THE BIOLOGICAL AND CLINICAL IMPLICATIONS OF PLASMINOGEN ACTIVATION IN PANCREATIC AND LUNG CARCINOMAS

by

Moamen Bydoun

Submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

February 2018

…Dedicated to baba

Table of Contents

List of tables

List of figures

Abstract

The blood protein plasminogen circulates as an inactive precursor of the serine protease plasmin. Plasmin prevents aberrant formation of blood clots and protects individuals from vascular/tissue damage. Plasmin is also an extracellular proteolytic agent that is often exploited by malignant cancers to facilitate their escape from the confinements of the extracellular matrix. The initiation of invasion and metastasis by cancer cells has been linked to epithelial to mesenchymal transition (EMT). EMT prompts cancer cells to lose epithelial proteins and acquire versatile characteristics thereby permitting mesenchymal migration and movement. The invasive process often associates increased plasminogen activation with mesenchymal cancer cells. However, the two distinct processes of plasminogen activation and EMT are not yet functionally linked. The **first objective** of this dissertation was to characterize differentially-expressed components of the plasminogen activation system in lung cancer cells undergoing EMT. This objective was addressed using various models of epithelial-like and mesenchymal-like cells. Specifically, we demonstrated that the plasminogen receptor S100A10, the plasminogen activator receptor uPAR and the plasminogen activation inhibitor PAI-1 were differentially regulated in epithelial vs. mesenchymal cells. The expression and localization of these proteins modulated plasminogen activation at the cell surface. Importantly, we demonstrated that epithelial cells and not mesenchymal cells display marked levels of plasminogen activation. The **second objective** was to assess genes involved in plasminogen activation as potential predictors of patient outcome in non-small cell lung cancer using hierarchical clustering strategies in merged patient cohorts. We identified a list of candidate markers of which four genes (*PLAUR*, *PLAU*, *ANXA2* and *S100A10*) emerged as strong predictors of overall survival. The **third objective** was to study the biological and clinical implications of S100A10 in pancreatic cancer. We showed that pancreatic carcinoma overexpressed S100A10 compared to early-stage lesions, stroma and normal tissues. S100A10 mRNA levels were also predictive of overall and recurrence-free survival in pancreatic cancer patients. The expression of S100A10 was largely driven by the oncogene *KRAS* and by DNA methylation of its promoter region. Together, these findings delineated a fundamental role of plasminogen activation, particularly that of S100A10 in lung and pancreatic carcinomas.

List of Abbreviations Used

Abbreviation	full description
4E-BP1	Eukaryotic translation initiation factor 4E-binding protein 1
$5\text{-}HT_{1B}$	5-hydroxytryptamine
5-mC	5-methylcytosine
5'UTR	5' untranslated region
A2M	Alpha-2-macroglobulin
A405	Absorbance at 405 nanometers
ACA	ε-aminocaproic acid
ACTA2	Actin, alpha 2, smooth muscle, aorta
ACTB	*β*-actin
ACVR1B	Activing A receptor type 1B
ADAM	A disintegrin and metalloproteinase
ADAMTS	ADAMs with thrombospondin motifs
ADEX	Aberrantly differentiated endocrine exocrine
ADP	Adenosine diphosphate
AFP	Serum alpha-fetoprotein
AHNAK	AHNAK nucleoprotein
AJ	Adherens junctions
Akt (PKB)	Protein kinase B
ALB	Albumin
ALK	ALK receptor tyrosine kinase
AML	Acute myeloid leukemia
Ap	Aprotinin
APC2	APC2, WNT signaling pathway regulator
Arg	Arginine
ARHGEF7	Rho guanine nucleotide exchange factor 7
ARID1A	AT-rich interaction domain 1A
ASCO	American society of clinical oncology
Asn	Asparagine
Asp	Aspartic acid
ATAD4 (PRR15L)	Proline rich 15 like
ATCC	American type culture collection
ATF2	Activating Transcription Factor 2
ATG9	Autophagy related 9
ATM	ataxia-telangiectasia mutated
ATRA	All-trans retinoic acid

bhFGF/H	Basic human fibroblast growth factor in heparin salt
BMP3	Bone morphogenetic protein 3
BRAF	B-Raf proto-oncogene, serine/threonine kinase
BRAF	Murine sarcoma viral oncogene homolog B
BRCA1	Breast cancer 1
BRCA2	Breast cancer 2
BTNL8	Butyrophilin like 8
C-1-P	Sphingolipid ceramide 1-phosphate
C.I	Confidence interval
CA-125	Cancer antigen 125
CASP10	Caspase 10
CCDC6	Coiled-coil domain containing 6
CCL13/CCL18	C-C motif chemokine ligand 13/18
CCLE	Cancer cell line encyclopedia
CCND1	Cyclin D1
CCR10	C-C motif chemokine receptor 10
CD4/8	Cluster of differentiation 4/8
Cdc42	Cell division control protein 42 homolog
CDH1	Cadherin 1 (also known as E-cadherin)
CDH10	Cadherin 10
CDH2	Cadherin 2 (also known as N-cadherin)
CDHA	Capital district health authority
CDK6	Cyclin-dependent kinase 6
CDKN2A	Cyclin-dependent kinase inhibitor 2A
cDNA	Complementary DNA
CFTR	Cystic fibrosis transmembrane conductance regulator
CHIP	Chromatin immunoprecipitation
CHRNA7	Cholinergic Receptor Nicotinic Alpha 7 Subunit
CK	Cytokeratin
CML	Chronic myelogenous leukemia
CNA	Copy-number aberrations
CNTN5	Contactin 5
COX2	Cytochrome c oxidase subunit 2
CpG	Cytosine-phosphate-guanine oligodeoxynucleotide
CSF1	Macrophage colony-stimulating factor 1 precursor
CT	Computed tomography
CTC	Circulating tumor cell
CTS	Cathepsin S
CTSB	Cathepsin B
CTSL	Cathepsin L

CYCC	Cyclophilin C
DAB	3,3'-diaminobenzidine
DAPL	Death-associated protein kinase
DDR	DNA damage response
DES	Desmin
DLBCL	Diffuse large B-cell lymphoma
DLC1	Deleted in liver cancer 1
DMEM	Dulbecco's modified eagle's media
DNA	Deoxyribonucleic acid
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3 alpha
DNMT3B	DNA methyltransferase 3 beta
DOCK2	Dedicator of cytokinesis 2
Doxy	Doxycycline
DSC1,2, 3	Desmocollin 1, 2, 3
DSG1, 2, 3, 4	Desmoglein 1, 2, 3, 4
E-cadherin	Epithelial cadherin
ECM	Extracellular matrix
ED	Extracellular domain
EDTA	Ethylenediaminetetraacetic acid
EF1A	Elongation factor 1α
EGAD	Enhanced Gene Analysis and Discovery
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EGTA	Ethyleneglycol-bis(β-aminoethylether)-N,N,N',N'-tetraacetic acid
EMBL-EBI	European molecular biology laboratory – European bioinformatics institute
EML4	Echinoderm microtubule associated protein like 4
EMT	Epithelial to mesenchymal transition
EMT-ATF	Epithelial to mesenchymal transition activating transcription factor
ENO1	Enolase 1
EORTC	European organization for research and treatment of cancer
ERBB2	Erb-B2 receptor tyrosine kinase 2
ERCC4/ERCC6	ERCC excision repair 4/6, endonuclease catalytic subunit
FAM3D	Family with sequence similarity 3 member D
FANCC	Fanconi anemia group C protein
FANCG	Fanconi anemia group G protein
FBS	Fetal bovine serum

FBXW7	F-Box and WD repeat domain containing 7
FDA	Food and drug administration
FFPE	Formalin-fixed and paraffin-embedded
FGF	Fibroblast growth factor
FGFR1	Fibroblast growth factor receptor 1
FHIT	Fragile histidine triad
FIG (GOPC)	Golgi associated PDZ and coiled-coil motif containing
FISH	Fluorescent in situ hybridization
FOXA2/3	Forkhead box A2/A3
FOXC2	Forkhead box C2
FoxD3	Forkhead box 3
FSP1	Fibroblast-specific protein 1
FUS1	FUS RNA binding protein
GAP	GTPase-activating proteins
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GATA6	GATA binding protein 6
GDC	Genomic data commons
GDP	Guanosine diphosphate
GEF	nucleotide exchange factor
GEO	Gene expression Omnibus
GJ	Gap junctions
GLI1/GLI3	GLI family zinc finger 1/3
GPCR	G-protein-coupled receptors
GPM6B	Glycoprotein M6B
Grb2	Growth factor receptor binding protein 2
GTP	Guanosine triphosphate
GTPase	Guanosine triphosphatase
H & E	Hematoxylin and eosin
h0(t)	Baseline hazard rate
H2A, H2B, H3, H4	Histone 2A, 2B, H3 and H4
HA	Hyaluronic acid
HCC	Hepatocellular carcinoma
HDAC	Histone deacetylase
HEK 293	Human embryonic kidney cells 293
HGF	Hepatocyte growth factor
HIBEC	Human intrahepatic biliary epithelial cells
HIF1α	Hypoxia inducible factor 1 alpha
HIP1	Huntingtin interacting protein 1
His	Histidine
HIST2H2BE	Histone cluster 2 H2B family member E

HM450	Human methylation (450,000 probes)
HMGB1	High mobility group box 1
HMLE	Human mammary epithelial cell line
HPRT	Hypoxanthine phosphoribosyl transferase
HPV16	Human papilloma virus 16
HR	Hazard ratio
ICGC	International cancer genome consortium
ID	Intracellular domain
IDT	Integrated DNA technologies
IFN-γ	Interferon gamma
IGF-1	Insulin-like growth factor 1
IGF2R	Insulin Like growth factor 2 receptor
IHC	Immunohistochemistry
IL-6	Interleukin 6
Ile	Isoleucine
ITGA4/ITGA9	Integrin A4/A9
ITGAM	Integrin subunit alpha M
JAK	Janus kinase
Jnk	c-Jun N-terminal kinase
KAT3A	CREB binding protein 3A
KAT3B	Histone acetyltransferase P300
KAT6A	Lysine acetyltransferase 6A
KAT6B	Lysine acetyltransferase 6B
Kd	Dissociation constant
kDa	Kilo daltons
KDM6A	Lysine demethylase 6A
Ki67	Antigen identifed by monoclonal antibody Ki-67
KIF5B	Kinesin family member 5B
KM	Kaplan meier
K_m	Michaelis constant
KRAS	Kirsten ras sarcoma
L-CAM	Liver cell adhesion molecule
LATS1	Large tumor suppressor kinase 1
LCNEC	Large cell neuroendocrine carcinoma
LLC	Lewis lung carcinoma
LMTK2	Lemur tyrosine kinase 2
lncRNA	Long-non-coding ribonucleic acid
LOH	Loss of heterozygosity
LY	Life saved
Lys	Lysine

MAP2K4	Mitogen-activated protein kinase kinase 4
MAPK	Mitogen activated protein kinase
MBD1, 2, 3	methyl-CpG binding domain protein 1, 2, 3
MCF-7	Michigan Cancer Foundation-7
MDCK cells	Madin-darby canine kidney cells
MeCp2	methyl-CpG binding protein 2
MEK1 (MAPKK)	dual specificity mitogen-activated protein kinase kinase 1
MEP1A	Meprin A subunit alpha
MET (gene)	MET proto-oncogene, receptor tyrosine kinase
MET (process)	Mesenchymal to epithelial transition
mGK6	Glandular kallikrein
mGluR5	Metabotropic glutamate receptor 5
MGMT	6-methylguanine DNA methyltransferase
miRNA	Micro ribonucleic acid
MLL2	Myeloid/lymphoid or mixed-lineage leukemia 2
MMP	Matrix metalloproteinase
MNX1	Motor neuron and pancreas homeobox 1
mRNA	Messenger ribonucleic acid
MT-MMP	Membrane-tethered matrix metalloproteinase
mTOR	Mammalian target of rapamycin
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium
MUC13	Mucin 13
Munc13-4	Protein unc-13 homolog D
MYC	MYC proto-oncogene, BHLH transcription factor
N-cadherin	Neural cadherin
Na(V)1.8	Tetradotoxin-resistant sodium channel
NaCl	Sodium chloride
NCI	National cancer institute
NDRG4	N-myc downstream regulated gene 4
NEUROD1	Neuronal Differentiation 1
NGS	Next-generation sequencing
NHERF	Na+/H+ exchanger regulatory factor
NKX2-2	NK2 Homeobox 2
NMuMG	Normal murine mammary gland
NNMF	Non-negative matrix factorization
NOD-SCID	Non-obese diabetic/severe combined immunodeficiency
Notch	Notch homolog 1, translocation associated (Drosophila)
NP-40	Nonyl phenoxypolyethoxylethanol
NR5A2	Nuclear receptor subfamily 5 group A member 2

NSCLC	Non-small-cell lung cancer
NT5E	5'-Nucleotidase Ecto
OS	Overall survival
p-S6K	Ribosomal protein S6 kinase beta-1
p14ARF	P14 ARF subunit of CDKN2A
p16INK4A	P16 INK4A subunit of CDKN2A
PAI-1	Plasminogen activator inhibitor 1
PAI-2	Plasminogen activator inhibitor 2
PAK2	P21 (RAC1) activated kinase 2
PALB2	Partner and localizer of BRCA2
PanINs	Pancreatic intraepithelial neoplasia
PAR6	Partitioning defective 6
PARK2 (PRKN)	Parkin RBR E3 ubiquitin protein ligase
PBS	Phosphate-buffered saline
PDAC	Pancreatic ductal adenocarcinoma
PDGF	Platelet-derived growth factor
PDGFR	platelet-derived growth factor receptor
PDX1	Pancreatic and duodenal homeobox 1
Pg	Plasminogen
PI3K	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PIK3R3	Phosphoinositide-3-kinase regulatory subunit 3
PKC	Protein kinase C
PLAT (tPA)	Tissue plasminogen activator
PLAU (uPA)	Urokinase plasminogen activator
PLCB3	Phospholipase C beta 3
PLGRKT	Plasminogen receptor with A C-terminal lysine
PML-RARα	Promyelocytic leukemia- retinoic acid receptor alpha
PNLIPRP2	Pancreatic-lipase-related protein 2
PPP2R3A	Protein phosphatase 2 regulatory subunit B alpha
PREX2	Phosphatidylinositol-3,4,5-trisphosphate dependent Rac Exchange factor 2
PRSS1	Protease serine 1
PRSS23	Protease, serine 23
PSA	Prostate-specific antigen
PTEN	Phosphatase and tensin Homolog
puro	Puromycin
QALY	Quality-of-life-adjusted year of life saved
QLQ-C30	Quality of life questionnaire core 30
QM	Quasi-mesenchymal

Rac1	Ras-related C3 botulinum toxin substrate 1
RASGRP3	RAS guanyl releasing protein 3
RASSF1A	Ras association domain family member 1
RB1	Retinoblastoma 1
RBPJL	Recombination signal binding protein for immunoglobulin Kappa J region like
REG1B	Regenerating islet-derived 1 beta
RET	Ret proto-oncogene
REV	Raw expression value
REV3L	REV3 like, DNA directed polymerase zeta catalytic subunit
RFI	Relative fluorescent intensity
RFS	Recurrence-free survival
Rho A	Ras homolog gene family member A
RMI	Risk of malignancy index
RNA	Ribonucleic acid
RNA-Seq	Ribonucleic acid sequencing
ROBO	Roundabout
ROBO2	Roundabout guidance receptor 2
ROCA	Risk of ovarian cancer algorithm
ROS	Reactive oxygen species
ROS1	ROS proto-oncogene 1, receptor tyrosine kinase
RP1	RP1, axonemal microtubule associated
RPA1	Replication protein A1
RPL32	Ribosomal protein L32
RPMI	Roswell park memorial institute
RSCL	Rotterdam symptom checklist
RUVBL1	RuvB like AAA ATPase 1
S100A1	S100 calcium binding protein A1
S6K1	Ribosomal protein S6 kinase beta-1
SAGE	Serial analysis of gene expression
SASH1	SAM and SH3 domain containing 1
SCLC	Small-cell lung cancer
scuPA	Single-chain pro-urokinase
SDS	Sodium dodecyl sulfate
Ser	Serine
SERPINB2 (PAI-2)	Serpin family B member 2
SERPINE1 (PAI-1)	Serpin family E member 1
SERPINE2 (PI7)	Serpin family E member 2 (protease nexin I)
SERPINF2 (α2-anti-plasmin)	Serpin family F member 2

SERPINI2 (PI12)	Serpin family I member 2 (neuroserpin)
ShcA (SHC1)	SHC-transforming protein 1
SHH	Sonic hedgehog
SHIP1	Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1
shRNA	Short hairpin RNA
siRNA	Short interfering RNA
SIRT	Sirtuin
SLC34A2	Solute carrier family 34 member 2
Slug (SNAI2)	Snail family transcriptional repressor 2
Smad	Mothers against decapentaplegic
SMC2	Structural maintenance of chromosomes protein 2
Snail (SNAI1)	Snail family transcriptional repressor 1
SnoRNA	Small nucleolar ribonucleic acid
SNP	Single nucleotide polymorphisms
SNV	Single nucleotide variant
Sos1	Son of sevenless homolog 1
SOX3	SRY-Box 3
SPINK1	Serine peptidase inhibitor, kazal type 1
STAT3	Signal transducer and activator of transcription 3
T-ALL	T-cell acute lymphoblastic leukemia
TAK1 (MAP3K7)	Mitogen-activated protein kinase kinase kinase 7
TAM	Tumor-associated macrophage
TASK-1	Tandem pore (2P) domain potassium channel
TBP	TATA sequence binding protein
TBX5	T-Box 5
TCGA	The Cancer Genome Atlas
T_cSS	Transcription start site
TERT	Telomerase reverse transcriptase
TFF1	Trefoil factor 1
TGFB1	Transforming growth factor beta 1
TGFα	Transforming growth factor alpha
TIF-1	Thyroid transcription factor 1
TIME	Telomerase immortalized microvascular endothelial
TIMP	Tissue inhibitor of metalloproteinases
TJ	Tight junctions
TKI	Tyrosine kinase inhibitor
TMA	Tissue microarray
TMEM45B	Transmembrane protein 45B
TNFα	Tumor necrosis factor alpha
TNM	Tumor, node, metastasis

TP53	Tumor protein 53
TP63	Tumor protein 63
TRAF6	TNF receptor associated factor 6
Tris	Trisaminomethane
TRPV5 and TRPV6	Transient receptor potential cation channel subfamily V member 5/6
TSC2	TSC complex subunit 2
TSGCT	Tenosynovial giant cell tumor
Twist1	Twist family BHLH transcription factor 1
TβRI (TGFBR1)	Transforming growth factor beta receptor 1
TβRII (TGFBR2)	Transforming growth factor beta receptor 2
UCA1	Urothelial cancer associated 1
ULK1	Unc-51 like autophagy activating kinase 1
UTX	Histone demethylase
Val	Valine
VE-cadherin	Vascular endothelial cadherin
VEGF	Vascular endothelial growth factor
VGLL1	Vestigial like family member 1
VIM	Vimentin
VWF	Von Wille-Brand factor
WHO	World Health Organization
Wnt	Wingle integration site
WPB	Weibel-Palade bodies
Z	Vector of co-variants
ZEB1/2	Zinc finger E-box binding homeobox 1/2
ZNRF3	Zinc and ring finger 3
ZO-1, -2	Zona occludens -1, -2
α-SMA	Alpha smooth muscle actin
β’	Vector of coefficients
β2M	*β2*-microglobulin
ACT	Adjuvant cisplatin/vinorelbine
REVIGO	Reduce and visualize gene ontology
7AAD	7-aminoactinomycin D
PI	Propidium iodide
HEPES	-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
FITC	Fluorescein isothiocyanate
BDNF	Brain derived neurotrophic factor
CCDC80	Coiled-coil domain containing 80
AKR1C1	Aldo-keto reductase family 1 member C1
TAFI	Thrombin activatable fibrinolysis inhibitor

Acknowledgements

It is my honor to acknowledge a series of individuals that have made my graduate experience an enjoyable and fruitful period of my life. First, I would like to thank my mentor Dr. David Waisman for this continuous support and encouragements through the ups and downs. I thoroughly enjoyed our hour-long conversations about everything from science to politics and many other topics. Thank you for five years that I will look back at and reminisce about the teachable moments.

Thank you to my committee members (Dr. Patrick Lee (past), Dr. Paola Marcato, Dr. Weei-Yuan Huang and Dr. Shashi Gujar) for their guidance, advice and constructive feedback throughout the years. It helped shape my dissertation the form it is now.

A thank you must also go to the funding agencies that allowed me to explore many projects and get a chance to travel the world and share my research with the scientific community. These agencies are the Beatrice Hunter Cancer Research Institute (BHCRI), Terry Fox Research Institute (TFRI), Nova Scotia Health Research Foundation (NSHRF) and the Killam laureates.

Andra, thank you for the support and encouragement. I could not have done this without you. Ali, thank you for everything and for being “there” for me since day 1. The ultimate thank you goes to my family. Thank you for being very understanding of my unpredictable schedule. Your unconditional love and support will always cherish my heart.

Chapter I: INTRODUCTION

1.1 Subchapter 1: Cancer and Implications on the Patient

1.1.1 Cancer

Cancer is a global health concern with approximately eight million deaths worldwide [1][2]. In Canada, cancer is the leading cause of death (30% of all deaths) with approximately 210 deaths every day [1]. In the USA, Cancer follows heart disease as is the second leading cause of death with 595,690 deaths in 2016 [3]. Beyond statistics, the word "cancer" usually invokes a series of emotions among the public characterized by fear, helplessness and a consensus on the lack of cure. These emotions are driven and rendered complex by the fact that cancer is personified and villainized by the patients and the people around them.

A simplistic definition of cancer is cell proliferation in an uncontrolled manner to form a tumor mass. To the biologist, cancer is a "sped up" version of evolution and a powerful example of Darwinism. In a seminal publication in 1976, Peter Nowell utilized his observations of cellular clonality in blood cancers to generate the theory of clonal evolution and explain cancer initiation and progression [4]. Clonality gives rise to populations that hijack growth signals, evade programs that suppress growth and resist cell death within a supportive primary tumor microenvironment [5]. However, the potential for primary tumor cells to metastasize is what warrants heightened concern and not the mere formation of a primary tumor.

Douglas Hanahan and Robert Weinberg published the original widely-known review "The hallmarks of cancer" in 2000, where they illustrated the complex nature of cancer biology in the form of six hallmarks that still stand true to this day. These hallmarks are self-sufficiency in pro-growth signals, unresponsiveness to anti-growth signals, inhibition of apoptotic signals, unlimited replication capacity, angiogenesis, invasion and metastasis [6]. This partly reductionist hallmark approach was triumphant during the early 2000s until the complexity, the unpredictability and the heterogeneity of cancer toppled such simplistic approach. Not surprisingly, these hallmarks were later expanded in their 2011 review to include two more hallmarks. These two additions include evasion of the immune response and the hijacking of cellular metabolism along two enabling characteristics in the form of genomic instability and highly-inflammatory tumor-promoting microenvironment [5]. Weinberg addressed the issue of trying to simplify the disease as being caused or enabled by six or eight key processes and a select group of driver genes by stating that the attempt to reduce causality into one or two assailants is always met with endless complexity [7].

Understanding the underlying molecular mechanisms of the disease is crucial due to the high prevalence of cancer. Despite the efforts to delineate the cellular and molecular mechanisms in cancer, knowledge gaps still plague the disease. Cancer cell dissemination represents a key process and a turning point in cancer progression. Once cancer cells become invasive and gain the ability to metastasize, patient prognosis and treatment efficacy decrease drastically. Establishing a complete model of the major signaling pathways involved in invasive escape would present a major advancement in the field of

cancer research and therapeutics. The goal is to attain a reasonable improvement in patient outcomes whilst maintaining an adequate quality of life.

1.1.2 Determinants of Patient Outcome

In 1993, the health services research committee of American Society of Clinical Oncology (ASCO) redefined the outcomes of pediatric and adult cancer patients and their responses to treatment. ASCO's guidelines are constantly refined considering new adjustments in current health care systems and new treatment modalities in the era of precision medicine [8]. These fundamental guidelines describe methods of assessing survival, quality of life, treatment toxicity, cost effectiveness as well as measures of patient response to treatment. Other guidelines address issues related to prioritization of patient outcome (i.e. what dictates how to proceed with treatment, patient quality of life vs. cancer response), the need to use multiple outcome measures to determine prospective treatment modalities and concerns regarding how to justify the benefit of treatment to patients, physicians and policy makers.

1.1.2.1 Survival

Patient survival is the most important determinant of patient outcome. Survival is represented through several measurements (summarized and defined in table 1) which include: overall survival, cancer-specific survival, event-free survival, progression-free survival, recurrence-free survival, median survival, disease-free survival, metastasis-free survival and others. Here, an important distinction must be made between survival of one patient and survival of a patient cohort. For instance, survival of a patient is a discrete

measure of how long that patient survives from diagnosis until an event occurs (e.g. death after 12 months from diagnosis). In contrast, the survival of a patient cohort, also dubbed survival "rate", is the percentage of patients that have not experienced an event within a specified duration of time after their diagnosis (e.g. 40% of patients are alive after 12 months of follow-up). This distinction in understanding survival is best explained using the Kaplan Meier estimator [9] (discussed next) (figure 1). Importantly, the larger a patient cohort size, the more representative is the survival function to that of the entire population of patients. Under such situation, the survival rate becomes equivalent to the probability of a patient experiencing an event (e.g. death) after a certain period of time has passed since diagnosis. Using the above example, a patient who survived to 12 months after diagnosis has a 40% chance of surviving their cancer.

As listed in table 1, examples of survival include cancer-specific survival rate which is the probability of a patient dying from their cancer within a period after diagnosis. Most common periods are one-, three- and five-year cancer-specific survivals (figure 1). Event-free survival is related to the absence/presence of any event or outcome related to the disease such as relapse, remission, death etc. [10]. Disease-free survival is the time between response to treatment (or surgery) and the recurrence or relapse of a tumor. Relative survival is a non-parametric measure that compares the number of events one would expect since the previous event if there was no difference between groups [9](table 1). Measures such as disease-free survival is relevant in the adjuvant setting to assess whether surgical and/or therapeutic interventions have been effective in preventing relapse. In contrast, progression-free survival is important in monitoring patients that have developed or at risk

Table 1. Measures of survival as determinants of patient outcome. Definitions of various types of survival as per clinical standards of a cancer patient. In clinical trials, these definitions are known as oncology endpoints.

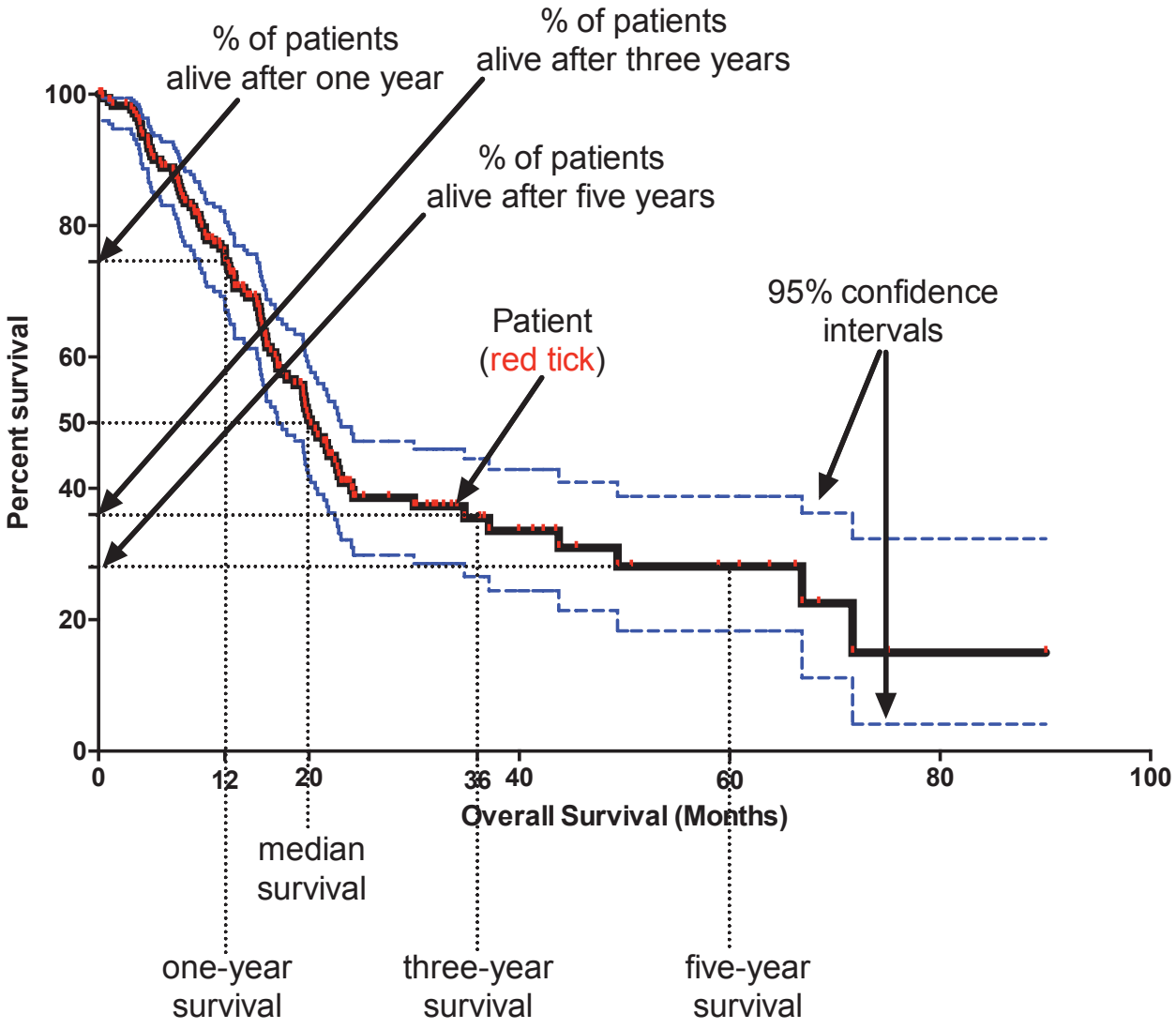
Survival	Description (*in the context of cancer)
Overall survival	A measure of how long a cancer patient survives until death regardless of the cause of death (cancer or unrelated cause)
Relative survival	A measure of overall survival of a cancer patient relative to overall survival of a cancer-free individual in a similar population
Cancer-specific survival	A measure of survival until patient death due to cancer
Median survival	A measure of the time at which half of patients have died
One-, three- and five-year survival	A measure of survival after one-, three- and five years post diagnosis
Disease-free survival	A measure of how long a cancer patient remains cancer-free after therapeutic intervention
Progression-free survival	A measure of how long a tumor remains stable (tumor-free or non-progressing tumor) after therapeutic intervention
Recurrence-free survival	A measure of how long it takes for a tumor to recur (relapse) after therapeutic intervention
Metastasis-free survival	A measure of how long it takes before a patient develops metastatic disease (including a recurring metastatic tumor)
Event-free survival	A measure of how long it takes for a patient to be subject to a pre-determined event (e.g. recurrence, therapy resistance, side effect etc).

of developing metastases [11]. Improvement in all types of survival is favorable and sufficient to justify further treatment while considering quality of life and cost.

1.1.2.2 Kaplan Meier Estimator

The Kaplan Meier (KM) estimator is a non-parametric test that estimates the survival function over time. The survival function on a KM plot offers information on the specific survival of a patient of interest and the percent of patients alive (or event-free) over time or at a particular point in time. The latter is equivalent to the probability of experiencing death (or any event) at that particular time point or over an extended period [12]. The figure legend contains specific information on how to read a KM plot (figure 1).

Figure 1. The Kaplan Meier (KM) estimator. The graph represents a overall survival function of a cohort of patients (178 pancreatic cancer patients in this case). The graph shows the probability of survival of this patient cohort at a designated time interval. The larger a patient cohort is, the closer is the survival function of this cohort (178 patients) to that of the entire population (i.e. all pancreatic cancer patients). Each tick represents a patient and is a measure of his/her survival (on x-axis) at the time of last follow-up. The declining shape of the curve is contributed by event occurrence (death in this case). There are concrete survival definitions that are displayed by a Kaplan Meier plot such as one-, three- and five-year survival of this patient cohort. The percentage of patients that survived is at the point of interception between the curve and a particular time point. Median survival is the time at which half (50%) of the patients in this cohort are still event-free (i.e. alive). The 95% confidence limits of the survivor function are shown. In practice, there are usually patients who are lost to follow-up or alive at the end of follow-up, and confidence limits are often wide at the tail of the curve, making meaningful interpretations difficult.


% of patients
alive after one year
% of patients
alive after three years
% of patients
alive after five years
Patient
(red tick)
95% confidence
intervals
Percent survival
100
80
60
50
40
20
0
0
12
20
36
40
60
80
100
Overall Survival (Months)
median
survival
one-year
survival
three-year
survival
five-year
survival

1.1.3 Quality of Life

The quality of life for a cancer patient is an important concern that not only addresses the 1) physical effects of cancer treatment but also the 2) psychological and 3) social aspects [13]. Physical effects are any symptoms caused by the cancer or by the toxicity of treatment. Toxicity considers multiple dimensions which include the frequency, duration and severity of the treatment and it may invoke disruptions to daily activities such as walking, talking, exercise etc. A classic example is the cardiotoxicity of anthracyclines where the acute and chronic cardiac dysfunction and the appropriate time frame for administering cardio-protective treatments should be considered [14]. The psychological effects are defined as any changes in the cognitive and emotional state of the patient due to cancer or treatment toxicity (depression, anxiety, stress etc.). Social effects are related to changes in social behavior and interpersonal relationships at home, workplace, school or community at large.

Although quality of life measurements are often subjective and a concrete scale may be impractical [15], focusing on the psychological and overall well-being of every individual is important. The term quality-adjusted survival is widely used in clinical trials and accounts for treatment side effects and overall cost. Cost-effectiveness is also a relevant "outcome" and is often represented as the cost per year of life saved (LY) or cost of quality-of-life-adjusted year of life saved (QALY). Cost-effectiveness evaluates the monetary cost of a cancer treatment and compares it to alternative treatment options whilst considering the effect on survival and quality of life [16]. Various methods of measurements have been developed to assess quality of life; these include EORTC (European Organization for

Research and Treatment of Cancer) quality of life questionnaire core 30 items (QLQ-C30), functional assessment of chronic illness therapy (FACIT) measurement system, Rotterdam symptom checklist (RSCL) and others [17] [18]. These measures are held to high standards of validity and reliability and are normally assessed prior to, during and after a treatment. These tests are also meant to be palpable to the patient and are easy to read and complete while remaining sensitive to subtle changes [19]. Quality of life of a cancer patient can also be affected by co-morbid conditions and their respective treatments. For instance, diabetic pancreatic cancer patients who are receiving anti-diabetic medications are at higher risk of dying from their cancer compared to diabetic cancer patients not receiving anti-diabetic medications [20]. Therefore, assessment of quality of life is fundamental in both randomized (Phase III) and non-randomized (Phase I and II) clinical trials [19].

