showed similar efficacy, however with lower side effects in case of the latter. In another clinical study, similar rates of partial and complete remission were observed between meisoindigo- and busulfan-treated groups of CML patients [156]. A

phase II clinical trial involving 134 newly diagnosed CML patients treated with meisoindigo (75-150 mg/d) demonstrated a rate of 32.1% and 48.5% of CR and hematological partial response (PR), respectively. The treatment also showed comparable CR and PR rates between newly diagnosed- and previously treated-CML patients in a phase III clinical trial [156, 158]. Based on a retrospective study with 274 CML patients who were followed over a period of 5 years, no major difference was observed regarding hematological response, median duration of chronic phase, median duration of survival, and blast crisis rate at 60 months from diagnosis between meisoindigo, hydroxyurea, and busulfan. However, combination therapy with meisoindigo and hydroxyurea was found to significantly improve all the aforementioned markers, compared with meisoindigo, hydroxyurea, and busulfan as single treatments [156, 158]. This suggests a possible synergy between meisoindigo and hydroxyurea in the treatment of CML. At present, indirubin is being evaluated in a multicenter randomized trial to improve the outcome of acute promyelocytic leukemia (APL) in children (***clinical trials.gov identifier: NCT02200978***).

1.3. Project narrative–rationale, hypothesis, and purpose

In light of the abovementioned data on the cellular mechanisms of indirubin analogues, which involve multiple signaling molecules with an essential role in the regulation of the TGFβ/BMP pathway (e.g., MAPKs, CDKs, and GSK3), we proposed if the new IRD–designated as E738–is able of interfering with TGFβ- and BMP signals (**see Figure 6**). Our previous knowledge could show the ability of indirubin derivatives to inhibit the signaling outcome though reducing the stability of basal R-Smads, which was found to occur by blocking deubiquitination (**Figure 6**). This has provided us with a less investigated theme in the regulatory mechanisms of the TGFβ/BMP signal transduction, i.e., **(i) the significance of steady-state R-Smad regulation–in addition to that of the receptor-activated R-Smads–in cellular responses to TGFβ- and BMP signals**, and **(ii) the importance of reversible ubiquitination in controlling total R-Smad pools, apart from its role in the regulation of activated R-Smads, which has been thoroughly investigated in literature.**

In the present research, we have focused on the effects of IRD E738, as it had shown the most efficient depletion of total R-Smads, among several tested derivatives. As previously discussed, a series of sequential post-translational modifications, including ubiquitination and phosphorylation (**see Figure 2 and Figure 6**), works cooperatively in order to regulate the half-life of R-Smad signaling. With a focus on TGFβ-related Smad2/3, we aimed at exploring **(i) how the abundance of total R-Smad pools is maintained through the dynamic interplay between phosphorylation and ubiquitination**, and **(ii) how accessible this regulatory mechanism is for small molecule intervention**. As discussed earlier, during cancer progression, the tumor-suppressive TGFβ/Smad signaling can shift to the oncogenic pathway; this malignant switch is associated with the aberrant phosphorylation pattern of R-Smads. Having this in mind, we sought of **(iii) translating the cellular effects of phospho-Smad signaling into a therapeutic strategy by using cholangiocarcinoma (CCA) cells as a cancerous model**. Finally, we intended to **(iv) investigate whether the kinase inhibitory activity of IRD E738 is**

able to inhibit the oncogenic phospho-Smad signaling in CCA, given that the phosphorylation status of R-Smads relies on the activity of several membrane bound- cytoplasmic and nuclear kinases.

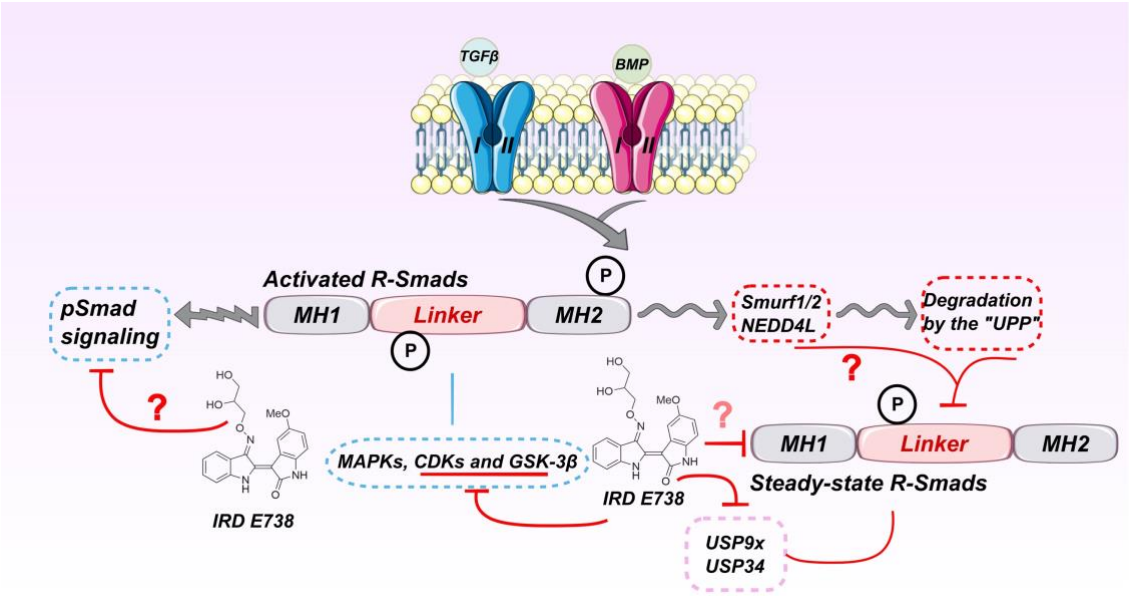


Figure 6. Rationale and working hypothesis of the present chapter. The schematic illustration summarizes our previous understanding of the regulation of *"steady-state"* (herein also referred to as basal- total- or non-activated) R-Smads by the ubiquitin-proteasome pathway (UPP) using small molecule intervention. The indirubin derivative (IRD), E738 depletes total R-Smads via inhibiting the deubiquitinating enzymes, USP9x and USP34, proposing a role for the "UPP" in maintaining steady-state R-Smad pools. Red question marks demonstrate the unaddressed issues of the small molecule's mode of action, and so the specific objectives of the current study. As shown by a blunt-headed arrow, IRD E738–similar to most of the indirubins–is a potent kinase inhibitor, targeting several kinases involved in R-Smad linker phosphorylation. Given that linker phosphorylation boosts R-Smad transcriptional activity (referred to as *"pSmad signaling"*), we asked if E738 (and/or other kinase inhibitors) is capable of interfering with pSmad signaling, with a particular implication in certain cancerous contexts. Modulation of Smad linker phosphorylation by the molecule, on the one hand, and its inhibitory effect on total Smad levels, on the other hand, pose another interesting question with regards to the control of TGFβ signaling; how relevant is linker phosphorylation to steady-state R-Smad stability? and if so, which ubiquitin ligases are involved? The following sections tackle these questions and discuss the findings and limitations of the study.

2. RESULTS

2.1. The indirubin derivative–E738–interferes with the TGFβ/BMP/Smad pathway at multiple separated levels.

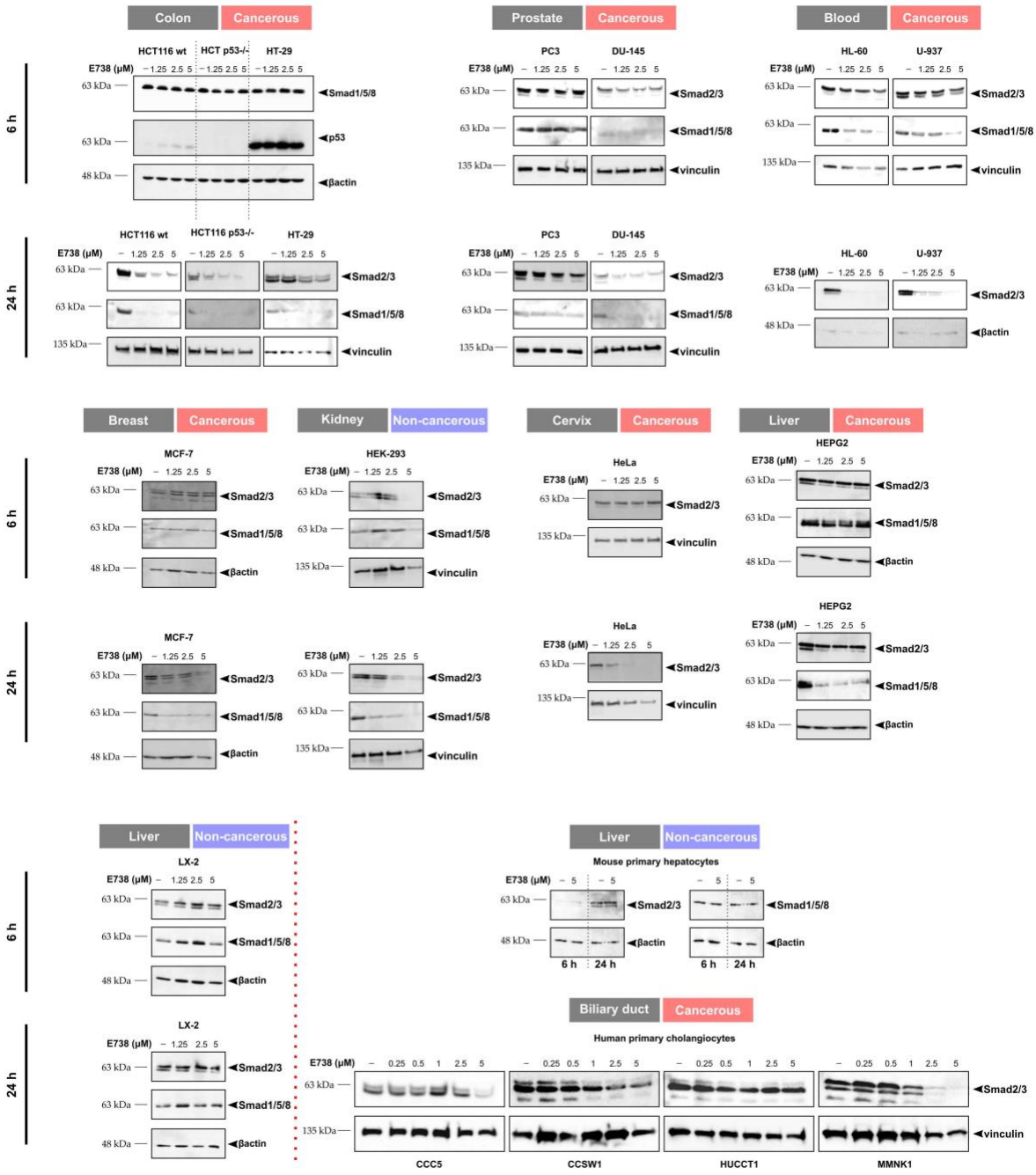


Figure 7. The indirubin derivative–E738–depletes steady-state R-Smad levels in cell lines of various tissue origin. Immunoblots showing the regulation of non-activated R-Smad levels in cells of the indicated cancerous- and non-cancerous tissue origin upon treatment with increasing concentrations of E738 for either 6 h or 24 h. All tested cell lines demonstrate a clear decrease in TGFβ-related Smad2/3 and BMP-related Smad1/5/8 in response to the small molecule, however, with different sensitivities. βactin and vinculin served as loading controls.

2.1.1. IRD E738 reduces the protein levels of steady-state R-Smads in various cell lines of both cancerous and non-cancerous tissue origin.

We had previously shown that the protein expression of Smad1/5/8 is reduced upon treatment with several indirubin derivatives in HeLa cells as well as in **human** foreskin fibroblasts (HFFs) [148]. Based on the structure-activity relationship (SAR) studies, a propanediol group on 3'-position and/or a methoxy group at 5-position causes the strongest depletion of total- and C-terminally phosphorylated Smad1/5/8 [148]. With this in mind, we chose IRD E738–carrying both substituents, a propanediol- and a methoxy group on 3'- and 5-positions, respectively–to study the regulation of R-Smad signaling. To find the most suitable *in vitro* model, we sought to perform a large screen encompassing cell lines of various cancerous- and non-cancerous origin. For this, immunoblotting was used and the protein levels of both TGFβ- and BMP-related R-Smads were determined upon treatment with increasing doses of E738 at short- and long-term time points. The results clearly showed a reproducible reduction in R-Smad protein levels, however, with different sensitivities among the tested cell lines (**Figure 7**).

In addition to E738, the IRD E860–carrying a methoxy- and a large basic group on 5- and 3'-positions, respectively (**Figure 8A**)–was tested in several cell lines, including HEK-293, T cells, as well as colorectal- and leukemia cancer cells, and HeLa. In comparison to E738, this analogue showed less efficiency in terms of R-Smad depletion (**Figure 8B**); this is consistent with the previous SAR studies, showing the positive effects of both propanediol- and methoxy substituents at 5- and 3'-positions on the indirubin-mediated inhibition of TGFβ/BMP signaling [148].

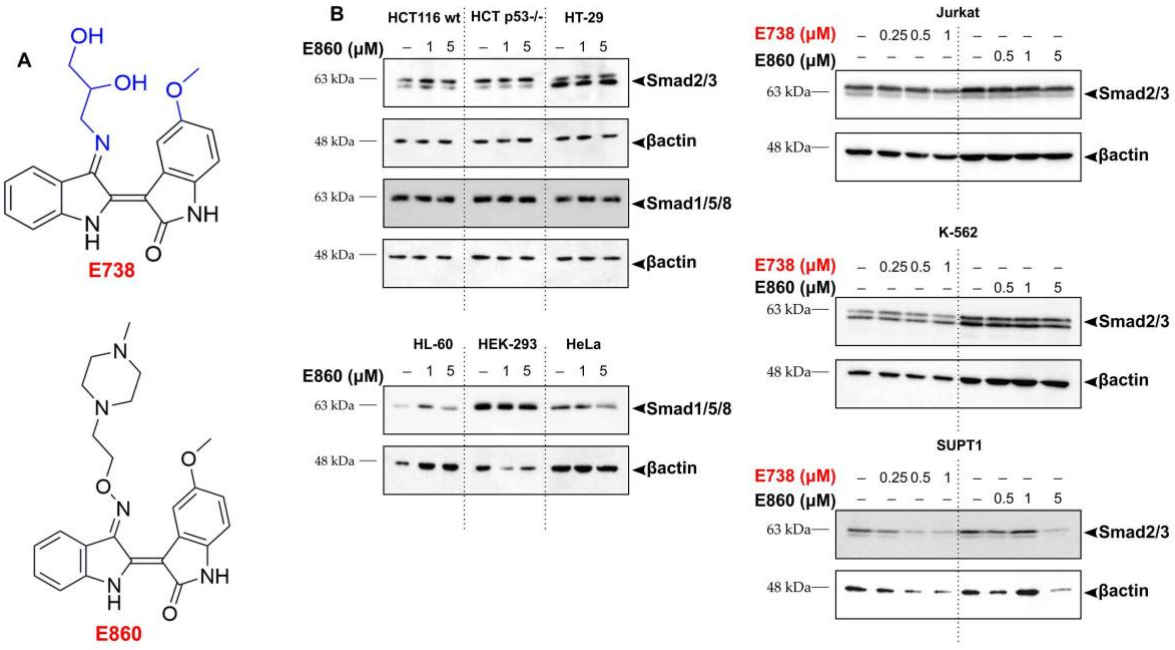


Figure 8. Comparison of IRD E860 with E738 in reducing steady-state R-Smad levels. (A) Chemical structures of the indirubin derivatives, E738 and E860. **(B)** Immunoblots compare the efficacy of E860 with E738 in depleting TGFβ-related Smad2/3 and BMP-associated Smad1/5/8 upon 24 h treatment with the small molecules at the indicated concentrations. In general, E860 does not show a clear decrease of R-Smad protein expression except for the leukemia cell line, HL-60 and T lymphoblast cells, SUPT1. βactin served as the loading control.

2.1.1.1. Biotinylated E738 displays similar regulation of steady-state R-Smad levels compared with its parental molecule.

To identify E738's interacting proteins, with a special focus on components of the ubiquitin-proteasome pathway, we sought to synthesize biotin-conjugated E738 (designated as E738 B). The schematic synthesis procedure is illustrated in **Figure 9A**. After successful characterization of E738 S–containing the spacer region (colored in blue)–we tested the new molecule's effects on total Smad2/3 levels, using E738 as a reference. Although in some cases higher concentrations were required as compared to that of the parental molecule, E738 S was able to significantly reduce Smad2/3 levels in the tested cell lines, HEK-293 and HCT116 (**Figure 9B and 9C**). We then completed the synthesis of E738 B (colored in red) by attaching the NHS-activated biotin to the spacer and assessed E738 B's regulation of Smad2/3 levels in the HCT116 colorectal cancer cell line. As shown in **Figure 9D and 9E**, the newly synthesized small molecule led to a profound depletion of Smad2/3 levels in HCT116 cells with similar or even higher efficacy compared with that of E738.

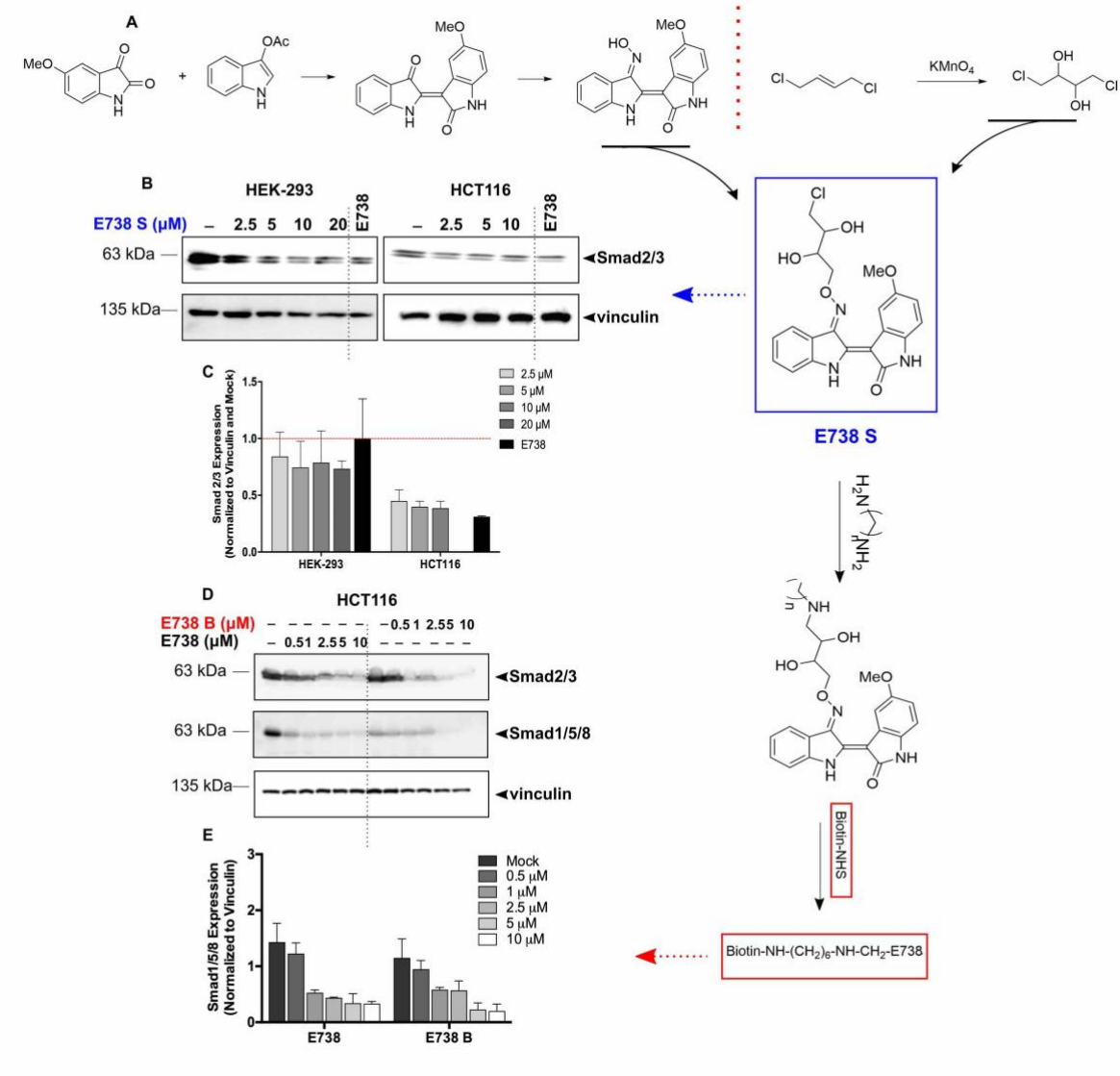


Figure 9. Biotinylated E738's regulation of R-Smad protein levels is comparable to that of the parental molecule. (A) schematic representation of the synthesis procedure of biotin-conjugated E738 (E738 B). **(B)** Immunoblots and the corresponding densitometric quantification **(C)**, showing a concentration-dependent decrease of Smad2/3 in HCT116 and HEK-293 cells after 24 h treatment with E738 carrying the spacer region (E738 S), as indicated. 24 h treatment with E738 (1 μM) was used as reference. **(D)** 24 h treatment with increasing concentrations of E738 B depletes non-activated Smad2/3- and Smad1/5/8 protein levels in a similar manner to that of the parental small molecule, E738. **(E)** Densitometric quantification of the Smad1/5/8 bands normalized to the respective loading control. Error bars indicate the SEM of at least two biological replicates, one of those is presented in **(D)** and **(B)**. Vinculin served as the loading control. 0.1% DMSO was used as mock treatment.

2.1.2. E738's effect on steady-state R-Smad levels is independent of that on receptor-activated R-Smads.

As described above, most of indirubin derivatives are potent protein kinase inhibitors, among others, are kinases involved in the duration of R-Smad activation, for example GSK3 and CDKs [113, 148]. We could previously show that IRD E738 delays the degradation of activated pSmad1/5/8 in HeLa cells, an effect which was attributed to its kinase inhibitory function [148]. In the present research, the focus was drawn to the TGFβ-associated Smad2/3. HEK-293 cells–which were found to be one of the most sensitive cells based on the initial screen (**see** Figure 7)–were stimulated with a pulse of TGFβ1 (10 ng/mL) for 1 h, followed by treatment with E738 (2.5 μM) or not (Figure 10A). Similar to the effects of the small molecule on C-terminally phosphorylated Smad1/5/8, activated pSmad2 and pSmad3 levels were found to be stabilized in the presence of E738, as determined by immunoblotting and the respective densitometric quantification (Figure 10B **and** 10C).

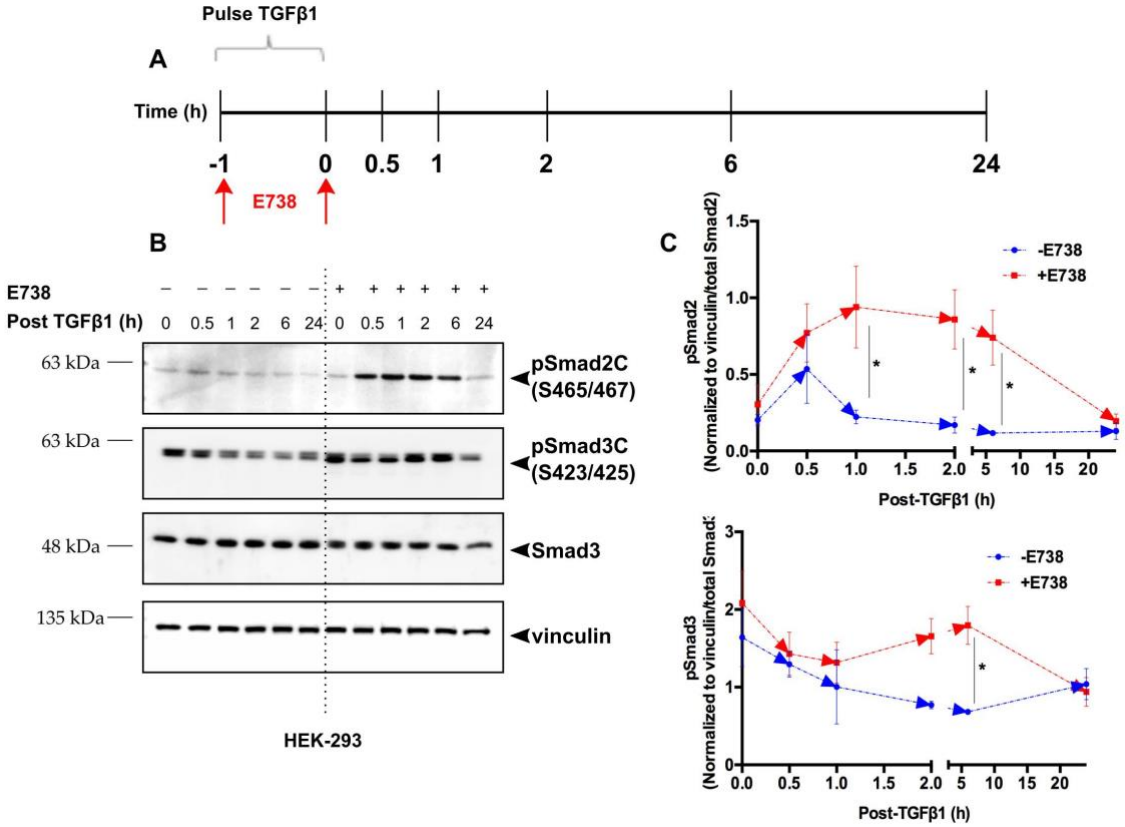


Figure 10. IRD E738 stabilizes the transcriptionally active Smad2/3. (A) schematic presentation of the treatment. HEK-293 cells were stimulated with TGFβ1 (10 ng/mL) 1 h prior to the addition of E738 (2.5 μM), thereafter were lysed at the indicated time points for immunoblotting. **(B)** The protein levels of C-terminally-phosphorylated Smad2/3 (pSmad2C and pSmad3C) are higher in the presence of E738 at most tested time points. **(C)** Densitometric quantification of pSmad2/3 bands normalized to total Smad2/3 levels and the respective loading control (vinculin). Error bars ± SEM; *n*=2. Multiple comparisons were made by a two-way ANOVA test, followed by a post-hoc Tukey test. * presents a *p*-value less than or equal to 0.05.

2.1.3. E738-mediated Smad2/3 reduction is linked to Smad stability.

Given that the proteasome-mediated degradation is one of the major routes through which R-Smad signaling is regulated [52], we sought to see if the E738's effect on total R-Smad levels is associated with the stability of the protein. We had previously shown that ubiquitin-proteasome-mediated degradation is required for E738's depletion of both activated and non-activated R-Smads of both TGFβ- and BMP pathways in HeLa and HFF cell lines [148]. Focusing on TGFβ-related Smad2/3 levels in HEK-293 cells, we found that co-treatment with the proteasome inhibitor, MG132, prevents E738-induced Smad2/3 depletion and increases the levels of C-terminally phosphorylated Smad2 (**Figure 11A-E**). In addition, the use of the protein synthesis inhibitor, cycloheximide (CHX) together with E738, further showed that the small molecule destabilizes total Smad2/3 protein levels, while it increases the stability of receptor-activated Smad2 (**Figure 11**).

2.1.4. IRD E738 regulates TGFβ's target gene transcription.

Our initial screen (**see Figure 7**) showed different degrees of E738-mediated R-Smad depletion among different cell lines. We therefore postulated that this may result from differential regulation of genes, among others, those encoding the key players of the ubiquitin-proteasome pathway. In light of this, comprehensive mRNA analyses using qRT-PCR revealed that most of the E3 ligases responsible for polyubiquitination of activated R-Smads, such as *Smurf1/2* and *NEDD4L* are reduced in response to E738 (**Figure 12A**). This suggests that E738's modulation of E3 ligases is possibly not involved in the observed degradation of total R-Smads, which is in line with our previous report [148]. On the other hand, mRNA levels of a number of ubiquitin proteases, namely *USP4*, *USP9x*, and *USP34* were found to be downregulated with treatment (**Figure 12A**), implicating a possible role for these enzymes in the small molecule's effects on R-Smad stability, as previously shown in case of USP9x and USP34 [148].

With regards to TGFβ/BMP target genes, we observed a clear reduction in the mRNA levels of *ID1/2/3* transcription factors across all tested cell lines (**Figure 12A**), consistent with the inhibition of the signaling. Inhibitors of cell differentiation, *ID*s, are well known TGFβ- and BMP target genes during both normal development and cancer progression [159, 160]. For example, *ID1* and *ID2* expression is positively regulated by BMP in myoblasts, osteoblast-like cells, multiple myeloma, breast- and lung cancer cells, whereas *ID3* gene expression is enhanced by TGFβ1 in progenitor- and mature B-lymphocytes, as well as in glioblastoma [159]. The extracellular matrix (ECM) protein, *CTGF*, was consistently upregulated in response to E738, in contrast to our previous report in HeLa cells (**Figure 12A**) [148]. CTGF is known to interact with a variety of molecules, including the growth factors; TGFβ,

BMP4, and BMP7. Its interaction with TGFβ positively regulates the signaling pathway and its target genes, which includes CTGF itself, in order to stimulate ECM synthesis, mediating pro-fibrotic actions of this cytokine. On the other hand, the interaction of CTGF and BMP is rather antagonistic, inhibiting downstream mediators, such as Smad1. Interestingly, the TGFβ- and BMP target gene and the transcriptional corepressor, *SKIL* (coding for SnoN), was found to be regulated in a p53-dependent manner, i.e., in cells harboring wild-type (wt) p53 (HEK-293, HCT116 wt, and HEPG2) a decrease in its mRNA expression was observed, whereas in cells lacking wt p53 expression (HL-60 and HCT116 p53-/-) the mRNA levels were induced (**Figure 12A**). SnoN and p53 have been shown to be related through several mechanisms, among others, is the interaction of SnoN with PML, which stabilizes p53, turning *SnoN* from being an oncogene into a tumor suppressor [161]. In addition to SnoN, we observed that several other genes, namely *Smad7*, *DUSP14*, *c-Myc*, and *CDKN1A* (coding for p21) are regulated in an opposite manner between wt- and p53-deficient cells (**Figure 12A**). To this end, we broadened our investigation by adding more p53 target genes (*MDM2* and *BBC3*, encoding Puma) as well as by including an additional p53-null cell line, HEP3B, and observed a similar pattern (**Figure 12A and 12B**). This suggests a role for the p53 pathway in E738's mode of action, in particular the small molecule-mediated cytotoxic and pro-apoptotic effects. The mRNA levels of *Runx2*, a cofactor of Smad transcriptional complexes and a well-known target of TGFβ1 and BMP2 [162] were reduced upon E738 treatment across all tested cell lines (**Figure 12A**). Finally, the mRNA expression of TGFβ- and BMP receptors, *ALK4* and *ALK3* [2], respectively as well as that of *R-Smads*, downstream of both major pathways was found to be downregulated (**Figure 12A**) similar to their protein expression (**see Figure 7**). With regards to other families of Smads, the common Smad, *Smad4* showed a slight downregulation (**Figure 12A**), whereas *Smad7* was found to be modulated p53-dependently, as mentioned earlier (**Figure 12A**). In view of this, it has been reported that p53 activation might mediate apoptosis in a Smad7-dependent manner [163].

2.2. "Linker" phosphorylation is involved in E738's effects on Smad 2/3 stability.

2.2.1. The linker domain of Smad2/3 plays an important role in the E738-mediated reduction of Smad protein expression.

As described above and schematically represented in **Figure 13**, R-Smad proteins share three structural regions: (i) DNA binding domain at their N-termini (MH1), (ii) the transcriptional activation domain at their C-termini (MH2), and (iii) the inter-domain linker region. We aimed to identify which of the three domains is most crucial to the regulation of R-Smad stability in response to IRD E738 with a focus on TGFβ-associated Smad2/3. For this, different constructs, including the wt protein, truncated forms, as well as Smad2/3 EPSM, possessing mutations in the Erk/proline-directed kinase sites were transiently over-expressed in HEK-923 cells, followed by treatment with E738 for 6 h. Taking a close look at the regulation of flag-tagged Smad expression in different mutants, we observed similar or even higher reduction of flag protein levels upon treatment in case of the C-terminally truncated Smad (**Figure 13B, 13C, 13F, and 13G**).

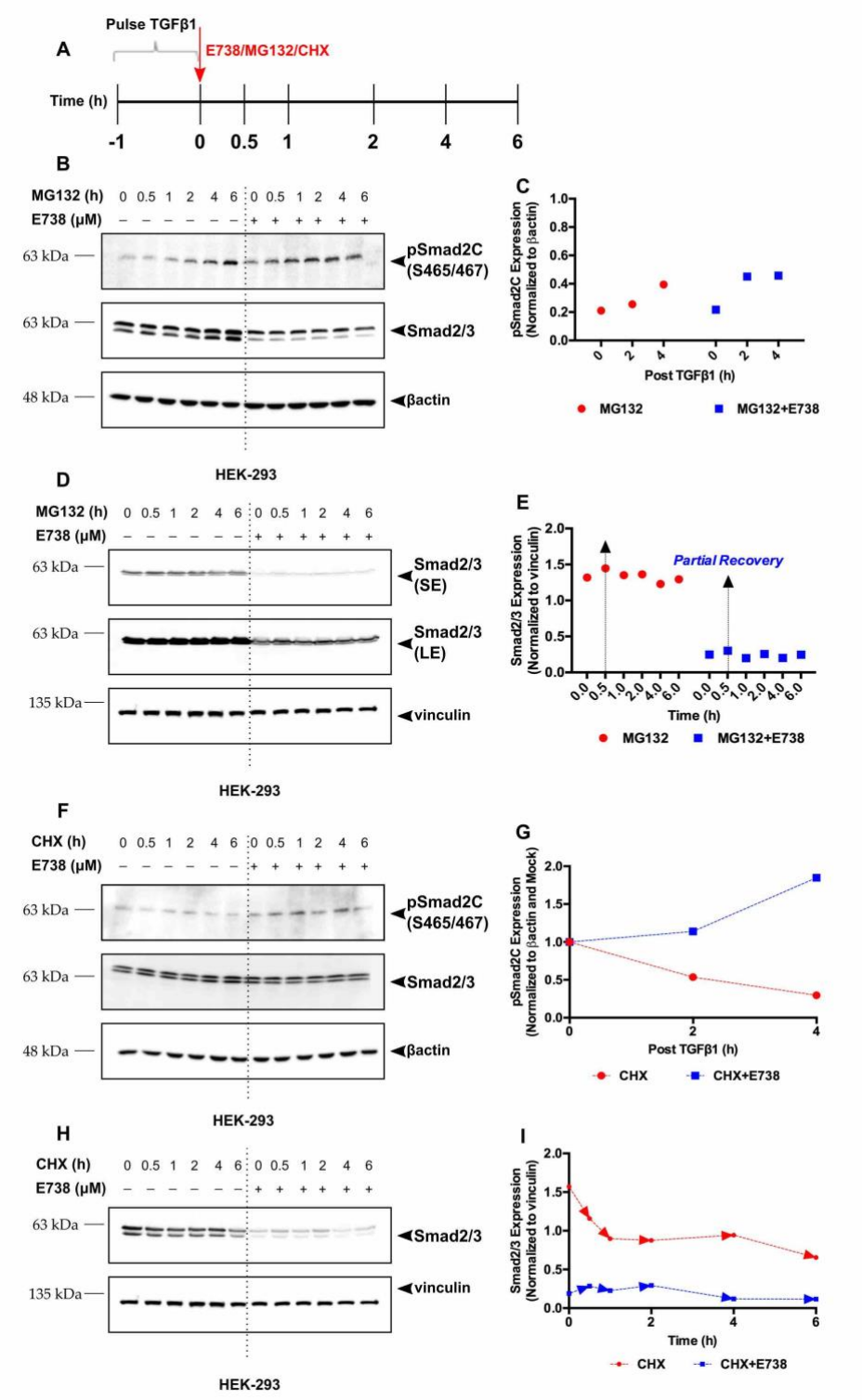


Figure 11. IRD E738-induced Smad2/3 depletion is most likely mediated via a reduction in R-Smad stability. (A) Treatment scheme performed in **(B)**, **(C)**, **(F)**, and **(G)**. HEK-293 cells were treated with a pulse of TGFβ1 (5 ng/mL) for 1 h after which the drugs, MG132 (10 μM), CHX (100 μg/mL), and E738 (2.5 μM) were added for the indicated time points. **(B)** and **(C)** Immunoblots and the respective densitometric quantification show that receptor-activated Smad2 (pSmad2C) is stabilized with E738 treatment, an effect which is enhanced in the presence of the proteasome inhibitor, MG132. On the other hand, steady-state Smad2/3 protein levels (without TGFβ stimulation) are clearly reduced in response to E738 and appear to be partially rescued upon co-treatment with MG132 for 0.5 h, as shown by

immunoblotting **(D)** and the corresponding densitometric analysis **(E)**. SE; short exposure and LE; long exposure. **(F)** and **(G)** Concomitant treatment with the protein synthesis inhibitor, CHX demonstrates a higher protein stability of pSmad2C in the presence of E738, further showing the stabilization of the receptor-activated Smad2 by the small molecule. This is independent of E738's regulation of non-activated Smad2/3 levels, which are reduced upon adding the molecule, and are further destabilized when cells are co-treated with CHX, as depicted by immunoblotting **(H)** and the respective quantification **(I)**. βactin and vinculin served as loading controls.

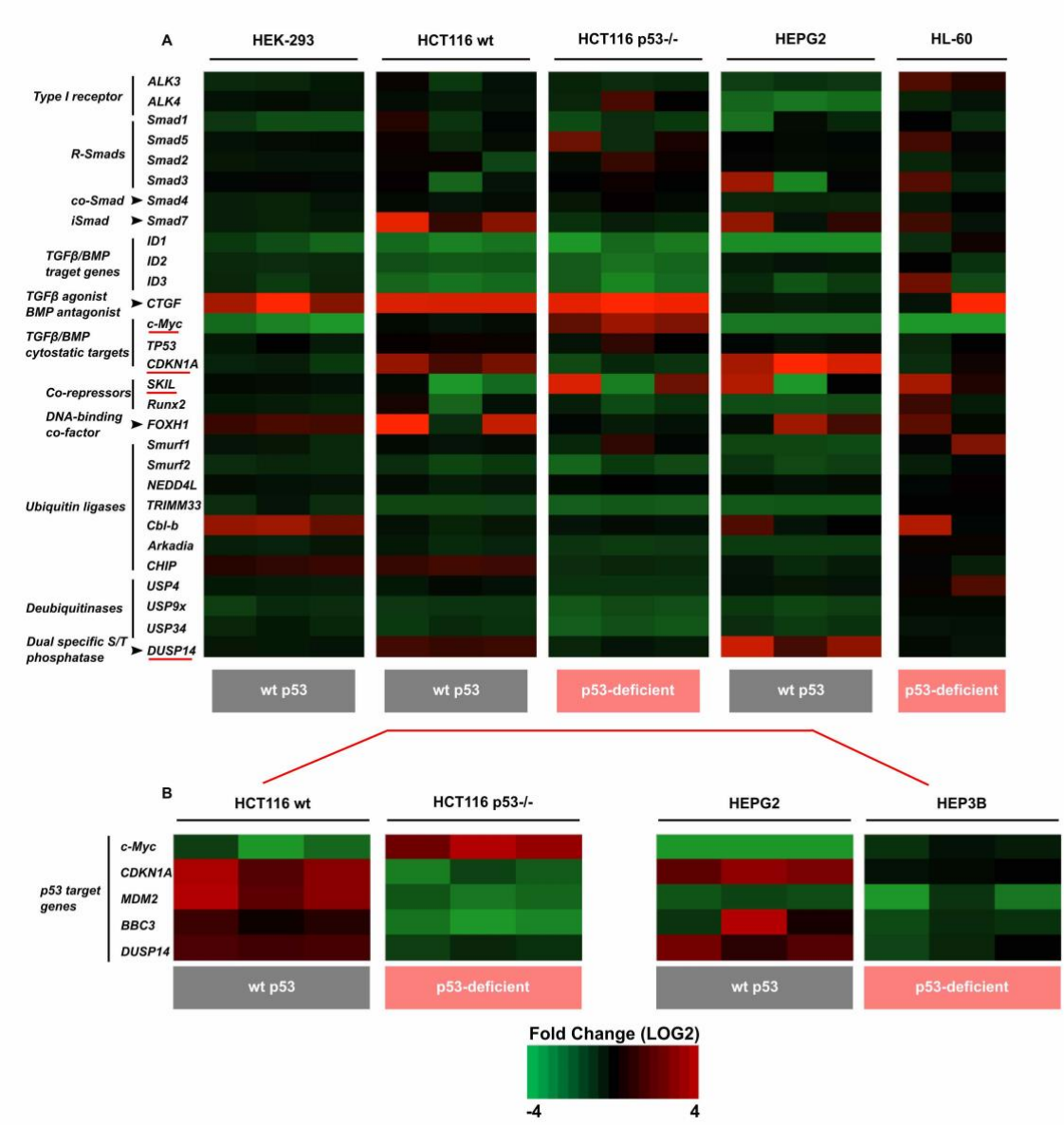


Figure 12. Regulation of TGFβ-related gene expression by IRD E738. **(A)** mRNA expression analyses of several signaling molecules associated with the TGFβ/BMP pathway, including target genes (e.g., *ID*s) as well as signaling molecules responsible for the regulation of the pathway (ubiquitin ligases and deubiquitinases), assessed by qRT-PCR analysis of the indicated cell lines treated with 1 μM of E738 for 24 h. **(B)** Multiple p53 target genes are differentially regulated between wild-type (wt) harboring cells and their p53-deficient counterpart upon 24 h treatment with E738 (1 μM), as determined by qRT-PCR. p53-related genes have been underlined in **(A)**. Relative expression was calculated by the ΔΔCt method, with vinculin and/or βactin as the housekeeping gene.

As aforementioned, C-tail phosphorylation of Smad2/3 at the SXS motif by the active receptor complex is a prerequisite step in R-Smad/Smad4 complex assembly and nuclear localization, which later on, in cooperation with other phosphorylated residues, targets activated phospho-Smads for proteasomal degradation [35]. With this in mind, a clear reduction in the ΔC truncates with treatment shows another proof of the E738's depletion of total Smad2/3 being independent of receptor-mediated activation (**also see Figure 10 and Figure 11**). Regarding the role of the MH1 (N-terminal) domain in the E738-mediated Smad depletion, we found less reduction in flag protein expression comparing Smad2 N and Smad3 N with wt Smad2/3 (**Figure 13B, 13C, 13F, 13G**). This suggests that the DNA binding domain, although being crucial to the transcription of target genes, is not sufficient for E738's regulation of Smad stability. Similar to Smad2 N and Smad3 N, forced expression of the MH2 (C-terminal) domain of Smad3 did not appear to cause a significant alteration in flag expression, while cells over-expressing the linker- and C-terminal regions (Smad3 LC) showed similar levels of flag reduction, compared with that of full-length Smad3 (**Figure 13F and 13G**). Considering the aforementioned data, it is likely that the presence of the linker region is required for the E738-mediated R-Smad degradation. The linker domain is of particular interest due to the presence of multiple phosphorylation sites that are essential for the regulation of R-Smad activity/stability, as well as its nuclear-cytoplasmic shuttling, as described earlier. In this regard, we observed that ectopic expression of Smad2/3 EPSM caused a greater E738-mediated reduction of flag expression, as compared to that of the wt protein, as well as several other truncates (**Figure 13B, 13C, 13F, and 13G**). Smad2- and Smad3 EPSM carry mutations in several S/T residues that are phosphorylated by various protein kinases (**see Figure 13A and 13D**), most of which are inhibited by IRD E738, based on the *in vitro* kinase profiling (**refer to Figure 19**). The role of these kinases in the stability of ligand-activated Smads has been thoroughly demonstrated [35], whereas their significance in the regulation of steady-state R-Smad levels is yet to be fully understood. Our results from transient over-expression of mutant constructs suggest a possible role for MAPKs, CDKs, and GSK3 in controlling the half-life of non-activated Smad2/3.

2.2.2. MEK1/Erk-mediated phosphorylation influences the stability of steady-state Smad2 and Smad3.

We had previously shown that two branches of the noncanonical (non-Smad) TGFβ signaling, p38 and JNK MAPKs are enhanced with E738 treatment [148]. In contrast to pp38 (T180/Y182) and pJNK, protein levels of active Erk1/2–pErk (T202/Y204)–were found to be significantly reduced in response to the indirubin derivative (**Figure 14A-C**) in HEK-293 as well as in CCA cell lines, which were used as a cancerous context. Having the E738-mediated inhibition of Erk1/2 in mind, we sought to further confirm the involvement of this kinase using known chemical inhibitors. In this regard, a more profound depletion of total Smad2 and Smad3 was observed when combining E738 with the MEK/Erk inhibitors, U0126 and PD0325901 (**Figure 14A, 14B, 14D, and 14E**).

Next, we tested if the proteasome inhibitor, MG132 is able of protecting Smad2/3 EPSM from E738-mediated degradation. As shown in **Figure 14F and 14G**, in the presence of MG132, both Smad2- and Smad3 protein levels were partially recovered, suggesting that E738-mediated reduction of the latter

constructs is most likely due to the proteasomal degradation. As aforementioned, Erk phosphorylation has been reported to stabilize Smad2 protein levels [21]. Using the translation inhibitor, CHX, we observed that the phosphorylation-defective mutant of Smad3 is less stable as compared to the wt protein, regardless of the treatment (**Figure 14H**). As expected, the addition of E738 further reduced the half-life of both wt Smad2/3 and Erk phosphorylation site mutants, compared with that of the non-treated cells (**Figure 14H-K**).

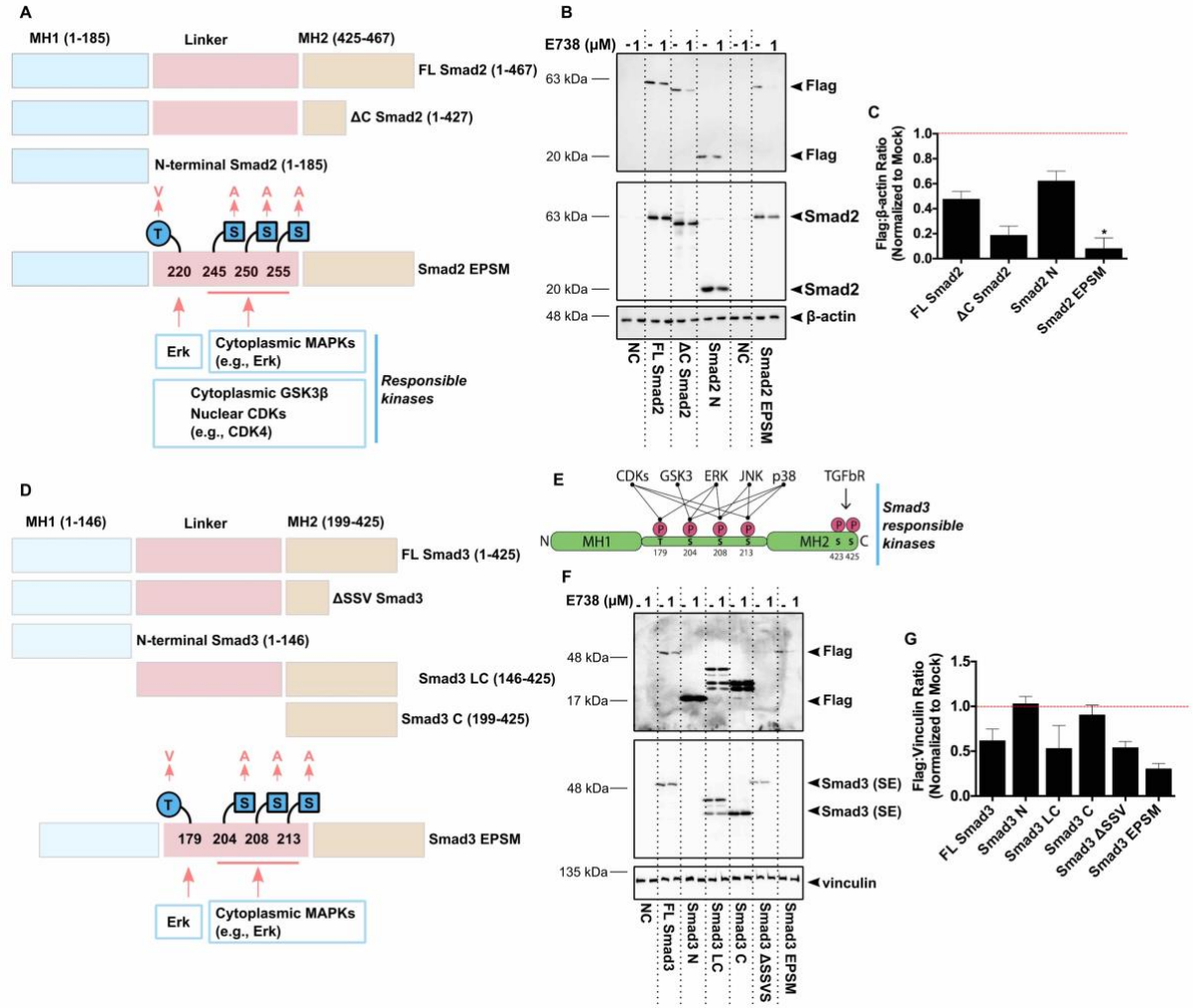


Figure 13. IRD E738-mediated Smad2/3 reduction relies on the linker (L) domain. (A) A schematic, illustrating the structural domains of full-length (FL) Smad2, as well as the three used mutant forms. The linker phosphorylation sites together with the involved kinases are depicted in Smad2 EPSM, which carries serine-to-alanine or threonine-to-valine (S-to-A/T-to-V) mutations in the indicated residues. All constructs are tagged with a flag sequence in their N-termini. **(B)** and **(C)** HEK-293 cells were transiently transfected with the constructs presented in **(A)**, followed by E738 treatment (1 μM, 6 h) 48 h post-transfection. Smad2 levels were then monitored by the expression of flag. As shown by immunoblots **(B)** and the respective quantification **(C)**, ectopic Smad2 EPSM demonstrates a significantly higher reduction in response to the small molecule, as compared to that of the FL protein. **(D)** Schematic presentation of Smad3 constructs used for the transient over-expression. C-tail- and linker phosphorylation residues as well as the responsible protein kinases are depicted in Smad3 EPSM schematic structure **(D)** and in wild-type Smad3 protein depicted in **(E)**. **(F)** and **(G)** Representative immunoblots **(F)** and quantification **(G)** showing a greater reduction of Smad3 EPSM compared with the FL protein at 6 h post-treatment with E738. Additionally, ectopic expression of constructs lacking the linker domain (Smad3 N and Smad3 C) abolished E738's depletion of Smad3, highlighting the significance of this region in the small molecule-

mediated R-Smad reduction. Error bars present the SEM of at least two biological replicates, one of which is shown in **(B)** and **(F)**. Multiple comparisons were performed by a two-way ANOVA test, followed by a post-hoc Tukey test. * indicates a *p*-value less than or equal to 0.05. 0.1% DMSO served as mock treatment.

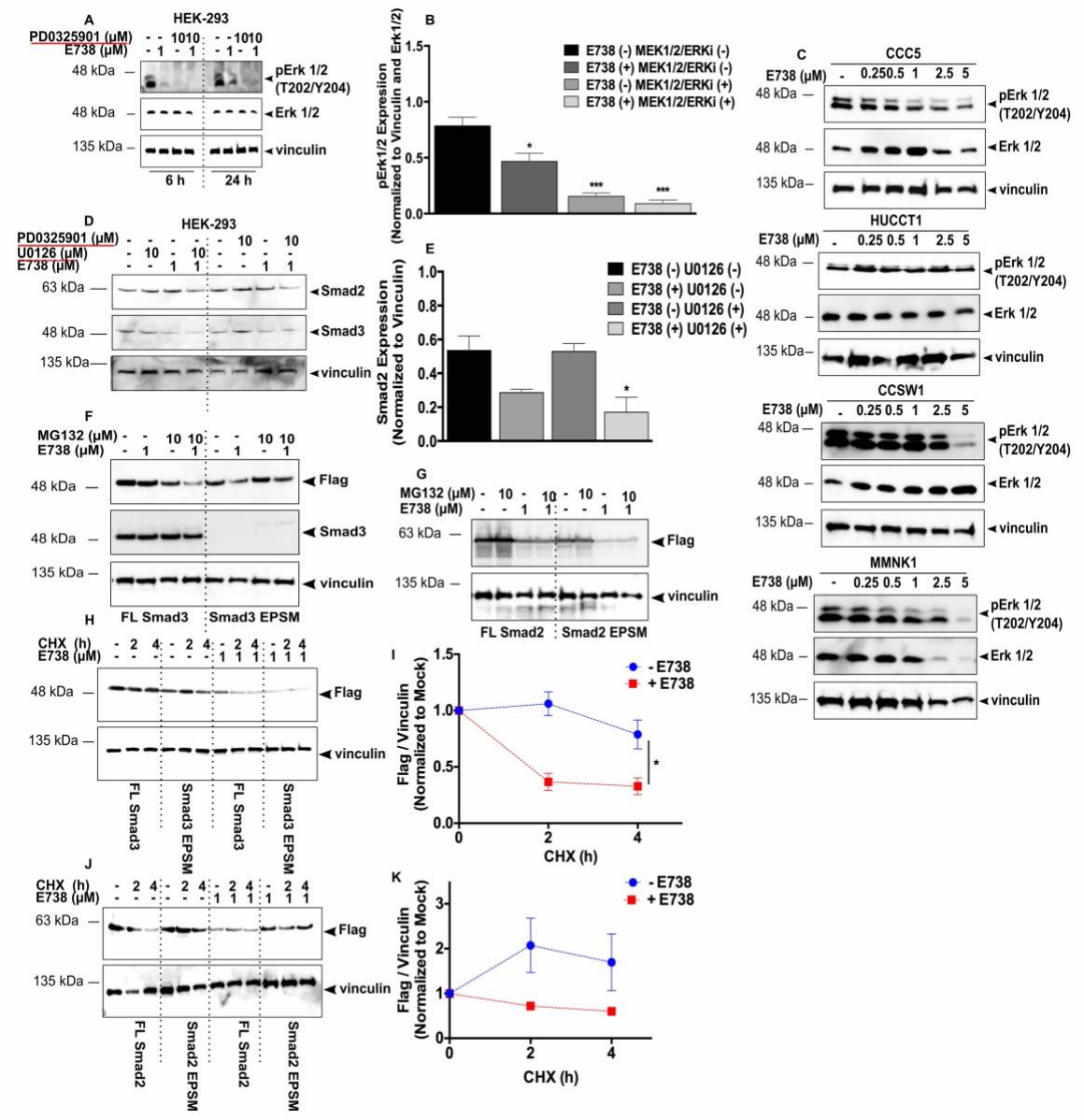


Figure 14. Erk1/2 most likely plays an important role in IRD E738's regulation of Smad2/3 stability through modulating the phosphorylation of multiple S/T residues in the linker domain. (A-C) immunoblots **(A)** and **(C)** and the densitometric analysis **(B)** show a profound inhibition of Erk activity by E738 (24 h treatment with the indicated concentrations) in HEK-293 **(A)** and **(B)** as well as in cholangiocarcinoma cells **(C)**. The small molecule inhibitor of Erk1/2/MEK, PD0325901 was used as reference. **(D)** and **(E)** Representative immunoblots **(D)** and the corresponding quantification **(E)**, showing a more profound reduction in steady-state Smad2- and Smad3 levels upon treatment with E738 (24 h) in combination with chemical inhibitors of Erk (PD0325901 and U0126). **(F)** and **(G)** The observed Smad depletion upon the ectopic expression of Smad2 EPSM- and Smad3 EPSM mutants is most likely due to Smad proteasomal degradation, judged by the partial rescue upon co-treatment with MG132. HEK-293 cells were transfected with either full-length (FL) Smad2/3 or the mutant construct (EPSM). At 48 h post-transfection, cells were treated with E738 for 24 h, MG132 was added 6 h before lysing the cells. **(H-K)** The use of CHX (100 μg/mL for the indicated time points) shows a reduction of Smad2/3

stability in HEK-293 cells over-expressing either FL- or mutant Smad2/3, an effect which is enhanced in the presence of E738, as determined by immunoblotting and the respective quantifications. Error bars ± SEM, *n* = 2. Multiple comparisons were made using a two-way ANOVA test, followed by a post-hoc Tukey test. *, **, and *** represent *p*-values less than or equal to 0.05, 0.01, and 0.001, respectively.

2.3. E738-mediated regulation of Smad2/3 stability is coupled to "phospho-Smad signaling".

As abovementioned, the role of TGFβ signaling in cancer involves the dysregulation of phospho-Smad2/3 signaling. The so-called "pSmad signaling" relies on linker phosphorylation (L) at multiple serine/threonine residues by CDKs, MAPKs, and GSK3, on the one hand, and the receptor-mediated C-tail phosphorylation (C), on the other hand (***reviewed in*** [164]). Based on the kinases involved, different phosphorylated forms of Smads are created (**see Figure 15A**), that will either enhance- or suppress TGFβ responses associated with Smad stability and cancer progression. Given the mechanism of E738-mediated regulation of Smad turnover, which was shown to involve modulation of specific phosphorylation sites in the Smad2/3 linker domain, we were inclined to explore whether there is a link between the small molecule's effects and the Smad phospho-isoform pathways.

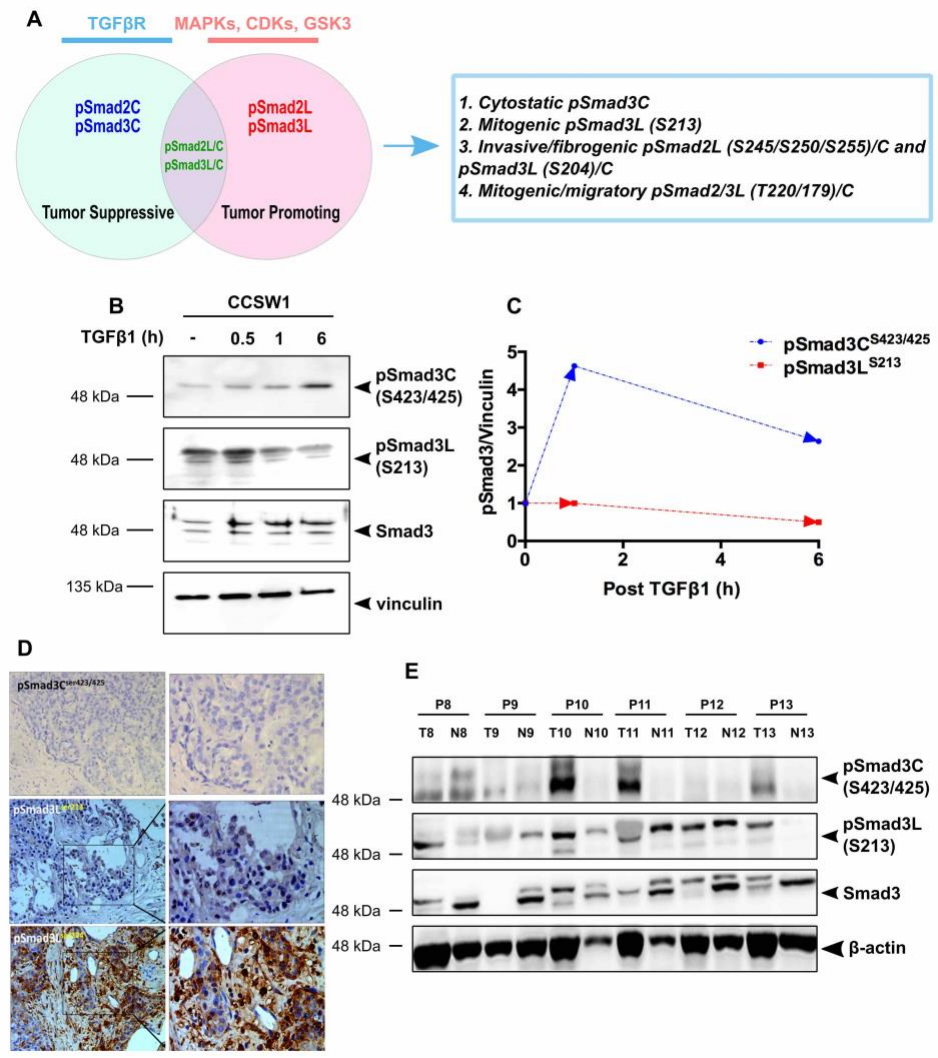


Figure 15. The regulation of phospho-Smad (pSmad) signaling in cholangiocarcinoma (CCA). (A) The Smad2/3 pathway shifts between tumor suppression and tumor progression based on the different

phosphorylation patterns, which are achieved by the action of various membrane bound- cytoplasmic- and nuclear kinases, as indicated. Such Smad phospho-isoforms induce distinct- and overlapping cellular phenotypes, as summarized in the blue box (***thoroughly reviewed in*** [164]). **(B)** and **(C)** In the intrahepatic CCA cell line, CCSW1, C-terminally- and linker (L)-phosphorylated Smad3 are regulated in an opposite manner to one another, as detected by immunoblotting **(B)** and the respective densitometric quantification **(C)**. **(D)** and **(E)** Immunohistochemistry **(D)** and immunoblotting **(E)** of a subgroup of CCA patient samples reveal a similar pattern of pSmad signaling compared with that of CCSW1 cell line in which the cytostatic pSmad3C- and the oncogenic pSmad3L levels are reduced and induced, respectively***. Patient samples were analyzed in collaboration with the group of Prof. Steven Dooley (Heidelberg University, Germany).***

2.3.1. Phospho-Smad3 signaling is impaired in cholangiocarcinoma.

Abnormal pSmad3 signaling has been shown to contribute to the pathogenesis of several human malignancies, such as colorectal- and liver cancers [164], however, its significance in the treatment of cholangiocarcinoma remains to be thoroughly investigated. We therefore sought to assess the relevance of different Smad3 phospho-isoforms to CCA, using healthy and tumorigenic tissues from patients, as well as a panel of patient-derived cell lines from both intra-(iCCA) and extra-hepatic (eCCA) origins. Importantly, we observed a cell-type specific regulation of pSmad3 signaling in the iCCA cell lines, CCSW1, showing a counter-regulation between the tumor-suppressive pSmad3C (S423/425)- and the oncogenic pSmad3L (S213) pathways (**Figure 15B and 15C**). Consistent with this, a subgroup of CCA patients displayed a similar pattern, as determined by immunohistochemistry and immunoblotting (**Figure 15D and 15E)**.

2.3.2. Erk/p38 MAPKs and GSK3 are involved in the regulation of phospho-Smad signaling in cholangiocarcinoma cells.

As described above, various S/T kinases stemming from other pathways (e.g., MAPKs and Wnt signaling) phosphorylate Smad2/3 linker residues (**see Figure 16A**). In light of this, we sought to perform a kinase inhibitor screen in the iCCA cell line, CCSW1. As depicted in **Figure 16B-D**, different small molecule inhibitors, including SB-431542 (targeting TGFβRI/ALK4/5/7), SP600125 (targeting JNK), U0126 (targeting MEK1/2/Erk), SB203580 (targeting p38), and CHIR (targeting GSK3) were incubated with the cells 1h prior to the treatment with TGFβ1 for 6h or not (**Figure 16B**). Looking into the phosphorylation status of the Smad3 linker domain, we found reduced levels of pSmad3 (S213) upon inhibition of p38/Erk MAPKs and GSK3, and a clear decrease of pSmad3 (S204) protein levels when targeting GSK3 in CCSW1 cells stimulated with TGFβ (**Figure 16C and 16D**). C-tail phosphorylated Smad3 (pSmad3C), on other hand, was found to be abolished upon treatment with the TGFβ receptor inhibitor (TGFβRi), as expected, and was induced with the rest of the inhibitors (**Figure 16C and 16D**). This could be explained by linker phosphorylation being a prerequisite step in the recognition of R-Smads by the ubiquitin ligases, followed by proteasome-mediated degradation [35]. Altogether, these results suggest the requirement of MEK1/2/Erk- and p38 MAPKs, as well as GSK3 in the Smad3 linker phosphorylation in CCA cells.

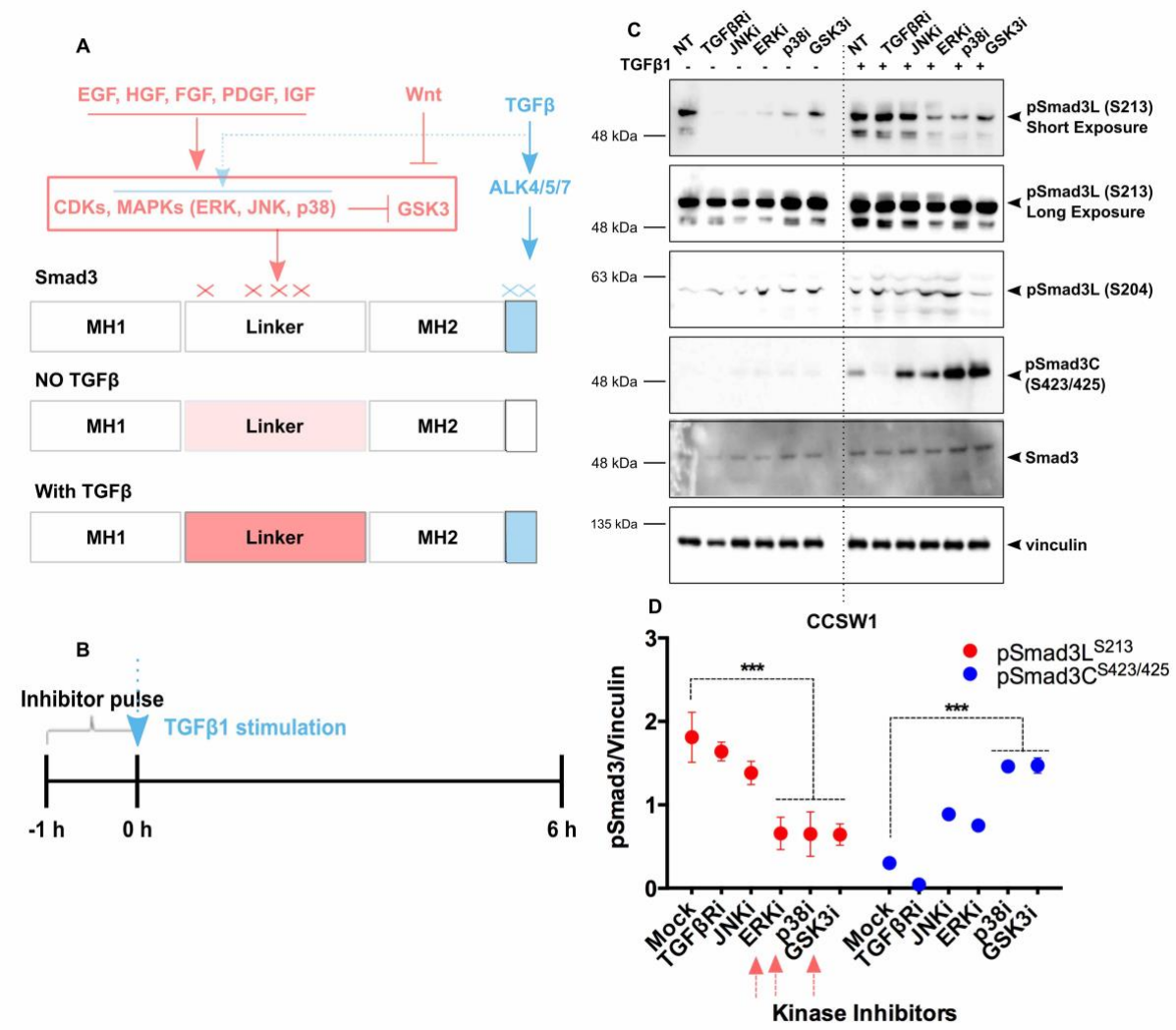


Figure 16. The oncogenic pSmad3L signaling in cholangiocarcinoma (CCA) cells is mediated via the action of various S/T kinases. (A) A scheme illustrating an intricate crosstalk between the canonical TGFβ signaling, which is mediated through the ALK4/5/7-mediated phosphorylation of Smad2/3 at the SXS motif located in the MH2 (C-terminal) domain, and the noncanonical pathways, such as MAPKs, which are induced by various growth factors and the TGFβ receptors themselves. The canonical- and noncanonical pathways come together onto Smads through post-translational modifications (e.g., phosphorylation) of the linker domain, as depicted. Although linker phosphorylation occurs regardless of TGFβ stimulation via the action of intracellular kinases (CDKs, MAPKs, and GSK3), it is enhanced in the presence of the ligand [165]. **(B)** Treatment scheme for the kinase screen shown in **(C)** and **(D)**. CCSW1 CCA cells were pre-treated for 1 h with the indicated inhibitors at a concentration of 10 μM, followed by the addition of TGFβ1 (5 ng/mL for 6 h) or not**. For more details regarding the tested kinase inhibitors see the text above. (C)** and **(D)** Representative immunoblots **(C)** and the respective quantifications **(D)**, showing the involvement of Erk- and p38 MAPKs as well as GSK3 in the regulation of pSmd3L signaling in CCA. Multiple comparison was performed using a two-way ANOVA test, and a post-hoc Tukey test. Error bars indicate the SEM of at least two biological replicates. *** denotes a *p*-value less than or equal to 0.001.

2.3.3. The oncogenic phospho-Smad3 signaling appears to be abrogated by the IRD E738.

As aforementioned, IRD E738 is a multiple kinase inhibitor ([148], **also see Figure19**), similar to most of indirubin analogues [113]. Considering the role of MAPKs and GSK3 in the regulation of Smad3 phospho-isoforms, we hypothesized that E738 can reduce pSmad3L levels through a mechanism

similar to that of the abovementioned kinase inhibitors (**see Figure 16**). We therefore used our panel of patient-derived CCA cell lines, and tested E738's effects on pSmad3C (S423/425) and pSmad3L (S204/213) protein levels. As shown in **Figure 17** Upon treatment with E738 and its biotinylated form– referred to as E738 B–a clear reduction in the protein levels of pSmad3L (S204/213) was observed in the absence/presence of TGFβ stimulation. The protein expression of pSmad3C, on the other hand, showed an opposite pattern, with significantly higher levels when compared to those of pSmad3L (**for more details see Figure 17**).

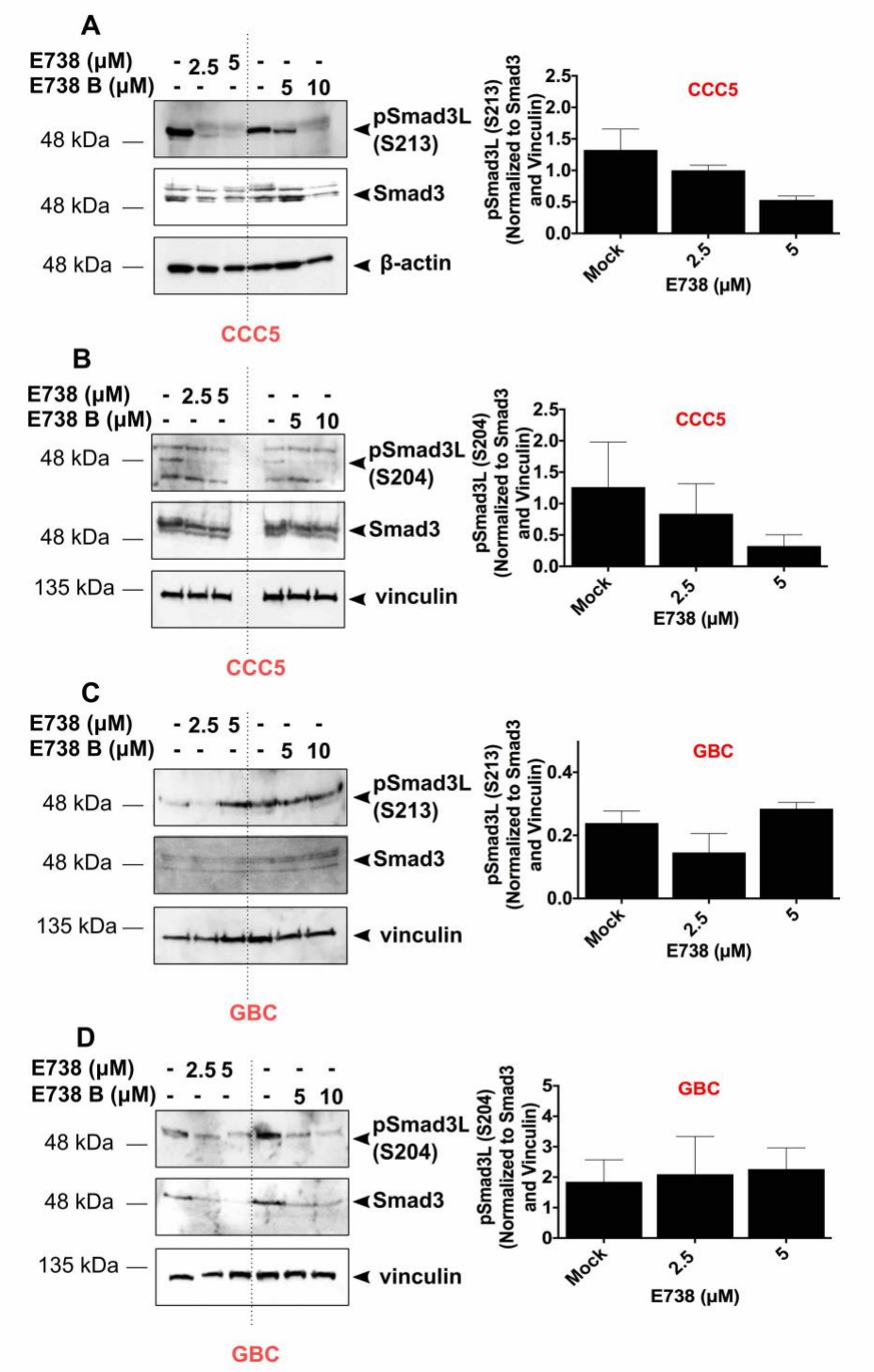


Figure 17. ***Cont.***

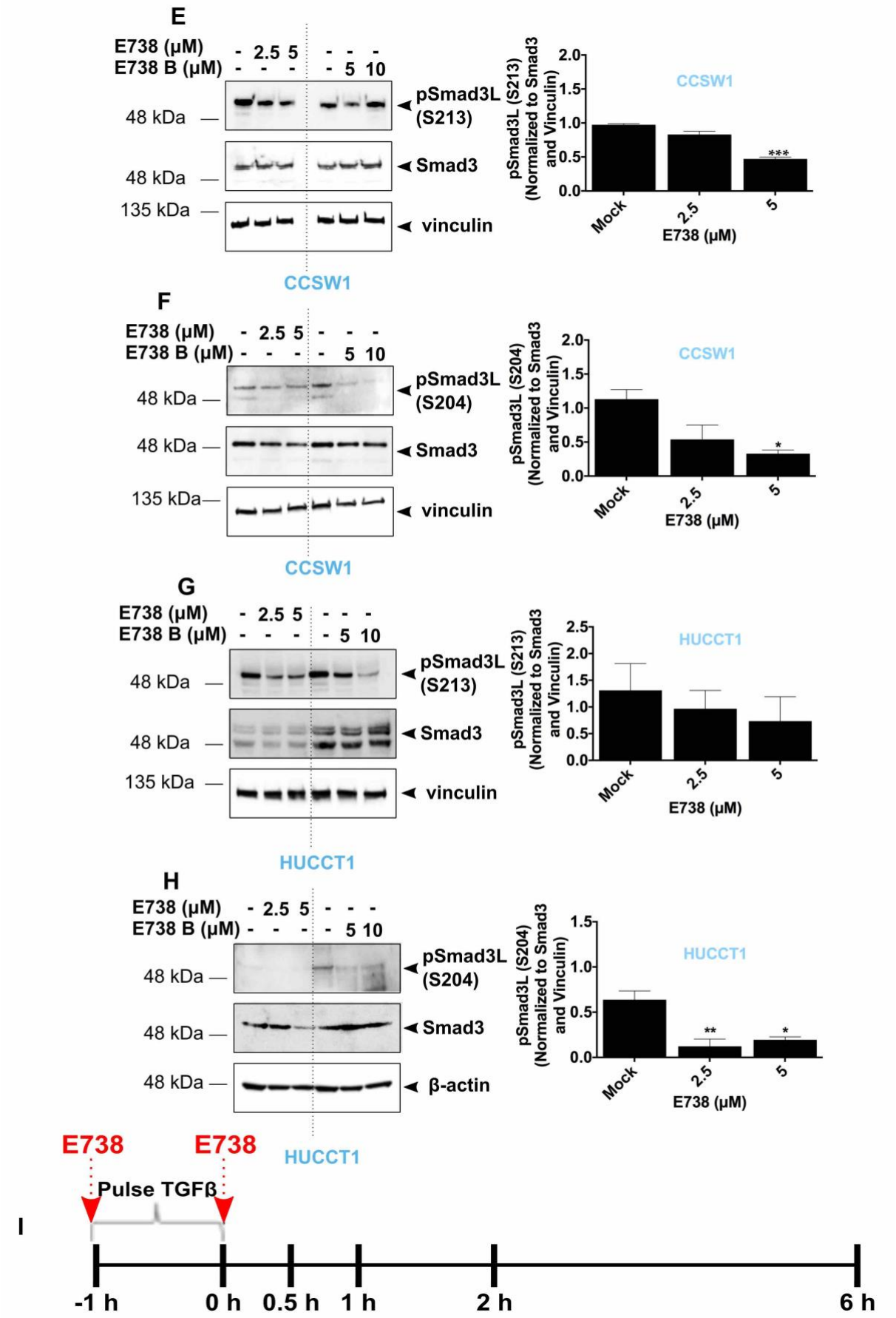


Figure 17. ***Cont.***

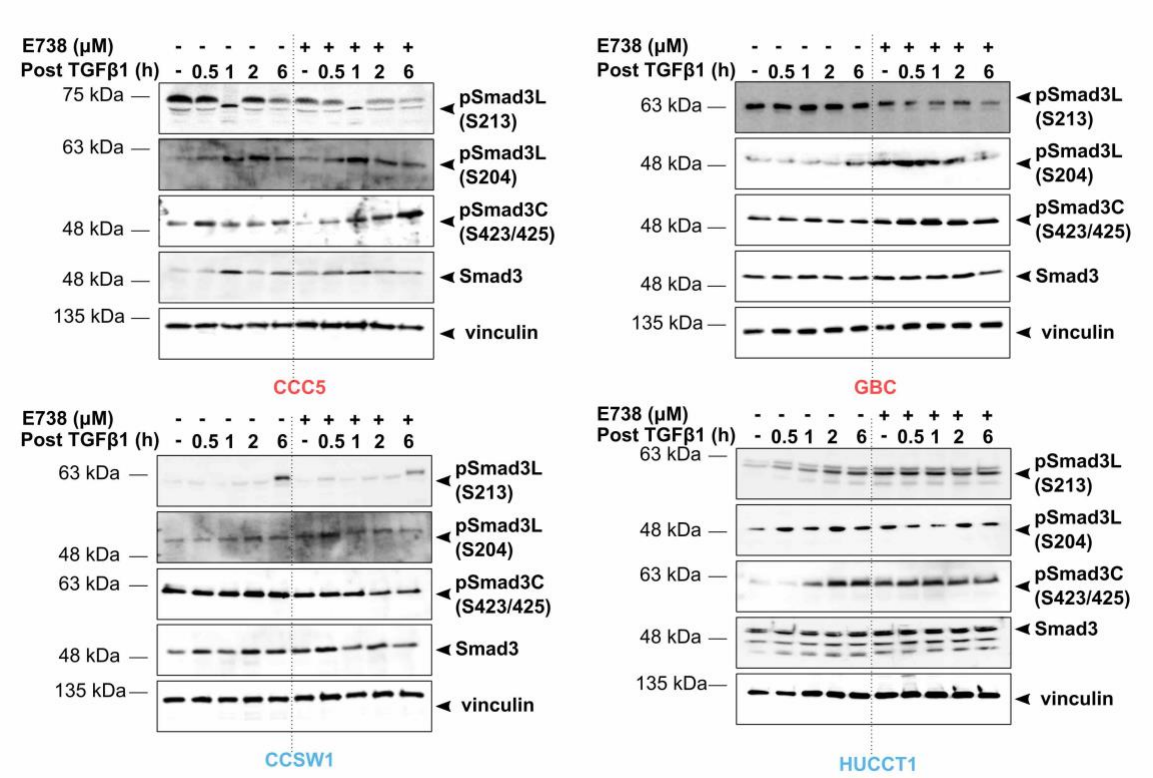


Figure 17. IRD E738 inhibits the carcinogenic Smad3 phospho-forms in a panel of patient-derived cholangiocarcinoma (CCA) cell lines. (A-D) 24 h treatment with the indirubin derivative, E738 and its biotinylated form (E738 B) has a strong inhibitory effect against Smad3 linker phosphorylation (204 and 213 serine residues) in extrahepatic CCA cells, CCC5 and GBC, as illustrated by the representative immunoblots and their respective densitometric analyses. **(E-H)** A similar pattern of E738's regulation of Smad phospho-isoforms was detected in case of the intrahepatic CCA cell lines, CCSW1 and HUCCT1. **(I)** The reduction of Smad3 linker phosphorylation could be also observed upon TGFβ1 stimulation (5 ng/mL) prior to the addition of E738 (5 μM) at the indicated time points. In contrast to pSmad3L levels, the protein levels of the cytostatic pSmad3C were found to either remain stable or even increased. Quantifications of the immunoblots were performed by normalizing the pSmad3 levels to the respective total Smad3 levels and loading controls (vinculin and βactin). Error bars represent the SEM of three biological replicates. Statistical significance between mock (0.1% DMSO) and E738 treatment was calculated using a one-way ANOVA test, and a post-hoc Tukey test. *, **, and *** indicate *p*-values less than or equal to 0.05, 0.01, and 0.001. A color code is used to distinguish between intra- and extrahepatic CCA cells, with blue indicating the former and red indicating the latter.

2.3.4. E738 synergizes with known kinase inhibitors, targeting MEK1/2/Erk, GSK3, and CDK to reduce pSmad3L protein levels.

Based on the results from the kinase inhibitor screen (**see** **Figure 16**), we were inclined to investigate the combinatory effects of pharmacological inhibitors of the involved kinases and IRD E738 on pSmad3L protein levels. As shown in **Figure 18**, the concomitant use of the Erk1/2 MAPK inhibitor (U0126) with E738 led to a more profound reduction of pSmad3L (S213) protein levels, as compared to single-treatments in HUCCT1 and CCSW1 iCCA cell lines. GSK3 and CDK inhibition showed similar effects on S213 phosphorylation status when combined with E738 in the CCSW1 cell line, whereas p38 blockage was not found to induce a major alteration in the pSmad3L levels in our experimental setting (**Figure 18**). With regards to the S204 residue, which is known to be phosphorylated by Erk and p38

MAPKs, as well as by GSK3, we observed a mild decrease in its phospho-levels upon treatment with most inhibitors, as expected, in CCSW1 cells. This effect, however, was not found to be clearly enhanced in the presence of E738 (**Figure 18**). Collectively, these results further confirm the involvement of Erk1/2/MEK, p38, and GSK3 in the regulation of pSmad3L signaling. Furthermore, they suggest a potential synergy between indirubin derivatives and small molecule inhibitors of the aforementioned kinases, which is most likely due to the simultaneous targeting of multiple protein kinases involved in Smad3 linker phosphorylation.

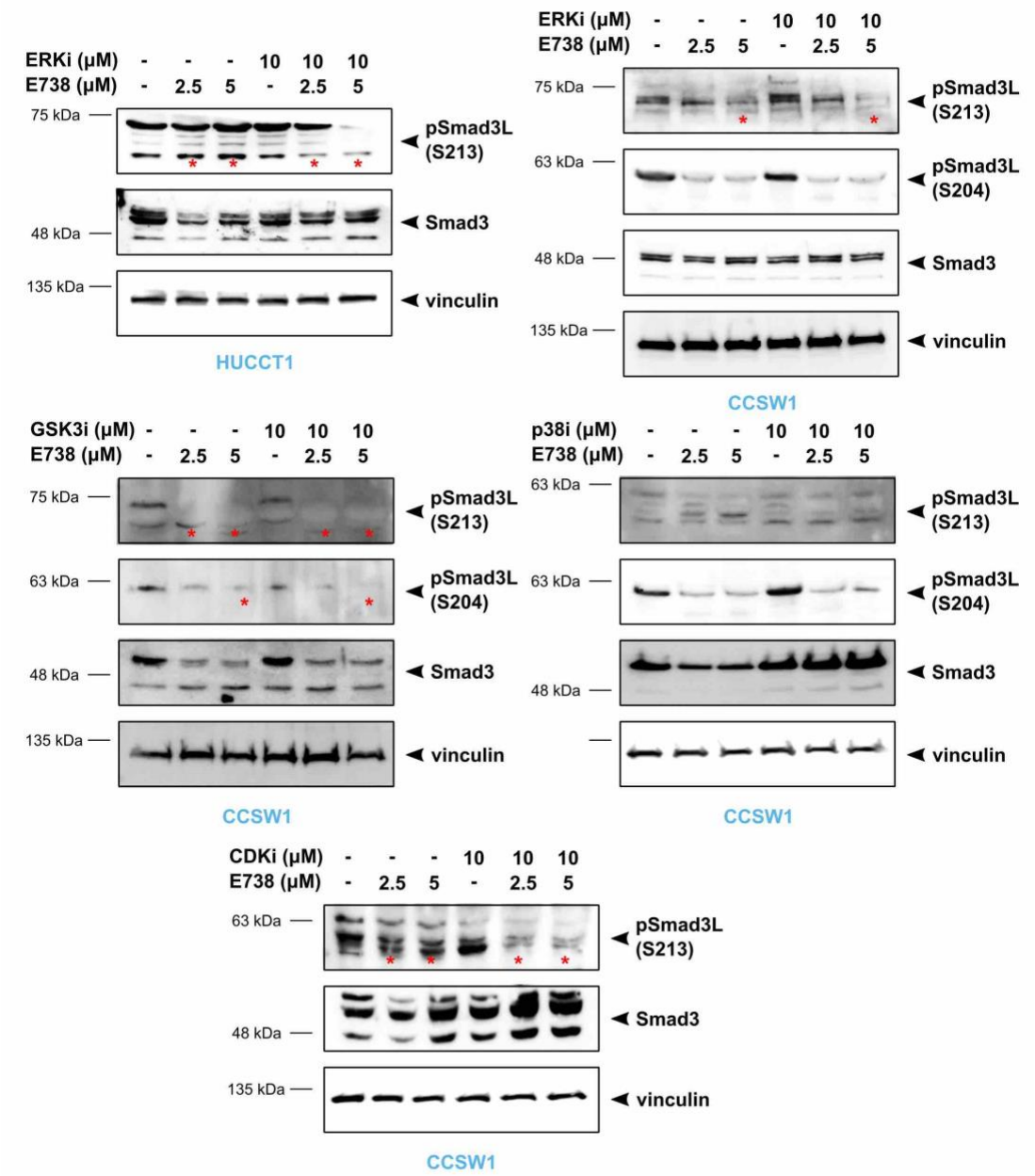


Figure 18. IRD E738 synergizes with other S/T kinase inhibitors to reduce Smad3 linker (pSmad3L) phosphorylation. The intrahepatic cholangiocarcinoma cell lines, HUCCT1 and CCSW1 were treated with E738 or the inhibitors of Erk1/2 (U0126), p38 (SB203580), GSK3 (BIO), and CDK (Flavopiridol, also known as Alvocidib) kinases at the indicated concentrations for 24 h with TGFβ1 (5 ng/mL) pre-treatment for 1 h. Immunoblotting was then performed using site-specific antibodies against pSmad3L. As indicated by red asterisks, the combination of the indirubin derivative (E738) and Erk-

GSK3- and CDK inhibitors causes a more profound reduction in pSmad3L levels. Vinculin was used as the loading control.

2.3.5. The kinase inhibitory activity is most likely indispensable to the E738-mediated Smad2/3 reduction, as well as to the inhibition of phospho-Smad3 signaling.

To further investigate whether the E738's kinase inhibitory function has a role in mediating R-Smad depletion and pSmad3L inhibition, we aimed to synthesize a kinase-inactive analogue of E738 by replacing the H-atom at 1-position with a methyl group, expecting to impair the binding ability of the small molecule to the kinases' ATP-binding pocket (**see Figure 19 and Figure 20)**. The resulting derivative–referred to as XC47–showed no significant inhibitory activity when screened against 410 kinases (**see the results from the *in vitro* kinase profiling, Figure19**). Looking at steady-state Smad2/3 protein levels, we observed no major alteration in response to various concentrations of XC47 as compared to its parental derivative, E738, in all tested CCA cell lines (**Figure 20**). Additionally, the small molecule failed to inhibit the oncogenic pSmad3L (S213) protein expression (**Figure 20**), suggesting that E738's inhibition of protein kinases is essential to the observed R-Smad depletion and the regulation of linker phosphorylation. This is consistent with our previously published results on two 7'-aza-indirubin derivatives, namely E846 and E847, which could neither deplete total R-Smads nor inhibit phosphorylation of Stat3, due to the lack of kinase inhibitory activity [148].

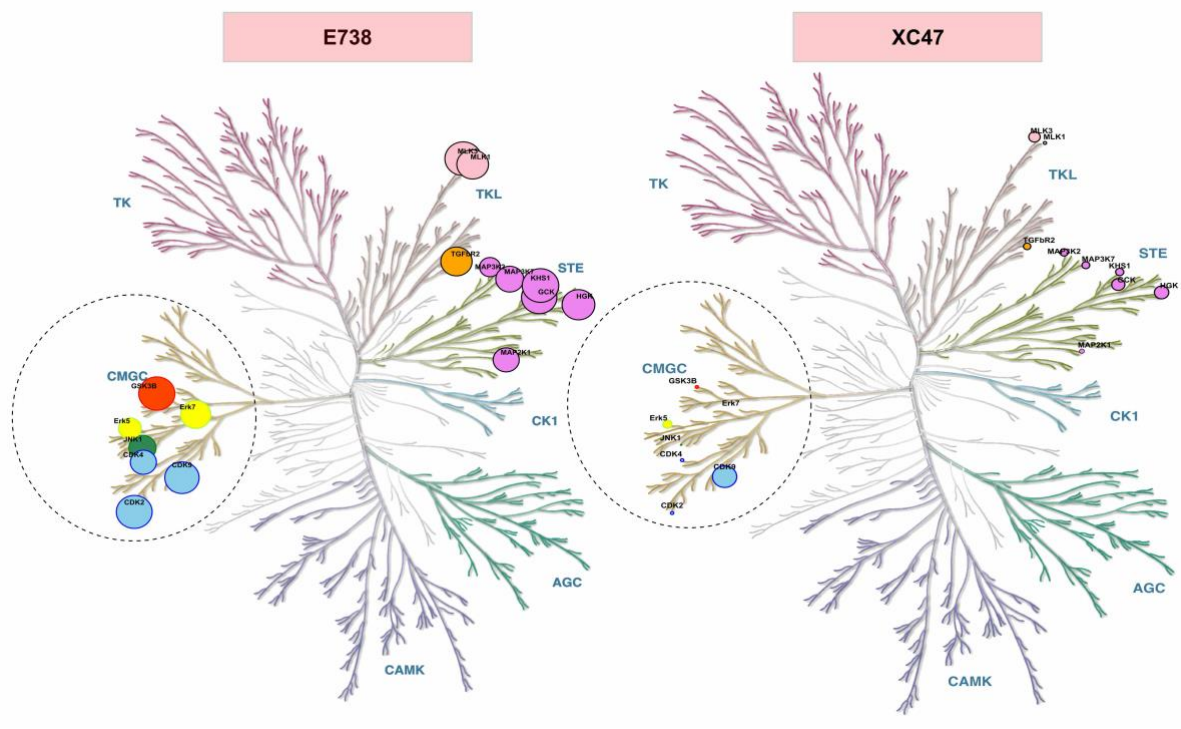


Figure 19. E738 is a strong kinase inhibitor, blocking the activity of several kinases responsible for Smad linker phosphorylation, an effect which is lost in case of XC47 (also see Figure 20). *In vitro* kinase profiling of 0.5 μM of E738 and its kinase-inactive counterpart, XC47 in a screen encompassing 410 wt- and mutant kinases (WildType+ 2Point-Profiler from ProQuinase). TGFβ signaling-related protein kinases with more than 50% inhibition by E738 are presented and the size of circles conversely correlates with %residual activity. A color code has been used for discrimination between different families of kinases, with blue indicative of CDKs, green indicative of JNK MAPKs,

yellow indicative of Erk MAPKs, red presenting GSK3, pink representative of MAPKKKs, orange presenting TGFβRII (TGFbR2), and violet showing MAPKKs. The dashed circle highlights the kinases directly involved in Smad linker phosphorylation. The map is generated using the open-access web-based tool described by Eid *et al.* [166] and the illustration reproduced courtesy of Cell Signaling Technology, Inc.

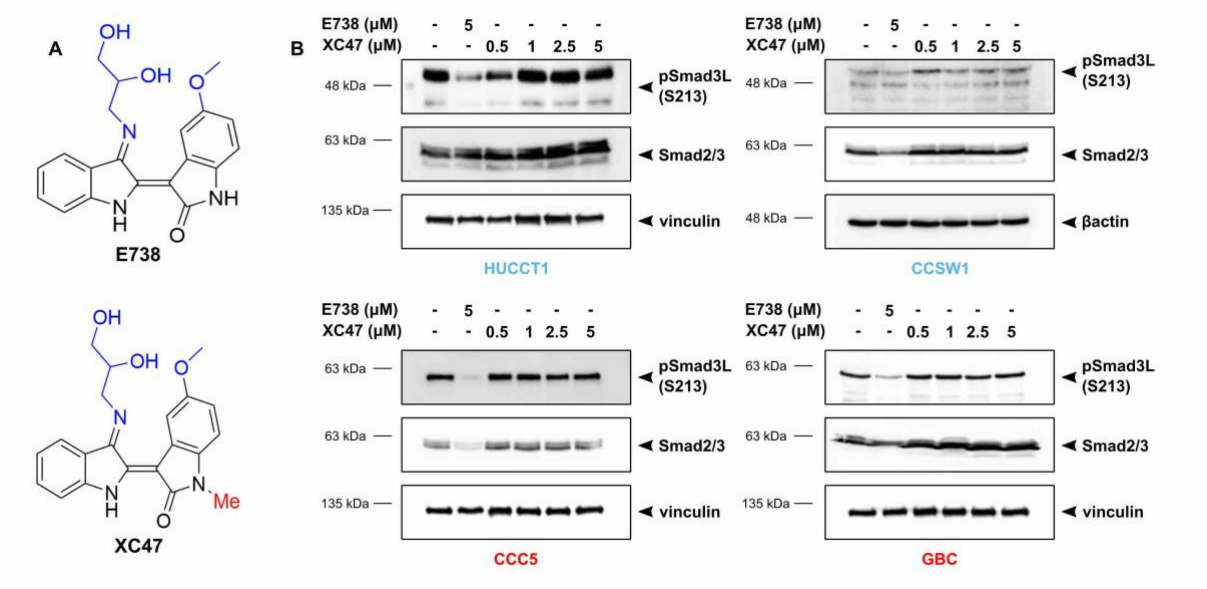


Figure 20. IRD E738's regulation of Smad3 linker phosphorylation is, at least in part, linked to total Smad3 degradation and the kinase inhibitory function of the molecule (also see Figure 19). **(A)** Chemical structures of E738 and its kinase-inactive analogue, XC47. The H-atom in E738 is replaced with a methyl group in XC47 (colored in red) in order to disturb its binding ability to the ATP-binding pocket of protein kinases. The methoxy- and propanediol moieties (colored in blue), which have been previously shown to enhance indirubin-mediated Smad2/3 reduction were kept in the new derivative [148]. **(B)** Intra- and extrahepatic cholangiocarcinoma cells (colored in blue and red, respectively) were treated with increasing concentrations of XC47 and 5 μM of E738 for 24 h, as indicated. Unlike E738, XC47 was not found to be capable of depleting non-activated Smad2/3 protein levels, as well as Smad3 linker phosphorylation, suggesting a role for the kinase inhibition in the observed effects. Vinculin and βactin served as loading controls.

2.4. phospho-Smad3L signaling in cholangiocarcinoma cells is associated with tumorigenic behavior.

2.4.1. Concomitant inhibition of TGFβ/Smad signaling, Erk/p38 MAPKs, and GSK3 appears to have synergistic cytotoxic effects in cholangiocarcinoma cells.

In addition to the inhibition of TGFβ signaling through depletion of steady state Smad2/3 pools, E738 blocks the activity of Erk1/2 (**see Figure 14**) and several other kinases, among others, GSK3 and CDKs (**see Figure 19**). Putting this in the context of cholangiocarcinoma, it has been reported that blocking the canonical TGFβ pathway, as well as the noncanonical signaling (e.g., Erk MAPKs) is an effective approach for inducing cytotoxic- and apoptogenic responses in human CCA cells [167, 168]. We therefore aimed to examine whether the E738-mediated inhibitory effects on Smad2/3 stability as well as pSmad3L signaling relate to the small molecule's cytotoxic response in CCA cells. Moreover, we sought to compare the E738-mediated cytotoxicity with that of other known kinase inhibitors, looking for

any potential combinatory effect. As shown in **Figure 21**, different eCCA and iCCA cells were treated with increasing concentrations of E738, BIO (GSK3 inhibitor), SB203580 (p38 MAPK inhibitor), and U0126 (Erk1/2 inhibitor) as mono-treatments and a combination of E738 with either of the three inhibitors. Cellular viability was then assessed using SRB and MTT assays. We observed lower cellular survival percentages when using E738 together with the pharmacological inhibitors of GSK3, Erk1/2- and p38 MAPKs in the intrahepatic HUCTT1 cell line as compared to single-treatments, assessed by SRB assay (**Figure 21A**). With regards to the extrahepatic GBC cells, changes in cellular survival were less profound as compared to other cells and were found to be significant only when E738 was combined with BIO and with 20 μM of U0126 (**see Figure 21A, results of the SRB assay**). We broadened our investigation by including more cell lines from either intra- or extra-hepatic origin and observed a similar cytotoxic response in the CCSW1 cells compared with that of HUCCT1, i.e., significantly reduced cellular survival when using a combination of E738 with either BIO or U0126 compared with either small molecule alone, as measured by MTT cytotoxicity assay (**Figure 21B and 21C**). Importantly, the combination of E738 and BIO led to the highest toxicity in all tested cells, and in CCC5 cells was found to be the only combination, which led to a significant decrease in cellular viability, as compared to single-treatments (**Figure 21B and 21C**). Additionally, we determined the coefficient of drug interaction (CDI) for combinatorial treatments to better describe the nature of drug combination outcomes. CDI is calculated as follows; (A+B)/A*B, a CDI of <1 was defined as synergistic, =1 as additive, and >1 as antagonistic. Several combinations were found to induce synergistic effects, however, as aforementioned, the ***"E738 and BIO combination"*** showed the lowest CDI across all tested cell lines (**Figure 21B**).

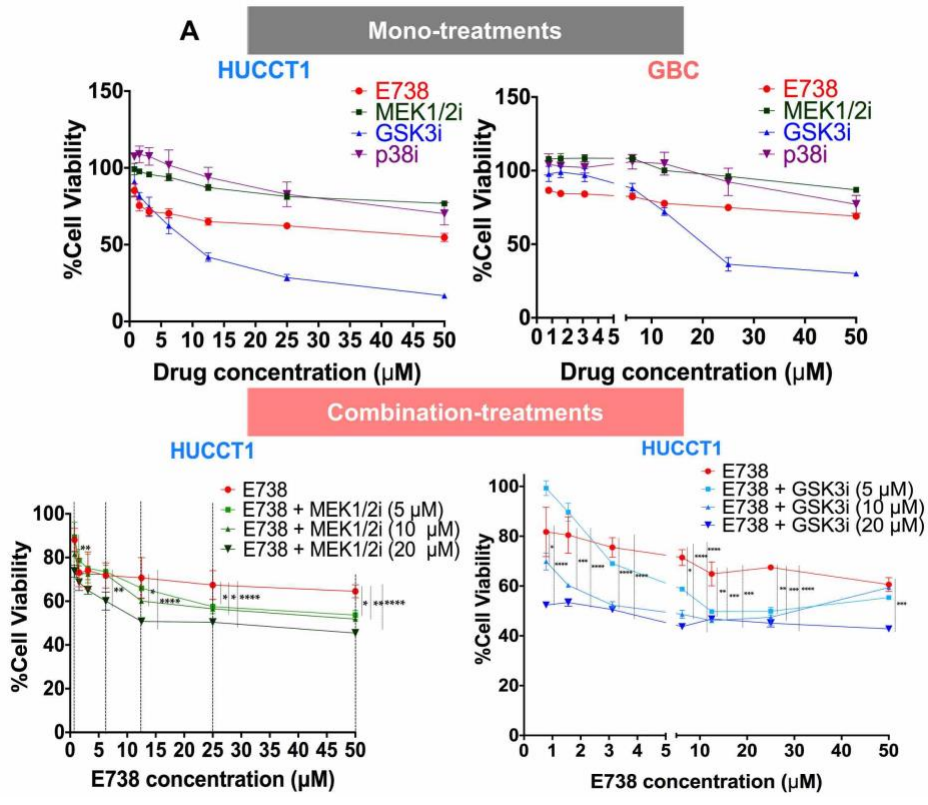


Figure 21. ***Cont.***

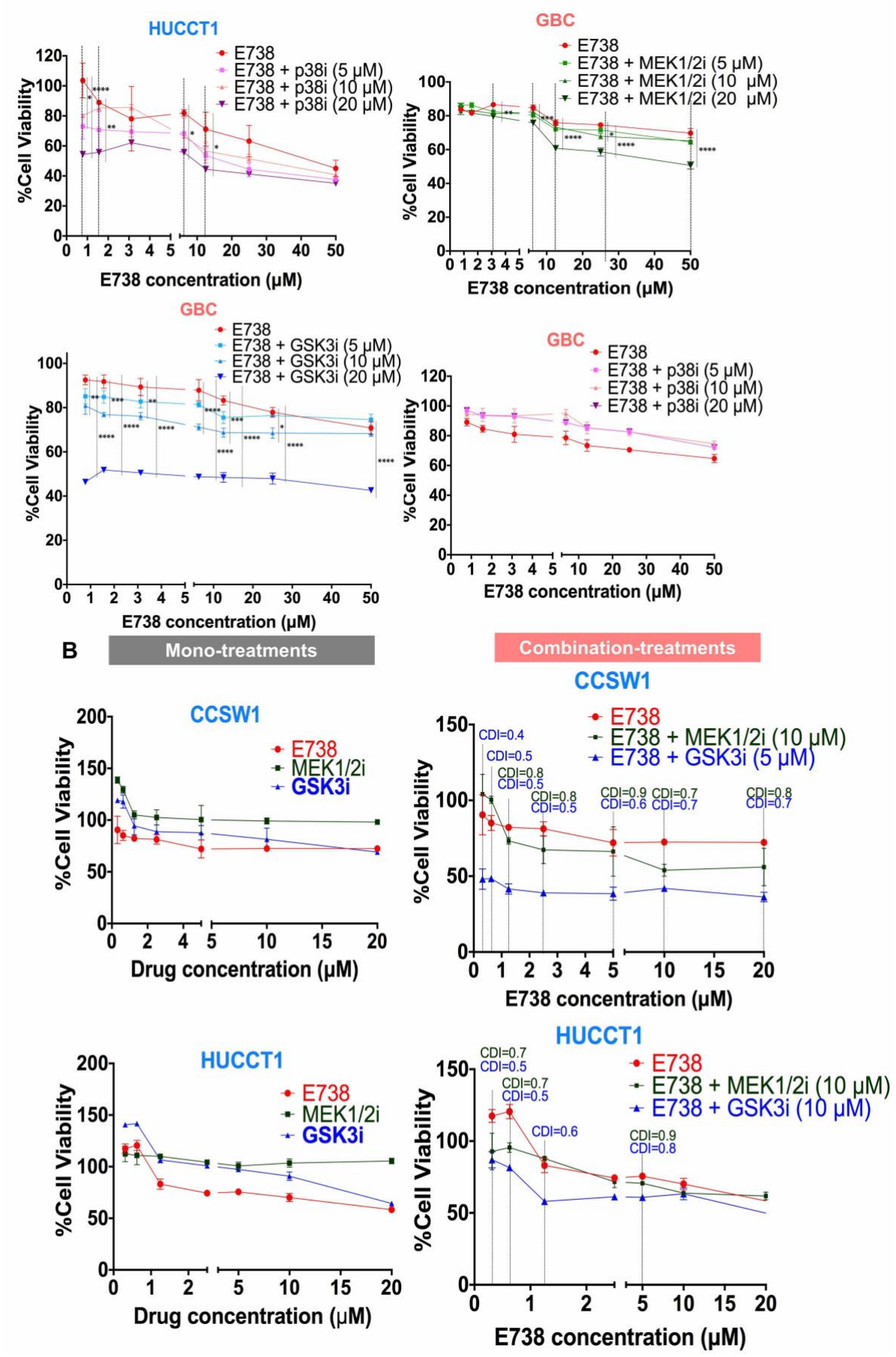


Figure 21. ***Cont.***

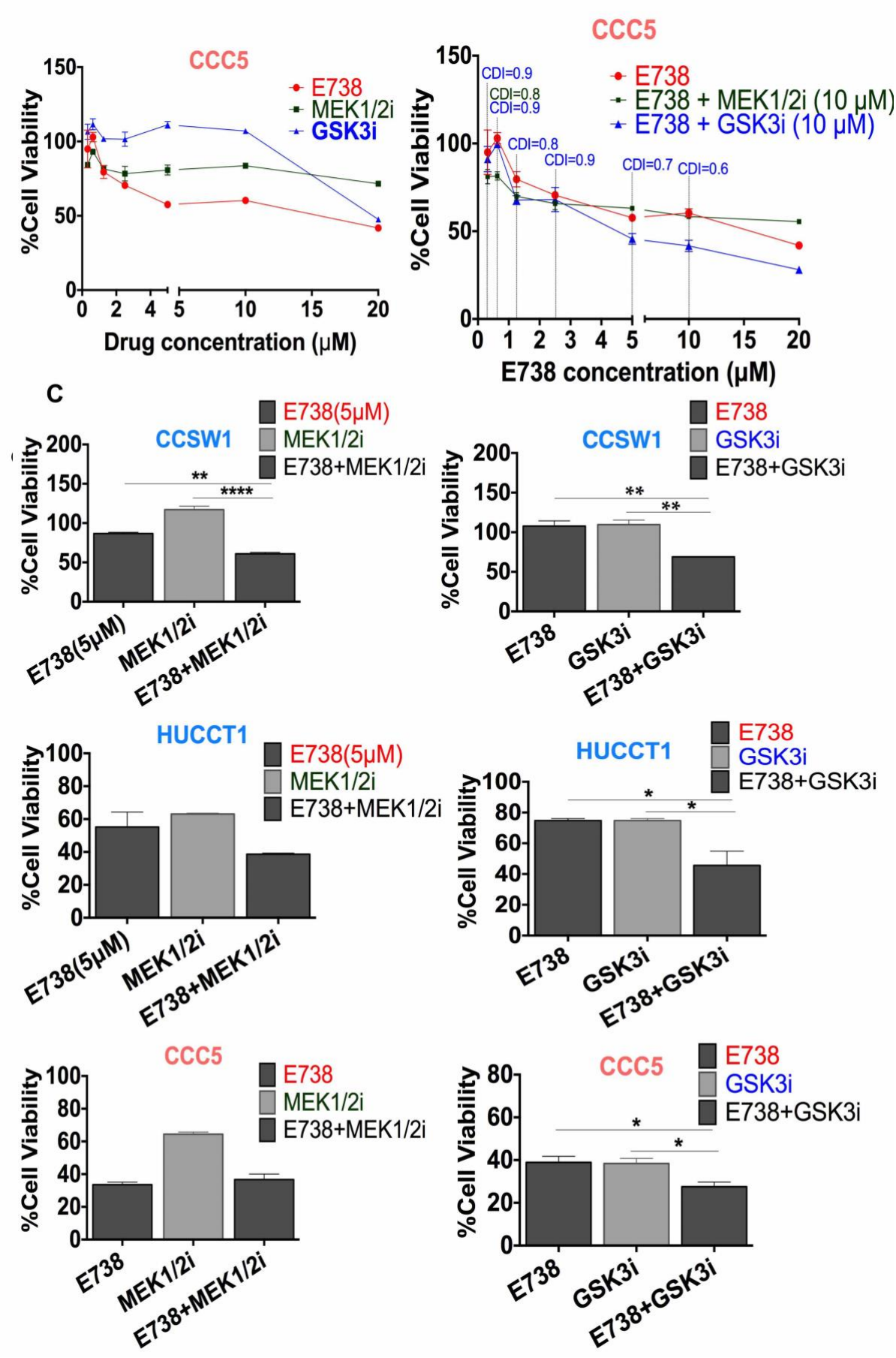


Figure 21. ***Cont.***

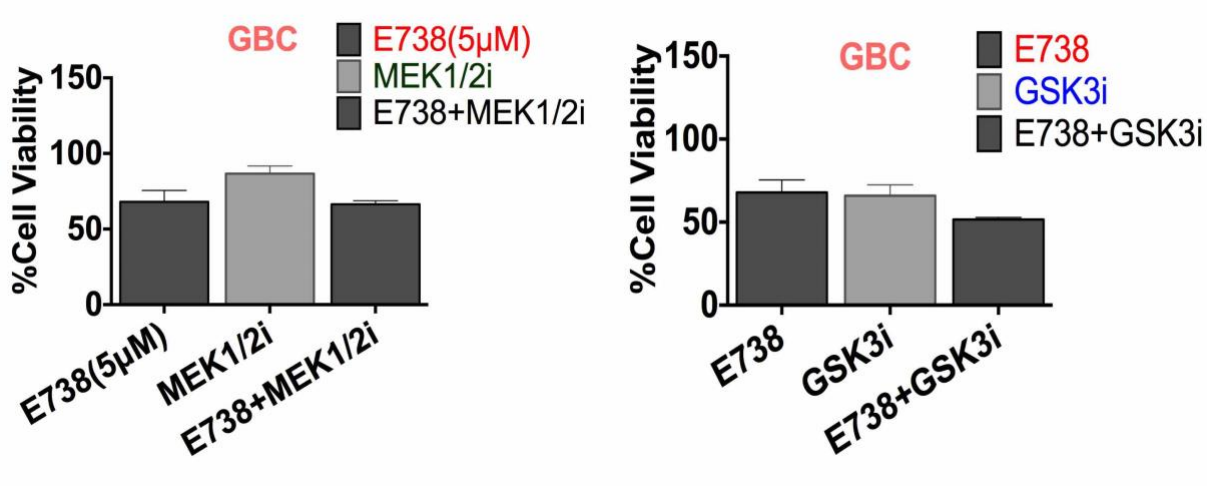


Figure 21. Inhibition of TGFβ signaling in combination with blocking Erk1/2- p38- and GSK3 kinases induces synergistic effects on cholangiocarcinoma cytotoxicity. (A) The intrahepatic CCA cells, HUCCT1 and the extrahepatic GBC cells were treated with various concentrations of E738 and small molecule inhibitors of Erk/MEK, p38, and GSK3 for 24 h. For combination treatments, increasing concentrations of E738 were combined with the indicated concentrations of each of the kinase inhibitors. **For more information regarding the name of the kinase inhibitors see the text in 2.4.1.** Cytotoxicity was then measured, using SRB assay. **(B)** The intrahepatic CCSW1 and HUCCT1 as well as the extrahepatic CCC5 cell lines were treated with increasing concentrations of E738 and the inhibitors targeting Erk/MEK- and GSK3 kinases, as single-treatments, and the combination of various concentrations of E738 and fixed concentrations of Erk/MEK- and GSK3 inhibitors, as indicated. MTT cytotoxicity assay was performed at 24 h post-treatment. The combination index (CDI), which is indicative of nature of drug combination, was then calculated using the "(A+B)/A*B" formula. A CDI less than 1 is considered a synergistic effect. **(C)** E738 (5 μM), MEK- and GSK3 chemical inhibitors (10 μM), as well as a combination of E738 with each of the latter drugs were employed on different cell lines and MTT assay was carried out after 24 h. Based on the data obtained from SRB and MTT assays, the combination of E738 and BIO (GSK3 inhibitor) appears to induce the highest synergy in all tested cell lines. Percentage cellular viability was determined as the number of treated- over control (0.1% DMSO) cells. Error bars indicate the SD of at least three independent experiments, each done in triplicates. Multiple comparisons were performed using a two-way ANOVA test, followed by a post-hoc Tukey test. *, **, ***, and *** denote *p*-values less than or equal to 0.05, 0.01, 0.001, and 0.0001, respectively. A color code is used to distinguish between intra- and extrahepatic CCA cell lines, with blue representative of the former and red representative of the latter.

2.4.2. IRD E738 induces apoptotic/necrotic cell death in cholangiocarcinoma cells, an effect which is also observed with GSK3 chemical inhibition.

Based on the promising data obtained from cytotoxicity assays (**see Figure 21**), we sought to see whether the E738-mediared pro-apoptotic effects, in addition to that of the cytotoxic effects, could be enhanced upon combining the small molecule with other known S/T kinase inhibitors involved in Smad3 linker phosphorylation. Using annexin V/propidium iodide (AV/PI) staining, we could show that E738 treatment results in the transition of cells through the early apoptotic quadrant (AV^+/PI^-) in CCSW1 and HUCCT1 intrahepatic CCA cell lines (**Figure 22A and 22B**). Chemical inhibition of MEK1/2/Erk using U0126 was not found to cause a significant change in the number of early- or late apoptotic/necrotic (AV^+/PI^+) cells. In addition, its combination with IRD E738 was not found to enhance the pro-apoptotic response of the latter in our experimental setting (**Figure 22A**). On the other hand, the indirubin derivative, BIO, which is known to be a potent GSK3 inhibitor (***recently reviewed and originally***

described in [113, 127]***, respectively***) showed a clear increase in the number of late apoptotic/necrotic cells at the tested concentration in both HUCCT1 and CCSW1 cell lines (**Figure 22A and 22B**). Combining E738 and BIO demonstrated a similar apoptogenic response compared to that of BIO alone but significantly higher than that of E738 (**Figure 22A and 22B**), suggesting that the observed apoptosis is mostly mediated by BIO. It has to be taken into concentration that the concentrations used and/or duration of the treatment have a crucial role in the small molecules' pro-apoptotic effects, and hence a concentration/time-dependent assay appears to be necessary in order to draw a precise conclusion with regards to the pro-apoptotic effects of each of the molecules alone and in combination. Despite this, we could show strong apoptosis in response to GSK3 chemical inhibition by BIO, which fits well with the results of SRB/MTT assays wherein BIO and its combination with E738 induced the most potent cytotoxic effects (**also see Figure 21**).