1.1.4 Measure of Cancer Response to Treatment

A measure of a tumor's response to treatment is considered a hallmark of disease progression. These measures include degree of tumor remission (partial or complete) and time to disease progression. For that purpose, multiple cancer biomarkers have been used to monitor cancer response (discussed later). Although many studies demonstrated that there is a positive correlation between increased quality of life and a favorable cancer outcome [18], the aforementioned treatment-induced toxicity can have a negative impact on quality of life despite tumor remission.

1.2 Subchapter 2: Cancer Research in the Era of "Big Data"

The most notable success stories which triggered drastic improvements in the outcome of cancer patients emerged from studies that identified a unique and targetable cancer-causing gene or event (e.g. *BRAF* mutations in melanoma patients) [21]. Early studies had utilized sanger sequencing and comparative genomic hybridization to identify a manageable number of cancer-causing events [22]. Such personalized treatments flourished upon the finalization of the human genome project in 2003 [23]. However, the subsequent advent and prompt availability of high-throughput high-resolution microarray and next generation sequencing revolutionized the unveiling of the genomic landscape of cancers and gave rise to the era of "big data". The big data era resulted in an arguably overwhelming body of genomic information that is now globally used to identify cancer-causing and cancer-promoting alterations to predict patient outcome and to better guide treatment regiments. The new era has also pushed aside the idea of dealing with a "single-gene" disease with small sample sizes and largely inaccessible clinical data. Rather this era has introduced accessibility to an expanding number of patient cohorts and a series of well-annotated clinical data. It should be noted that the broad term "genetic alterations" encompasses single nucleotide variants (SNVs), single nucleotide polymorphisms (SNPs), chromosomal translocations and aberrations, somatic copy-number aberrations (CNAs), transcriptional profiles and epigenetic changes. These alterations have revealed a significant degree of tumor divergence among and within individuals, as well as divergence in different stages of tumor development. This heterogeneity is not easily addressed by standardized clinical tests resulting in hindrances in the applicability of new personalized approaches in cancer treatment.

1.2.1 Types of Genomic Data

1.2.1.1 Genomics and Transcriptomics

The advent of whole genome sequencing or targeted sequencing of enriched regions enabled us to detect many of the aforementioned genetic alterations. While whole genome DNA sequencing produces a detailed snapshot of the genomic landscape of patient tumors, it is a time-consuming process especially with a large sample size. In contrast, targeted sequencing (e.g. exome sequencing of only the protein coding region of genomic DNA) offers a lower resolution by sequencing enriched regions using pre-determined primers. At the RNA level, microarray analysis offers new insights into the gene expression levels without sequencing the coding regions but is capable of extracting information about both gene expression and copy number aberrations. Microarrays revolutionized the classification of cancer into multiple subtypes with unique expression profiles and clinical behaviors [24]. Recently, the advent of RNA sequencing (RNA-Seq) in 2009 enabled researchers to not only quantify gene expression of non-coding and coding RNA but also to detect single nucleotide polymorphisms, copy number aberrations, post-transcriptional modifications, gene fusions and alternative splicing [25]. RNA-Seq and other next-generation sequencing (NGS) tools encouraged cancer scientists worldwide to examine thousands of tumors from most cancer types from various parts of the world. The need to share data among researchers was confined by geographical boundaries and the ineffectiveness of the physical storage of data. Consequently, multiple initiatives were taken to improve data accessibility. These efforts culminated in the formation of international consortia such as the Cancer Genome Atlas (TCGA), International Cancer

Genome Consortium (ICGC) and the European Molecular Biology Laboratory –European Bioinformatics institute (EMBL-EBI) [26]. TCGA contains genomic profiles of over 11,000 late-stage tumor samples from 33 different cancer types. These genomic profiles include copy number aberrations, somatic mutations, DNA methylation, mRNA expression (both microarray and RNA-Seq), miRNA (microRNA) as well as protein expression. Analysis of these databases allowed researchers to decipher genetic events and signaling pathways that drive malignancy in patients. The genomic profiles and signaling events have helped understand the molecular mechanisms of malignant disease, address molecular and genetic heterogeneity and identify biomarkers for cancer diagnosis and progression, response to treatment and outcome predictors. Table 2 summarizes examples of such resources and databases, the type of analyses provided and the strengths and limitations of each resource (table 2). One limitation of the TCGA cohorts is that tumors are predominately late stage tumors, which minimizes the ability to study early events during cancer development or relapse and importantly undermines findings that may not be applicable to early-stage patients. For that purpose, a new initiative dubbed the Pre-cancer Genome Atlas is ongoing, which encourages genomic profiling of pre-cancerous lesions and the surrounding microenvironment [27].

1.2.1.2 Epigenomics

Chromatin is the macromolecular complex consisting of DNA and histones. It packages DNA into a compact form, sustains mitosis, prevents DNA damage and modulates DNA replication and gene expression. The fundamental functional unit of chromatin is the nucleosome which contains 147 base pairs enfolded by four pairs of the

histones H2A, H2B, H3 and H4. At any given time during the cell's lifetime, chromatin exists in 2 forms. These 2 forms include: 1) heterochromatin that is a highly condensed form containing inactive genes inaccessible to transcription factors and 2) euchromatin, which maintains an open structure accessible to transcription factors. Both components of the nucleosome, DNA and histones are subject to biochemical modifications mediated by chromatin-modifying enzymes in a tightly-orchestrated process [28].

In 1956, in a publication in the journal *Evolution*, Conrad Waddington first used the term epigenetics to describe heritable modifications that affect cellular functions without affecting the DNA genomic sequence. He exposed eggs of *Drosophila Melanogaster* eggs to environmental stimuli in the form of ether. This exposure assimilated the bithorax complex phenotype (doubling of wings, thorax and stomach) in less than 30 generations without changes in DNA [29]. Up until 1982, many research groups had observed that gene silencing was linked to DNA methylation in various tissues [30]. However, the first observation of epigenetic modifications in cancer was reported in 1983 by Andy Feinberg and Bert Vogelstein who demonstrated using southern blotting that CpG dinucleotides in many DNA sequences were hypo-methylated in tumor tissues compared to normal tissues [31]. Later that year, Gama-Sosa *et al*. utilized high performance liquid chromatography to show that the overall amount of 5-methylcytosine (5mC) was reduced in tumors, a phenomenon called "global hypo-methylation" [32]. Subsequent studies revealed that the high frequency of hypo-methylation at CpG sites is seen across many cancer types including cancers of the pancreas [33], colon [34], lung [35] and stomach [36].

Table 2. Online resources for genome-wide analyses of tumor biology and patient outcome. The table summarizes select examples of resources and databases available for cancer researchers, the type analyses that could be performed as well as strengths and limitations of each resource.

Database	Genetic analyses available	Description
International Cancer Genome Consortium (ICGC)	- mRNA - microRNA - Mutations - Gene-copy number Methylation	- Requires prior knowledge in bioinformatics and manipulation of large files.
Cbioportal	- mRNA - microRNA - Mutations - Gene-copy number - Methylation	- Includes analyses of 1000s of tumor samples from various cancer types. - Offers interactive easy-to-use interface.
Oncomine	- mRNA - Gene-copy number	- Offers unique comparisons between tumor and normal tissues. - Allows inclusion of comparison of several tumor features (e.g. drug resistance, recurrence vs. primary, primary vs. metastasis).
Firebrowse	- mRNA	- Offers unique visual comparisons between tumor and normal tissues and across tumors. - Has limited input and does not allow data download and analysis.
MEXPRESS	- Methylation - mRNA	- Allows visualization of overall methylation profiles across multiple patient cohorts.
MethHC	- Methylation - mRNA	- Allows visualization AND analysis of overall methylation profiles across multiple patient cohorts.
Wanderer maplab	- Methylation - mRNA	
OncoLnc	Kaplan Meier survival analysis based on mRNA and miRNA expression	- Requires predetermined expression cut-offs to plot survival curves.
Cancer Cell Line Encyclopedia (CCLE)	- mRNA - microRNA - Mutational - Gene-copy number - Methylation **(Cell lines only)**	Contains build-in resources: - Integrative-genomics viewer (IGV): visualization tool for interactive exploration of large integrated datasets. - Differential expression analysis - Gene co-expression - Gene Set Enrichment Analysis (GSEA): analysis of curated pathways that correlate with gene/s of interest.

Since then, multiple epigenetic alterations have been described including 4 modifications affecting DNA [37] and 16 affecting histones [38][39]. These modifications can not only change chromatin structure by affecting the non-covalent interactions between nucleosomes but also form novel binding sites for proteins that are specific for the modified regions. Examples of DNA modifications include: methylation, hydroxymethylation, formylation (addition of formyl group), and carboxylation (addition of carboxyl group). In contrast, histone modifications are more diverse and consist of: acetylation, methylation of lysine and arginine, phosphorylation of serine/threonine or tyrosine, ubiquitination, ADP ribosylation, sumoylation, deamination, crotonylation, proline isomerization, propionylation, butyrylation, formylation, hydroxylation and O-GlcNAcylation of serine/threonine [37]. DNA methylation and histone acetylation will be discussed next due to their relevance in cancer and relatedness to the dissertation objectives.

1.2.1.2.1 DNA Methylation

The importance of hypo-methylation in cancer is manifested through reactivation of proto-oncogene expression, which would normally be methylated in non-neoplastic tissues [40](figure 2). 5mC is the most studied form of DNA methylation where the carbon at position 5 of the nitrogenous base cytosine is subject to methylation or demethylation (figure 3). Details of the chemical reactions involved in the methylation and demethylation of cytosine are discussed in the figure legend of figure 3. Modifications of methylation can not only affect protein-coding genes but also non-protein coding genes such as microRNAs (miRNA) and long-non-coding RNAs (lncRNA) that play key roles in oncogenesis [37]. 5mCs that are part of CpG dinucleotides aggregate in gene promoters forming CpG islands

directly influencing gene expression (figure 4). Over two thirds of mammalian promoters contain CpG islands highlighting the relevance of DNA methylation in modulating gene expression [37] [41]. Up to 10% of unmethylated CpG islands in gene promoters exhibit hyper-methylation in cancers. CpG shores which are upstream or downstream of CpG islands and are highly conserved sequences have also been implicated in regulating gene transcription. Unlike methylation in CpG islands, CpG shore hyper-methylation is usually linked with increased gene expression indicating that spatial and contextual methylation is to be considered while studying gene regulation [37]. Spatial representations of the CpG islands, shores and shelves are depicted in figure 4 ((figure 4).

DNA methyltransferases are responsible for the addition or removal of methyl groups; 3 of them have been characterized in eukaryotic cells. DNMT3A and DNMT3B are *de novo* methyltransferases that add methyl groups to unmethylated DNA. *DNMT3A* mutations are found in 25% of patients with AML where the mutation affected the catalytic domain responsible for the addition of methyl groups [42]. In contrast, DNMT1 is a maintenance DNA methyltransferase which recognizes hemi-methylated (cytosine methylated on one strand) sequences generated during DNA replication and adds a methyl group to the newly-synthesized cytosine on the opposite strand [43]. DNMT3A and DNMT3B are also involved in the sustenance of the tightly-regulated methylation processes during embryonic development [44]. Methylated DNA generates new docking sites for methyl-binding proteins such as the methyl-CpG binding protein 2 (MeCp2) and the methyl-CpG binding domain proteins MBD1, MBD2, and MBD3. These proteins further recruit histone-modifying proteins (e.g. histone deacetylases), which in turn trigger chromatin remodeling, gene silencing and inaccessibility [45].

Figure 2. Impact of CpG island methylation on gene expression. Unmethylated CpG islands permit binding of transcription factors to the promoter regions to initiate transcription. In contrast, methylated CpG islands hinder transcription factor binding and consequently repress gene expression. CpG islands are often found within the promoter regions upstream of the gene TSS.

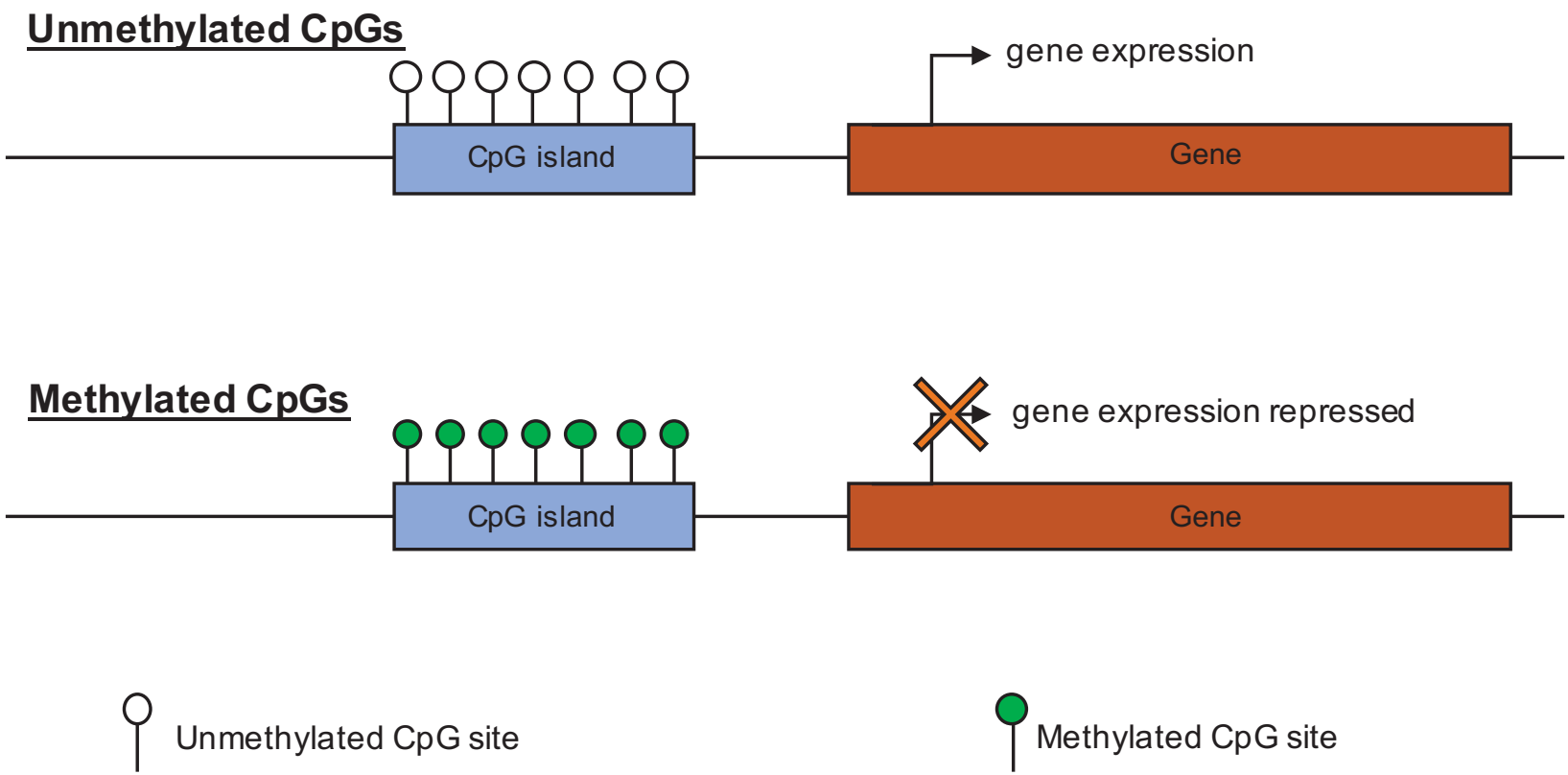

Unmethylated CpGs
gene expression
CpG island
Gene
Methylated CpGs
gene expression repressed
CpG island
Gene
Unmethylated CpG site
Methylated CpG site

Figure 3. Cytosine methylation and demethylation. DNMT3A and DNMT3B are de novo methyltransferases that add methyl groups to unmethylated DNA. In contrast, DNMT1 is a maintenance DNA methyltransferase which maintain methylation of the newly synthesized strand during cell division. The methyl group is "donated" by S-adenosyl methionine which is converted to S-adenosyl homocysteine upon loss of the methyl group. DNA demethylation of 5-methylcytosine (5mC) can occur passively during DNA replication where the newly synthesized strand is not methylated due to reduction in activity or absence of DNMT1. Demethylation of 5mC can also occur via hydroxylation by TET (ten-eleven translocation) enzymes (TET1, 2, 3) to form 5-hydroxymethylcytosine (5hmC) which is in turn further oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). The latter is then converted back to cytosine by the DNA glycosylases (TDG and SMUG1). 5hmC can also be deaminated by the AID and APOBEC family of deaminases to form 5-hydroxymethyluracil (5hmU). The latter is then converted to cytosine by TDG or SMUG1.

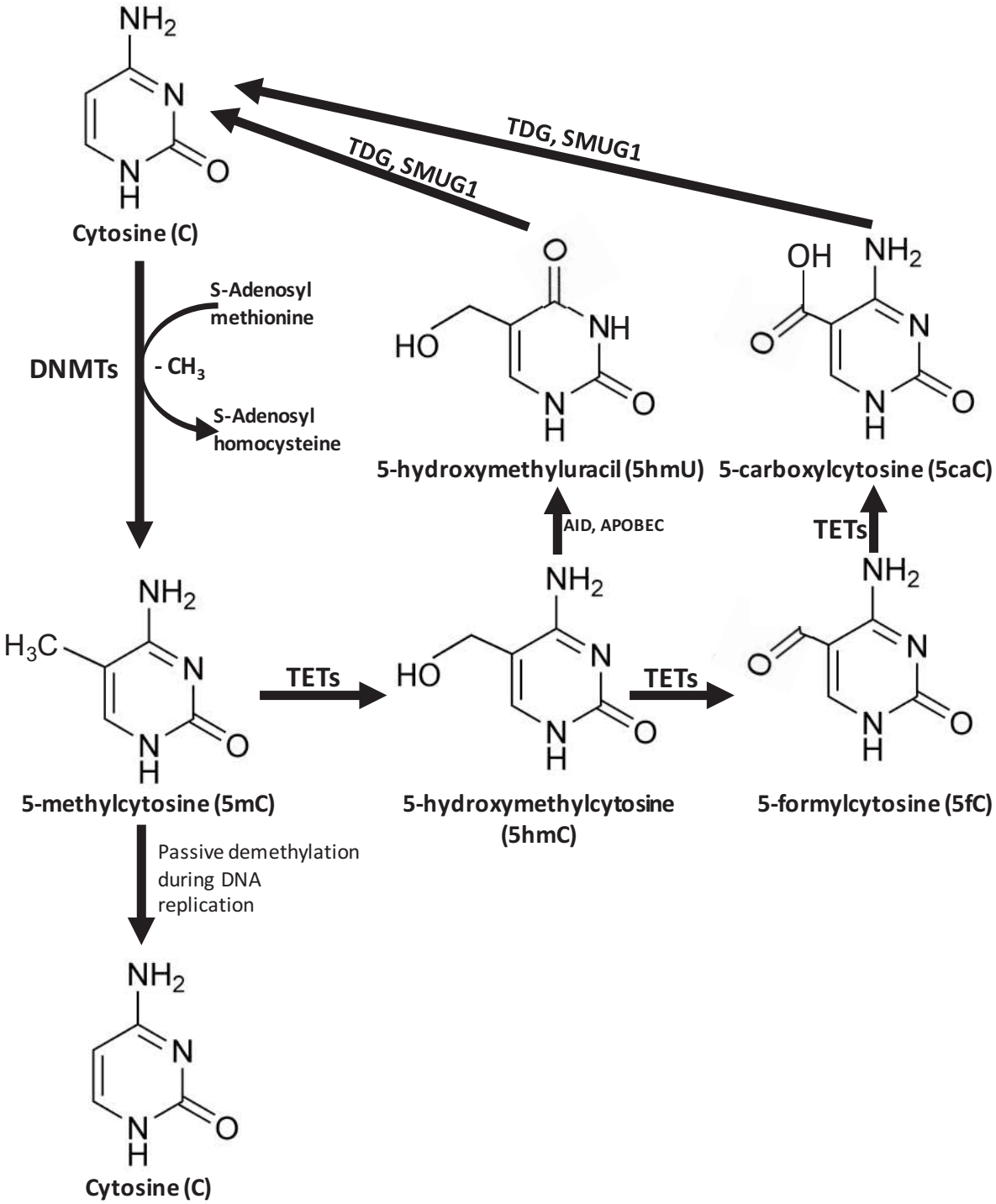
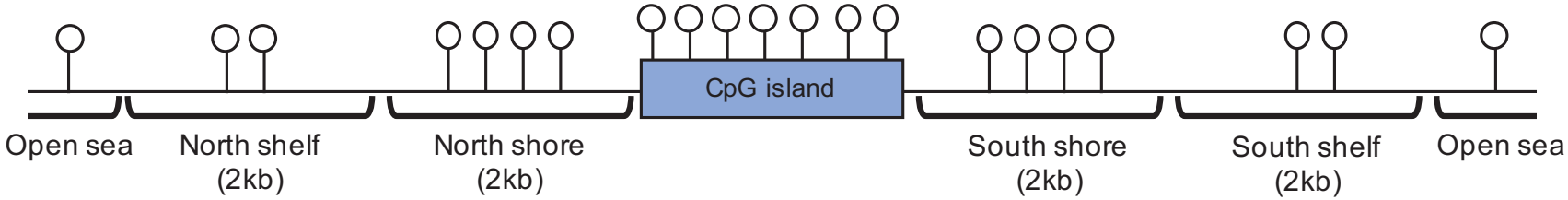

NH2
N
N
H
O
Cytosine (C)
TDG, SMUG1
TDG, SMUG1
S-Adenosyl
methionine
DNMTs
- CH3
S-Adenosyl
homocysteine
O
NH
HO
N
H
O
5-hydroxymethyluracil (5hmU)
OH
NH2
N
O
N
H
O
5-carboxylcytosine (5caC)
AID, APOBEC
TETs
NH2
H3C
N
N
H
O
5-methylcytosine (5mC)
TETs
NH2
HO
N
N
H
O
5-hydroxymethylcytosine
(5hmC)
TETs
NH2
O
N
N
H
O
5-formylcytosine (5fC)
Passive demethylation
during DNA
replication
NH2
N
N
H
O
Cytosine (C)

Figure 4. Map of CpG site distribution. The CpG sites are mapped based on their proximity to the CpG island which is the most CpG-rich site in the genome. Shores and shelves are respectively less rich in CpG sites and are less likely to modulate gene expression. The open sea constitutes all DNA sequences beyond the shelf regions until a shelf of another CpG island is reached.


CpG island
Open sea
North shelf
(2kb)
North shore
(2kb)
South shore
(2kb)
South shelf
(2kb)
Open sea

1.2.1.2.2 Histone Acetylation

In 1964, Allfrey *et al*. first demonstrated that histone acetylation (and to a lesser extent methylation) directly affected gene transcription in the calf thymus *in vitro* [46]. Histone modifications are diverse and can influence many processes including gene transcription, DNA repair, chromatin condensation, and DNA replication [47]. The acetylation of lysine residues on histones is one of the main methods of histone modification. The acetyl group neutralizes the positive charge of lysine, which subsides the electrostatic interactions between histones and the negatively-charged DNA. This leads to the reduction of chromatin condensation and creates an open formation. Histone acetylation takes place around promoter regions but can also occur at upstream enhancer sequences or downstream intragenic sequences [48]. The acetylation creates a binding site for proteins with bromo-domain, which binds acetylated lysines [49]. Acetylation is regulated by two groups of enzymes: histone lysine acetyltransferases (KATs) which add acetyl groups to lysines and histone deacetylases (HDACs) that remove acetyl groups.

Type A KATs (e.g. *KAT3A*) are nuclear enzymes that are responsible for nucleosomal histones while type B are cytoplasmic and they acetylate free histones. The expression of KATs have been reported to be altered in many cancers [50]. Mutations and chromosomal translocations are seen in *KAT3A* in both hematological and solid cancers [51][52]. In contrast, HDACs serve to de-acetylate lysines and restore their positive charge. There are four subclasses of HDACs grouped together based on sequence homologies: 1) Class I (HDAC1-3 and HDAC8), class II (HDAC4-7, HDAC9 and HDAC10), class III (sirtuin proteins; SIRT1 through 7) and class IV (HDAC11) [53]. The catalytic activity of

SIRTs is NAD^+-dependent, which is different from the remaining three subgroups that do not require a cofactor but rely on zinc ions [54]. Oncogenic events such as gene fusions in leukemia (e.g. *PML-RARα,* promyelocytic leukemia retinoid acid receptor alpha) have been shown to preferentially recruit the N-CoR (nuclear receptor co-repressor) deacetylase complex to promote silencing of retinoic acid-responsive tumor suppressor genes [55]. Although mutations in genes encoding HDACs are rare in human malignancies, their expression is altered in many cancers [56]. HDAC inhibitors are also being considered for clinical use as seen with Vorinostat, an FDA-approved product for use in cutaneous T cell lymphoma [57].

1.2.1.3 Other "-omics"

Other types of "-omic" analyses include proteomics and non-coding transcriptomics which are less studied but are becoming increasingly relevant in diseases, particularly cancer. Proteomics uses several variations of mass spectrometry to identify global expression of up to 10,000 proteins. Non-coding transcriptomic analyses assess the global expression of microRNA, long non-coding RNA (lncRNA) and small nucleolar RNA (SnoRNA) and have redefined the pathologic landscape of cancer development and metastasis [58].

1.3 Subchapter 3: Cancer Biomarkers

1.3.1 The "Ideal" Biomarker

Ideally, clinicians require a biomarker that allows them to predict the behavior and outcome of a tumor during its early stages with high specificity and sensitivity. This will enable them to tailor treatment regimens and cautionary measures appropriately. A biomarker test must offer considerably faster turnover time with minimally invasive procedures. For instance, a blood biomarker test will likely increase patient compliance in clinical trials compared to biomarkers tests that require tumor biopsies. It will also facilitate the characterization of targetable causative events.

Several approaches have introduced novel biomarkers that attain some but not all the characteristics of the ideal biomarker. The advent of cancer genomics has helped develop such biomarkers that are more personalized and considerably less invasive. These new cancer biomarkers could replace or complement existing markers. An extensive list of biomarkers has been approved for clinical use at various points of cancer progression and treatment regiments and are presented in table 3. The table includes cancer type, required sample type and the outcome determined by the biomarker measurement (table 3). Below is a summary of the five types of cancer biomarkers, their definitions and a few corresponding examples.

1.3.2 Screening Biomarkers

Cancer screening represents a crucial stage of patient care that can mitigate worsening outcomes by offering early intervention. Screening for most common solid

cancers such as breast, lung and colon is a standardized clinical assessment for at-risk individuals. For instance, Cologuard (Exact Sciences; Madison, WI) is a recently FDA-approved non-invasive screening test for colorectal cancer patients. The test could be performed at three-year intervals which is a step forward compared to the colonoscopy's 2-year interval. Cologuard examines the KRAS mutation status, methylation levels of BMP3 and NDRG4 promoter regions as well as an immunochemical assay for hemoglobin [59]. Another example is the Epi proColon 2.0 test (Epigenomics AG; Berlin, Germany) which is a circulating DNA screening test for hyper-methylated DNA of the Septin 9 gene [60] in colon cancer.

Table 3. A list of current cancer biomarkers. The table contains a conservative list of cancer biomarkers, the required sample type and their clinical use.

Biomarker	Cancer type	Sample type	Outcome determined
ALK fusion or overexpression	Non-small cell lung cancer Anaplastic large cell lymphoma	Primary tumor	Response to therapy Cancer progression
Alpla-fetoprotein (AFP)	Hepatocellular carcinoma Germ cell tumors	Blood	Diagnosis Response to treatment Cancer progression staging
Beta2 microglobulin (B2M)	Chronic lymphocytic leukemia, multiple myeloma, some lymphoma	Blood, urine or CSF (cerebrospinal fluid)	Cancer progression Response to treatment
Beta-human chorionic gonadotropin (BhCG)	Choriocarcinoma and germ cell tumors	Blood, urine	Response to therapy Cancer progression staging
***BRCA1/2* mutations**	Ovarian cancer and breast cancer	blood	Response to targeted treatment
***BCR-ABL* fusion (Philadelphia chromosome)**	Acute myelogenous leukemia, acute lymphoblastic leukemia, chronic myeloid leukemia	Blood, bone marrow aspirate	
***BRAF V600E* mutation**	Melanoma, colorectal cancer	tumor	Response to targeted treatment
C-kit/CD117	Mucosal melanoma, gastrointestinal stromal tumors	tumor	Diagnosis Response to treatment
CA15-3/CA27.29	Breast cancer	blood	Response to treatment Recurrence
CA-19-9	Pancreatic cancer, bile duct cancer, gastric cancer, gallbladder cancer	blood	Response to treatment
CA-125	Ovarian cancer	blood	Diagnosis Response to treatment recurrence
Calcitonin	Medullary thyroid tumors	blood	Diagnosis, response to treatment, recurrence
Carcinoembryonic antigen (CEA)	Colorectal cancer	blood	Response to treatment, recurrence
CD20	Non-Hodgkin's lymphoma	blood	Response to targeted treatment

Table 3 (continued)			
Biomarker	**Cancer type**	**Sample type**	**Outcome determined**
Chromogranin (CgA)	Neuroendocrine tumors	blood	Diagnosis, response to treatment, recurrence
Polysomy of chromosome 3, 7 and 17 and deletion of 9p21	Urothelial carcinoma	urine	Diagnosis, recurrence
Cytokeratin fragment 21-1	Lung cancer	blood	Recurrence
***EGFR* amplification/mutation**	Non-small cell lung cancer	tumor	Diagnosis, response to treatment
Estrogen and progesterone receptors (ER/PR)	Breast cancer	tumor	Response to hormone therapy
HER2/Neu amplification or overexpression	Breast cancer, gastroesophageal adenocarcinoma, gastric cancer	tumor	Response to targeted treatment
Fibrinogen/fibrin	Bladder cancer	urine	Cancer progression, response to treatment
Human Epididymis protein 4 (HE4)	Ovarian cancer	blood	Response to treatment, cancer progression, recurrence
***KRAS* mutations**	Colorectal cancer, non-small cell lung cancer	tumor	Response to targeted therapy
Lactate dehydrogenase	Lymphomas, leukemia, melanoma, germ cell tumors, neuroblastoma	blood	Cancer progression, response to treatment, staging
Neuron-specific enolase (NSE)	Small cell lung cancer, neuroblastoma	blood	Diagnosis, response to treatment
Nuclear matrix protein 22	Bladder cancer	Urine	Response to treatment
Programmed death ligand 1 (PDL1)	Non-small cell lung cancer	tumor	Response to targeted treatment
Prostate-specific antigen (PSA)	Prostate cancer	blood	Diagnosis, response to treatment, recurrence
Thyroglobulin	Thyroid cancer	blood	Response to treatment, recurrence
Urokinase plasminogen activator (uPA) and plasminogen activator inhibitor 1 (PAI-1)	Breast cancer	tumor	Response to treatment Cancer progression
Mammaprint® (70-gene signature)	Breast cancer	tumor	recurrence
OncotypeDX® (21-gene signature)	Breast cancer	tumor	recurrence
OVA1® (5-protein signature)	Ovarian cancer	Blood	diagnosis
CELLSEARCH® (Circulating tumor cells)	Metastatic colorectal, breast and prostate cancers	blood	progression

1.3.3 Diagnostic Biomarkers

Cancer diagnosis typically requires sample biopsies for clinicians to assess disease pathology and complete diagnosis. For instance, serial measurements of the serum biomarker CA-125 (cancer antigen 125) are routinely used to diagnose patients with ovarian cancer [61]. A measurement of 30-35 U/ml is considered as the threshold, which when collected through serial measurements, creates the Risk of Ovarian Cancer Algorithm (ROCA) to predict the risk or likelihood of having an ovarian tumor. ROCA stratifies patients into low-, intermediate- and high-risk groups based on their CA-125 scores [62] [63]. The risk of malignancy index (RMI) combines serum CA-125 levels, menopausal status and ultrasound to determine if the elevated CA-125 levels represent a benign pelvic mass or an ovarian carcinoma [64].

1.3.4 Progression Biomarkers

These markers are also called prognostic markers and are designed to indicate how aggressive a tumor is and its likelihood of progression. For instance, serum AFP (alpha-fetoprotein) is used to predict the outcome of patients with hepatocellular carcinoma (HCC). Higher AFP levels correlate with increased tumor size and volume at diagnosis. HCC patients with AFP greater than 400 ng/ml have a higher chance of bi-lobe involvement and portal vein thrombosis, which leads to worse outcome compared to those with AFP < 400ng/ml. Survival is poorer when AFP levels exceed 1000 ng/ml. Some exceptions have been reported where patients with AFP > 1000 ng/ml had a significantly better prognosis than what was predicted based on AFP levels [65]. In untreated HCC patients, AFP levels increase over time in tandem with a progressing tumor. Interestingly,

patients that have normal AFP levels at diagnosis will maintain below threshold levels regardless of tumor progression [66] delineating inconsistencies in the predictive power of AFP.

1.3.5 Response to Therapy Biomarkers

These biomarkers are used to monitor patients who are being treated for cancer. In general, a marker that is known to be at higher levels prior to treatment and is considerably lower after treatment indicates that the therapy is effective. The lack of change or even increase in marker levels indicates that the cancer is not responding. For instance, a 20% or more decrease in AFP serum levels in HCC patients was indicative of a response to tyrosine kinase inhibitor sorafenib and correlated with better survival [67]. A >20% decrease in serum levels of CEA (carcinoembryonic antigen) strongly correlated with a favorable response to radiation therapy in colorectal cancer patients with liver metastases [68]. In addition, response to therapy can be based on a binary identifier of a sensitizing marker. A classic example of markers to predict response to targeted therapy is the *BRAFV600E* mutation which is present in 40-60% of melanoma patients. *BRAF*-positive melanoma patients are sensitive to first generation RAF kinase inhibitor sorafenib [69] and the highly-specific second-generation inhibitors PLX4032 [70] and GSK2118436 [71] that target mutant BRAF only.

1.3.6 Recurrence Biomarkers

Recurrence biomarkers are also known as relapse markers. They are utilized as tools to detect if cancer recurs (i.e. returns) after surgical resection or therapeutic

intervention. Examples include CA-125 in ovarian cancer, PSA (prostate-specific antigen) in prostate cancer and HCG (human chorionic gonadotropin) in islet cell tumors, choriocarcinomas, germ cell tumors and others. In prostate cancer, a PSA level of 0.2 ng/ml or higher on two consecutive tests is considered (when combined with other clinical features) as an indicator of recurrence in patients that underwent prostatectomy. The other clinical features that are considered include pre-operative PSA levels, Gleason score, tumor stage, age and percentage of PSA positive biopsies, together they generate a Prostate Risk Assessment score [72].