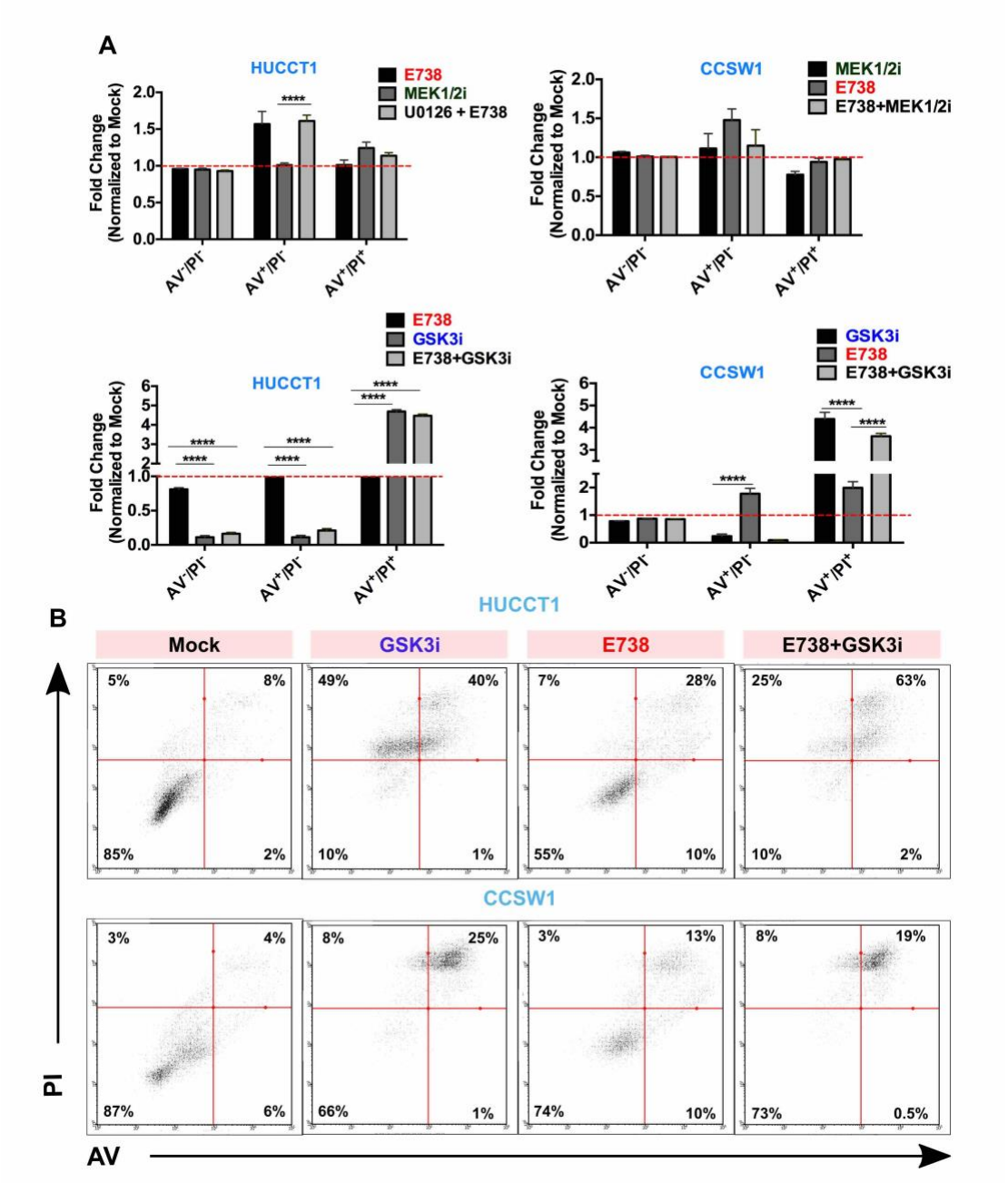


Figure 22. The indirubin derivatives–E738 and BIO–induce pro-apoptotic effects in cholangiocarcinoma cell lines. (A) 24 h treatment with 5 μM of E738 increases the number of early apoptotic cells (AV^+/PI^-) in HUCCT1 and CCSW1 cells, as determined by flow cytometric analysis of annexin V/propidium iodide (AV/PI) staining. BIO, the GSK3 small molecule inhibitor (designated as GSK3i) induces late apoptotic/necrotic cell death (AV^+/PI^+) at the tested concentration and time point (10 μM for 24 h). U0126, which targets Erk MAPKs (designated as MEK1/2i) was not found to cause a major change in the apoptotic population. Percentage cell population in each quadrant was normalized to the respective mock (0.1% DMSO) treatment. **(B)** Representative density plots illustrate a clear transition towards the AV^+/PI^- and AV^+/PI^+ quadrants upon treatment with E738 and BIO (GSK3i), respectively. Data in **(A)** present mean ± SD of three biological replicates, one of which is depicted in **(B)**. Multiple comparisons were made using a two-way ANOVA test, and a post-hoc Tukey test. **** denotes a *p*-value less than or equal to 0.0001.

2.4.3. IRD E738 regulates mRNA expression of multiple signaling molecules associated with apoptosis and EMT in cholangiocarcinoma cells.

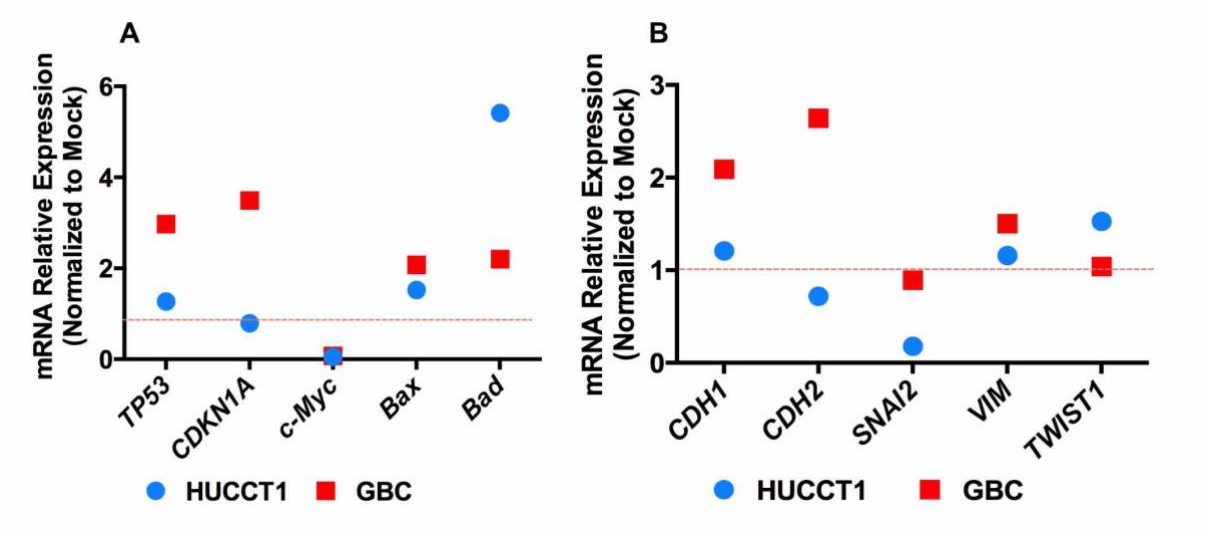


Figure 23. IRD E738's regulation of genes involved in apoptosis and EMT in cholangiocarcinoma cells. (A) HUCCT1 and GBC cholangiocarcinoma cells were treated with 5 μM of E738 for 24 h, after which qRT-PCR was performed for assessing the mRNA levels of genes related to cytotoxicity and apoptosis. In GBC, E738 appeared to induce the pro-apoptotic molecules, p53 (coded by *TP53*), p21 (coded by *CDKN1A*), *Bax*, and *Bad*. In HUCCT1, a similar pattern was observed with an exception for *CDKN1A*, which showed no significant alteration with E738 treatment. *c-Myc*, a well-known oncogene and a crucial mediator of TGFβ-mediated cellular invasion, was found to be strongly reduced by the small molecule in both cell lines. **(B)** E738 (5 μM, 24 h) up/downregulates/ E/N-cadherin (coded by *CDH1/2*, respectively) as well as the pro-EMT molecule, SNAIL2 (also known as Slug, coded by *SNAI2*) in HUCCT1 cells, suggesting the possibility of EMT suppression in cholangiocarcinoma by the indirubin derivative. Relative mRNA expression was calculated using the ΔΔ Ct method where the Ct values of each gene were normalized to those of the housekeeping gene (vinculin) and mock (0.1% DMSO) treatment. Data represent mean of two biological replicates, each done in triplicate. A color code is used to distinguish between intra- and extrahepatic CCA cells, with blue presenting the former and red presenting the latter.

Based on the results obtained from MTT/SRB cytotoxicity assays and flow cytometric analysis of apoptosis (**refer to Figure 21 and Figure 22**), we aimed at investigating the regulation of multiple genes related to pro-apoptotic and anti-proliferative effects in response to E738. As shown in **Figure 23A**, the small molecule was found to upregulate the mRNA levels of several pro-apoptotic- and anti-proliferative

genes in HUCCT1 and GBC cell lines, namely *TP53* (coding for the tumor suppressor, p53), p53's downstream gene and the cell cycle regulator, *CDKN1A* (coding for p21), as well as the anti-survival BCL2 family members, *Bax* and *Bad*. In addition to this, we observed a strong downregulation of the proto-oncogene and one of the key molecules of the TGFβ-induced cytostatic program, *c-Myc* **(Figure 23A)**, altogether suggesting an induction of the apoptogenic program by the molecule. Additionally, we sought to assess the status of genes mediating the epithelial-to-mesenchymal transition. As thoroughly described in the introduction, EMT is one of the key biological features of the TGFβ pathway in cancer and development, which is mostly mediated through crosstalks with TGFβ's noncanonical signaling networks, such as Erk/p38/JNK MAPKs [79] (**for more details see sections 1.1.3. and 1.1.5.2.**). In cholangiocarcinoma, inhibition of canonical- and noncanonical TGFβ signaling has been shown to be an effective therapeutic approach through blocking TGFβ-induced EMT [167, 169-171]. In light of this, we looked into the regulation of multiple EMT-associated genes by E738 and found increased and decreased mRNA levels of *CDH1* (coding for E-cadherin) and *SNAI2* (coding for SNAIL2, also known as Slug), respectively (**Figure 23B**). This might suggest the ability of the small molecule in inhibiting the EMT process in CCA cell lines, which however requires in depth investigation of the signaling molecules involved, as well as the functional assays such as the cellular migration/invasion analyses. Other tested EMT-associated genes, *VIM* (encoding vimentin) and *TWIST1* remained either stable or slightly increased in response to E738, and *CDH2*, which codes for N-cadherin was found to be downregulated by E738 in case of HUCCT1 and upregulated in case of GBC (**Figure 23B**).

2.5. The direct interaction of IRD E738 with components of the ubiquitin-proteasome pathway appears to have a role in the control of R-Smad stability.

2.5.1. Biotinylated E738 displays decreased and increased binding to several subunits of the proteasome complex.

As described in the section **2.1.1.1.**, we have synthesized the biotin-conjugated small molecule (designated as E738 B) and confirmed its biological functions with respect to TGFβ/Smad signaling (**see Figures 9 and Figure 17**). We then performed pull-down assays *in vitro* and in cells followed by mass-spectrometry analyses in order to identify E738 B's interacting partners. As depicted in **Figure 24A**, we could find some hits with key roles in the proteasome complex, showing either increased or decreased binding to biotinylated E738, as compared to that of biotin- and mock-treated cells. The proteasome complex consists of two subcomplexes: (i) the 20S catalytic core and (ii) the 19S regulatory particles [172]. The former includes four hetero-heptameric β rings that are flanked on either side by two α rings; PSMA3 and PSMA6 are α subunits of the 20S core particle and PSMD4 is a subunit of the regulatory particle (**see Figure 24B**) [172]. To validate these hits, we used siRNA-mediated knockdown of the aforementioned subunits and monitored for Smad2/3 reduction in response to E738 B treatment. As determined by immunoblotting and the respective densitometric quantification, upon inhibition of PSMA3, the small molecule-mediated Smad2/3 degradation was found to be less efficient (**Figure 24C and 24D**), suggesting that this subunit may contribute to E738's mode of action with regards to modulation of R-Smad stability. In case of PSMD4, the basal levels of Smad2/3 appeared to be

decreased, thus causing a further reduction in the protein levels upon drug treatment (**Figure 24C and 24D**).

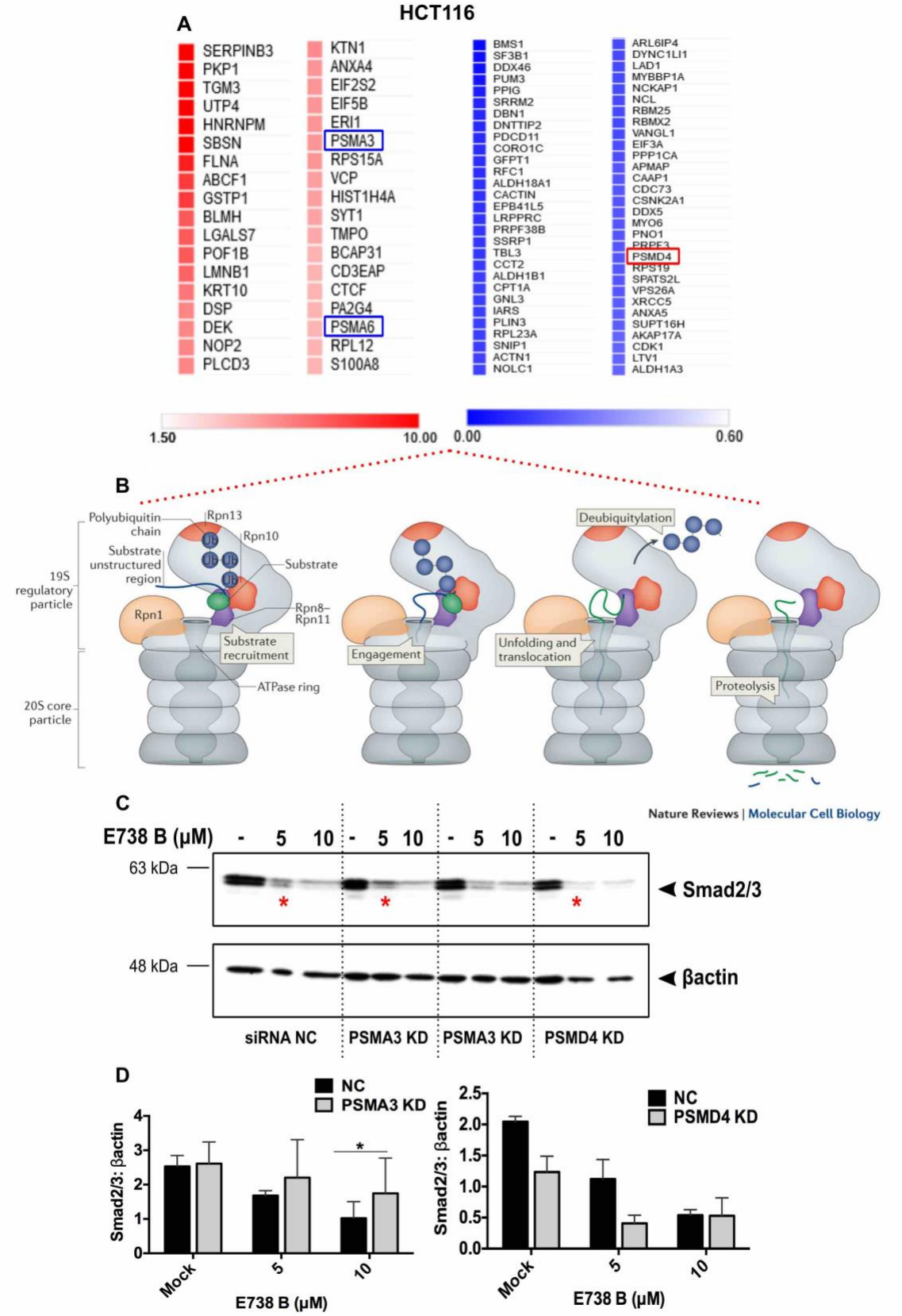


Figure 24. Proteasome complex subunits might be responsible for IRD E738-mediated steady-state Smad2/3 degradation. (A) HCT116 cells were incubated with E738 (5 μM) for 24 h, after which cell lysates were pulled down using immunoprecipitation by streptavidin beads, ran on SDS-PAGE, and analyzed by mass-spectrometry. Red and blue in the heatmaps are indicative of proteins with more- and less binding ability to biotin-conjugated E738, respectively. Data were normalized to the respective biotin-only-treated cells and mock (0.1% DMSO) treatment. **(B)** A schematic taken from Bhattacharyya *et al.* [172], illustrating the proteasome complex structure, which consists of the 20S catalytic core and the 19S regulatory particles. Hits obtained from our proteomics analysis, PSMA3/6 are α subunits of the 20S particle, and PSMD4 is a subunit of the 19S regulatory particle (Rpn10). **(C)** and **(D)** *PSMA3/6* and *PSMD4* were knocked-down (KD) in HCT116 cells using the respective siRNA. A non-targeting siRNA was used as the negative control (designated as siRNA NC). At 48 h post-transfection, cells were treated with E738 B at the indicated concentrations for 24 h. As shown by representative immunoblots **(C)** and the corresponding densitometric quantifications **(D)**, knocking down *PSMA3* leads to a less profound Smad2/3 reduction, whereas genetic inhibition of *PSMD4* causes a greater depletion of total Smad2/3 levels by E738 B. βactin served as the loading control. Error bars ± SEM, *n* = 3. Statistical significance was made between the negative control and knockdown cells using a one-way ANOVA test, and a post-hoc Tukey test. * denotes a *p*-value less than or equal to 0.05.

2.5.2. Biotinylated E738 demonstrates a direct interaction with several E3 ubiquitin ligases.

Next, we looked for hits from the family of E3 ubiquitin ligase enzymes in the *in vitro* proteomics analyses (**Figure 25A**), and found several E3 ligases, showing a direct interaction with biotinylated E738. These include HERC5 from the HECT-domain containing family of E3 ligases as well as the RING-domain enzymes, RNF2 and RBBP6. Using siRNA-mediated repression of the respective genes followed by E738 B treatment or not, we found significantly higher Smad2/3 protein levels when knocking down *RBBP6* compared with that of the non-targeting siRNA (**Figure 25B**). Inhibition of *HERC5* and *RNF2* genes demonstrated no major alteration either in basal Smad2/3 levels or in its levels upon small molecule treatment (**data not shown**).

In addition to the aforementioned enzymes, we included two other E3 ubiquitin ligases in our siRNA screen, MDM2 and VHL. MDM2 is a RING domain-containing protein ligase with the tumor suppressor, p53 being its primary target [173]. We could previously show a p53-dependent regulation of cytostatic genes with E738 treatment, which was further confirmed by an opposite pattern of p53 target gene expression–including *MDM2*–between cells harboring wt p53 or not (**see Figure 12**). Additional experiments showed a strong accumulation of p53 protein levels with treatment due to an increase in the stability of the protein by E738, which most likely involves p53's negative regulator, MDM2 (**data not shown**). This prompted us to look for the effects of MDM2 on E738-mediated Smad2/3 depletion, using genetic repression of *MDM2*. As shown in **Figure 26A**, in *MDM2*-knockdown cells, basal levels of Smad2/3 appear to be elevated regardless of E738 treatment. A similar pattern was observed in case of *VHL* knockdown, where baseline Smad2/3 showed a profound increase, which in turn led to higher protein levels in E738-treated cells compared with that of the negative control (**Figure 26B**). VHL is the substrate recognition subunit of the von Hippe-Lindau (VHL) E3 ligase complex with hypoxia-inducible factor 1α (HIF 1α) being its main substrate [174]. The increased Smad2/3 protein levels upon genetic inhibition of *VHL* stimulated us to investigate whether this E3 ligase is related to R-Smad protein stability, irrespective of E738 treatment. For this, we first knocked down both E3 ligases–*MDM2* and *VHL*–whose repression was shown to increase baseline Smad2/3 protein status. As demonstrated in **Figure 26C**,

upon dual inhibition of *MDM2* and *VHL*, we observed a clear induction in Smad2/3 protein levels, which was found to be clearly higher than *MDM2* knockdown, but not significantly different than that of *VHL*, suggesting that the effects are most likely resulting from *VHL* repression. In addition, we checked for the protein levels of BMP-related Smad1/5/8 and found higher levels in case of *VHL* knockdown compared with negative control, however, the effect was less profound than that of Smad2/3 (**Figure 26C**). We then tested TGFβ target gene expression in cells transfected with siRNA against *VHL*, *MDM2*, or both and compared it with that of the negative control, using qRT-PCR. As shown in **Figure 26D**, most of TGFβ target genes namely, *ID1*, *ID2*, *ID3*, and *CTGF* were found to be upregulated in *MDM2*/*VHL*-knockdown cells. *ID3* and *CTGF* showed the highest fold change in mRNA expression compared with negative control in case of both single- and double knockdown cells (**Figure 26D**). Importantly, Smad2- and Smad3 mRNA levels were found to be either unaffected or reduced, respectively (**Figure 26D**), implicating that the strong accumulation of proteins is most probably not controlled transcriptionally, but rather directly at the protein level. Consistent with the upregulation of TGFβ target genes, protein levels of the transcriptionally active Smad3–pSmad3 (S423/425)–were found to be increased in *VHL*-knockdown HEK-293 cells (**Figure 26E**).

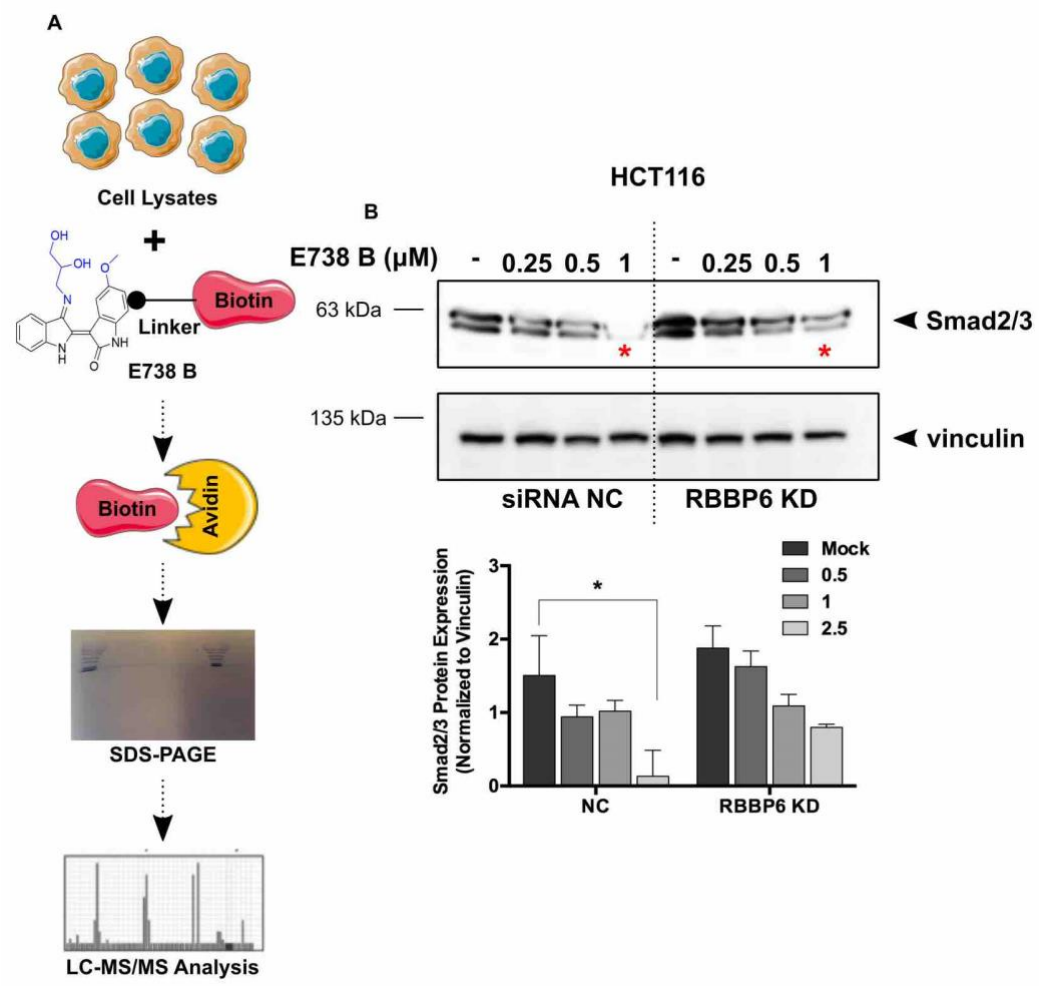


Figure 25. The E3 ubiquitin ligase, RBBP6 might be responsible for E738-mediated Smad2/3 degradation through a direct protein-small molecule interaction. (A) Schematic illustration of the *in vitro* streptavidin-based immunoprecipitation. HCT116 cell lysates were incubated with biotin-conjugated

E738 (E738 B) at a concentration of 100 μM for 24 h, then pulled down by streptavidin beads, followed by SDS-PAGE and mass-spectrometry (MS) analyses. **(B)** One of the hits obtained from MS analysis, the E3 ligase *RBBP6*, was genetically repressed (KD) in HCT116 cells using siRNA. 48 h after transfection, cells were treated with increasing concentrations of E738 B for 24 h. A non-targeting siRNA was used as the negative control (denoted as siRNA NC). Immunoblots (top) and the respective quantification (bottom) demonstrate a more profound reduction of Smad2/3 in the presence of RBBP6, suggesting the involvement of this E3 ubiquitin ligase in the E738's regulation of Smad2/3 stability. Vinculin served as the loading control. Red asterisks highlight the difference in Smad2/3 protein expression between the negative control and *RBBP6*-knockdown cells. Statistical significance was calculated between mock (0.1% DMSO) and E738 B treatment using a one-way ANOVA test, and a post-hoc Tukey test. Error bars represent the SEM of three biological replicates. * indicates a *p*-value less than or equal to 0.05.

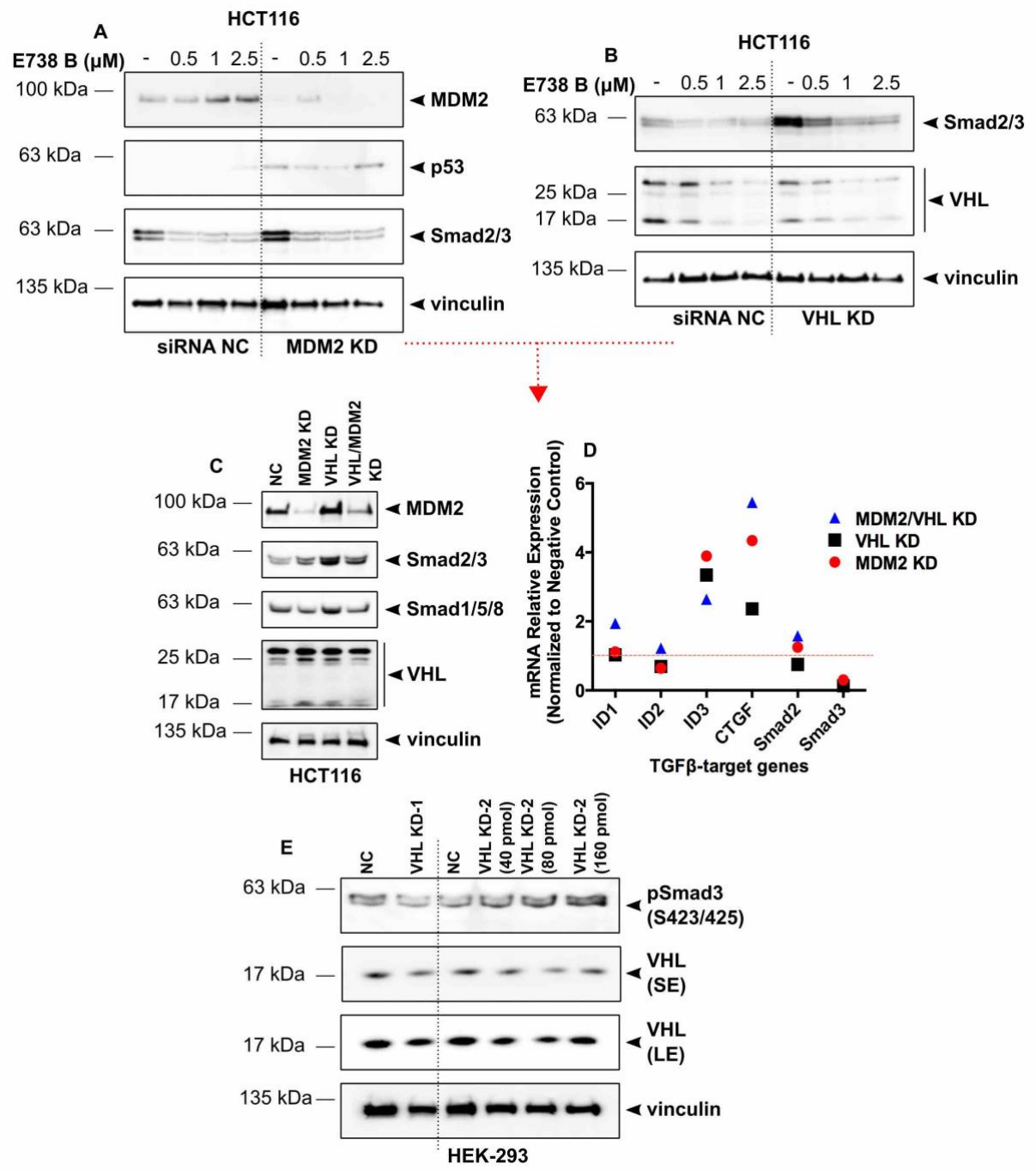


Figure 26. Smad2/3 might be a new substrate for the E3 ubiquitin ligases, MDM2 and VHL. (A-C) The RING-finger ubiquitin ligase, *MDM2*, as well as *VHL* were knocked-down (KD) either individually or simultaneously in HCT116 cells. At 48 h post-transfection, cells were treated with various concentrations of biotin-conjugated E738 (E738 B) for 24 h, then immunoblotting was performed. Cells were transfected with a non-targeting siRNA (denoted as siRNA NC) as the negative control. **(A)** We found that in *MDM2*-knockdown cells, steady-state Smad2/3 protein levels are increased regardless of the treatment. p53 was used as reference, as it is the most well-known substrate for MDM2. **(B)** A similar pattern (regarding the Smad2/3 baseline status) was observed upon siRNA-mediated suppression of *VHL*. Importantly, VHL protein levels were concentration dependently decreased with E738 B treatment. **(C)** Dual inhibition of *MDM2* and *VHL* revealed that the increase in basal Smad2/3 protein expression is most likely more efficient in *VHL*-knockdown cells. **(D)** qRT-PCR analysis of TGFβ target genes as well as *Smad2/3* upon genetic inhibition of *MDM2*, *VHL* or both, as indicated. We found a profound upregulation of *ID3* and *CTGF* in *VHL*-knockdown cells and a mild upregulation of *ID1/2* in the double knockdown cells, suggesting the activation of the TGFβ/Smad pathway in the absence of the E3 ligases. mRNA expression of *Smad2* and *Smad3*, on the other hand, was found to be either unchanged or decreased, respectively, implicating that the observed accumulation of proteins occurs independently of the transcriptional regulation. Data present the mean of two biological replicates, each done in triplicates. **(E)** Knocking down *VHL* in HEK-293 pre-treated with TGFβ1 (5 ng/mL, 1 h) demonstrates an increase in the protein levels of the receptor-phosphorylated Smad3 (pSmad3 (S423/425)), further showing the activation of the signaling.

Altogether, our in vitro and cell-based proteomics analyses suggest RBBP6 as a direct binding partner of E738, which most probably plays a critical role in E738-mediated R-Smad degradation. Additionally, our results from the siRNA screen implicate Smad2/3 as a potential new substrate for MDM2 and VHL E3 ubiquitin ligases. In both cases, further experiments are required in order to confirm ubiquitination of Smad2/3 by the abovementioned protein ligases upon E738 treatment or at the basal level. In addition to this, the biological outcome of increased Smad2/3 protein levels (in the context of development and/or cancer) as a result of VHL repression is an exciting topic for future experiments.

3. DISCUSSION

The idea of the work presented in this chapter was initially sparked by our previous knowledge in the group regarding the contribution of total R-Smads to the control of TGFβ- and BMP signal transduction pathways. As thoroughly discussed in the introduction (**see section 1.1.4.**), TGFβ/BMP signaling is regulated at several separated levels and through various mechanisms, among others, is the ubiquitination of receptor-activated R-Smads, which marks them for degradation, and hence limits the duration of TGFβ/BMP/Smad signaling. Unlike transcriptionally active R-Smads, steady-state (total or non-activated) pools of TGFβ- and BMP central mediators have not been considered to be a major regulatory mechanism of the pathways. The results of this chapter illuminate this poorly investigated matter using small molecule approaches. **In particular, we could show the interference of multiple indirubin derivatives with TGFβ signaling in the following ways: (i) reduction of steady-state Smad2/3 pools and (ii) stabilization of transcriptionally active Smad2/3. Detailed analyses of the small molecules' actions revealed that the two aforementioned effects, at least in part, reconcile on the kinase inhibitory function of indirubin derivatives.** Interestingly, the latter quality of indirubin analogues was shown to antagonize the oncogenic R-Smad signaling, using patient-derived cholangiocarcinoma cell lines as a cancerous context. The following sections discuss in detail the above findings with regards to the modulation of TGFβ/Smad signaling by indirubins, on the one hand, and the translation of these effects into a potential therapeutic strategy for certain tumorigenic contexts, on the other hand.

3.1. Regulation of TGFβ/Smad signaling: *insights from small molecule approaches*

Our results from a large screen encompassing cell lines of various cancerous and non-cancerous origin showed a profound and reproducible R-Smad reduction upon treatment with the indirubin analogues; E738, E738 B, and E860, with the first molecule showing the strongest effect (**see Figure7, Figure 8 and Figure 9**). E738-mediated depletion of non-activated R-Smads at the protein level was in line with reduced transcription of *R-Smads*, as well as the TGFβ- and BMP target genes (e.g., *IDs*) in response to the small molecule (**see Figure 12**), altogether suggesting IRD E738 as an inhibitor of the TGFβ- and BMP pathways. These findings are consistent with a previous study of the group, showing the ability of indirubin analogues–particularly E738–in inhibiting BMP/Smad1/5/8 signaling through transcriptional regulation as well as protein degradation [148]. The latter is known to be a major mechanism by which R-Smad signaling is terminated [52]. Having that in mind and with a special focus on TGFβ-related Smad2/3, we could show that the E738-induced reduction of steady-state Smad2/3 pools is, at least in part, due to a decrease in protein stability, as co-treatment with proteasome- and protein synthesis inhibitors decreased and increased the molecule's effects, respectively. This highlights the importance of the ubiquitin-proteasome pathway not only in the control of transcriptionally active Smad2/3 but also in the maintenance of total Smad2/3 levels. Although most studies have focused on the ubiquitination of receptor-activated R-Smads in the control of TGFβ signaling (***reviewed in*** [40]),

there is evidence supporting the idea of non-activated Smad2/3 being regulated through similar dynamics. Related examples are the ubiquitination of baseline Smad2/3 by Smurf2 [43, 75, 76] and the control of steady-state Smad3 stability by the U-box-containing E3 ligase, CHIP (also known as STUB1) [77].

The activity and duration of the TGFβ/Smad pathway is not only determined by the abundance of ligands and receptors, but also through an intricate interplay between post-translational modifications of Smads, including phosphorylation, ubiquitination, and deubiquitination [52]. Through undergoing sequential steps of phosphorylation: (i) TGFβRI-induced phosphorylation of the SXS residues in the MH2 domain followed by (ii) phosphorylation of (S/T) P motifs in the linker region, Smad2 and Smad3 reach their maximum transcriptional potential, on the one hand, and are marked for proteasomal degradation, on the other hand [34, 35]. ***The "last minutes of Smads" prior to their degradation is therefore characterized by boosted transcriptional activity due to phosphorylation, which is essential for canonical TGFβ signaling-induced target gene expression, differentiation, and tissue formation.*** In this context, the indirubin derivative, E738 was found to stabilize the receptor-activated pools of Smad2 and Smad3, as determined by increased levels of C-tail phosphorylated forms of the proteins in the presence of the small molecule alone, or in combination with chemical inhibitors of protein synthesis (**see Figure 10**). This observation may look paradoxical, considering the profound decrease in the protein stability of non-activated Smad2/3. The explanation, however, underlies the molecular mode of action of most indirubin analogues; their strong inhibitory function against several protein kinases, with many of those being required for the Smad linker phosphorylation (e.g., GSK3 and CDKs) (***reviewed in*** [113]). As described earlier, this modification eventually marks Smads for recognition by the respective E3 ligases, leading to their destruction. IRD E738, therefore, is most likely able to prolong the half-life of receptor-phosphorylated Smad2/3 through inhibiting their subsequent linker phosphorylation by S/T kinases. Consistent with TGFβ-associated Smad2/3, we could previously show the short-term stabilization of C-terminally phosphorylated Smad1/5/8 by E738 upon BMP stimulation [148].

The protein kinase inhibitory function appears to give sufficient rationale for the transient phospho-Smad2/3C stabilization by indirubins, but not for the strong inhibition of TGFβ signaling observed at the level of non-activated Smad2/3 pools. The minimal effect of other pharmacological inhibitors targeting various S/T kinases, for example CDKs and GSK3 on total Smad2/3 levels (**also see Figure 18**), suggests that kinase inhibition is possibly not the major route by which E738 inhibits R-Smad expression. To address this issue, we adopted two strategies; (i) synthesis of a kinase-inactive analogue of E738 (designated as XC47) and (ii) over-expression studies using various truncated- and mutant forms of Smad2 and Smad3. Taking the first approach, we observed a significantly less efficient Smad2/3 reduction upon treatment with XC47 compared with that of E738 (**see Figure 20**), which implicates a potential role for kinase inhibition in E738's reduction of Smad2/3. This conclusion, however, has to be drawn cautiously, since the two small molecules may have other differences in their

mode of action in addition to that of protein kinase inhibition. This may include the negative regulation of deubiquitinating enzymes (e.g., USP34), which has been previously proposed as a mechanism through which E738 attenuates total R-Smad levels [148].

Our results from the transient over-expression of different Smad2/3 mutants revealed multiple important facts with regards to E738's regulation of steady-state R-Smad pools (**also see Figure 13**); (i) E738's depletion of non-activated Smad2/3 occurs independently of TGFβRI-induced Smad activation, as over-expressing truncated forms lacking the SXS motif resulted in similar or even a higher reduction in response to the indirubin derivative, (ii) the linker domain of Smad2 and Smad3 is essential for the E738's regulation of Smad2/3 stability, as overexpressing truncated proteins lacking this region caused almost none Smad2/3 reduction, and (iii) linker phosphorylation influences E738-mediated degradation of Smad2/3, as forced expression of Smad2/3 constructs carrying mutations in the conserved (S/T) P motifs leads to a more profound reduction of Smad2/3 half-life. The latter finding associates E738's regulation of signaling modules stemming from other pathways with its effects on Smad2/3 degradation. Such pathways include p38- JNK- and Erk MAPKs, as well as Wnt signaling, all of those are either directly or indirectly involved in Smad linker phosphorylation (**see Figure 16**) [20, 165, 175-178]. We could previously show the activation of the two branches of noncanonical TGFβ signaling, p38 and JNK, by E738 [148]. On the other hand, Erk activity (unlike the rest of MAPKs) was found to be attenuated in response to the indirubin derivative in HEK-293 and a set of primary CCA cell lines (**see Figure 14**). Putting this next to the observation, showing a stronger Smad2/3 depletion in cells over-expressing constructs with Erk site-mutants compared with wt Smad, we hypothesized that the E738-mediated Erk inhibition might contribute to the stability of non-activated Smad2 and Smad3. In support of this, combining E738 with known chemical inhibitors of MEK1/2/Erk was found to have a synergistic effect on reducing the protein levels of Smad2/3. Consistent with our results, it has been previously reported that Erk phosphorylation stabilizes Smad2 protein levels, and hence enhances TGFβ responses [21]. Additionally, the study showed that the amount of exogenous Smad2 protein levels is decreased and increased in case of phosphorylation-defective- and phosphorylation-mimicking mutants, respectively [21]. Despite the combinatory effect of E738 and Erk chemical inhibitors on Smad2/3 depletion, single-treatments with the two tested inhibitors, U0126 and PD0325901, showed no significant alteration in Smad2 or Smad3 levels, suggesting that E738's regulation of Smad2/3 pools–although related to Erk-mediated phosphorylation–involves additional molecular targets. One possibility could be the involvement of other kinases required for Smad linker phosphorylation, such as CDKs and GSK3, whose activity was shown to be strongly inhibited by E738 in our *in vitro* kinase profiling or p38/JNK MAPKs, which are known to be activated by the indirubin derivative. In support of the contribution of GSK3 to non-activated Smad stability, is the work of Guo *et al.* who showed that GSK3-mediated phosphorylation of Smad3, but not Smad2, promotes its ubiquitination, followed by proteasomal degradation [78]. Considering this, GSK3 inhibitory function of E738 is most likely not a determinant of reduced Smad3 stability by the molecule. Whether other kinases, including CDK, p38,

and JNK, which phosphorylate distinct and overlapping residues in the linker domain are involved in baseline properties of Smad2 and Smad3 proteins could be an interesting topic for further investigations.

3.2. Regulation of "phospho-Smad signaling": *implications in cholangiocarcinoma*

The malignant switch in the TGFβ pathway involves, among other mechanisms, abnormal phosphorylation patterning of R-Smads (**see also sections 1.1.5. and 2.3.**). The so-called *"pSmad signaling"* relies on linker phosphorylation on conserved (S/T) P motifs by various kinases originating from other pathways, on the one hand, and C-tail phosphorylation by the active receptor complex, on the other hand [164]. Depending on the kinases involved, different phosphorylation forms of R-Smads are created with either anti- or pro-cancer effects [164]. Our finding with regards to E738's effects on Smad turnover, which was shown to involve modulation of specific linker residues prompted us to explore a potential connection between the small molecule's mode of action and the pSmad pathway. Aberrant Smad phosphorylation has been shown to play a role in the pathogenesis of some human malignancies, such as liver- and colorectal cancer [164, 179, 180]. However, its role in the diagnosis and treatment of cholangiocarcinoma remains to be investigated. With that in mind, we thought of using CCA as a cancerous model for studying the relevance of pSmad signaling to human cancer, on the one hand, and the accessibility of this signaling pathway for small molecule intervention, on the other hand.

In close collaboration with Prof. Dooley's lab (Heidelberg University, Germany), we detected impaired pSmad signaling in CCA patients and cell lines, i.e., a decrease in the tumor-suppressive pSmad3C protein levels and an increase in those of the oncogenic pSmad3L (S204 and S213 residues) (**see Figure 15**). Considering the primary mode of action of indirubins–protein kinase inhibition–we tested the effects of different kinase inhibitors, including IRD E738, on pSmad3L protein levels. Promising data not only identified several S/T kinases, such as MEK/Erk, p38, and GSK3, as regulators of pSmad3L levels in CCA (**Figure 16**), but also demonstrated potent inhibition of the pro-cancer Smad3 phospho-isoforms by E738 (**Figure 17 and Figure 18**). By including the kinase-inactive analogue of E738–XC47–we could provide further evidence that modulation of pSmad3L signaling relies on the protein kinase inhibitory function of E738, although it does not fully rule out the possibility of alternative mechanisms.

At the molecular level, several signaling networks have been shown to drive the pathogenesis of cholangiocarcinoma, which could be classified as follows: (i) signaling pathways related to cellular survival (e.g., Ras-MAPK and PI3K/Akt pathway), (ii) signaling pathways related to cellular fate and differentiation (e.g., Wnt and TGFβ pathways), as well as (iii) metabolic- and epigenetic-related signaling modules (e.g., isocitrate dehydrogenase (IDH)1/2) (***recently reviewed in*** [168]). Focusing on TGFβ signaling, multiple reports have shown that both canonical- and noncanonical branches of this pathway could be exploited for the treatment of CCA [181]. For example, blocking TGFβ-activated ERK1/2 has been reported to suppress TGFβ-mediated invasion and metastasis in intrahepatic CCA cell lines, while maintaining its tumor-suppressive functions [167]. Another example with regards to the *in vitro* effects of TGFβ signaling modulation on CCA is the work of Lustri *et al.* who showed that chemical blockage of

TGFβRI is capable of reducing cellular migration in human CCA cell lines [170]. The above reports together with our mechanistic insight on the involvement of various kinases–including those mentioned in the CCA treatment–in the control of pSmad3L signaling inspired us to investigate whether the inhibition of oncogenic pSmad3L could alter the tumorigenic features of our panel of patient-derived CCA cell lines. In light of this, the combination of IRD E738 and pharmacological inhibitors of the involved kinases, namely ERK1/2, p38, GSK3, and CDKs was found to induce synergistic effects on the inhibition of pSmad3L protein levels, on the one hand (**see Figure 18**), as well as cellular viability and apoptogenic responses, on the other hand (**see Figure 21 and Figure 22**). In view of our results, several lines of studies have recently shown the beneficial outcome of targeting the aforementioned kinases and related signaling modules in the treatment of CCA. With regards to the MAPK pathway, anti-MEK therapies have shown promising results in the control of intrahepatic CCA. Interesting examples are the use of chemical inhibitors of Erk, PD901 and U0126, which showed high efficacy against *in vitro* and *in vivo* iCCA models, in particular in tumors harboring *KRAS* mutations, through reducing cellular proliferation, induction of apoptosis, as well as effects on the stroma [182]. Small molecule inhibitors of CDK4/6, such as ribociclib and palbociclib have also shown to be a promising approach in the treatment of iCCA [183, 184]. In particular, concomitant administration of the CDK4/6 inhibitor, palbociclib with MLN0128, which inhibits mTOR was found to induce profound and synergistic growth inhibitory functions *in vitro* and *in* vivo [183]. Sparse reports on the regulation of p38 signaling in CCA have suggested a tumor-promoting role for this pathway [168], implicating the potential benefit of p38 inhibitors either as mono-treatment or in combination with other drugs in CCA therapy [185]. The use of p38 inhibitors gets more desirable, given their activation by TGFβ through the noncanonical (non-Smad) pathways (**also see section 1.1.3.**). This may explain the observed combinatory effect of p38 inhibitors and E738, as the latter is known to promote p38 activation, while inhibiting TGFβ signaling. By combining p38 inhibitors with the indirubin derivative, both molecular mechanisms could be targeted, which will then enhance the cytotoxic response as compared to either molecule alone. Finally, the role of GSK3 in CCA formation and progression appears to be highly context-dependent similar to other cancer types. Although GSK3α/β phosphorylation at S21/9, which is mediated by Akt in order to inactivate GSK3, has been reported to be increased in CCA tumorigenic areas compared with normal biliary tissue, the cancerous contexts in which GSK3 inhibition is beneficial remains to be thoroughly investigated [186]. What is clear, however, is the key role of this enzyme in cholangiocarcinoma, as it bridges the PI3K/Akt and Wnt/β-catenin pathways, both of which are part of CCA driver signaling network. Based on our results, we demonstrated a potent cytotoxic- and apoptogenic response to GSK3 chemical inhibition in CCA cell lines, and the strongest synergy, among other inhibitors, when combined with IRD E738 (**see Figure 18, Figure 21, and Figure 22**). Altogether, our findings suggest a potential synergy upon simultaneous targeting of TGFβ signaling, MEK/Erk, p38, and GSK3 in terms of reducing oncogenic pSmad3L levels, as well as inducing cytotoxic effects.

3.3. Concluding remarks and future perspectives

The results presented in this chapter emphasize on the less-studied regulatory mechanisms of the TGFβ/Smad signal transduction, i.e., the significance of non-activated R-Smad pools in the control of TGFβ's responses. For this, we adopted a chemical approach and showed a multi-layer modulation of the pathway in response to indirubin derivatives; (i) inhibition of TGFβ signaling through proteasome-mediated degradation of steady-state Smad2/3, (ii) short-term stabilization of receptor-activated Smad2/3, and (iii) regulation of noncanonical TGFβ signaling, including Erk inhibition and p38/JNK activation. Our results from transient over-expression studies demonstrated that modulation of certain phosphorylation sites in Smad2/3 linker domain by IRD E738 is most likely involved in the reduced smad2/3 stability in response to the small molecule. Mechanistically, we could show the importance of Erk-mediated phosphorylation in E738's regulation of total Smad2/3 levels. What remains to be addressed in this regard is the role of other kinases, such as CDKs and GSK3 in the control of steady-state Smad2/3 levels in the absence or presence of the indirubin derivative. This becomes particularly important, considering the strong inhibition of the aforementioned kinases by IRD E738. Additionally, our results suggested that proteasomal degradation is most probably involved in reducing protein levels of Smad2/3 carrying mutations in the conserved (S/T) P motifs of the linker region. ***Whether linker phosphorylation through the action of various S/T kinases plays a major role in the ubiquitination of non-activated Smad2/3, and if so, which E3 ubiquitin ligases are involved in this process presents an exciting topic for future projects.***

Our mechanistic insight with regards to the role of certain linker phosphorylation residues in E738's regulation of Smad turnover sparked the idea of the potential modulation of *"pSmad signaling"* by IRD E738, as well as other derivatives. Based on our promising preliminary data on the dysregulation of pSmad signaling in cholangiocarcinoma, and due to a lack of knowledge regarding the diagnostic- and therapeutic value of this pathway in CCA, we sought to use a panel of patient-derived CCA cell lines as a cancerous context in which we could study the relevance of E738's mode of action to Smad phospho-isoforms pathways. This approach provided us with multiple important facts with regards to pSmad3 signaling and CCA; (i) the most predominant Smad3 linker phosphorylation sites in CCA are the Serine residues, S204 and S213, which could counteract the cytostatic pSmad3C (S423/S425) signaling, and hence induce a malignant switch in TGFβ's responses, (ii) several S/T kinases, namely MEK/Erk, p38, GSK3, and CDKs are most likely responsible for Smad3 linker phosphorylation in CCA cell lines, (iii) the indirubin derivative, E738, causes a profound decrease in the oncogenic pSmad3L (S204/S213) protein levels, most likely through its kinase inhibitory function, and (iv) the combination of E738 and pharmacological inhibitors of the involved kinases causes a stronger inhibition of pSmad3L signaling, an effect which coincides with a synergy in cytotoxic- and pro-apoptotic effects. ***Nonetheless, several issues remain to be addressed with regards to the translation of pSmad3 signaling into a therapeutic strategy against CCA, which we are currently working on. Firstly, the anti-tumor efficacy of indirubin derivatives as well as other inhibitors in CCA cell lines and its connection***

to pSmad3L signaling requires more investigations. In particular, cellular migration/invasion has to be taken into consideration, as it is one of the major features of TGFβ signaling overactivation in CCA. Additionally, knockin mutant cell lines, carrying phosphorylation-defective residues are planned to be generated in order to study their tumorigenic features in the absence/presence of treatment. Secondly, the molecular mechanism behind the in vitro effects of E738 remains to be further studied. Through including the kinase-inactive analogue of E738, we could propose the kinase inhibitory activity as a mechanism by which the molecule reduces pSmad3L levels. However, this requires further validation, for example by comparing the differentially regulated pathways between a none/less-responsive cell line and the most-responsive cell line (CCSW1) upon E738 treatment using RNA-sequencing. Lastly, the in vivo anti-tumor effects of IRD E738, other inhibitors, and their combination could be tested in a xenograft mouse model, using CCSW1, which showed the most promising results in the in vitro assays.

CHAPTER [2]

Small Molecule Approaches for Enhancing Transcription Factor-Based Cellular Reprogramming

1.INTRODUCTION

1.1. Transcription factor-mediated cellular reprogramming towards pluripotency

The first example of cellular reprogramming was reported by Sir John Gurdon in 1962 through somatic cell nuclear transfer (SCNT). SCNT is the process of transferring the nucleus of a somatic cell to an enucleated egg, which then starts to divide, and generates an embryo, that is genetically identical to the donor of the somatic cell [187, 188]. Gurdon's discovery suggests that all genetic information is maintained in the nuclei of somatic cells, thus these cells can be reprogrammed to an embryonic, pluripotent state. Following experiments by several groups led to the establishment of embryonic stem cells (ES cells), with a developmental capacity to give rise to all cell types in our body, a property which is known as *"pluripotency"* [189, 190]. Finally, in the late 20th century, the development of cloned animals such as, mice and sheep was repeatedly reported using SCNT [191, 192]. During this time and later on, several studies suggested the existence of *"reprogramming factors"* that can change cellular fate. Early evidence came from cell fusion experiments, showing that the profile of gene expression in somatic cells can be modified through fusion with other cell types, leading to the reprogramming of these cells [193, 194]. The prospect of the existence of reprogramming factors was further shown by several other studies, reporting the induction of transdifferentiation via introducing a single transcription factor, for example the conversion of mouse fibroblasts to myoblasts through ectopic expression of MYOD [195] and the generation of megakaryocyte and erythrocyte precursors from myeloblasts by exogenously-expressed GATA1 [196]. The increasing evidence finally led to the influential discovery of induced pluripotent stem cells (iPSCs), which are somatic cell-derived pluripotent cells with a close resemblance to ES cells (***the history of reprogramming research has been summarized in an excellent review by Takahashi K. and Yamanaka S. in*** [197]).

From a technical perspective, the generation of iPSCs from terminally differentiated somatic cells represents a remarkably simpler approach compared with SCNT, i.e., ectopic expression of a defined combination of transcription factors–OCT4, SOX2, KLF4, and MYC (collectively known as *"OSKM"*)–in the host cell. The first evidence of the ability of this cocktail in the induction of pluripotency came from the generation of mouse iPSCs (miPSCs) through introducing a retroviral transduction system in mouse embryonic fibroblasts (MEFs) [198]. For this, Yamanaka's group first aimed at identifying critical factor(s) that contribute to the pluripotent identity of ES cells, using *in silico* differential display. This led to the recognition of ES cell-specific genes, referred to as ES cell-associated transcripts (ECATs), with *Nanog* being the first identified gene that has a crucial role for the maintenance of pluripotency in ES cells and early embryos [199, 200]. They followed the experiments by testing each- and a combination of all 24 candidates using the "Fbxo15–β-geo" reporter system in which β-galactosidase expression correlates with the promoter activity of F-box only protein 15 (Fbxo15)–one of the ECATs–The results showed that none of the individual factors is able to support colony formation, whereas 22 colonies were observed in cells transduced with all 24 genes, which were similar to those of ES cells. Next, they narrowed down the combination by eliminating a single factor at a time, which eventually led to the finding that

elimination of four genes, encoding OSKM factors, inhibits ES cell-like colony formation. Further testing of OSKM-transduced MEFs revealed that this combination represents the minimal set of reprogramming factors necessary for the generation of miPSCs [198]. Human iPSCs could also be developed with a similar approach, however, with several optimizations due to the poor transduction efficiency [201]. Simultaneously, a group led by James Thomson reported the generation of human iPSCs with a cocktail containing four alternative factors, namely OCT3/4, SOX2, NANOG, and LIN28 [202], which later on was used by several other groups and has even shown synergistic effects in generating human iPSCs when combined with OSKM [203, 204].

1.1.1. Transcription factor-mediated reprogramming–enhancers

One of the major concerns with regards to OSKM-mediated reprogramming is its time-consuming process (3-4weeks) as well as the extremely low efficiency (<0.01%) [197]. In this context, *"reprogramming enhancers"* have been identified and used in order to increase the reprogramming efficiency.

1.1.1.1. Reprogramming enhancers–pluripotency-associated genes

Further investigations of the minimal required reprogramming factors suggested that except for OCT3/4, the rest of the transcription factors (SKM) could be replaced; for example, SOX2 can be substituted by SOX1 or SOX3, and KLF4 is replaceable with KLF2 or KLF5 [197]. In case of MYC, it was shown that among different isoforms, L-MYC has the lowest proto-oncogenic activity, thus provides a safer approach for the generation of human iPSCs with clinical potential [205]. The redundancy in the role of "SKM" pluripotency factors allows different combinations in addition to the OSKM cocktail. Aside from this, reprogramming efficiency was shown to be enhanced by using certain transcription factors–with high expression levels in ES cells–together with OSKM. Of those, we can mention the role of cell transcription factor 1 (UTF1), Sal-like protein 4 (SALL4), LIN28, and the E3 ubiquitin ligase, TRIM71 in facilitating the generation of human iPSCs, as well as the ability of NANOG in replacing KLF4 in the OSKM combination [197]. ***The rationale behind the use of the aforementioned reprogramming enhancers for the induction of pluripotency comes from the relationship between these factors. In fact, many of these enhancers are most likely part of the gene signature induced by OSKM. Therefore, studying the downstream targets of OSKM not only provides mechanistic insights into the reprogramming process, but also allows the improvement of OSKM-mediated reprogramming protocols.***

One of the characteristics of iPSCs is their infinite capacity for proliferation, giving iPSCs the self-renewal ability in order to differentiate into any cell type. Having this in mind, it is easy to assume that cell cycle regulating genes may promote the generation of iPSCs, as it is indeed the case for one of the original pluripotency factors, MYC. The first evidence in this regard came from the inhibitory role of the tumor suppressor, p53 in reprogramming, which was followed by several studies showing that repressing p53 strongly enhances the generation of human and mouse iPSCs (***examples are;*** [206-

208]). Further supporting examples are the positive and negative effects of other cell cycle promoters (e.g., cyclin D1) and -inhibitors (e.g., ARF) on the reprogramming efficiency, respectively [209, 210].

1.1.1.2. Reprogramming enhancers–epigenetic modifiers

Transcription factor-mediated reprogramming is largely dependent on epigenetic modifications. In fact, the epigenetic signature of the somatic cell has to be removed during the process of conversion to iPSCs, so the cells can adopt a stem cell-like epigenome. These epigenetic changes include chromatin remodeling, DNA demethylation of promoter regions of reprogramming factors, such as *OCT3/4*, *Sox2*, and *NANOG*, reactivation of the somatically-repressed X-chromosome, as well as genome-wide reorganization of histones' modifications (***reviewed in*** [211, 212]). The latter is regulated by enzymes, which mediate the post-translational modifications of histones; lysine methylation and -acetylation, to name the most frequent ones. Unlike DNA demethylation and X-chromosome reactivation, changes in histone methylation- and acetylation patterns represent an early event during the reprogramming process, implicating that resetting of histone marks is necessary in initiation of the process. In light of this, it has been reported that immediately upon introducing the reprogramming factors, histone H3 dimethylated at lysine 4 (H3K4Me2) accumulates at promoter and enhancer regions of many pluripotency-related genes (e.g., *SALL4*), that are rich for SOX2 and OCT4 binding sites but deficient for H3K4Me1 or H3K4Me3 marks. In an opposite manner to H3K4Me2, H3K27Me3 is known to be depleted, showing a promoter hypomethylation in regions important for the induction of pluripotency. This stage is parallel to the loss of H3K4Me2 at the promoters and enhancers of somatic genes, leading to hypermethylation and silencing at later stages of reprogramming [211, 212]. ***Altogether, concomitant changes in histone marks of key somatic factors and early pluripotency genes, which result in down- and upregulation of mesenchymal- and pluripotency genes, respectively are among the early events in reprogramming, driving the cellular fate in a particular direction (either towards pluripotency or differentiation).***

The existence of specific histone marks during OSKM-mediated reprogramming raises the following questions; (i) what are the chromatin modifier enzymes, which participate in reshaping the epigenetic landscape of somatic cells during the process of conversion to stem cells? and (ii) how are these chromatin modifiers targeted to the pluripotency- and/or somatic genes responsible for the reprogramming process? One way to answer this is to assume that genome-wide OSKM binding sites mark the regions that are meant to be epigenetically modified. An example is the interaction of OCT4 with the WD repeat protein 5 (WDR5) on pluripotency gene promoters in order to maintain H3K4 methylation, which in turn facilitates the activation of the respective genes [213]. Another example is the interaction of H3K27 demethylase enzyme–UTX–with OSK (OCT3/4, SOX2, and KLF4) to remove the repressive histone mark, H3K27Me3 from early pluripotency genes (e.g., *SALL4* and *UTF1*). Accordingly, inhibition of UTX has been reported to be associated with H3K27 hypermethylation, and hence attenuation of reprogramming efficiency [214]. In addition to WDR5 and UTX, modifying the expression of several other epigenetic modifier enzymes, such as SUV39H1,

SETDB1, EZH2, and DOT1L has been reported to either enhance or decrease the efficiency of reprogramming [215, 216].