1.3.7 Limitations and Precautions for Biomarker Studies

1.3.7.1 Clinical Limitations

First, the ability of a biomarker to offer concrete predictive evidence of cancer is always challenged by inter-patient heterogeneity as well as variability within individual samples (intra-patient heterogeneity). Second, a major hurdle is that most markers are expressed at high levels in late-stage cancers but not in early-stage cancers rendering early intervention a difficult task [73][74]. Conversely, when a marker is expressed in early stage patients, the error rate is much higher. For example, CA-125 is markedly less sensitive (60%) in early stage patients, which increases false positives and negatives [61], [75]. Third, some patients may express normal levels of a biomarker despite a progressing cancer. For instance, 20% of prostate cancer patients express normal levels of PSA (< 4ng/ml). Fourth, reliability of a cancer biomarker is also challenged by factors such as expression of most tumor-associated markers in non-neoplastic cells. Fifth, confounding conditions other than cancer can also increase levels of a biomarker as seen in pancreatic

cancer patients where the serum levels of CA-19-9 are often affected by cholestasis [76] and jaundice [77], which are common complication of this cancer type.

1.3.7.2 Logistical Limitations

In solid tumor biopsies, sample collection is often invasive and introduces inflammation-induced changes in tumors. Biopsy-based diagnosis can also experience delays primarily due to long wait times before biopsy appointments. Other technical challenges arise from sample collection, processing, storage, measurement methodology and center-to-center variability. In addition, screening biomarkers also require population-wide screening to be able to detect a small percentage of high-risk individuals. This will increase the work and economic burden on the healthcare system. The usage of screening tools with these shortcomings will cause diagnostic hesitation on part of the clinicians as well as frustration and anxiety on part of the patient [78].

1.4 Subchapter 4: Cancer Cell Invasion and Migration

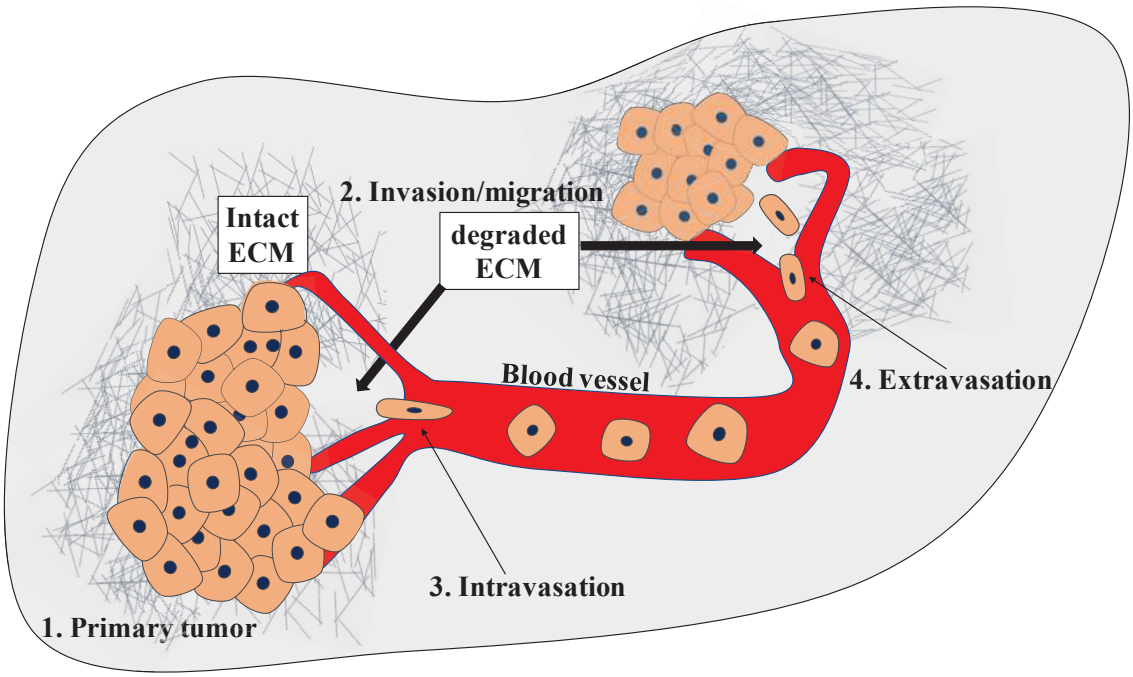
The identification of robust cancer biomarkers that will withstand rigorous clinical testing requires that these biomarkers are functionally and mechanistically linked to cellular changes pertaining to a tumor phenotype. Proteins involved in cancer development, migration, invasion, metastasis, angiogenesis, and DNA repair are often targets for biomarker discovery and testing. A plethora of studies have focused on identifying the underlying mechanisms of cell migration and invasion in neoplastic and non-neoplastic cells as means of biomarker discovery [79]. Earlier studies established that dynamic changes in the cytoskeletal structure and modulation of cellular adhesion are crucial steps for successful migration and invasion [80] (figure 5). Later efforts to pinpoint a specific pathway that initiated the migratory and invasive program proved to be an arduous task. In fact, multiple processes have been implicated in driving migration and invasion that revolve around protease-dependent and independent interactions with the extracellular matrix (ECM). The elements that drive the above processes belong to an ever-expanding list of proteins, non-protein coding genes and signaling pathways, making them eligible for biomarker testing. The lack of a universal mechanism provided insights into the complexity of cancer cell migration and invasion indicating a high degree of plasticity [81].

1.4.1 Migration versus Invasion

Although the terms migration and invasion are used interchangeably in the literature partly because they occur in tandem, there are important distinctions that ought to be made. Migration is the mere physical movement of cells within confinements of an ECM and often requires dynamic cytoskeletal rearrangements. Cytoskeletal

rearrangements are largely mediated by the activation of Rho GTPases (Guanosine triphosphatases), which regulate actin polymerization and depolymerization, myosin activity, integrin interactions with the ECM and reorganization of microtubules and intermediate filaments [82]. These dynamic changes generate a highly motile and agile cell. In contrast, invasion is a process by which cells activate a protease-dependent program promoting ECM degradation and remodeling. Invading cells often express markers of mesenchymal cells, lack apical-basal polarity and undergo dynamic changes in the cytoskeleton and at cell junctions [83].

Figure 5. The metastatic cascade. The initial step occurs at the primary site where a normal cell (or cells) undergoes genetic changes that prompt its neoplastic transformation. Transformed cells must proliferate and establish a primary tumor. The second step involves a few cancer cells acquiring migratory and invasive properties that enable them to degrade the underlying basement membrane and extracellular matrix (ECM). The third step necessitates cells to "squeeze" through the endothelial lining of blood vessels in a process known as intravasation. Once in circulation, most cancer cells fail to survive except a few that adhere to blood vessels adjacent to prospective metastatic site. Cells then undergo the fourth step of exiting the vasculature and invading the new site in a process known as extravasation. Finally, cells that successfully extravasated must then colonize and proliferate within a supportive microenvironment to give rise to micro- and macro-metastases.


2. Invasion/migration
Intact ECM
degraded ECM
Blood vessel
4. Extravasation
3. Intravasation
1. Primary tumor

1.4.2 Cellular Junctions

Cellular junctions are key membrane-associated structures that are subject to drastic changes which in turn dictate the fate of a neoplastic cell. They are multi-protein complexes that sustain contact and communication between neighboring cells and between cells and the ECM. A prerequisite to invasion is an acquired ability to alter expression of cell adhesion proteins that promote the disintegration of cellular junctions. Invading cells must first disengage cellular junctions and then detach from neighboring cells and the underlying basement membrane for successful invasion.

1.4.2.1 Tight Junctions

Tight junctions (TJ) are the first barrier that cancer cells need to overcome. They are located at the most apical position of the intercellular membrane space and serve as a cellular barrier and a site of cell attachment. Cancer cells and endothelial cells induce or repress proteins involved in TJs [84]. Early studies have demonstrated that less differentiated cancers, which are typically more aggressive than well-differentiated cancers, are associated with lower expression of TJ proteins [85]. For example, factors such as HGF (hepatocyte growth factor) reduced trans-epithelial resistance and enhanced para-cellular permeability in breast cancer cell lines by decreasing the expression of TJ proteins such as ZO-1, ZO-2, occludin, claudin-1 and claudin-7 [86]. Downregulation of occludin is also associated with a higher chance of metastatic disease in breast cancer patients due to the loss of TJ integrity [86]. ZO-1 downregulation is linked to poor differentiation and higher grade and TNM (tumor, node, metastasis) staging [87].

1.4.2.2 Adherens Junctions

Another important type of junctions is the adherens junctions (AJ) which are located below the TJs in the intercellular space. The cadherin family of proteins are abundant and essential in providing structure for AJs. E-cadherin is the most abundant AJ protein in epithelial cells while VE-cadherin is characteristic of endothelial cell AJ. AJs form the zonula adherens (or adhesion belt), which surrounds cells along with the intracellular actin belt. Other AJ proteins include armadillo proteins and plakins [88]. This architecture provides structural support for epithelial cells while maintaining a fluidic environment due to its association with actin filaments [89]. Dysregulation of AJ architecture results in major implications in cellular transformation and cancer cell invasion [90](discussed later). Of note, armadillo proteins are characterized by armadillo repeat/s which is a repetitive amino acid sequence containing 40 residues [91]. These amino acids form two alpha helices in the shape of a hairpin. Tandem repeats of armadillos are ubiquitous which in turn results in an alpha solenoid structure [92]. Examples of armadillo proteins are β-catenin, plakoglobin, α-importin and others [88].

1.4.2.3 Desmosomes

Desmosomes are fundamental for tissue integrity, cell-to-cell communication and establishment of an intercellular adhesive framework between the cytoskeleton and plasma membrane. The framework involves anchoring the intermediate filaments in the cytoskeleton to the cytoplasmic and extracellular parts of the desmosomes via a series of protein complexes [93]. These proteins include cadherins, plakins and catenins. Two types of cadherins that are unique to desmosomes are represented by the desmogleins (DSG1 to

4) and desmocollins (DSC1 to 3), which serve as anchors for the keratin intermediate filaments in nearby cells [94]. It has been reported that alterations in desmosomes proteins precede those of AJs to allow the early onset of invasion [95].

1.4.2.4 Gap Junctions

Gap junctions (GJ) act as cell-to-cell channels for the diffusion of ions, metabolites and second messengers. Connexin proteins are present in GJ with connexin 43 as the most abundant [96]. Connexins assemble into heteromeric hemi-channels (called connexon) which then interact with connexons on adjacent cells to form the complete intercellular gap junction. The permeability characteristics of each GJ are dictated by the type of connexins involved [97]. High expression of connexin 43 is linked to better patient prognosis and vice versa in various cancers including pancreatic [98], prostate [99], colorectal [100], breast [101] and non-small cell lung [102] cancers.

1.4.3 Mechanisms of Cell Migration

Cancer cells can migrate either individually or collectively. Individual cell migration results from significant loss of cell-cell adhesion while collective cell migration involves the retention of some, but not all of the cell-adhesion capacity, manifested as multi-cellular bodies [81]. Figure 6 illustrates the different types of individual and multicellular migration and the subcategories of each (figure 6) (discussed next).

1.4.3.1 Individual Cell Migration

During individual migration, cells will initially induce actin polymerization to form pseudopod protrusions at the leading edge [103]. "Leading" cells will then interact with ECM substrates and localize cell adhesion molecules and cell surface receptors, together activating a forward motion, referred to as traction force [104]. The small GTPases Rac and Cdc42 mediate the formation of these protrusions, which interact with the ECM [79]. A few micrometers behind the leading edge, the cell surface becomes engaged in active proteolysis (discussed next) which remodels the surrounding ECM and allows cellular advancement [81]. To mediate forward movement, Rho GTPase activates myosin II that initiates contraction by actomyosin (complex of actin and myosin). Finally, the cell will disengage adhesion molecule interaction at the trailing end forming micro-tracks (10-15 μm). If multiple cells are migrating in an individual manner, the "leading cell" will form the initial micro-track where the ECM has been proteolytically cleaved. The following cells will further widen the micro-track by shear mechanical force and proteolytic cleavage [105] (figure 6).

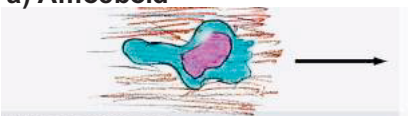
Figure 6. Models of cancer migration and invasion. Cancer cell migration can occur as an individual cell or collectively (multi-cellular) based on expression of specific cell-cell junction proteins and the contractility of the cytoskeleton. Individual cell migration is further subdivided into single cell migration which lack cell-cell adhesion molecules or multi-cellular streaming which retains some cell adhesion. The cytoskeletal contractility dictates whether individual cell migration will involve amoeboid or mesenchymal cell movement. In contrast, multicellular migration can be collective or expansive. The figure represents whether each migration mechanism requires surface proteolytic activity. Adapted from [106].

Individual cell migration

Proteolytic activity required

1) Single-cell migration

a) Amoeboid



NO

b) Mesenchymal

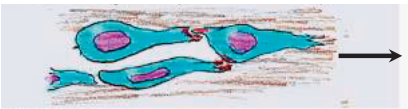
YES

2) Multicellular streaming (minimal cell-cell adherions)

a) Amoeboid

NO

b) Mesenchymal



YES

Multicellular migration

Proteolytic activity required

1) Collective cell migration

a) Cluster

YES

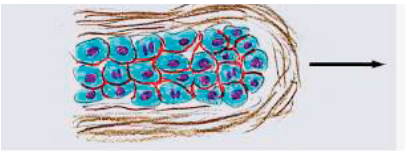
b) Solid strand

YES

c) Strand with lumen

YES

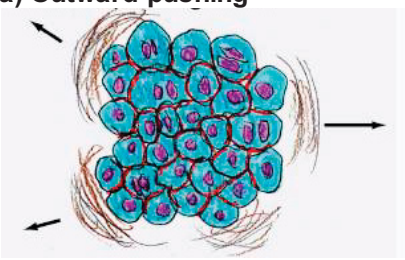
d) Strand with protrusion



YES

2) Expansive growth

a) Outward pushing



NO

There are two types of single cell migrations: **amoeboid-like migration** and **mesenchymal cell migration**. Amoeboid-like movement is accomplished by a round-like cell which either has 1) short thin protrusions with no membrane blebbing hence utilizing a higher migratory velocity (0.4-5µm/min), 2) a bleb-rich membrane, which causes disoriented movement at lower velocities or 3) small membrane protrusions with high surface protease activity and slow velocities (0.1µm/min) [83]. In contrast, mesenchymal cell movement involves an elongated fibroblast-like, spindle morphology and considerably large membrane protrusions. These protrusions are called "invadopodia" [107]. Invadopodia are cancer-specific protrusions that were initially observed on the baso-lateral side of cancer cells cultured *in vitro* [108] (figure 6).

1.4.3.2 Multicellular Migration or "Streaming"

Streaming is achieved by groups of cells that are loosely attached to each other and that migrate together on the same and often straight path at moderate velocity (1-2µm/min). Both amoeboid and mesenchymal movements can be displayed by these cells [106]. Streaming typically takes place in response to chemokine signals within the surrounding tissue like that seen in neural crest devolvement during embryogenesis [109]. Roussos *et al*. described a chain-like movement of mammary neoplastic cells displaying multicellular migration [110] (figure 6).

1.4.3.3 Collective Invasion

Collective invasion mandates strong cell-to-cell adhesion and concomitant activation of a migratory phenotype. This form of invasion typically involves cells forming

small strands or clusters at the interface between the leading edge of a tumor and the surrounding stroma. The maintenance of cellular adhesion indicates that some cells retain their epithelial polarity which in some cases allow the formation of gland-like structures resembling the tissue of origin. However, the leading edge of the collectively invading tumor cluster/strand will ultimately become mesenchymal to produce a Rho-mediated forward traction force and activate surface proteolysis [111] (figure 6).

1.4.3.4 Expansive Growth

Expansive growth takes place where the surrounding tissue does not exert any physical containment of the growing tumor mass. This leads to multicellular clusters of cells with virtually unaltered cell adhesion to exert a forward push in the absence of active migration or ECM proteolysis [112]. The cellular cluster will then form a capsule-like structure surrounded by collagen fibers [113]. Expansive growth typically does not require proteolytic activity for successful migration (figure 6). However, Ilina *et al.* and Weigelin *et al.* demonstrated that expansive growth can be coupled with active migration, which in turn exacerbates invasion, particularly collective invasion [114] [115].

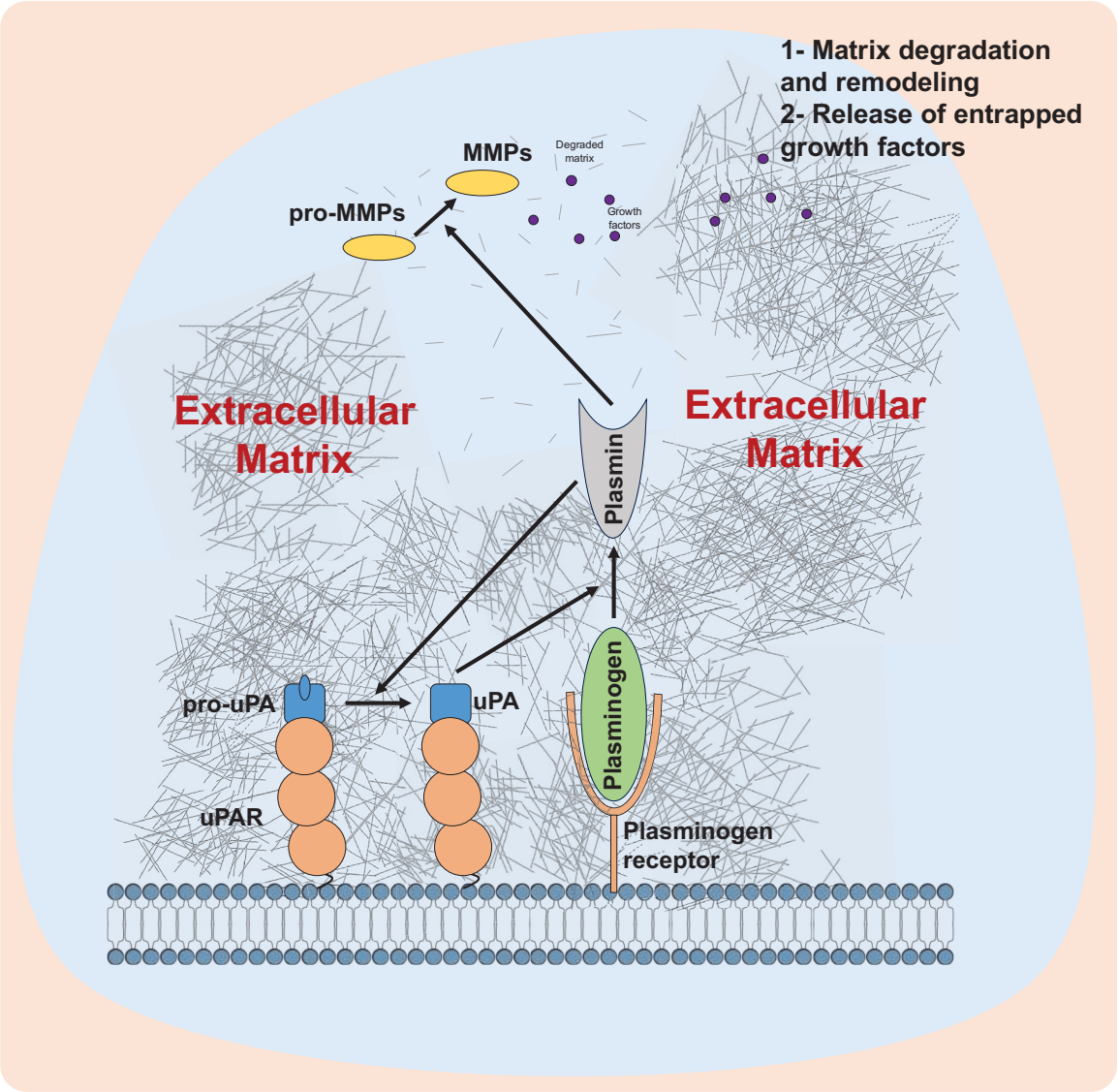
1.4.4 ECM Proteolysis: The Act of Invasion

Cell surface proteolysis is an essential part of cellular migration through which a series of active proteases are produced to degrade and remodel the ECM. These proteases are mostly serine (e.g. plasmin), cysteine (e.g. cathepsins), aspartic (e.g. cathepsin D) and metalloproteases (e.g. matrix metalloproteinases or MMPs) and act on a range of overlapping protein substrates (figure 8). Proteases are also commonly upregulated during

neoplastic transformation [116]. These proteases mediate ECM substrate breakdown through three major mechanisms of action. First, MT-MMPs (membrane-tethered matrix metalloproteinases) and ADAMs (a disintegrin and metalloproteinase) mediate contact-dependent pericellular proteolysis of ECM substrates such as collagen, fibronectin, laminin and others [105] [117]. Second, the cleavage of ECM proteins by MMPs (e.g. MMP2) remodel the ECM by creating migration-promoting gaps [117]. MMP2 also cleaves fibronectin and vitronectin which exposes new protein fragments that promote peritoneal adhesion as seen in ovarian cancer cells [118]. Third, ADAMs and MT-MMPs are also capable of activating growth factor receptors and adhesion surface receptors (e.g. integrins) via a cleavage-dependent event.

Plasmin and MMPs also drive ECM degradation as well as cleavage-mediated activation of sequestered growth factors such as TGFβ1 (transforming growth factor β 1), VEGF (vascular endothelial growth factor) and IGF-1 (insulin-like growth factor 1) within the matrix [119] (figure 7). Metastasizing tumors cells associate with endothelial cells through weak surface carbohydrate interactions followed by stronger adhesion-molecule-mediated bonds. This interaction will allow the already enhanced protease activity to penetrate the endothelial layer and the basement membrane by proteolytic cleavage, leading to extravasation [119]. The proteolytic network of protease interactions within themselves is very extensive and multidirectional as depicted in figure 8 (figure 8).

Figure 7. ECM remodeling and proteolysis during cancer cell invasion. Through its C-terminal lysine (or internal lysine), a cell surface plasminogen receptor binds plasminogen which induces an activation-susceptible conformation in plasminogen by the urokinase plasminogen activator uPA. The urokinase-plasminogen activator (uPA) is bound to its receptor (u-PAR) and forms the uPA/u-PAR complex that co-localizes with the plasminogen-plasminogen receptor complex. This colocalization results in accelerated cleavage of plasminogen into plasmin. Plasmin in turn activates pro-uPA into uPA forming a positive feedback loop. Plasmin is a multifunctional serine protease, that 1) cleaves extracellular matrix components, 2) releases trapped growth factors within the matrix, 3) activates other proteases such as pro-MMPs (matrix metalloproteinases) into active MMPs. Active plasmin and MMPs degrade impeding obstacles in the ECM, and mediate tumor cell invasion.


1- Matrix degradation
and remodeling
2- Release of entrapped
growth factors
MMPs
Degraded
matrix
pro-MMPs
Growth
factors
Extracellular
Matrix
Extracellular
Matrix
Plasmin
Plasminogen
pro-uPA
uPA
uPAR
Plasminogen
receptor

1.4.5 Metastasis

Metastasis is a series of events that are characterized by the spread of cancer cells from a primary site of growth via the hematogenous or lymphatic route to tissues and organs where they form secondary and tertiary foci of micro- and macro-metastasis [120][5] (figure 5). Once a cancer cell successfully colonizes a secondary site, patient prognosis is markedly reduced. Considering metastasis is largely responsible for the morbidity and mortality of cancer patients, it is not surprising that significant research efforts have addressed the molecular underpinnings of metastasis [121][122]. Although macro-metastasis is detectable by conventional detection methods, patients can also develop micro-metastasis or possess dormant tumors, which cannot be easily detected using standard imaging techniques. Micro-metastasis and dormancy are often responsible for cancer relapse post-surgery or chemotherapy [123].

The sequence of events that give rise to metastasis is known as the metastatic cascade and is divided into three broad steps are shown in figure 5 and described below (figure 5):

(1) Invasion: Invasion is initiated by the loss of cell-cell adhesion, which enables cancer cells to dissociate from the primary tumor and trigger protease-mediated alterations in cells' interactions with the ECM. This allows cancer cells to invade the surrounding stroma until a lymphatic or hematologic vessel is encountered and intravasation can occur. Importantly, invasion relies on the production of proteases to degrade the underlying basement membrane and ECM (discussed above), the activation of motility/migration proteins and suppression of detachment-induced apoptosis [124]. The succeeding event is the initiation of intravasation.

(2) Intravasation: Intravasation is characterized by the physical penetration of cancer cells through the endothelial cell barriers into the blood or lymph circulation. Vessel density and diameter drastically affect the efficiency of intravasation. The number of circulating tumor cells is often used as quantifiable metric for the effectiveness of intravasation [125]. Furthermore, intravasation is facilitated by potent pro-angiogenesis signals produced by tumor cells such as VEGF that promotes blood vessel formation and endothelial cell expansion [126][127]. Interaction between tumor cells and the matrix is also crucial for initiation of angiogenesis [128]. The rapid formation of an extensive blood supply network allows tumors to grow beyond the 2-mm threshold and sustain further growth. The 2-mm threshold is the maximum dimension that still permits local diffusion of nutrients and waste into and out of the tumor core [129].

(3) Extravasation: The final step of metastasis is extravasation which depends on the ability of the cancer cells to successfully exit the circulation and extravasate into the surrounding tissue. This process involves adhesion and interaction of tumor cells with the endothelial lining (e.g. MCAM, melanoma cell adhesion molecule on endothelial cells) followed by trans-endothelial migration of individual cells to reach the prospective metastatic site [130]. Additionally, cancer cells can also become arrested in small capillaries [131] at which point they proliferate and then extravasate [132][133].

Many genetic alterations and cellular phenomena have been linked to promoting or inhibiting different steps of the metastatic cascade. Two specific processes will be discussed next which constitute the **plasminogen activation system** and **epithelial to mesenchymal transition**. Although both processes have been extensively studied, little is known about the interactions and modes of regulation between the two.

1.5 Subchapter 5: The Plasminogen Activation System

As eluded to earlier, tumor cell invasion and metastasis involve a cascade of interdependent events, which require proteases to degrade the basement membrane and render it conducive for the invasive escape of tumor cells. The serine protease plasmin plays a key role in orchestrating an invasive program that promotes the breakdown of ECM and allows cells to leave the primary tumor and metastasize. The proteolytic network, to which plasminogen/plasmin belongs to is a complex and intricate network that is tightly controlled and affected by a series of proteases which will ultimately execute the act of invasion by degrading the matrix. These proteins include plasminogen (*PLG*), plasminogen activators (*PLAU, PLAT*), plasminogen activator receptors (*PLAUR*), plasminogen activation inhibitors (e.g. *SERPINE1, SERPINB2, and SERPINF2*), MMPs (e.g. *MMP1*, *MMP2*, *MMP9*), ADAMs (e.g. *ADAM 1*, *ADAM2*), kallikreins (KLKs), cathepsins (e.g. *CTSB,* and *CTSL*), tissue inhibitors of metalloproteinases (TIMPs) and plasminogen receptors (e.g. *ENO1, S100A10, RUVBL1, HIST2H2BE* and *PLGRKT*) ((table 4 and table 5). Figure 8 depicts the interactions among the four known types of proteases and corresponding inhibitors (figure 8). Below are detailed descriptions of the above components.

1.5.1 Plasminogen

1.5.1.1 Activation Site and Catalytic Activity

Plasminogen is a circulating (1.6 μM) zygmogen that is produced in the liver with a half-life of two days. Human plasminogen contains 791 amino acids, which creates 24

disulphide bonds 16 of which help form five homologous loops called kringles [134]. Law *et al* characterized the crystal structure of precursor plasminogen which circulates in a closed activation-resistant form. This closed conformation is maintained by the Pan-apple (PAp) domain, serine protease domain at the carboxyl terminus and availability of chloride ions, all of which interact with the kringle domains thus preventing their cellular interactions [135].

Plasminogen activation into plasmin is the result of a single proteolytic cleavage of the peptide bond between Arg561 and Val562 by the plasminogen activators uPA and tPA. The amino terminus of plasminogen which contains the PAp domain and the five kringle domains, mediates its interactions with other regulatory proteins. Kringle 1 (K1) and kringle 4 (K4) can bind lysines on fibrin, plasminogen receptors, α2-antiplasmin and other ECM proteins [136][137] with high (K1) and low (K4) affinities [138]. In contrast, the carboxyl terminus contains the active protease site of plasmin [139].

1.5.1.2 Glu- to Lys-plasminogen Conversion

Plasminogen circulates in the Glu-plasminogen (glutamic acid at amino-terminus) form and is cleaved by plasmin at the carboxyl end of Lys62, Arg68 and Lys77 [140][141] and at some basic residues in the hinge region of plasminogen [142]. These cleavage events generate new amino termini on plasminogen and is hence termed Lys-plasminogen. Lys-plasminogen does not normally circulate in plasma and is usually found on cellular surfaces [143] where is more readily activated by plasminogen activators [144][145]. Gong *et al.* demonstrated that the conversion of Glu-plasminogen to Lys-plasminogen is necessary for

the maximum activation of plasminogen by tPA and uPA at the cell surface of endothelial cells [146].

1.5.1.3 Glycosylation

Post translational modification of plasminogen results in two glycosylated forms (form 1 and form 2), which not only dictates plasminogen's binding specificity to receptors and binding partners but also its degradation mechanism. Edelberg demonstrated that human prenatal plasminogen was more heavily glycosylated (form 1-like) than adult plasminogen and as a result was less able to be activated by the tissue plasminogen activator tPA [147]. The less glycosylated form 2 is one degree of magnitude better at being activated by tPA than form 1 [148]. More specifically, N-glycosylation of K3 decreases the stability of the plasminogen-plasminogen activator complex, which hinders its activation and disrupts its interaction with fibrin [149].

Table 4. Components of the plasminogen activation system. Some components have been directly linked to plasminogen activation based on literature review while others are members of the same family. Plasminogen (PLG) is not shown in the table. Plasminogen receptors are summarized in table 5.

Plasminogen activators	
PLAU	PLAT

Plasminogen activator receptor
PLAUR

Cathepsins			
CTSA	CTSE	CTSK	**CTSV**
CTSB	CTSF	**CTSL**	CTSW
CTSC	CTSG	CTSO	CTSZ
CTSD	CTSH	CTSS	

Plasminogen activation inhibitors						
SERPINA1	SERPINA4	SERPINB1	SERPINB2	SERPINB6	SERPIND1	SERPING1
SERPINA10	SERPINA5	SERPINB10	SERPINB3	SERPINB7	SERPINE1	SERPINH1
SERPINA12	SERPINA6	SERPINB11	SERPINB3/B4	SERPINB8	SERPINE2	SERPINI1
SERPINA2	SERPINA7	SERPINB12	SERPINB4	SERPINB9	SERPINF1	SERPINI2
SERPINA3	SERPINA9	SERPINB13	SERPINB5	SERPINC1	SERPINF2	

Matrix metalloproteinases (MMPs)				
MMP1	MMP9	MMP14	MMP20	MMP25
MMP2	MMP10	MMP15	MMP21	MMP26
MMP3	MMP11	MMP16	MMP23A/B	MMP27
MMP7	MMP12	MMP17	MMP24	MMP28
MMP8	MMP13	MMP19	MMP24-AS1	

A disintegrin and metalloproteinase (ADAMs)			
ADAM1	ADAM10	ADAM18	ADAM23
ADAM2	ADAM11	ADAM19	ADAM28
ADAM7	ADAM12	ADAM20	ADAM29
ADAM8	ADAM15	ADAM21	ADAM30
ADAM9	ADAM17	ADAM22	ADAM33

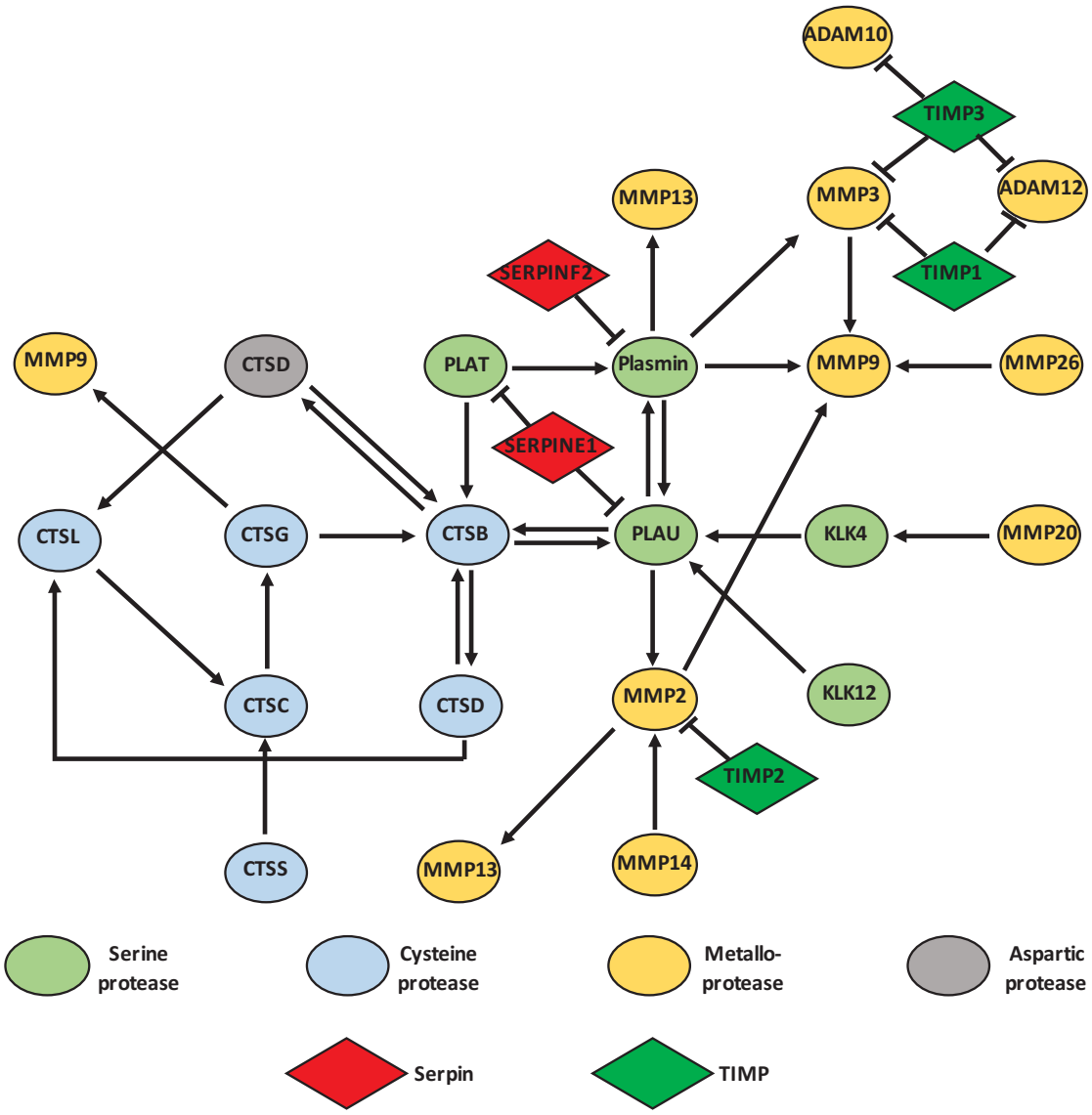
Kallikreins (KLKs)			
KLK1	KLK5	KLK9	KLK13
KLK2	KLK6	KLK10	KLK14
KLK3	KLK7	KLK11	KLK15
KLK4	KLK8	KLK12	

Tissue inhibitors of metalloproteinases (TIMPs)			
TIMP1	TIMP2	TIMP3	TIMP4

Table 5. Plasminogen receptors. The Table contains all 12 well-established plasminogen receptors, their corresponding gene name, cellular localization and C-terminal lysine status.