Apart from the enzymes involved in the regulation of histone modifications, histone composition–in particular oocyte-enriched histone variants, TH2A and TH2B–has been reported to improve the quality of both mouse- and human iPSCs generated by OSKM-mediated reprogramming [217]. In addition to this, removal of histone H1 from chromatin was found to reduce chromatin condensation, resulting in enhanced reprogramming efficiency in mice [215].

Lastly, the methylation status of the DNA itself serves as an epigenetic modifier, influencing the reprogramming progression. Although *de novo* DNA methylation by the DNA methyltransferases, DNMT3a and DNMT3b has been reported to be dispensable for reprogramming [218], oxidative demethylation of DNA appears to be an essential step in the path to pluripotency during the early phases of reprogramming [219].

As implied earlier, histone modification is a reversible process. Focusing on H3K4 methylation as an example, it is known that the complex proteins-associated with SET1 (COMPASS)-like proteins serve as the *"writers"* of H3K4 methylation, whereas members of the KDM (also known as JMJC and JARID) and LSD families of histone demethylases counteract this process, performing as *"erasers"* of methylated H3K4. Histone demethylases or the *"erasing enzymes"* may be selective for one or more specific histone proteins (e.g., H3 or H4), location of methylation sites i.e., lysine residues (e.g., K4, K9, and K27), as well as for the degree of methylation at these sites (e.g., Me1, Me2, and Me3). For example, LSD1 catalyzes demethylation of H3K4Me1, H3K4Me2, as well as H3K9Me1 and HeK9Me2, whereas KDM5s (also known as JARIDs) are thought to be specific to H3K4Me2/3 (***reviewed in*** [220]). Additionally, the JMJC-domain containing enzyme, NO66 (also referred to as MAPJD) is known to demethylase H3K4 not only when it is di- or tri-methylated (which is the case for KDM5s), but at all three states of methylation [220]. This dynamic regulation of histone methylation through the action of demethylase enzymes provides even more opportunities in the use of epigenetic modifiers as enhancers of the reprogramming process. In view of this, small molecule-mediated inhibition of LSD1 has been reported to promote the generation of human iPSCs [221]. Another example in this respect is the role of vitamin C in enhancing reprogramming efficiency through increasing the activity of KDM2A and KDM2B, which are responsible for demethylation of H3K36 [222].

1.1.2. Transcription factor-mediated reprogramming–models and mechanisms

1.1.2.1. Models of somatic cell reprogramming

Early after the generation of cloned animals by SCNT, and later the discovery of transcription factor-mediated induction of pluripotency, the *"elite"* model was proposed in order to explain the extremely low efficiency of cellular reprogramming [223]. According to this model, only a small number of cells with progenitor and/or stem cell-like qualities–the *"elites"*–within the somatic cell population have the capacity

for conversion to iPSCs, and the differentiated cells are resistant to reprogramming [223]. However, several lines of lineage tracing studies and clonal analyses ruled out the possibility of *"elite"* model by showing that all cells, including the terminally differentiated cells (e.g., T and B lymphocytes and albumin-expressing hepatocytes) have the potential to generate iPSCs [224, 225]. ***We now know that the low efficiency of cellular reprogramming is not related to the heterogenicity of the somatic cell population, but rather to the fact that most of cells never complete the path towards pluripotency, and only a small fraction of cells becomes iPSCs*** (***for more details regarding the molecular mechanisms of somatic cell reprogramming refer to*** [226]).

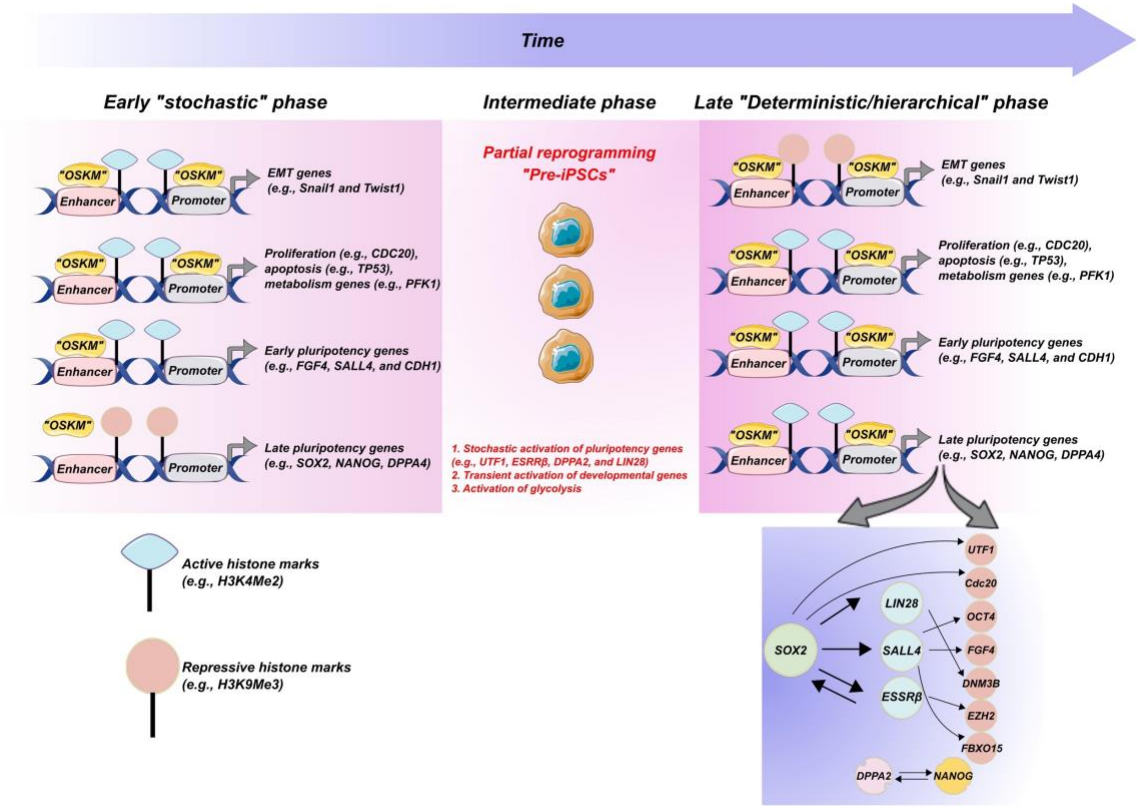


Figure 1. "Stepwise model" of somatic cell reprogramming. The schematic illustration puts the somatic cell reprogramming process in the so-called "stochastic-deterministic" model. Doing so, reprogramming will be divided into 3 phases, the early (initiation) phase, the intermediate phase, and the late (maturation) phase, as depicted. After the introduction of OSKM factors into a given somatic cell, cells undergo a stochastic gene activation and assume one of the many possible fates, such as apoptosis, transdifferentiation, transformation, senescence, and reprogramming. Among a mixed cell population, the reprogrammable cells will then increase their proliferation rate and undergo some epigenetic changes (e.g., enrichment of active histone marks, such as H3K4Me2) at the promoters/enhancers of somatic genes in order to facilitate the mesenchymal-to-epithelial (MET) transition, as well as DNA repair, RNA processing, and metabolic rewiring. In parallel to this, genes important for the somatic cell identity, as well as the epithelial-to-mesenchymal (EMT) process lose the active histone marks and start to gain the repressive modifications (e.g., H3K9Me3). These cells then enter an intermediate phase in which the stochastic gene activation could lead to the expression of some predictive markers, for example UTF1 and LIN28, as indicated. This will then instigate the initiation of

the deterministic phase through which the so-called "pre-iPSCs" stabilize into a mature pluripotent state. The gene expression in this phase is seemingly more hierarchical, starting with the activation of *SOX2*. The promoter/enhancer regions of early pluripotency genes in this phase contain high levels of H3K4Me2 and other active epigenetic modifications. Simultaneously, late pluripotency genes are switched on, which can possibly loosen the compacted chromatin through the removal of the repressive epigenetic marks and the enrichment of active histone marks.

As described earlier, a successful reprogramming process requires many steps, which occur either sequentially or in parallel. Some events, such as the mesenchymal to epithelial transition (MET), as well as suppression and activation of somatic genes and early pluripotency genes, respectively must occur during early phases, whereas other events, for example activation of late pluripotency-associated genes or repression of tissue-specific transcription factors and developmental genes occur during later stages of reprogramming. Starting from a mixed cell population in which every cell is permissive to OSKM factors, reprogramming can be induced and completed in either a *"stochastic"* or a *"deterministic"* fashion [223, 227]. According to the latter, all cells are reprogrammed with a so-called *"fixed latency"*, meaning that every cell follows the exact same path to reach pluripotency. According to the former, iPSCs are generated with *"variable latencies"*, which means that transition steps vary within the somatic cell population, and hence it is not possible to predict if the progeny of somatic cells becomes an iPSC, and if so when it will complete the process. In support of the stochastic model, it has been shown by single-cell cloning studies that sister cells from an early colony give rise to iPSCs with variable latency, with some sister cells never finishing the process. Although one cannot anticipate whether and when the reprogramming process is completed, activation of certain genes, for example *UTF1* (as abovementioned) can serve as a marker that rare early cells are on their way towards a pluripotent state, as it implicates that their promoters are available for OSKM. In contrast, pluripotency loci marked with H3K9me3 (e.g., *SOX2* and *NANOG*) are activated later in reprogramming and are not responsive to OSKM binding at early stages. Therefore, H3K9 methylation represents an epigenetic roadblock for *"pre-iPSCs"*, that has to be overcome in order to generate fully reprogrammed iPSCs (**also see Figure 1**).

Applying the "stochastic-deterministic" model to the process of somatic cell reprogramming, one may breakdown the pluripotency path into smaller steps; (i) early (initiation) phase, which is a long stochastic step of gene activation, (ii) intermediate phase where the partially reprogrammed cells–pre-iPSCs–are generated, and (iii) late (maturation and stabilization) phase, that is a shorter, hierarchical, and deterministic phase, starting with the activation of the SOX2 locus (***schematically illustrated in* Figure 1**)***.*** In the early phase of reprogramming, OSKM bind to genomic loci, including enhancers, then the promoters of somatic and pluripotency genes in order to repress and induce their expression, respectively. This phase appears to be stochastic and inefficient most probably due to the presence of many permissive and repressive histone marks as well as other epigenetic markers. In fact, any factor that can destabilize the compacted chromatin (typical of somatic cells) allows the initial OSKM-genome interaction [228]. Supporting evidences are several studies, showing that general DNA methyltransferases- and histone deacetylase inhibitors, such as 5-azacitidine

and valproic acid enhance reprogramming in combination with OSKM (***reviewed in detail in*** [229]). Next, the reprogrammable cells enter an intermediate phase wherein they undergo a stochastic activation of pluripotency markers, which in some cases can lead to the expression of predictive markers (e.g., UTF1 and LIN28, as mentioned earlier), driving the so-called *"second wave"* of reprogramming–a seemingly more predictable and hierarchical event–During this late phase, the expression of late pluripotency genes takes place through the interaction of OSKM with late pluripotency loci marked with methyl histones (e.g., H3K4, as abovementioned). This initiates with the activation of SOX2–one of a group of pluripotency initiating factors–and is followed by the hierarchical activation of a series of genes, which eventually allows cells to stabilize into the pluripotent state. Consistent with this *"stepwise model"* of reprogramming, it has been suggested that transcription factors other than OSKM are sufficient for generating iPSCs due to the hierarchical network of gene activation. In light of this, it has been reported that a combination of OSKM downstream targets; ESRRβ, LIN28, DPPA2, and SALL4 is able of inducing iPSCs from MEFs [227].

1.1.2.2. Markers of somatic cell reprogramming

When discussing reprogramming markers, it is very important to keep in mind that due to the stochastic nature of somatic cell reprogramming, no set of molecular markers is able to predict whether a given cell early in the reprogramming path will eventually produce a pluripotent progeny or not. Another important concern in this regard is the ability to differentiate the partially-reprogrammed cells from established iPSC. For example, although activation of some genes, such as stage-specific embryonic antigen 1 (SSEA) and/or alkaline phosphatase is considered to be an essential step during the reprogramming process, such markers are not necessarily confined to the cells destined to become fully-reprogrammed iPSCs (***reviewed in*** [211]).

Despite the abovementioned issues, global gene expression- and proteomics analyses of enriched population of cells at different phases of reprogramming (**see Figure 1**) suggest a set of transcription factors as the representative of each stage [230-233]. In light of this, it has been shown that (i) *Fbxo15*, *FGF4*, *SALL1*, fucosyltransferase 9 (*Fut9*), chromodomain helicase DNA binding protein 7 (*Chd7*) and E-cadherin (*Cdh1*) are indicative of the initiation phase of reprogramming, (ii) genes coding for NANOG, OCT4, SOX2, SALL4, NODAL, ESRRβ, and ERAS are normally activated during the intermediate phase where the so-called "pre-iPSCs" are generated, and (iii) genes including *REX1*, *GDF3*, *DPPA2*, *DPPA3*, and *UTF1* most likely mark the late- or maturation phase (**also refer to Figure1**). Importantly, there are some limitations associated with the information obtained from gene expression/proteomics analysis of heterogenous populations owing to the small number of cells that are destined to become full iPSCs. In this regard, technological advancements, such as single-cell expression analysis have redefined our understanding of the mechanisms- and markers of reprogramming. An example is the single-cell analysis of SSEA1-positive pre-iPSCs, which identified early- intermediate- and late reprogramming markers, including epithelial cell adhesion marker (EPCAM), representative of the early phase; KIT receptor, representative of the intermediate phase; and platelet endothelial cell adhesion molecule

(PECAM1), indicative of the late stabilization phase. Consistently, sorting SSEA1- and EPCAM-positive early cells showed a slight increase in the reprogramming efficiency [232]. Using two complementary single-cell approaches, Buganim *et al.* identified *UTF1*, *DPAA2*, *LIN28*, and *ESRRβ* as the ***"predictive genes"*** that are expressed in a small subpopulation of cells early in the process, marking those that are meant to become iPSCs (**also refer to Figure 1**) [227]. Taken together, the critical endogenous reprogramming factors (e.g., OCT4 and SALL4), which are conventionally used as reprogramming markers appear to be poor predictors of iPSC generation. What supports this idea the most is that although these markers are expressed only in rare cells early in the process, they are as well activated in partially-reprogrammed cells.

1.1.2.3. Mechanisms of somatic cell reprogramming–role of the OSKM factors, levels, and stoichiometry

How ectopic expression of the OSKM combination drives the pluripotency path is not thoroughly understood. Putting the process of reprogramming in the *"two-step"* model (as described above), it has been shown that the first transcriptional wave is mainly mediated by MYC, which occurs in all transduced cells, while the second wave is limited to reprogrammable cells and is induced by a slow increase in OCT3/4 and SOX2 target gene expression, leading to the activation of other pluripotency genes. Finally, KLF4 is thought to support both phases through repressing somatic genes in the early phase and enhancing the expression of pluripotency genes during the late phase (***reviewed in*** [211]). As abovementioned, immediately upon OSKM transduction, these factors occupy open and accessible chromatin, binding promoters of genes marked with activating- or repressing histone modifications. In this regard, it has been reported that MYC binding–significantly more than the rest of the factors–is associated with an active epigenetic state of the genomic loci. An example is the occupation of methylated H3K4- or acetylated H3K27-marked loci (a sign of open chromatin) by MYC for repressing somatic genes (***the role of MYC in transcription regulation was revisited as a global gene amplifier in*** [234, 235]). This finding could let us assume that MYC does not initiate the reprogramming process, but rather potentiates the action of the other three factors. In addition to active chromatin being a pre-requisite for MYC binding, there are other lines of evidence, which support the above hypothesis in differentiating between the role of "OSK" and "MYC" in reprogramming. In view of this, it has been demonstrated that the OSK factors bind multiple distal sites of many genes with no or minimal pre-existing epigenetic modifications, which shows that these initially OSKM-bound sites are distinct from those occupied by OSK in ESCs. On this basis, it has been suggested that "OSK" act as *"pioneering reprogramming factors"*, meaning that they are able to bind closed chromatin on their own without the help of other transcription factors. They will then open the inactive chromatin regions for the activation of pluripotency genes. MYC, on the other hand, only facilitates this process, as it largely relies on pre-existing chromatin modifications (***for more details refer to*** [228], ***which raises the notion of OSK proteins being the pioneer factors of reprogramming through promiscuous binding to distal elements***).

The comparison of two doxycycline-inducible transgenic reprogrammable mice strains in their ability to generate all-iPSC mice was one of the early evidences, showing a crucial role for the appropriate stoichiometry of OSKM factors in the induction of pluripotency [236, 237]. This idea was supported by the fact that the only difference between the two transgenic systems was a different factor stoichiometry, with the high quality iPSCs being associated with the strain producing 10- to 20-fold higher protein levels of OCT3/4 and KLF4 and lower levels of SOX2 and MYC. This was further confirmed by two other studies, that showed high and low OCT3/4 and SOX2 expression corresponds to the highest reprogramming efficiency, respectively (***reviewed in*** [211]).

In addition to factor stoichiometry, the expression levels of the OSKM transgenes influence the success rate of reprogramming. In this regard, continuous OSKM expression has been reported to be essential for iPSC generation until the intermediate phase of partially reprogrammed cells with unstable fates has been passed. Therefore, disrupting OSKM expression before this stage (~8-12 days after OSKM transduction in mouse cells) leads to a failure in iPSC generation (***reviewed in*** [197]). Another example of the different transgene expression between intermediate- and fully-reprogrammed iPSCs is the OSKM-genome interaction. With regards to this, it was shown that partially-reprogrammed iPSCs express a unique set of genes bound with more reprogramming factors compared with that of ESCs, for example promoter or enhancer genes bound with only OCT3/4 and SOX2 in ESCs are bound with OSKM in the intermediate state. In contrast, genes that are highly expressed in ESCs were found to be bound with fewer reprogramming factors in the intermediate cells [238]. Another parameter which contributes to the generation of fully-reprogrammed cells is the timing of factor expression. As an example, during the early phase, the *"OCT3/4^{high}/SOX2^{low} stoichiometry"* leads to the activation and suppression of the mesodermal- and ectodermal genes, respectively. This expression pattern appears to be also important in the late phase wherein the endogenous factors are expressed in a similar manner, with OCT4 levels being high, whereas SOX2 levels remaining low. Therefore, the *"OCT3/4^{high}/SOX2^{low}"* stoichiometry determines mesodermal features, which is essential in the establishment and development of intermediate cells on the path to fully-reprogrammed iPSCs [197, 239]. ***These observations are the basis of the so-called "seesaw model" of reprogramming, which emphasizes on the importance of a balanced expression of lineage-specifying factors for a successful reprogramming route.***

1.1.2.4. Mechanisms of somatic cell reprogramming–signaling networks

Multiple signaling pathways, including TGFβ, PI3K, MAPK, and Wnt have been shown to come together in a finely-balanced manner in order to induce and maintain the stem cell identity during the process of reprogramming. From a simplistic viewpoint, these signaling networks must interact and crosstalk to sustain a long-term proliferative state, on the one hand, and to repress differentiation, on the other hand (***signaling networks involved in somatic cell reprogramming as well as the small molecules targeting such pathways are summarized in two excellent reviews;*** [229, 240] ***and*** **Figure 2**).

The transition from using MEF as a feeder-layer for culturing PSCs to feeder-free culture systems represents a turning point in our understanding of self-renewal signaling. The latter approach utilizes Matrigel or laminin as attachment matrices and substitutes the MEF feeder layer with MEF-conditioned media, fetal calf serum or knockout serum replacement supplement (KSR), and FGF2 for maintenance [241]. Early on, it was reported that Activin A–from the TGFβ superfamily of cytokines–is secreted by MEF feeder layers and that its addition to hPSC culture media maintains their pluripotency for more than 20 passages [242]. Later, it was shown that Activin A could be replaced with other TGFβ family members, including TGFβ and Nodal, which signal through the Smad2/3 pathway, targeting pluripotency genes, such as *NANOG* [243]. The role of PI3K/Akt signaling in hPSC culturing came to our understanding immediately upon using fetal calf serum or KSR supplement in feeder-free formulations, as all these components contain factors, such as insulin or IGF, which are known to be a potential activator of the canonical PI3K pathway [244, 245]. Other studies have suggested a critical role for the canonical Wnt signaling in pluripotency [246]. ***Altogether, FGF2 through Erk MAPK, Activin A through Smad2/3, insulin/IGF through PI3K, as well as canonical Wnt signaling form the basis of our current understanding of self-renewal signaling. Importantly, what establishes a stable pluripotent state is the crosstalk of these signaling pathways, which will be described in detail in the following sections.***

FGF2 and MAPK/Erk signaling. As abovementioned, under feeder-free culture systems, FGF2 is used in order to supplement the activities of the feeder layer in MEF conditioned media. Initially, it was thought that FGF2 functions through the MAPK/Erk pathway, which we now know has a more complicated scenario. Several lines of studies have shown conflicting roles for the Erk pathway in hPSCs, which could be potentially reconciled (***reviewed in*** [240]). In fact, a minimal level of Erk is required for cell cycle progression, cell proliferation, and inhibition of apoptosis. It has been shown that above a certain threshold, Erk signaling represses pluripotency genes and activates developmental genes, and hence an intermediate level of Erk activation is optimal for maintaining pluripotency. With regards to its relationship with FGF2, it has been reported that both high- and low concentrations of the cytokine are capable of sustaining the basal Erk activity–compatible with pluripotency–however through distinct mechanisms. Low levels (<10 ng/mL) of FGF2 support self-renewal through a mild activation of Erk, while higher levels (>50 ng/mL) activate the PI3K pathway in parallel to Erk induction, which in turn suppresses MAPK/Erk signaling via a crosstalk mechanism [247]. This explains how Erk activity is adjusted at an optimal level under a wide range of FGF2 concentrations. Despite these observations, more recent studies have questioned the absolute necessity of FGF2 for reprogramming, as hPSCs could be generated in its absence [247]. The reason might be due to the secretion of FGF2 from hPSCs themselves, which leads to the activation of Erk signaling through an autocrine/paracrine signaling.

Activin A and canonical TGFβ signaling. Activin A represents the first bioactive molecule derived from MEF feeder layers, which is essential for hPSC maintenance, as described earlier. In addition, new generations of culture methods, such as most of the chemically-defined media, also use Activin A or

other TGFβ family members to support hPSC generation and culturing [240]. In this context, Activin A is assumed to signal through the canonical TGFβ pathway–Smad2/3–which then activates pluripotency-associated genes, with *NANOG*, being the most well-known target [243].

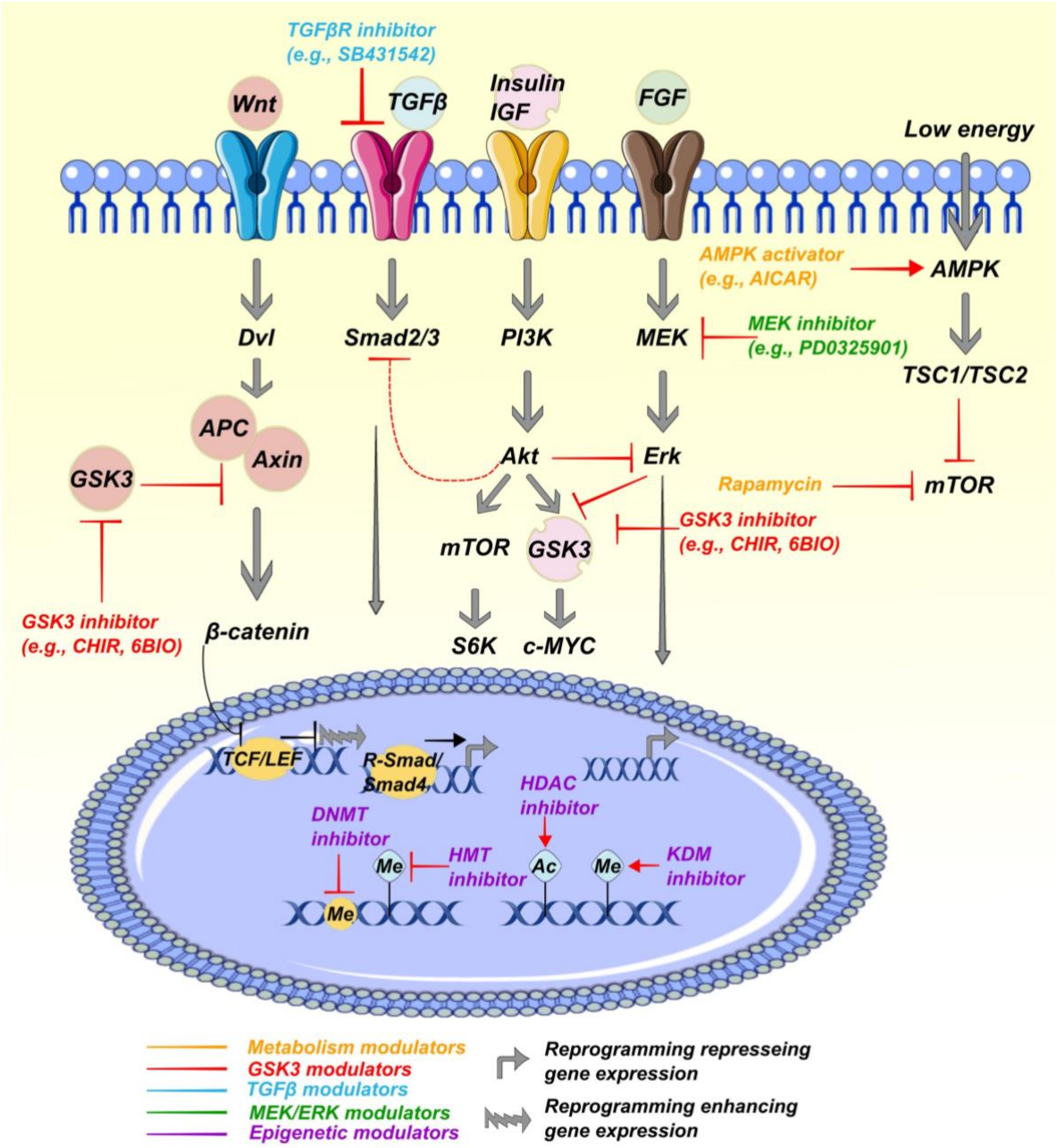


Figure 2. Schematic summary of key signaling pathways and epigenetic modifications associated with somatic cell reprogramming and their availability for chemical modulation. The schematic illustration summarizes the reprogramming-related signaling networks; TGFβ- Wnt- PI3K- MAPKs- and AMPK, as well as the crosstalk between the first four pathways. In addition, the key epigenetic modifications; DNA methylation and histone methylation/acetylation are shown. A color code has been used for discrimination between small molecules targeting different pathways and epigenetic changes, as indicated. **For detailed information refer to section 1.1.2.4.**

Despite this, a major concern regarding the role of TGFβ/Smad signaling in hPSC maintenance was long-existed due to its opposing functions in pluripotency and early differentiation [244]. This discrepancy has been resolved through a crosstalk mechanism between Smad2/3 and PI3K/Akt, which controls the balance between self-renewal and differentiation [247]. Simply, when PI3K/Akt is active, it establishes conditions in which Activin A/Smad2/3 has a pro-self-renewal function by activating target genes, such as *NANOG*. On the other hand, when PI3K/Akt activity is low, Wnt signaling is activated and works together with Smad2/3 to promote differentiation [247].

PI3K/Akt signaling. How PI3K/Akt regulates hPSC maintenance could be easily clarified by understanding its central role in an elaborate crosstalk signaling network between (i) PI3K/Akt, (ii) Raf/MEK/Erk, and (iii) the canonical Wnt pathway, which converges on Smad2/3. Besides impacting the level of Smad2/3 activation (as pointed out above), PI3K/Akt inhibits Erk, which in turn suppresses GSK3, and by doing so the PI3K/Akt pathway reaches over to Wnt signaling [247]. In addition to this, GSK3 activity could be directly inhibited through Akt-mediated phosphorylation (at Ser9), an effect which in this context leads to the stabilization of MYC, a key regulator of PSC generation and maintenance. In light of this, it has been reported that small molecule inhibition of PI3K (e.g., by LY294002) or removal of PI3K activators, as well as the inhibition of mTOR–a downstream kinase of Akt–by rapamycin leads to the loss of pluripotency and triggers differentiation [244, 248]. Interestingly, several evidences have shown that usually more than one PI3K activator is required for maintaining PSCs. For example, StemPro-based media use both IGF and the EGF family member, heregulin, which both signal through the PI3K pathway, but one is insufficient in the absence of the other [240]. The immediate explanation could be that these factors have functions outside of PI3K/Akt signaling, however, another possibility would be that different ligands signal with different temporal kinetics, such as duration of the signaling and changes in the signal intensity overtime. Detailed time-course experiments regarding the effects of the individual factors on PI3K signaling are therefore necessary to address this question.

Wnt signaling. The role of the Wnt pathway as a pro-self-renewal signaling has always been controversial, as there is little evidence, supporting the importance of the pathway in the induction of pluripotency in human, and more drastically its positive impact on promoting differentiation [247, 249]. The latter is also consistent with numerous lines of *in vivo* studies, showing that Wnt is required for germ layer formation during gastrulation. This confusion appears to have been generated by overlooking different pools of GSK3 [250], and hence their different sensitivities to small molecule inhibitors frequently used to improve the reprogramming process. For example, blocking GSK3 activity using the indirubin derivative, BIO has been shown to maintain human- and mouse pluripotent state, and this has been interpreted as Wnt being a positive regulator of pluripotency [246]. However, another way to interpret this would be that inhibition of the GSK3 pool downstream of Akt (not that of Wnt signaling) is required for stem cell maintenance, which involves stabilization of several transcription factors, such as MYC (as mentioned above). In view of these observations, it has been reported that when all pools of GSK3 are inhibited, Wnt signaling is activated through β-catenin, leading to differentiation of hPSCs

[247, 251]. ***In summary, it seems that different pools of GSK3 have distinct roles in determining PSCs' cellular fate, and the general belief is that suppression of the canonical Wnt pathway stabilizes hPSCs.*** In support of this, OCT4 has been shown to repress β-catenin, promoting self-renewal and pluripotency [249].

1.2. Pharmacological reprogramming towards pluripotency

Reprogramming can be achieved by genetic means, such as forced expression of pluripotency transcription factors (**as described in depth in section 1.1.**), which could be further enhanced or even avoided using pharmacological means, such as small molecules and growth factors. Chemical reprogramming becomes even more desirable when considering the slow process of transcription factor-mediated reprogramming as well as its extremely low efficiency. From a therapeutic perspective and in the context of regenerative medicine, pharmacological methods of reprogramming allow a safer approach–compared with viral transduction–to generate patient's own iPSCs, which can replicate indefinitely and differentiate into any somatic cell type, giving rise to unlimited number of therapeutic cells for cell transplantation.

1.2.1. Pharmacological reprogramming–small molecule epigenetic modulators

Gene expression is largely dependent on epigenetic regulatory mechanisms, involving chemical modifications of chromatin, such as DNA methylation as well as histone methylation and -acetylation (***reviewed in*** [212]). Therefore, small molecules that modulate the activity of the enzymes responsible for these modifications may change the chromatin organization in a manner that it would be suitable for transcription of genes required for cellular reprogramming. In light of this, chemical epigenetic modifiers have been reported to enhance somatic cell reprogramming towards pluripotency. For example, BIX-01294, an inhibitor of histone methyl transferase–G9a–has been shown to induce the generation of mouse iPSCs from neural progenitor cells when combined with the ectopic expression of "SKM" pluripotency factors [252]. Another example is the generation of iPSCs from MEF using a combined genetic and chemical approach by means of the DNA methyl transferase (DNMT) inhibitor, RG108 and BIX-01294 in cells only expressing two of the core pluripotency factors, OCT4 and KLF4 [253].

In addition to RG108, which is a non-nucleoside DNMT inhibitor, 5-azacitidine (known as 5-aza, **also see section 1.1.2.1.**), as well as its related analogue, 5-aza-2'-deoxycitidine (5-aza-dc) represent two examples from the family of nucleoside DNMT inhibitors and have been shown to enhance OSKM-mediated cell reprogramming (***reviewed in*** [229]). Additionally, 5-aza was shown to be able to complete the reprogramming process of partially-reprogrammed cells (pre-iPSCs) [254]. 5-aza-dc has the advantage of incorporating exclusively into DNA, compared with 5-aza, which can incorporate in both DNA and RNA when used at high doses. This allows the use of lower concentrations of 5-aza-dc for inhibition of DNMTs, thus reduces its toxicity [255]. Despite this, both analogues share a similar mechanism by incorporating into DNA, which can induce DNA damage. Therefore, further investigation

is required to better determine their pharmacological activity and the optimal concentration for safer and more efficient generation of iPSCs.

1.2.2. Pharmacological reprogramming–small molecule modulators of signaling networks

Modulating signaling pathways involved in reprogramming (**see section 1.1.2.4.**) provides another approach to rewire the gene transcription in favor of pluripotency. In view of this, it has been reported that dual inhibition of TGFβ- and MEK/Erk signaling using SB431542 and PD0325901, respectively is able to induce pluripotency by promoting the process of MET. Other MET-relevant chemicals include A-83-01 and RepSox, both of which are known to blunt the TGFβ pathway [229]. The use of SB431542 and RepSox has been found to substitute SOX2 in the process of iPSCs generation [256], and A-83-01 has been used in chemical cocktails, which mediate cellular reprogramming in combination with *OCT4* ectopic expression, replacing the other three reprogramming factors [257].

GSK3 inhibitors have been repeatedly shown to enhance reprogramming. A good example is the role of CHIR99021, a small molecule inhibitor of GSK3 and activator of Wnt, in enhancing reprogramming. The proposed mechanism is that activation of Wnt stabilizes β-catenin, which then transfers to the nucleus and exploits homeodomain-interacting protein kinase 2 to phosphorylate TCF3. By doing so, the repressive effect of TCF3 on the promoter regions of reprogramming factors is relieved, facilitating the reprogramming process [258]. Additionally, CHIR99021, in combination with other small molecules, has been shown to replace SOX2 and MYC in inducing ground state pluripotency of mouse brain-derived neural stem cells [259], as well as human keratinocytes [260].

In addition to the MEK/Erk- TGFβ- and GSK3 small molecule modulators, activation of cAMP signaling has been long-established for iPSC generation. In light of this, 8-bromo-cAMP, a cell-permeable analogue of cAMP, was shown to increase the efficiency of OSKM-mediated reprogramming of human neonatal foreskin fibroblasts by 2-fold, and its combination with valproic acid was found to synergistically increase the efficiency by 6.5-fold [261]. Another interesting example of cAMP signaling modulators is forskolin–an adenylyl cyclase activator–which was found to functionally replace OCT4 for the generation of iPSCs from MEF when combined with the other three transcription factors, "SKM" (***the first example of chemically-generated miPSCs;*** [262]). Further investigations suggested that the downstream exchange protein directly activated by cAMP (EPAC)–rather than protein kinase A (PKA) signaling–is necessary and sufficient for lowering the reprogramming barrier through enhancing MET and proliferation [263].

1.2.3. Pharmacological reprogramming–small molecule modulators of cell metabolism

During reprogramming, a major metabolic switch from mitochondrial oxidative phosphorylation towards glycolysis is required (***for more details on the metabolic regulation during reprogramming refer to*** [264]). The early evidence with regards to metabolic modulation in cellular reprogramming came from the use of a small molecule activator of pyruvate dehydrogenase kinase 1 (PDK1), PS48, for

generating human iPSCs in combination with *OCT4* expression, as well as other epigenetic- and signaling modulators [257]. Consistent with this finding, it was reported that promoting glycolytic metabolism using various small molecules, such as fructose 2,6-bisphosphate (through activating phosphofructokinase 1) and quercetin (through promoting HIF1 activity) enhances reprogramming. In contrast, inhibition of glycolytic metabolism blocks reprogramming; the examples are the effects of 2-deoxyglucose (2DG, a glycolysis inhibitor), 3-bromopyruvic acid (a hexokinase II inhibitor), as well as dichloroacetate (a PDK inhibitor) in blunting the reprogramming process [265].

The necessity of glycolytic metabolism for iPSC generation was further confirmed by showing a negative role for rapamycin (a mechanistic target of rapamycin (mTOR) inhibitor) and a positive role for AICAR (a small molecule activator of AMP-activated protein kinase (AMPK)) in OSKM-mediated iPSC generation [266]. Importantly, this study shed new light on the mechanism of metabolic reprogramming associated with the induction of pluripotency, as it suggested that the noncanonical autophagy pathway (ATG5-independent) plays a crucial role in mitochondrial clearance, harboring the oxidative phosphorylation machinery [266]. ***This finding is particularly interesting, as it shows that iPSC generation and tumorigenesis share a similar metabolic switch, which appears to be accessible for small molecule modulation.***

1.2.4. Pharmacological reprogramming–combined chemical treatments

The chemical approaches described in the previous sections demonstrate that multiple mechanisms, such as epigenetic- and signaling pathways could be targeted concomitantly using certain small molecules or their combination in order to compensate for the absence of most pluripotency transcription factors. In this regard, a combination of A83-01 (targeting TGFβ signaling) and AMI-5 (a protein arginine methyltransferase inhibitor) was shown to enable OCT4-induced reprogramming of MEF to iPSCs [267]. In addition, AMI-5 has been used in combination with the MEK/Erk inhibitor (PD0325901) to induce pluripotency in human keratinocytes with ectopic expression of *OCT4* and *KLF4*. This was further improved when adding PS48 (**also see section 1.2.3.**) and sodium butyrate (NaB; an inhibitor of histone deacetylase) to the cocktail, which enabled OCT4-only induction of pluripotency [268]. Adding the GSK3 inhibitor, CHIR99021 (**also see section 1.2.2.**) and parnate, which is a small molecule inhibitor of lysine-specific demethylase 1 (LSD1) to this cocktail was found to further improve the reprogramming of human keratinocytes, which are known to be resistant to reprogramming. On the basis of these findings, an optimized protocol has been developed using the combination of these six small molecules with forced expression of *OCT4* alone for the generation of iPSCs from human epidermal keratinocytes (***reviewed in*** [269]). ***Altogether, it seems that screening small molecule enhancers of reprogramming not only provides us with new tools for generating iPSCs, but also unravels previously unknown mechanisms underlying the reprogramming process.***

1.3. Project narrative–rationale, hypothesis, and purpose

Several lines of preclinical- and clinical evidence have pinpointed the need of safer and more efficient approaches for the generation- and maintenance of iPSCs. In light of this, pharmacological reprogramming using chemical enhancers of pluripotency appears to offer an alternative route, which can complement, reduce, or even avoid the use of genetic tools for example, delivery of exogenous transcription factors in somatic cells. The ability to generate such chemically induced pluripotent stem cells (referred to as ciPSCs) may provide us with a safer reprogramming method for various applications, including iPSC-based cell therapies as well as drug development in the field of regenerative medicine.

One way to identify small molecules that are able to facilitate reprogramming is functional screen of known drugs, i.e., searching for chemicals sharing similar backbones with approved drugs in order to have advantages in terms of pharmaco-kinetic and -dynamic properties. We therefore adopted this strategy with the aim of **(i) finding new candidates capable of inducing pluripotency-associated markers.** The focus was particularly drawn to one of the *"master genes"* of reprogramming, *OCT4*, as it has been frequently shown to be irreplaceable when using chemical cocktails for the induction of pluripotency in human somatic cells (**see section 1.2.**). In view of this, mouse fibroblasts have been successfully reprogrammed to iPSCs by a combination of seven small molecule compounds, showing that the master pluripotency genes are dispensable for the generation of mouse iPSCs. Despite this breakthrough and as mentioned earlier, small molecule cocktails for the generation of human chemically-induced iPSCs are yet to be developed mostly due to the lack of a sufficiently strong inducer of OCT4. Having this in mind, we aimed at **(ii) optimizing a chemical condition using our small molecule hits from the screen in order to substitute OCT4 in the context of OSKM-meditated reprogramming and referred to them as OCT4 inducing compounds (O4I).**

Reprogramming is a complexed process involving many genetic- and epigenetic alterations and intricate crosstalks between multiple cellular signaling pathways (as described in depth, previously). After more than a decade of great advances in the field of transcription factor-mediated reprogramming, molecular mechanisms underlying this process remain to be thoroughly understood. Upon the identification of new small molecules able to enhance OSKM-based reprogramming, we aimed at **(iii) exploring the mode of action of O4I compounds employing different methods, ranging from DNA microarray to RNA-sequencing and ATAC-sequencing.** Our findings not only led to the development of new chemical tools for the induction of pluripotency in human somatic cells, but also shed new light on the molecular events during reprogramming as well as their availability for small molecule intervention.

2. RESULTS

2.1. Cell-based high throughput screening identifies new small molecules able to induce pluripotency-associated markers.

Looking for potential small molecule inducers of pluripotency, we sought to perform a high throughput screen (HTS) from a chemical library containing ~250,000 compounds. The screen was carried out in HEK-293 cells stably transfected with a reporter construct in which the expression of luciferase is under the control of human *OCT3/4* promoter (**Figure 3**). Doing so, we identified several lead structures showing potent activity in enforcing *OCT4* expression, among others, are 2-[4-[(4-methoxyphenyl)methoxy]phenyl]acetonitrile (referred to as O4I1) and 2-((4-chlorophenyl)amino)-thiazole-4-carboxylate (referred to as O4I2), as previously described [270] (**Figure 3, also see Figure 4A and 4B for the structures of O4I2 and its derivatives**).

2.1.1. The metabolically-stable O4I2 derivatives are potent OCT4 inducers.

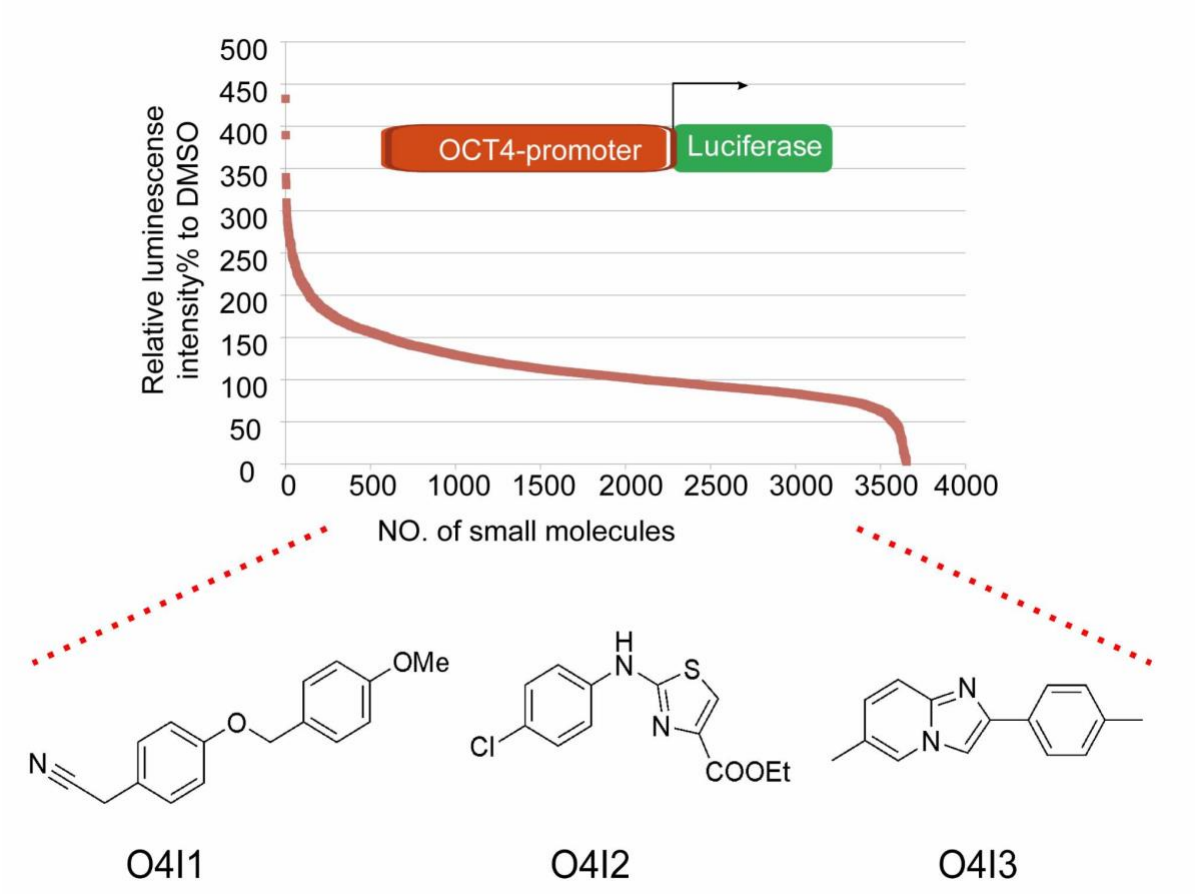


Figure 3. Cell-based high throughput screening identifies a series of OCT4-inducing compounds (O4Is). Top; HEK-293-OCT4 reporter cell line in which luciferase expression is controlled by the promoter activity of the master pluripotency gene, *OCT4* was used to screen a chemical library containing ~250,000 small molecules. Bottom; chemical structures of the hits (herein O4I1/2/3) obtained from the screen.

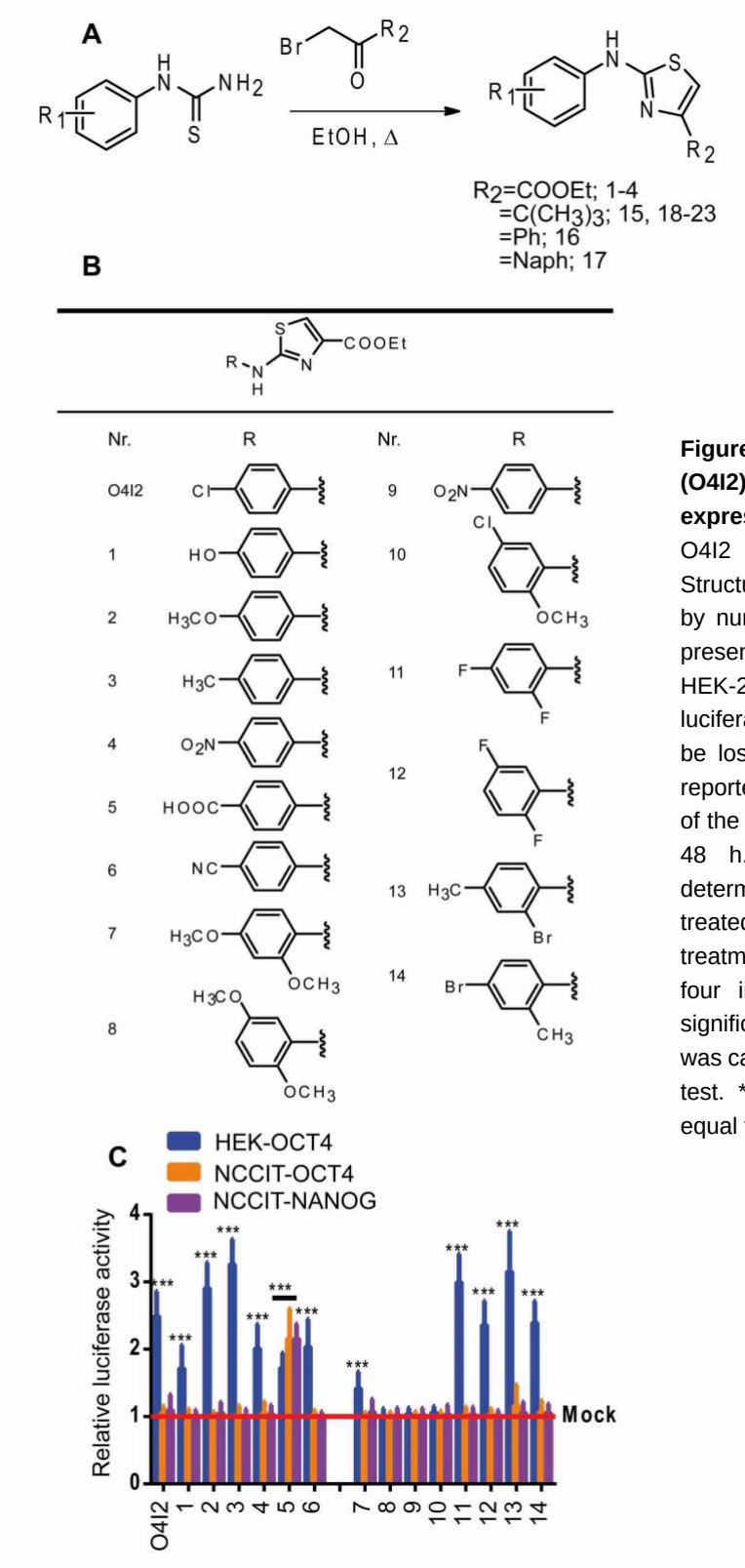


Figure 4. OCT4-inducing compound 2 (O4I2) and its derivatives induce OCT4 expression. (A) Synthesis procedure of O4I2 and its carboxylic analogues. **(B)** Structures of the O4I2 derivatives, indicated by numbers. **(C)** O4I2 and the derivatives presented in **(B)** induce OCT4 expression in HEK-293 reporter cells, as determined by luciferase activity. This effect was found to be lost in the NCCIT-OCT4 and -NANOG reporter cell lines. A concentration of 20 μM of the chemicals was applied to the cells for 48 h. Relative luciferase activity was determined by normalizing the values of the treated cells to mock (0.2% DMSO) treatment. Data present the mean ± SD of four independent experiments. Statistical significance between mock and treated cells was calculated using a two-tailed student's *t*-test. *** indicates a *p*-value less than or equal to 0.001.

In addition to HEK-OCT4 cells, we sought to expand our readout by including two additional reporter cell lines, human embryonic carcinoma cells stably expressing luciferase under the control of promoter regions of either *OCT4* or *NANOG* (designated as NCCIT-OCT4 and NCCIT-NANOG, respectively). Almost all analogues were found to activate OCT4 in HEK-OCT4 reporter cells, while they failed to do the same in NCCIT cells (**Figure 4C**).

We have previously reported that O4I2 converts to 2-(4-chlorophenylamino)thiazole-4-carboxylic acid (O4I2-COOH), a metabolic product that was shown to be inactive in HEK-OCT4 cells [270]. This prompted us to hypothesize that acid hydrolysis, for example via the action of carboxylesterase 1 (CES1) enzyme, which is responsible for ester-to-acid conversion contributes to the inactivation of O4I2 derivatives in NCCIT cells. In line with our hypothesis, we observed a significantly higher expression of CES1 in NCCIT cells and PSCs when comparing the expression levels of CES1-5 isozymes among selected cell lines, using previously published gene expression data [271-273] (**Figure 5A**). We then replaced the ethyl ester with CES1-resistant groups (**Figure 5B**), recovering the activity of the derivatives in the NCCIT reporter cells (**Figure 5C**). Based on these results, we identified the most potent compound (**No. 23**), and designated it as O4I4.

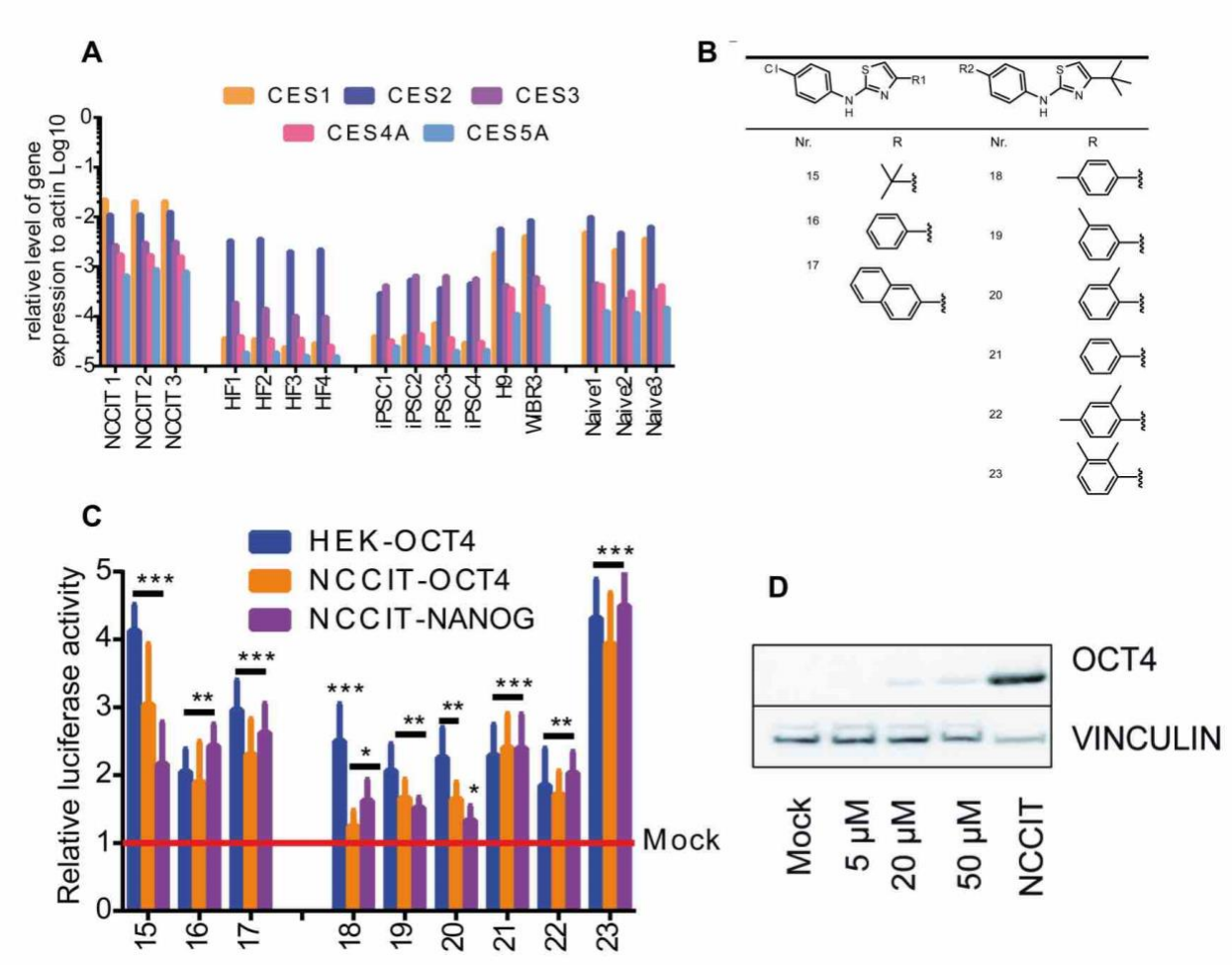


Figure 5. Metabolic stability is necessary for OCT4-inducing ability of O4I2 and its derivatives. **(A)** Re-analysis of previously published gene expression data [271-273] for comparing the expression levels of different isozymes of carboxylesterase (CES) in NCCIT cells, human fibroblasts (HFs), iPSCs, as well as human pluripotent naive stem cells. **(B)** Structures of the non-carboxylic derivatives of O4I2, indicated by numbers. **(C)** OCT4-inducing activity of the compounds presented in **(B)** was measured in the HEK-293 and NCCIT reporter cell lines treated with 20 μM of the small molecules for 48 h. Relative luciferase activity was calculated respective to mock (0.2% DMSO) treatment. **(D)** A 5-day treatment with the most potent non-carboxylic analogue of O4I2 (No. 23, designated as O4I4) induces OCT4 protein levels in HFs concentration-dependently, as determined by immunoblotting. NCCIT cells were included as reference. Vinculin served as the loading control. Data in **(C)** present mean ± SD of four independent

experiments. Statistical significance was calculated between mock and treated cells using a two-tailed student's *t*-test. *, **, and *** are indicative of *p*-values less than or equal to 0.05, 0.01, 0.001, respectively.

2.1.2. Imidazopyridine derivatives are potent OCT4 inducers.

As our next step, we intended to search for small molecules with similar structures to those of approved drugs, a strategy which could ultimately lead to *"drug repurposing"*. This approach offers various advantages for example, in terms of safety and time over developing an entirely new molecule for a given application [274]. In light of this, our initial screen identified more than 4000 imidazopyridine derivatives sharing a similar backbone with the FDA-approved sedative-hypnotic drug, zolpidem (**Figure 3B**). These analogues showed significantly higher luciferase activity in the HEK-OCT4 reporter cell line, as compared to that of the non-treated cells (**Figure 3A**). In order to study this class of compounds in depth, we synthesized the lead structure, 6-methyl-2-(p-tolyl)imidazo[1,2-a]pyridin by heating ω-bromoacetophenone with an excess of 2-aminopyridine for 3 h to achieve a yield of >95%, thereafter referred to the compound as O4I3 (**Figure 6A**).

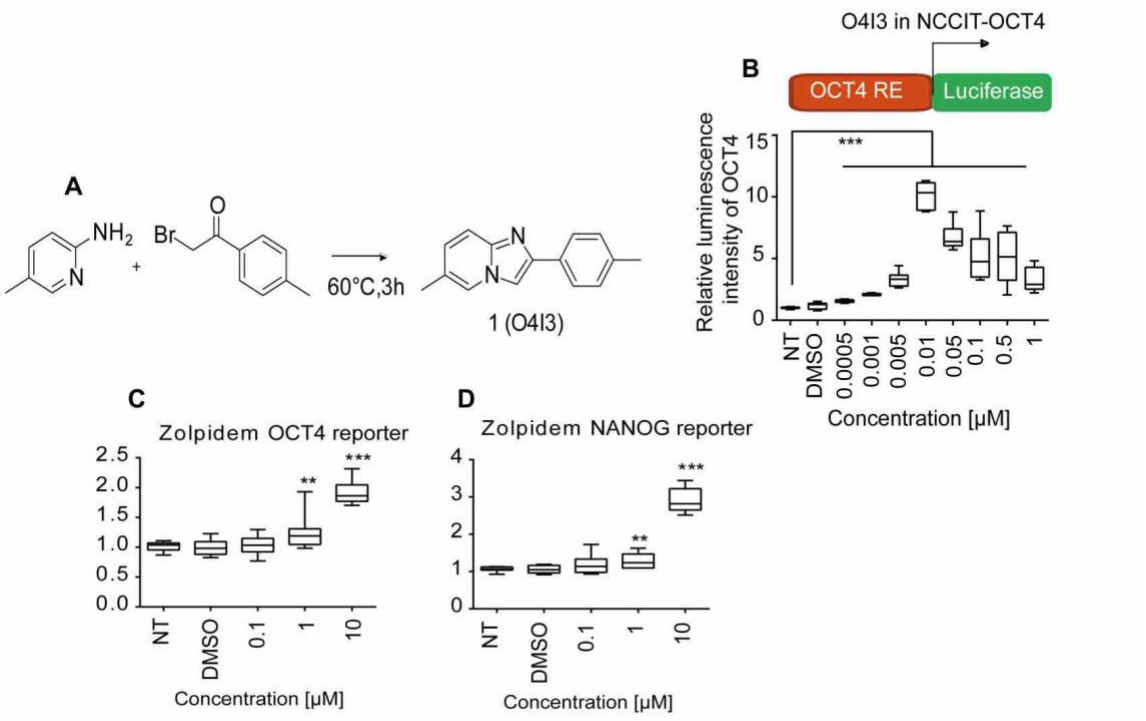
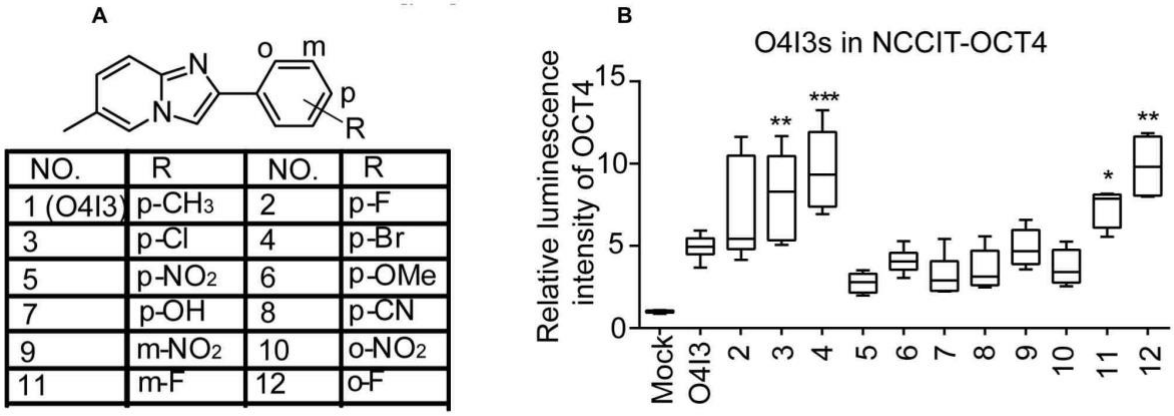


Figure 6. The imidazopyridine analogues, O4I3 and zolpidem, show potent OCT4-inducing ability. **(A)** The synthesis procedure of O4I3, the lead structure of imidazopyridine derivatives identified by the cell-based high throughput screening. **(B)** 48 h treatment with various concentrations of O4I3 induces a significant increase in OCT4 expression in NCCIT reporter cells. **(C)** and **(D)** Similar to O4I3, zolpidem–also carrying the imidazopyridine backbone–was found to activate the pluripotency-associated genes, *OCT4* and *NANOG* in NCCIT reporter cells at 48 h post-treatment with the indicated concentrations. *OCT4* promoter activation was measured by luciferase assay. 0.1% DMSO was used as mock treatment. Data in **(B-D)** present mean ± SD of at least three independent experiments. Statistical significance was compared to non-treated (NT) cells and mock using a two-tailed student's *t*-test. ** and *** are representative of a *p*-value less than or equal to 0.01 and 0.001, respectively.

In addition to HEK-293, we tested O4I3 in NCCIT-OCT4 reporter cells and found ~10-fold induction in the expression of this master pluripotency gene upon treatment with various concentrations of O4I3 (**Figure 6B**). To our expectation, zolpidem, which carries the imidazopyridine backbone showed an increase in *OCT4* and *NANOG* expression, although to a lower extent (~2-fold) and at higher concentrations (10 μM) compared with O4I3 (**Figure 6C and 6D**).

A major concern of large screens is pan-assay interference compounds (PAINS), which refer to structures from several classes of compounds with non-drug-like mechanisms (***discussed by Baell J. and Walters M.A. in*** [275]). To avoid this, O4I3 was chemically expanded by replacing the methyl group at the phenyl moiety with several other substituents on the aryl moiety (**Figure 7A**). These analogues were then tested in NCCIT-OCT4 reporter cells and were found to activate OCT4 expression in a structure-dependent manner (**Figure 7B**).



NO.	R	NO.	R
1 (O4I3)	p-CH_3	2	p-F
3	p-Cl	4	p-Br
5	p-NO_2	6	p-OMe
7	p-OH	8	p-CN
9	m-NO_2	10	o-NO_2
11	m-F	12	o-F

Figure 7. O4I3 analogues induce OCT4 activity in a similar manner to the parental molecule. (A) Chemical structures of O4I3 derivatives, indicated by numbers. **(B)** NCCIT reporter cells were treated with 50 nM of the derivatives listed in **(A)** or O4I3 for 48 h, after which activation of *OCT4* promoter was measured using luciferase assay. Data present the mean of at least three independent experiments, Error bars ± SD. 0.1% DMSO was used as mock treatment. Statistical significance was made in comparison to O4I3 treatment using a two-tailed *t*-test. *, **, and *** denote *p*-values less than or equal to 0.05, 0.01, and 0.001, respectively.

2.2. O4I3 supports the generation and maintenance of human iPSCs.

2.2.1. O4I3 maintains the pluripotency of human PSCs.

Our initial functional annotation and enrichment analysis of DNA microarray in O4I3-treated human PSCs demonstrated that a concentration of 10 nM of the compound is able to promote PSCs' homeostasis as well as repressing their differentiation (**Figure 8A**). Consistent with this, gene expression analysis using qRT-PCR confirmed the stabilization of several pluripotency-associated genes (e.g., *OCT4*, *SOX2*, and *NAONG*) together with the inhibition of differentiation markers (**Figure 8B**). In addition, cellular viability of human iPSCs was increased by ~4-fold, as compared to the small molecule inhibitor of Rho-associated kinase (Rocki; known as Y-27632), which is used as a reference compound for maintaining iPSCs [276] (**Figure 8C**). As expected, zolpidem was found to have comparable effects to O413 on iPSCs' cellular survival, however at higher doses (**Figure 8C**).

2.2.2. O4I3 promotes the OSKM-mediated reprogramming process in resistant human fibroblasts.

To investigate whether O413 has a positive impact on OSKM-mediated reprogramming, we first treated fibroblasts (HFs) with the compound, and analyzed the expression of reprogramming-associated genes, using qRT-PCR. As demonstrated in **Figure 9A and 9B**, O4I3 treatment led to ~10- and ~100-fold induction in *OCT4* and *SOX2* mRNA levels, respectively, compared with non-treated HFs. This was found to be ~0.1% of OCT4 expression and ~1% of SOX2 expression in human PSCs, which were included as reference (**Figure 9A and 9B**).

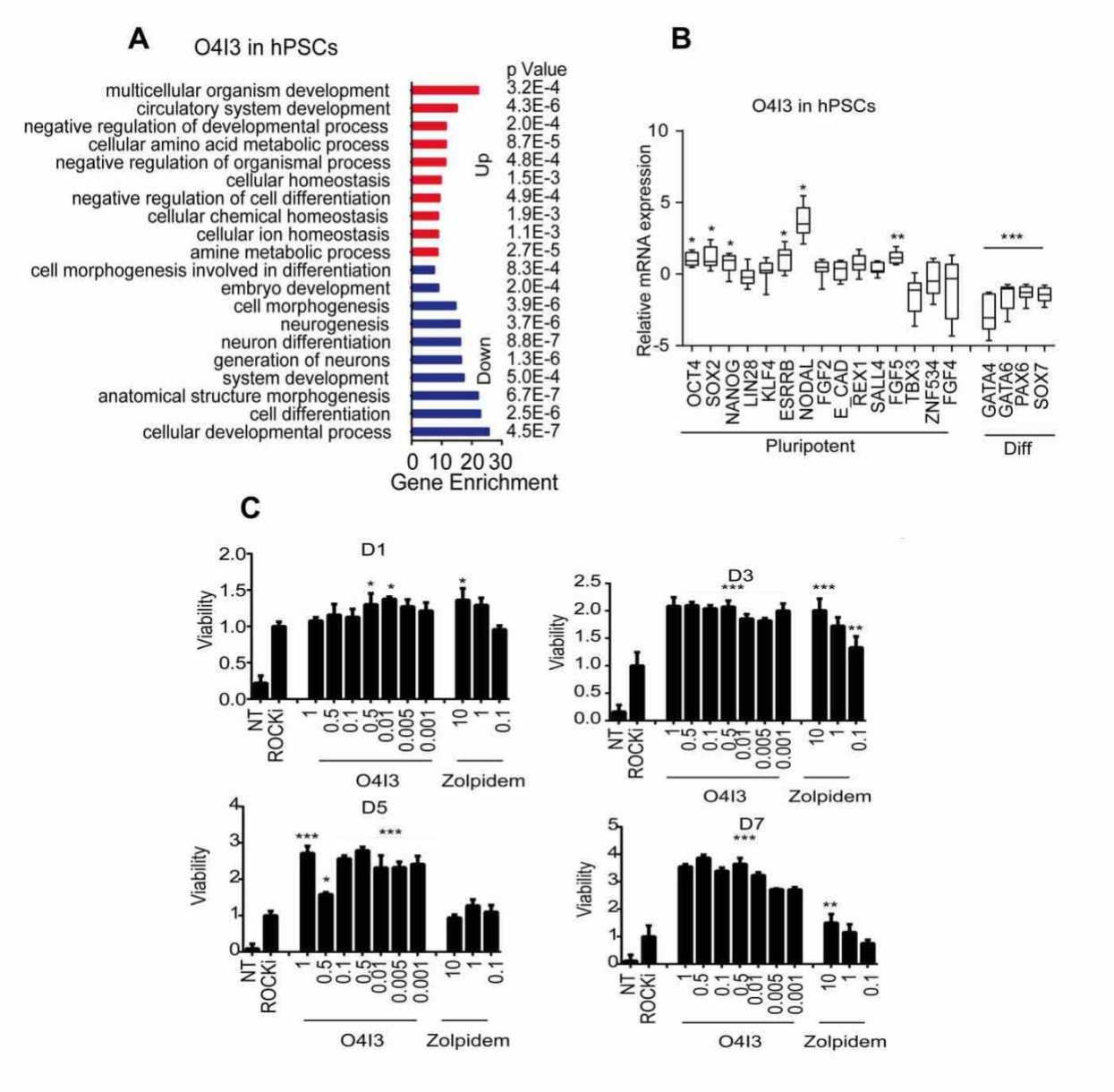


Figure 8. O4I3 supports iPSC pluripotency and viability. (A) Functional enrichment analysis of DNA microarray in human pluripotent stem cells (PSCs) treated with 50 nM of O4I3 for 24 h demonstrates a significant enrichment in gene ontology terms related to pluripotency, development, and homeostasis. A color code has been used for up- and downregulated genes, with red indicative of the former and blue indicative of the latter. **For more details regarding the analysis see the method section; 4.2.17. (B)** qRT-PCR analysis of the mRNA levels of several pluripotency- and differentiation-related genes

shows a clear increase and decrease with O4I3 treatment (10 nM, 48 h) in human PSCs, respectively. Relative mRNA expression was calculated using the ΔΔ Ct method where the Ct values of each gene were normalized to those of the housekeeping gene (βactin). **(C)** O4I3 enhances the viability of iPSCs after single-cell expansion using 10 μM of Rho kinase inhibitor (designated as ROCKi) for the first 24 h, followed by continuous cultivation for 7 days (D, days) in the presence of various concentrations of O4I3 or zolpidem. Cellular viability was determined as the number of treated cells over mock, using MTT cytotoxicity assay. Zolpidem was found to induce a similar pattern in increasing the viability of PSCs, however at higher tested doses compared with O4I3. Data in **(B)** and **(C)** present mean ± SD of at least three independent experiments. Statistical significance was performed in comparison to mock (0.1% DMSO) treatment or ROCKi-treated cells using a two-tailed student's *t*-test. *, **, and *** indicate a *p*-value less than or equal to 0.05, 0.01, and 0.001, respectively.

Having confirmed this, we combined O4I3 with a commercially available episome-based reprogramming cocktail, which enforces the expression of OCT4, SOX2, KLF4, and L-MYC (or the "OSKM"), as well as LIN28 and suppresses the expression of p53 (for simplicity, we will refer to this cocktail as OSKM throughout this section, although it contains additional vectors (LIN28 and p53) compared with the original OSKM cocktail). Interestingly, O4I3 synergized with OSKM and increased the number of TRA-1-60-positive colonies by 10- to 20-fold in four different human primary fibroblasts (HF1, HF2, HF3, and HF4) when compared with OSKM-only treated HFs (**Figure 9C**). Through various mechanisms, such as senescence donor fibroblasts might confer resistance to transcription factor-meditated reprogramming. In view of this, we observed that adding O4I3 to the OSKM cocktail not only promotes the generation of TRA-1-60-positive colonies in reprogramming-resistant fibroblasts, (designated as HF3 and HF4, **Figure 9C**), but also enables the generation of iPSC-like colonies, as detected by EGFP-expressing cells using reporter fibroblasts carrying a LTR7-EGFP construct (**Figure 9D**). Similar to O4I3, zolpidem was able to convert resistant fibroblasts into iPSCs in combination with *OSKM* ectopic expression (**Figure 9E**).

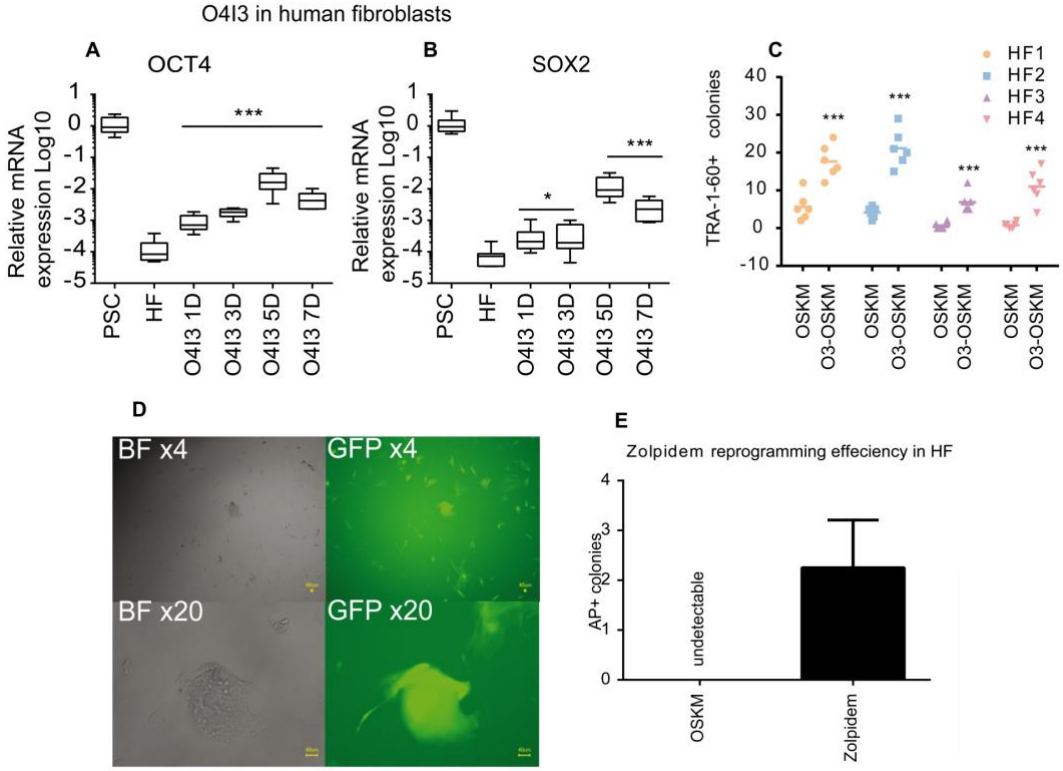


Figure 9. Imidazopyridines promote OSKM-mediated reprogramming of human fibroblasts. (A) and **(B)** O4I3 induces the mRNA levels of the master pluripotency genes, *OCT4* **(A)** and *SOX2* **(B)**, as determined by qRT-PCR analysis. Human fibroblasts (HFs) were treated with 50 nM of the small molecule for the indicated duration (D, days). Relative expression was determined using the ΔΔ Ct method in which the Ct values of *OCT4* and *SOX2* were normalized to those of the housekeeping gene (βactin). **(C)** OSKM-mediated reprogramming efficiency with O4I3 (designated as O3-OSKM, 50 nM) or without the small molecule, as determined by TRA-1-60 staining. Reprogramming-resistant fibroblasts are denoted as HF3 and HF4. **(D)** Representative fluorescence images, showing the generation of ESC-like colonies from resistant fibroblasts using a combination of O4I3 and *OSKM* forced expression, as determined by EGFP expression. The used HFs carry the LTR7-EGFP constructs in which the expression of EGFP is under the control *LTR7* promoter, which is known to be activated during reprogramming [277]. **(E)** Reprogramming-resistant fibroblasts were transfected with the OSKM cocktail in the absence or presence of the imidazopyridine analogue, zolpidem (10 μM), after which ESC-like colonies were detected using alkaline phosphatase (AP) staining. Data in **(A)**, **(B)**, **(C)**, and **(E)** present mean ± SD of three biological replicates. Statistical significance was calculated in comparison to mock treatment (0.1% DMSO) in **(A)** and **(B)** and to the OSKM-only treatment in **(C)** using a two-tailed student's *t*-test. * and *** present a *p*-value less than or equal to 0.05 and 0.001, respectively.

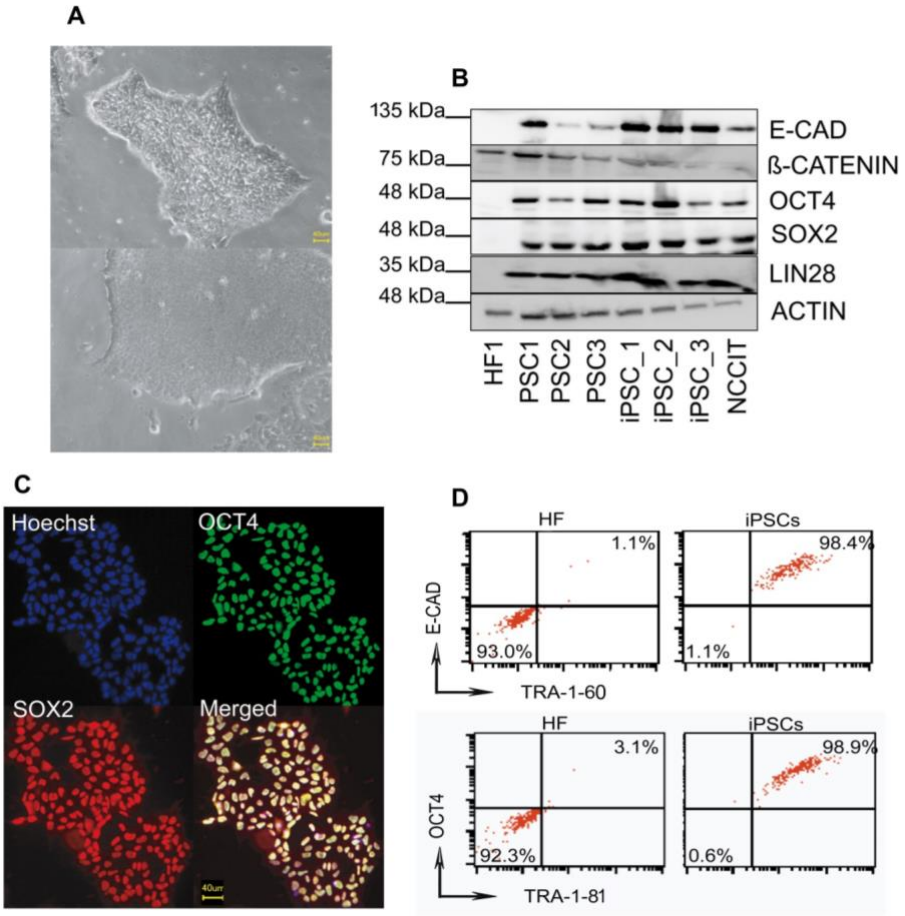


Figure 10. iPSCs generated by O4I3 and OSKM combination express pluripotency markers in a similar manner to human PSCs. (A) Bright field images showing the ESC-like morphology of the newly generated iPSCs. **(B)** Immunoblots demonstrating the expression of several pluripotency markers in new iPSCs (denoted as iPSC_1-3) in a comparable manner to those in human PSCs and NCCIT cells, which were included as reference. Human fibroblasts (HFs) served as the negative control, as they do not express the indicated pluripotency proteins. Actin served as the loading control. **(C)** Immunocytochemistry of the master pluripotency factors, OCT4 and SOX2 in the new iPSCs. **(D)** Representative density plots illustrating co-expression of the pluripotency-associated makers, E-

cadherin and TRA-1-60 (top) as well as OCT4 and TRA-1-81 (bottom) in the new iPSCs, assessed by flow cytometry analysis.

2.2.3. O4I3-generated iPSCs resemble human ESCs.

Next, we aimed to expand and characterize the new iPSC cell lines generated by means of O4I3 and *OSKM* forced expression. For this, iPSC-like colonies were isolated and cultured under feeder-free conditions for >30 passages. The newly established iPSCs showed a similar morphology to that of human PSCs (**Figure 10A**) and were found to have a similar expression pattern of pluripotency-associated markers (e.g., OCT4 and SOX2) with that of human PSCs, as shown by immunoblotting and immunocytochemistry (**Figure 10B and 10C**). In addition to this, flow cytometry analysis of OCT4, E-Cadherin, TRA-1-60, and TRA-1-81 demonstrated more than 95% of iPSC population, co-expressing E-Cadherin and TRA-1-60, as well as OCT4 and TRA-1-81 (**Figure 10D**). Comparative DNA microarray analyses showed an up- and downregulation of pluripotency-associated genes and fibroblast-related genes in two cell lines of the new iPSCs (indicated as iPSC_1 and iPSC_2), respectively (**Figure 11A**).

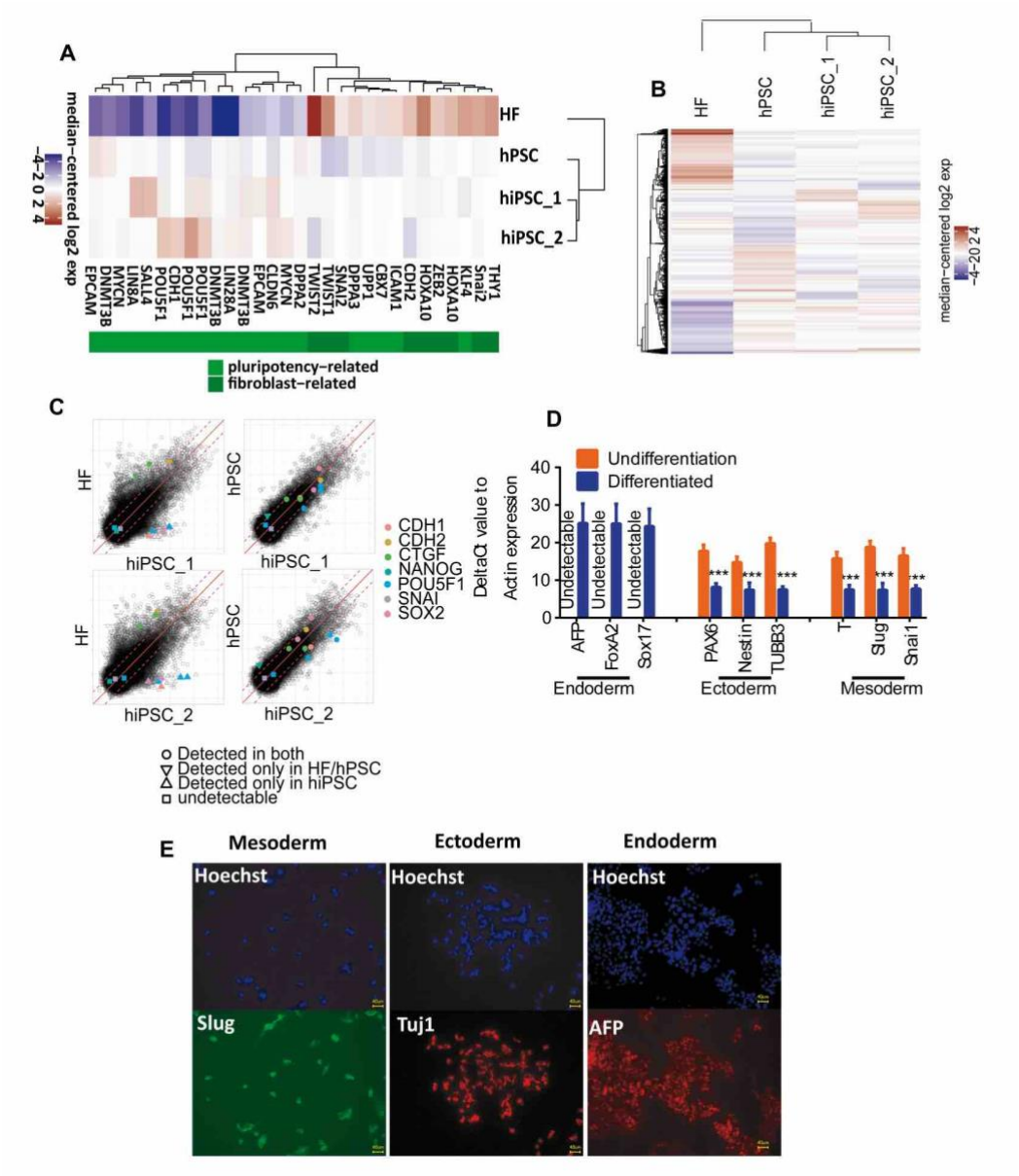


Figure 11. iPSCs generated by O4I3 and OSKM combination resemble human PSCs in their self-renewal and differentiation properties. (A) Heatmaps of gene expression profiles of human fibroblasts (HFs), hPSCs, and the newly generated iPSCs (denoted as iPSC_1-2). hPSCs, iPSC_1, and iPSC_2 show a high similarity in the expression pattern of the selected pluripotency- (light green) and differentiation-related (dark green) genes, which is distinct to that of HFs. **For more details regarding the analysis refer to the method section; 4.2.17. (B)** and **(C)** Heatmaps- **(B)** and scatter plots **(C)** of global gene expression profiling in HFs, hPSCs, and the new iPSCs. Heatmaps and scatterplots are based on log2-transformed gene expression values. Red dashed lines in **(C)** depict the 2-fold change in expression values. **(D)** qRT-PCR analysis of differentiation markers indicative of each germ layer. Data present the Δ Ct values of each gene respective to the housekeeping gene (βactin). Error bars ± SD, *n* = 3. Statistical significance was made in comparison to the undifferentiated iPSCs using a two-tailed student's *t*-test. **(E)** Immunocytochemistry of selected differentiation markers representative of the indicated lineages demonstrate the differentiation ability of the newly generated iPSCs. Scale bar: 40 μM. *** presents a *p*-value less than or equal to 0.001.

Furthermore, comparison of global gene expression profiles of iPSC_1 and iPSC_2 with human PSCs and HFs revealed a high resemblance of the gene expression pattern among the first three, but not with that of HFs, as determined by both heatmap- and scatterplot analyses (**Figure 11B and 11C**). Finally, the newly-generated iPSCs were found to be able to differentiate into the three germ layers (mesoderm, ectoderm, and endoderm)–one of the unique properties of PSCs–as determined by qRT-PCR and immunocytochemistry analyses of lineage-specific markers (**Figure 11D and 11E**).

2.2.4. O4I3 is an epigenetic modifier.

2.2.4.1. O4I3 protects the methylation of H3K4.

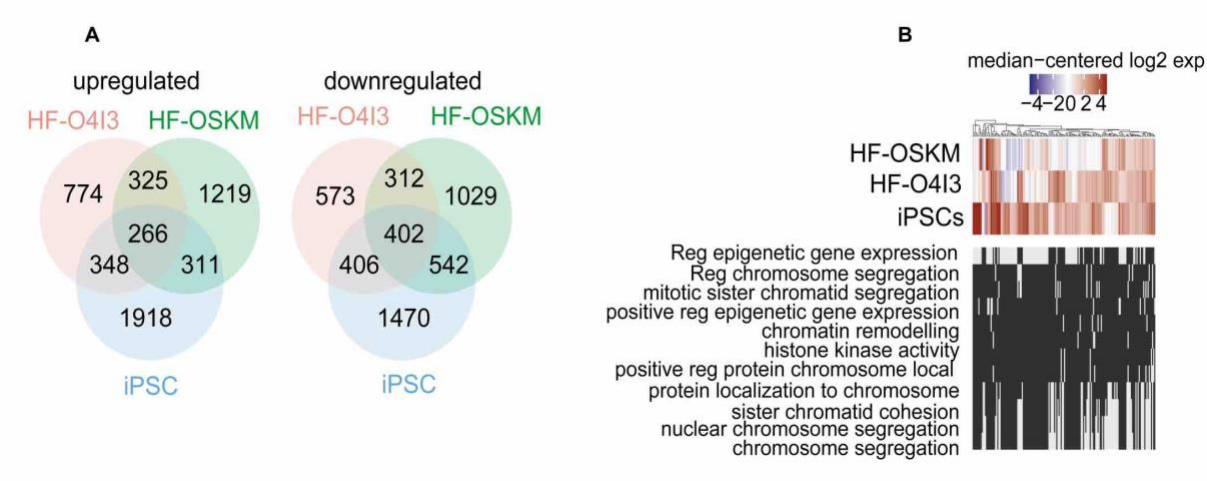


Figure 12. Epigenetic genes are on top of the O4I3-regulated genes. (A) Venn diagrams comparing the up- and downregulated genes in human fibroblasts (HFs) treated with 50 nM of O4I3 for 24 h (designated as HF-O4I3), HFs transfected with OSKM (HF-OSKM), and iPSCs. **(B)** Heatmaps of gene expression profiles demonstrate that among similarly regulated genes between the three groups, genes involved in epigenetic modification and remodeling are on top of the regulated terms. **For more details regarding the analysis refer to the method section; 4.2.17.**

To gain further insight into the mode of action of O4I3, we sought to compare the gene expression profiles of HFs, HFs treated with O4I3 (HF-O4I3), HFs transduced with OSKM (HF-OSKM), as well as human iPSCs (as positive control). More than 40% of all the regulated genes were found to be similar

among the three groups, of those, ~30% appeared to be involved in epigenetic modification and chromatin remodeling (**see Figure 12**). This encouraged us to study the epigenetic regulatory effects of O4I3 and its potential relevance to the reprogramming efficiency. As thoroughly discussed above, transcription factor-mediated reprogramming is largely dependent on epigenetic changes, among others, rearrangements of histone modifications (***summarized in an excellent review*** [278]). With that in mind, we focused on three well-known histone modifications associated with the promoter of *OCT4*, namely H3K4Me3, H3K9Me3, and H3K27Me3 (***thoroughly reviewed in*** [212]). Hypermethylation of H3K4 and demethylation of H3K9 and H3K27 represent some of the early epigenetic events during the course of reprogramming, which are required for loosening the compacted heterochromatin, and hence facilitating the binding of pluripotency transcription factors to the "open" chromatin [212, 278]. In light of this, chromatin immunoprecipitation-qPCR (ChIP-qPCR) analysis of reprogrammable fibroblasts (designated as HF1 and HF2) and reprogramming-resistant fibroblasts (designated as HF3 and HF4) revealed that O4I3 (in combination with OSKM) is capable of protecting H3K4 methylation at the promoter of *OCT4* at significantly higher levels compared with those treated with OSKM alone in HF2 as well as the resistant HFs (**Figure 13A**). H3K27Me3 occupation levels remained to be mostly unchanged in our experimental setting (**Figure 13A and 13C**). Additionally, we also analyzed global expression of H3K4Me3, H3K9Me3, and H3K27Me3 by immunocytochemistry and immunoblotting and observed an increase in case of H3K4Me3, a decrease in case of H3K9Me3, and no significant change in case of H3K27Me3 in response to O4I3 (**Figure 13B-D**). Moreover, an *in vitro* methylation assay demonstrated the ability of O4I3 in protecting trimethylated H3K4 with an IC_{50} value of ~20 nM (**Figure 13E**), further implicating the regulation of histone methylation as O4I3's mode of action with regards to cellular reprogramming.

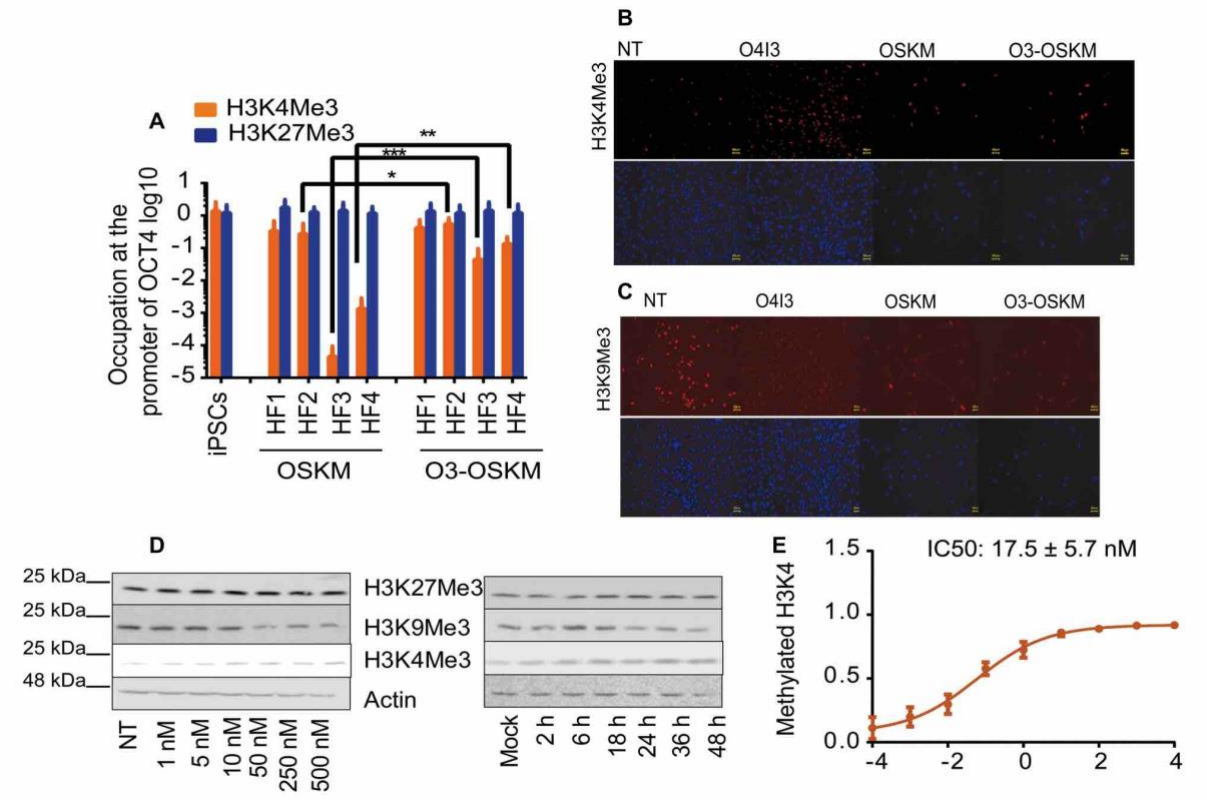


Figure 13. O4I3 is an epigenetic regulator through supporting H3K4 methylation. (A) ChIP-qPCR analysis of H3K4Me3- and H3K27Me3 occupation at the promoter of *OCT4* in reprogrammable and reprogramming-resistant human fibroblasts (HF1/2 and HF3/4, respectively) in the presence of O4I3 (designated as O3-OSKM, 50 nM) or in the absence of the molecule. The small molecule was found to increase the levels of H3K4Me3 at *OCT4* promoter, whereas methylated H3K27 remained to be mostly unchanged with treatment. **(B-D)** O4I3 increases the global levels of H3K4 **(B)** and decreases those of H3K9 **(C)**, as determined by immunocytochemistry of HFs treated with 250 nM of the molecule transfected with the OSKM cocktail or not. Representative immunoblots **(D)** show a similar pattern of protein expression of the aforementioned histone marks in HFs treated with increasing concentrations of O4I3 for 24 h or with 250 nM of the small molecule for the indicated time points. 0.1% DMSO served as mock and βactin was used as the loading control. **(E)** O4I3 protects H3K4 methylation with an IC_{50} value of ~20 nM, as measured by an *in vitro* methylation assay. Statistical significance in **(A)** was performed in comparison to OSKM-only treated fibroblasts using a two-tailed student's *t*-test. Error bars ± SD, *n* = 3. *, **, and *** are indicative of a *p*-value less than or equal to 0.05, 0.01, and 0.001.

2.2.4.2. O4I3 strongly inhibits the histone demethylase enzyme, KDM5A.

As discussed in the **section 1.1.1.2.**, histone methylation is a reversible process, meaning that one or more of the methyl groups existing on a particular lysine residue could be erased through the action of histone demethylases, for example KDM5s and LSD1 in case of H3K4 (***reviewed in*** [220]). This sparked the idea of O4I3 acting through the inhibition of specific histone demethylase enzymes. Focusing on the H3K4-related HDMs, we sought to test the inhibitory effect of O4I3 on LSD1 and KDM5, as well as KDM4, which is known to be selective for demethylation of H3K9 and H3K36 [279]. Our results from an *in vitro* demethylation assay revealed a strong inhibition of KDM4 and KDM5, however, with ~300-fold less efficiency in case of the former, as determined by the calculated IC_{50} values, 249 nM and 0.79 nM, respectively **(Figure 14**). Unlike the KDM family members, O4I3 did not show any major inhibitory effect on LSD1 even when using concentrations as high as 100 μM (**Figure 14A**). Having confirmed the potent affinity of O4I3 to KDM5, we were interested to assess whether there is a significant alteration in the small molecule's effects on different isozymes of the KDM5 family, namely KDM5A, KDM5B, KDM5C, and KDM5D. In this regard, O4I3 was found to have the most selectivity against KDM5A (with an IC_{50} value of ~0.19 nM), followed by KDM5D, KDM5C, and finally KDM5B (**Figure 14B**). Consistent with our initial hypothesis, suggesting a similar mode of action between zolpidem and O4I3 due to the existence of an imidazopyridine backbone, we observed an increase in H3K4 methylation and a decrease in KDM4/5 activity by zolpidem, however at higher concentrations compared with those of O4I3 (**Figure 14C**).

2.2.5. KDM5A represents a targetable epigenetic barrier during the process of reprogramming.

As our next step, we aimed at comparing the effects of known small molecule inhibitors of KDMs on the expression of pluripotency-associated markers. In light of this, CPI-455 [280] and JIB-04 [281], targeting KDM5 showed a clear induction of OCT4 in NCCIT reporter cells (**Figure 15A)**, an effect which could not be observed in cells treated with daminozide (inhibiting KDM2/7) [282] or GSK-J4 (a selective jumonji H3K27 demethylase inhibitor) [283]. Consistent with the results of reporter cells, JIB-04 promoted the mRNA expression of *OCT4*, *SOX2*, and *E-cadherin* in HFs (**Figure 15B and 15C**).

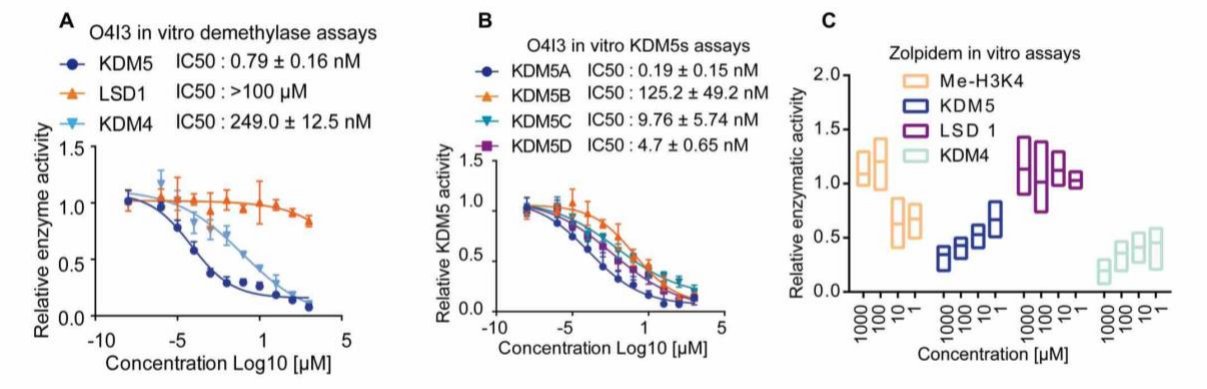


Figure 14. Imidazopyridines are strong inhibitors of H3K4 demethylase enzymes, KDM5s. (A) O4I3's inhibitory effect on histone demethylases, KDM4, KDM5, and LSD1, using an *in vitro* demethylation assay. **(B)** Among the four different isozymes of KDM5, O4I3 was found to have the most potent inhibitory activity against KDM5A *in vitro*. **(C)** The imidazopyridine analogue, zolpidem protects the methylation of H3K4 and inhibits its demethylases, KDM5s *in vitro* in a similar manner to O4I3. Data in **(A)** and **(B)** are mean ± SD of three biological replicates.

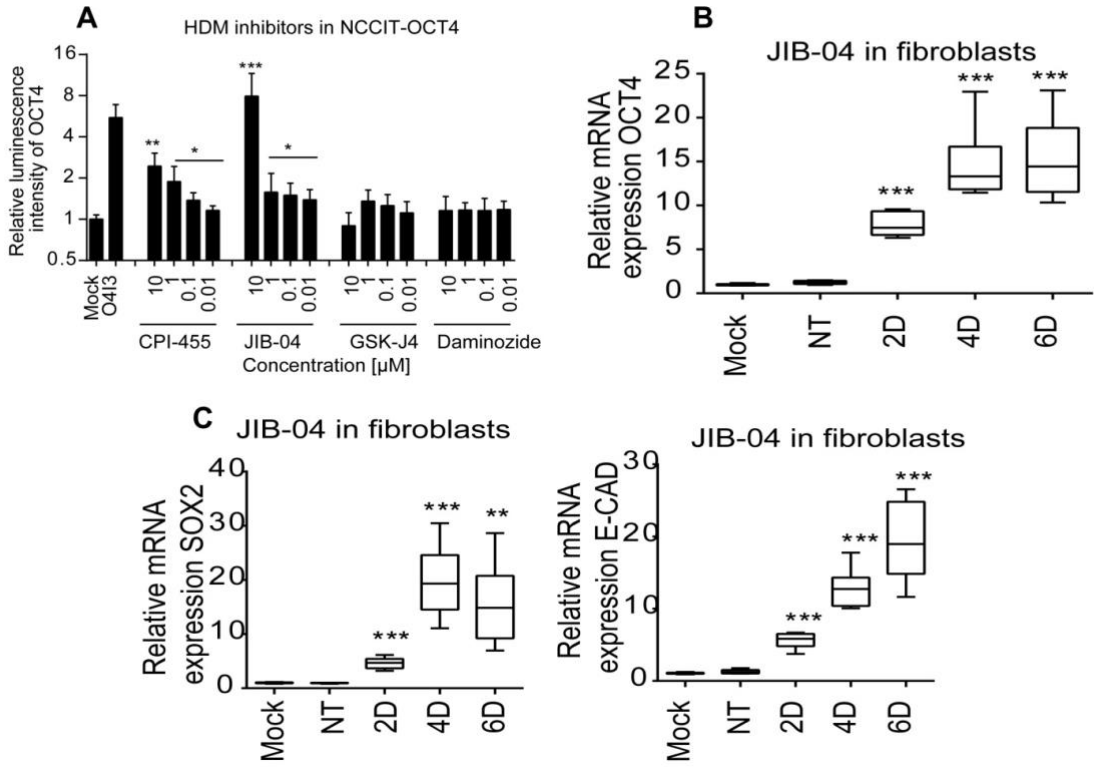


Figure 15. KDM5 chemical inhibition induces pluripotency-associated factors. (A) Various small molecule inhibitors of histone demethylases were employed on NCCIT-OCT4 reporter cells for 48 h, after which *OCT4* promoter activity was measured by luciferase assay. The selective KDM5 inhibitor, JIB-04 shows the highest OCT4 induction compared with mock (0.1% DMSO) treatment. **(B)** and **(C)** 5 μM of JIB-04 induces the mRNA levels of the master pluripotency gene, *OCT4* **(B)** as well as *SOX2* and *E-cadherin* (denoted as E-CAD) in human fibroblasts, assessed by qRT-PCR analysis (D, days). Relative mRNA expression was determined using the ΔΔ Ct method where the Ct values of each gene were normalized to those of the housekeeping gene (βactin). Data present mean ± SD of three biological replicates. Small molecule-treated cells were compared with mock treatment using a two-tailed student's *t*-test. *, **, and *** present a *p*-value of 0.05, 0.01, and 0.001, respectively.

Based on our observation with chemical inhibitors (**see Figure 15A**), showing a higher OCT4 induction in case of JIB-04, which is more selective towards the KDM5A isozyme compared with CPI-455, which is known to be a pan-KDM5 inhibitor [280, 281], we were intrigued to investigate the differential role of KDM5 family members in promoting pluripotency. For this, we compared the expression of KDM5A-D in NCCIT cells, HFs (both reprogramming-sensitive and -resistant), HFs reprogrammed by OSKM, as well as iPSCs, and found that KDM5A expression is higher in reprogramming-resistant HFs (designated as HF4), whereas KDM5B is rather expressed in cells with a pluripotent state (**Figure 16A**). This was consistent with the previously-published gene expression profiles [239] (**Figure 16B**) as well as our immunoblotting data (**Figure 16C**), altogether implicating KDM5A as an epigenetic barrier during the process of reprogramming.

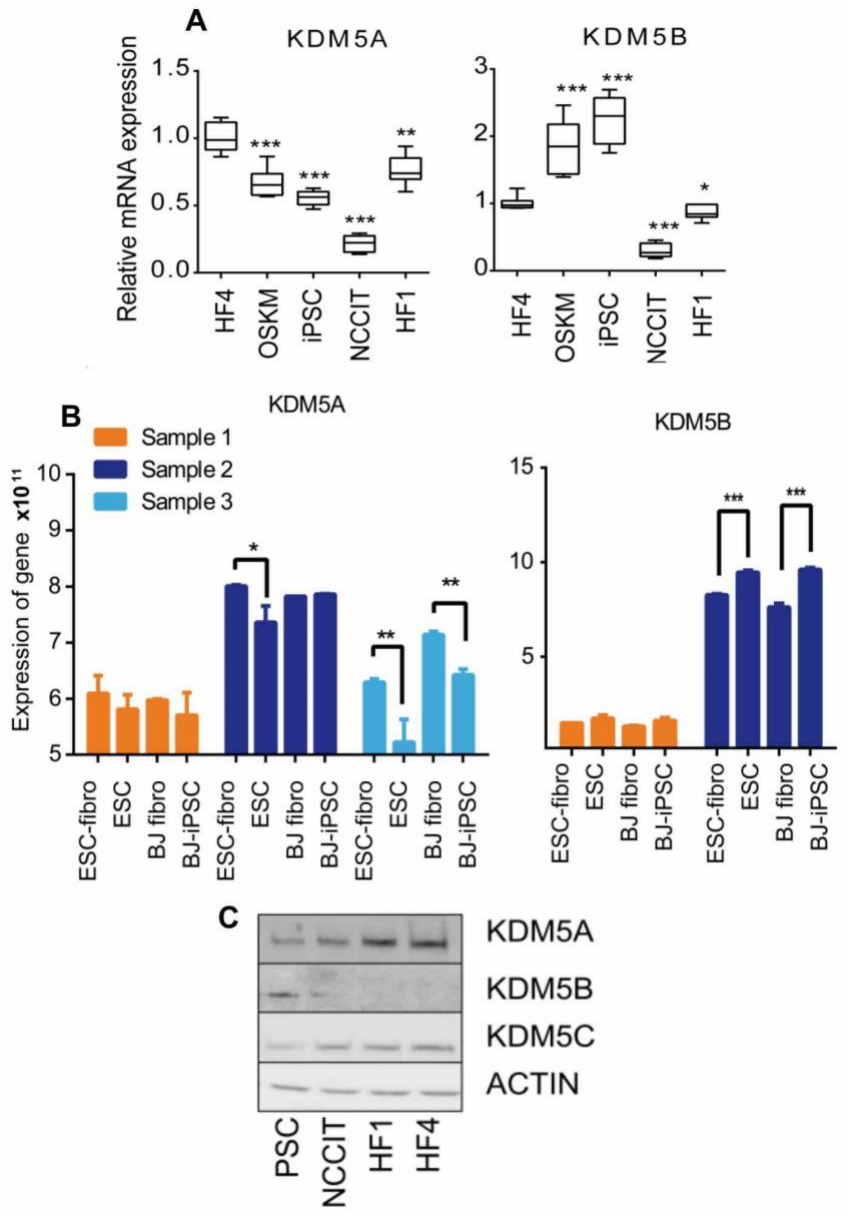


Figure 16. KDM5A is oppositely regulated in differentiated- and pluripotent states. (A) comparison of KDM5A and KDM5B mRNA expression in reprogrammable and reprogramming-resistant fibroblasts (HF1 and HF4, respectively), fibroblasts transfected with the OSKM cocktail, as well as iPSCs and NCCIT cells, assessed by qRT-PCT analysis. Relative mRNA expression was determined using the ΔΔ

Ct method where the Ct values of KDM5A/B were normalized to those of the housekeeping gene (β actin). KDM5A and KDM5B were found to have the highest and the lowest expression in HF4, respectively. Error bars ± SD, *n* = 3. **(B)** Consistent with qRT-PCR results, re-analysis of previously published gene expression data [239] revealed that KDM5B is mostly associated with a pluripotent state, while KDM5A has the highest expression in fibroblasts. **(C)** Immunoblots showing higher KDM5A protein levels in HFs as compared to PSCs and NCCIT cells. In contrast, KDM5B has the highest expression in pluripotent stem cells. βactin served as the loading control. In **(A)** statistical significance was performed respective to HF4 using a two-tailed student's *t*-test and in **(B)** multiple comparisons were employed by a two-way ANOVA test followed by a post-hoc Tukey test. *, **, and *** are representative of a *p*-value less than or equal to 0.05, 0.01, and 0.001, respectively.

In support of the repressing role of KDM5A in pluripotency, we found that genetic inhibition of KDM5A by siRNA significantly increases OCT4 expression in NCCIT reporter cells as well as *OCT4* and *SOX2* mRNA levels in HFs, an effect which could also be observed in case of KDM5D knockdown cells, but not in cells transfected with KDM5B/C siRNA (**Figure 17A and 17B**). Additionally, KDM5A knockdown cells demonstrated increased global methylation of H3K4 in HFs both in the absence and presence of OSKM (**Figure 17C**). As expected, *KDM5A* ectopic expression counteracted these effects mediated either by O4I3 or siRNA knockdown, as determined by significantly less OCT4 induction in HEK-293 and NCCIT reporter cells (**Figure 17D**). In addition to this, combining KDM5A chemical inhibitors (JIB-04 and O4I3) with *KDM5A* over-expression significantly abrogated OCT4-inducing ability of the former (**Figure 17E**). Finally, *KDM5A* over-expression was found to reduce the reprogramming efficiency of resistant HFs, as indicated by the number of TRA-1-60-positive colonies (**Figure 17F**).

Having confirmed KDM5A as an epigenetic barrier of reprogramming, we were inclined to further study the relevance of H3K4 methylation as well as its demethylation (which is mostly mediated by KDM5A) to the reprogramming efficiency. First, we compared H3K4Me3 occupation at the promoter of *OCT4* among reprogramming-resistant HFs (HF4), HF4 treated with JIB-04, as well as HF4 reprogrammed by OSKM in the absence- and presence of genetic/chemical inhibition of KDM5A and siRNA-mediated knockdown of *KDM5B-D* (**see Figure 18A**). As determined by ChIP-qPCR analysis, *OSKM* forced expression increases the levels H3K4Me3 at the promoter of *OCT4*, an effect which is further enhanced in combination with siRNA-mediated knockdown of *KDM5A*, but not with that of other members of the KDM5 family (**Figure 18A**). Chemical inhibition of KDM5A (using JIB-04) showed a similar pattern either as single treatment or in combination with ectopic expression of *OSKM* (**Figure 18A**). Consistent with this, OSKM-transduced HF3 cells were found to express *OCT4*, *SOX2*, and *E-Cadherin* mRNAs at significantly higher levels in the presence of anti-*KDM5A* siRNA or small molecule inhibitors, O4I3 and JIB-04 (**Figure 18B**). This implicates that ectopic expression of *OCT4* (using the OSKM cocktail) is not sufficient to initiate reprogramming in resistant fibroblasts. Using 11 fibroblast cell lines derived from different patients, we showed a negative correlation between *KDM5A* expression and reprogramming efficiency, as measured by the number of AP-positive colonies (**Figure 18C**). Finally, transient repression of KDM5A using either siRNA or the chemical inhibitor, JIB-04, was found to facilitate OSKM-mediated reprogramming in both sensitive- (HF1 and HF2) and resistant (HF3 and HF4) fibroblasts, as determined by increased number of TRA-1-60-positive cell population (**Figure 18D**).

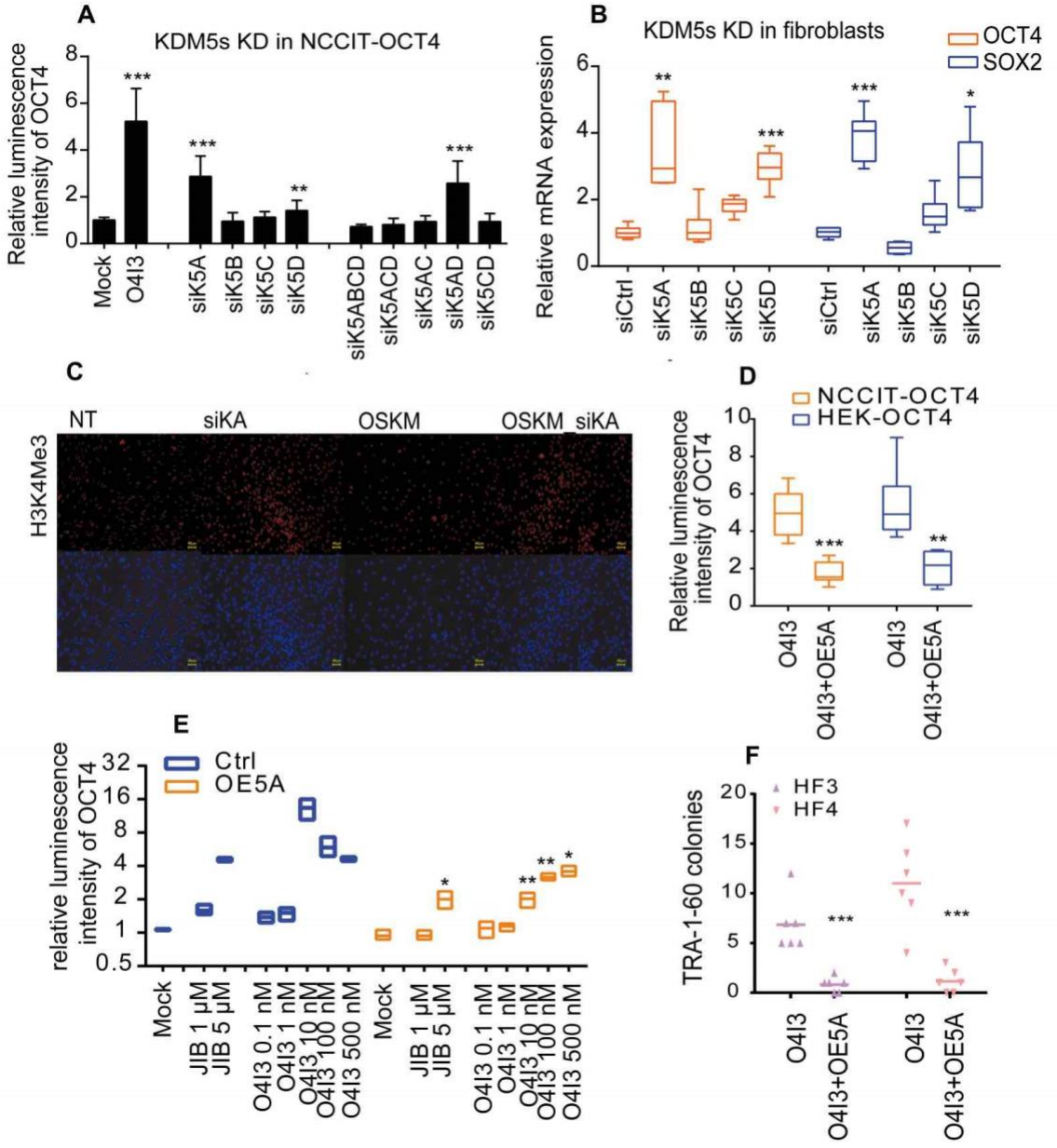


Figure 17. O4I3-mediated induction of pluripotency is associated with its inhibitory activity against KDM5A. (A) siRNA-mediated knockdown (KD) of *KDMA/D* significantly induces OCT4 transcription in NCCIT-OCT4 reporter cells, as measured by luciferase assay. **(B)** *OCT4* and *SOX2* mRNA levels are upregulated upon knocking down *KDM5A* and *KDM5D*, as detected by qRT-PCR analysis. A non-targeting siRNA (indicated as siCtrl) served as the negative control. Relative mRNA expression was calculated using the ΔΔ Ct method where the Ct values of each gene were normalized to those of the housekeeping gene (βactin). **(C)** Immunocytochemistry of human fibroblasts revealed a global increase in H3K4Me3 levels in cells transfected with siRNA against *KDM5A* (denoted as siKA) in the absence or presence of *OSKM* ectopic expression. **(D)** In contrast to siRNA-mediated repression of *KDM5A*, its over-expression (denoted as OE5A) compromises O4I3's induction of OCT4 in HEK-293 and NCCIT reporter cells, further proving KDM5A as the small molecule's target. Cells were transfected with the over-expression construct and were treated with O4I3 at a concentration of 100 nM for 48 h. **(E)** The combination of pharmacological inhibition of KDM5A by O4I3 or JIB-04 and *KDM5A* ectopic expression antagonizes the former's effects on inducing OCT4 expression. NCCIT-reporter cells were transfected with the over-expression construct and were treated with various concentrations of O4I3 for 48 h. An empty vector was used as the negative control (denoted as Ctrl). **(F)** Reprogramming efficiency was

found to be significantly reduced in resistant fibroblasts in the presence of *KDM5A* forced expression, as determined by TRA-1-60 staining. Error bars ± SD, *n* = 3. Statistical significance was made in comparison to mock (0.1% DMSO) in **(A)**, to the negative control in **(B)** and **(E)**, and to O4I3-only treated cells in **(D)** and **(F)** using a two-tailed student's *t*-test. *, **, and *** denote a *p*-value less than or equal to 0.05, 0.01, and 0.001, respectively.