Plasminogen receptor	Gene name	C-terminal lysine	Most prominent cellular localization
actin	*ACTB*	No	cytoplasm
αMβ2 integrin	*ITGM* and *ITGB2*	No	Surface (integral membrane protein)
αVβ3 integrin	*ITGAV* and *ITGB3*	No	Surface (integral membrane protein)
$\alpha II_b\beta 3$ integrin	*ITGA2B* and *ITGB3*	No	Surface (integral membrane protein)
α-enolase	*ENO1*	Yes	cytoplasm
Cytokeratin 8	*KRT8*	Yes	cytoplasm
Histone H2B	*HIST2H2BE*	Yes	Nucleus/surface
HMGB1	*HMBGB1*	No	Nucleus/cytoplasm/surface
Plg-rKT	*PLGRKT*	Yes	Surface (integral membrane protein)
p11	*S100A10*	Yes	Cytoplasm/surface
TIP49α	*RUVBL1*	Yes	nucleus
GAPDH	*GAPDH*	No	Surface, cytoplasm

Figure 8. Proteolytic networks. The figure is a simplified illustration of the interactions between the four types of proteases at the cell surface. Light green circles represent serine proteases, blue circles represent cysteine proteases (e.g. cathepsins except cathepsin D, CTSD), yellow circles represent metalloproteinases (e.g. MMPs) and gray circles represent aspartic proteases (Cathepsin D, CTSD). Dark green rhombuses represent TIMPs (e.g. TIMP1, TIMP2) while red rhombuses represent serpins (e.g. PAI-1 (SERPINE1) and **α**2-antiplasmin (SERPINF2)).


ADAM10
TIMP3
ADAM12
MMP13
MMP3
TIMP1
SERPINF2
MMP9
CTSD
PLAT
Plasmin
MMP9
MMP26
SERPINE1
CTSL
CTSG
CTSB
PLAU
KLK4
MMP20
CTSC
CTSD
MMP2
KLK12
TIMP2
CTSS
MMP13
MMP14
Serine protease
Cysteine protease
Metallo-protease
Aspartic protease
Serpin
TIMP

1.5.2 Plasmin

Plasmin contains a catalytic triad which is common in serine proteases and is formed by His602, Asp645 and Ser740. This triad gives plasmin its broad-spectrum protease activity. Excessive plasmin production by cancer cells was first observed in 1925 by A. Fischer where cancer cells completely degraded the underlying fibrin matrix while normal cells failed to do so [150]. Plasmin is important in the regulation of ECM remodeling, a characteristic which is frequently exploited by malignant tumors to proteolytically cleave ECM components such as laminin and fibronectin [151]. Plasmin can also activate various MMPs and growth factors further degrading the ECM to allow tumor cell progression [152][153]. It is proposed that cell surface-associated plasmin acts to proteolytically cleave membrane-associated MMPs such as MMP3, MMP9 and MMP13 in the pericellular environment [154]. Only MT-MMPs and furin-activated MMPs (MMP11, 21, 28) are cleaved and activated intracellularly independent of plasmin [155].

1.5.3 Plasminogen Activators

Plasminogen is activated into plasmin via two specific serine proteases termed tissue plasminogen activator (tPA) and urokinase plasminogen activator (uPA). The overall structure of the catalytic domain of both plasminogen activators displays a typical serine protease fold with multiple insertion loops surrounding the active site cleft. The structure of these insertion loops is what determines their specificity to plasminogen [156]. Generally, tPA-mediated activation of plasminogen is implicated in fibrin clot dissolution where fibrin serves as a binding partner for both plasminogen and tPA [157]. In contrast,

uPA-mediated activation of plasminogen is frequently affiliated with extracellular tissue remodeling and cellular motility [158].

1.5.4 Urokinase Plasminogen Activator (uPA)

1.5.4.1 scuPA

The urokinase plasminogen activator uPA is encoded by the *PLAU* gene as a single 411-amino acid precursor called single-chain pro-urokinase (scuPA). ScuPA consists of three domains, a growth factor domain for binding to the uPA receptor (uPAR) [159], an active protease domain for plasminogen cleavage [158][160] and a kringle domain that binds αvβ3 integrin [161][162]. ScuPA can undergo post translational modifications including phosphorylation on Ser138 and Ser303 [163], N-glycosylation of Asn302 and fucosylation on Thr18 [164].

1.5.4.2 Single-chain to Two-chain uPA

uPAR, plays a central role in recruiting scuPA to the cell surface for plasminogen activation. scuPA-uPAR binding allows cleavage of scuPA at the peptide bond of Lys158-Ile159 by plasmin, cathepsin B or glandular kallikrein mGK6 (KLK1) [165]. The cleavage event allows the formation of a disulfide bond between two scuPAs giving rise to the two-chain uPA. The cleavage/activation of scuPA by plasmin is also known as reciprocal zymogen activation where plasmin promotes a positive feedback loop to exacerbate plasminogen activation [166] (figure 7).

1.5.5 Tissue Plasminogen Activator (tPA)

Like uPA, tPA is produced as a single chain form (sctPA) which is then cleaved by plasmin into the active two-chain tPA [167]. tPA is a 70kDa glycoprotein encoded by the *PLAT* gene and is predominantly produced by endothelial cells under physiological conditions [168], neurons and microglia [169] and cancer cells[170]. tPA circulates at a relatively low concertation of 5 ng/ml with a plasma half-life of 5 minutes. Upon release from endothelial cells under normal conditions, tPA is rapidly bound to the inhibitor PAI-1 (discussed later) and to a lesser extent to α2-microglobulin (discussed later) after which it is cleared by the liver. When a vascular trauma or ischemia occurs, endothelial cells (or at times neuronal terminals adjacent to the vasculature) dramatically enhance tPA production [171]. tPA maintains blood vessel patency via activating plasminogen into plasmin and degrading intravascular fibrin clots after the injury. *PLAT*-null mice display an incapability to degrade fibrin clots and have exacerbated fibrin deposition in tissues including the brain, which can lead to brain injuries under stroke-inducing conditions [172][173].

The amino-terminus of tPA contains a finger domain that binds to fibrin with high affinity at two binding sites (K_d = 31 nmol/L and 244 nmol/L respectively) [174]. tPA binding to plasminogen is stronger than that of uPA (0.3 uM) and is zinc-dependent [175] suggesting that tPA is a more effective fibrinolytic activator. tPA can also bind ECM proteins such as fibronectin, laminin and insulin growth factor. The binding of tPA to fibrin is mediated through two kringle domains increasing tPA's ability to activate plasminogen

by 3-fold [176]. Fibrin and plasminogen receptors can protect tPA from the circulating inhibitor PAI-1 [177][178].

1.5.6 Urokinase Plasminogen Activator Receptor (uPAR)

uPAR is a glycosyl-phosphatidylinositol membrane-anchored receptor for uPA. Intact uPAR contains three homologous domains (I, II and III) of which domain I can be cleaved by uPA. The cleaved uPAR with domains II and III remains on the surface but can also be shed as a soluble form. As a result, three soluble forms of uPAR are released: uPAR-I-II-III, uPAR-I and uPAR-II-III [179].

1.5.7 Plasminogen Activation Inhibitors

Plasminogen activation is tightly regulated through a balance between activators and inhibitors to prevent or abort aberrant production of plasmin. Serpins are a family of serine protease inhibitors with a wide range of inhibitory capabilities that control proteolytic events such as the coagulation cascade [180]. Serpins possess a reactive center loop that mimics the protease substrate, which upon binding to the protease forms an inactive complex [181][182]. Seven serpins have been demonstrated to regulate plasminogen activation including SERPINE1 (PAI-1) [183], SERPINE2 (PI7; protease nexin I) [184], SERPINB2 (PAI-2) [185], SERPINF2 (α2-anti-plasmin) [186], SERPINI2 (PI12, neuroserpin) [187] and A2M (α2-macroglobulin) [188].

PAI-1 is released into the extracellular space to act as an inhibitor of uPA, tPA, plasmin and thrombin. PI7 is also extracellular and inhibits both uPA and tPA. PAI-2 is largely intracellular and is an inhibitor of uPA [185]. A small percentage of intracellular

PAI-2 is glycosylated and enters the secretory pathway to be released into the extracellular space through a process known as facultative translocation [189]. α2-anti-plasmin is extracellular and is a potent inhibitor of plasmin. The binding of α2-anti-plasmin to plasmin is mediated through Lys436 and Lys452 with possible involvement of internal lysines in initiating binding to the kringle domains of plasmin [190]. Lastly, neuroserpin is extracellular and is an inhibitor of uPA, tPA and plasmin [180].

1.5.8 Plasminogen Receptors

1.5.8.1 Binding Plasminogen

Plasminogen receptors are a heterogeneous group of proteins that share a common ability to bind plasminogen. The plasminogen binding capability is predominantly mediated through a carboxy-terminal lysine residue, which is either part of the uncleaved receptor or is exposed after a proteolytic cleavage event [191]. The C-terminal lysine sensitizes these receptors to cleavage by carboxypeptidases (e.g. carboxypeptidase B) or inhibition by lysine analogs such as ε-aminocaproic acid [192]. However, some plasminogen receptors such as integrins αvβ3, αMβ2 and αIIβ2 lack the canonical C-terminal lysine and are not well characterized. Table 5 summarizes the 12 well-established plasminogen receptors, their corresponding gene names, cellular localization and C-terminal lysine status (table 5). The two most recent additions to the list of plasminogen receptors are PLGRKT [193] and GAPDH [194]. The discovery of novel plasminogen receptors is an ongoing field of research which is mostly accomplished by the identification of new cell surface proteins that are were previously considered intracellular or nuclear proteins.

Although both tPA and uPA can activate circulating plasminogen into plasmin at a very low rate, the localization of plasminogen/plasmin to the cell surface ensures its proximity to the plasminogen activators. This reduces Michaelis constant (K_m) by 60-fold and therefore enhances proteolytic and fibrinolytic activity. In addition, the binding of plasminogen to receptors also promotes its conversion from the "activation-resistant" glu-plasminogen into the "activation-sensitive" lys-plasminogen. Once activated, receptor-bound plasmin is protected from α2-anti-plasmin [195].

1.5.8.2 Tissue Expression

Plasminogen receptors are ubiquitously expressed across various tissues cell surfaces at high surface densities of 10^5 to 10^7 receptors per cell [196]. These receptors are present in a wide range of cell types including monocytic and lymphocytic immune cells, neurons, platelets, endothelial cells and epithelial cells [195] but not on red blood cells [197]. Plasminogen receptors can either be anchored to the plasma membrane (tailed receptors) or bound to an anchoring partner protein (tail-less receptors) both of which are capable of binding plasminogen. Tailed plasminogen receptors such as integrins (αvβ3, αMβ2) are more ubiquitous on immune cells where they also transmit cell adhesion and migration signals and activate intracellular survival signaling pathways [198] (table 5).

1.5.8.3 Broad Functions

A plethora of evidence documented the relevance of plasminogen receptors in cell surface regulation of plasmin production using physiological and pathological models [195][199]. For instance, cytokeratin 8 is expressed at the cell surface of breast cancer cells

and is important for activation of plasminogen by tPA [200]. In inflammation, plasminogen receptors mainly function to mediate immune cell recruitment and promote matrix degradation and proteinase activation. Enolase 1 promoted plasminogen-mediated recruitment of monocytes to sites of acute inflammation in the lungs [201]. Plg-R_{KT} is also required for the migration of macrophages under inflammatory conditions [202].

Apart from the capacity to bind plasminogen at the cell surface, plasminogen receptors vary largely in structure and distribution. Various plasminogen receptors have been studied in recent years as proteins dysregulated in disease. These include cytokeratin-8, α-enolase, Plg-R_{KT}, H2B, S100A4, and HMGB-1, with involvements in cell invasion and cancer metastasis through multiple mechanisms [199]. Cytokeratin 8 expression positively correlated with enhanced invasiveness of breast cancer cells [200] and increased expression has been observed in pancreatic [203], colorectal [203], and oral squamous cell carcinomas [204].

1.5.9 Matrix Metalloproteinases and their Inhibitors

Matrix metalloproteinases or MMPs belong to the zinc-dependent family of endopeptidases and are involved in several processes such as organogenesis, wound healing [205], inflammation [206] and oncogenesis [207]. Overexpression of MMPs by both tumor and stromal cells has been shown to contribute to carcinogenesis [207]. Mechanistically, MMPs degrade the physical barriers presented by the ECM to promote invasion and are often recruited to invadopodia where they mediate matrix breakdown at the invasive fronts [208]. MMPs can also promote cell proliferation by increasing shedding of membrane-anchored EGFR ligands such as heparin-bound EGF and TGFα [209]. MMPs

also increase the shedding of E-cadherin which releases β-catenin and allows its translocation to the nucleus to promote proliferation [210].

A close family of MMPs is a group of proteins called ADAMs (a disintegrin and metalloproteinase) which are either membrane-bound ADAMs or ADAMs with thrombospondin motifs (ADAMTS). ADAMs share similar functions with MMPs in their ability to proteolytically cleave ECM but they also possess non-proteolytic functions related to integrin-mediated adhesion [211][212]. Inhibitors of matrix proteinases, known as TIMPs (tissue inhibitors of metalloproteinases) are endogenous regulators of the ECM remodeling and turnover. Four paralogous genes encode TIMP1 through 4 of which TIMP3 is an ADAM and ADAMTS inhibitor [213]. TIMP1 is a potent inhibitor of MMP3 and MMP7 as well as ADAM10 and ADAM12 [214]. TIMP3 and TIMP4 inhibit ADAM17 [215] (figure 8). TIMPs have been positively and negatively implicated in cell cycle control, apoptosis, angiogenesis, synaptic plasticity and cellular differentiation [213].

1.5.9 Cathepsins and Kallikreins

In humans, the cysteine cathepsin family consists of 11 members which are mainly endopeptidases (except cathepsins C and Z) [216]. Many cathepsins have been implicated in modulating the tumor microenvironments by degrading the ECM [217][218], activating growth factors [219] and shedding cell-cell adhesion molecules [217] all of which contributing to enhanced invasion and metastasis [220][221]. Kallikreins (KLKs) are a family of trypsin-like serine proteases encoded by 15 structurally similar genes in humans (KLK1 through 15). Physiological functions of KLKs include cellular growth and tissue remodeling. However, multiple KLKs have been found to be upregulated (e.g. KLK11 in

neuroendocrine carcinoma and KLK10 in pancreatic cancer) or downregulated (e.g. KLKs 2,3,5,6,10 and 13 in prostate cancer) in several malignancies [222].

1.5.10 Plasminogen Activation: An Orchestrated Process Between Plasminogen, its Activator and its Receptor (*in the Eyes of a PhD Student*)

Circulating Glu-plasminogen first binds to carboxyl-terminal lysines on plasminogen receptors. The binding alters Glu-plasminogen from an activation-resistant to an activation-prone conformation. The cleavage of Glu-plasminogen by plasmin into Lys-plasminogen further increases the susceptibility of plasminogen to activation by plasminogen activators. Plasminogen activators tPA and uPA then cleave plasminogen into the active protease plasmin. Plasmin can reciprocally activate pro-uPA into active uPA creating a positive feedback loop. In addition, plasmin cleaves and activates MMPs thereby degrading the ECM and activating a series of matrix-sequestered growth factors. Importantly, the activity of uPA is dependent on binding its receptor uPAR, which induces the clustering of uPAR in the plasma membrane into cholesterol- and sphingolipid-rich areas (figure 7). Enhanced surface expression of pro-uPA and uPA along with the concomitant increase in plasminogen receptors accelerate the generation of plasmin. Notably, uPA-mediated activation of plasminogen that is not bound to a receptor is markedly lower than when plasminogen is bound to a receptor and is in the activation-prone form. The proteolytic activities of plasmin and uPA are inhibited by serpins such as PAI-1, PAI-2 and α2-antiplasmin (figure 8).

1.6 Sub-chapter 6: The Plasminogen Receptor S100A10

1.6.1 Structure

S100A10 or p11 belongs to the S100 family of small calcium-binding proteins with molecular weights ranging from 9 to 13 kDa [223]. The family includes 20 members; 16 members (S100A1-A16) are encoded by separate genes in a defined region of chromosome1q21. The remaining four members S100B, S100G, S100P, S100Z are located outside the 1q21 region [224]. The calcium binding function of S100 proteins is attributed to two calcium binding loops called EF-hand motifs. One EF-hand motif (EF1) is located on the carboxyl-terminus and is shared with all other calcium-binding proteins such as calmodulin and troponin [225]. EF1 contains the canonical 12-amino acid calcium-binding sequence (DXDGDGTIXXXE) with highly acidic side chains of aspartic (D) and glutamic acid (E). The other EF-hand motif (EF2) is a S100-specfic motif and is located on the N-terminus. This motif is unconventional in that it is 14 amino acids long and binds calcium through the carbonyl backbone of amino acids and the carboxyl group of glutamic acid [226].

The C-termini of S100 proteins exhibit the most variability throughout evolution and it is the main distinguishing factor in their different functions [227]. The S100 family is considered a relatively young group of proteins having emerged about half a billion years ago from a calmodulin-like protein. This is supported by the fact that S100 proteins have been only found in vertebrates and not in invertebrate eukaryotes [228]. Despite that, S100 proteins have been demonstrated to be highly adaptive proteins with a large degree of "interactivity" potential. Permyakov *et al*. coined the term "intrinsic disorder" to describe

the structural adaptability and functional versatility of these proteins [229]. This is further illustrated by the wide range of interacting proteins (discussed later).

The uniqueness of S100A10 within the S100 family of proteins arises from the fact that the S100A10 EF hand motifs cannot bind calcium. This is due to three deletions in the linker regions between H1 and H2 and two mutations that gave rise to glutamic acid and asparagine substitutions in the EF domain (figure 9). Interestingly, although these substitutions render S100A10 incapable of binding calcium, the resultant conformational change resembles a calcium-bound state i.e. a constitutively active [230].

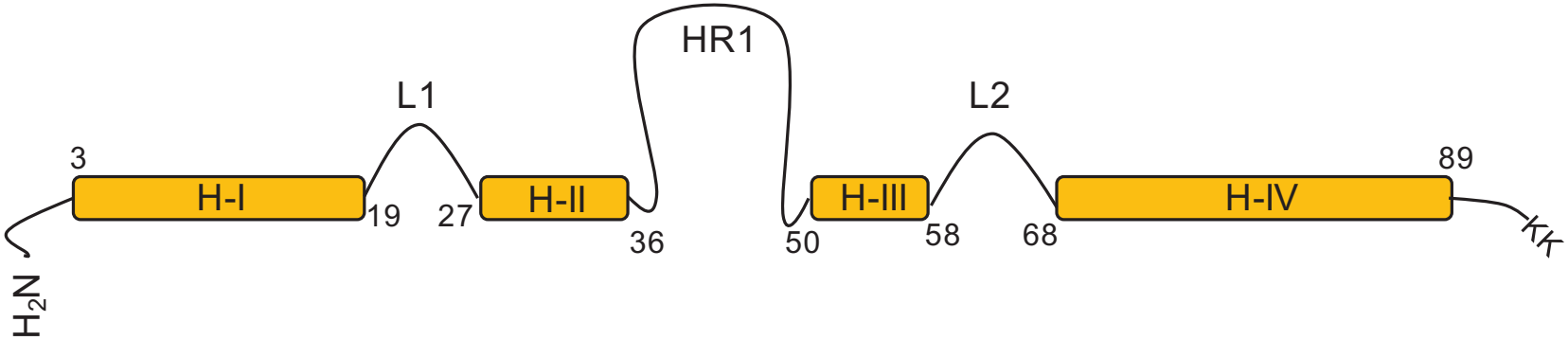
Monomeric S100A10 contains four alpha-helices known as H-1 (residues Q3-A19), H-2 (residues K27-K36), H-3 (residues A50-L58) and H-4 (residues F68-H89). The helices are separated by two loops L1 and L2 which form the calcium-binding loops. The calcium EF1 loop is located between H-1 and H-2 (residues A19-L30) while the canonical EF2 loop is located between H-3 and H-4 (D59-S70) [230]. The region between H-2 and H-3 is known as the hinge region (HR, residues P39-N44), providing S100A10 with its conformational flexibility (figure 9).

1.6.2 A Putative Plasminogen Receptor

As mentioned previously, plasminogen receptors are increasingly relevant in regulating various diseases including stroke, inflammation and cancer. The activation of plasminogen by plasminogen activators is amplified upon binding to a receptor at the cell surface. S100A10 meets all the criteria of a putative plasminogen receptor. Firstly, it binds plasminogen with a binding affinity (K_d) of 1.81 μM. Secondly, S100A10 possesses a

carboxy-terminal lysine, which has been shown to be essential in binding plasminogen [231]. Thirdly, S100A10 alone or in partnership with annexin A2 (discussed later) binds plasminogen inducing a conformational change into the open activation-prone conformation in cell-free conditions [177]. Lastly, S100A10 can protect plasmin and plasminogen activators (tPA) from their inhibitors, α2-antiplasmin and PAI-1 respectively [232]. S100A10 also binds and/or localizes to plasminogen activators and their receptors. It is believed that this localization greatly enhances plasminogen activation and is driven by oncogenic events [233]. S100A10 binds tPA (K_d=0.45μM) and accelerates plasminogen activation [177], allows localized proteolysis, while also protecting plasmin from inhibition by α2-antiplasmin [177]. S100A10 also localizes with uPAR at the cell surface of HT1080 fibrosarcoma [234] and Colo222 colorectal [235] cancer cells. Loss of S100A10 from the extracellular surface of cancer cells results in a significant decrease in plasmin generation in macrophages [236], HT1080 fibrosarcoma [234], Colo222 colorectal [235], NB4 leukemic [237] and Lewis-lung carcinoma [238] cancer cells.

Figure 9. Structure of the S100A10 monomer. Each monomer contains of four α-helices (H-I, H-II, H-III, and H-IV). Two L1 and L2 separate H-I from H-II and H-III from H-IV respectively. H-II and H-III are linked by a flexible hinge region (HR1). The C-terminal lysines are also shown and represent the binding sites for plasminogen and tissue plasminogen activator (tPA).


HR1
L1
L2
3
H-I
19
27
H-II
36
50
H-III
58
68
H-IV
89
H_2N

1.6.3 Role in Fibrinolysis

Vascular fibrinolysis is a fundamental process where endothelial cells lining the blood vessels play a key role in preventing blood clotting via the production of plasmin. Given its role in plasminogen activation, S100A10 is also a player in clot breakdown [239]. S100A10 knockdown in TIME cells (telomerase immortalized microvascular endothelial cells) resulted in a dramatic decrease in their ability to bind and activate plasminogen. The S100A10 knockout mice display signs of aberrant fibrinolytic activity and accumulate fibrin in various tissues (lung, spleen, liver and kidney) compared to wild-type mice. These mice were inefficient in breaking down batroxobin-induced blood clots, a consequence of reduced fibrinolysis. The absence of S100A10 *in vivo* not only affected blood clot breakdown and fibrin tissue deposition but also the formation of blood vessels or angiogenesis. S100A10 knockout mice showed reduced CD31 staining indicating impaired vascularization [240].

1.6.4 Role in Cancer

A series of studies in the last 20 years have addressed the role of S100A10 as a plasminogen receptor in various cancer models [241][242]. For instance, S100A10 promoted the activation of plasminogen and invasiveness of macrophages [236], HT1080 fibrosarcoma cancer cells [234], Colo 222 colorectal cancer cells [235] and NB4 leukemic cells [237]. Consequently, the loss of S100A10 dramatically reduced surface plasmin generation and the invasive capacity of these cancer cells. The loss of S100A10 also reduced the metastatic burden in the lungs of mice intravenously injected with HT1080 cells [234]. A more recent report showed that S100A10 also regulated the infiltration of

tumor-associated macrophages (TAMs) into tumor sites and was essential for the growth of a tumor in a xenograft model of mouse Lewis lung carcinoma (LLC). In fact, LLC tumors failed to grow in S100A10-null mice and were less angiogenic, potentially due to failure of recruitment of these macrophages. Tumor growth was restored when S100A10-null macrophages were intra-tumorally injected (and not intravenously) into LLC tumors of S100A10-null mice suggesting that S100A10 was required for the infiltration step via blood vessel walls. Collectively, these studies highlighted the importance of the plasminogen activation system, mediated by stromal cell and cancer cell S100A10, in tumor growth [238].

1.6.5 A Role in a Hyper-Fibrinolytic Cancer

Patients of acute promyelocytic leukemia (APL) suffer from severe bleeding complications primarily caused by a hyper-fibrinolytic vasculature and low-platelet counts. Hyper-fibrinolysis or aberrant fibrinolysis is mediated through the accelerated capability to generate plasmin, which breaks down fibrin clots formed at wound sites. A recent report by O'Connell *et al*. showed that S100A10 depletion in NB4 leukemia cells resulted in over 70% decrease in the plasminogen-binding and activation at the cell surface. The ability to degrade fibrin was also hampered in S100A10-depleted cells and in S100A10-null mice, which manifested as increased fibrin deposition in various tissues. Moreover, induced expression of *PML-RARα*, the fusion oncogene responsible for APL, upregulated the expression of cell surface S100A10 in the myeloid precursor PR9 cells. ATRA (all-trans retinoic acid), a standard treatment for APL patients, downregulated S100A10 expression [237]. This study provides a potential mechanism for plasmin contributing to the hyper-

fibrinolytic phenotype of APL patients [240]. Additionally, it indicated that S100A10 is a regulator of plasmin-mediated fibrinolysis.

1.6.6 Interaction with Annexin A2

The heterotetrameric complex formed between annexin A2 and S100A10 represents a unique example of how plasminogen receptors can be implicated in physiological and pathological conditions including inflammation, stroke and cancer [242]. It is generally believed that plasminogen activation is localized in glycosphingolipid-rich plasma membrane micro-domains, called lipid rafts. uPA and uPAR localize with the S100A10/annexin A2 heterotetramer to promote plasminogen activation at these sites [243][244]. The heterotetramer also binds to the kringle domains of tPA and plasminogen via the S100A10 subunits [177][245]. Carboxypeptidase B treatment, which cleaved the C-terminal lysines of the native annexin A2 heterotetramer, led to an 80% decrease in plasminogen activation [246]. Noteworthy, the C-terminal lysines of S100A10 are also sensitive to other carboxypeptidases including carboxypeptidase N and TAFI (Thrombin activatable fibrinolysis inhibitor) [246]. The combination of wild-type annexin A2 with either wild-type S100A10 or a mutant S100A10 (S100A10ΔKK) which lacks the two C-terminal lysines revealed that the mutant complex possessed minimal plasminogen activation capacity (12%) compared to the wild-type heterotetramer. These findings emphasized the importance of the two C-terminal lysines of S100A10 in plasminogen binding and activation.

1.6.6.1 S100A10 Stability

The relationship between annexin A2 and S100A10 is predominantly dictated by the dependence of S100A10's stability on the presence or sustained expression of annexin A2. For instance, transient knockdown of annexin A2 in MDCK cells resulted in reduction of both annexin A2 and S100A10 protein expression [247]. Annexin A2-null mice express considerably lower levels of S100A10 protein. The N-terminus of annexin A2 protects S100A10 by masking the C-terminus region which contains poly-ubiquitination sites hence preventing S100A10 degradation [248]. This permits the translocation of S100A10 to the cell surface [248][249][250][251][252]. Interestingly, depletion of S100A10 by siRNA did not result in a decrease in annexin A2 protein expression [253][234].

1.6.6.2 Sites of Interaction

In the 1990s, studies utilized site-directed mutagenesis to identify the amino acids required for S100A10 association with its binding partner annexin A2 [254]. The binding is mediated through a four amino-terminal amphipathic helix (V3, I6, L7, L10) on annexin A2. This helix binds to the hydrophobic cleft formed by the hinge region (HR) and H-1 of one S100A10 monomer and the H-3 region of the other monomer [255][256]. Annexin A2 forms multiple points of contact with the S100A10 monomers making this interaction highly favorable and specific. The four amino acids form seven, two and nine sites of interactions with the H-1, HR and H-3 respectively (reviewed in [224]).

1.6.6.3 Role in Auto-proteolysis of Plasmin

In addition to directly cleaving matrix substrates such as fibrin, fibronectin and laminins, plasmin can also undergo auto-proteolysis. Furthermore, the annexin A2-S100A10 heterotetramer can stimulate plasmin auto-proteolysis [257]. The self-destruction phenomenon is believed to be evolutionary means of mitigating collateral tissue damage resulting from uncontrolled accumulation of plasmin in tissues.

1.6.6.4 Addressing an Enduring Ambiguity

Many reports have suggested that annexin A2 is also a putative receptor for plasminogen [258]. It has been challenging to attribute any plasminogen-dependent cellular changes to annexin A2 and/or S100A10. Based on the evidence presented below, this challenge is at least partially addressed, and concludes that S100A10 is the sole receptor for plasminogen within the heterotetramer [241]. A study by Kwon *et al*. utilized site-directed mutagenesis of plasminogen to change a serine residue in the plasmin catalytic site into cysteine which was subsequently labeled with fluorescein. Results showed that the purified heterotetramer induced a conformational change in glu-plasminogen (K_d = 1.26 μM). However, purified monomeric annexin A2 failed to induce a conformational change suggesting that either annexin A2 did not bind plasminogen or that the proposed interaction of plasminogen with annexin A2 was mechanistically distinct from that involving the entire heterotetramer [259]. The heterotetramer and S100A10 monomer are proficient at mediating tPA-dependent activation of plasminogen, 3 to 4-fold (respectively) higher than that of monomeric annexin A2 [232][177]. Furthermore, Fog *et al*. generated recombinant heterotetramers formed by wild-type annexin A2 with either wild-type S100A10 or a

mutant S100A10 that lacks two carboxyl-terminal lysines. The S100A10-mutant heterotetramer possessed around 10% of the activity of the wild-type heterotetramer, which emphasized the importance of the two carboxyl-terminal lysines of S100A10 in plasminogen binding [246]. Lastly, surface plasmon resonance (SPR) studies revealed that while the heterotetramer bound both tPA (K_d=0.68μM) and plasminogen (K_d=0.11μM), monomeric annexin A2 failed to do so [253].

1.6.7 Interactors

Although S100A10's interaction with annexin A2 is the most-studied and well-established interaction especially in relation to disease, multiple studies have demonstrated various interacting partners for S100A10. The plasticity of S100 proteins renders them very promiscuous in their binding capacity. The S100A10-interacting proteins discussed below are summarized in table 6. Figure 10 also summarizes the various functions of S100A10 in physiological and pathological models based on the proposed interactions (figure 10).

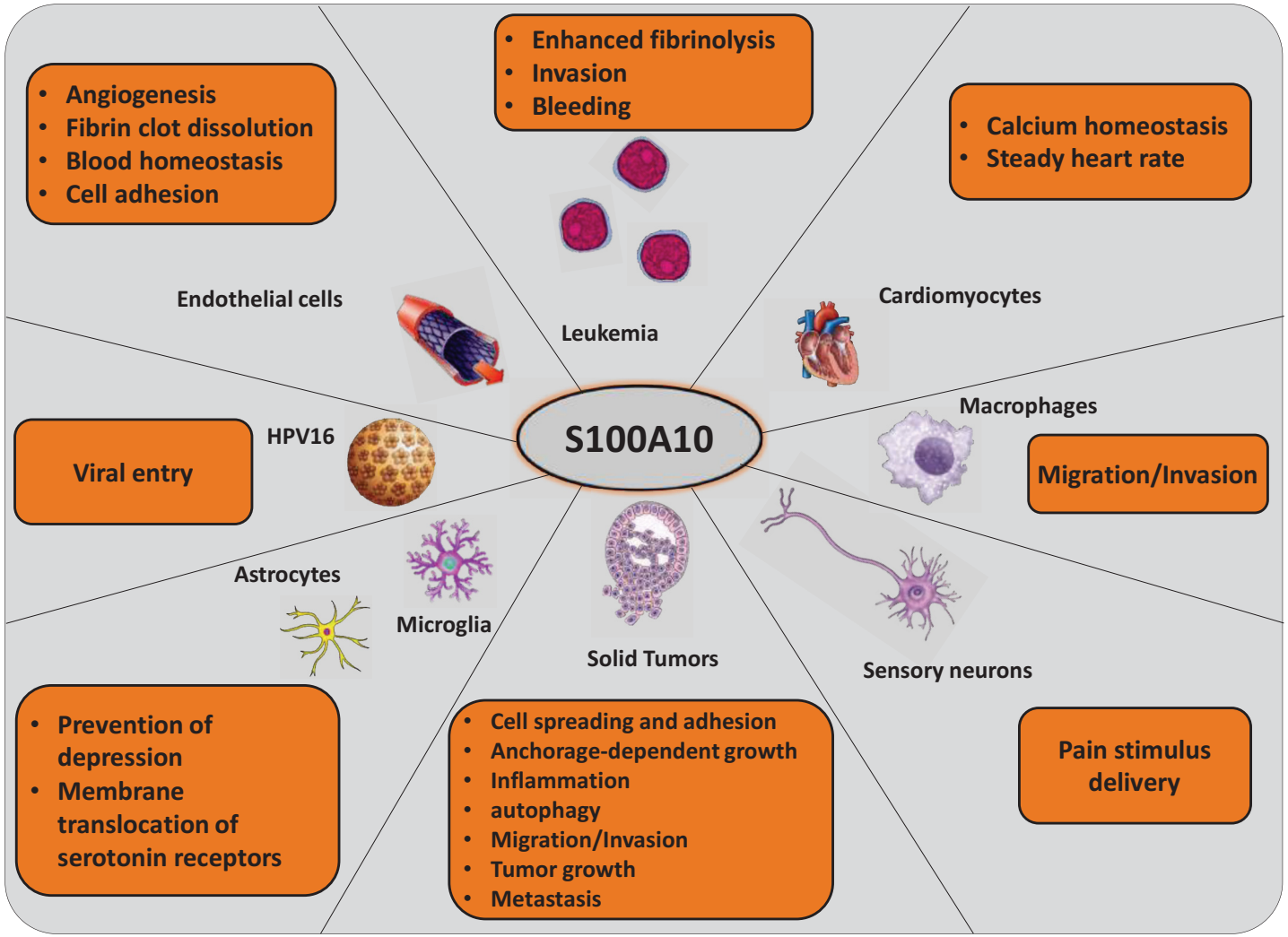
Yang *et al*. demonstrated that the protein DLC1 (deleted in liver cancer 1) competes with annexin A2 to bind S100A10 and in turn promotes ubiquitin-mediated degradation of S100A10. This interaction led to decreased cellular invasion, migration, colony formation and anchorage-independent growth of lung cancer cells in a Rho GTPase-dependent manner [260]. In a slightly different context, S100A10 was shown to bind the carboxy-terminal cytoplasmic tail of the chemokine receptor CCR10 and regulate its surface localization [261]. CCR10 belongs the GPCR family of proteins and is involved in mediating inflammatory responses [262]. A recent report by Chehab *et al*. examined Weibel-Palade bodies (WPBs) that are secretory granules storing the pro-coagulant von

Wille-Brand factor (VWF). The exocytosis of WPBs from endothelial cells is dependent on their recruitment to the plasma membrane upon insult. The study identified that Munc13-4 (protein unc-13 homolog D) directly interacted with S100A10 (within the heterotetramer) to form a complex that was essential for cell membrane recruitment and exocytosis of WPBs [263]. Chen *et al*. recently examined the role of S100A10 in autophagy and autophagosome formation in bronchial epithelial cells in response to interferon gamma (IFN-γ). The group demonstrated that S100A10 was essential for autophagosome formation via interactions with the serine/threonine-protein kinase (ULK1) promoting its localization to the autophagosome formation sites [264]. Herein, ULK1 phosphorylates and activates essential autophagy-related proteins such as ATG9 and Beclin1 [265][266]. S100A10 and annexin A2 also interact with sphingolipid ceramide 1-phosphate (C-1-P) and help facilitate cellular invasion [267]. Furthermore, both S100A10 and Annexin A2 interact with a large protein called AHNAK that is involved in membrane repair [268].