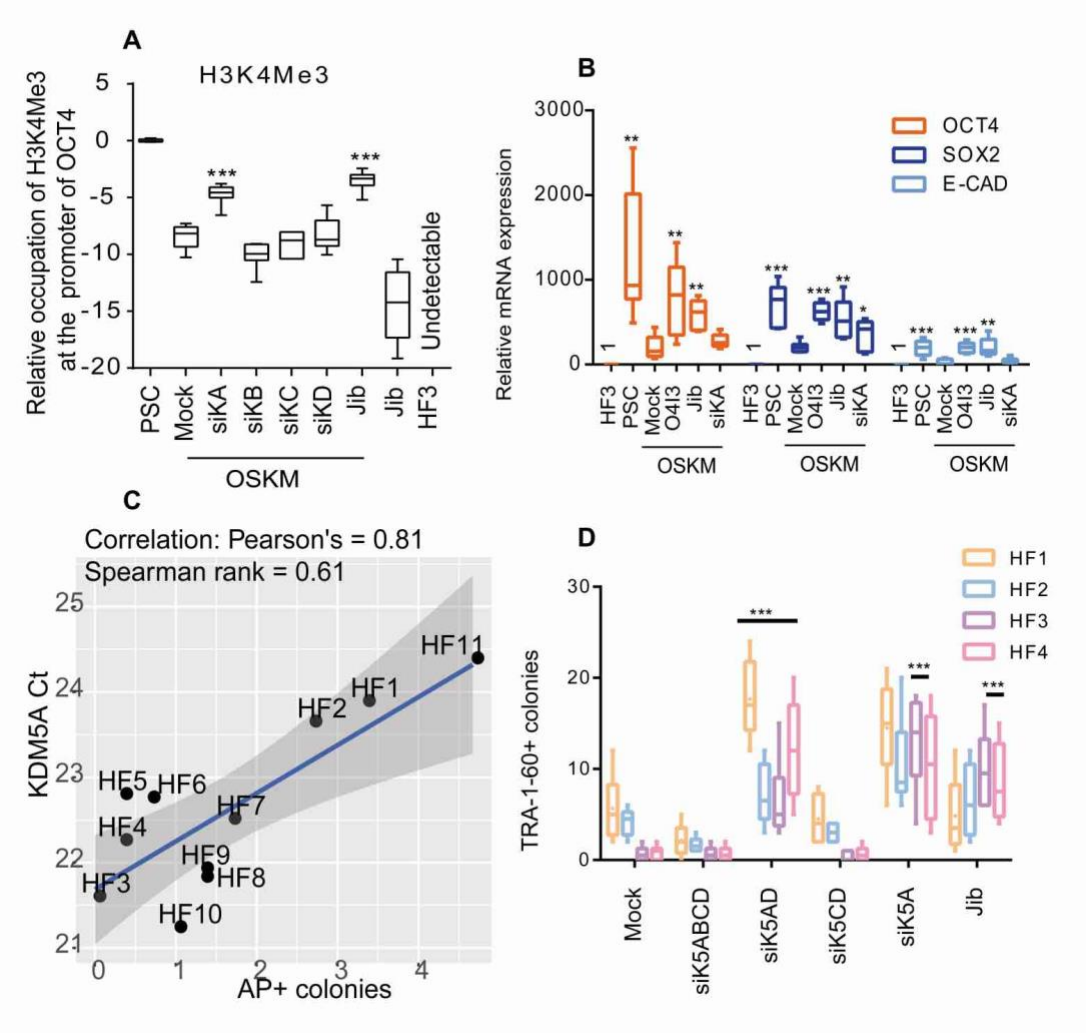


Figure 18. KDM5A presents a targetable epigenetic barrier during the course of OSKM-based reprogramming. (A) siRNA-mediated suppression of *KDM5A* as well as its chemical inhibition using JIB-04 (5 μM, 48 h) supports H3K4Me3 occupation at *OCT4* promoter in reprogramming-resistant fibroblasts (designated as HF3), as determined by ChIP-qPCR analysis. **(B)** In line with the enrichment of methylated H3K4 at the promoter of *OCT4*, chemical/genetic suppression of KDM5A upregulates the indicated pluripotency-associated genes in HF3 cells when combined with *OSKM* ectopic expression, as detected by qRT-PCR analysis. Relative mRNA expression was determined using the ΔΔ Ct method where the Ct values of the pluripotency genes were normalized to those of the housekeeping gene (βactin). **(C)** A positive correlation was observed between *KDM5A* expression and reprogramming efficiency (assessed by the number of AP-positive colonies) in 11 human primary fibroblasts. Similar amount of cDNA was used for qRT-PCR analysis; Ct values are indicative of *KDM5A* expression. **(D)** siRNA-mediated- and pharmacological inhibition of KDM5A facilitates OSKM-based reprogramming of sensitive- (HF1/2) and resistant (HF3/4) fibroblasts, assessed by TRA-1-60 staining. HFs were transfected with the respective siRNA oligos (si) against KDM5 family members (K5A/B/C/D) for the first 5 days of reprogramming or treated with JIB-04 (5 μM) for the first 48 h. In **(A)** and **(B)** and **(D)**, statistical significance was made respective to mock, which is the OSKM-only treated HFs, using a two-tailed

student's *t*-test. Data are presented as mean ± SD, *n* = 3. ** and *** indicate a *p*-value less than or equal to 0.01 and 0.001, respectively.

2.3. O4I3 and O4I4 combination is able to generate iPSC-like colonies in the absence of exogenous OCT4.

2.3.1. O4I4 promotes OCT4-associated signaling in various cell lines.

Having studied the mode of action of O4I3 in detail, we aimed at optimizing a chemical condition in which exogenous OCT4 could be completely replaced. For this, we returned to our initial hits from the cell-based HTS (**see section 2.1.1.**) and selected the most metabolically-stable derivative of the compound O4I2–herein O4I4–for the upcoming experiments. As described in the **section 2.1.1.**, O4I4 was able to induce the expression of pluripotency factors, including OCT4 and NANOG in HEK-293 and NCCIT reporter cells. Further experiments in O4I4-treated NCCIT cells showed that the small molecule increases the mRNA levels of several pluripotency-related genes, as determined by qRT-PCR (**Figure 19A**) and increases cell populations with the expression of OCT4, SOX2, E-Cadherin, and NANOG, as indicated by FACS analysis (**Figure 19B**). In addition to NCCIT cells, iPSC cells deprived of bFGF, which is known to be essential for the maintenance of pluripotency [284], were found to rescue in the presence of O4I4, as indicated by comparable mRNA levels of several pluripotency genes between O4I4-treated iPSCs in the absence of bFGF and iPSCs cultured in standard media (**Figure 19C**). Consistently, primary HFs treated with the small molecule demonstrated an increase in the mRNA expression of several pluripotency genes (**Figure 19D**). Finally, promising data obtained from functional enrichment analysis of HF cells treated with O4I4 revealed high enrichment of biological processes associated with suppressing differentiation, on the one hand, and promoting cellular proliferation and adhesion, on the other hand (**Figure 19E**).

2.3.2. O4I3/4 enables the generation of human iPSC-like colonies in the absence of exogenous OCT4.

Based on the exciting results from O4I3's mode of action (**see section 2.2.**), showing a significant increase in the efficiency of OSKM-mediated reprogramming through inducing endogenous OCT4 expression, we thought of combining both chemicals, O4I3 and O4I4, with episomal transfection of "SKML" (SOX2, KLF4, MYC, and LIN28) and without *OCT4* forced expression–collectively designated as *"CSKML"*– for the generation of human iPSCs (**Figure 20A**). We then applied CSKML to reporter HF cells expressing GFP under the control of *LTR7* promoter and observed few ESC-like colonies expressing GFP with higher intensity compared with that of "SKML" and "O4I4-SKML"-treated cells (**Figure 20B**). These cells were successfully isolated, expanded, and were used for TRA-1-60 staining, which revealed a comparable number of positive cells between CSKML- and OSKML-treated fibroblasts (**Figure 20C**). qRT-PCR and immunoblotting analyses showed similar mRNA and protein levels of several pluripotency-related markers (e.g., OCT4 and SOX2) in CSKML-iPSCs and the standard iPSCs (**Figure 20D and 20E**). Moreover, comparison of global gene expression profiles in CSKML-iPSCs, iPSCs, and fibroblasts revealed a similar gene expression pattern among the first two, which was distinct from that of HFs (**Figure 20F**).

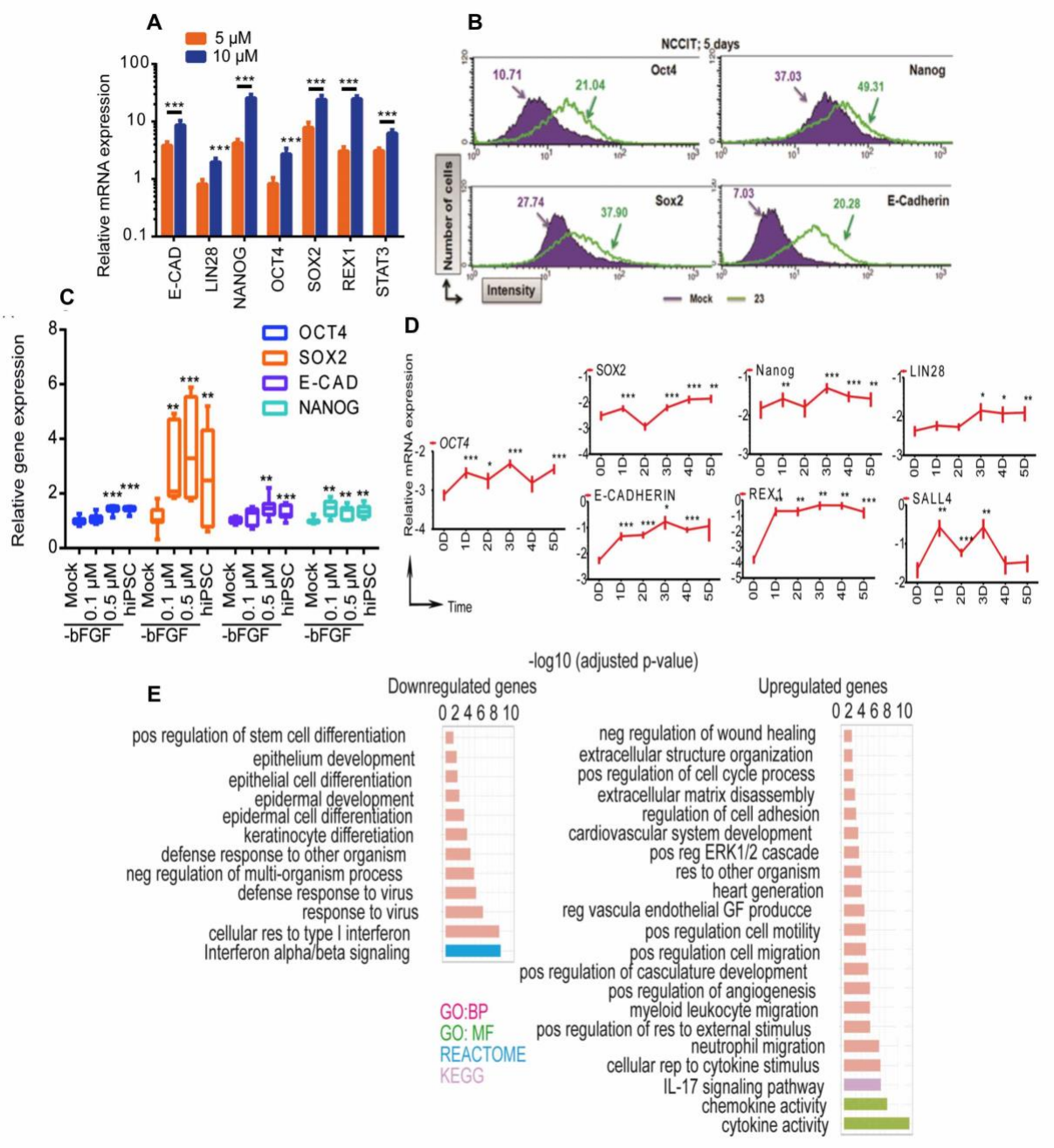


Figure 19. O4I4 induces pluripotency-related signaling molecules. (A) qRT-PCR analysis of the mRNA levels of several pluripotency-associated genes in NCCIT cells treated with the indicated concentrations of O4I4 for 24 h. **(B)** 10 μM of O4I4 treatment for 5 days increases the number of cells expressing the indicated pluripotency-associated markers in NCCIT cells compared with mock (0.1% DMSO) treatment, as detected by flow cytometry analysis. O4I4 is denoted as compound No. 23 (**see Figure 5**). **(C)** O4I4 is able to protect iPSC pluripotency in the absence of bFGF, as determined by qRT-PCR analysis of multiple pluripotency genes. Human iPSCs were cultured in the presence of 10 ng/mL bFGF for 24 h and were further cultivated with the indicated concentrations of O4I4 instead of bFGF for another 24 h. 0.1% DMSO served as mock and iPSCs cultured in standard media were used as the positive control. **(D)** 20 μM of O4I4 was employed on human fibroblasts for the indicated durations (D, days), after which qRT-PCR was performed for the assessment of pluripotency gene expression. **(E)** Functional enrichment analysis of up- and downregulated genes upon O4I4 treatment in human fibroblasts. The height of the bars is directly proportional to the significance of the enrichment. Relative mRNA expression in **(A)**, **(C)**, and **(D)** was determined using the ΔΔ Ct method where the Ct values of each gene were normalized to those of the housekeeping gene (βactin). Error bars present the SD of

three independent experiments. Statistical significance was made in comparison to mock treatment in **(A)** and **(C)** and to the starting time point (day 0, indicated as 0D) in **(D)** using a two-tailed student's *t*-test. *, **, and *** are indicative of a *p*-value less than or equal to 0.05, 0.01, and 0.001, respectively.

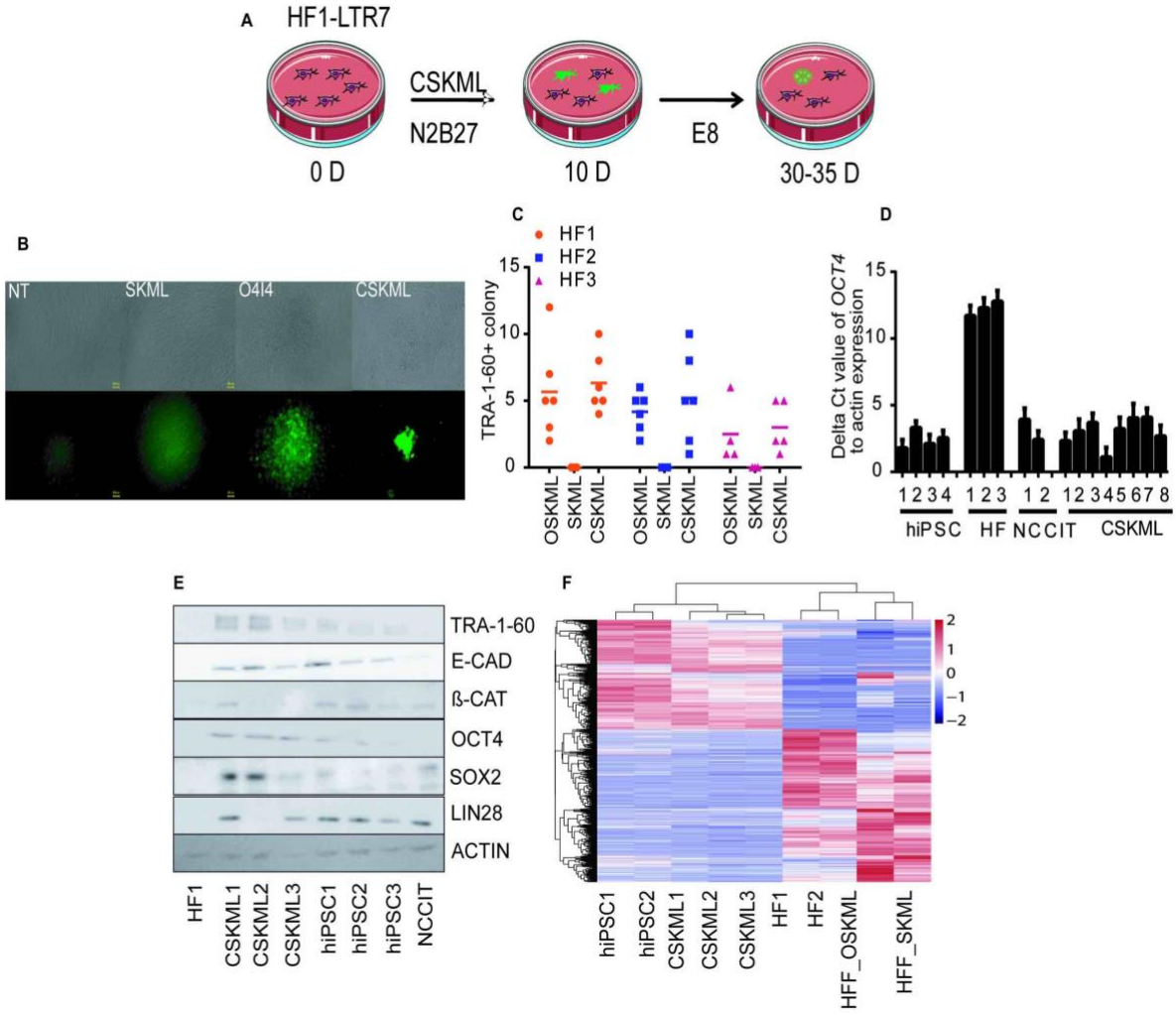


Figure 20. The combination of O4I3 and O4I4 is able to convert human fibroblasts to iPSC-like colonies in the context of transcription factor-based reprogramming and without the need for *OCT4* ectopic expression. (A) A schematic illustrating the reprogramming process of human fibroblasts carrying the LTR7-EGFP promoter (denoted as HF1-LTR7) using a combination of O4I3/4 and the ectopic expression of "SKML" cocktail, containing *SOX2*, *KLF4*, *MYC*, and *LIN28*, without exogenous *OCT4*. HF1 cells were pre-incubated with 250 nM of O4I3 for 2 days, followed by 5 μM of O4I4 for 7 days in combination with *SKML* over-expression. The small molecule-containing medium was refreshed daily. At day 10, the medium was replaced with E8 medium, as indicated. D, days. **(B)** GFP-positive colonies were detected to have a higher intensity in CSKML-treated HF1 and O4I4-SKML treated cells compared with non-transfected cells and cells over-expressing SKML alone. **(C)** Reprogramming efficiency was compared between human fibroblasts (HF1/2/3) transfected with SKML, OSKML, and CSKML cocktails, using TRA-1-60 staining. **(D)** qRT-PCR analysis of *OCT4* mRNA expression using relative Ct values; DeltaCt = Ct_{gene}-Ct_{actin}. βactin served as the housekeeping gene. Error bars ± SD, *n* = 3. **(E)** Immunoblots showing comparable protein levels of pluripotency-related factors in CSKML-reprogrammed HFs to hiPSCs and NCCIT cells, which were used as positive controls. HF1 served as the negative control. βactin was used as the loading control. **(F)** Heatmaps obtained from DNA microarray analyses of iPSCs, HF1, HFs reprogrammed by CSKML, OSKML, and SKML demonstrate a similar gene expression pattern between iPSCs and CSKML-treated HFs.

Given the cellular mechanism of O4I3 (**see Figure 13 and Figure 14**), we analyzed the global methylation status at the *OCT4* promoter, using bisulfite sequencing, and observed a general hypomethylation in iPSCs generated by means of CSKML compared with HF cells, which could suggest reactivation of endogenous OCT4 in the former (**Figure 21**).

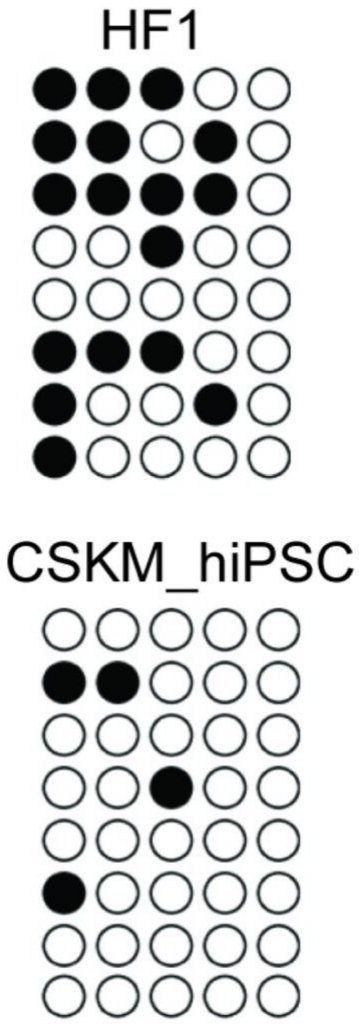


Figure 21. Hypomethylation of *OCT4* gene promoter in iPSCs generated by means of CSKM compared with human fibroblasts (HF1), as detected by bisulfite sequencing. Each horizontal line is representative of the sequencing results of one sub-clone. Methylated- and unmethylated CpG sites are depicted as solid- and open circles, respectively.

2.3.3. O4I3/4-mediated activation of endogenous OCT4 involves the BMP/Smad pathway.

2.3.3.1. "CSKML"-mediated iPSC generation is dependent on HMGA1 expression.

To gain some mechanistic insight into the O4I3/4's effects on cellular reprogramming, we sought to perform RNA-sequencing (RNA-seq) on HFs, HFs treated with either O4I3 or O4I4 and the combination (O4I3/4), as well as HFs reprogrammed by either OSKML or CSKML. We then ranked the genes, which were upregulated in O4I3/4-treated fibroblasts and compared them with those of iPSCs. As shown in **Figure 22A**, high-mobility group A1 (HMGA1) was found to be one of the top hits and has also been previously reported to support pluripotency [285, 286]. Consistent with the RNA-seq results, ATAC-sequencing revealed higher chromatin accessibility at *HMGA1* promoter in HF cells treated with CSKML compared with those transfected with OSKML, hinting on the possibility that *HMGA1* transcription is

relevant to CSML-mediated reprogramming (**Figure 22B**). We then analyzed the timely regulation of OCT4 and HMGA1 protein expression in O4I3/4-treated HFs by immunoblotting and observed that HMGA1 accumulation occurs prior to that of OCT4 (**Figure 22C**). In addition to this, small molecule-treated HF cells showed similar or slightly higher HMGA1 protein levels compared with iPSCs (**Figure 22D**). In line with our expectation, fibroblasts transfected with "SKML" failed to express both HMGA1 and OCT4, as determined by immunoblotting (**Figure 22D**). Furthermore, HMGA1 was not detectable in cells transfected with *"OSKML"* for 4 days, whereas it was clearly expressed in CKML-treated HFs (**Figure 22D**).

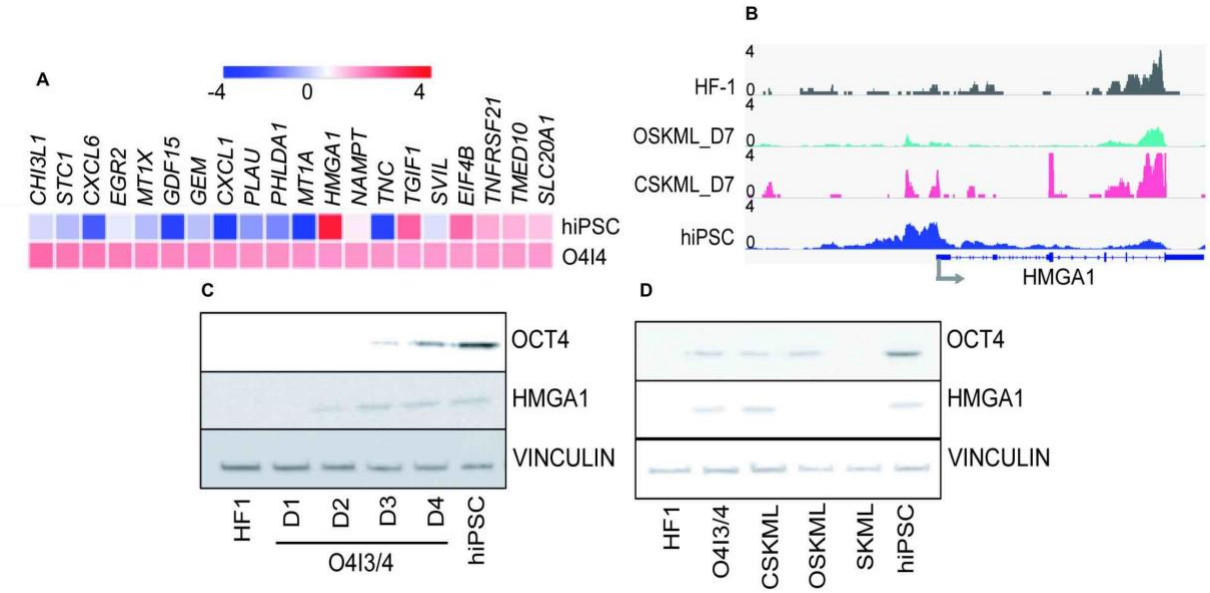


Figure 22. HMGA1 is involved in CSKML-mediated reprogramming. (A) Heatmaps obtained from RNA-sequencing, showing the top upregulated genes in human iPSCs as well as O4I4-treated fibroblasts. **(B)** ATAC-sequencing demonstrates chromatin accessibility of *HMGA1* gene promoter in human fibroblasts (HF1), HF1 cells treated with either OSKML or CSKML for 7 days (D, days), as well as iPSCs, which were included as reference. **(C)** A time-dependent analysis of HMGA1 and OCT4 protein levels in O4I3/4-treated HF1 cells reveals that HMGA1 expression precedes that of OCT4, detected by immunoblotting. O4I3 and O4I4 were used at a concentration of 0.5 μM and 10 μM for the indicated time points (D, days), respectively. **(D)** Immunoblots showing the concomitant expression of OCT4 and HMGA1 in HF1 treated with O4I3/4 (0.5/10 μM for 4 days), CSKML, as well as in human iPSCs but not in OSKML- and SKML-treated cells. Vinculin was used as the loading control.

To further support our hypothesis regarding the role of HMGA1 in CSKML-mediated activation of OCT4 during reprogramming, we knocked down this gene in fibroblasts using siRNA and found that *HMGA1*-deficient HF cells fail to promote OCT4 expression upon CSKML treatment, as determined by immunoblotting and qRT-PCR analyses (**Figure 23A and 23B**). Further evidence came from the ChIP-qPCR experiment, showing the occupation of *HMGA1* at the promoter of *OCT4* (**Figure 23C**), which was in good agreement with a previous report [287]. Finally, we aimed at replacing O4I4 with *HMGA1* over-expression for reprogramming of primary fibroblasts towards pluripotency. As indicated in **Figure 23D**, forced expression of *HMGA1* from day 2 to day 9 showed the highest efficiency among other tested durations, as determined by TRA-1-60 staining. We then designated the new iPSCs generated by means of *HMGA1* ectopic expression in combination with *SKML* as *"HSKML"*, which were found to

express the iPSC marker, TRA-1-60, however to a lower extent than that of OSKML-induced iPSCs (**Figure 23E**).

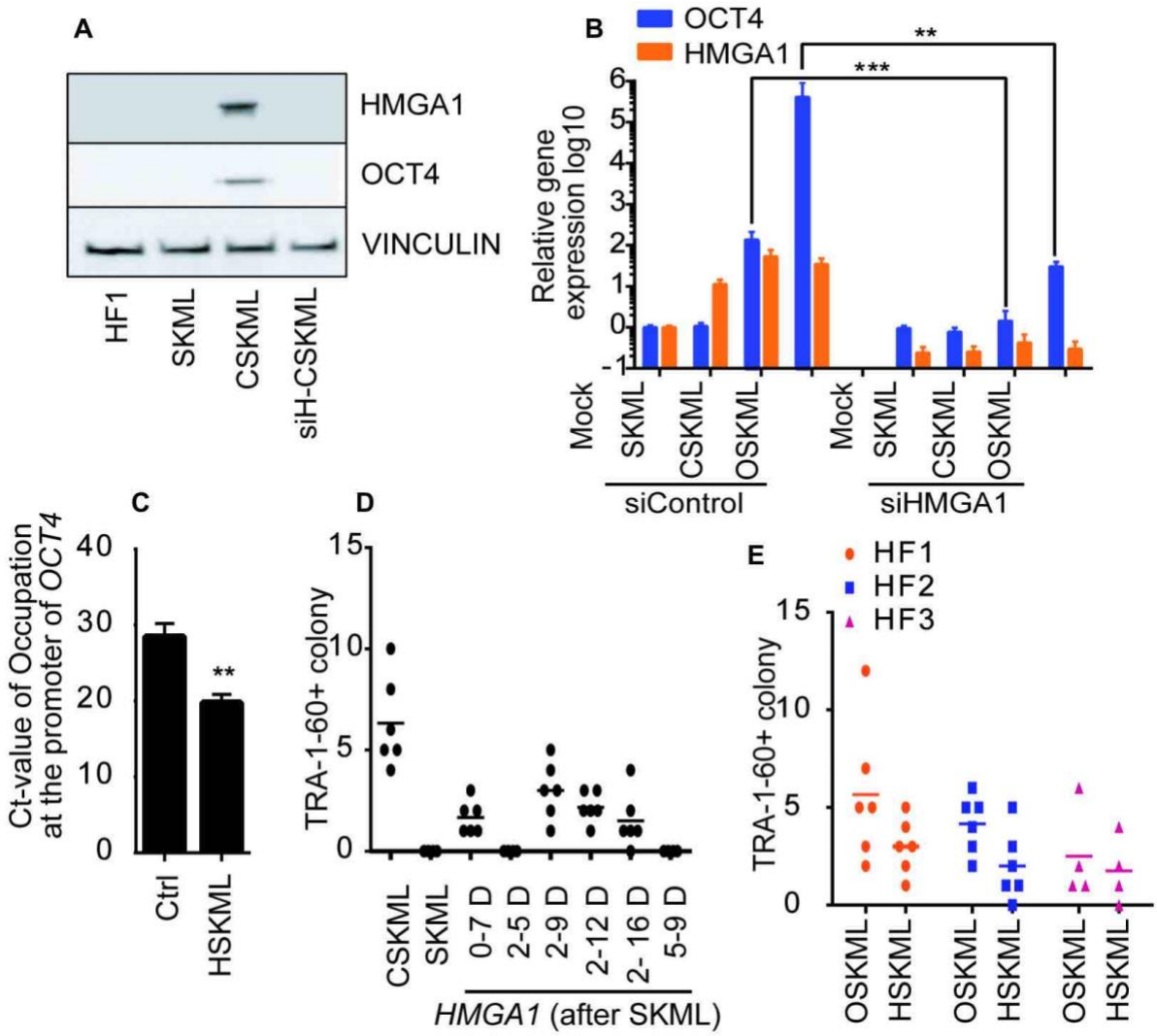


Figure 23. HMGA1 is involved in CSML-mediated reprogramming. (A) *HMGA1* knockdown using siRNA (designated as siH) abolishes CSKML-induced OCT4 induction in human fibroblasts (HF1), detected by immunoblotting. Anti-*HMGA1* siRNA was co-transfected with the SKML cocktail (**see Figure 20 for the treatment scheme**). Vinculin was used as the loading control. **(B)** Similar to its protein levels, mRNA expression of *OCT4* is significantly reduced upon genetic repression of *HMGA1* in both CSKML- and OSKML-treated HFs, assessed by qRT-PCR analysis. A non-targeting siRNA (designated as siControl) was used as the negative control. Relative gene expression was calculated using the ΔΔ Ct method where the Ct values of the target genes were normalized to those of the housekeeping gene (βactin). **(C)** ChIP-qPCR analysis showing the occupation of *HMGA1* at the promoter of *OCT4*. **(D)** Reprogramming efficiency of HF cells treated with the SKML cocktail and transfected with the *HMGA1* over-expression construct at the indicated days (D, days) was determined as the number of TRA-1-60-positive colonies. Transient transfection was refreshed every 2-3 days. **(E)** TRA-1-60 staining demonstrates the efficiency of the reprogramming process mediated by *HMGA1* over-expression in combination with *SKML* ectopic expression (referred to as HSKML). Error bars are the SD of three independent experiments. In **(B)** multiple comparisons were performed using a two-way ANOVA test and post-hoc Tukey test and in **(C)** statistical significance was calculated using a two-tailed student's *t*-test. ** and *** are representative of a *p*-value less than or equal to 0.01 and 0.001, respectively.

2.3.3.2. "CSKML"-mediated HMGA1 activation relies on the BMP/Smad pathway.

To understand the mechanisms by which HMGA1 is activated upon CSKML treatment, we looked into the differentially expressed genes among fibroblasts treated with O4I4, DMSO (as mock), OSKML, and CSKML, using RNA-seq. As demonstrated in **Figure 24A**, three genes, namely *PDGFRL*, *BMP6* and *DACT1*–all of which have been previously linked to stemness signaling pathways–were found to be induced only in case of CSKML-treated HFs as compared to SKML-transfected cells. The BMP family of cytokines have been frequently shown to promote transcription factor-mediated- and chemical reprogramming [288, 289]. Having that in mind, we focused on the potential role of BMP signaling in CSKML-mediated reprogramming. As thoroughly described in the **section 1.1. of chapter 1**, the BMP pathway initiates with receptor-mediated phosphorylation of Smad1/5/8, enabling their complex formation with Smad4. The R-Smad-Smad4 complexes are then transferred to the nucleus where they regulate the transcription of various downstream genes, among others, *ID*s (***basics of TGFβ/BMP/Smad signaling are thoroughly reviewed in*** [2]). In light of this, treatment of fibroblasts with increasing concentrations of O4I4 was found to significantly induce the mRNA levels of *ID1* and *ID2* (**Figure 24B**). In addition to this, immunoblotting showed a clear increase in Smad1/5/8 C-tail phosphorylation, as well as ID2 protein levels upon CSKML treatment (**Figure 24C**). In support of this, ectopic expression of *ID1/2* appeared to activate HMGA1 and to cause a slight increase in *OCT4* mRNA expression, an effect which was clearly enhanced in combination with O4I3/4 or CSKML (**Figure 24D**).

To further support the involvement of BMP signaling in CSKML's mode of action, we sought of suppressing the signaling using (i) the BMP receptor inhibitor (dorsomorphin; designated as DM), (ii) the BMP antagonist (noggin), and (iii) siRNA-mediated knockdown of *Smad4* and monitoring the effects of CSKML. Promising data obtained from immunoblotting showed that inhibiting the BMP pathway not only abrogates Smad1/5/8 receptor activation and ID2 expression but is also capable of counteracting CSKML-mediated HMGA1- and OCT4 induction (**Figure 24C)**. Similarly, knocking down *ID1/2* using siRNA resulted in the inhibition of CSKML-mediated HMGA1 and OCT4 activation, as determined by immunoblotting and qPCR analyses, respectively (**Figure 24E and 24F**). In contrast to its genetic inhibition, *ID1/2* ectopic expression caused a profound increase in *HMGA1* and *OCT4* mRNA levels (**Figure 24D**). As expected, the addition of BMP6 was found to promote *OCT4* mRNA expression (**Figure 24F)**. Consistent with the above data, BMP6 addition and/or *ID1/2* forced expression was able to increase the reprogramming efficiency, as indicated by a higher number of TRA-1-60-positive colonies (**Figure 24G**).

Moreover, we also tested the effects of other members of the BMP family, namely BMP2 and BMP9. The former showed similar effects to those of BMP6, as indicated by increased C-terminal phosphorylation of Smad1/5/8, as well as the induction of HMGA1- and OCT4 protein levels in CSKML-treated fibroblasts (**Figure 25**). BMP9, on the other hand, failed to promote the expression of HMGA1 and OCT4 (**Figure 25**). ***Altogether, the above results demonstrate the activation of endogenous OCT4 by CSKML through the BMP-Smad-ID-HMGA1 signaling axis.***

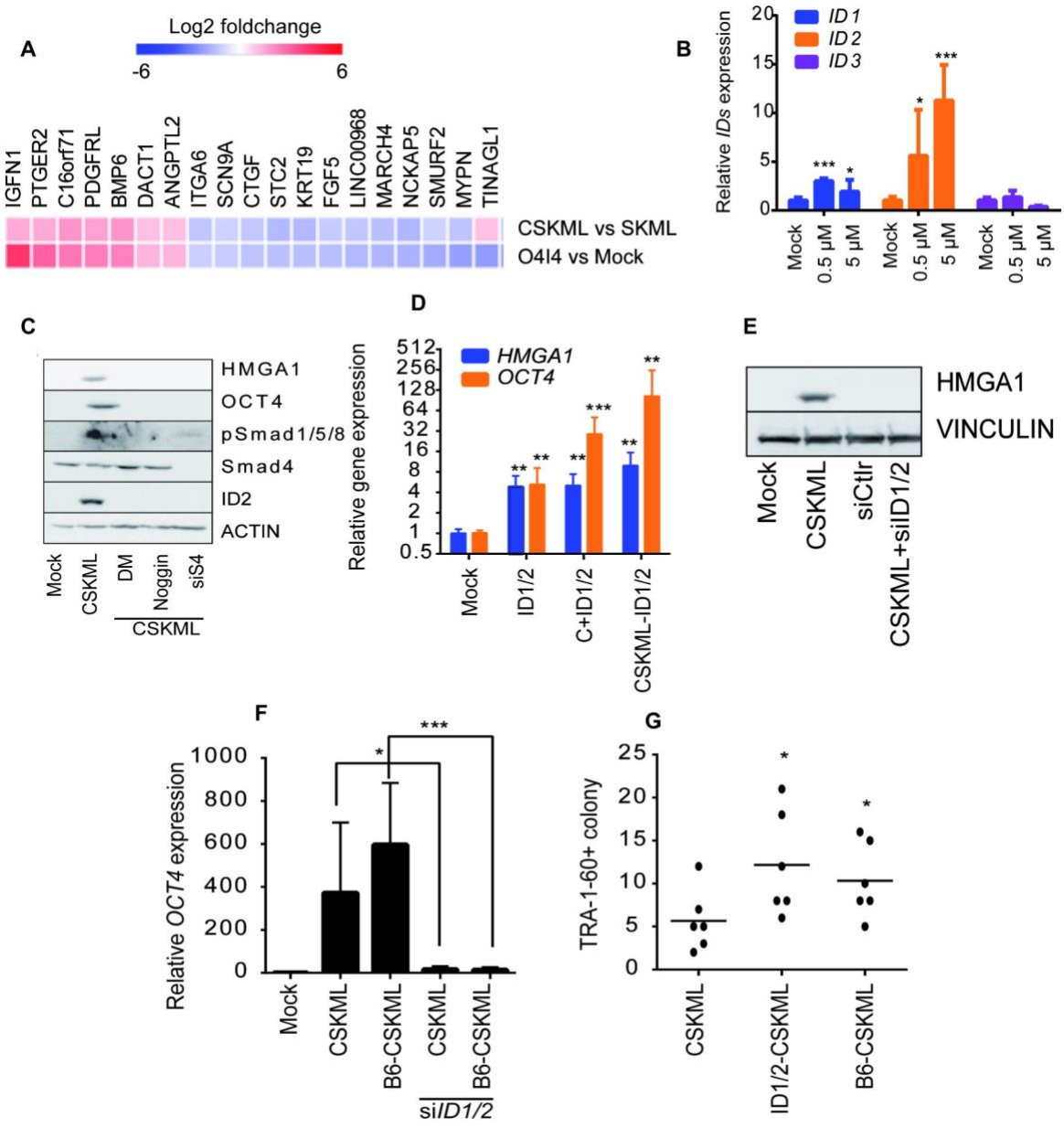


Figure 24. HMGA1 activation by CSKML involves BMP/Smad/ID signaling. (A) RNA-seq analysis of up- and downregulated genes in CSKML- *vs.* SKML treatments, as well as in mock (0.1% DMSO) *vs.* O4I4 treatments. Human fibroblasts were transfected with SKML alone or in combination with 10 μM of O4I4 (CSKML) for 48 h. In the other group, HFs were treated either with 0.1% DMSO or O4I4 (10 μM, 48 h). **(B)** O4I4 induces mRNA levels of *ID* genes in HFs treated with the indicated concentrations of the molecule for 48 h. **(C)** Immunoblots showing the activation of the BMP/Smad/ID pathway by CSKML treatment, as determined by an increase in the protein levels of receptor-phosphorylated Smad1/5/8 and ID2. This was found to be counteracted in the presence of 5 μM of BMP receptor inhibitor (dorsomorphin; designated as DM) and 200 ng/mL of the BMP antagonist, noggin, as well as the siRNA-mediated suppression of the common Smad, *Smad4*. Treatment with the chemical inhibitors and siRNA transfection were done together with CSKML treatment for 48 h. Vinculin was used as the loading control. **(D)** Forced expression of the BMP downstream genes, *ID1/2* enhances the mRNA expression of *HMGA1* and *OCT4* in the absence or presence of O4I3/4 (denoted as C) and O4I3/4-SKML (CSKML) combination, analyzed by qRT-PCR. HF1 cells were transfected with either *ID1/2* over-expression plasmids alone or in combination with the chemicals and the SKML cocktail for 48 h. **(E)** siRNA-mediated repression of *ID1/2* abrogates CSKML-induced HMGA1, as determined by immunoblotting. A non-

targeting siRNA (designated as siCtrl) served as the negative control. Vinculin was used as the loading control. **(F)** The addition of BMP6 (denoted as B6) in the first 72 h of CSKML treatment further increases CKML-induced *OCT4* mRNA expression, an effect which is antagonized by genetic inhibition of *ID1/2*. Relative OCT4 gene expression was detected by qRT-PCR. **(G)** Reprogramming efficiency of CSKML is enhanced in the presence of BMP6 or by over-expressing its targets, *ID1/2*, as identified by TRA-1-60 staining. In **(B)**, **(D)**, and **(F)** relative mRNA levels were determined by the ΔΔ Ct method in which the Ct values of the target genes were normalized to those of the housekeeping gene (βactin). Error bars ± SD, *n* = 3. 0.1% DMSO served as mock treatment. In **(B)** and **(D)** statistical significance was made respective to mock using a two-tailed student's *t*-test and in **(F)** multiple comparisons were performed using a one-way ANOVA test followed by a post-hoc Tukey test. *, **, and *** denote a *p*-value less than or equal to 0.05, 0.01, and 0.001, respectively.

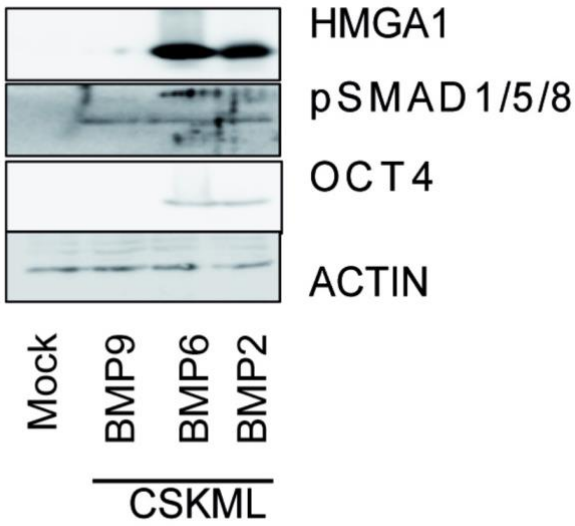


Figure 25. BMP2, but not BMP9, has a similar role to BMP6 in the context of CSKML-induced pluripotency. Immunoblots showing that BMP2 is able to promote HMGA1 and OCT4 protein expression in a comparable manner to that of BMP6. BMP9, on the other hand, failed to activate the HMGA1/OCT4 axis. 0.1% DMSO served as mock treatment. βactin was used as the loading control.

3. DISCUSSION

In parallel to the work presented in **chapter 1**, I had a chance to be majorly involved in the exciting field of *"cellular reprogramming"* where we focused on establishing new protocols for chemical conversion of human somatic cells into a pluripotent state. The ideas of this project came from previous reports of the group on the development of new chemicals for the induction of pluripotency factors [270, 290], as well as the lack of strong small molecule inducers of OCT4, which has been long challenging the generation of human chemical iPSCs. The results presented in this chapter introduce newly synthesized small molecules–herein O4I3 and O4I4–as potent activators of pluripotency-associated markers and strong enhancers of the transcription factor-mediated reprogramming. Based on my interest and experience in cellular signal transduction, I was particularly intrigued by the molecular mechanisms underpinning the reprogramming process, which are known to be mediated by an intricate crosstalk between several signaling modules including those of the TGFβ/BMP pathway. While investigating the mode of action of O4I3 and O4I4, we could identify previously uncharacterized molecular events during the process of reprogramming at both genetic- and epigenetic levels. In this section, I discuss the effects of the two small molecules in the context of OSKM-mediated reprogramming as well as their mode of action, which unraveled the significance of the epigenetic modifier, KDM5A (in case of O4I3), and the HMGA1/BMP/Smad axis (in case of O4I4) in the path towards pluripotency.

3.1. New chemical enhancers of transcription factor-mediated reprogramming: *from inducing pluripotency-associated genes to a profound increase in reprogramming efficiency*

Transcription factor-mediated reprogramming–first reported by Yamanaka *et al.*–offers a simple approach, i.e., forced expression of a defined set of pluripotency factors in the somatic cell, for the generation of induced pluripotent stem cells (***reviewed by Takahashi K. and Yamanaka S. in*** [291]). However, as thoroughly described before (**also see sections 1.1. and 1.2.**), this technology is associated with several disadvantages, such as the extremely low efficiency and safety concerns regarding the use of such stem cells (produced by viral transduction) for therapeutic purposes [292]. In view of this, many research groups have focused on the generation of iPSCs using small molecules, a process which is well-known as *"chemical (or pharmacological) reprogramming" (**reviewed in*** [269]). Looking for small molecule inducers of pluripotency, we screened a library of ~250,000 chemicals using cell-based assays and identified a series of OCT4 inducing compounds [270, 290]. Among these molecules were the analogues of imidazopyridine (***we have recently reported the results in*** [293]), which were found to be particularly interesting, as they share the imidazopyridine backbone with the FDA-approved sedative-hypnotic drug, zolpidem. Of those, a promising candidate–herein O4I3–was selected and synthesized. Further investigations revealed the ability of O4I3 in the induction of pluripotency-associated markers (particularly OCT4), as well as in the maintenance of pluripotency through repressing differentiation of iPSCs (**see Figure 6, Figure 7, and Figure 8**). Intrigued by these results, we thought of testing O4I3's effects on OSKM-mediated reprogramming. Promising data

showed that O4I3 increases the reprogramming efficiency (from 0.01% to 0.1%) and enables the conversion of reprogramming-resistant fibroblasts to iPSCs when combined with ectopic expression of the *"OSKM"* cocktail (**see Figure 9, Figure 10, and Figure 11**).

Although O4I3 treatment caused a profound induction of OCT4 in HEK-293 and NCCIT reporter cell lines as well as in fibroblasts (**see Figure 6, Figure 7, and Figure 9**), it failed to replace exogenous OCT4 in the context of OSKM-mediated reprogramming. Chemical replacement of *OCT4*–more than the rest of the pluripotency genes–has been long a challenge in establishing reprogramming cocktails for the generation of chemical iPSCs. This barrier was finally overcome by the work of Hou *et al.* who showed that all four exogenous *"master genes"*, including *OCT4*, are dispensable for the generation of mouse pluripotent stem cells, using a combination of seven small molecules with forskolin (a cAMP agonist) as the OCT4 inducer [262]. In addition to forskolin, we and others have reported on a few number of molecules able to promote endogenous OCT4 expression in mouse- and human somatic cells, namely pyrrolo[2,3-b]pyridine based OCT4-activating compound 1 (OAC1) [294], 4-(benzyloxy)phenyl derivatives (O4I1) [270], and 2-aminothiazoles (O4I2) [290]. Nonetheless, all these chemicals failed to activate OCT4 to a level sufficient for the generation of human iPSCs in the absence of ectopic *OCT4* expression. In view of this challenge, we sought to chemically modify our previously identified hits [290] in order to generate new small molecules, which are capable of substituting exogenous OCT4. The focus was drawn to O4I2 and its analogues, which have previously shown promising results with regards to OCT4 activation. Although being a strong OCT4 inducer in the HEK-293 reporter cell line, it was observed that this activity is lost in NCCIT-OCT4 cells. We have previously reported that O4I2 undergoes acid hydrolysis, producing a carboxylic acid product (O4I2-COOH), which was found to be inactive in the HEK-OCT4 reporter cells [290]. This sparked the idea of *"metabolic instability"* being responsible for abrogating the OCT4-inducing ability of O4I2, and hence the potential capability of *"metabolically-stable"* analogues to recover the small molecule's pluripotency-promoting effects.

Among the enzymes catalyzing the "ester-to-acid" conversion is carboxylesterase 1, which is well-known for metabolism and biotransformation of drugs [295]. In view of this, we sought of analyzing CES1-5A expression pattern, using previously published gene expression data and found high levels of these enzymes in PSCs and NCCIT cells. This finding not only fits with our hypothesis regarding metabolic instability underlying inactivation of O4I2 in NCCIT cells, but also suggests that enzymes produced by partially- or fully-reprogrammed cells might contribute to destabilization of chemicals, which in turn reduces the efficiency of chemical reprogramming. In support of this, replacing the ethyl ester moiety with CES1-resistant groups recovered O4I2-mediated OCT4 induction in NCCIT reporter cells (**see Figure 5**). In addition to this, the most active derivative of the stable compounds–herein O4I4–was found to strongly induce pluripotency-associated signaling networks in human fibroblasts, highlighting the significance of metabolic stability of chemicals for a successful reprogramming process.

In light of the promising results from O4I3, which showed a significant increase in reprogramming efficiency through activating endogenous OCT4, on the one hand, and the ability of O4I4 in promoting OCT4-associated signaling, on the other hand, we thought of using a combination of both chemicals (O4I3/4) in the context of OSKM-mediated reprogramming. Promising data from fibroblasts treated with *"CSKML"* (in which C stands for the combination of O4I3/4) demonstrated a clear increase in reprogramming efficiency when compared with that O4I4-only treated cells. Importantly, iPSCs generated by CSKML were found to resemble those obtained from OSKML-mediated reprogramming in terms of global gene expression patterns, epigenetic status, as well as their potential for differentiation (**see Figure 20 and Figure 21**). As implicated earlier, reprogramming protocols for the generation of chemically-induced human iPSCs are yet to be established, however, several chemical cocktails have been shown to successfully convert human somatic cells to iPSCs when combined with exogenous OCT4 expression [257]. We, however, failed to produce human chemical pluripotent stem cells using a combination of O4I3/4 with the previously reported chemicals, suggesting that further optimizations (e.g., dosages and duration of the treatment) of the chemical cocktails are necessary.

3.2. Understanding molecular mechanisms of somatic cell reprogramming: *new perspectives from the mode of action of OCT4-inducing compounds*

3.2.1. Is the histone demethylase enzyme KDM5A an obstacle in the path to pluripotency?

An exciting result in the investigations of this chapter is the identification of KDM5A as an epigenetic barrier during the course of reprogramming. As thoroughly described in the introduction (**also see section 1.1.1.2.**), transcription factor-induced pluripotency requires coordinated regulation of different epigenetic pathways (***summarized in*** [278]). Such epigenetic events during the transition between differentiated- and pluripotent cells can be summarized in three main categories, (i) genomic demethylation (both DNA and histones), (ii) histone acetylation, and (iii) reduced levels of heterochromatin, altogether contributing to an *"open chromatin"* state. Among several epigenetic regulators involved in the control of chromatin landscape during the process of reprogramming are histone methyl transferases and their countering enzymes, histone demethylases, which determine the type of histone marks, and hence the access and efficiency of the transcriptional machinery (***thoroughly discussed in two excellent reviews;*** [212, 296]). In view of this, some of the major histone modifications have been identified, with H3K4Me3 being indicative of active transcription, whereas H3K9Me3 and H3K27Me3 acting as repressive marks [212]. Through investigations into the molecular mechanism of O4I3, we observed a global increase of H3K4Me3 upon treatment with the small molecule as well as a significant occupation of this active histone modification at the promoter of *OCT4*. In support of this, it has been shown that at the initial phases of reprogramming, OCT4, SOX2, and KLF4–known as the "pioneer pluripotency factors"–bind to the nucleosomal pluripotency regions to enhance H3K4 methylation at the enhancer and promoter sites. This will then pave the way for the activation of those genes that are necessary for the establishment of a pluripotent state [228, 297]. In addition to this, genome-wide methylation analyses have revealed the gain of H3K4Me3 at promoters of pluripotency-

associated genes during the process of somatic cell reprogramming [254]. On the basis of these observations, it is easy to assume that chemicals able to protect H3K4 methylation could lift this epigenetic barrier and enhance the reprogramming efficiency.

Histone methylation is a reversible process, meaning that one or more of the methyl groups existing on a particular lysine residue could be erased through the action of histone demethylases (HDMs), for example KDM5s and LSD1 in case of H3K4 (***reviewed in*** [220]). This sparked the idea of O4I3 acting through the inhibition of specific HDMs, which could explain the increased levels of methylated H3K4 upon treatment with the small molecule. Focusing on HDMs related to H3K4 demethylation, we found a potent inhibition of KDM5, in particular the isozyme KDM5A, using an *in vitro* demethylase assay, an effect which was also observed in case of zolpidem, however, at higher concentrations. Follow up experiments revealed that both pharmacological inhibition of KDM5A (using O4I3 or a known inhibitor, JIB-04) and siRNA-mediated knockdown are able to promote the expression of several pluripotency markers at the mRNA level as well as in reporter NCCIT and HEK-293 cell lines. Additionally, chemical and genetic repression of KDM5A increased the efficiency of OSKM-mediated reprogramming (assessed by TRA-1-60 staining), mimicking O4I3 treatment, whereas its over-expression abrogated the compound's effects. In light of our results, a number of chemical epigenetic modifiers have been reported to facilitate the reprogramming process of human somatic cells, for example the small molecule inhibitor of LSD1, parnate [260], histone deacetylase inhibitor, NaB [268], and the DNA methyltransferase inhibitor, RG108 [252] (**also see** **section 1.2.1.**). In particular, parnate-induced pluripotency could be explained through the inhibition of H3K9 demethylation, which may block the expression of differentiation-associated genes, on the one hand, as well as its protective effect on H3K4 methylation, and hence activation of pluripotency genes, on the other hand [298, 299]. In case of O4I3 (and zolpidem), we could not observe a significant inhibitory function on LSD1, a histone demethylase, which is–unlike KDM5s–shared between H3K4 and H3K9 marks [298]. This further supports our hypothesis with regards to the role of H3K4 methylation in imidazopyridine-induced pluripotency.

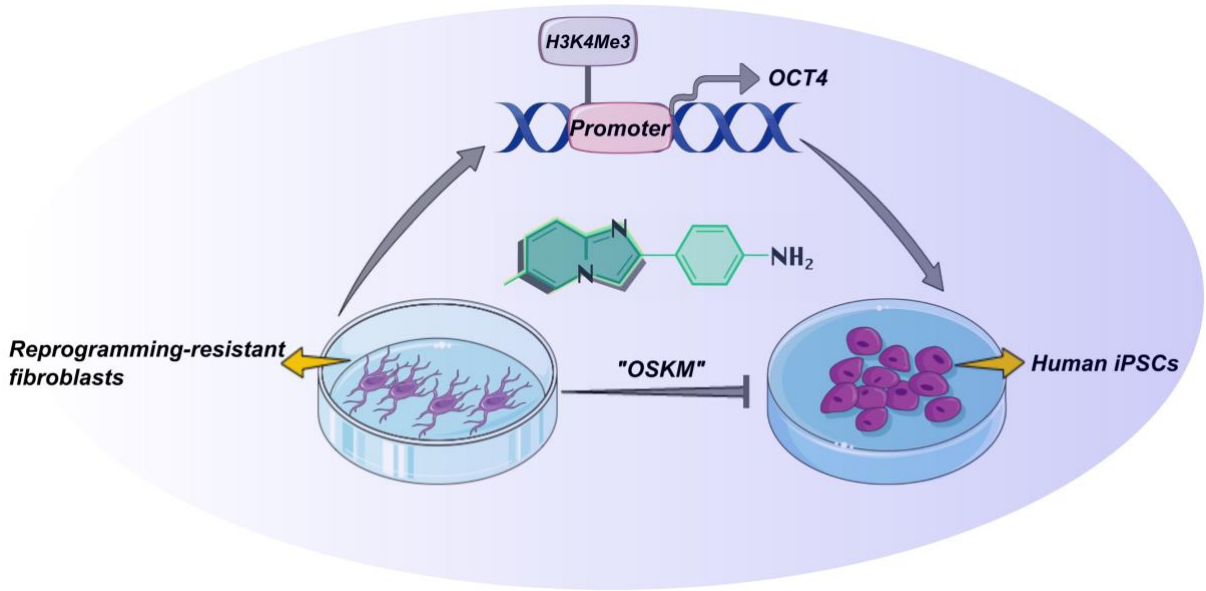


Figure 26. Imidazopyridine analogues as chemical enhancers of OSKM-based somatic cell reprogramming. A scheme showing the ability of OCT-4 inducing compound 3–O4I3– in converting the reprogramming-resistant human fibroblasts to iPSCs when combined with *"OSKM"* forced expression. We propose that inhibition of the histone demethylase enzyme, KDM5A by the small molecule increases the active histone mark, methylated H3K4 at *OCT4* promoter, and hence lifts an important epigenetic barrier of cellular reprogramming. The imidazopyridine backbone has been highlighted in gray in the chemical structure of O4I3. The presented image is modified from [293].