Table 6. S100A10 interactors. The table describes the contributions of S100A10 to its interactors and the impact on their cellular functions.

Interactor	Functional contribution of S100A10 in interaction
Annexin A2	- Enhanced endothelial cell fibrinolysis, cell adhesion, cellular spreading, cancer cell invasion and metastasis
DLC1	- Decreased cell migration and invasion, colony formation and anchorage-independent growth
5-HT_{1B}	- Is required for membrane translocation - Prevents depression
mGLuR5	
5HT_4R	- Is required for membrane translocation - Aberrant increase leads disrupted calcium ion handling and increase in heart rate
TRPV5 and TRPV6	- Is required for membrane translocation - Calcium homeostasis
TASK1	- Is required for membrane translocation - pH homeostasis
Na(V)1.8	- Is required for membrane translocation - Maintains delivery of pain signals in sensory neurons
CCR10	- Is required for membrane translocation - Inflammation and cancer progression
Munc13-4	- Required for exocytosis of Weibel-Palade bodies (WPBs) and subsequent release of Von Wille-brand factor (VWF)
ULK1	- Formation of autophagosomes and mediation of autophagy
C-1-P	- Cellular invasion
AHNAK	- Cell membrane repair
L2 minor capsid protein	- Human papilloma virus 16 (HPV16) infection

Figure 10. Functions of the S100A10 protein. The figure summarizes the various functions of S100A10 in physiological and pathological conditions. The cell models include leukemia, solid tumors, cardiomyocytes, macrophages, sensory neurons, microglia and astrocytes and human papilloma virus 16 (HPV16).


S100A10
Endothelial cells
• Angiogenesis
• Fibrin clot dissolution
• Blood homeostasis
• Cell adhesion
Leukemia
• Enhanced fibrinolysis
• Invasion
• Bleeding
Cardiomyocytes
• Calcium homeostasis
• Steady heart rate
Macrophages
Migration/Invasion
Sensory neurons
Pain stimulus delivery
Solid Tumors
• Cell spreading and adhesion
• Anchorage-dependent growth
• Inflammation
• autophagy
• Migration/Invasion
• Tumor growth
• Metastasis
Microglia
Astrocytes
• Prevention of depression
• Membrane translocation of serotonin receptors
HPV16
Viral entry

S100A10 also interacts with the L2 minor capsid protein of the human papilloma virus 16 (HPV16) and is required for the internalization and infection of epithelial cells [269].

S100A10 has also been implicated in brain function under both physiological (e.g. neuronal function) and pathophysiological conditions (e.g. depression). A seminal study by Svenningsson and colleagues utilized the yeast-two hybrid screening system to identify proteins that interact with the serotonin receptor 1B (5-hydroxytryptamine (5-HT_{1B}) receptor), which is involved in serotonin neurotransmission [270][271]. S100A10 was the predominant prey clone that interacted with 5-HT_{1B} but not with other serotonin receptors (e.g. 5-HT_{1A}, 5-HT_{2A}, 5-HT_{5A}). In fact, S100A10 was responsible for the translocation of 5-HT_{1B} to the cell surface. Importantly, the study also revealed that *S100A10*-null mice display a depression-like phenotype, reduced responsiveness to serotonin receptor agonists and an incomplete response to antidepressants. Consequently, S100A10 was increased by long-term treatment of mice with the tricyclic antidepressant imipramine [270]. Other groups showed that S100A10 modulated the membrane translocation of serotonin receptor $5HT_4R$ in rat ventricular cardiomyocytes [272] and the metabotropic glutamate receptor 5 (mGluR5) in murine GABAergic neurons [273]. S100A10 also binds the amino terminus of the tetradotoxin-resistant sodium channel (Na(V)1.8) which is implicated in transmission of pain signals in sensory neurons. S100A10 binding promotes Na(V)1.8 translocation to the plasma membrane [274]. Another study by Van der Graaf *et al*. showed that S100A10 (and the entire heterotetramer) interacts with the calcium channels TRPV5 and TRPV6 (Transient receptor potential cation channel subfamily V member 5 and member 6) and was also required for routing these channels to the plasma membrane [275][276]. Similarly, Girard *et al*. identified that S100A10 binds the tandem pore (2P)

domain potassium channel TASK-1, which masks the endoplasmic reticulum retention signal and mediates its translocation to the plasma membrane [277]. Collectively, these results revealed a broad function of S100A10 that involves the translocation of a plethora of proteins to the cell surface.

1.7 Subchapter 7: Epithelial to Mesenchymal Transition (EMT)

In the 1970s, Markwald and colleagues were studying the types of cells that constitute the ECM (called "cardiac cushion" in the heart) of the atrio-ventricular canal of a chicken embryo heart. The authors utilized a video camera to track the first recorded transition of endothelial cells lining the cardiac vasculature from an epithelial to a mesenchymal morphology that then constituted the cardiac cushion [278]. Later reports described similar transitions in different types of tissues where cells undergo morphological and functional changes under both physiological and pathological conditions Epithelial to mesenchymal transition (EMT) gradually became a fundamental biological mechanism by which epithelial cells migrate during embryonic development (type I), tissue repair/fibrosis (type II) and cancer cell dissemination (type III). In 2007, EMT was officially classified into three distinct biological types during a conference in Poland [279]. All three types share similar biochemical pathways often activated via various signals such as TGFβ (discussed later), PDGF (platelet-derived growth factor), EGF (epidermal growth factor) and HGF (hepatocyte growth factor) [279]. The dynamic nature of EMT and MET (mesenchymal to epithelial transition) and the associated cellular changes are depicted in figure 11 and figure 12 respectively (figure 11, figure 12).

1.7.1 Types of EMT

1.7.1.1 Type I EMT

Type I EMT was originally characterized during heart morphogenesis of chicken embryos [278]. However, it is now implicated in various other embryonic processes including implantation, gastrulation and organ morphogenesis [279]. For instance, cells in the parietal endoderm undergo EMT that prompts implantation of the embryo and formation of the placenta [280]. Epithelial cells in the epiblast layer undergo EMT generating the primitive streak, the first step of gastrulation. This primitive streak is what ultimately gives rise to the three germ layers that generate all tissues of the body [281].

The epithelial cells forming the neuroectoderm express mesenchymal transcription factors such as Slug, Snail, Sox and FoxD3, which stimulate these cells to undergo EMT [282]. These epithelial cells become the migratory cells of the neural crest [283]. The migratory neural cells then dissociate from the neural folds to undergo differentiation into various cell types [284].

Figure 11. dynamic nature of EMT. EMT represents a spectrum of events that ranges from a highly epithelial, polarized and specialized cell to a mesenchymal motile cell with stem-like properties. Within that spectrum, cells can possess a partial EMT status where they retain expression of some epithelial markers while acquiring new mesenchymal markers. MET is the reverse process of EMT and is known as mesenchymal to epithelial transition.

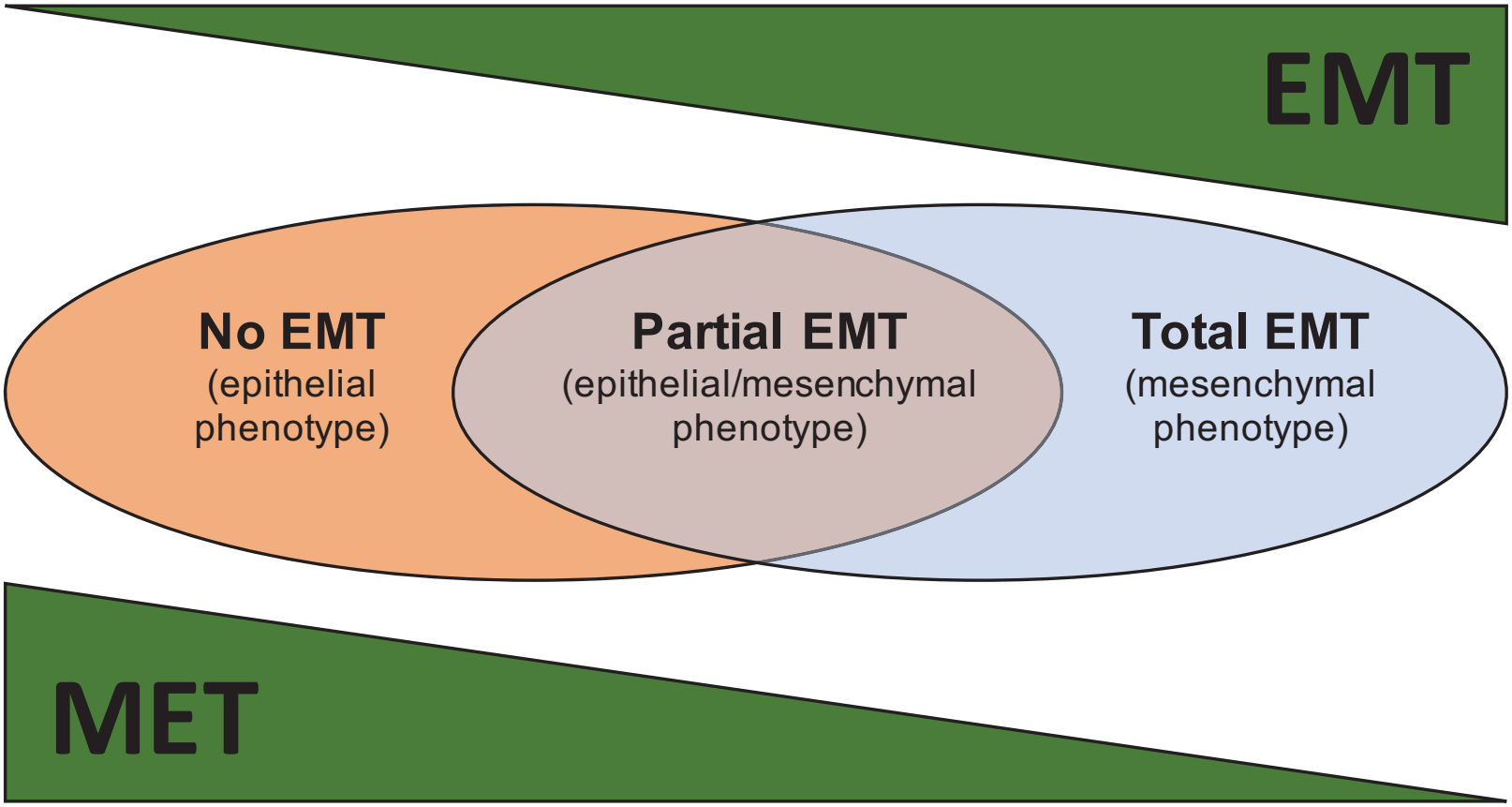
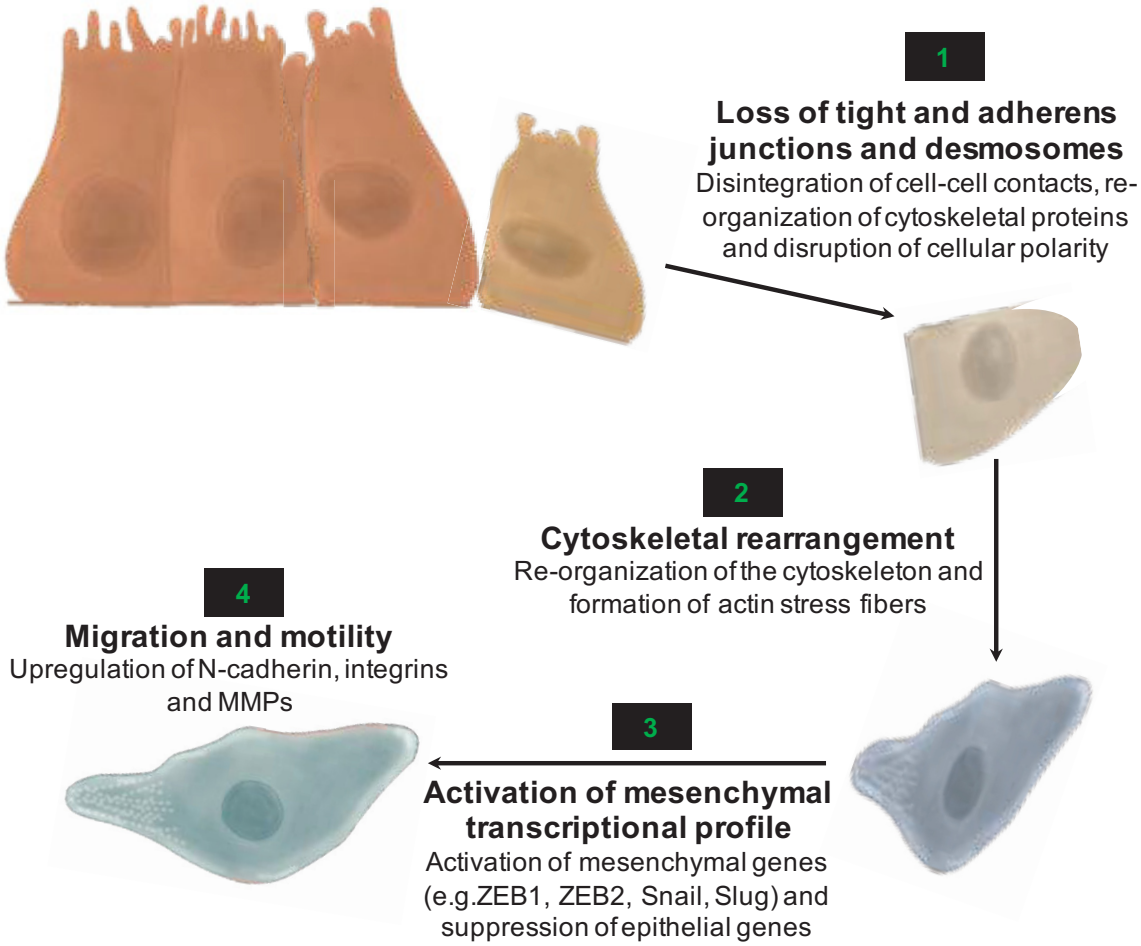

EMT
No EMT
(epithelial
phenotype)
Partial EMT
(epithelial/mesenchymal
phenotype)
Total EMT
(mesenchymal
phenotype)
MET

Figure 12. Progressive changes in epithelial cells undergoing EMT. EMT is defined by the progressive cellular changes that are initiated by loss of cell-cell junctions followed by cytoskeletal rearrangements and formation of stress fibers. The latter, when combined with elevated expression of EMT transcription factors, promote cell movement and migration. The migratory phenotype is supported by the expression of motility markers (N-cadherin) as well as invasive markers (e.g. MMPs).


1
Loss of tight and adherens junctions and desmosomes
Disintegration of cell-cell contacts, re-organization of cytoskeletal proteins and disruption of cellular polarity
2
Cytoskeletal rearrangement
Re-organization of the cytoskeleton and formation of actin stress fibers
3
Activation of mesenchymal transcriptional profile
Activation of mesenchymal genes (e.g.ZEB1, ZEB2, Snail, Slug) and suppression of epithelial genes
4
Migration and motility
Upregulation of N-cadherin, integrins and MMPs

1.7.1.2 Type II EMT

Type II EMT is linked to tissue and organ fibrosis mediated by fibroblasts, pro-inflammatory cells and a series of ECM components including collagen and laminin [279]. Tissues such as the liver, lung, intestine and kidneys are examples of organs where type II EMT takes place. Fibroblasts release fibroblast-specific protein 1 (FSP1) [285], collagen I, alpha smooth muscle actin (α-SMA), vimentin and others, which can be used as markers of chronic tissue fibrosis [286]. Meanwhile, the epithelial cells of the affected tissue also acquire mesenchymal markers such as FSP1 while maintaining expression of some epithelial markers such as E-cadherin and cytokeratin resulting in a "partial EMT" phenotype (Figure 11), as demonstrated in renal fibrosis [287]. This phenotype is sufficient for these cells to detach from the basement membrane and neighboring cells and then migrate to the interstitial space where they accumulate and complete their "partial EMT" to become fully mesenchymal and fibroblast-like [279]. Another example of type II EMT is pulmonary fibrosis, a lung condition characterized by irreversible destruction of lung architecture. Pulmonary fibrosis is primarily driven by TGFβ1 signaling, which mediates fibroblast proliferation and their migration to fibrotic sites [288]. TGFβ1 also induces EMT in pulmonary fibroblasts and myofibroblasts which are responsible for the aberrant extracellular matrix deposition [289].

1.7.1.3 Type III EMT

Unlike type I and II, the outcome of type III EMT is markedly different and unpredictable because it is driven/coupled with genetic events that occur in cancer cells. Cancer cell dissemination is initiated by the movement of cancer cells into the blood

vasculature, which is thought to be triggered by EMT. EMT enables an epithelial cancer cell to acquire an motile invasive phenotype [290] resulting in the downregulation of epithelial genes and the upregulation of mesenchymal genes [291]. Mesenchymal cells are commonly observed at the invasive front of primary tumors, displaying a poorly differentiated morphology and are thought to be the drivers of invasive escape of cancer cells giving rise to metastasis [292].

Type III EMT is characterized by the progressive loss of epithelial characteristics, mainly through the deconstruction of tight junctions and other cell-cell contact structures and reorganization of the actin cytoskeleton, both leading to the subsequent loss of apical-basal polarity and gradual dissociation from the basement membrane. E-Cadherin (discussed later) is a major protein component of intercellular junctions whose encoding gene is frequently repressed by a plethora of EMT transcription factors. Consequently, cells become more motile as they express specific cytoskeletal motility proteins, such as vimentin and N-cadherin (discussed later). Concurrently, MMPs are also activated, which degrade the impeding extracellular matrix proteins such as collagen and fibronectin. Mesenchymal cancer cells are then able to intravasate and subsequently exit the bloodstream at a secondary site, where they undergo mesenchymal to epithelial transition (MET) and may form secondary epithelial tumors or metastases [279].

1.7.1.4 Epithelial and Mesenchymal Markers

The list of epithelial and mesenchymal markers is always expanding as new markers are being readily discovered (table 7). Table 7 represents a compilation of epithelial and mesenchymal markers based on their cellular function and/or localization. These markers were divided into four different categories: cytoskeletal proteins, extracellular matrix proteins, cell surface proteins, transcription factors, miRNAs and lncRNAs (Table 7). The involvement of each marker is either universal as seen in the downregulation of E-cadherin and upregulation of N-cadherin during EMT or unique such as the requirement of Twist1 activation in specific cancer models but not others. Below is a detailed discussion of the two most-studied markers, E-cadherin and N-cadherin.

1.7.1.4.1 E cadherin

E-cadherin, also known as epithelial cadherin or cadherin 1, is encoded by the *CDH1* gene and is a member of the cadherin superfamily. The murine equivalent of human E-cadherin, uvomoulin shares an 80% nucleotide and amino acid sequence homology [293]. E-cadherin is a 120kDa glycoprotein and a calcium-dependent cell adhesion molecule (CAM) composed of a substantial extracellular domain (ED), a single transmembrane domain and a short cytoplasmic intracellular domain (ID). The latter interacts with α-, β-, γ-catenins that link the ID to the actin cytoskeleton [294]. The ID contains a highly conserved series of 150 amino acid residues (juxta-membrane region) that have been demonstrated to modulate the cell to cell adhesion function of the ED through its interactions with the actin cytoskeleton [295]. The ED contains five folded repeats of 110 amino acids each, which contain the Ca^{2+} binding sites and dictate the

hemophilic interaction with EDs of E-cadherins on other cells [296]. Cadherins first homo-dimerize (cis-dimerization) on one cell surface followed by trans-dimerization with other cadherins on the neighboring cells [295].

E-cadherin is arguably the most studied cell-cell adhesion protein and has been shown to be required for the formation and sustenance of epithelial linings. This was first demonstrated in chicken embryos by Gallin and colleagues and was originally named L-CAM (liver cell adhesion molecule) [297]. It localizes to the surface of epithelial cells at sites of cell to cell contact primarily at adherens junctions.

Table 7. Epithelial and mesenchymal markers. The Table represents a compilation of E/M markers based on their cellular function and/or localization. These markers were divided into four different categories: cytoskeletal proteins, extracellular matrix proteins, cell surface proteins, transcription factors and miRNA and lncRNA.

Category	Epithelial markers	Mesenchymal markers
Cytoskeletal proteins	Cytokeratin	Vimentin β-catenin α-SMA (α-smooth muscle actin) FSP1
Extracellular matrix proteins	Collagen IV (α1) Laminin 1	Collagen I (α1) Collagen III (α1) Fibronectin Laminin 5
Cell surface proteins	E-cadherin ZO-1 (Zonula occludens 1)	N-cadherin αVβ6 integrin α5β1 integrin Syndecan-1
Transcription factors	FOXA1/2 (Song et al 2010) GATA3 (Yan et al 2010) TP53 (Chao et al 2011)	Twist1 ZEB1/2 Snail1/2 Ets-1 Goosecoid LEF-1 CBF-A/KAP-1 complex
miRNA and lncRNA	miR-200s	miR10b miR-21

During type I EMT, E-cadherin levels are repressed accordingly during gastrulation, neurulation and organogenesis [298]. Downregulation, complete loss or mutations in E-cadherin have also been linked to malignant transformation and are known to interfere with the stability of adherens junctions. E-cadherin downregulation is achieved through various mechanisms including genetic alterations (mutations, loss of heterozygosity etc.), epigenetic changes through DNA methylation and transcriptional control [299][300]. *CDH1* Mutations have been identified in gastric, ovarian and breast cancers [301]. Loss of heterozygosity (LOH) of *CDH1* has been observed in various cancers including breast, prostate and liver [302]. LOH is a chromosomal event where a mutated allele results in the loss of the other allele (e.g. *RB1* and *BRCA1* mutations leading to loss of the wild-type allele). Transcriptional repression is however the most studied mechanism and has been implicated in EMT and EMT-like changes [281]. A series of zinc-finger-family of transcription factors such as *ZEB1/2*, Twist, Snail and Slug can bind the *CDH1* promoter and repress its transcription [303][304]. Collectively, the repression of E-cadherin correlated with loss of epithelial polarity [305], poor-differentiation [306], higher grade [307], enhanced metastatic potential [308] and ultimately worse patient prognosis [309].

1.7.1.4.2 N-cadherin

N-cadherin is encoded by the *CDH2* gene and is known as cadherin-2 or neural cadherin. The discovery of N-cadherin was a serendipitous incident during the examination of the effect of an anti-neutrophil monoclonal antibody NCD-1 on mouse embryonic brain cells. Cells treated with NCD-1 failed to form compact structures, which was later

attributed to the inhibition of N-cadherin leading to reduction in cell adhesion [310]. N-cadherin, like other cadherins, contain the five folded repeats capable of cis- and trans-dimerization.

N-cadherin is first detected during the gastrulation and neurulation stages where cells undergo EMT to form the mesoderm and neural crest respectively. To do so, these cells upregulate N-cadherin and downregulate E-cadherin, a process known as E- to N-cadherin switch [311]. The switch is important for the epiblast cells to ingress through the primitive streak and the neural crest away from the neural tube [312][313]. In adult cells, N-cadherin is crucial for maintaining the structural and adhesive properties of cells especially in neurons and during synapse formation [310].

Twist, a repressor of E-cadherin, can also activate N-cadherin expression [314]. E-cadherin expression is highly dependent on the availability of p120-catenin which serves to stabilize E-cadherin. The downregulation of E-cadherin by TGFβ1 or EMT transcription factors (e.g. Twist), frees up the p120-catenin, which then binds another cadherin, likely N-cadherin that is concomitantly upregulated. Cadherins compete for binding to catenins to mediate their stability [315]. Interestingly, forced expression of one cadherin can downregulate expression of other cadherins. For instance, forced expression of N-cadherin in epithelial cells downregulated E-cadherin by increasing its degradation [316]. N-and R-cadherin promote the endocytosis and subsequent degradation of E-cadherin via competition for p120-catenin binding [317].

Cellular behavior is also influenced by E- to N-cadherin switching. N-cadherin binds fibroblast growth factor receptor (FGFR) and serves as a stabilizer of the receptor on

the cell surface by promoting its dimerization [318][319]. N-cadherin also promotes cancer cell interactions with endothelial and mesenchymal cells (e.g. fibroblasts). The small scaffold protein NHERF links N-cadherin and β-catenin to the platelet-derived growth factor receptor (PDGFR) at sites of lamellipodia formation, which in turn increases motility [320]. N-cadherin expression increases steady-state levels of the Rho GTPases Rac1, RhoA and Cdc42 in the active GTP-bound form leading to enhanced cell motility [321][322].

1.7.1.5 EMT Signaling

Accumulating evidence has demonstrated that EMT is inducible by multiple factors (e.g. TGFβ, EGF, WNT, FGF, Notch, BMP). However, TGFβ remains the most potent inducer of EMT not only during embryogenesis and tissue fibrosis but also during cancer progression [279]. Physiologically, TGFβ is a crucial regulator of cell proliferation, differentiation, migration and apoptosis. Mutations in genes involved in the TGFβ pathway (e.g. TGFβ receptors) have been associated with cancer occurrence and TGFβ overexpression has been linked to highly metastatic tumors and poor patient prognosis [323]. It is worth noting that EMT is generally considered an anti-proliferative mechanism by which cancer cells temporarily sacrifice an increased growth capability for the benefit of acquiring motile, drug resistant and stem cell-like characteristics [324][325].

1.7.1.6 Canonical Smad TGFβ1 Signaling

TGFβ1 binds two types of transmembrane serine/threonine kinase receptors, designated as type I and type II TGFβ receptors (TβRI and TβRII). Binding of TGFβ to the Type II receptor results in receptor activation and auto-phosphorylation of both receptors,

which then phosphorylate Smad factors (Smad2 and 3). Phosphorylated Smad2/3 dissociate from the receptors and form a complex with Smad4 [326]. The Smad2/3/4 complex then translocates to the nucleus where Smads act as activators or repressors of transcription factors to modulate gene expression [327].

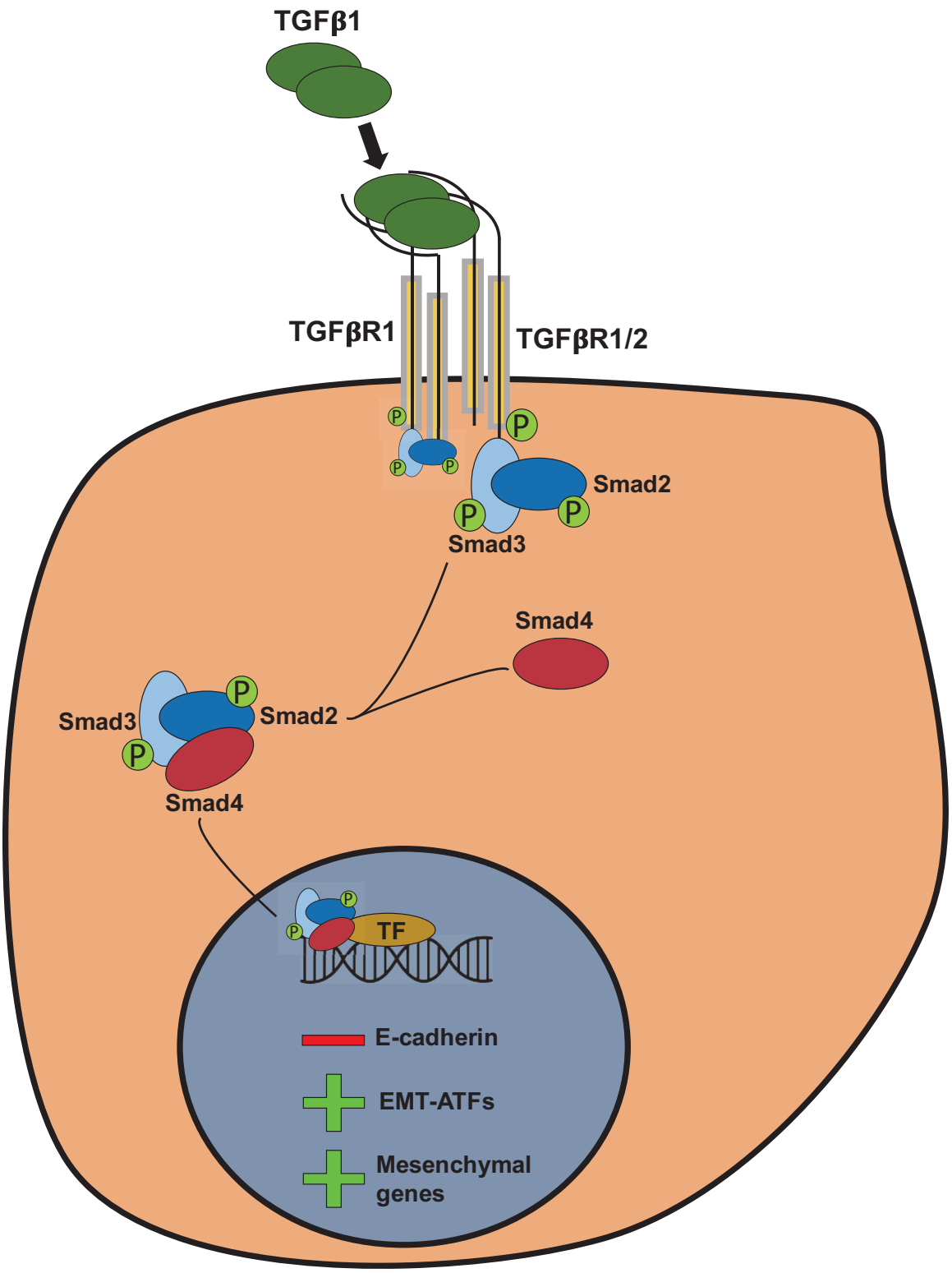
Smad proteins were originally described as part of EMT activation during tissue development [328]. Smads can exist as receptor-regulated (R-Smads) which consist of Smad2 and Smad3. Both Smads are direct phosphorylation substrates of TβRI and activin type I receptors [329][330]. Smads can additionally amplify the EMT response by increasing autocrine TGFβ production [331]. Smad4 is part of a class of Smads that are required for R-Smad signaling but are not direct substrates of TβRs. Smad4 association with Smad3 is also required for repression of E-cadherin and occludin in response to TGFβ1 [332] (figure 13). The expression of an inactive Smad4 or decreased expression of Smad4 also inhibited TGFβ1-indcued EMT in breast cancer cells [327]. Additionally, the Smad3-Smad4 complex interacts with *ZEB1*, *ZEB2* and *SNAI1* in response to TGFβ1 to further exacerbate EMT activation [333][332].

In addition to Smads, TGFβ receptors also directly phosphorylate other major families of EMT-ATFs (EMT-activating transcription factors), including the Snail family of zinc finger proteins. Snail is activated by TGFβ, Notch, COX2, EGF, Wnt, and other factors and can directly induce other EMT-ATFs. However, it also cooperates with Smad3/Smad4 to repress epithelial markers such as E-cadherin, desmoplakin, occludins, and cytokeratins. Concurrently, Snail stimulates expression of mesenchymal markers such as N-cadherin, vimentin and MMPs (reviewed in [333]). As a result, Snails play a pivotal

role in the dissociation of cell-to-cell attachments and subsequent invasiveness seen in cancer cells having undergone EMT.

Genetic changes in Smad-encoding genes have been implicated in cancer. For example, loss or mutation of the *SMAD4* gene on human chromosome 18q21.1 is found in more than 50% of pancreatic carcinomas [334] as well as in breast and ovarian cancers [335]. Studies on allelic *SMAD4* loss showed carcinoma development after 6-12 months in heterozygous *SMAD4* mice (*SMAD4* +/-). The second allele was subsequently lost at later stages of tumor progression, suggesting that loss of one allele is sufficient to promote tumor initiation while loss of function of both alleles (as seen in LOH) is important in subsequent progression of malignant tumors [336]. Allelic loss on chromosome 15q21-22, which harbors the *SMAD3* gene, is also common in breast, colorectal, and pancreatic tumors. LOH at the *SMAD3* locus was found in 73% of non-metastatic and 90% of metastatic breast carcinomas [337].

Figure 13. Canonical TGFβ1 signaling. TGFβ1 binds two types of transmembrane serine/threonine kinase receptors, designated as type I and type II TGFβ receptors (TβRI and TβRII). Binding of TGFβ to a Type II receptor results in receptor activation and phosphorylation of type I and type II receptors, which is in turn activated and further phosphorylates Smad factors (Smad2 and 3). Phosphorylated Smad2/3 then dissociate from the receptors and form a complex with Smad4. The complex then translocates to the nucleus where Smads bind to Smad-binding elements (SBEs) in DNA and act as activators or repressors of transcription factors. Smads can additionally amplify the EMT response by increasing autocrine TGFβ production [331]. The Smad complex serves to repress E-cadherin, activate expression of EMT activating transcription factors (EMT-ATFs) and induce expression of mesenchymal genes (e.g. N-cadherin).


TGFβ1
TGFβR1
TGFβR1/2
P
Smad2
Smad3
Smad4
Smad3
Smad2
Smad4
TF
E-cadherin
EMT-ATFs
Mesenchymal genes

1.7.1.7 Non-Smad TGFβ Signaling

TGFβ signaling through TβRI and TβRII also activates non-Smad pathways that include the MAP kinase, PI3K/Akt, p38/Jnk and Rho GTPase pathways. Those pathways are discussed below.