The role of enzymes regulating trimethylation of H3K4 in stem cell self-renewal and differentiation has been incompletely understood. A number of reports have shown the involvement of KDM5B in the aforementioned processes [300, 301], for example Xie *et al.* reported the positive regulation of self-renewal-associated gene expression by KDM5B, which is correlated with methylated H3K36, an active transcriptional mark [301]. This appears to be contrasting the conventional role of KDM5B as a transcriptional repressor, due to its capability of removing promoter-associated H3K4Me3 (a histone mark normally indicative of active transcription) [298]. The example of KDM5B shows that the contribution of enzymes responsible for regulating histone methylation levels is highly dependent on their substrates, and hence could lead to various and even opposing effects in different contexts. We here show that KDM5A is highly expressed in fibroblasts and is decreased during the course of cellular reprogramming (**see Figure 16**). Consistent with this, suppression of KDM5A either genetically or pharmacologically mimicked O4I3's effects in the induction of pluripotency markers and improving reprogramming efficiency, whereas its forced expression was found to have an opposite outcome, altogether suggesting a negative role for KDM5A in the path towards pluripotency (**summarized in Figure 26**). In contrast to our results, de-repression of KDM5A using miRNA 212/132 has been reported to enhance reprogramming of mouse embryonic fibroblasts [302]. One way to explain this discrepancy could be through the differences between mouse- and human pluripotent states (***reviewed in*** [303]). Additionally, the identification of epigenetic pathways (e.g., various histone marks) by which KDM5A leads to an increase or a decrease in the reprogramming efficiency is crucial for understanding its precise role in somatic cell reprogramming.

3.2.2. The *"BMP-Smad-HMGA1"* axis as a positive modulator of somatic cell reprogramming

High mobility group proteins A1–coded by *HMGA1*–function as architectural chromatin-binding proteins, promoting the expression of various genes [304, 305]. In the context of somatic cell reprogramming, HMGA1 is known to have high expression levels in human embryonic stem cells and fully-reprogrammed iPSCs, while showing low expression levels in terminally-differentiated cells and fibroblasts (***recently reviewed in*** [306]). During the course of transcription factor-mediated reprogramming, HMGA1 has been shown to synergize with OSKM factors in enhancing the reprogramming efficiency through binding to the promoters of pluripotency genes, *SOX2*, *LIN28*, and *c-MYC* and inducing their expression [286]. Our investigations in the mode of action of O4I3/4 revealed that chemically-induced HMGA1 could reactivate endogenous OCT4 to initiate the reprogramming process. This suggests an important role for HMGA1 as a driver of the pluripotent state via activating self-renewal-associated networks.

Further investigations comparing the differentially-expressed genes between *"CSKML"* and *"OSKML"*-treated fibroblasts (**see Figure 24**) illustrated potential upstream pathways responsible for the activation of HMGA1. In particular, we could show that the BMP/Smad axis is indispensable for HMGA1-mediated OCT4 activation induced by CSKML. The BMP pathway has been previously reported to enhance both transcription factor- and chemically-induced reprogramming [307]. For example, BMPs were able to replace KLF4 in the context of OSKM-mediated reprogramming of mouse embryonic fibroblasts, to increase the efficiency of OCT4/SOX2-mediated reprogramming of MEFs to ~1%, and finally to promote OCT4-only-based reprogramming [289]. More recently, Yamanaka *et al.* reported that the BMP/Smad/ID pathway inhibits the P16-meidated senescence, and hence overcomes a major barrier during reprogramming [288]. Accordingly, interruption of the pathway via dorsomorphin and LDN1931890 or transducing inhibitory Smads (Smad6/7) was found to attenuate the positive effects on iPSC generation [288]. This is in line with our results, showing the countering effects of both chemical- and genetic inhibitors of BMP/Smad/ID signaling on HMGA1 and OCT4 activation as well as on TRA-1-60-positive colonies (**see Figure 24**). On the contrary, the addition of BMP6 and forced expression of *ID1/2* enhanced the reprogramming efficiency, as expected. In line with these data, it has been previously reported that adding BMP4 at early times (but not at later times) during the reprogramming process increases the efficiency of iPSC generation [288]. Altogether, our results further confirm the positive role of the BMP/Smad/ID pathway in the induction of pluripotency and suggest a previously unrecognized mechanism of this signaling in the context of OSKM-based reprogramming through the activation of HMGA1.

3.3. Concluding remarks and future perspectives

In this chapter, we aimed at enhancing the process of OSKM-based reprogramming using newly synthesized small molecules. For this, we initially performed a cell-based high throughput screen encompassing ~250,000 chemicals, which led to the identification of several lead structures as OCT4-inducing compounds. Of those, more than 4000 molecules were found to be analogues of imidazopyridine, with a similar backbone to the FDA-approved sedative-hypnotic drug, zolpidem. A promising candidate, herein O4I3, was then selected and synthesized. Comprehensive analyses of O4I3's cellular effects revealed that the molecule is capable of (i) inducing the pluripotency-associated networks and (ii) supporting the generation and maintenance of iPSCs through repressing differentiation. Investigating the impact of O4I3 on OSKM-mediated reprogramming, we observed (iii) a significant increase in reprogramming efficiency upon the addition of O4I3 to the OSKM cocktail. Of note, the above effects were found to be reproducible with zolpidem, further supporting our hypothesis with regards to the use of imidazopyridines as enhancers of somatic cell reprogramming.

Through investigations into the molecular mode of action of O4I3 using comparative gene expression analyses, we found that more than 30% of the similarly regulated genes between O4I3-treated HFs, OSKM-treated HFs and PSCs are involved in epigenetic modifications. This sparked the idea of O4I3 acting through the epigenetic pathways in order to promote the generation of iPSCs. Focusing on the

most well-known histone modifications, we observed a profound increase in trimethylated H3K4 levels at the promoter of *OCT4* and globally in response to O4I3. The methylation status of specific histone marks is determined via the countering action of histone methyl transferases and histone demethylases, for example KDM5s and LSD1 in case of H3K4. In light of this, we found a (iv) potent inhibition of the histone demethylase enzyme, KDM5A upon O4I3 treatment *in vitro* and in cells, an effect which was also observed in case of zolpidem. Additional experiments further confirmed (v) the role of KDM5A as an epigenetic barrier of reprogramming, as its genetic/chemical suppression mimicked O4I3's effects in inducing pluripotency genes and increasing the OSKM-based reprogramming efficiency. Despite this, ***there are several questions remaining to be addressed with regards to the impact of O4I3 on KDM5A and its relevance to the induction of pluripotency. For example, has O4I3 any significant effect on epigenetic marks other than H3K4 as well as their regulating enzymes? and what is the precise mode of interaction of O4I3 with KDM5 demethylases? These issues, among others, present interesting topics for future projects.***

Although inducing OCT4 expression significantly, O4I3 was not able to substitute exogenous OCT4 during the course of reprogramming, a long-lasting challenge in the generation of chemically-induced human iPSCs. This encouraged us to test different combinations of small molecules by which exogenous OCT4 could be replaced. After several attempts with our initial hits from the screen, we selected the most potent metabolically-active derivative of O4I2, herein O4I4, for additional experiments. (vi) Using a combination of O4I3, O4I4 (designated as O4I3/4), and the ectopic expression of *"SKML"* (altogether referred to as "CSKML"), we were able to generate partially-reprogrammed iPSCs without the need of *OCT4* transduction. ***This finding, although promising, requires further experimental support, particularly with regards to the characterization of the newly-reprogrammed iPSCs, including the long-term expansion capacity, clone formation assay after passaging, colony morphology, and teratoma assays.***

Our mechanistic analyses of the O4I3/4's mode of action in human somatic cell reprogramming revealed (vii) the involvement of BMP/Smad signaling upstream of HMGA1, which in turn activates endogenous OCT4 and initiates the reprogramming process. We could show that over-expression of *HMGA1* together with *"SKML"* (herein referred to as *"HSMKL" iPSCs*) or the addition of BMP6 and the downstream signaling molecules, ID1/2 increases the reprogramming efficiency in a similar manner to that of "CSKML" treatment. ***There remained unanswered questions related to the signaling program triggered by O4I3/4, for example if the BMP/Smad/HMGA1 axis can replace the full role of the small molecules in the generation of iPSCs.*** Nonetheless, our findings provide new insights into the molecular events during reprogramming and present new chemical tools able to enhance the transcription factor-based reprogramming of human somatic cells.

4. MATERIALS AND METHODS

4.1. Materials

4.1.1. Consumables

Table 1. List of the general consumables used in the described experiments of chapter 1 and chapter 2.

Item	Supplier
96-well plates (round bottom)	Greiner Bio-One, Germany
Accutase®	Sigma-Aldrich, Germany
BD FACSFlow™	Becton Dickinson, Germany
BDTM FACSclean	Becton Dickinson, Germany
Cell culture flasks (T-25, -75, -175)	Greiner Bio-One, Germany
Cell culture petri-dishes	Greiner Bio-One, Germany
Cell culture plates (6-, 12-, 24-, 48-, and 96-well plates)	Greiner Bio-One, Germany
DMEM-F12 Knockout medium	Gibco, Germany
Dulbecco's Modified Eagle Medium GlutaMAXTM	Gibco, Germany
Dulbecco's Phosphate-Buffered Saline (D-PBS)	Gibco, Germany
Eppendorf tubes (1.5 and 2.0 mL)	Greiner Bio-One, Germany
Essential® 8 medium	Thermo Fisher Scientific, Germany
FACS tubes	Becton Dickinson, Germany
Falcon/Centrifuge tubes	Greiner Bio-One, Germany
Fetal Calf Serum (FCS)	Gibco, Germany
Opti-MEM Reduced Serum Medium	Gibco, Germany
PageRulerTM Prestained Protein Ladder	Nippon Genetics Europe GmbH, Germany
Parafilm® M	Sigma-Aldrich, Germany
Penicillin-streptomycin	Gibco, Germany
Polyvinylidene fluoride (PVDF) membranes	GE Healthcare, Germany
Ponceau S	Sigma-Aldrich, Germany
RPMI 1640-GlutaMAXTM	Gibco, Germany
Serological pipettes (2.0, 5.0, 10.0, 25.0 mL)	Costar corning incorporated
Syringe filters (0.2 µm pore size)	GE Healthcare, Germany
Syringes (50 mL)	Becton Dickinson, Germany
TrypLETM Express	Gibco, Germany
WhatmanTM 3mm CHR (45x57 cm)	GE Healthcare, Germany
White 96-well microplate (clear bottom)	Costar corning incorporated

4.1.2. Chemicals, drugs, and cytokines

Table 2. List of chemicals, growth factors, and drugs used in the described experiments of chapter 1 and chapter 2.

Chemical	Supplier
1,4-Dithioththeriotol (DTT)	Carl Roth GmbH, Germany
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Carl Roth GmbH, Germany
Activin	Thermo Fisher Scientific, Germany
Alvocidib (Flavopiridol)	Selleckchem, Germany
Aprotinin	AppliChem, Germany
bFGF	Thermo Fisher Scientific, Germany
BMP4	Thermo Fisher Scientific, Germany
Bovine Serum Albumin (BSA)	Carl Roth GmbH, Germany
Bromophenol blue	AppliChem, Germany
$CaCl_2$	AppliChem, Germany
CHIR99021	Sigma-Aldrich, Germany
Chloroform	Zentralbereich Theoretikum Chemikalienlager, Heidelberg University.
CPI-455	Selleckchem, Germany
Cycloheximide (CHX)	Sigma-Aldrich, Germany
Daminozide	Selleckchem, Germany
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, Germany
Ethanol	VWR Chemicals, Germany
Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid (EGTA)	Sigma-Aldrich, Germany
Ethylenediaminetetraacetic acid (EDTA)	Merck, Germany
Glycerol	Sigma-Aldrich, Germany
Glycine	Sigma-Aldrich, Germany
GSK-J4	Selleckchem, Germany
IGEPAL-CA630	Sigma-Aldrich, Germany
Isopropanol	Zentralbereich Theoretikum Chemikalienlager, Heidelberg University.
JIB-04	Selleckchem, Germany
KCl	AppliChem, Germany
LY294002	Selleckchem, Germany
Methanol	Sigma-Aldrich, Germany
MG-132	Sigma-Aldrich, Germany
$MgCl_2$	Sigma-Aldrich, Germany
Milk	Carl Roth GmbH, Germany
Na_3VO_4	AppliChem, Germany
Na_4P2O_7	AppliChem, Germany
NaCl	AppliChem, Germany

NaF	AppliChem, Germany
Paraformaldehyde (PFA)	Sigma-Aldrich, Germany
PD0325901	Selleckchem, Germany
Pepstatin	AppliChem, Germany
Phenylmethane sulfonyl fluoride (PMSF)	AppliChem, Germany
Rotiphorese® Gel 30/ Acrylamide mix	Carl Roth GmbH, Germany
SB-431542	Selleckchem, Germany
SB203580	Selleckchem, Germany
Sodium dodecyl sulfate (SDS)	Merck, Germany
SP600125	Selleckchem, Germany
ß-mercaptoethanol	Merck, Germany
Tetramethylethylenediamine (TEMED)	Carl Roth GmbH, Germany
TGFβ1	Biomol, Germany
Trichloroacetic acid	VWR Chemicals, Germany
Tris	Carl Roth GmbH, Germany
Tris-HCl	Carl Roth GmbH, Germany
Triton X-100	Carl Roth GmbH, Germany
Tween 20	Carl Roth GmbH, Germany
U0126	Selleckchem, Germany
Urea	Carl Roth GmbH, Germany
Y-27632 (ROCK inhibitor)	Sigma-Aldrich, Germany
Zolpidem	Selleckchem, Germany

4.1.3. Commercially-available reagents and kits

Table 3. List of the commercially-available reagents and kits used in the described experiments of chapter 1 and chapter 2.

Reagent/Kit	Supplier
Annexin V/propidium iodide	BD Biosciences, Germany
Beetle juice kit	PJK, Germany
Bradford reagent	Sigma-Aldrich, Germany
Cellartis iPSC differentiation kit	Takara, Germany
Epigenase jarid demethylase activity/inhibition assay kit	Epigentek, Biocat, Germany
Epigenase LSD1 demethylase activity/inhibition assay kit	Epigentek, Biocat, Germany
EpiQuik histone demethylase (H3K4 specific) activity/inhibition assay kit	Epigentek, Biocat, Germany
Episomal OSKM (Epi5) expression kit	Thermo Fisher Scientific, Germany
EpiTect Bisulfite Kit	Qiagen, Germany
Geltrex	Thermo Fisher Scientific, Germany
Hoechst 33342	Sigma-Aldrich, Germany
LightCycler® 480 SYBR Green I Master	Roche, Germany

Lipofectamine® 3000	Thermo Fisher Scientific, Germany
Matrigel	Thermo Fisher Scientific, Germany
NucleoSpin® RNA plus	
Protein A/G magnetic beads	Thermo Fisher Scientific, Germany
ProtoScript® First Strand cDNA Synthesis Kit	New England Biolabs, Germany
QIAzol lysis reagent	Qiagen, Germany
SimpleChIP Plus Enzymatic Chromatin IP Kit	Cell Signaling Technology, Germany
SRB sodium salt	Santa Cruz Biotechnology, Germany
StemDiff differentiation kit	Stemcell Technologies, Germany
Streptavidin magnetic beads	Thermo Fisher Scientific, Germany
Thiazolyl blue tetrazolium bromide dye (MTT)	Sigma-Aldrich, Germany
TransNGS® Tn5 DNA library pre-Kit for Illumina	
Western lightning^{TM} plus-ECL	Perkin Elmer, Germany

4.1.4. Antibodies

Table 4. List of the primary/secondary antibodies used for immunoblotting, immunoprecipitation, flow cytometry, immunocytochemistry, and immunohistochemistry experiments described in chapter 1 and chapter 2.

Target protein	Catalogue No.	Supplier
E-Cadherin	#3195	Cell Signaling Technology, Germany
Erk1/2	SC-135900	Santa Cruz Biotechnology, Germany
Flag	#8146	Cell Signaling Technology, Germany
Goat anti-mouse Alexa Flor 594	#115-585-003	Dianova, Germany
Goat anti-rabbit Alexa Flor 488	#005-540-003	Dianova, Germany
H3K27Me3	#9733	Cell Signaling Technology, Germany
H3K4Me3	#9751	Cell Signaling Technology, Germany
H3K9Me3	#13969	Cell Signaling Technology, Germany
HMGA1	ab4078	Abcam, Germany
Horse reddish peroxidase (HRP)-linked anti-mouse IgG	#7076	Cell Signaling Technology, Germany
HRP-linked anti-rabbit IgG	#7074	Cell Signaling Technology, Germany

ID2	SC-398104	Santa Cruz Biotechnology, Germany
KDM5A	#3876	Cell Signaling Technology, Germany
KDM5B	#3273	Cell Signaling Technology, Germany
KDM5C	#5361	Cell Signaling Technology, Germany
KDM5D	PA5-40120	Thermo Fisher Scientific, Germany
LIN28	#3334-1	Epitomics, Germany
MDM2	#86934	Cell Signaling Technology, Germany
NANOG	#4893	Cell Signaling Technology, Germany
OCT4	#2840	Cell Signaling Technology, Germany
p53	SC-126	Santa Cruz Biotechnology, Germany
Phospho-Erk1/2 (T202/Y204)	#5683	Cell Signaling Technology, Germany
Phospho-Smad1/5/8	#9511	Cell Signaling Technology, Germany
Phospho-Smad2 (S465/467)	#12747 (Smad2/3 antibody sampler kit)	Cell Signaling Technology, Germany
Phospho-Smad3 (S204)	PA5-36877	Thermo Fisher Scientific, Germany
Phospho-Smad3 (S213)	ab63403	Abcam, Germany
Phospho-Smad3 (S423/425)	#12747 (Smad2/3 antibody sampler kit) ab52903	Cell Signaling Technology, Germany Abcam, Germany
Smad1/5/8	SC-6031-R	Santa Cruz Biotechnology, Germany
Smad2	#12747 (Smad2/3 antibody sampler kit)	Cell Signaling Technology, Germany
Smad2/3	#8685	Cell Signaling Technology, Germany
Smad3	#12747	Cell Signaling Technology, Germany

	(Smad2/3 antibody sampler kit) MA5-15663	Thermo Fisher Scientific, Germany
Smad4	#9515	Cell Signaling Technology, Germany
SOX2	#4900	Cell Signaling Technology, Germany
ßactin	SC-47778	Santa Cruz Biotechnology
ßcatenin	#8480	Cell Signaling Technology, Germany
TRA-1-60	#4746	Cell Signaling Technology, Germany
TRA-1-81	#4745	Cell Signaling Technology, Germany
VHL	ab140989	Abcam, Germany
Vinculin	SC-73614	Santa Cruz Biotechnology

4.1.5. Primers

Table 5. List of primer pairs used for qRT-PCR experiments of chapter 1 and chapter 2. All sequences were synthesized by Eurofins Genomics (Germany). Lyophilized primers were dissolved in nuclease-free water to a final concentration of 100 pmol/µL and were further diluted (1:20) for qRT-PCR analysis.

Target gene	Primer sequence (forward; 5'→3')	Primer sequence (reverse; 5'→3')
ACTB (Actin)	TGCTCCAGCAGGTCCTAAGT	TGGTAGCTGAGGGCAGAACT
AFP	CTGACAGAGGCACCACTGAA	ATCTCCAGAGTCCAGCACA
BMPR1A (ALK3)	ATCTAGCCACATCTTGGAGGA	ACTGGACTACACACCAAGGG
ACVR1B (ALK4)	ATGCTGCGCCATGAAAACAT	CCCCTTCCTCCCTTGTATTGG
RNF111 (Arkadia)	ACTTATAAAGGTCCCAGGATCAC	AATTTGTGCCTCACAGTATCTGC
c-Myc	TTCGGGTAGTGGAAAACCAG	CAGCAGCTCGAA TTTCTTCC
Cbl-b	CAGTGCCTGGAAGGATTTGTG	AGTTAAAACTGTGCCGCCTG
E-cadherin (CDH1)	GAGAGACTGGGTTATTCCTC	GATGCTGTAGAAAACCTTGCC
N-cadherin (CDH2)	GAAGATGGCATGGTGTATGC	CTCCTCAGTTAAGGTTGGCTT
CDKN1A (p21)	GACACCACTGGAGGGTGACT	CAGGTCCACATGGTCTTCCT
STUB1 (CHIP)	GGCCGCGCGATCACC	GAAGTGCGCCTTCACAGACT
CCN2 (CTGF)	CTATGAGTGGGACAGGAACACTGC	TAAGGGAAGATCAAACGCGAACGGC
DUSP14	CATGGCCCCTCGGATGATTT	AATGCAGGTGATGCCACGA
ESRRβ	CCTTCCAATCAGCTGCCTTC	GATCTGCCTGCCTCTCTCAT
FGF2	AGGAGTGTGTGCTAACCGTT	CAGTTCGTTTCAGTGCCACA
FGF4	TGAGTGCACGTTCAAGGAGA	GACACTCGGTTCCCCTTCTT

FGF5	AGTGGTATGTGGCCCTGAAT	CTGCTCCGACTGCTTGAATC
FoxA2	CTACTCCTCCGTGAGCAACA	GACGACATGTTCATGGAGCC
FOXH1	GGATATCCAGCCCCCATAGT	CTTCCCTGGTCAGGTTTCAGG
GATA4	TCCAAACCAGAAAATGGAAG	CTGTGCCCGTAGTGAGATGA
GATA6	CCACAACACAACCTACAGCC	ACGCCTATGTAGAGCCCATC
HMGA1	GAAAAGGACGGCACTGAGAA	CTTCCTGGAGTTGTGGTGGT
ID1	CATGAACGGCTGTTACTCAC	GTTCCAACTTCGGATTCCGAG
ID2	GACTGCTACTCCAAGCTCAAG	GTGATGCAGGCGACAATAGTG
ID3	CAGCTTAGCCAGGTGGAAATC	GTCGTTGGAGATGACAAGTTC
KDM5A	GCTCTGCAAAGACTGGTTCC	AGTCTCTAGCCTGGGCCTTC
KDM5B	TTTTGAGCTGTTGCCAGATG	TGATGCAGGCAAACAAGAAG
KLF4	CGCCGCTCCATTACCAAGA	GCGAATTTCCATCCACAGCC
LIN28A	GTGGATGTCTTTGTGCACCAG	GACACGGATGGATTCCAGAC
NANOG	AATACCTCAGCCTCCAGCAGATG	TGCGTCACACCATTGCTATTCTTC
NEDD4L	TCTCGGGCACCCTCACC	GTTCTCCCCTCTGCCTCCA
NES (Nestin)	TCCAGGAACGGAAAATCAAG	GCCTCCTCATCCCCTACTTC
NODAL	GTCGACATCACTTGCCAGAC	TGGTCGGATGAAACTCCTCC
POU5F1 (OCT4)	GAAGGATGTGGTCCGAGTGT	GTGAAGTGAGGGCTCCCATA
POU5F1 (OCT4) promoter1	GTTTGTGTGTATGCATGCCA	AAGCTGCTAAGTTCTGGGTT
POU5F1 (OCT4) promoter2	GGATATAGCACGGAGGCCTT	CACCCTCTCAGCTCCTCAAA
PAX6	GACTTCAGCTGAAGCGGAAG	GGTAGATCTATTTTGGCTGC
ZFP42 (REX1)	TGGACACGTCTGTGCTCTTC	GTCTTGGCGTCTTCTCGAAC
Runx2	CCACATGATTCTGCCTCTCC	GGAGAAAGTTTGCACCGCAC
SALL4	ATTTGTGGGACCCTCGACAT	GGTAAAAGCTCGCCCACAAA
SKIL (SnoN)	GTGCTGTTTTCCATTTTCATGCC	AGCTCACACACGATACTGAAA
Smad1	GTGCTGGTTCCAAGACACAG	ATTGGGAGAGTGAGGAAACG
Smad2	AGAAGTCAGCTGGTGGGTCT	TGAGTGGTGAGGCTTTCTC
Smad3	GGGCTTTGAGGCTGTCTACC	CAGAGTCACCTGGAGTTGGG
Smad4	TGCCATAGACAAGGTGGAG	AGCCTCCCATCCAATGTTC
Smad5	TTGTCAAGGTGAGTTGTGACC	CCTGCCGGTGATATTCTGCT
Smad7	CAGTTGAGTTGGCGTCTTCC	CAAGAAGGAGGGATGTCTGGG
Smurf1	GATACCGGATACCAGCGTTTG	GGTTCCTATTCTGTCTCGTG
Smurf2	ACAGAGGACAACGCAACAAG	ATTACGGATCTCCCATCAG
SNAI1 (SNAIL1)	CCTCTACTTCAGTCTCTTCC	CTTTCGAGCCTGGAGAATTCCTTG
SNAI2 (Slug)	CTCCTCTTTCCGGATACTCCTC	GTAGTCTTTCCTCTGAGCCAC
SOX17	CAGAATCCAGACCTGCACAA	GCGGCCGGTACTTGTAGTT
SOX2	AACCCCAAGATGCACAACTC	GCTTAGCCTCGTCGATGAAC
SOX7	ACCCCAACTACAAGTACCGG	TACTCACCCCTGTCCTCCTT
STAT3	GGCCATCTTGAGCACTAAGC	CGGACTGGATCTGGGTCTTA
TBXT (T)	ACCCAGTTCATAGCGGTGAC	GGATTGGGAGTACCCAGGTT
TBX3	CATCGCTGTGACTGCATACC	TCTCTTCGGCCATTTCCAGT
TP53	AATCTCAGAGCTTCCCGTCA	TGCCTCTGCCAGTTTTTCTT

TRIMM33	AGTGGCCATTTTCACCCTTATC	GCCTGTGATGTCATTCTGCTG
TUBB3 (β-tubulin)	CAGATGTTCGATGCCAAGAA	GGGATCCACTCCACGAAGTA
TWIST1	CAGATGTTCGATGCCAAGAA	GGGATCCACTCCACGAAGTA
USP34	GCCGAACGATGTGCGAGA	GCCTCTGTGTCCAGGAATTGAT
USP4	TGGGCTATACCAGGGTCAGG	GCTAGGCGCAGTGCTTGATT
USP9x	CCAGGACCGTGATACACAAA	GGCGCTTCACCAAACAAACT
VCL (vinculin)	CAGTCAGACCCTTACTCAGTG	CAGCCTCATCGAAGGTAAGGA
ZNF534	GGGGAAGCATTGTTCAGAAA	TGGGATGGGGGTATTACAGA

4.1.6. Cell lines

Table 6. List of cell lines used in the described experiments of chapter 1 and chapter 2.

Cell line	Supplier
CCC5	Generated from patient tissue by the group of Prof. Dr. Steven Dooley (Heidelberg University, Germany).
CCSW1	Generated from patient tissue by the group of Prof. Dr. Steven Dooley (Heidelberg University, Germany).
DU-145	ATCC, USA
GBC	Generated from patient tissue by the group of Prof. Dr. Steven Dooley (Heidelberg University, Germany).
HCT116 p53-/-	Kind gift from Prof. Dr. Thomas G. Hofmann (DKFZ, Heidelberg, Germany), originally introduced by Bunz *et al.* [308].
HCT116 wt	ATCC, USA
HEK-293	ATCC, USA
HEK-293/NCCIT reporter cell lines	Established by the group of Prof. Dr. James Adjaye (Heinrich Heine University Düsseldorf, Düsseldorf, Germany).
HeLa	ATCC, USA
HEP3B	Kind gift from Prof. Dr. med. Thomas Longerich (Heidelberg University, Germany).
HEPG2	ATCC, USA
HL-60	ATCC, USA
HT-29	ATCC, USA
HUCCT1	Generated from patient tissue by the group of Prof. Dr. Steven Dooley (Heidelberg University, Germany).
Human foreskin fibroblasts (HFs)	Patient-derived HFs were provided by Prof. Dr. med. Jochen Utikal (DKFZ, Heidelberg, Germany).
Human pluripotent stem cells (hPSCs)	Human iPSCs were provided by Prof. Andreas Kurtz (Charité-Universitätsmedizin Berlin, Berlin, Germany)
Jurkat	ATCC, USA
K-562	ATCC, USA

LX-2	Kind gift from Prof. Dr. Steven Dooley (Heidelberg University, Germany).
MCF-7	ATCC, USA
MMNK1	Generated from patient tissue by the group of Prof. Dr. Steven Dooley (Heidelberg University, Germany).
NCCIT	ATCC, USA
PC3	ATCC, USA
SUPT1	ATCC, USA
U-937	ATCC, USA

4.1.7. Instruments

Table 7. List of the instruments utilized to perform the described experiments of chapter 1 and chapter 2.

Instrument	Supplier
AccuBlock Tm Digital Dry Bath	Labnet International, Inc., USA
Analytical balance 440-33N	Kern und Sohn, Germany
Biorevo fluorescence microscope BZ9000	Keyence, USA
BlueFlash-L	SERVA Electrophoresis, Germany
Centrifuge 5415R	Eppendorf, Germany
Centrifuge 5702	Eppendorf, Germany
Centrifuge 5804	Eppendorf, Germany
EnduroTM Power Supplies	Labnet International, Inc., USA
FACS Calibur	Becton Dickinson, Germany
Guava easyCyte HT sampling flow cytometer	Guava Technologies, USA
Hemocytometer	Brand, Germany
HERA cell 150i incubator	Thermo Fisher Scientific, Germany
LAS3000 Imaging System	Fujifilm, Germany
Magnetic rack	Nippon Genetics Europe GmbH, Germany
Micropipettes (10, 20, 100, 200, 1000 µL)	Eppendorf, Germany
Microscope CKX41	Olympus Life Science, Germany
Miltidrop Combi	Thermo Fisher Scientific, Germany
Mixer mill MM 300	Retsch, Germany
Multichannel pipettes (100 and 200 µL)	Eppendorf, Germany
NanoDrop 2000 spectrophotometer	Thermo Fisher Scientific, Germany
Orbital shaker/Belly Dancer	Stovall Life Science Inc., USA
Pipetboy	Integra Biosciences, Germany
Power Station 200	Labnet International, Inc., USA
Problot hybridization oven	Labnet International, Inc., USA
Real-time thermal cycler qTower	Analytik Jena AG, Germany
Roller mixer SRT6	Stuart, Germany
Spectrafuge Mini	Labnet International, Inc., USA

Tecan Infinite F200 Pro	Tecan, Germany
Tecan Ultra plate reader	Tecan, Germany
Thermomixer comfort	Eppendorf, Germany
Vacuum pump	ZM Engineering, Germany
VIAFLO ASSIST automated multi-channel pipettes	Integra Biosciences, Germany
Vortex-Genie®2	Scientific Industries, Inc., USA
Water bath	GFL, Germany

4.1.8. Buffers and solutions

Table 8. List of buffers/solutions and the corresponding recipes used in the described experiments of chapter 1 and chapter 2. If not listed in the presented table, the recipe has been indicated in the respective method section (**also see section 4.2.**).

Buffer/solution	Component	Concentration
3x Gel loading buffer	Tris HCl (pH = 6.8, 25°C)	187.5 mM
	SDS	6% (w/v)
	Glycerol	30% (w/v)
	DTT	150 mM
	Bromophenol blue	0.03% (w/v)
	(in distilled water)	
5x Gel loading buffer	Tris HCl (pH = 6.8, 25°C)	10 mM
	SDS	4% (w/v)
	Glycerol	20% (w/v)
	Bromophenol blue	0.2% (w/v)
	ß-mercaptoethanol	200 mM
	(in distilled water)	
Annexin v binding buffer	HEPES (pH = 7.4)	10 mM
	NaCl	140 mM
	$CaCl_2$	2.5 mM
	BSA (pH = 7.4)	1% (w/v)
Anode buffer	Methanol	10% (w/v)
	Tris HCl	0.3 M
	(in distilled water)	
Blocking buffer	Goat serum	5% (w/v)
	BSA	1% (w/v)
	Triton X-100	0.3% (w/v)
	(in D-PBS)	
Cathode buffer	Methanol	10% (w/v)
	Tris HCl	25 mM
	(in distilled water)	
Luciferase lysis buffer	Tris phosphate	25 mM
	EGTA	4 mM

	Triton X-100 Glycerol (Filtered through 0.45 μM sterile filter) DTT (freshly added prior to usage)	1% (w/v) 10% (w/v) 2 mM
Running buffer for SDS-Polyacrylamide gel electrophoresis (PAGE)	Tris HCl Glycine SDS (in distilled water)	25 mM 250 mM 0.1% (w/v)
Stacking gel for Tris-glycine SDS-polyacrylamide gel (5%)	30% Acrylamide mix 1.0 M Tris (pH = 6.8) 10% SDS 10% Ammonium persulfate TEMED Distilled water	17% 12.5% 1% 1% 0.1% 68%
TBS/Tween	NaCl KCl Tris HCl Tween 20 (in distilled water)	8 g/L 0.2 g/L 3 g/L 0.1%
Tris-glycine SDS-Polyacrylamide gel (10%)	30% Acrylamide mix 1.5 M Tris (pH = 8.8) 10% SDS 10% Ammonium persulfate TEMED Distilled water	33% 25% 1% 1% 0.04% 40%

4.1.9. siRNAs

Table 9. List of siRNA oligonucleotide sequences used for transient knockdown experiments of chapter 1 and chapter 2. All small interfering RNA sequences were purchased from Thermo Fisher Scientific (Germany), unless otherwise stated in the text (**for more details regarding the oligo IDs, suppliers, as well as the transfection protocol refer to section 4.2.3.**).

Target gene	Sense sequence (5'→3')	Antisense sequence (5'→3')
HMGA1	AAUGUUCAUCCUCAUUGCCTT	GGCAAUGAGGAUGAACAUUTG
ID1	GGUAAACGUGCUGCUCUACTT	GUAGAGCAGCACGUUUACCTG
KDM5A	GGUCACCCGAAGACGAAAATT	UUUUCGUCUUCGGGUGACCTC
KDM5B	GGGCAUUAUGAACGAAUUCTT	GAAUUCGUUCAUAAUGCCCTC
KDM5C	GGUGGUUAUGAAGCUAUCUTT	AGAUAGCUUCAUAACCACCTT
KDM5D	GGUGGGAACGUGUCAUCAATT	UUGAUGACACGUUCCCACCTC
MDM2	GCCAUUGCUUUUGAAGUUATT	UAACUUCAAAAGCAAUGGCTT
PSMA3	GCUAUUGGAAUCAGAUGCATT	UGCAUCUGAUUCCAAUAGCTG

PSMA6	GGUCUACACACAGAAUGCUTT	AGCAUUCUGUGUGUAGACCTG
PSMD4	GGAUGCUGUCAACAUAGUUTT	AACUAUGUUGACAGCAUCCTG
RBBP6	GCGAUGGCAACUACAAAAGTT	CUUUUGUAGUUGCCAUCGCTG
VHL	GGAGCGCAUUGCACAUCAATT	UUGAUGUGCAAUGCGCUCCTG

4.1.10. Computer software

Table 10. List of the computer software used in preforming the described experiments of chapter 1 and chapter 2 and analyzing the associated data.

Software	Supplier
CellQuest Pro analysis software	Becton Dickinson, Germany
GraphPad PRISM	GraphPad Software Inc., USA
Image Reader LAS-3000	Fujifilm, Germany
ImageJ	National Institutes of Health, USA
Microsoft Excel	Microsoft, USA
NanoDrop 2000 software	Thermo Fisher Scientific, Germany
Inkscape	https://inkscape.org
Adobe Illustrator	Adobe Inc., USA
Photoshop	Adobe Inc., USA
GuavaSoft 3.1.1 software	Guava Technologies, USA
R 3.3.2 computing environment	R Core Team, Austria
qPCRsoft 3.1	Analytik Jena, Germany

4.2. Methods

4.2.1. Cell culture

Cell lines listed in **Table 6** were cultured in DMEM GlutaMAX™ supplemented with 10% (v/v) FCS and 1% (v/v) penicillin/streptomycin, with the exception of hPSCs and suspension cell lines, HL-60, U-937, K-562, and SUPT1. The latter were maintained in RPMI GlutaMAX™ containing 10% (v/v) and 1% (v/v) FCS and penicillin/streptomycin, respectively. The former as well as the induced pluripotent stem cells were cultivated in Essential® 8 medium or in DMEM-F12 Knockout medium with 20% knockout serum and 10 ng/mL bFGF in cell culture plates coated with Geltrex or Matrigel. All cells were kept in a standard cell culture incubator at 37°C under 5% CO_2 in a humidified atmosphere. Cells were passaged at 70-80% confluency. Adherent cell lines were sub-cultured by aspirating the old medium, washing once with D-PBS, then incubating with TrypLE™ Express for 3-8 mins depending on the cell line. For example, human primary cholangiocarcinoma cells (CCC5, CCSW1, HUCCT1, GBC, and MMNK1) require longer incubation periods (~8 mins) with TrypLE™ Express. Detached cells were then collected in standard DMEM and were centrifuged for 3 mins at 200-300 g. Medium on top of the cellular pellet was aspirated, and the cellular pellet was resuspended in fresh media. Medium containing a known cell number, depending on the culture flask used as well as the growth rate of each cell line, was then transferred back to the flask

containing fresh medium. Suspension cells were sub-cultured by collecting them from the flask and transferring them to a 15 mL falcon tube, after which the same protocol described above for the adherent cells was used. With regards to hPSCs, a similar protocol as the adherent cells was applied, however Accutase® was replaced with TrypLE™ Express for detaching the cells and D-PBS (instead of fresh medium) was used to harvest the detached cells. Additionally, the small molecule ROCK inhibitor was used in the first 24 h after passaging in order to support hPSC proliferation and viability [276]. After seeding at the desired density in designated culture plates, cells were maintained in the incubator overnight prior to treatments. All treatments were carried out in standard media with 10% FCS and 1% antibiotics or Essential® 8 medium supplemented with 10 ng/mL bFGF (in case of hPSCs), with the exception of TGFβ1 treatment, which was performed in the presence of 1% FCS. The concentrations of cytokines, small molecules, and drugs as well as the duration of treatments are indicated in the corresponding results' section.

4.2.2. Cell viability assays

MTT- and SRB cytotoxicity assays were employed to assess the impact of various treatments on cellular proliferation. Cells were seeded in 96-well- 24-well- or 12-well plates at a density of 10,000-50,000 or 150,000, in 150 µL- 500 µL- and 1 mL medium/well, respectively. On the following day, cells were treated as indicated in the respective results' section. For SRB assay, cells were fixed at the end of the treatment periods using 10% ice-cold trichloroacetic acid, followed by incubation at 4 °C for at least one hour. Plates were then washed three times and were dried using a hybridization oven set at 60°C for 30 mins. SRB solution (0.054 % (w/v) SRB in 1 % (v/v) acetic acid in water) was then added to the wells and incubated for 30 mins in dark at room temperature. Plates were then washed three times with 1 % acetic acid and were dried at 60 °C for 30 mins. Finally, the SRB dye was solubilized using 10 mM Tris buffer (pH = 10.5) and measurements were performed at 535 nm absorbance wavelength using the Tecan Ultra plate reader. For MTT assay, treatment medium was replaced with DMEM (1% FCS), containing 0.5 mg/mL of the MTT dye, followed by 1-2 h incubation at 37°C. Intracellular formazan pellets were then dissolved in DMSO, and absorbance was measured at 595 nm wavelength using the Tecan Ultra plate reader.

4.2.3. Transient transfection studies: ectopic gene expression and RNA interference

Small interfering RNA oligos against *KDM5A* (oligo ID: 107607), *KDM5B*, (oligo ID: 108065), *KDM5C* (oligo ID: 14315), *KDM5D* (oligo ID: 107493), *HMGA1* (oligo ID: 217139), *ID1* (oligo ID: 11072), *PSMA3* (oligo ID: 105604), *PSMA6* (oligo ID: 105047), *PSMD4* (oligo ID: 12139), *RBBP6* (oligo ID: 143012), *VHL* (oligo ID: 138745), *MDM2* (oligo ID: 122296) as well as the non-targeting siRNA were obtained from Thermo Fisher Scientific (Darmstadt, Germany). **For more information regarding the siRNA sequences refer to Table 9.** Anti-*Smad4* oligos were synthesized using previously published sequences [65] and riboxx siRNA-design (Riboxx, Radebeul, Germany). To knockdown the aforementioned genes, 40 pmol of the respective siRNA were diluted in 100 µL/well

Opti-MEM reduced serum medium in a 24-well plate. 1.5 µL/well Lipofectamine® 3000 were added, and the plate was left to incubate at room temperature for 10-20 mins (not more than 30 mins). 60,000 cells suspended in DMEM containing 10% FCS and no antibiotics were then added to each well and cells were maintained in the incubator overnight. Medium was changed to fresh DMEM (containing 10% FCS and 1% PS) on the following day and small molecule treatments were performed at 48 h post-transfection, otherwise cells were used directly for the described experiments. For transient over-expression experiments, the following constructs were used; Smad2 (Addgene plasmid #14042), Smad2 N (#14929), Smad2 ΔC (#12624), Smad2 EPSM (#14933), Smad3 (#14052), Smad3 ΔSSV (#12639) Smad3 N (#14827), Smad3 C (#14830), Smad3 LC (#14829), Smad3 EPSM (#14963), KDM5A (#14799), ID1 (#16061), and ID2 (#98394). HMGA1 was a kind gift from the group of Prof. Robert Hock (University of Würzburg, Germany) [309]. DNA-lipid complexes were prepared in 200/100 µL/well Opti-MEM reduced serum in a 6-well/24-well plate using 2.5/0.5 µg/well plasmid DNA and 5/1.5 µL/well of the transfection reagents, Lipofectamine® 3000 and P3000. Similar to siRNA-mediated knockdown, the mixture was left at room temperature for 10-20 mins (not more than 30 mins), after which cell suspension was added at a density of 800,000 cells/well (in case of 6-well plates) and 60,000 cells/well (in case of 24-well plates). 24 h later, medium was refreshed and at 48 h post-transfection, cells were either treated with small molecules or used directly for the respective experiments. In all plasmid transfections, the corresponding empty vector served as the negative control.

4.2.4. RNA extraction, reverse transcription, and qRT-PCR

As previously reported [310], after the described treatments, cells were harvested from 6-well plates in Eppendorf tubes using the QIAzol lysis reagent (150 µL for 5 mins at room temperature), followed by the addition of 30 µL chloroform, mixing for 15 secs, and placing on ice for 15 mins. Cells were then centrifuged at 13,000 g for 15 mins at a temperature of 4°C. The aqueous phase on top was then transferred to fresh tubes and RNA was precipitated by adding 75 µL isopropanol per tube and incubation for 10 mins at room temperature, followed by centrifugation under the same conditions. The RNA pellet was then washed twice with 70% (v/v) ethanol in water. After centrifugation (7500 g for 5 mins; at 4°C), RNA pellets were left to dry at room temperature, and were dissolved in 20-50 µL of RNAase-free water. Alternatively, total RNA was extracted using NulceoSpin®RNA Plus, according to the manufacturer's instruction. The quality of the resulting RNA samples was then determined by NanoDrop 2000 UV-Vis spectrophotometer.

cDNA synthesis was performed by reverse transcription of 500-1000 ng of total RNA using ProtoScript II first strand cDNA synthesis kit. The mRNA expression of the genes listed in **Table 5** was determined using real-time thermal cycler q-Tower. The amplification reaction solutions (5 µL) were prepared with 2.5× of ready to use master mix LightCycler® 480 SYBR Green I, 1× of nuclease-free H_2O and cDNA templates, as well as 1× of primer mixtures listed in **Table 5**. The thermal cycler

conditions used are as follows; an initial denaturation step at 95°C for 5 mins, 45 cycles of denaturation (95°C, 10 secs/cycle), annealing at 60°C for 10 secs, and a final elongation step at 72°C for 20 secs. Melting curve analysis was performed after each cycle in order to exclude wells with false positive results.

4.2.5. Immunoblotting

As we have thoroughly described [311], cells were lysed at the end of the respective treatments by 40-100 μL (depending on the cell number and desired protein concentration) of a urea-based lysis buffer (6M urea, 1 mM EDTA, 5 mM NaF, and 0.5% Triton X-100) supplemented with the following protease- and phosphatase inhibitors; 10 μg/mL pepstatin, 0.1 mM PMSF, 10 μg/mL aprotinin, 1 mM sodium orthovanadate, and 2.5 mM sodium pyrophosphate. Eppendorf tubes containing cellular lysates were then placed on ice for 10-20 mins, followed by centrifugation at 4°C for 20-30 mins at maximum speed. The supernatant was then transferred to fresh tubes and protein concentration was determined with Bradford reagent. Sample preparation was performed by mixing 5x gel loading buffer (**see Table 8**) with volumes of cell lysates corresponding to equal amounts of protein (~20-50 μg of total protein depending on the investigated target protein), which were normalized by diluting with 6M urea buffer. Samples were then denatured at 95°C for 5 mins, and were loaded equally on SDS-PAGE gels, followed by gel electrophoresis (~80-100 V for ~2 h). Protein transfer was carried out using the semi-dry blotting and the sequential placing of the following layers on the anode plate of the transfer machine; 3 layers of filter papers soaked in anode buffer (**see Table 8**), then the methanol-activated PVDF membrane soaked in anode buffer, the SDS-PAGE gel soaked in cathode buffer (**see Table 8**), and finally 3 layers of cathode-soaked filter papers. The trapped air bubbles were carefully removed upon the addition of each layer for efficient transferring and an electrical current of 2.5 mA/cm^2 was applied for 30-40 mins. The membrane was then washed once with 1x Tris-buffered saline (TBS)/Tween (**see Table 8**) for 5 mins on an orbital shaker, followed by incubation with the Ponceau S solution for 5 mins and moderate washing with TBS/Tween, thereafter transfer efficiency was documented by imaging. After thorough washing with TBS/Tween for the removal of Ponceau S, membrane was blocked with 5% (w/v) non-fat dry milk in TBS/Tween for at least 2 h at room temperature or overnight at 4°C. Membrane was then washed shortly with TBS/Tween prior to overnight incubation at 4°C with the desired primary antibody solution prepared with 5% (w/v) BSA (or milk) in TBS/Tween at dilutions recommended by the supplier (**also see Table 4**). After washing once for 5 mins with TBS/Tween, membrane was incubated for 1 h at room temperature with the corresponding anti-rabbit or -mouse HRP-conjugated secondary antibody dissolved in 5% (w/v) milk in TBS/Tween. The membrane was then washed at least three time (each time for 5 mins) with TBS/Tween, and the target protein was finally visualized using the HRP-substrate, Western Lightning Plus-ECl™ (Perkin Elmer) and imaged by the LAS3000 imaging system.

4.2.6. Immunoprecipitation: streptavidin-biotin pull-down assays

HCT116 cells were seeded in 10 cm-culture dishes until they reach 70-80% confluency. For *in vitro* pull-down assay, cells were harvested with 500 µL of a lysis buffer containing 20 mM Tris HCl (pH = 8), 150 mM NaCl, 1% Nonidet P-40 (NP-40), and 2 mM EDTA and supplemented with the aforementioned protease- and phosphatase inhibitors (**see section 4.2.5.**). Upon the addition of ice-cold lysis buffer, petri dishes were left on ice for 30 mins, after which cells were scraped off the dish using a plastic cell scraper. Cell lysates were then transferred to pre-cold Eppendorf tubes, and were centrifuged at 4°C for 30 mins at maximum speed. The supernatant was then transferred to fresh tubes and was pre-washed by a 20 min-incubation with 0.25 mg of streptavidin magnetic beads at room temperature, followed by separation on a magnetic rack. Pre-cleaned lysates were then incubated with 100 µM of biotin for 30 mins at 4°C with rotation. Streptavidin beads were then added to the samples at a dilution of 1:10 for 30 mins at 4°C with gentle agitation and were separated from the supernatant using a magnetic rack. The supernatant was subsequently transferred to fresh tubes and was incubated with biotin (as the negative control) and the biotin-conjugated small molecule (herein E738 B) at a final concentration of 100 µM at 4°C overnight. This was followed by the addition of the streptavidin beads (at 1:10 dilution) and incubation with rotation at 4°C for 1-3 h, then separation on a magnetic rack. The supernatant was then discarded and the pellets (beads) were washed at least three times with 500 µL of the aforementioned lysis buffer containing protease/phosphatase inhibitors. To elute the protein from the beads, we mixed the beads with 150 µL 3x SDS gel loading buffer (**see Table 8**) and heated the mixture for 3-5 mins at 96°C. The beads were then precipitated using a magnetic rack and the supernatant was transferred to fresh tubes and was run on SDS-PAGE. Samples were analyzed in collaboration with Dr. Thomas Ruppert at the ZMBH core facility for mass spectrometry and proteomics (Heidelberg, Germany) using the in-gel digestion by trypsin method, followed by MALDI TOF MS analysis. The pull-down assays in cells were carried out using a similar protocol, with the only difference of incubating HCT116 cells with biotinylated small molecule (E738 B, 10 µM for 24 h), biotin (10 µM for 24 h) or 0.1% DMSO; the last two were included as the negative controls of the experiment. Briefly, treated cells were then used for immunoprecipitation by adding the beads (at 1:10 dilution for 30 mins at 4°C), followed by separation on a magnetic rack, three times washing with the lysis buffer, and elution as described earlier.

4.2.7. Luciferase reporter assay

NCCIT-OCT4 and -NANOG, as well as HEK-OCT4 reporter cells were seeded in 24-well plates at a density of 100,000 cells/well. On the following day, cells were treated with the indicated small molecules, and at the end of the respective treatment, luciferase assay was performed using the Beetle juice kit according to the manufacturer's protocol. Briefly, cells were harvested by adding 80 µL/well of the luciferase lysis buffer (**see Table 8**). After incubation with the lysis buffer at 37°C for 15-30 mins, cells were shortly kept on ice and the contents were transferred to Eppendorf tubes,

thereafter centrifuged for 10 mins at 4°C at maximum speed to separate cell debris. 20 μL of the supernatant was then loaded onto a white 96-well plate with transparent bottom and the plate was centrifuged at 2000 rpm for 1 min. Immediately after this, 100 μL of the Beetle juice reaction mixture containing luciferin and ATP were added to each well pre-loaded with samples. The plate was then incubated in dark for 5 mins at room temperature, followed by luminescence measurement using the Tecan Ultra plate reader. In addition to this, protein concentration was measured using the Bradford reagent. Luciferase values were normalized to those of the respective Bradford values, and finally to the respective mock treatment, resulting in the fold change of luciferase activity. Bradford ratios below 0.1 (compared to blank) were considered as toxic conditions, and hence were excluded from further analysis.

4.2.8. Immunocytochemistry

Human fibroblasts and PSCs were seeded in 96-well plates at a density of 10,000 cells/well. At the end of the respective treatment, cells were fixed using 50 μL/well of 4% PFA in D-PBS for 10-15 mins (not more than 20 mins) at room temperature, after which washed at least two time (each time for 5 mins) with 200 μL/well D-PBS. Fixed permeable cells were then incubated for 30 mins with 50 μL/well of the blocking buffer (**see Table 8**) with rotation at room temperature, followed by overnight incubation with rotation at 4°C with 20-30 μL/well of the desired primary antibody solution (**also see Table 4**) prepared in the blocking buffer at dilutions recommended by the manufacturer. The primary antibodies were then replaced with a 100 μL/well mixture, containing the matching Alexa Flor 488 goat anti-rabbit or Alexa Flor 594 anti-mouse secondary antibodies, as well as 1 μg/mL of the Hoechst dye in order to visualize nuclei. After 1 h of incubation with rotation in dark at room temperature, cells were sufficiently washed with D-PBS (at least three time for 5 mins), and fluorescence signals were detected using the BIOREVO fluorescence microscope.

4.2.9. Flow cytometric analysis of cell death: annexin V and propidium iodide staining

As previously described [312], 200,000 cells/well were seeded in 12-well plates and kept in the incubator overnight. On the following day, cells were treated for the indicated durations, after which 250,000 cells/well were harvested by trypsinization (using TrypLE™ Express) in Eppendorf tubes. Cells were then washed once with 0.5-1 mL of the annexin V binding buffer (**see Table 8**), centrifuged at 200 g for 3 mins, thereafter the pellets were incubated with 50 μL/sample of the binding buffer containing 5 μL of both FITC-conjugated annexin V and propidium iodide. After 10-15 mins incubation with rotation in dark at room temperature, each sample was resuspended in binding buffer or phenol red-free DMEM up to a final volume of 200-500 μL. This was followed by loading the samples on round-bottom 96-well plates for FACS analysis, which was performed by the Guava easyCyte HT sampling flow cytometer. Data were analyzed using the GuavaSoft 3.1.1 software.

4.2.10. Flow cytometric analysis of pluripotency-associated markers

For analyzing the expression of pluripotency-related markers by FACS, NCCIT reporter cells or iPSCs were seeded in 6-well plates (in case of iPSCs coated with Geltrex) at a density of 300,000-500,000 cells/well and were left in the incubator overnight. On the following day, they were treated for the indicated periods, otherwise harvested directly. A similar protocol to that of immunocytochemistry was used to fix and permeabilize the cells, followed by incubation with primary- and secondary antibody solutions (**for more details refer to section 4.2.8.**). Fluorescence signals were then measured by FACSCalibur and data were analyzed using the software, CellQuest Pro.

4.2.11. Analysis of reprogramming efficiency: TRA-1-60+/AP+ colony counting

To determine the number of TRA-1-60- and AP-positive colonies (as a measure of reprogramming efficiency), iPSCs generated by means of episomal OSKM transfection in the absence or presence of OCT4-inducing small molecules and/or forced expression of *KDM5A*, *HMGA1*, and *ID1/2*, we performed immunostaining using the respective antibodies in living cells after 32 days (**for a detailed protocol on immunocytochemistry refer to section 4.2.8.**). Transfection efficiency was determined by quantifying the fluorescence signal resulting from the simultaneous over-expression of the pCXLE-EGFP (Addgene plasmid #27082).

4.2.12. iPSCs differentiation into the three germ layers

Mono-layer iPSCs seeded in Geltrex-coated 24-well plates and at a confluency of 60-80% were used for differentiation using the following kits, according to the manufacturer's instructions (**also see Table 3**); (i) Cellartis iPSC to hepatocyte differentiation system was utilized to differentiate iPSCs to endoderm, and (ii) StemDiff neuron differentiation kit was used for differentiation into the ectoderm germ layer. For mesoderm generation, iPSCs were seeded in Geltrex-coated plates as single cells till they reach 80% confluency, after which Essential® 8 medium was replaced with N2B27 containing freshly-added 100 ng/mL of Activin, 20 ng/mL of bFGF, 10 ng/mL of BMP4, 10 µM of LY294002 and 5 µM of CHIR99021. The medium was refreshed every 2-3 days until 14 days.

4.2.13. Chromatin-immunoprecipitation assay

SimpleChIP Plus Enzymatic Chromatin IP Kit (**see Table 3**) was used for DNA isolation precipitated with H3K4M3 and H3K27Me3 antibodies, according to the manufacturer's protocol. Non-targeting rabbit IgG served as the negative control and total histone H3 antibodies (provided by the manufacturer) were used as reference. Recruited DNA was then subjected to qRT-PCR using ChIP primer sets, and human RPL30 exon 3 primer pairs were used as positive control.

4.2.14. *In vitro* measurements of H3K4 methylation and demethylase activity of the histone demethylases, KDM4, KDM5s, and LSD1

For the *in vitro* measurement of H3K4 methylation status in response to O4I3 and zolpidem, the Epiquik histone demethylase (H3K4 specific) activity/inhibition assay kit (**see Table 3**) was used, according to the manufacturer's instructions. The *in vitro* inhibitory activity of the small molecules on KDM5s, KDM4 (also known as JMJD2), and LSD1 was determined using the following kits, respectively; (i) Epigenase JARID demethylase activity/inhibition assay kit, (ii) Epigenase JMJD2 demethylase activity/inhibition assay kit, and (iii) EpiQuik histone demethylase LSD1 activity/inhibition assay kit (**see Table 3 for detailed information of the suppliers**).

4.2.15. Bisulfate sequencing

The bisulfite conversion of DNA was performed using the EpiTect Bisulfite Kit (**see Table 3**), according to the manufacturer's instructions. 5 µL of bisulfite-treated genomic DNA was used to amplify the *OCT4* promoter region in human iPSCs and the parental fibroblasts. PCR products were cloned into pGEM-T vectors (Promega) and sequenced using the M13 forward primer. Primers used to amplify the *OCT4* promoter region have been previously described [313] (**also see Table 5 for the forward- and reverse primer sequences**).

4.2.16. ATAC-sequencing

After lysing cells with a lysis buffer containing 10 mM Tris HCL (pH = 7.4), 10 mM NaCl, 3 mM $MgCl_2$, and 0.1% IGEPAL-CA630, DNA library was prepared by the TransNGS® Tn5 DNA library pre-Kit for Illumina, according to the manufacturer’s instructions.

4.2.17. Gene expression profiling and associated data analyses

Gene expression profiling was conducted on the Illumina Human Sentrix-12v4 BeadChip array by the genomics and proteomics core facility at the German cancer research center (DKFZ). Gene expression analyses were performed in collaboration with the group of Prof. Miguel A. Andrade-Navarro (University of Mainz, Germany), using R.3.3.2. computing environment and packages from the open-source software development project Bioconductor 3.4. [314]. In addition, we used the standalone tool STEM 1.3.11 for clustering of gene profiles and functional enrichment analysis. Quality assessment and preprocessing of raw data were done with the package beadarray 2.24.2. [315]. Preprocessing involved default image processing (with median-based local background), default Illumina removal for outlying observations, mean summarization of bead-level observations into probe-level data, quantile normalization of probe-level data and log2 transformation. The resulting probes were matched to genes using annotations from the package illuminaHumanv4.db 1.26.0. We filtered probes, that poorly matched the annotated genes (quality status of ‘no-match’ and ‘bad’) as well as low-expression probes (detection score > 0.05) across all samples.