1.7.1.7.1 MAPK/Erk Pathway

The phosphorylated TβRs serve as docking sites for various proteins other than Smads such as proteins containing phosphotyrosine-binding domains and src homology domains (e.g. Grb2 (growth factor receptor binding protein 2)). Grb2 is normally complexed with another adaptor protein called Sos in the cytoplasm. The receptor phosphorylation recruits the Grb2/Sos1 complex where Sos activates Ras proteins by exchanging bound GDP for GTP. Active Ras binds Raf to activate a series of MAP (mitogen-activated protein) kinases leading to the activation of MEK1, which ultimately phosphorylates Erk [338]. Erk activation is required for the disintegration of cell junctions and cell motility as well as enhanced interaction with the ECM [339]. The earliest evidence of TGFβ-induced activation of MAPK was observed in rat intestinal cells where TGFβ treatment induced an increase in p21(Ras) levels, which is upstream of the MAPK pathway [340]. Later reports demonstrated that TGFβ activates the MAPK pathway through Raf and MEK1, which in turn promotes the phosphorylation of Erk in fibroblasts [341], epithelial cells [342] and cancer cells [343]. Knockdown of Grb2 or ShcA in normal breast epithelial cells and cancer cells renders these cells unresponsive to TGFβ with limited migratory and invasive capabilities [344].

1.7.1.7.2 PI3k/Akt Pathway

TGFβ also induces the activation of PI3k and the subsequent phosphorylation of Akt [345][346][347], independently of Smad signaling [348]. The association of the p85 subunit of PI3k with TβRI is the initiating event upon TGFβ treatment. In contrast, p85 also associates with TβRII but this association does not require TGFβ. Regardless, the phosphorylation of both receptors and their kinase capacity is the determinant of the activation of PI3k [349]. TGFβ also induces PI3k indirectly by activating the expression of TGFα, which in turn activates EGFR-mediated activation of PI3k [347]. PI3k has been demonstrated as necessary for the re-organization of the actin cytoskeleton as well as cell migration during TGFβ-induced EMT. mTOR (mammalian target of rapamycin), a downstream effector protein of PI3k signaling mediates the phosphorylation of 4E-BP1(Eukaryotic translation initiation factor 4E-binding protein 1) and S6K1 (Ribosomal protein S6 kinase beta-1) in response to TGFβ in NmuMG murine epithelial cells and HaCAT keratinocytes [350]. In contrast, TGFβ has also been shown to inhibit PI3k/Akt signaling through a Smad-dependent mechanism, via the expression of the lipid phosphatase SHIP1 which dephosphorylates Akt [351]. Whether PI3k-mediated and Smad-mediated activation of EMT act synergistically or antagonistically remains elusive and is context dependent (addressed in the discussion section).

1.7.1.7.3 P38/Jnk Pathway

The p38/Jnk pathway is one of most well-established non-Smad signaling pathways. p38 and Jnk phosphorylation is also mediated through MAP kinases specifically MMK3/6 and MMK4 respectively [352]. TAK1 (Transforming Growth Factor-Beta-

Activated Kinase 1) and TRAF6 (TNF Receptor Associated Factor 6) are two adaptor proteins that associate with the phosphorylated TβRs in response to TGFβ and serve as activators of MMK3/6 and MMK4. Early studies reported that TGFβ treatment of various cancer cells such pancreatic, colorectal, breast, fibrosarcoma and lung cancers induced activation of p38 and Jnk [343][353][354]. These effects were independent of Smad3 and Smad4 [354]. The activation of the p38/Jnk pathway is a known mechanism by which TGFβ suppress growth and induces apoptosis [355]. However, the p38/Jnk pathway also plays a role in mediating EMT-associated changes in the actin cytoskeleton and cell morphology in NmuMG cells [356][357].

1.7.1.7.4 Rho GTPase Pathway

The Rho family of GTPases consist of small 21kDa proteins that are a subfamily of the Ras superfamily. RhoA, Cdc42 and Rac1 are the most well-characterized Rho GTPases and they play essential roles in cytoskeletal rearrangement, organelle development, cell motility as well as other functions [358]. Rho GTPases also play a key role in TGFβ-induced EMT by dynamic regulation of the actin cytoskeleton, stress fiber formation and acquisition of the motile mesenchymal phenotype in epithelial cells [359]. Like the aforementioned non-Smad pathways, the Rho GTPase activation by TGFβ is independent of Smads [360]. However, RhoA activation by TGFβ can be delayed in situations where Smads are required for the transcriptional activation of NET1, a GEF essential for RhoA activation [359][361]. Cdc42 also interacts with phosphorylated TβRs, which then associates with Pak2 and other proteins including occludin. Occludin then localizes the TβRs/Cdc42/Pak2 to the tight junctions where Pak2 phosphorylates and inactivates cofillin

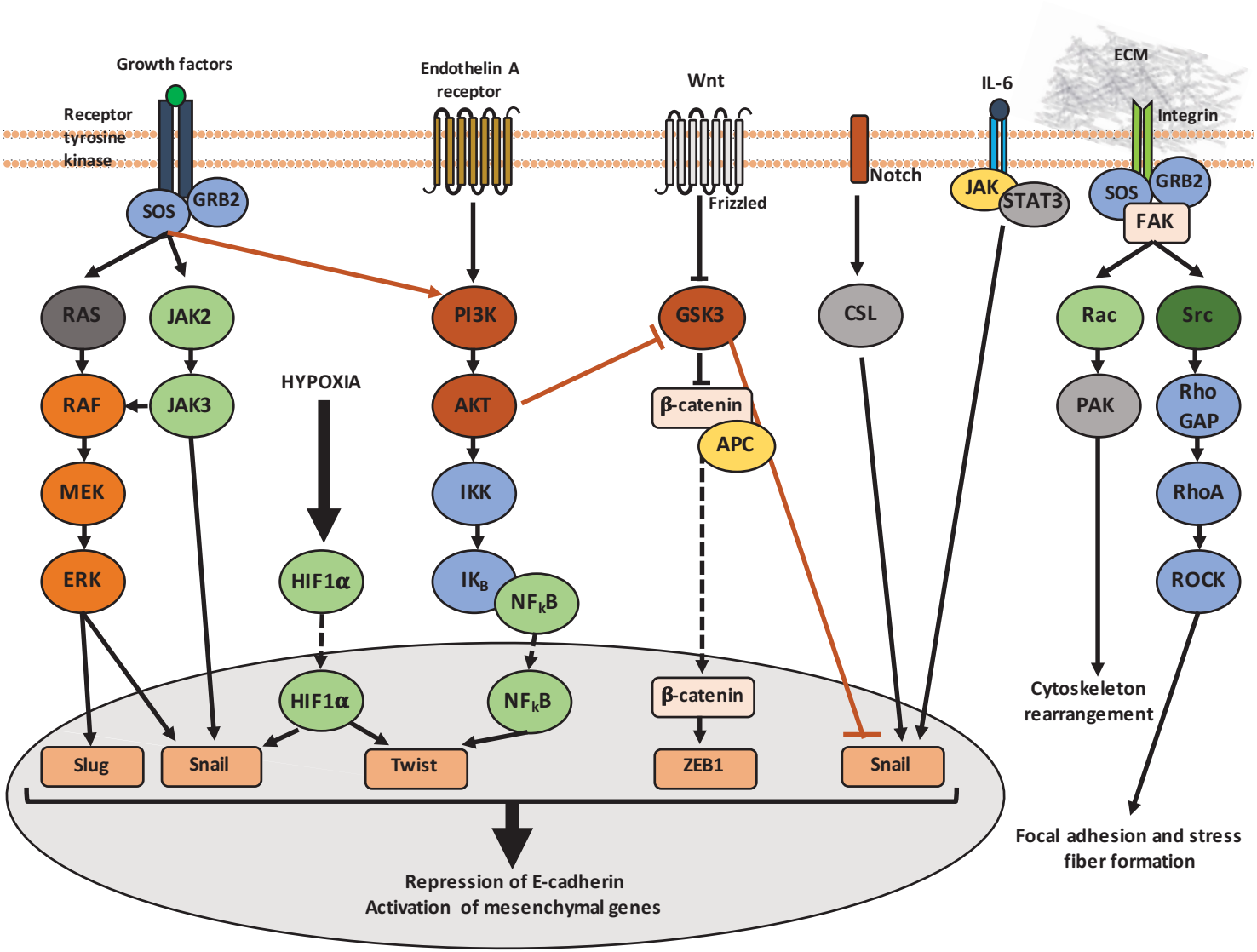
[362]. The inactivation of cofillin leads to increased actin polymerization and consequently promote tight junction dissolution [363][364]. Similarly, RhoA activates Rock that activates LIM kinase leading to the inhibition of cofillin [365].

In contrast, TGFβ can inhibit RhoA activation at tight junctions where phosphorylated TβRs phosphorylate a scaffold protein called Par6 which associates with Smurf1 to form a complex [366]. Complex formation promotes the ubiquitination of RhoA at tight junctions enabling the disintegration of these junctions during TGFβ-induced EMT [367].

1.7.1.8 Non-canonical EMT Pathways

EMT can also be induced through pathways independently of TGFβ and are illustrated in figure 14. Some of these pathways were described above as non-Smad TGFβ signaling pathways, however they can also be activated by other ligands. PI3K/Akt and Rho GTPase pathways are activated in response to growth factors (e.g. IGF-1, VEGF) binding to receptor tyrosine kinases [368], endothelin receptor activation [369] and interaction with matrix-bound integrins [370]. Other pathways involve IL-6, which induces EMT via the activation of Snail through the JAK/STAT3 pathway in human intrahepatic biliary epithelial cells (HIBEC) [371], MCF-7 breast cancer cells [372] and various non-small cell lung cancer cell lines [373]. HIF1α activation due to hypoxia stimulates Snail [374][375] and Twist [376] expression to promote metastasis. Wnt also activates an EMT program via binding its receptor Frizzled, which alleviates the inhibition of β-catenin hence allowing its nuclear translocation and activation of Wnt target genes [377][378].

Figure 14. Non-canonical EMT pathways. The figure represents the non-canonical signaling pathways that activate an EMT program which culminates in the repression of E-cadherin, activation of mesenchymal genes, changes in focal adhesion, stress fiber formation and cytoskeletal rearrangement. These pathways include the receptor tyrosine kinase-activated Ras/Raf/MEK/Erk, JAK2/JAK3 and PI3K/Akt/GSK3 pathways. Endothelin A receptor also activates Nf_KB pathway and promotes its nuclear localization via the PI3K/Akt arm. In addition, hypoxic conditions activate HIF1**α** which in turn translocates to the nucleus. Similarly, **β**-catenin dissociates from its inhibitor APC in response to the activation of the Frizzled receptor. The Notch pathway also activates EMT via CSL-mediated activation of Snail. Integrin interaction with the extracellular matrix recruits FAK (focal adhesion kinase) which in turn activates the Rac/PAK and Src/RhoA pathways.


Growth factors
Receptor tyrosine kinase
SOS
GRB2
RAS
JAK2
RAF
JAK3
MEK
ERK
HYPOXIA
HIF1α
Endothelin A receptor
PI3K
AKT
IKK
IKB
NFkB
Wnt
Frizzled
GSK3
β-catenin
APC
β-catenin
ZEB1
Notch
CSL
IL-6
JAK
STAT3
ECM
Integrin
SOS
GRB2
FAK
Rac
Src
PAK
Rho GAP
RhoA
ROCK
Cytoskeleton rearrangement
Focal adhesion and stress fiber formation
Slug
Snail
HIF1α
Twist
NFkB
Snail
Repression of E-cadherin
Activation of mesenchymal genes

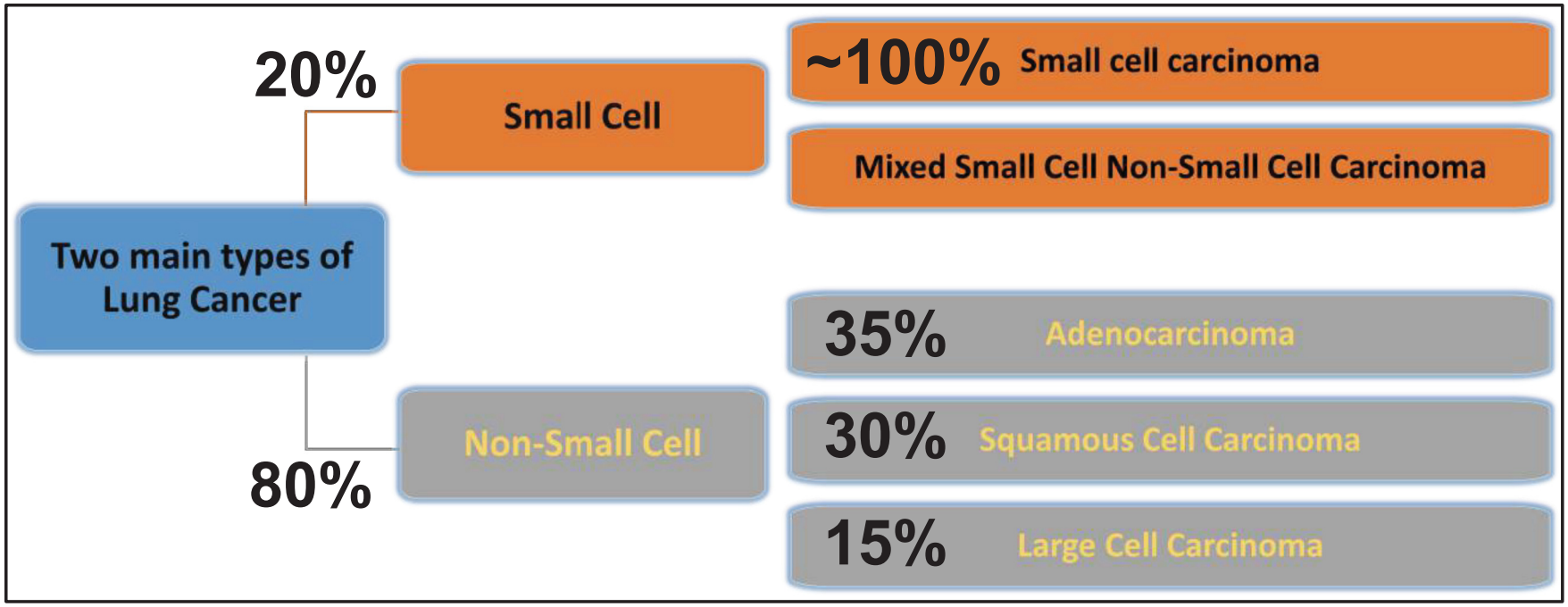
1.8 Sub-chapter 8: Lung Cancer

Lung cancer is the leading cause of cancer-related deaths in North America and worldwide [379]. Lung cancer is also the most prevalent cancer worldwide with almost two million new cases every year [380]. Despite recent advances in lung cancer screening, lung cancers are often diagnosed at advanced stages at which point patient prognosis is not favorable. Non-small-cell lung cancers (NSCLC, 85%) and small-cell lung cancers (SCLC, 15%) comprise the two most common types of lung cancer (figure 15).

1.8.1 Initiation and Clonal Evolution

The cell of origin that gives rise to lung cancer remains largely unknown. It is however accepted that SCLC primarily originates from neuroendocrine cells in the proximal airway (e.g. bronchus) while NSCLC arises from more distal regions of the lung (e.g. bronchioles and alveoli) [381]. At the cellular level, pre-neoplastic lesion will undergo hyperplasia (or dysplasia or metaplasia) of which very few lesions will progress into an invasive carcinoma. More importantly and unlike pancreatic cancers (discussed in subchapter 9), the sequence of genetic alterations that give rise to a lung tumor has not been fully understood. Numerous reports documented that many of these alterations affecting both oncogenes and tumor suppressors are present in pre-malignant stages before tumors become clinically-detectable [382].

Figure 15. Lung cancer histological classifications. Small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) compromise the most common types of lung cancer. The image shows the prevalence of each histological subtype within lung cancer patients.


Two main types of Lung Cancer
20%
Small Cell
~100% Small cell carcinoma
Mixed Small Cell Non-Small Cell Carcinoma
80%
Non-Small Cell
35% Adenocarcinoma
30% Squamous Cell Carcinoma
15% Large Cell Carcinoma

1.8.2 Environmental Risk Factors

Smoking remains the leading risk factor for the development of lung cancer representing a 20-fold increase in risk compared to non-smokers [383]. However, the probability of smoking affecting one's risk of lung cancer varies from one individual to another. Smoking is strongly associated with SCLC and squamous cell carcinoma (one subtype of NSCLC) (discussed in 1.8.5). Other environmental factors include exposure to radiation, asbestos and radon gas [384]. Lung cancer etiology is a multi-faceted field where multiple genetic and environmental factors cooperate to enhance the risk of cancer development. Although many environmental factors have been identified, assessing new factors is subject to the same assessment criteria that involves derivation of its relative risk (RR). The latter is a comparison between the influence of exposure to a factor in at-risk population (i.e. lung cancer patients) and the influence of the same factor on the general population. The interaction of multiple factors can be calculated based on the combinatorial population attributable risk. For instance, smoking is linked to around 90% of lung cancer cases, however up to 15% of that is "attributed" to exposure to workplace factors such as radon and air pollution [385].

1.8.3 Germline Genetic Factors

Etiological and epidemiological studies demonstrated strong association between family history and risk of lung cancer. Germline mutations in *TP53* [386], *RB1* [387] and *EGFR* [388] have been reported to increase cancer incidence including that of lung cancer. Bailey-Wilson *et al*. identified a major lung cancer susceptibility locus at 6p23-25 where a genetic linkage of a series of genes including tumor suppressor genes (*IGF2R*, S*ASH1*,

PARK2, *LATS1*) was identified [389]. Single-nucleotide polymorphisms (SNPs) at 15q24-15q25.1 were also associated with increased risk of familial lung cancer [390]. The region contains two genes encoding two subunits of the *CHRNA7* gene (cholinergic receptor nicotinic alpha 7), a gene that encodes a receptor bound by nicotine [391].

1.8.4 Somatic Genetic Factors

Mutations in *KRAS* and *EGFR* are early events in lung carcinogenesis (discussed in 1.8.6). Altered expression of genes involved in DNA repair and pro-inflammatory pathways has been observed in both patients with lung cancer and heavy smokers with no clinically-detectable tumors [392]. Genetic alterations have also been detected in histologically "normal" tissues surrounding lung tumors. These alterations include *EGFR* (epidermal growth factor receptor) amplification [393] and *TP53* alterations (mutation, LOH or hyper-methylation) [394], c-myc amplifications [395] and microsatellite instability [396]. Collectively, these reports render determination of surgical margins difficult since such tissue may appear normal to the surgeons and pathologists.

In NSCLC, a key genomic event is the loss of heterozygosity in putative tumor suppressor genes at the following loci: 3p21 (contains the *RASSF1A* (Ras association domain family member 1) and *FUS1* (FUS RNA binding protein), 9p21 (*P16INK4A*), 17p13 (*TP53*) and 3p14 (*FHIT*, fragile histidine triad) (table 8)[397]. Hyper-methylation of tumor suppressors in stage I lung cancer has also been reported in the gene promoters of *FHIT*, *MGMT* (6-methylguanine DNA methyltransferase), *P16INK4A*, *RASSF1A* and *DAPL* (death-associated protein kinase) [398][394][399]. In fact, the co-hyper-methylation of *P16INK4A* and *FHIT* is predictive of tumor recurrence in surgically resected stage I

adenocarcinoma patients [398]. Table 8 summarizes and compares the main genetic alterations in NSCLC and SCLC and their respective prevalence (table 8).

Table 8. Prevalence of common lung cancer genetic alterations in SCLC vs. NSCLC. The table summarizes well-established mutations, deletions, amplifications, fusions or overexpression of driver and passenger genes.

Genetic alteration	SCLC	NSCLC
Chromosomal deletions	3q, 5p, 8q, 19q	3p, 4p, 4q, 5q, 8p, 10q, 13q, 17p
Chromosomal gains	1p, 1q, 3q,5p, 7p, 7q, 8q, 11q, 12q	3p,6q, 8p, 9p, 9q, 13q, 17p, 18q, 19q, 21q, 22q
ERBB1 (EGFR) overexpression	60%	rare
KRAS mutation	rare	40%
FHIT deletion or mutation	80%	40%
TP53 inactivating mutation or deletion	85%	50
Rb deletion or inactivating mutation	90%	15-30%
P16INK4A inactivating mutation or deletion	rare	70%
ERBB2 (HER2/Neu) overexpression	rare	20%
BCL-2 overexpression	75-95%	10-35%
Myc amplification	15-30%	5-10%
RET fusion	rare	1-2% (RET-KIF5B fusion)
ALK fusion	none	7%
ROS1 fusion	none	2%
MET amplification	none	11%
MET activating mutation	12.5%	3%
PIK3CA mutation	none	1-5%
PTEN mutation or reduced expression	10%	74%
FGFR1 amplification	rare	3-21%

1.8.5 Histological Subtypes

1.8.5.1 Small Cell Lung Cancer (SCLC)

Small cell lung cancer (SCLC) is histologically characterized by small cells with a relatively high nucleus to cytoplasm ratio and a high proliferative index. Almost 90% of SCLCs are neuroendocrine in nature and express neuroendocrine markers. SCLC cells contain neurosecretory granules and produce hormones, a hallmark of neuroendocrine differentiation. SCLC usually arises peri-bronchially where it infiltrates the underlying mucosa. Some SCLCs contain tumor regions that are representative of non-small cell lung cancers and may often be diagnosed as "mixed" if the representation is not equivocal [400] (figure 15).

1.8.5.2 Non-small Cell Lung Cancer (NSCLC)

NSCLC consists of three distinct subtypes: adenocarcinoma, squamous cell carcinoma and large cell carcinoma (figure 15).

1.8.5.2.1 Adenocarcinoma

Lung adenocarcinoma consists of four histological subtypes namely solid, acinar, papillary and bronchioloalveolar. All adenocarcinoma subtypes display the typical glandular structures except for solid and bronchioloalveolar adenocarcinomas, which may not display any glandular structures [401]. Bronchioloalveolar adenocarcinomas are also further classified into mucinous and non-mucinous measured by the periodic acid Schiff histochemical staining of cytoplasmic mucin [402].

Diagnosis of adenocarcinoma is usually achieved using hematoxylin and eosin staining (H and E), which can be further confirmed using immunohistochemistry for specific markers. TIF-1 (thyroid transcription factor 1) is highly expressed in NSCLC compared to other lung cancer histological subtypes and it is commonly used as a marker for this cancer. TIF-1 is also a marker for colorectal and thyroid cancers [403]. Other diagnostic markers have been developed such as cytokeratins (CK) 5 and 6, mucin and p63 [404][405]. Adenocarcinomas are positive for TIF1 and CK7 staining and negative for CK5/6 and p63 [404].

1.8.5.2.2 Squamous Cell Carcinoma

Well-differentiated squamous cell carcinoma has a very distinct histological appearance characterized by squamous-like differentiation, intercellular bridges and keratinization. However, poorly-differentiated squamous cell carcinomas present a diagnostic problem that prompts the use of IHC markers. Equivocal NSCLC lung tumors are considered squamous cell carcinoma if they are positive for CK5/6 and p63 and negative for TIF1 and CK7 [404].

1.8.5.2.3 Large Cell Carcinoma

Large cell carcinomas represent about 10% of lung tumors and were initially labelled as undifferentiated tumors with no resemblance to any of the other subtypes. However, the 2015 WHO classification of lung tumors implied that large cell carcinoma consists of several histological variants (e.g. basoloid carcinoma, lymphoepithelioma-like carcinoma, clear cell carcinoma and large cell neuroendocrine carcinoma (LCNEC)).

These variants may also resemble, at diagnosis, solid adenocarcinomas or non-keratinizing squamous cell carcinoma based on positive staining of the markers TTF-1 or p40 respectively [406].

1.8.6 Driver and Passenger Alterations in Lung Cancer

A standard definition of a driver alteration is an alteration that offers a selective advantage to a tumor or a tumor cell population. In contrast, a passenger alteration is an alteration that is passed on to daughter cells due to the mere co-occurrence of a driver alteration. Noteworthy, a common misconception is that passenger alterations offer no selective advantage to a tumor. In fact, these passenger alterations are part of a myriad of perturbations (e.g. DNA repair breach) that cause both driver and passenger alterations to occur.

1.8.6.1 Driver Genetic Alterations

1.8.6.1.1 EGFR

Activating mutations in *EGFR* are driver events in the development of NSCLC. These mutations are mainly present in adenocarcinoma patients and are less common in squamous cell carcinoma and large cell carcinoma. *EGFR* mutations are independent of smoking history in adenocarcinoma patients [407]. Interestingly, *EGFR* mutations are more prevalent in Asian (40%) populations compared to Caucasian population [408]. The presence of an *EGFR* mutation sensitizes NSCLC patients to tyrosine kinase inhibitors (TKIs). Multiple prospective phase III trials showed that tumors in never-treated patients with an *EGFR* mutation are highly sensitive to treatment with the reversible TKIs erlotinib

and gefitinib and have increased progression-free survival compared to patients treated with the standard platinum-based chemotherapy [409][410][411]. Erlotinib is approved by the FDA as the first line of defense for NSCLC patients that tested positive for the Cobas® *EGFR* mutation test. If patients have received platinum chemotherapy, they are placed on gefitinib monotherapy [412]. Second generation irreversible TKIs such as dacomitinib and afatinib have also been approved. Dacomitinib is primarily used in *EGFR*-mutated patient tumors with the T790M substitution and that are resistant to erlotinib or gefitinib [413][414][415]. In a recent phase II clinical trial, dacomitinib improved progression-free survival of *EGFR*-mutated NSCLC patients compared to those treated with erlotinib [416][417].

1.8.6.1.2 KRAS

KRAS is a 21 kDa small-GTPase that cycles between an active GTP-bound form and inactive GDP-bound form. Activating mutations in *KRAS* leads to a constitutively active protein that triggers pro-growth, anti-apoptotic and migratory signals (figure 15). In lung cancer, activating mutations in *KRAS* occur predominantly in adenocarcinomas (up to 40% of NSCLC) [418], to a much lesser frequency in squamous cell carcinoma [419] and almost never in SCLC. Interestingly, *KRAS* mutations occur at higher frequency in adenocarcinoma patients that are smokers compared to those that are non-smokers.[407]. Mutations in codons 12, 13 and 61 have been identified with codon 12 being the most common. In addition, mutations that involve G/T and G/C transversions have been associated with tobacco exposure [420]. A large proportion (40%) of *KRAS* mutations with the G12C amino substitution resulted from G/T transversions [421].

Therapeutic targeting of the KRAS protein was proven to be an arduous task primarily due to the high affinity of GTP to the binding pocket on RAS proteins, which renders it difficult to design a small-molecule inhibitor with a high competitive binding capacity [422][423]. Alternatively, the *KRAS* status in lung cancer has been used as a prognostic or a response to treatment marker. For instance, the G12D form is associated with better long-term outcome than the G12R and G12V forms [424]. The good prognosis is not necessarily applicable in other cancer models as seen in colorectal cancer where the G12D form is predictive of increased resistance to chemotherapy [425]. NSCLC patients with a wild-type *KRAS* are more sensitive to cisplatin and vinorelbine compared to those with mutated *KRAS*. Interestingly, *KRAS* mutations are also associated with increased resistance to EGFR inhibition in NSCLC [426]. This may seem counterintuitive since EGFR is upstream of KRAS and its inhibition should not affect the activity of mutant KRAS. However, Eberhard *et al*. demonstrated that combinatorial treatment of *KRAS*-mutated NSCLCs with chemotherapy and the EGFR inhibitor erlotinib results in shorter overall and progression-free survival [426]. These observations added to the complex nature of KRAS signaling in lung cancer and the importance of assessing the *KRAS* status in NSCLC patients as standard diagnostic test.

1.8.6.1.3 ALK

ALK is a transmembrane tyrosine-kinase receptor that is highly expressed in the brain, testes and small intestine but not in lungs. In 2007, Soda *et al*. documented, for the first time, that the activation of ALK signaling in lung tumors is mediated by an oncogenic fusion event between *ALK* and the microtubule-associated protein *EML4* [427]. The *ALK*-

EML4 fusion occurs in about 7% of NSCLC patients, which are primarily adenocarcinoma cases with no smoking history [428][429]. Subsequently, ALK inhibitors such as crizotinib were FDA-approved in *ALK*-positive patients as a first line of defense when tested using fluorescent *in situ* hybridization (FISH) [430] or IHC [431] methods. The same trial then tested crizotinib as a second line of defense in *ALK*-positive patients and demonstrated an improved progression-free survival compared to patients who received chemotherapy alone [432].

1.8.6.1.4 RET

RET is another tyrosine-kinase receptor encoded by a proto-oncogene, which when altered can introduce cellular changes in growth, proliferation, migration, invasion and differentiation. RET activation is mediated via fusions with other genes such as *KIF5B* as seen in 2% of adenocarcinoma patients who tend to be young and have no smoking history or family history of lung cancer [433]. *CCDC6* has also been identified as a fusion target of RET [434]. Additionally, a recent phase II clinical trial revealed that NSCLC tumors with RET fusions are sensitive to the TKI cabozantinib [435][436]. RET fusions are often mutually exclusive with other driver events affecting *EGFR*, *KRAS* or *ALK* [437].

1.8.6.1.5 ROS1

ROS1 is also a tyrosine-kinase receptor which is part of the insulin receptor family of proteins. Like *RET* and *ALK*, *ROS1* undergoes fusion events that drive the progression of NSCLC. *ROS1* fusions have been reported with multiple other genes such as *FIG*, *SLC34A2* and *CD74* [438][439]. *ROS1* fusions are prevalent in about 2% of NSCLC with

the majority occurring in adenocarcinoma patients who have no smoking history [438]. *RET* fusions are also mutually exclusive with other driver events including *EGFR*, *ALK* or *KRAS* [437]. The TKI crizotinib demonstrated great efficacy in *ROS1*-positve NSCLC patients in a recent phase I clinical trial [440].

1.8.6.2 Passenger Events

1.8.6.2.1 MET

MET encodes a transmembrane tyrosine-kinase receptor that is activated by the hepatocyte growth factor (HGF) [441]. MET amplifications have been reported in up to 11% of NSCLC patients who present with a high proliferative index and is also predictive of poor patient outcome [442][443]. *MET* amplifications are present in 20% of *EGFR*-positive NSCLC patients and are linked to *MET*-mediated resistance of TKIs [444]. Small molecule MET inhibitors (cabozantinib and tivantinib) [445], humanized monoclonal antibody (Onartuzumab) [446] and specific TKIs (crizotinib) [438] have been developed to target MET-amplified NSCLC tumors. The monoclonal antibody onartuzumab combined with the TKI erlotinib improved overall survival and progression-free survival compared to erlotinib alone in *MET*-amplified NSCLC [447].

1.8.6.2.2 PIK3CA

PIK3CA encodes the 110 kDa catalytic subunit of the PI3K protein. The catalytic subunit utilizes ATP to phosphorylate phosphatidylinositols. Mutations in *PIK3CA* have been reported in about 5% of NSCLC patients [448] and usually co-occur with other genetic alterations such as *KRAS* mutations, *EGFR* mutations or *ALK* fusion [449]. Many

PI3K inhibitors have been developed, however their clinical success has been only noticeable when in combination with chemotherapy [450].

1.8.6.2.3 PTEN

The tumor suppressor gene *PTEN* encodes a phosphatase that dephosphorylates phosphoinositide substrates (e.g. phosphatidylinositol-3,4,5-trisphosphate or PIP_3) and hence acts as a negative regulator of PI3K signaling [451]. Inactivating mutations or deletions in *PTEN* have been predominantly reported in squamous cell carcinoma [452]. *PTEN* loss in NSCLC has been associated with increased resistance to TKIs in *EGFR*-positive tumors. Additionally, the TKI vandertanib has shown promising results in targeting *EGFR*-positive NSCLC tumors with *PTEN* deficiency [453]. A recent meta-analysis demonstrated that *PTEN* mutations were differentially linked to ethnic backgrounds where it was found in 10% of squamous cell carcinoma and 2% of adenocarcinoma NSCLC patients of Asian ethnicity (China, Japan, Taiwan). In contrast, *PTEN* mutations were found in 6% of adenocarcinoma and none in squamous cell carcinoma patients from other populations (i.e. Europe, North America and Australia) [454].

1.8.6.2.4 FGFR1

FGFR1 encodes the fibroblast growth factor receptor 1 which is a receptor tyrosine kinase that signals through various pathways including RAS/Erk, PI3K/Akt and PKC (protein kinase C) [455]. *FGFR1* amplifications have been reported in 25% of squamous cell carcinoma, 25% of large cell carcinoma and less than 3% of adenocarcinoma [456].

Although *FGFR1* amplifications exert oncogenic effects on cell proliferation and angiogenesis, the presence of this alteration has shown no or little correlation with overall survival and progression-free survival of NSCLC patients [457][458]. Several inhibitors of FGFR1 have been tested and the specific FGFR inhibitors ponatinib [459] and AZD4547 [460] demonstrated the most promising results in *FGFR1*-positive NSCLC patients.

1.9 Sub-chapter 9: Pancreatic Cancer

1.9.1 Epidemiology and Clinical Presentation

Pancreatic ductal adenocarcinoma (PDAC), the predominant form of pancreatic cancer (≈95%), is a fatal cancer with a five-year survival rate of 4% [461]. In 2016, over 53,000 individuals were diagnosed with PDAC in North America most of who will succumb to their disease in 5 years [462] largely due to metastases to the liver, lungs and/or peritoneal cavity [463]. Due to early dissemination of pancreatic tumor cells and late manifestation of symptoms, 92% of the patients are diagnosed with locally advanced or metastatic disease [464]. In this late stage, surgery is rarely curable and often not recommended to avoid post-operative complications. As a result, only 10-15% are considered eligible for curative surgery [465] and will receive adjuvant chemotherapy with or without radiation, which results in a 15-30% chance of surviving to five years [466][467]. The gold standard for predicting PDAC patient outcome is TNM staging that performs adequately in late stage (stage III and IV) patients where tumors are usually not resectable. However, the prognostic performance of TNM staging is below par in early stage (stage I and II) resectable patients [468]. The consequence of this poor performance is a tendency to undertreat patients who have a high risk of recurrent disease and over-treating patients who are at low risk of recurrence.

1.9.2 PDAC Progression

The ***histological progression*** of PDAC had been adequately characterized by pathologists from the neoplastic transformation of normal ductal epithelial cells to the

advancement through pancreatic intraepithelial neoplasia (PanINs) and culminating in ductal carcinoma. The histological changes associated with such progression are summarized in figure 16. The development of pancreatic cancer, or any cancer, could also be described by ***clonal evolution*** of cells. This has been made possible through the advent of next-generation sequencing platforms [469], which not only decoded the evolutionary path of PDAC but also unraveled remarkable heterogeneity between patients (inter-patient heterogeneity) and within the same tumor (intra-tumoral heterogeneity) [470]. In a seminal review, Alvin Makohon-Moore and Christine Iacobuzio-Donahue divided the progression of PDAC into three major steps: initiation, clonal expansion and exposure to foreign microenvironments (i.e. stroma, metastases site and immune system) [471].

1.9.2.1 Initiation

Darwinian evolution dictates that normal cells will acquire random mutations after which positive selection can occur [472]. A particular mutation must not hinder cell division in order for the mutation to be passed on to daughter cells (figure 17). Bozic *et al*. reported that the average number of somatic mutations is around three single nucleotide variants (SNVs) for every cell division [473]. However, given that pancreatic tissue in an adult organism has a low proliferative or regenerative capacity [474], the probability of driver mutations (discussed in 1.9.3) is extremely low. Yachida *et al*. described the genetic evolutionary landscape of patients who develop non-familial PDAC. The report predicted that the first driver mutation in a normal pancreatic cell probably occurred at least two decades prior to diagnosis [470]. This is further supported by the fact that having familial genetic variants (addressed next) that increase the risk of development of PDAC will only

lower the onset of disease by 5 years compared to non-familial PDAC. This can be compared to breast cancer where a high-risk variant (e.g. *BRCA1* mutation) will decrease onset for up to 20 years compared to non-familial breast cancer partly due the highly proliferative and hormone-sensitive nature of breast tissue [475]. In addition, although the number of cell divisions by pancreatic stem cells predicted the overall risk of developing pancreatic cancer [476], the low number of divisions suggested that extrinsic factors may play a more significant role in PDAC development. In fact, Wu *et al.* concluded that extrinsic factors such as carcinogens and radiation are more influential than intrinsic genetic factors (e.g. errors in DNA replication) [477].

Figure 16. PDAC progression timeline. The figure illustrates the histological, genetic and molecular alterations that occur during PDAC progression. The genetic alterations are classified into activating alterations that pertain oncogenic roles and inactivating alterations that suppresses the anti-tumor signals. Telomere shortening is believed to be one of the earliest events that predispose cells to become immortalized prior to any irreversible genetic events [478].

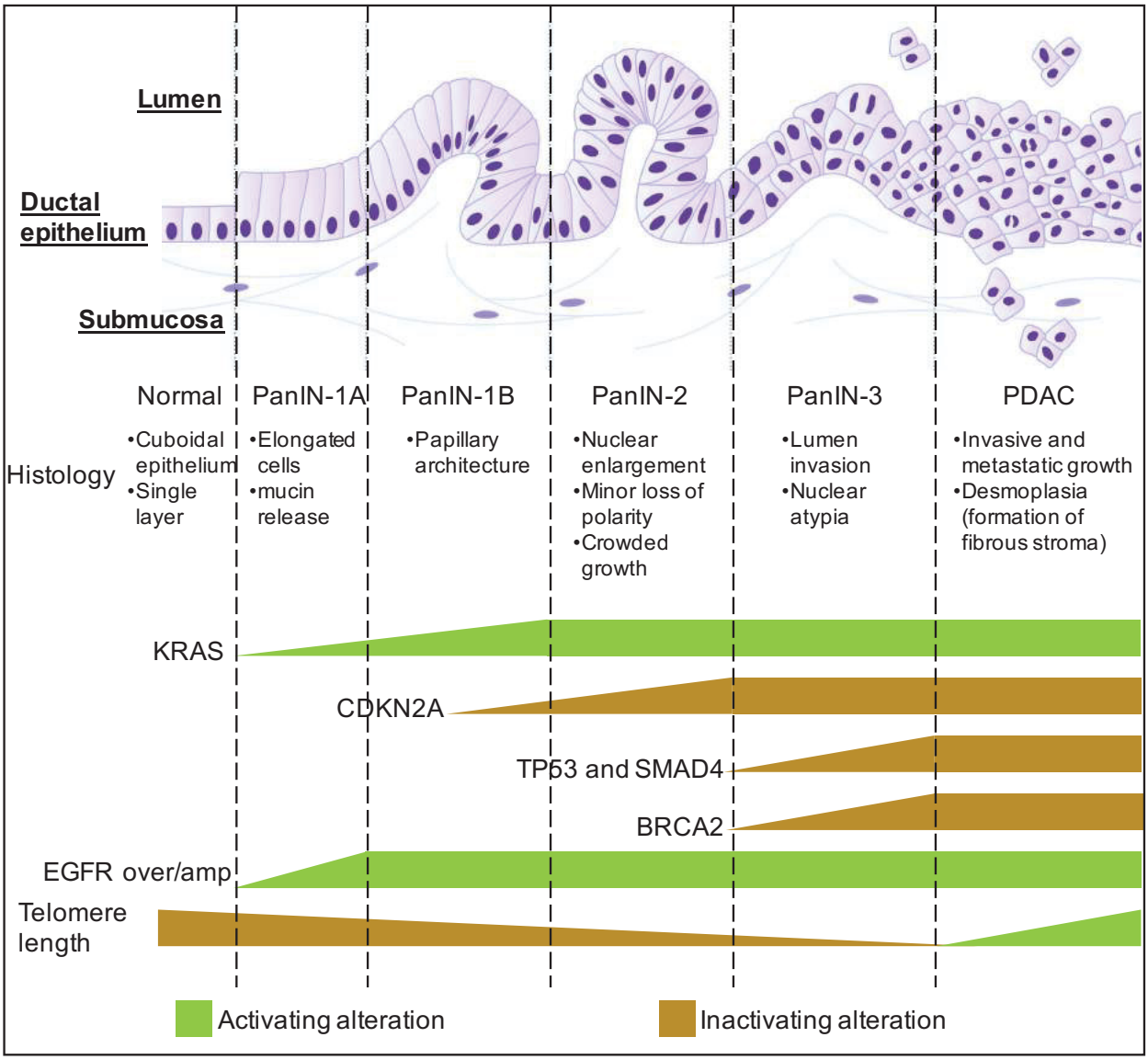
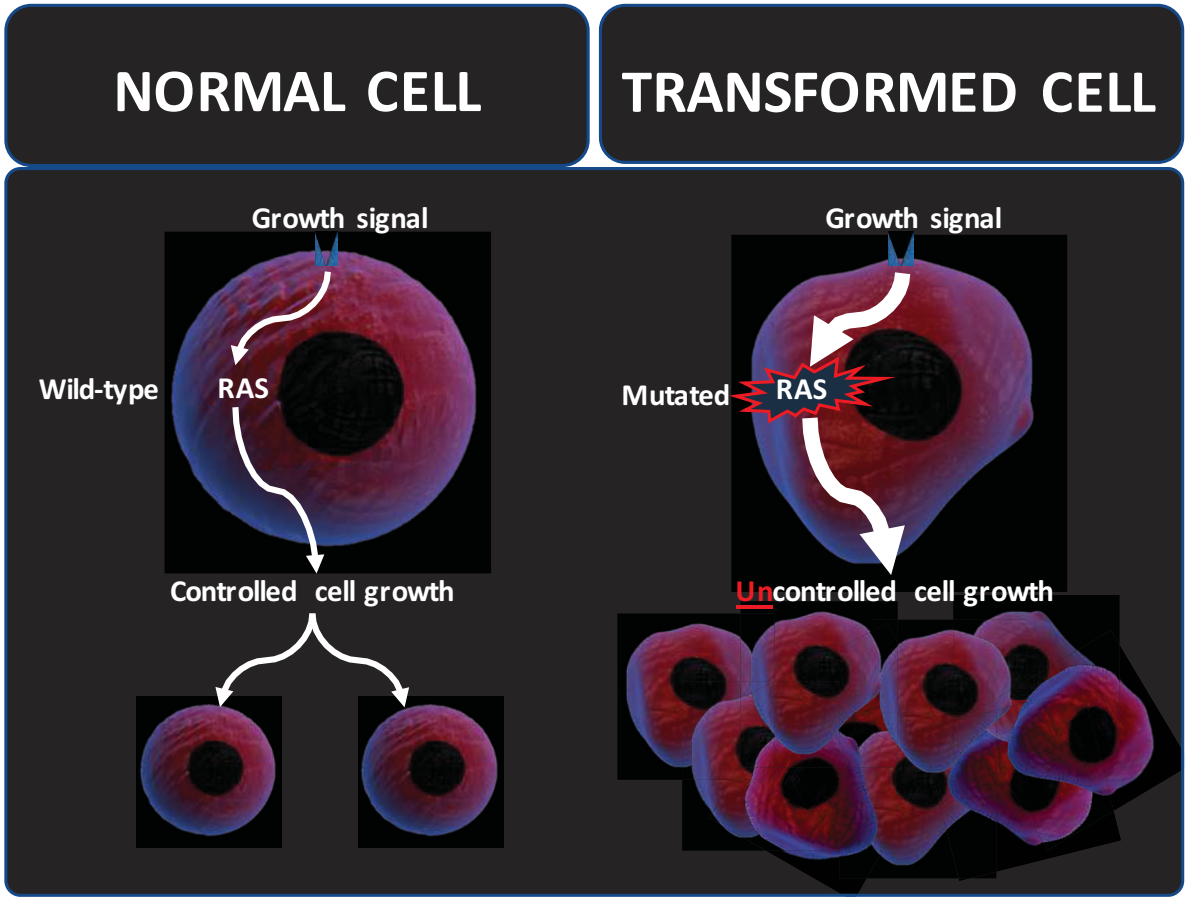

Lumen
Ductal epithelium
Submucosa
Normal
PanIN-1A
PanIN-1B
PanIN-2
PanIN-3
PDAC
Histology
•Cuboidal epithelium
•Single layer
•Elongated cells
•mucin release
•Papillary architecture
•Nuclear enlargement
•Minor loss of polarity
•Crowded growth
•Lumen invasion
•Nuclear atypia
•Invasive and metastatic growth
•Desmoplasia (formation of fibrous stroma)
KRAS
CDKN2A
TP53 and SMAD4
BRCA2
EGFR over/amp
Telomere length
Activating alteration
Inactivating alteration

Figure 17. Impact of KRAS mutation on cell growth. Wild-type KRAS is activated by adequate levels of growth factors in normal cells which initiates controlled cell growth. In contrast, an activating mutation in KRAS triggers uncontrolled cell growth and the formation of a tumor mass.


NORMAL CELL
TRANSFORMED CELL
Growth signal
Wild-type
RAS
Controlled cell growth
Growth signal
Mutated
RAS
Uncontrolled cell growth

Unlike most solid tumors, the initiating genetic event in PDAC is well-characterized. Activating mutations in *KRAS* are found in over 90% of PDAC patients and is the earliest genetic breach in the low-grade pre-cancerous lesions PanINs (intraepithelial neoplasia) [479](figure 16). In contrast, familial pancreatic cancer is characterized by multiple genetic variants that appear to cooperate and increase the lifetime risk of developing the disease. The most characterized genetic germline variants are mutations in *BRCA1*, *BRCA2*, *ATM*, *SMC2*, *FANCC*, *FANCG*, *CDKN2A* [480], *TERT* [475], *NR5A2* [481] and *ZNRF3* [482]. Germline *BRCA1*, *BRCA2* and *ATM* gene mutations promote genomic instability during DNA repair and increase the incidence of somatic mutations in genes like *KRAS* [483]. *CDKN2A* germline mutations result in perturbation of the G1/S cell cycle checkpoint leading to uncontrolled cell growth [484]. These germline mutations, particularly those related to DNA repair, will increase the rate by which somatic mutations will occur including the driver gene mutations.

Mutations in the serine protease *PRSS1* and the serine peptidase *SPINK1* have been also linked to the development of pancreatitis in patients, which also increases the lifetime risk of developing pancreatic cancer. This can be partly attributed to the inflammatory response (e.g ROS, reactive oxygen species) during pancreatitis and to the aberrant increase in cellular proliferation due to tissue damage [485]. ROS production and aberrant cellular division will increase the rate of somatic mutation occurrence. Other factors such as obesity, smoking and type II diabetes have been linked to increasing risk of developing pancreatic cancer [471].

1.9.2.2 Clonal Expansion

Acquisition of the initiating *KRAS* mutation does not always warrant the development of pancreatic cancer. In fact, almost one third of pancreatic tissue from routine autopsies display the pancreatic precursor lesions PanINs supporting the concept that not all precancerous lesions will develop into an invasive ductal carcinoma [486]. Almost all stage I PanINs lesions have *KRAS* mutations. However, the percentage of cells that have the mutation within one PanIN lesion depends on the grade, with high grade PanINs containing a higher percentage of *KRAS*-mutant cells. Subsequent somatic mutations are then acquired either gradually (i.e. linear progression) or accelerated (i.e. punctuated progression).

1.9.2.2.1 Linear Progression Model

The linear progression model suggests that the predominant clone will likely contain most of the genetic variants as cells acquire new alterations during their progression from early- to late-stage PanINs. As a result, driver genetic alterations in *KRAS*, *TP53*, *CDKN2A* and *SMAD4* become more frequent in higher stage PanINs [487]. For instance, *CDKN2A* and *SMAD4* losses are considerably higher in PanIN-3 than PanIN-2 [479][488]. *TP53* mutations that lead to its accumulation in the nucleus are higher in PanIN-3 and PDAC compared to early stage lesions [489]. These observations support the concept of clonal expansion via the gradual acquisition of genetic alterations depicted in figure 16.

1.9.2.2.2 Punctuated Progression Model

This model suggests that catastrophic genetic events during the cell cycle generates many structural chromosomal alterations during the early stages of transformation. Chromothripsis is a phenomenon where thousands of genomic rearrangements occur in a confined region involving a few chromosomes. Wadedell *et al*. highlighted that chromothripsis occurs in about 10% of pancreatic cancer patients [490]. The abrupt nature of punctuated expansion may not be the dominant form by which pancreatic cancers develop; however, it does contribute via providing selective advantage when these alterations promote oncogene activation/expression or disrupt tumor suppressor expression.

1.9.2.3 Exposure to Foreign Environments

A vital pillar of PDAC progression is the interaction of neoplastic cells with the surrounding environments and the confounding selective pressures exerted by the surrounding *stroma*, the prospective *metastatic sites* as well as the *immune system*.

1.9.2.3.1 PDAC Stroma

The initial clues of the fundamental role of pancreatic stroma in disease pathology became evident during studies that examined wound healing in patients with chronic pancreatitis [491]. The Type II EMT program utilizes TGFβ to modulate the tissue repair process by activating fibroblasts and creating an immunosuppressive environment that allows remodeling of the ECM and ultimately triggers regeneration of healthy epithelia [492]. Considering the similarities between neoplastic growth and wound healing [493]

and the dense fibrotic stroma in pancreatic tissue, the effect of a desmoplastic environment on tumor development and progression was relevant.

Paracrine signals from stromal cells such as myofibroblasts contribute to neoplastic growth. Myofibroblasts are a highly proliferative cell type derived from pancreatic stellate cells that transdifferentiated to express α-SMA (alpha smooth muscle actin). Myofibroblasts not only produce ECM components (e.g. hyaluronic acid, HA) to increase stromal density but also secrete the immunosuppressive cytokine TGFβ and growth factors such as PDGF [494]. In response, neoplastic cells produce TGFβ, PDGF and SHH (sonic hedgehog) to further support the growth of the myofibroblasts [495]. Despite the involvement of the stroma in PDAC biology, the role of these factors is not definitive. For instance, some studies demonstrated that calcipotriol (vitamin D analogue) [496], SHH inhibition [497] or short-term HA inhibition [498][499] all led to stromal collapse, reduction in tumor growth and enhanced penetrance of chemotherapeutic drugs. In contrast, genetic deletion of SHH [500] or α-SMA [501] resulted in larger and more metastatic tumors. It has been proposed that the divergent effect of different stromal components on neoplastic growth contributes to intratumoral heterogeneity and the emergence of favorable clones. This was adequately demonstrated in the case of HA, a large hydrophilic negatively charged glycosaminoglycan. Interaction with water increased hydrostatic pressure and interstitial fluid pressure which stressed collagen fibers that are associated with both tumor cells and endothelial cells [502]. This led to collapse of the vasculature and reduction of blood perfusion to the tumor, resulting in poor drug delivery to PDAC tumor beds. The poor blood flow also caused the physical isolation of nutrient-restricted tumor cells [497]. This will result in divergent evolution (allopatric evolution) of

specific cell populations driving tumor heterogeneity [498][503]. Stromal pressures and limited resources exert "evolutionary refinement" prior to the onset of any invasive processes. Cells that survive the refinement will become the most successful at invasion and metastasis.

1.9.2.3.2 PDAC Immune Surveillance

The immune system represents a credible determinant of pancreatic neoplastic growth. Generally, the immune microenvironment of PDACs is a highly immunosuppressive that was established during the clonal expansion of PanINs. The intervention of the immune system at different stages of clonal expansion might have created spatial and temporal bottlenecks that gave rise to highly immuno-heterogeneous tumor cell populations [504]. The immunosuppressive nature of PDACs is driven by a series of immune cell types including regulatory T cells, myeloid-derived cells and alternatively activated tumor-associated macrophages (TAMs) or M2 macrophages [505][506][507]. CD8 and CD4 T cells can be present in the PDAC microenvironment and are a unique target for potential cancer immunotherapies [506]. The interactions between different immune cell types and PDACs are of complex nature and are beyond the scope of this dissertation.

1.9.2.3.3 PDAC Metastasis

The rise of metastatic disease represents a clinically significant determinant of patient outcome, eligibility for resection and treatment options. Examining metastasis as

an evolutionary event suggested a form of competition between primary tumor clones with different degrees of cellular fitness and metastatic propensity [471].

In a seminal study, Rhim *et al* proposed that cellular dissemination from primary PDAC tumors is not necessarily the crucial step in metastasis. In fact, pancreatic epithelial cells from the presumably non-invasive PanINs disseminate and spread before primary tumor formation [508]. This is also consistent with the observation that less than 1% of disseminating cells will survive the treacherous conditions in circulation [509]. Additionally, the representation of the four main genetic alterations in PDAC (*KRAS*, *TP53*, *SMAD4* and *CDKN2A*) was comparable between primary tumor clones and metastatic tumor clones indicating that these alterations do not offer selective advantages for metastasis. Instead, the metastatic propensity is determined by genetic alterations during the expansion phase prior to the onset of invasion *per se* [510]. Campbell *et al*. and more recently Maddipati *et al*. demonstrated that, in patients with metastatic PDAC, certain subclones in primary tumors have acquired a set of unique structural arrangements and passenger mutations that were enriched at the metastatic sites [511][512]. Yachida *et al* identified novel passenger mutations in *CNTN5*, *LMTK2*, *DOCK2* and *MEP1A* which are involved in cell adhesion, tyrosine phosphorylation, cellular motility and surface proteolysis respectively. These mutations were found in metastatic lesions of PDAC with late stage (stage IV) PDAC patients when compared to primary tumors of early stage (stage II) patients with no clinically-diagnosed metastatic disease. However, these mutations were all present but to a lesser clonal representation (except one) in the matched primary tumors of the late stage patients indicating that they are pro-metastasis genes and not metastasis-specific genes. Yachida *et al*. proposed that the poor vascularity of PDACs creates a highly

hypoxic microenvironment that is fertile for cells to acquire these passenger mutations [470]. The same study performed mathematical modelling of PDAC tumor evolution based on the accumulation of passenger mutations. The rationale is based on the putative assumption that passenger mutations are neutral events and do not alter the tumor evolution and thus accumulate independently in each individual cell lineage. Conservative estimates revealed that it takes an average of 11.7 years between the rise of a potentially neoplastic cell from a normal epithelial cell to the emergence of a “founder” clone. Another 6.8 years are required for the founder clone to become a neoplastic lesion some of which will have metastatic propensities. From that point, it estimated that another 2.7 years are expected for these lesions to cause patient death [470].

1.9.3 Driver Genetic Alterations

Alterations in *KRAS*, *CDK2A*, *TP53* and *SMAD4* occur at a high frequency and they are seen in all PanINs. Table 9 summarizes these four alterations and the functions of the altered forms in PDAC compared to their wild-type counterparts in normal tissue (Table 9). Figure 16 also demonstrates the occurrence and prevalence of these alterations across the PDAC progression timeline (figure 16). Lower frequency events have also been characterized [490][513]. These events included activating mutations in *ARID1A*, *KDM6A* and *PREX2*, inactivating mutation in tumor suppressor *ROBO2*, focal amplifications in *ERBB2*, *MET*, *CDK6*, *FGFR1*, PIK3CA and *PIK3R3* and inactivation of DNA repair genes *PALB2, BRCA1* and *BRCA2* [490]. The prevalence of the four genomic alterations indicates that the there is a low likelihood of discovering another high-frequency driver event and more importantly infers that the development of pancreatic cancer has a limited number of evolutionary routes driven by these alterations. Less prevalent alterations are however essential in understanding the interactions between multiple signaling pathways that support pancreatic cancer development and progression.

1.9.3.1 KRAS

The proto-oncogene *KRAS* encodes a 21kDa small GTPase, which alternates between an inactive GDP-bound form and a GTP-bound active form. The generation of the active form is mediated through nucleotide exchange factors (GEFs) that replace bound GDP with GTPs. In contrast, GTPase-activating proteins (GAPs) serve to inactivate KRAS by promoting the hydrolysis of the bound GTP by KRAS. Mutations affecting codon 12 represents around 98% of all K*RAS*-mutant PDACs (figure 18). Rare mutations affecting

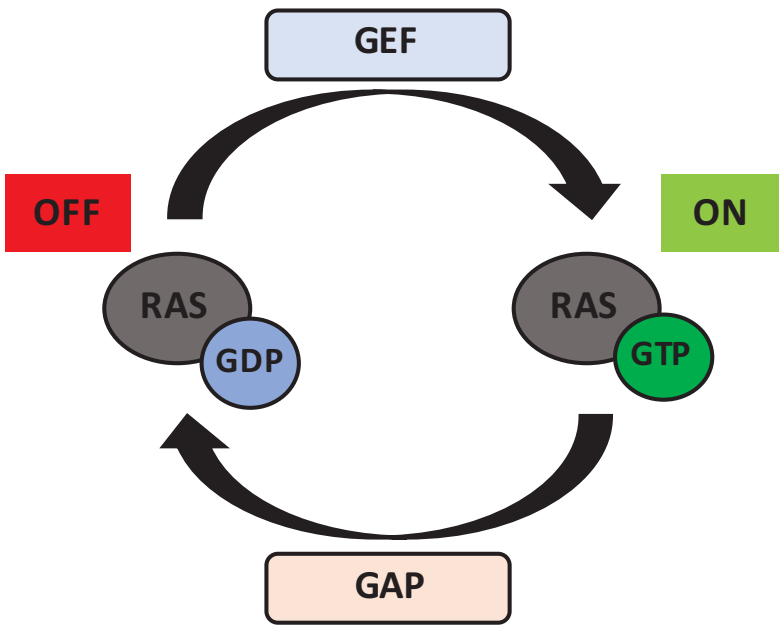
codons 13 and 61 have also been reported. These mutations can both inhibit the intrinsic GTPase activity of KRAS and hinder the association of KRAS with GAPs. This gives rise to a constitutively active GTP-bound KRAS protein and aberrant activation of downstream pathways [514]. These pathways include Raf/MEK/Erk, PI3K/Akt, RalGDS and TIAM/RAC1 modulating survival, proliferation, vesicular trafficking and cytoskeletal rearrangements respectively (figure 19).

Despite the fundamental role of *KRAS* in driving and in many cases sustaining PDAC oncogenesis, therapeutic targeting by direct inhibition of KRAS has proven to be unsuccessful due to the high affinity of GTP to its binding pocket in KRAS (as discussed earlier). The current consensus is that targeting upstream or downstream proteins of KRAS is more likely to succeed clinically [423]. Notably, $KRAS^{G12C}$ retains the ability to remove GTP from its binding pocket, which renders it not constitutively active and a potential drug target using allele-specific inhibitors in G12C-positive patients [515].

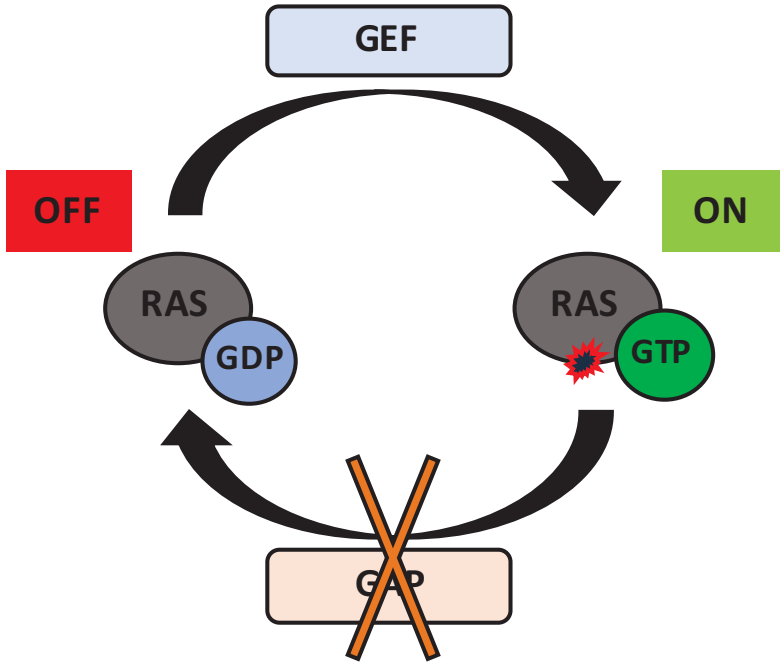
Table 9. The four fundamental genetic alterations in PDAC. The table lists the prevalence of each genetic alteration in PDAC patients, their chromosomal location as well as the role of the wild-type and altered protein.

Genetic event	prevalence	Chromosomal location	Wild-type protein function	Altered protein function
KRAS mutation	90-95%	12p12	- Small GTPase - Cell survival, proliferation, cytoskeletal remodeling	- Activating mutation leads to constitutively active GTPase (except KAS^{G12C} mutation) (Lito 2016).
CDKN2A mutation or loss	>90%	9p21	- P16 transcript: Cell cycle inhibition at G1/S checkpoint (Bertoli *et al* 2013) - P14 transcript: induces cell cycle arrest independently of CDKs (Sharpless *et al* 1999)	- Mutations occur in exon 1 of p16. Homozygous deletions affect both transcripts (Sharpless et al 1999). - Loss of cell cycle checkpoint control leads to aberrant CDK4/6 activity and subsequent telomere shortening and genomic instability (Campbell et al 2010). - Loss of p14 negates apoptosis induced by wild-type *TP53* (sharpless et al 1999)
TP53 mutation or loss	85%	17p13	- DNA damage and stress response. - Modulation of G1/S checkpoint - G2/M arrest to allow DNA repair, or if damage is too severe to induce apoptosis (Vogelstei et al 2000)	- Most mutations are missense mutations that affect its DNA binding capability. - Frameshift mutations and homozygous deletions have been reported.
SMAD4 mutation or loss	55%	18q21	- Co-transcription factor in TGFβ1 signalling - Cell growth and differentiation	- Homozygous deletions (30%) or mutation with LOH (25%) - Mutated in PanIN-3. - Mutation co-exists with gain of function mutation in *TP53* - Wildtype coexists with loss-of function TP53 mutation

Figure 18. Regulation of the activity of the GTPase KRAS. KRAS alternates between an inactive GDP-bound form and a GTP-bound active form. The generation of the active form is mediated through nucleotide exchange factors (GEFs) which replace bound GDP with GTPs. In contrast, GTPase-activating proteins (GAPs) serve to inactivate KRAS by promoting the hydrolysis of the bound GTP. Mutations affecting KRAS (primarily codon 12) can both inhibit the intrinsic GTPase activity of KRAS and hinder the association of KRAS with GAPs. This leads to a constitutively active GTP-bound KRAS protein leading aberrant activation of downstream pathways.

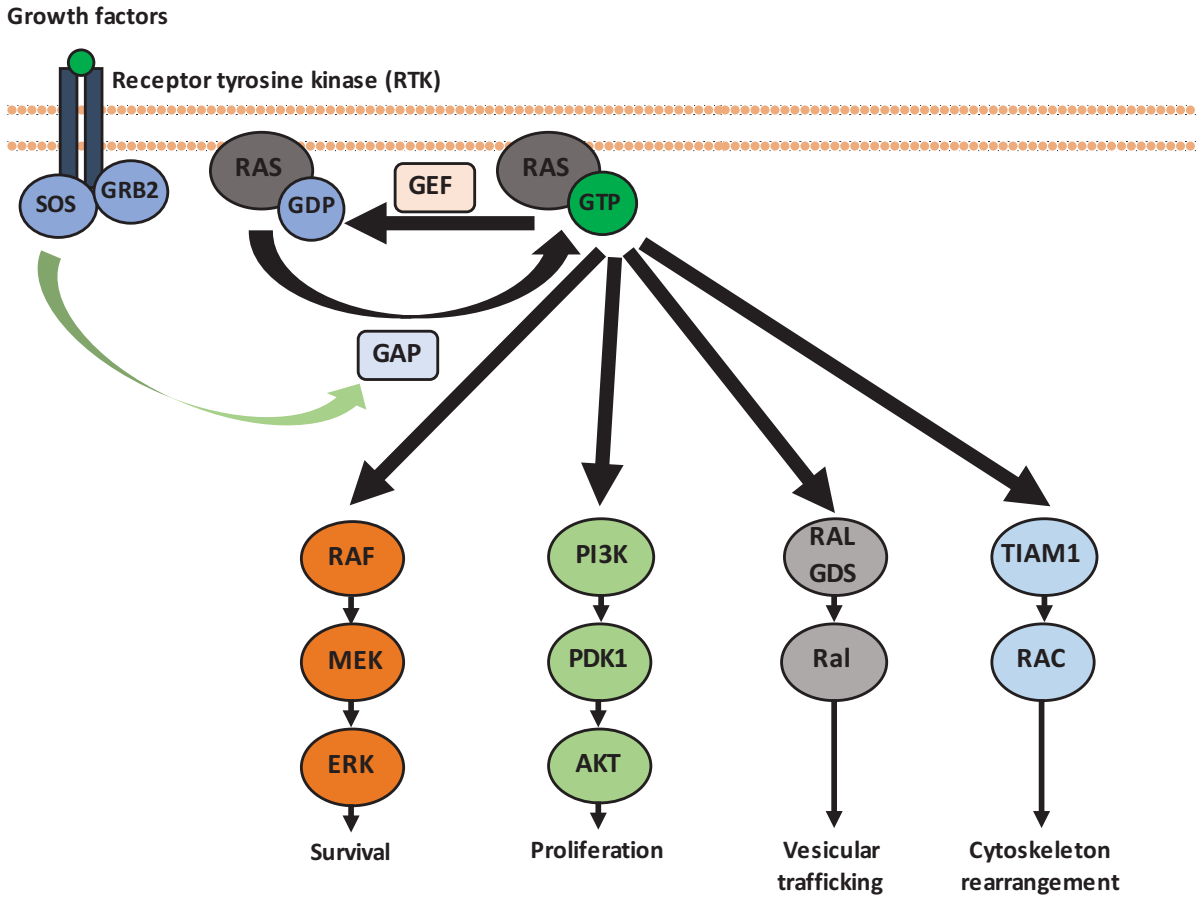


Tightly regulated
Self-limiting



Constitutive
Proliferation, survival, differentiation

Figure 19. KRAS signaling. Activation of a receptor tyrosine kinase (e.g. EGFR) via binding of a growth (e.g. EGF) promotes auto-phosphorylation of the cytoplasmic domain of the receptor. The phosphorylation event creates a docking site for the SOS/GRB2 complex. SOS is a GTPase-activating proteins (GAP) that exchanges a GDP for a GTP. Active KRAS-GTP signals through four major downstream pathways: the RAF/MEK/ERK, PI3K/PDK1/AKT, RalGDS/Ral and TIAM/Rac pathways. These pathways control fundamental cellular processes namely survival, proliferation, vesicular trafficking and cytoskeletal rearrangement.


Growth factors
Receptor tyrosine kinase (RTK)
SOS
GRB2
RAS
GDP
GEF
RAS
GTP
GAP
RAF
MEK
ERK
Survival
PI3K
PDK1
AKT
Proliferation
RAL
GDS
Ral
Vesicular
trafficking
TIAM1
RAC
Cytoskeleton
rearrangement

1.9.3.2 CDKN2A

The cyclin-dependent kinase inhibitor 2A (*CDKN2A*) is a tumor suppressor gene that encodes two transcripts, *p16INK4A* and *p14ARF*. The two transcripts share the same second and third exons while they differ in their first exon (exon 1α for *p16* and exon 1β for p14). Additionally, *p16INK4A* and *p14ARF* have different reading frames for exon 2, which delineates that they are not isoforms. The loss of *CDKN2A* (*p16INK4A* transcript) alleviates the inhibition of cyclin-dependent kinases CDK4 and 6 at the G1/S cell cycle checkpoint leading to aberrant cell proliferation and telomere shortening. The latter increases genomic instability and promotes the formation of structural rearrangements [516].

1.9.3.3 TP53

TP53 encodes a 43.7 kDa tumor suppressor and a transcription factor primarily induced in response to cellular stress or DNA damage. In the presence of such stimuli, *TP53* inhibits the cell cycle at the G1/S checkpoint and promotes G2/M arrest. This allows the DNA damage response (DDR) to initiate DNA repair. If the DNA damage is too severe, *TP53* will initiate apoptosis [517]. *TP53* harbors an inactivating (of tumor suppressor function) mutation in almost 85% of pancreatic cancer patients of which 66% affect its DNA binding capability [510]. These mutations are also associated with deletions in the other allele of *TP53.* Certain mutations of *TP53* can also impose oncogenic roles [517], which are often concomitant with its nuclear accumulation. Mutations that lead to the loss of *TP53* protein expression are found in almost 50% of advanced-stage pancreatic cancer patients[518]. In cases where TP53 is wildtype (15%), other genes that are linked to *TP53*

signaling are often mutated. For instance, the gene encoding the DNA double stranded break sensor protein ATM is often mutated in *TP53*-wildtype tumors [519]. ATM is responsible for TP53 phosphorylation upon DNA damage and is frequently mutated in familial pancreatic cancer [520].

1.9.3.4 SMAD4

As mentioned earlier, *SMAD4* is part of the Smad family of proteins and it acts as a co-activator or co-repressor of transcription factors during TGFß1 signaling. TGFβ, a potent inducer of EMT binds TβRs resulting in receptor activation and subsequent phosphorylation of Smad2/3 proteins [521][327]. Alterations in *SMAD4* are found in 55% of pancreatic cancers, 30% of which are caused by homozygous deletions while the remaining are due to mutations and loss of heterozygosity [334]. The role of TGFβ signaling pathway in PDAC is dualistic in which it initially acts as an anti-proliferative mechanism that inhibits the dysplastic growth of PanIN-1 and PanIN-2 at which point *SMAD4* is still wild-type. Upon loss of *SMAD4* in PanIN-3 (figure 16), TGFβ drives oncogenic growth. In 10% of pancreatic cancer patients, tumors that possess a wild-type *SMAD4* acquire other inactivating mutations in the TGFβ signaling pathway such as *TGFBR1*, *TGFBR2*, *SMAD3* or *ACVR1B* (activin A receptor type 1B) [522].