Dose experiment data analysis; 14108 probes were available for clustering analysis with STEM, a tool for clustering short-time series data, that can differentiate between real- and random temporal expression patterns. For this purpose, we selected one probe per gene (the one with highest variance, if multiple probes matched to single gene) and converted the data into fold changes with respect to the not-treated cells. We analyzed the data as short-time series of length three (corresponding to dose levels) by treating the dose level as a time factor. We applied STEM in default mode, except for the following settings; at least one measurement with absolute fold change of 0.65; significance of model profiles corrected by false discovery rate (FDR) method; cellular component terms, terms with evidence code IEA and NAS or terms that had less than 10 genes in common with the identified model profiles (or clusters of similar model profiles) were excluded from the Gene Ontology (GO) enrichment analysis [316]; multiple testing correction of GO term significant values was determined by the randomization test, with 5000 randomly sampled gene sets per term. Expression profiles of selected gene sets (related to significantly enriched GO terms) were visualized by heatmaps with the packages heatmap 1.0.8.

Fibroblast/iPSC experiment data analysis; approximately 18,000 probes, after filtering, were available for differential expression analysis. The latter was based on fold changes i.e., the probes with at least two-fold expression relative to the untreated fibroblast were selected as differentially expressed; since single biological replicates per treatment were used, statistical tests were not applicable. Global gene expression comparison between pairs of samples was visualized by scatter plots, plotted with the package ggplot2 2.2.1, while the heatmaps of gene expressions were visualized with the package ComplexHeatmap 1.12.0. In addition, columns and rows in the heatmaps were reordered with hierarchical clustering based on average linkage and Euclidean distance or Pearson correlation. Functional enrichment analysis on differentially expressed probes was performed with gProfileR version r1732_e89_eg36 via the R interface in the package gProfileR version 0.6.1. More specifically, the query lists were treated as unordered human gene lists, and the set of probes retained after filtering was supplied as a background. Multiple testing correction was done by the FDR method according to Benjamini and Hochberg *et al.* [317] with 0.05 FDR cut-off, including biological processes and molecular functions from the Gene Ontology database, KEGG and REACTOME pathway databases with a size of 5 up to 500 genes (***all database references could be found in*** [318].).

4.2.18. Statistics

statistical analyses were performed using Microsoft Excel, GraphPad Prism, and the R software. Densitometric quantifications were done by the ImageJ software. Statistical significance was calculated using a two-tailed student's *t*-test, as indicated in the respective figure legend. Multiple comparisons were employed either by one-way- or two-way ANOVA, followed by a post-hoc Tukey test, as determined in figure legends. *p*-values less than or equal to 0.05, 0.01, 0.001, and 0.0001

are denoted as *, **, ***, and **** on the figures, respectively. Error bars represent either the SD or SEM, as indicated in the corresponding figure legends.

REFERENCES

1. Derynck, R. and E.H. Budi, *Specificity, versatility, and control of TGF-beta family signaling.* Sci Signal, 2019. **12**(570).

2. Schmierer, B. and C.S. Hill, *TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 970-82.

3. Massague, J., J. Seoane, and D. Wotton, *Smad transcription factors.* Genes Dev, 2005. **19**(23): p. 2783-810.

4. Wrighton, K.H., X. Lin, and X.H. Feng, *Phospho-control of TGF-beta superfamily signaling.* Cell Res, 2009. **19**(1): p. 8-20.

5. Zhang, Y.E., *Non-Smad Signaling Pathways of the TGF-beta Family.* Cold Spring Harb Perspect Biol, 2017. **9**(2).

6. Olsson, N., et al., *Transforming growth factor-beta-mediated mast cell migration depends on mitogen-activated protein kinase activity.* Cell Signal, 2001. **13**(7): p. 483-90.

7. Yan, Z., S. Winawer, and E. Friedman, *Two different signal transduction pathways can be activated by transforming growth factor beta 1 in epithelial cells.* J Biol Chem, 1994. **269**(18): p. 13231-7.

8. Mulder, K.M. and S.L. Morris, *Activation of p21ras by transforming growth factor beta in epithelial cells.* J Biol Chem, 1992. **267**(8): p. 5029-31.

9. Lee, M.K., et al., *TGF-beta activates Erk MAP kinase signalling through direct phosphorylation of ShcA.* EMBO J, 2007. **26**(17): p. 3957-67.

10. Yoon, S. and R. Seger, *The extracellular signal-regulated kinase: multiple substrates regulate diverse cellular functions.* Growth Factors, 2006. **24**(1): p. 21-44.

11. Xie, L., et al., *Activation of the Erk pathway is required for TGF-beta1-induced EMT in vitro.* Neoplasia, 2004. **6**(5): p. 603-10.

12. Zavadil, J., et al., *Genetic programs of epithelial cell plasticity directed by transforming growth factor-beta.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6686-91.

13. Moustakas, A. and C.H. Heldin, *Induction of epithelial-mesenchymal transition by transforming growth factor beta.* Semin Cancer Biol, 2012. **22**(5-6): p. 446-54.

14. Lim, M., C.M. Chuong, and P. Roy-Burman, *PI3K, Erk signaling in BMP7-induced epithelial-mesenchymal transition (EMT) of PC-3 prostate cancer cells in 2- and 3-dimensional cultures.* Horm Cancer, 2011. **2**(5): p. 298-309.

15. Hsu, Y.L., et al., *Lung tumor-associated osteoblast-derived bone morphogenetic protein-2 increased epithelial-to-mesenchymal transition of cancer by Runx2/Snail signaling pathway.* J Biol Chem, 2011. **286**(43): p. 37335-46.

16. Kang, M.H., et al., *Metastatic function of BMP-2 in gastric cancer cells: the role of PI3K/AKT, MAPK, the NF-kappaB pathway, and MMP-9 expression.* Exp Cell Res, 2011. **317**(12): p. 1746-62.

17. Lai, C.F. and S.L. Cheng, *Signal transductions induced by bone morphogenetic protein-2 and transforming growth factor-beta in normal human osteoblastic cells.* J Biol Chem, 2002. **277**(18): p. 15514-22.

18. Gallea, S., et al., *Activation of mitogen-activated protein kinase cascades is involved in regulation of bone morphogenetic protein-2-induced osteoblast differentiation in pluripotent C2Cl2 cells.* Bone, 2001. **28**(5): p. 491-498.

19. Kretzschmar, M., J. Doody, and J. Massague, *Opposing BMP and EGF signalling pathways converge on the TGF-beta family mediator Smad1.* Nature, 1997. **389**(6651): p. 618-22.

20. Kretzschmar, M., et al., *A mechanism of repression of TGF beta/Smad signaling by oncogenic Ras.* Genes & Development, 1999. **13**(7): p. 804-816.

21. Funaba, M., C.M. Zimmerman, and L.S. Mathews, *Modulation of Smad2-mediated signaling by extracellular signal-regulated kinase.* J Biol Chem, 2002. **277**(44): p. 41361-8.

22. Javelaud, D. and A. Mauviel, *Crosstalk mechanisms between the mitogen-activated protein kinase pathways and Smad signaling downstream of TGF-beta: implications for carcinogenesis.* Oncogene, 2005. **24**(37): p. 5742-5750.

23. Weston, C.R. and R.J. Davis, *The JNK signal transduction pathway.* Curr Opin Cell Biol, 2007. **19**(2): p. 142-9.

24. Sorrentino, A., et al., *The type I TGF-beta receptor engages TRAF6 to activate TAK1 in a receptor kinase-independent manner.* Nat Cell Biol, 2008. **10**(10): p. 1199-207.

25. Yamashita, M., et al., *TRAF6 mediates Smad-independent activation of JNK and p38 by TGF-beta.* Mol Cell, 2008. **31**(6): p. 918-24.

26. Jung, S.M., et al., *Smad6 inhibits non-canonical TGF-beta1 signalling by recruiting the deubiquitinase A20 to TRAF6.* Nat Commun, 2013. **4**: p. 2562.

27. Di Guglielmo, G.M., et al., *Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover.* Nat Cell Biol, 2003. **5**(5): p. 410-21.

28. Yang, Y.A., et al., *Smad3 reduces susceptibility to hepatocarcinoma by sensitizing hepatocytes to apoptosis through downregulation of Bcl-2.* Cancer Cell, 2006. **9**(6): p. 445-57.

29. Valderrama-Carvajal, H., et al., *Activin/TGF-beta induce apoptosis through Smad-dependent expression of the lipid phosphatase SHIP.* Nat Cell Biol, 2002. **4**(12): p. 963-9.

30. Bakin, A.V., et al., *p38 mitogen-activated protein kinase is required for TGFbeta-mediated fibroblastic transdifferentiation and cell migration.* J Cell Sci, 2002. **115**(Pt 15): p. 3193-206.

31. Yoo, J., et al., *Transforming growth factor-beta-induced apoptosis is mediated by Smad-dependent expression of GADD45b through p38 activation.* J Biol Chem, 2003. **278**(44): p. 43001-7.

32. Zhang, L., et al., *TRAF4 promotes TGF-beta receptor signaling and drives breast cancer metastasis.* Mol Cell, 2013. **51**(5): p. 559-72.

33. Takaesu, G., et al., *TAK1 is critical for IkappaB kinase-mediated activation of the NF-kappaB pathway.* J Mol Biol, 2003. **326**(1): p. 105-15.

34. Chen, Y.G. and X.F. Wang, *Finale: the last minutes of Smads.* Cell, 2009. **139**(4): p. 658-60.

35. Massague, J., *TGFbeta signalling in context.* Nat Rev Mol Cell Biol, 2012. **13**(10): p. 616-30.

36. Sapkota, G., et al., *Balancing BMP signaling through integrated inputs into the Smad1 linker.* Mol Cell, 2007. **25**(3): p. 441-54.

37. Matsuura, I., et al., *Cyclin-dependent kinases regulate the antiproliferative function of Smads.* Nature, 2004. **430**(6996): p. 226-31.

38. Fuentealba, L.C., et al., *Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal.* Cell, 2007. **131**(5): p. 980-93.

39. Aragon, E., et al., *A Smad action turnover switch operated by WW domain readers of a phosphoserine code.* Genes Dev, 2011. **25**(12): p. 1275-88.

40. Lonn, P., et al., *Regulating the stability of TGFbeta receptors and Smads.* Cell Res, 2009. **19**(1): p. 21-35.

41. Zhu, H., et al., *A SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation.* Nature, 1999. **400**(6745): p. 687-93.

42. Suzuki, C., et al., *Smurf1 regulates the inhibitory activity of Smad7 by targeting Smad7 to the plasma membrane.* J Biol Chem, 2002. **277**(42): p. 39919-25.

43. Zhang, Y., et al., *Regulation of Smad degradation and activity by Smurf2, an E3 ubiquitin ligase.* Proc Natl Acad Sci U S A, 2001. **98**(3): p. 974-9.

44. Gao, S., et al., *Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF-beta signaling.* Mol Cell, 2009. **36**(3): p. 457-68.

45. Moren, A., et al., *Degradation of the tumor suppressor Smad4 by WW and HECT domain ubiquitin ligases.* J Biol Chem, 2005. **280**(23): p. 22115-23.

46. Nagano, Y., et al., *Arkadia induces degradation of SnoN and c-Ski to enhance transforming growth factor-beta signaling.* J Biol Chem, 2007. **282**(28): p. 20492-501.

47. Stroschein, S.L., et al., *Smad3 recruits the anaphase-promoting complex for ubiquitination and degradation of SnoN.* Genes Dev, 2001. **15**(21): p. 2822-36.

48. Fukuchi, M., et al., *Ligand-dependent degradation of Smad3 by a ubiquitin ligase complex of ROC1 and associated proteins.* Mol Biol Cell, 2001. **12**(5): p. 1431-43.

49. Wan, M., et al., *Smad4 protein stability is regulated by ubiquitin ligase SCF beta-TrCP1.* J Biol Chem, 2004. **279**(15): p. 14484-7.

50. Li, L., et al., *CHIP mediates degradation of Smad proteins and potentially regulates Smad-induced transcription.* Mol Cell Biol, 2004. **24**(2): p. 856-64.

51. Kwon, Y.T. and A. Ciechanover, *The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy.* Trends Biochem Sci, 2017. **42**(11): p. 873-886.

52. Hata, A. and Y.G. Chen, *TGF-beta Signaling from Receptors to Smads.* Cold Spring Harb Perspect Biol, 2016. **8**(9).

53. Dupont, S., et al., *Germ-layer specification and control of cell growth by Ectodermin, a Smad4 ubiquitin ligase.* Cell, 2005. **121**(1): p. 87-99.

54. Xi, Q., et al., *A poised chromatin platform for TGF-beta access to master regulators.* Cell, 2011. **147**(7): p. 1511-24.

55. Tang, L.Y., et al., *Ablation of Smurf2 reveals an inhibition in TGF-beta signalling through multiple mono-ubiquitination of Smad3.* EMBO J, 2011. **30**(23): p. 4777-89.

56. Herhaus, L. and G.P. Sapkota, *The emerging roles of deubiquitylating enzymes (DUBs) in the TGFbeta and BMP pathways.* Cell Signal, 2014. **26**(10): p. 2186-92.

57. Kim, S.Y. and K.H. Baek, *TGF-beta signaling pathway mediated by deubiquitinating enzymes.* Cell Mol Life Sci, 2019. **76**(4): p. 653-665.

58. Zhang, L., et al., *USP4 is regulated by AKT phosphorylation and directly deubiquitylates TGF-beta type I receptor.* Nat Cell Biol, 2012. **14**(7): p. 717-26.

59. Al-Salihi, M.A., et al., *USP11 augments TGFbeta signalling by deubiquitylating ALK5.* Open Biol, 2012. **2**(6): p. 120063.

60. Eichhorn, P.J., et al., *USP15 stabilizes TGF-beta receptor I and promotes oncogenesis through the activation of TGF-beta signaling in glioblastoma.* Nat Med, 2012. **18**(3): p. 429-35.

61. Kwon, S.K., M. Saindane, and K.H. Baek, *p53 stability is regulated by diverse deubiquitinating enzymes.* Biochimica Et Biophysica Acta-Reviews on Cancer, 2017. **1868**(2): p. 404-411.

62. Zhao, Y., et al., *USP2a Supports Metastasis by Tuning TGF-beta Signaling.* Cell Rep, 2018. **22**(9): p. 2442-2454.

63. Zou, Q., et al., *USP15 stabilizes MDM2 to mediate cancer-cell survival and inhibit antitumor T cell responses.* Nat Immunol, 2014. **15**(6): p. 562-70.

64. Inui, M., et al., *USP15 is a deubiquitylating enzyme for receptor-activated SMADs.* Nature Cell Biology, 2011. **13**(11): p. 1368-U187.

65. Dupont, S., et al., *FAM/USP9x, a deubiquitinating enzyme essential for TGFbeta signaling, controls Smad4 monoubiquitination.* Cell, 2009. **136**(1): p. 123-35.

66. Lim, J.H., et al., *CYLD negatively regulates transforming growth factor-beta-signalling via deubiquitinating Akt.* Nat Commun, 2012. **3**: p. 771.

67. Zhao, Y., et al., *The deubiquitinase CYLD targets Smad7 protein to regulate transforming growth factor beta (TGF-beta) signaling and the development of regulatory T cells.* J Biol Chem, 2011. **286**(47): p. 40520-30.

68. Herhaus, L., et al., *OTUB1 enhances TGFbeta signalling by inhibiting the ubiquitylation and degradation of active SMAD2/3.* Nat Commun, 2013. **4**: p. 2519.

69. Wicks, S.J., et al., *The deubiquitinating enzyme UCH37 interacts with Smads and regulates TGF-beta signalling.* Oncogene, 2005. **24**(54): p. 8080-4.

70. Cutts, A.J., et al., *Early phase TGFbeta receptor signalling dynamics stabilised by the deubiquitinase UCH37 promotes cell migratory responses.* Int J Biochem Cell Biol, 2011. **43**(4): p. 604-12.

71. Itoh, F., et al., *Promoting bone morphogenetic protein signaling through negative regulation of inhibitory Smads.* EMBO J, 2001. **20**(15): p. 4132-42.

72. Ibarrola, N., et al., *Cloning of a novel signaling molecule, AMSH-2, that potentiates transforming growth factor beta signaling.* BMC Cell Biol, 2004. **5**: p. 2.

73. Fan, Y.H., et al., *USP4 targets TAK1 to downregulate TNFalpha-induced NF-kappaB activation.* Cell Death Differ, 2011. **18**(10): p. 1547-60.

74. Liu, X., et al., *USP18 inhibits NF-kappaB and NFAT activation during Th17 differentiation by deubiquitinating the TAK1-TAB1 complex.* J Exp Med, 2013. **210**(8): p. 1575-90.

75. Lin, X., M. Liang, and X.H. Feng, *Smurf2 is a ubiquitin E3 ligase mediating proteasome-dependent degradation of Smad2 in transforming growth factor-beta signaling.* Journal of Biological Chemistry, 2000. **275**(47): p. 36818-36822.

76. Inoue, Y., et al., *Contribution of the constitutive and inducible degradation of Smad3 by the ubiquitin-proteasome pathway to transforming growth factor-beta signaling.* J Interferon Cytokine Res, 2004. **24**(1): p. 43-54.

77. Xin, H., et al., *CHIP controls the sensitivity of transforming growth factor-beta signaling by modulating the basal level of Smad3 through ubiquitin-mediated degradation.* Journal of Biological Chemistry, 2005. **280**(21): p. 20842-20850.

78. Guo, X., et al., *Axin and GSK3-beta control Smad3 protein stability and modulate TGF-beta signaling.* Genes & Development, 2008. **22**(1): p. 106-120.

79. Seoane, J. and R.R. Gomis, *TGF-beta Family Signaling in Tumor Suppression and Cancer Progression.* Cold Spring Harb Perspect Biol, 2017. **9**(12).

80. Edlund, S., et al., *Transforming growth factor-beta1 (TGF-beta)-induced apoptosis of prostate cancer cells involves Smad7-dependent activation of p38 by TGF-beta-activated kinase 1 and mitogen-activated protein kinase kinase 3.* Mol Biol Cell, 2003. **14**(2): p. 529-44.

81. Ramjaun, A.R., et al., *Upregulation of two BH3-only proteins, Bmf and Bim, during TGF beta-induced apoptosis.* Oncogene, 2007. **26**(7): p. 970-81.

82. Bhowmick, N.A., et al., *TGF-beta signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia.* Science, 2004. **303**(5659): p. 848-51.

83. Massague, J., *TGFbeta in Cancer.* Cell, 2008. **134**(2): p. 215-30.

84. Massague, J., S.W. Blain, and R.S. Lo, *TGFbeta signaling in growth control, cancer, and heritable disorders.* Cell, 2000. **103**(2): p. 295-309.

85. Sjoblom, T., et al., *The consensus coding sequences of human breast and colorectal cancers.* Science, 2006. **314**(5797): p. 268-74.

86. Levy, L. and C.S. Hill, *Alterations in components of the TGF-beta superfamily signaling pathways in human cancer.* Cytokine Growth Factor Rev, 2006. **17**(1-2): p. 41-58.

87. Jaffee, E.M., et al., *Focus on pancreas cancer.* Cancer Cell, 2002. **2**(1): p. 25-8.

88. Barrett, M.T., et al., *Allelic loss and mutational analysis of the DPC4 gene in esophageal adenocarcinoma.* Cancer Res, 1996. **56**(19): p. 4351-3.

89. Lei, J., et al., *Infrequent DPC4 gene mutation in esophageal cancer, gastric cancer and ulcerative colitis-associated neoplasms.* Oncogene, 1996. **13**(11): p. 2459-62.

90. Dowdy, S.C., et al., *Overexpression of the TGF-beta antagonist Smad7 in endometrial cancer.* Gynecol Oncol, 2005. **96**(2): p. 368-73.

91. de Boeck, M., et al., *Smad6 determines BMP-regulated invasive behaviour of breast cancer cells in a zebrafish xenograft model.* Sci Rep, 2016. **6**: p. 24968.

92. Cerutti, J.M., et al., *Expression of Smad4 and Smad7 in human thyroid follicular carcinoma cell lines.* Journal of Endocrinological Investigation, 2003. **26**(6): p. 516-521.

93. Kleeff, J., et al., *Smad6 suppresses TGF-beta-induced growth inhibition in COLO-357 pancreatic cancer cells and is overexpressed in pancreatic cancer.* Biochem Biophys Res Commun, 1999. **255**(2): p. 268-73.

94. Zhu, Q., et al., *Dual role of SnoN in mammalian tumorigenesis.* Mol Cell Biol, 2007. **27**(1): p. 324-39.

95. Sanjabi, S., S.A. Oh, and M.O. Li, *Regulation of the Immune Response by TGF-beta: From Conception to Autoimmunity and Infection.* Cold Spring Harb Perspect Biol, 2017. **9**(6).

96. Arteaga, C.L., et al., *Reversal of tamoxifen resistance of human breast carcinomas in vivo by neutralizing antibodies to transforming growth factor-beta.* J Natl Cancer Inst, 1999. **91**(1): p. 46-53.

97. Fontana, A., et al., *Modulation of the immune response by transforming growth factor beta.* Int Arch Allergy Immunol, 1992. **99**(1): p. 1-7.

98. McDonald, P.P., et al., *Transcriptional and translational regulation of inflammatory mediator production by endogenous TGP-beta in macrophages that have ingested apoptotic cells.* Journal of Immunology, 1999. **163**(11): p. 6164-6172.

99. Maman, S. and I.P. Witz, *A history of exploring cancer in context.* Nat Rev Cancer, 2018. **18**(6): p. 359-376.

100. Derynck, R. and R.A. Weinberg, *EMT and Cancer: More Than Meets the Eye.* Developmental Cell, 2019. **49**(3): p. 313-316.

101. Hao, Y., D. Baker, and P. Ten Dijke, *TGF-beta-Mediated Epithelial-Mesenchymal Transition and Cancer Metastasis.* Int J Mol Sci, 2019. **20**(11).

102. Erez, N., et al., *Cancer-Associated Fibroblasts Are Activated in Incipient Neoplasia to Orchestrate Tumor-Promoting Inflammation in an NF-kappaB-Dependent Manner.* Cancer Cell, 2010. **17**(2): p. 135-47.

103. Zeisberg, E.M., et al., *Discovery of endothelial to mesenchymal transition as a source for carcinoma-associated fibroblasts.* Cancer Res, 2007. **67**(21): p. 10123-8.

104. David, C.J., et al., *TGF-beta Tumor Suppression through a Lethal EMT.* Cell, 2016. **164**(5): p. 1015-30.

105. Baccelli, I. and A. Trumpp, *The evolving concept of cancer and metastasis stem cells.* J Cell Biol, 2012. **198**(3): p. 281-93.

106. Anido, J., et al., *TGF-beta Receptor Inhibitors Target the CD44(high)/Id1(high) Glioma-Initiating Cell Population in Human Glioblastoma.* Cancer Cell, 2010. **18**(6): p. 655-68.

107. Ikushima, H., et al., *Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors.* Cell Stem Cell, 2009. **5**(5): p. 504-14.

108. Penuelas, S., et al., *TGF-beta increases glioma-initiating cell self-renewal through the induction of LIF in human glioblastoma.* Cancer Cell, 2009. **15**(4): p. 315-27.

109. You, H., W. Ding, and C.B. Rountree, *Epigenetic regulation of cancer stem cell marker CD133 by transforming growth factor-beta.* Hepatology, 2010. **51**(5): p. 1635-44.

110. Naka, K., et al., *TGF-beta-FOXO signalling maintains leukaemia-initiating cells in chronic myeloid leukaemia.* Nature, 2010. **463**(7281): p. 676-80.

111. Scheel, C., et al., *Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast.* Cell, 2011. **145**(6): p. 926-40.

112. Gao, H., et al., *The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites.* Cell, 2012. **150**(4): p. 764-79.

113. Dabiri, Y., Song, G., Cheng, X, *Indirubins As Multi-target Anti-Tumor Agents*, in *Herbal Medicine: Back to the Future*, F. Murad, Rahman, A, Bian, K Editor. 2019, Bentham Science: Singapore. p. 148-160.

114. Cheng, X. and K.H. Merz, *The Role of Indirubins in Inflammation and Associated Tumorigenesis.* Adv Exp Med Biol, 2016. **929**: p. 269-290.

115. Hoessel, R., et al., *Indirubin, the active constituent of a Chinese antileukaemia medicine, inhibits cyclin-dependent kinases.* Nat Cell Biol, 1999. **1**(1): p. 60-7.

116. Jautelat, R., et al., *From the insoluble dye indirubin towards highly active, soluble CDK2-inhibitors.* Chembiochem, 2005. **6**(3): p. 531-40.

117. Davies, T.G., et al., *Inhibitor binding to active and inactive CDK2: the crystal structure of CDK2-cyclin A/indirubin-5-sulphonate.* Structure, 2001. **9**(5): p. 389-97.

118. Heshmati, N., et al., *Physicochemical characterization and in vitro permeation of an indirubin derivative.* Eur J Pharm Sci, 2013. **50**(3-4): p. 467-75.

119. Heshmati, N., et al., *Enhancement of oral bioavailability of E804 by self-nanoemulsifying drug delivery system (SNEDDS) in rats.* J Pharm Sci, 2013. **102**(10): p. 3792-9.

120. Heshmati, N., et al., *In vitro and in vivo evaluations of the performance of an indirubin derivative, formulated in four different self-emulsifying drug delivery systems.* Journal of Pharmacy and Pharmacology, 2014. **66**(11): p. 1567-1575.

121. Eisenbrand, G., et al., *Molecular mechanisms of indirubin and its derivatives: novel anticancer molecules with their origin in traditional Chinese phytomedicine.* J Cancer Res Clin Oncol, 2004. **130**(11): p. 627-35.

122. Marko, D., et al., *Inhibition of cyclin-dependent kinase 1 (CDK1) by indirubin derivatives in human tumour cells.* Br J Cancer, 2001. **84**(2): p. 283-9.

123. Liao, X.M. and K.N. Leung, *Indirubin-3'-oxime induces mitochondrial dysfunction and triggers growth inhibition and cell cycle arrest in human neuroblastoma cells.* Oncol Rep, 2013. **29**(1): p. 371-9.

124. Cohen, P. and S. Frame, *The renaissance of GSK3.* Nat Rev Mol Cell Biol, 2001. **2**(10): p. 769-76.

125. Henriksen, E.J., *Dysregulation of glycogen synthase kinase-3 in skeletal muscle and the etiology of insulin resistance and type 2 diabetes.* Curr Diabetes Rev, 2010. **6**(5): p. 285-93.

126. Leclerc, S., et al., *Indirubins inhibit glycogen synthase kinase-3 beta and CDK5/p25, two protein kinases involved in abnormal tau phosphorylation in Alzheimer's disease. A property common to most cyclin-dependent kinase inhibitors?* J Biol Chem, 2001. **276**(1): p. 251-60.

127. Meijer, L., et al., *GSK-3-selective inhibitors derived from Tyrian purple indirubins.* Chem Biol, 2003. **10**(12): p. 1255-66.

128. McCubrey, J.A., et al., *GSK-3 as potential target for therapeutic intervention in cancer.* Oncotarget, 2014. **5**(10): p. 2881-911.

129. Williams, S.P., et al., *Indirubins decrease glioma invasion by blocking migratory phenotypes in both the tumor and stromal endothelial cell compartments.* Cancer Res, 2011. **71**(16): p. 5374-80.

130. Braig, S., et al., *Indirubin derivative 6BIO suppresses metastasis.* Cancer Res, 2013. **73**(19): p. 6004-12.

131. Du, Z. and C.M. Lovly, *Mechanisms of receptor tyrosine kinase activation in cancer.* Mol Cancer, 2018. **17**(1): p. 58.

132. Nam, S., et al., *Indirubin derivatives inhibit Stat3 signaling and induce apoptosis in human cancer cells.* Proc Natl Acad Sci U S A, 2005. **102**(17): p. 5998-6003.

133. Hoelbl, A., et al., *Stat5 is indispensable for the maintenance of bcr/abl-positive leukaemia.* EMBO Mol Med, 2010. **2**(3): p. 98-110.

134. Nam, S., et al., *Indirubin derivatives induce apoptosis of chronic myelogenous leukemia cells involving inhibition of Stat5 signaling.* Mol Oncol, 2012. **6**(3): p. 276-83.

135. Sen, B., et al., *Sustained Src inhibition results in signal transducer and activator of transcription 3 (STAT3) activation and cancer cell survival via altered Janus-activated kinase-STAT3 binding.* Cancer Res, 2009. **69**(5): p. 1958-65.

136. Nam, S., et al., *Dual inhibition of Janus and Src family kinases by novel indirubin derivative blocks constitutively-activated Stat3 signaling associated with apoptosis of human pancreatic cancer cells.* Mol Oncol, 2013. **7**(3): p. 369-78.

137. Liu, L., et al., *6-Bromoindirubin-3'-oxime inhibits JAK/STAT3 signaling and induces apoptosis of human melanoma cells.* Cancer Res, 2011. **71**(11): p. 3972-9.

138. Liu, L., et al., *A novel 7-bromoindirubin with potent anticancer activity suppresses survival of human melanoma cells associated with inhibition of STAT3 and Akt signaling.* Cancer Biology & Therapy, 2012. **13**(13): p. 1255-1261.

139. Ribas, J., et al., *7-Bromoindirubin-3'-oxime induces caspase-independent cell death.* Oncogene, 2006. **25**(47): p. 6304-18.

140. Nicolaou, K.A., et al., *Induction of discrete apoptotic pathways by bromo-substituted indirubin derivatives in invasive breast cancer cells.* Biochem Biophys Res Commun, 2012. **425**(1): p. 76-82.

141. Kim, J.K., et al., *Indirubin-3'-monoxime, a derivative of a chinese antileukemia medicine, inhibits angiogenesis.* J Cell Biochem, 2011. **112**(5): p. 1384-91.

142. Zhang, X., et al., *Indirubin inhibits tumor growth by antitumor angiogenesis via blocking VEGFR2-mediated JAK/STAT3 signaling in endothelial cell.* Int J Cancer, 2011. **129**(10): p. 2502-11.

143. Chan, Y.K., et al., *An indirubin derivative, E804, exhibits potent angiosuppressive activity.* Biochem Pharmacol, 2012. **83**(5): p. 598-607.

144. Shin, E.K. and J.K. Kim, *Indirubin derivative E804 inhibits angiogenesis.* BMC Cancer, 2012. **12**: p. 164.

145. Zhen, Y., et al., *Indirubin-3'-monoxime inhibits autophosphorylation of FGFR1 and stimulates ERK1/2 activity via p38 MAPK.* Oncogene, 2007. **26**(44): p. 6372-85.

146. Janssen, J.A. and A.J. Varewijck, *IGF-IR Targeted Therapy: Past, Present and Future.* Front Endocrinol (Lausanne), 2014. **5**: p. 224.

147. Cheng, X., et al., *Identification of a Water-Soluble Indirubin Derivative as Potent Inhibitor of Insulin-like Growth Factor 1 Receptor through Structural Modification of the Parent Natural Molecule.* J Med Chem, 2017. **60**(12): p. 4949-4962.

148. Cheng, X., et al., *Indirubin derivatives modulate TGFbeta/BMP signaling at different levels and trigger ubiquitin-mediated depletion of nonactivated R-Smads.* Chem Biol, 2012. **19**(11): p. 1423-36.

149. Sethi, G., et al., *Indirubin enhances tumor necrosis factor-induced apoptosis through modulation of nuclear factor-kappa B signaling pathway.* J Biol Chem, 2006. **281**(33): p. 23425-35.

150. Cheng, X., et al., *7,7'-Diazaindirubin--a small molecule inhibitor of casein kinase 2 in vitro and in cells.* Bioorg Med Chem, 2014. **22**(1): p. 247-55.

151. Clara, J.A., et al., *Targeting signalling pathways and the immune microenvironment of cancer stem cells - a clinical update.* Nat Rev Clin Oncol, 2019.

152. Cheng, X., et al., *Methylisoindigo preferentially kills cancer stem cells by interfering cell metabolism via inhibition of LKB1 and activation of AMPK in PDACs.* Mol Oncol, 2016. **10**(6): p. 806-24.

153. Guo, R., et al., *Novel MicroRNA Reporter Uncovers Repression of Let-7 by GSK-3beta.* PLoS One, 2013. **8**(6): p. e66330.

154. Wu, Y., et al., *GSK3 inhibitors CHIR99021 and 6-bromoindirubin-3'-oxime inhibit microRNA maturation in mouse embryonic stem cells.* Sci Rep, 2015. **5**: p. 8666.

155. Zhang, Z.N., et al., *Treatment of chronic myelocytic leukemia (CML) by traditional Chinese medicine and Western medicine alternately.* J Tradit Chin Med, 1985. **5**(4): p. 246-8.

156. Xiao, Z., et al., *Indirubin and meisoindigo in the treatment of chronic myelogenous leukemia in China.* Leuk Lymphoma, 2002. **43**(9): p. 1763-8.

157. Chen, F., et al., *Imatinib achieved complete cytogenetic response in a CML patient received 32-year indirubin and its derivative treatment.* Leuk Res, 2010. **34**(2): p. e75-7.

158. Xiao, Z.J., et al., *Meisoindigo for the treatment of chronic myelogenous leukaemia.* British Journal of Haematology, 2000. **111**(2): p. 711-712.

159. Roschger, C. and C. Cabrele, *The Id-protein family in developmental and cancer-associated pathways.* Cell Commun Signal, 2017. **15**(1): p. 7.

160. Clement, J.H., et al., *Bone morphogenetic protein 2 (BMP-2) induces sequential changes of Id gene expression in the breast cancer cell line MCF-7.* Journal of Cancer Research and Clinical Oncology, 2000. **126**(5): p. 271-279.

161. Tecalco-Cruz, A.C., et al., *Transcriptional cofactors Ski and SnoN are major regulators of the TGF-beta/Smad signaling pathway in health and disease.* Signal Transduct Target Ther, 2018. **3**: p. 15.

162. Wu, M., G. Chen, and Y.P. Li, *TGF-beta and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease.* Bone Res, 2016. **4**: p. 16009.

163. Zhang, S., et al., *TGFbeta1-induced activation of ATM and p53 mediates apoptosis in a Smad7-dependent manner.* Cell Cycle, 2006. **5**(23): p. 2787-95.

164. Matsuzaki, K., *Smad phospho-isoforms direct context-dependent TGF-beta signaling.* Cytokine Growth Factor Rev, 2013. **24**(4): p. 385-99.

165. Ooshima, A., J. Park, and S.J. Kim, *Phosphorylation status at Smad3 linker region modulates transforming growth factor-beta-induced epithelial-mesenchymal transition and cancer progression.* Cancer Sci, 2019. **110**(2): p. 481-488.

166. Eid, S., et al., *KinMap: a web-based tool for interactive navigation through human kinome data.* BMC Bioinformatics, 2017. **18**(1): p. 16.

167. Sritananuwat, P., et al., *Blocking ERK1/2 signaling impairs TGF-beta1 tumor promoting function but enhances its tumor suppressing role in intrahepatic cholangiocarcinoma cells.* Cancer Cell Int, 2017. **17**: p. 85.

168. Fouassier, L., et al., *Signalling networks in cholangiocarcinoma: Molecular pathogenesis, targeted therapies and drug resistance.* Liver Int, 2019. **39 Suppl 1**: p. 43-62.

169. Araki, K., et al., *E/N-cadherin switch mediates cancer progression via TGF-beta-induced epithelial-to-mesenchymal transition in extrahepatic cholangiocarcinoma.* Br J Cancer, 2011. **105**(12): p. 1885-93.

170. Lustri, A.M., et al., *TGF-beta signaling is an effective target to impair survival and induce apoptosis of human cholangiocarcinoma cells: A study on human primary cell cultures.* Plos One, 2017. **12**(9).

171. Sato, Y., et al., *Epithelial-mesenchymal transition induced by transforming growth factor-{beta}1/Snail activation aggravates invasive growth of cholangiocarcinoma.* Am J Pathol, 2010. **177**(1): p. 141-52.

172. Bhattacharyya, S., et al., *Regulated protein turnover: snapshots of the proteasome in action.* Nat Rev Mol Cell Biol, 2014. **15**(2): p. 122-33.

173. Wade, M., Y.C. Li, and G.M. Wahl, *MDM2, MDMX and p53 in oncogenesis and cancer therapy.* Nat Rev Cancer, 2013. **13**(2): p. 83-96.

174. Gossage, L., T. Eisen, and E.R. Maher, *VHL, the story of a tumour suppressor gene.* Nat Rev Cancer, 2015. **15**(1): p. 55-64.

175. Alarcon, C., et al., *Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways.* Cell, 2009. **139**(4): p. 757-69.

176. Kamaraju, S.K. and A.B. Roberts, *Role of Rho/ROCK and p38 MAP kinase pathways in transforming growth factor-beta-mediated Smad-dependent growth inhibition of human breast carcinoma cells in vivo.* Journal of Biological Chemistry, 2005. **280**(2): p. 1024-1036.

177. Yamagata, H., et al., *Acceleration of Smad2 and Smad3 phosphorylation via c-Jun NH2-terminal kinase during human colorectal carcinogenesis.* Cancer Research, 2005. **65**(1): p. 157-165.

178. Yu, J.S., et al., *PI3K/mTORC2 regulates TGF-beta/Activin signalling by modulating Smad2/3 activity via linker phosphorylation.* Nat Commun, 2015. **6**: p. 7212.

179. Matsuzaki, K., T. Seki, and K. Okazaki, *TGF-beta signal shifting between tumor suppression and fibro-carcinogenesis in human chronic liver diseases.* J Gastroenterol, 2014. **49**(6): p. 971-81.

180. Matsuzaki, K., et al., *Smad2 and Smad3 Phosphorylated at Both Linker and COOH-Terminal Regions Transmit Malignant TGF-beta Signal in Later Stages of Human Colorectal Cancer.* Cancer Research, 2009. **69**(13): p. 5321-5330.

181. Papoutsoglou, P., C. Louis, and C. Coulouarn, *Transforming Growth Factor-Beta (TGF beta) Signaling Pathway in Cholangiocarcinoma.* Cells, 2019. **8**(9).

182. Wang, P., et al., *MEK inhibition suppresses K-Ras wild-type cholangiocarcinoma in vitro and in vivo via inhibiting cell proliferation and modulating tumor microenvironment.* Cell Death Dis, 2019. **10**(2): p. 120.

183. Song, X., et al., *Combined CDK4/6 and Pan-mTOR Inhibition Is Synergistic Against Intrahepatic Cholangiocarcinoma.* Clin Cancer Res, 2019. **25**(1): p. 403-413.

184. Sittithumcharee, G., et al., *Dependency of Cholangiocarcinoma on Cyclin D-Dependent Kinase Activity.* Hepatology, 2019. **70**(5): p. 1614-1630.

185. Zhang, M.X., et al., *S100A11 promotes cell proliferation via P38/MAPK signaling pathway in intrahepatic cholangiocarcinoma.* Mol Carcinog, 2019. **58**(1): p. 19-30.

186. Mancinelli, R., et al., *Multifaceted Roles of GSK-3 in Cancer and Autophagy-Related Diseases.* Oxid Med Cell Longev, 2017. **2017**: p. 4629495.

187. Gurdon, J.B., T.R. Elsdale, and M. Fischberg, *Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei.* Nature, 1958. **182**(4627): p. 64-5.

188. Gurdon, J.B., *The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles.* J Embryol Exp Morphol, 1962. **10**: p. 622-40.

189. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.

190. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. **78**(12): p. 7634-8.

191. Wilmut, I., et al., *Viable offspring derived from fetal and adult mammalian cells.* Nature, 1997. **385**(6619): p. 810-3.

192. Wakayama, T., et al., *Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei.* Nature, 1998. **394**(6691): p. 369-74.

193. Tada, M., et al., *Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells.* Curr Biol, 2001. **11**(19): p. 1553-8.

194. Cowan, C.A., et al., *Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells.* Science, 2005. **309**(5739): p. 1369-73.

195. Davis, R.L., H. Weintraub, and A.B. Lassar, *Expression of a single transfected cDNA converts fibroblasts to myoblasts.* Cell, 1987. **51**(6): p. 987-1000.

196. Kulessa, H., J. Frampton, and T. Graf, *Gata-1 Reprograms Avian Myelomonocytic Cell-Lines into Eosinophils, Thromboblasts, and Erythroblasts.* Genes & Development, 1995. **9**(10): p. 1250-1262.

197. Takahashi, K. and S. Yamanaka, *A decade of transcription factor-mediated reprogramming to pluripotency.* Nat Rev Mol Cell Biol, 2016. **17**(3): p. 183-93.

198. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-676.

199. Mitsui, K., et al., *The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells.* Cell, 2003. **113**(5): p. 631-642.

200. Chambers, I., et al., *Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells.* Cell, 2003. **113**(5): p. 643-655.

201. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-872.

202. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.

203. Yu, J., et al., *Human induced pluripotent stem cells free of vector and transgene sequences.* Science, 2009. **324**(5928): p. 797-801.

204. Tanabe, K., et al., *Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts.* Proc Natl Acad Sci U S A, 2013. **110**(30): p. 12172-9.

205. Nakagawa, M., et al., *Promotion of direct reprogramming by transformation-deficient Myc.* Proc Natl Acad Sci U S A, 2010. **107**(32): p. 14152-7.

206. Kawamura, T., et al., *Linking the p53 tumour suppressor pathway to somatic cell reprogramming.* Nature, 2009. **460**(7259): p. 1140-4.

207. Hong, H., et al., *Suppression of induced pluripotent stem cell generation by the p53-p21 pathway.* Nature, 2009. **460**(7259): p. 1132-5.

208. Banito, A., et al., *Senescence impairs successful reprogramming to pluripotent stem cells.* Genes Dev, 2009. **23**(18): p. 2134-9.

209. Edel, M.J., et al., *Rem2 GTPase maintains survival of human embryonic stem cells as well as enhancing reprogramming by regulating p53 and cyclin D1.* Genes Dev, 2010. **24**(6): p. 561-73.

210. Li, H., et al., *The Ink4/Arf locus is a barrier for iPS cell reprogramming.* Nature, 2009. **460**(7259): p. 1136-9.

211. Buganim, Y., D.A. Faddah, and R. Jaenisch, *Mechanisms and models of somatic cell reprogramming.* Nat Rev Genet, 2013. **14**(6): p. 427-39.

212. Papp, B. and K. Plath, *Epigenetics of reprogramming to induced pluripotency.* Cell, 2013. **152**(6): p. 1324-43.

213. Ang, Y.S., et al., *Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network.* Cell, 2011. **145**(2): p. 183-97.

214. Mansour, A.A., et al., *The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming.* Nature, 2012. **488**(7411): p. 409-13.

215. Onder, T.T., et al., *Chromatin-modifying enzymes as modulators of reprogramming.* Nature, 2012. **483**(7391): p. 598-U119.

216. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.* Genes Dev, 2002. **16**(22): p. 2893-905.

217. Shinagawa, T., et al., *Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells.* Cell Stem Cell, 2014. **14**(2): p. 217-27.

218. Pawlak, M. and R. Jaenisch, *De novo DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state.* Genes Dev, 2011. **25**(10): p. 1035-40.

219. Hu, X., et al., *Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming.* Cell Stem Cell, 2014. **14**(4): p. 512-22.

220. Hyun, K., et al., *Writing, erasing and reading histone lysine methylations.* Exp Mol Med, 2017. **49**(4): p. e324.

221. Wang, Q., et al., *Lithium, an anti-psychotic drug, greatly enhances the generation of induced pluripotent stem cells.* Cell Research, 2011. **21**(10): p. 1424-1435.

222. Wang, T., et al., *The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner.* Cell Stem Cell, 2011. **9**(6): p. 575-87.

223. Yamanaka, S., *Elite and stochastic models for induced pluripotent stem cell generation.* Nature, 2009. **460**(7251): p. 49-52.

224. Hanna, J., et al., *Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency.* Cell, 2008. **133**(2): p. 250-64.

225. Aoi, T., et al., *Generation of pluripotent stem cells from adult mouse liver and stomach cells.* Science, 2008. **321**(5889): p. 699-702.

226. Smith, Z.D., C. Sindhu, and A. Meissner, *Molecular features of cellular reprogramming and development.* Nat Rev Mol Cell Biol, 2016. **17**(3): p. 139-54.

227. Buganim, Y., et al., *Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase.* Cell, 2012. **150**(6): p. 1209-22.

228. Soufi, A., G. Donahue, and K.S. Zaret, *Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome.* Cell, 2012. **151**(5): p. 994-1004.

229. Liu, K., et al., *Chemical Modulation of Cell Fate in Stem Cell Therapeutics and Regenerative Medicine.* Cell Chem Biol, 2016. **23**(8): p. 893-916.

230. Samavarchi-Tehrani, P., et al., *Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming.* Cell Stem Cell, 2010. **7**(1): p. 64-77.

231. Golipour, A., et al., *A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network.* Cell Stem Cell, 2012. **11**(6): p. 769-82.

232. Polo, J.M., et al., *A molecular roadmap of reprogramming somatic cells into iPS cells.* Cell, 2012. **151**(7): p. 1617-32.

233. Hansson, J., et al., *Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency.* Cell Rep, 2012. **2**(6): p. 1579-92.

234. Lin, C.Y., et al., *Transcriptional Amplification in Tumor Cells with Elevated c-Myc.* Cell, 2012. **151**(1): p. 56-67.

235. Nie, Z.Q., et al., *c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells.* Cell, 2012. **151**(1): p. 68-79.

236. Carey, B.W., et al., *Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells.* Cell Stem Cell, 2011. **9**(6): p. 588-98.

237. Stadtfeld, M., et al., *Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells.* Nature, 2010. **465**(7295): p. 175-81.

238. Sridharan, R., et al., *Role of the murine reprogramming factors in the induction of pluripotency.* Cell, 2009. **136**(2): p. 364-77.

239. Takahashi, K., et al., *Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm.* Nat Commun, 2014. **5**: p. 3678.

240. Dalton, S., *Signaling networks in human pluripotent stem cells.* Curr Opin Cell Biol, 2013. **25**(2): p. 241-6.

241. Xu, C., et al., *Feeder-free growth of undifferentiated human embryonic stem cells.* Nat Biotechnol, 2001. **19**(10): p. 971-4.

242. Beattie, G.M., et al., *Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers.* Stem Cells, 2005. **23**(4): p. 489-95.

243. Xu, R.H., et al., *NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs.* Cell Stem Cell, 2008. **3**(2): p. 196-206.

244. McLean, A.B., et al., *Activin a efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed.* Stem Cells, 2007. **25**(1): p. 29-38.

245. Bendall, S.C., et al., *IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro.* Nature, 2007. **448**(7157): p. 1015-U3.

246. Sato, N., et al., *Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor.* Nat Med, 2004. **10**(1): p. 55-63.

247. Singh, A.M., et al., *Signaling Network Crosstalk in Human Pluripotent Cells: A Smad2/3-Regulated Switch that Controls the Balance between Self-Renewal and Differentiation.* Cell Stem Cell, 2012. **10**(3): p. 312-326.

248. Zhou, J., et al., *mTOR supports long-term self-renewal and suppresses mesoderm and endoderm activities of human embryonic stem cells.* Proc Natl Acad Sci U S A, 2009. **106**(19): p. 7840-5.

249. Davidson, K.C., et al., *Wnt/beta-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4.* Proc Natl Acad Sci U S A, 2012. **109**(12): p. 4485-90.

250. Singh, A.M., et al., *Reconciling the different roles of Gsk3beta in "naive" and "primed" pluripotent stem cells.* Cell Cycle, 2012. **11**(16): p. 2991-6.

251. Sumi, T., et al., *Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling.* Development, 2008. **135**(17): p. 2969-79.

252. Shi, Y., et al., *A combined chemical and genetic approach for the generation of induced pluripotent stem cells.* Cell Stem Cell, 2008. **2**(6): p. 525-8.

253. Shi, Y., et al., *Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds.* Cell Stem Cell, 2008. **3**(5): p. 568-74.

254. Mikkelsen, T.S., et al., *Dissecting direct reprogramming through integrative genomic analysis.* Nature, 2008. **454**(7200): p. 49-55.

255. Zhao, Y., et al., *A XEN-like State Bridges Somatic Cells to Pluripotency during Chemical Reprogramming.* Cell, 2015. **163**(7): p. 1678-91.

256. Ichida, J.K., et al., *A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog.* Cell Stem Cell, 2009. **5**(5): p. 491-503.

257. Zhu, S., et al., *Reprogramming of human primary somatic cells by OCT4 and chemical compounds.* Cell Stem Cell, 2010. **7**(6): p. 651-5.

258. Wray, J., et al., *Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation.* Nat Cell Biol, 2011. **13**(7): p. 838-45.

259. Silva, J., et al., *Promotion of reprogramming to ground state pluripotency by signal inhibition.* PLoS Biol, 2008. **6**(10): p. e253.

260. Li, W., et al., *Generation of human-induced pluripotent stem cells in the absence of exogenous Sox2.* Stem Cells, 2009. **27**(12): p. 2992-3000.

261. Wang, Y. and J. Adjaye, *A cyclic AMP analog, 8-Br-cAMP, enhances the induction of pluripotency in human fibroblast cells.* Stem Cell Rev Rep, 2011. **7**(2): p. 331-41.

262. Hou, P., et al., *Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds.* Science, 2013. **341**(6146): p. 651-4.

263. Fritz, A.L., et al., *cAMP and EPAC Signaling Functionally Replace OCT4 During Induced Pluripotent Stem Cell Reprogramming.* Mol Ther, 2015. **23**(5): p. 952-963.

264. Zhang, J., et al., *Metabolic regulation in pluripotent stem cells during reprogramming and self-renewal.* Cell Stem Cell, 2012. **11**(5): p. 589-95.

265. Folmes, C.D., et al., *Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming.* Cell Metab, 2011. **14**(2): p. 264-71.

266. Ma, T.H., et al., *Atg5-independent autophagy regulates mitochondrial clearance and is essential for iPSC reprogramming.* Nature Cell Biology, 2015. **17**(11): p. 1379-1387.

267. Yuan, X., et al., *Brief Report: Combined Chemical Treatment Enables Oct4-Induced Reprogramming from Mouse Embryonic Fibroblasts.* Stem Cells, 2011. **29**(3): p. 549-553.

268. Mali, P., et al., *Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes.* Stem Cells, 2010. **28**(4): p. 713-20.

269. Xie, M., et al., *Pharmacological Reprogramming of Somatic Cells for Regenerative Medicine.* Acc Chem Res, 2017. **50**(5): p. 1202-1211.

270. Cheng, X., et al., *Identification of 2-[4-[(4-Methoxyphenyl)methoxy]-phenyl]acetonitrile and Derivatives as Potent Oct3/4 Inducers.* J Med Chem, 2015. **58**(12): p. 4976-83.

271. Jin, B., et al., *Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells.* Cell Rep, 2012. **2**(5): p. 1411-24.

272. Gafni, O., et al., *Derivation of novel human ground state naive pluripotent stem cells.* Nature, 2013. **504**(7479): p. 282-6.

273. Loewer, S., et al., *Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells.* Nat Genet, 2010. **42**(12): p. 1113-7.

274. Pushpakom, S., et al., *Drug repurposing: progress, challenges and recommendations.* Nat Rev Drug Discov, 2019. **18**(1): p. 41-58.

275. Baell, J. and M.A. Walters, *Chemical con artists foil drug discovery.* Nature, 2014. **513**(7519): p. 481-483.

276. Watanabe, K., et al., *A ROCK inhibitor permits survival of dissociated human embryonic stem cells.* Nat Biotechnol, 2007. **25**(6): p. 681-6.

277. Wang, J., et al., *Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells.* Nature, 2014. **516**(7531): p. 405-9.

278. Krishnakumar, R. and R.H. Blelloch, *Epigenetics of cellular reprogramming.* Curr Opin Genet Dev, 2013. **23**(5): p. 548-55.

279. *Writing R extensions*, in *Vienna, Austria: R Foundation for Statistical Computing*. 2003.

280. Vinogradova, M., et al., *An inhibitor of KDM5 demethylases reduces survival of drug-tolerant cancer cells.* Nat Chem Biol, 2016. **12**(7): p. 531-8.

281. Wang, L., et al., *A small molecule modulates Jumonji histone demethylase activity and selectively inhibits cancer growth.* Nat Commun, 2013. **4**: p. 2035.

282. Rose, N.R., et al., *Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases.* J Med Chem, 2012. **55**(14): p. 6639-43.

283. Kruidenier, L., et al., *A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response.* Nature, 2012. **488**(7411): p. 404-8.

284. Wang, G., et al., *Noggin and bFGF cooperate to maintain the pluripotency of human embryonic stem cells in the absence of feeder layers.* Biochem Biophys Res Commun, 2005. **330**(3): p. 934-42.

285. Richards, M., et al., *The transcriptome profile of human embryonic stem cells as defined by SAGE.* Stem Cells, 2004. **22**(1): p. 51-64.

286. Shah, S.N., et al., *HMGA1 Reprograms Somatic Cells into Pluripotent Stem Cells by Inducing Stem Cell Transcriptional Networks.* Plos One, 2012. **7**(11).

287. Pells, S., et al., *Novel Human Embryonic Stem Cell Regulators Identified by Conserved and Distinct CpG Island Methylation State.* PLoS One, 2015. **10**(7): p. e0131102.

288. Hayashi, Y., et al., *BMP-SMAD-ID promotes reprogramming to pluripotency by inhibiting p16/INK4A-dependent senescence.* Proc Natl Acad Sci U S A, 2016. **113**(46): p. 13057-13062.

289. Chen, J., et al., *BMPs functionally replace Klf4 and support efficient reprogramming of mouse fibroblasts by Oct4 alone.* Cell Res, 2011. **21**(1): p. 205-12.

290. Cheng, X.L., et al., *Ethyl 2-((4-Chlorophenyl)amino)thiazole-4-carboxylate and Derivatives Are Potent Inducers of Oct3/4.* Journal of Medicinal Chemistry, 2015. **58**(15): p. 5742-5750.

291. Takahashi, K. and S. Yamanaka, *A decade of transcription factor-mediated reprogramming to pluripotency.* Nature Reviews Molecular Cell Biology, 2016. **17**(3): p. 183-193.

292. Yamanaka, S. and H.M. Blau, *Nuclear reprogramming to a pluripotent state by three approaches.* Nature, 2010. **465**(7299): p. 704-712.

293. Dabiri, Y., et al., *Imidazopyridines as Potent KDM5 Demethylase Inhibitors Promoting Reprogramming Efficiency of Human iPSCs.* iScience, 2019. **12**: p. 168-181.

294. Li, W., et al., *Identification of Oct4-activating compounds that enhance reprogramming efficiency.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(51): p. 20853-20858.

295. Di, L., *The Impact of Carboxylesterases in Drug Metabolism and Pharmacokinetics.* Curr Drug Metab, 2019. **20**(2): p. 91-102.

296. Gaspar-Maia, A., et al., *Open chromatin in pluripotency and reprogramming.* Nat Rev Mol Cell Biol, 2011. **12**(1): p. 36-47.

297. Soufi, A., et al., *Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming.* Cell, 2015. **161**(3): p. 555-568.

298. Kooistra, S.M. and K. Helin, *Molecular mechanisms and potential functions of histone demethylases.* Nat Rev Mol Cell Biol, 2012. **13**(5): p. 297-311.

299. Chen, J., et al., *H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs.* Nat Genet, 2013. **45**(1): p. 34-42.

300. Schmitz, S.U., et al., *Jarid1b targets genes regulating development and is involved in neural differentiation.* EMBO J, 2011. **30**(22): p. 4586-600.

301. Xie, L., et al., *KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription.* EMBO J, 2011. **30**(8): p. 1473-84.

302. Pfaff, N., et al., *Inhibition of miRNA-212/132 improves the reprogramming of fibroblasts into induced pluripotent stem cells by de-repressing important epigenetic remodelling factors.* Stem Cell Res, 2017. **20**: p. 70-75.

303. Davidson, K.C., E.A. Mason, and M.F. Pera, *The pluripotent state in mouse and human.* Development, 2015. **142**(18): p. 3090-9.

304. Sumter, T.F., et al., *The High Mobility Group A1 (HMGA1) Transcriptome in Cancer and Development.* Curr Mol Med, 2016. **16**(4): p. 353-93.

305. Deplancke, B., D. Alpern, and V. Gardeux, *The Genetics of Transcription Factor DNA Binding Variation.* Cell, 2016. **166**(3): p. 538-554.

306. Parisi, S., et al., *HMGA Proteins in Stemness and Differentiation of Embryonic and Adult Stem Cells.* Int J Mol Sci, 2020. **21**(1).

307. Wang, R.N., et al., *Bone Morphogenetic Protein (BMP) signaling in development and human diseases.* Genes Dis, 2014. **1**(1): p. 87-105.

308. Bunz, F., et al., *Requirement for p53 and p21 to sustain G2 arrest after DNA damage.* Science, 1998. **282**(5393): p. 1497-501.

309. Harrer, M., et al., *Dynamic interaction of HMGA1a proteins with chromatin.* J Cell Sci, 2004. **117**(Pt 16): p. 3459-71.

310. Dabiri, Y., et al., *The essential role of TAp73 in bortezomib-induced apoptosis in p53-deficient colorectal cancer cells.* Scientific Reports, 2017. **7**(1): p. 5423.

311. Dabiri, Y., et al., *p53-Dependent Anti-Proliferative and Pro-Apoptotic Effects of a Gold(I) N-Heterocyclic Carbene (NHC) Complex in Colorectal Cancer Cells.* Front Oncol, 2019. **9**: p. 438.

312. Dabiri, Y., et al., *A Ruthenium(II) N-Heterocyclic Carbene (NHC) Complex with Naphthalimide Ligand Triggers Apoptosis in Colorectal Cancer Cells via Activating the ROS-p38 MAPK Pathway.* Int J Mol Sci, 2018. **19**(12).

313. Grabundzija, I., et al., *Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells.* Nucleic Acids Res, 2013. **41**(3): p. 1829-47.

314. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biology, 2004. **5**(10): p. R80.

315. Dunning, M.J., et al., *beadarray: R classes and methods for Illumina bead-based data.* Bioinformatics, 2007. **23**(16): p. 2183-4.

316. Ashburner, M., et al., *Gene Ontology: tool for the unification of biology.* Nat Genet, 2000. **25**.

317. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.

318. Reimand, J., et al., *g:Profiler-a web server for functional interpretation of gene lists (2016 update).* Nucleic Acids Res, 2016. **44**(W1): p. W83-9.