1.9.4 Co-occurrence of PDAC Driver Events

Considering the molecular pathways that are affected by the four genes, it is unlikely that the effects caused by these alterations act independent of each other. Yachida *et al.* demonstrated an intriguing relationship between *TP53* and *SMAD4*. *SMAD4*

inactivating mutations had a strong positive correlation with *TP53* gain of function changes. In contrast pancreatic tumors with wild-type *SMAD4* harbored a loss-of function alteration in *TP53* [510]. The question of how the interdependence of *SMAD4* and *TP53* status is linked to the molecular and genetic profiles (discussed in 1.9.5) of pancreatic cancers is yet to be fully addressed.

In early stage PanINs where *SMAD4* is likely to have a wild-type status, mutated *KRAS* serves to inhibit TGFβ/Smad signaling by promoting the degradation of Smad4 [523][524]. In contrast, wildtype TP53 can associate with Smads to mediate TGFβ-induced changes in gene expression [525]. Mutant *KRAS* also inhibits serine 9 phosphorylation of TP53, which in turn prevents TP53 interaction with TGFβ-activated Smads [526]. Conversely, when *TP53* is mutated, TGFβ and *KRAS* cooperate where Smads serve as platforms for mutant TP53 and wildtype TP63 (tumor suppressor) to form a complex that antagonizes TP63 functions. TP63 inhibition reduces the expression of TP63-induced tumor suppressor genes leading to an increase in TGFβ1-induced metastasis [527]. Collectively, these driver events form a complex interdependent network of signaling molecules that undermine anti-neoplastic mechanisms within the cell.

1.9.5 PDAC Subtypes

Analyses of somatic mutations, chromosomal structural variants, epigenetic alterations and gene expression have generated a series of approaches to stratify PDAC patients. These approaches were contributed by five seminal studies which are discussed below and are summarized in table 10. Although the applicability of these techniques into clinical settings can be compromised by differences in data processing and the statistical algorithms used to obtain patient subgroups and to for sample preparation and processing, they still offer concrete evidence of the existence of genetically and molecularly distinct subtypes of PDAC.

1.9.5.1 Mutational and Transcriptional Profiling

1.9.5.1.1 The Jones Classification

In 2008, Jones and colleagues reported the first comprehensive analyses to define the mutational, copy number and transcriptional landscape of PDAC [522]. The authors performed genome-wide sequencing of 24 PDAC tumors to identify somatic mutations as well as homozygous deletions and amplifications. The results revealed that each tumor contained an average of 63 genetic alterations. The authors then utilized SAGE (serial analysis of gene expression) to measure gene expression, a quantification method that compensates for preferential amplification bias towards larger transcripts as seen in traditional microarray analysis [528]. Combination of SAGE results with the mutational/copy number landscape of these tumors uncovered that the 63-alteration average affected 12 core signaling pathways concomitantly altered in almost two thirds of

the tumors. These pathways include KRAS signaling (*KRAS*, *MAP2K4*, *RASGRP3*), apoptosis (e.g. *HIP1*, *CASP10*), DNA repair (e.g. *TP53*, *ERCC4*, *ERCC6*), G1/S checkpoint (e.g. *CDKN2A*, *FBXW7*, *APC2*), hedgehog signaling (e.g. *TBX5*, *SOX3*, *GLI1*, *GLI3*), cell adhesion (e.g. *CDH1*, *CDH2*, *CDH10*), integrin signaling (e.g. *ITGA4*, *ITGA9*), Jnk signaling (e.g. *TNF*, *ATF2*), invasion (e.g. *ADAM11*, *ADAM12*, *PRSS23*), small GTPase signaling (*ARHGEF7*, *PLCB3*, *RP1*), TGFβ signaling (e.g. *SMAD4*, *SMAD3*, *TGFBR2*) and Wnt/Notch signaling (e.g. *MYC*, *PPP2R3A*, TSC2) [522] (table 10). Although this study identified the key genetic alterations and molecular pathways implicated in PDAC, the small sample size did not allow any meaningful subtyping.

1.9.5.1.2 The Collisson Classification

In 2011, Collisson *et al*. published the first attempt at molecular subtyping of PDAC. The study combined the gene expression data from multiple studies including human and mouse cell lines in order to maximize sample sizes. Multivariate analysis of the non-negative matrix factorization (NNMF) of differentially expressed genes and their clustering patterns supported the identification of three transcriptionally-distinct subtypes; classical, quasi-mesenchymal (QM) and exocrine-like. The study also developed a gene signature, called *PDAssigner*, which consisted of 62 genes whose expression was sufficient to distinguish between the three subtypes. The classical subtype was characterized by high expression of adhesion molecules and epithelial markers such as *TFF1* (trefoil factor 1), *MUC13* (Mucin 13) and *TMEM45B* (transmembrane protein 45B). This subtype had the best long-term survival compared to the other two subtypes. Patients in the quasi-mesenchymal subtype expressed high levels of mesenchymal genes (*GPM6B*, glycoprotein

M6B; *NT5E*, 5' nucleotidase) with very poor prognosis. The exocrine-like subgroup uniquely expressed genes encoding digestive enzymes such as *REG1B* (regenerating islet-derived 1 beta), *CFTR* (cystic fibrosis transmembrane conductance regulator) and *PNLIPRP2* (pancreatic-lipase-related protein 2) (table 10). The exocrine-like subtype had an improved short-term survival compared to the quasi-mesenchymal subtype but a relatively similar long-term survival [529].

Table 10. The five PDAC classification studies. The table enlists five genome-wide studies that examined the genetic landscape of PDAC. These studies were named based on the first author of the respective publication. The table also lists the various subtypes derived from each study and their defining characteristics.

classification	subtypes	Defining characteristic/s
The Jones classification	N/A	- Average of 63 alterations per PDAC - Affected 12 core signaling pathways
The Collisson classification	1) Quasi-mesenchymal	- High levels of mesenchymal genes (*GPM6B*, *NT5E*)
	2) Classical	- High levels of adhesion molecules and epithelial markers (*TFF1*, *MUC13*, *TMEM45B)*
	3) Exocrine-like	- High levels of digestive enzyme genes (*REG1B*, *CFTR*, *PNLIPRP2*)
The Moffit classification	1)basal-like tumor with normal stroma	- Basal-like tumors express *S100A1*, *UCA1* and *VGLL1* - Classical tumors express *FAM3D*, *ATAD4* and *BTNL8* - Normal stroma expresses stellate cell markers such as *ACTA2*, *DES* and *VIM* - Activated stroma expresses a macrophage-like gene signature (e.g. *ITGAM*, *CCL13*, *CCL18*)
	2) basal-like tumor with activated stroma	
	3) classical tumor with normal stroma	
	4) classical tumor with activated stroma	
The Bailey classification	1) squamous	- *TP53* and *KDM6A* mutations - Increased methylation of endodermal genes
	2) pancreatic progenitor	- Expression of pancreatic development genes (e.g. *PDX1*, *MNX1*)
	3) immunogenic	- Immunosuppressive gene expression profile
	4) aberrantly differentiated endocrine exocrine (ADEX)	Expression of: - KRAS signaling genes - Endocrine cell differentiation genes (*NKX-2* and *NEUROD1*) - Exocrine cell differentiation genes (*RBPJL* and *NR5A2*)
The Waddell classification	1) stable	- <50 SVs (structural variants) - Global aneuploidy - 20% of PDAC
	2) locally-rearranged	- Focal amplifications - 30% of PDAC
	3) scattered	- Non-random chromosomal damage - <200 SVs - 36% of PDAC
	4) unstable	- >200 SVs - Deficiencies in DNA repair - 14% of PDAC

1.9.5.1.3 The Moffit Classification

The Collisson *et al*. study utilized micro-dissected samples that enriched for PDAC tumors cells while minimizing the contamination with stromal cells and normal ductal tissue. Despite the added benefit of using micro-dissection, it bypasses the ability to examine stromal and normal cell transcriptional profiles that might be contributing to PDAC development. In fact, PDAC is characterized by a dense fibrotic stroma that have been shown to enhance the aggressive nature of cancer cells and contribute to chemotherapy [530][531]. A 2015 study by Moffitt *et al*. utilized virtual microdissection instead of mechanical microdissection. The study collected tumor-associated samples (including tumor, tumor-associated stroma and normal tissue) along with non-tumor-associated normal and stromal tissues from various organs including pancreas, liver and immune cells. The non-tumor-associated tissue was used to create a normal cell gene signature as well as a normal stromal cell signature; both were then compared to the tumor cell, tumor-associated stroma and tumor-associated normal tissue signatures. They identified stroma-specific genes that allowed the distinction of two types of tumor associated-stroma, "normal" and "activated". "Normal" stroma genes included *ACTA2*, *DES* and *VIM*, markers of pancreatic stellate cells. In contrast, "activated" stroma expressed higher levels of integrins (e.g. *ITGAM*) and chemokines (*CCL13*, *CCL18*) and resembled macrophage-like gene signatures. Other genes overexpressed in activated stroma were Wnt signaling and MMP genes suggesting their potential involvement in PDAC progression. Importantly, PDAC patients with an activated stroma had a poorer survival than those with normal stroma [532].

Moffitt *et al*. also characterized two tumor subtypes; basal-like subtype that expresses high levels of *S100A1*, *UCA1* and *VGLL1* and a classical subtype expressing *FAM3D*, *ATAD4* and *BTNL8*. Classical PDAC patients had a better overall survival than basal-like patients. Basal-like and classical subtypes possessed both normal and activated stroma giving rise to four subtypes: 1) basal-like tumor with normal stroma, 2) basal-like tumor with activated stroma, 3) classical tumor with normal stroma and 4) classical tumor with activated stroma (table 10). Patients with classical tumors and normal stroma had the best prognosis compared to the other three subtypes [532].

1.9.5.1.4 The Bailey Classification

A 2016 study by Bailey *et al* expanded the PDAC mutational and transcriptional profile analysis using whole-genome and deep-exome sequencing into a larger cohort of 456 PDAC patients. The mutational landscape of these tumors revealed 32 driver mutations that affected 10 core signaling pathways. These included KRAS, TGβ, NOTCH, ROBO/SLIT and WNT signaling as well as G1/S checkpoint, chromatin modification, SWI-SNF nucleosome remodeling, DNA repair and RNA processing [533]. The mutated genes and the affected pathways strongly resemble the mutational and pathway analyses patterns originally described by Jones and colleagues [522]. Bailey *et al*. also performed RNA-Seq gene expression and methylation analyses and identified four distinct molecular subtypes: 1) squamous, 2) pancreatic progenitor, 3) immunogenic and 4) aberrantly differentiated endocrine exocrine (ADEX) (table 10). Squamous tumors were characterized with *TP53* and *KDM6A* mutations, increased methylation of the pancreatic endodermal genes and upregulation of the proto-oncogene *TP63*-ΔN, which lacks the transactivation

domain at the NH2 terminus. In contrast, pancreatic progenitors expressed genes that are unique to early pancreatic development and include *MNX1*, *PDX1* and *FOXA2/3*. Immunogenic tumors displayed gene networks representative of an immunosuppressive environment. ADEX tumors expressed genes involved in KRAS signaling, endocrine cell differentiation (*NKX2-2* and *NEUROD1*) and exocrine cell differentiation (*RBPJL* and *NR5A2*). Patients with squamous tumors had the worst prognosis among all four subtypes [533].

1.9.5.2 Structural Variants Profiling

Mutational and transcriptional investigations yielded comprehensive coverage of PDAC genetic alterations that drive tumorigenesis. However, another contributing factor to PDAC are somatic structural rearrangements of chromosomes. These structural rearrangements or variants (SVs) include deletions that lead to gene disruptions, copy number gains and amplifications. This may result in oncogene overexpression and gene fusions, which generate oncogenic fusion proteins. SVs are potentially catastrophic events that can directly drive cancer development and progression [534]. The prevalence of SVs in PDAC was observed as early as 1995 where karyotyping displayed consistent chromosomal abnormalities [535]. Later studies confirmed a high degree of genomic instability in PDAC [512]. Recently, Waddell and colleagues performed whole-genome sequencing to discern chromosomal SVs in PDAC [490].

1.9.5.2.1 The Waddell Classification

Waddell *et al*. combined SVs and point mutation analyses of 100 PDAC tumors, which resulted in increasing the prevalence of inactivating events that involved *TP53* (74% of all PDAC with 3 SVs and 71 point mutations), *SMAD4* (31% of all PDAC with 9 SVs and 22 point mutations) and *CDKN2A* (35% of all PDAC with 11 SVs and 24 point mutations). Waddell *et al*. also identified two novel genes (*PREX2* and *KDM6A*) that were frequently mutated or affected by structural rearrangements in around 10% of all PDACs. Furthermore, the study derived four subtypes based on patterns of chromosomal SVs. A "stable" subtype represented 20% of PDACs, contained less than 50 SVs and displayed global aneuploidy indicating a deficiency in cell cycle control. *TP53* mutations in "stable" tumors were less frequent compared to the other subtypes. The second subtype was the "locally-rearranged" subtype which was found in 30% of all PDACs and displayed marked focal amplification events on one or two chromosomes. This subtype contained copy number gains in putative oncogenes such as *KRAS*, *GATA6*, *ERBB2*, *MET*, *CDK6*, *PIK3R3*, *PIK3CA* and *SOC9*. The third subtype was named the "scattered" subtype as it displayed a moderate number of non-random chromosomal damage with less than 200 SVs and was present in 36% of PDAC patients. The last subtype was classified as "unstable" and was present in 14% of PDAC patients (table 10). Patients with unstable genomes exhibited a significant number of SVs (more than 200) that was largely attributed to deficiencies in DNA repair. The latter was driven by mutations in genes involved in the BRCA pathway (*BRCA1*, *BRCA2*, *PALB2*, *ATM*, *TP53*, *REV3L* and *RPA1*) and sensitized these patients to the DNA-damaging platinum therapy [490].

1.10 Subchapter 10: Conceptual Framework

The activation of plasminogen at the surface of cancer cells is a crucial step in mediating cancer cell invasion and promoting an aggressive tumor phenotype. The latter has been closely linked to the ability of cancer cells to undergo EMT. However, **the role of proteins that drive the plasminogen activation process is of utmost importance in order to understand the biological mechanism of cancer cell escape from primary tumors and subsequent formation of metastasis especially in the context of EMT**. Herein, the dissertation attempts to answer the above question using three overarching objectives:

1.10.1 Objective I: Plasminogen Activation and EMT

Question: How does the epithelial or mesenchymal state of a cancer cell alter its surface plasminogen activation?

Hypothesis: Mesenchymal cells have enhanced plasminogen activation capabilities compared to epithelial cells.

Methodology:

1) Utilize models of epithelial-like and mesenchymal-like lung cancer cells to study plasminogen activation *in vitro*.
2) Employ the above models to decipher the signaling pathways regulating the expression of major proteins involved in plasminogen activation.

1.10.2 Objective II: Plasminogen Activation and Lung Cancer

Question: Do plasminogen activation genes serve as predictors of lung cancer patient outcome?

Hypothesis: Plasminogen activation genes are potential predictors of overall survival in NSCLC patients?

Methodology:

1) Develop a strategy to systematically assess expression of genes involved in plasminogen activation.
2) Employ hierarchical clustering methods and Pearson correlation comparisons to identify the most differentially-expressed plasminogen genes.
3) Perform Kaplan Meier survival analyses to assess the predictive capacity of the respective differentially-regulated genes in different histological subtypes of lung cancer.
4) Generate a predictive gene signature.

1.10.3 Objective III: Plasminogen Receptor S100A10 and Pancreatic Cancer

Question What is the role of the plasminogen receptor S100A10 in the biological and clinical presentation of PDAC?

Hypothesis: S100A10 is a potential predictor of PDAC patient survival and a driver of PDAC tumorigenesis and invasiveness.

Methodology:

1) Assess transcript and protein expression of S100A10 in normal, PanINs and PDAC using published mRNA datasets and tissue microarrays of PDAC patients.
2) Apply univariate and multivariate regression models to examine the predictive power of S100A10 as a novel biomarker of outcome.
3) Assess the role of S100A10-mediated plasminogen activation on cancer cell growth and invasion *in vitro* using our well-established plasminogen activation and invasion assays.
4) Decipher the molecular mechanisms that modulate S100A10 expression in PDAC *in vitro*.
5) Study the effect of S100A10 depletion on *in vivo* tumor growth using a PDAC mouse model.

CHAPTER 2: METHODS

2.1 Cell lines

All cell lines were purchased from the American Type Culture Collection (ATCC) (except HMLE, BxPC-3 and AsPC-1) and tested negative for mycoplasma. A549 (CCL-185, male), NMuMG (CRL-1636, female), Panc-1 (CRL-1469, male) and MCF-7 (HTB-22, female) cells were supplemented with Dulbecco's Modified Eagle's Media (DMEM, Hyclone) containing 10% fetal bovine serum (FBS) (Hyclone). BEAS-2B (CRL-9609, male) were supplemented with LHC-8 media (Thermo-fisher scientific) with and without FBS (Hyclone, Canada, characterized). Panc 10.05 (CRL-2547, male), BxPC-3 (CRL-1687, female), AsPC-1 (CRL-1682, female) and HPAF-II (CRL-1997, male) were supplemented with Roswell Park Memorial Institute (RPMI) media with 10% FBS. The AsPC-1 (female) and Bx-PC3 (female) cell lines were a generous gift from Dr. David Hoskin (Dalhousie University, Halifax, Nova Scotia, Canada). The human mammary epithelial cell line (HMLE, female) was a generous gift from Dr. Robert Weinberg (Whitehead Institute for Biomedical Research, Cambridge, Massachusetts) and was cultured in a 1:1 ratio of DMEM F12 1:1 and mammary epithelial cell growth medium (MEGM, Lonza) supplemented with 13 µg/mL bovine pituitary extract, 20 µg/mL human epidermal growth factor, 10 µg/mL insulin, 1 µg/mL gentamicin/amphotericin and 2 µg/mL hydrocortisone (Lonza Clonetics) and 10% FBS. All cells were cultured in the presence of 1% pencillin/streptomycin (Hyclone) and were maintained at 37°C with 5% CO_2.

2.2 Chemical reagents

All reagents were optimized for ideal dosage and time courses to minimize cellular toxicity while maximizing response of proteins of interest. Zarnestra (Tipifarnib) (Selleckchem, S1453, 10 μM) and decitabine (Sigma-Aldrich, A3656, 10 μM), Rapamycin (Tocris, 10 μM), A83-01(Tocris, 2939, 25μM), Tiplaxtinin (Tocris, 5565/10, 10 μM) and LY294002 (Santa Cruz Biotechnology, 154447-36-6, 50 μM) were reconstituted in DMSO. Doxycycline (Clontech, 631311, 1μg/mL), bhFGF-1 (R&D systems, 233-FB-025, 0 to 200 ng/ml) and heparin sodium salt (Tocris, 2812/100, 100 ug/ml) was reconstituted in tissue-culture grade water. Plasminogen (Sigma-Aldrich, 528180, 0.5 μM), S2251 (Chromogenix, 82033239, 5 μM), ε-aminocaproic acid (Sigma, A2504, 100mM) and aprotinin (Pentapharm 2.2μM) were reconstituted in PBS. TGFβ1 (Peprotech, 20 ng/ml unless indicated) was reconstituted in 10mM citric acid.

2.3 Antibodies

The sources and dilutions of antibodies are as follows:

- β-actin (Sigma Aldrich mouse monoclonal anti-β-actin, A2228, 1:2000)
- N-cadherin (BD Biosciences mouse monoclonal anti-N-cadherin, 610921, 1:2000)
- E-cadherin (BD Biosciences mouse monoclonal anti-E-cadherin, 610181, 1:2000)
- Vimentin (Sigma-Aldrich goat polyclonal anti-Vimentin, V4630, 1:1000)
- S100A10 (BD Biosciences mouse monoclonal anti-S100A10, 610070, 1:2000)
- Annexin A2 (BD Biosciences mouse monoclonal anti-Annexin II, 610069, 1:2000)
- GAPDH (Biochain mouse monoclonal anti-GAPDH, Y3322, 1:2000)

- p-S6K (Cell signaling rabbit monoclonal anti-pS6K, 9205S, 1:1000)
- FOXC2 (Bethyl laboratories rabbit polyclonal anti-FOXC2, A302-383A, 1:1000)
- PAI-1 (Cell signaling rabbit monoclonal anti-PAI-1 D9C4, 11907, 1:2000)
- uPAR (Santa Cruz rabbit polyclonal anti-uPA FL-290, sc-10815, 1:300)
- p-Erk (Cell signaling rabbit monoclonal anti-pErk (Thr202/Tyr204), 9101, 1:1000)

2.4 Plasmids

The *S100A10* shRNA1 knockdown construct was designed by cloning the following dsRNA oligo 5'-GAT CCC CGT GGG CTT CCA GAG CTT CTT TCA AGA GAA GAA GCT CTG GAA GCC CAC TTT TTA-3' and 5'-AGC TTA AAA AGT GGG CTT CCA GAG CTT CTT CTC TTG AAA GAA GCT CTG GAA GCC CAC GGG-3' into the pSUPER-retro-puro vector plasmid (OligoEngine). The non-silencing siRNA (4390843) and S100A10 siRNA (s12429) were purchased from the Ambion Silencer Select pre-designed and validated siRNA library (ThermoFisher Scientific). The pGIPZ SMAD4 and FOXC2 constructs were obtained from EGAD (enhanced Gene Analysis and Discovery) core facility at Dalhousie University. The pBabe-puro control (#1764), $KRAS^{G12D}$ (#58902) and pBabe-puro-FOXC2 (#15535) constructs were obtained from the plasmid depository Addgene. The transfected clones were selected in 1 μg/ml puromycin.

2.5 Stable Retroviral Transfection

To establish stable *S100A10*-depleted and FOXC2-overexpressing cell lines, Phoenix cells (in 6-well plates) were first transfected with 4 μg of the pSUPER-retro scramble control, pSUPER-retro-S100A10 shRNA1, pBabe-puro control and pBabe-puro

FOXC2 plasmids using the lipofectamine 2000 transfection reagent (Invitrogen). 10 µl of lipofectamine 2000 reagent was incubated with 240 µl of Opti-MEM for each well for 5 min at room temperature. The plasmids (in 250 µl Opti-MEM) and lipofectamine solutions were then mixed and incubated for 20min at room temperature. The total of 500 µl was then added to 1.5 ml of culture media (no antibiotics added). Retroviral supernatants were collected at 24hrs and 48hrs post transfection. Cells of interest were then transduced with the retroviral supernatants (with 10 µg/ml polybrene). Puromycin selection started at 48hr post infection.

2.6 Stable Lentiviral Transfection

To establish the pGIPZ SMAD4 shRNA and FOXC2 shRNA cell lines, a mix of 6µg of the pGIPZ lentiviral vector, 4.3 µl of the trans-lentiviral packaging mix and 15 µl of CaCl2 and 150µl of 2X HBSS as per manufacturer's instructions (Dharmacon, TLP5912). The mix was incubated for 3min at room temperature then added into one well of HEK293T cells (6-well plate) containing 2 ml of antibiotic-free media. Lentiviral supernatants were collected at 24hrs and 48hrs post transfection. Cells of interest were then transduced with the lentiviral supernatants (with 10µg/ml polybrene). Puromycin selection started at 48hr post infection.

2.7 Transient Transfection

$3.5x10^4$ cells were seeded into 6-well plates overnight. 4ug of non-targeting or S100A10 siRNAs were reconstituted in 250 µl Opti-MEM. 10 µl of lipofectamine 2000 reagent was incubated with 240 µl of Opti-MEM in each well for 5 min at room

temperature. The siRNA and lipofectamine solutions were then mixed and incubated for 20min at room temperature. The 500 µl total was then added to 1.5 ml of culture media (no antibiotics added). Transfection media was not removed until cells were trypsinized 48 hours after transfection and seeded for further analysis.

2.8 Western Blotting

Cells were washed with PBS and lysed in lysis buffer (1% NP-40, 150 mM NaCl, 20 mM Tris, pH 7.0, 1 mM EDTA and 1 mM EGTA) containing 2X Halt protease and phosphatase inhibitors (Thermo Scientific). Samples were subject to SDS-polyacrylamide gel electrophoresis then transferred onto a nitrocellulose membrane. Membranes were incubated with primary antibodies overnight at 4°C or one hour at room temperature. Li-COR secondary antibodies used to visualize bands using a LI-COR Odyssey imaging scanner. Relative band intensities per lane were determined for each protein and normalized to intensities of GAPDH or actin bands. Band intensity was measured using the Odyssey Li-COR software V3.0. The intensity was then subtracted from background intensity (above or below band). Relative band intensities per lane were determined for each protein and normalized to intensities of GAPDH or β-actin bands. Noteworthy, a consistent upregulation of β-actin was observed in A549 cells in response to TGFβ1 treatment (figure 20b). Protein expression was therefore normalized relative to GAPDH under conditions where A549 cells were treated with TGFβ1. All gels were cropped for clarity. Molecular weights of proteins are listed under the antibodies section.

2.9 Quantitative RT-PCR

RNA was extracted using TRIzol as per standard procedure (Qiagen). 2 µg of RNA was used for the synthesis of cDNA using Superscript II (Invitrogen). *S100A10* (p11) gene expression was amplified using gene-specific primers on the CFX96™ platform. All primer sequences are listed in supplemental table 22. The primers were designed with high specificity, purchased from IDT and then verified for optimal amplification. Relative mRNA expression was calculated using the Livak and Schmittgen's $2^{-\Delta\Delta CT}$ method and normalized to GAPDH as a reference gene [536].

2.10 Plasminogen Activation Assay

Cells were seeded overnight into 96-well plates at $1x10^5$ cells/well (A549, BEAS-2B, iKRAS) or $5x10^3$ cells/well (Panc-1). Cells were then washed with Dulbecco's PBS (Hyclone), incubated with 0.5 µM (in 75 µl) plasminogen for 10 min and then incubated with 0.5 mM S2251 (in 75 µl) (chromogenic plasmin substrate, Chromogenix, Diapharma Group) (figure 62b). The rate of plasmin generation was quantified based on the absorbance at 405 nm every 4 min for 4hrs using the Spectra M3 plate reader (Molecular Devices). A405 was subtracted from A600 to account for turbidity. The rate of plasmin generation was determined from the slope of the A405nm vs $time^2$ of the kinetic curve.

2.11 Surface Expression Measurement by Flow Cytometry

Cells were washed with PBS, gently lifted with a cell lifter and then blocked with 2% FBS in PBS. Cells were then incubated with primary antibodies at room temperature for 30 min, washed 3 times with PBS then incubated with FITC- or PE-conjugated

secondary antibodies for 30min at room temperature. Cells were then washed with PBS and analyzed on a BD FACSCalibur flow cytometer. Surface expression was quantified based on relative fluorescent intensities (RFIs) using the Flowing Software 2 [537]. Mean RFI of S100A10-stained samples was subtracted from an isotype-stained control). RFI was calculated by subtracting the mean fluorescence intensity of samples incubated with the anti-S100A10 antibody from that of samples incubated with IgG1 isotype control.

2.12 Surface Expression Measurement by Surface Biotinylation

Cells were seeded into 150-cm cell culture plates until 90% confluency. Cells were then washed twice with ice-cold PBS and incubated with 1mg/ml Sulfo-NHS-SS-Biotin (Pierce, Thermo Scientific) for 30 min at 4°C. The reaction was quenched with 100 μM glycine in PBS, then washed twice with ice-cold PBS. Cells were then lysed in RIPA lysis buffer. 500 μg of protein lysates were incubated with 100 μl of Dynabeads M-280 streptavidin (Invitrogen) for 2hrs at 4°C with rotation. Biotinylated proteins were separated from unlabeled proteins using a magnet with five washes of the lysis buffer. Biotinylated proteins were then suspended in protein sample buffer, boiled at 95°C for 10 min and subjected to gel electrophoresis.

2.13 H&E Staining

Cells were seeded on Poly-L-Lysine slides then fixed and permeabilized using 1:1 ratio of methanol and acetone. Fixed cells were then stained with hematoxylin, washed with PBS, then stained with eosin. Glass slides were mounted for bright-field microscope imaging (Zeiss).

2.14 Gene Array Analysis and Normalization

RNA Seq V2 RSEM expression values for the TCGA tumors (Supplemental Fig. S1A) as well as CCLE Z-scores (Figure 53b) were downloaded from Cbioportal. RNA Seq V2 REVs were normalized by dividing by the mean expression value [538]. Z-scores were compared using the z-ratio equation as previously described [539]. z-ratio=z-$score_{avg}$ (cell type)– z-$score_{avg}$ (CML) / SD of z-score differences. z-$score_{avg}$ (cell type) is the average of the z-scores of all the cell lines within a particular tumor type (CML: chronic myelogenous leukemia). z-$score_{avg}$ (CML) is the average of the z-scores of CML cell lines which had the lowest average z-score and was used as a control. SD of z-score differences is the standard deviation (SD) of the [z-ratio=z-$score_{avg}$ (cell type)– z-$score_{avg}$ (CML)] values of each tumor type. A z-ratio of 1.96 or higher is considered equivalent to a *p*-value =< 0.05. For normal/tumor data normalization**,** expression values were retrieved from the GEO (Gene expression Omnibus) as per corresponding accession numbers (GSE16515[540], GSE22780[541], GSE3654[542], GSE1542[543], GSE15471[544] and GSE28735[545]) log-transformed and median-centered per array (Figure 54). Expression values from Segara *et al* [546] and Logsdon *et al*[547] gene arrays were extracted from Oncomine[548] as median centered intensities.

2.15 CDHA Patient Cohort

Ethics approval was received from the Capital Health Research Ethics Board of Capital District Health Authority (CDHA) on Oct 09 2014 (CDHA-RS/2012-206). 89 samples were collected from pancreatic adenocarcinoma patients admitted to the Queen Elizabeth Hospital (Halifax, NS) between 2001 and 2009. All patients underwent surgical

resection at which point samples were collected prior to adjuvant chemotherapy/radiation. Samples were formalin-fixed and paraffin-embedded (FFPE).

2.16 Tissue Microarray (TMA) Construction and Immunohistochemistry

2mm areas containing both tumor and stroma from each sample were used as a single core. Normal, precancerous and cancerous cores were collected from each sample. Cancerous cores were only collected in triplicates. 11 TMAs were constructed with 40 cores/TMA. TMA blocks were then sectioned and subject to immunohistochemical staining (IHC). Primary rabbit anti-human S100A10 antibody (1:800, Proteintech 11250-1-AP) was used to stain TMA using the Ventana automated staining platform (Roche) followed by DAB (3,3'-Diaminobenzidine) stain to visualize staining areas.

2.17 DAB Quantification

TMAs were scanned on the Aperio AT2 high volume digital whole slide scanning system (Leica Biosystems) at 20X magnification. Three representative images of tumor and stroma in each core were captured for staining quantification. Images were subject to color deconvolution in ImageJ as previously described in Varghese *et al*. Briefly, color deconvolution yields three images, hematoxylin (counter stain), DAB, and an additional image. Stained areas were manually highlighted by the selection tool, color de-convoluted and quantified using the IHC profiler plugin. The plugin was developed by Varghese *et al*. [549]. The profiler is ImageJ-compatible and analyzes cytoplasmic signals from de-convoluted DAB images. The profiler also generates a pixel intensity histogram which plots the pixel intensity values of the brown DAB color from the darkest (intensity value =

0) to the lightest (intensity = 255) shades. Pixel intensity values were divided into four sub-categories: 0-60, 61-120, 121-180, and 181-255. The plugin then outputs the percentage of pixels in each category of the highlighted area (figure 55a).

2.18 H-scoring

The scoring assignment of selected DAB-stained areas was accomplished via H-scoring [550] using the following formula: H-score = (% of pixels in 0-60 category * 3) + (% of pixels in 61-120 category * 2) + (% of pixels in 121-180 category * 1) + (% of pixels in 181-255 category * 0). H-scores range from 0 to 300. To generate cut-off classifiers, we considered an H-score <100 to be negative/weak staining, H-score of 100 to 200 to be low positive and H-score of >200 to be high positive values (supplemental table 16). The H-score was then normalized to the average of all intensities.

2.19 Kaplan Meier Survival

Survival percentage was calculated non-parametrically based on observed overall survival times. At the time of last follow-up, live patients were assigned a zero (0) due to absence of event (i.e. death). Deceased individuals were assigned a one (1) since the event of death occurred. Recurrence-free survival (RFS) was represented by the duration between a complete response to treatment and the status of disease at the time of last follow-up i.e. disease free (0) or progressive disease (1). Log-rank (Mantel-Cox test) was used to compare relative risk in Kaplan Meier plots with binary classifiers (median and optimal cut-offs). Multiple comparisons testing was applied to ternary classifier and an adjusted *p*-value was

calculated based on Bonferroni-corrected threshold. The $p\text{-value}_{adj} = p\text{-value}_{raw}/k$, where raw p-value = 0.05 and represents k the number of comparisons made.

2.20 Univariate and Multivariate Analysis

Univariate and multivariate regression models were fitted to the overall (OS) and recurrence-free survival (RFS) of the TCGA PDAC patient cohort. The variables/predictors were: *S100A10* mRNA (RNA Seq V2 RSEM), gender, race, age, grade, tumor dimension, stage, metastasis, smoking, alcohol consumption. A natural logarithm (ln) was applied to the *S100A10* mRNA raw expression values (REVs). The fitted single-variable model included all variables listed. The fitted multivariate model included all variables except smoking history and alcohol consumption due to high number of missing values. A semi-parametric proportional hazard regression model was fitted to identify variables that are predictors of overall and recurrence-free patient survival times. The model assumes: $H(t|Z) = h0(t) \exp(\beta' Z)$ where h0(t) is an arbitrary baseline hazard rate, β' is a vector of coefficients, Z is a vector of co-variants or variables. We fit the semiparametric proportional hazards regression model for each single variable. The univariate and multivariate analyses results are summarized in tables 11 through 14.

2.21 Normalization of GDC Tumor RNA-Seq and CCLE Microarray Gene Expression Data

RNA Seq V2 RSEM expression values of GDC (Genomic Data Commons) tumors (figure 53a) and expression Z-scores of Cancer Cell line Encyclopedia (CCLE) cell lines (figure 53b) were downloaded from Cbioportal and were normalized to the mean

expression value [538]. For pancreatic normal/tumor data normalization**,** expression values were retrieved from the Gene expression Omnibus (GEO as per corresponding accession numbers (GSE16515 [540], GSE22780 [541], GSE3654 [542], GSE1542 [543], GSE15471[544] and GSE28735 [545]) log-transformed and median-centered per array (figure 54 and supplemental figure 11). Expression values from Segara et al [546] and Logsdon *et al* [547] gene arrays were extracted from Oncomine [548] as median centered intensities.

2.22 KM Plot

Expression data and overall survival times from 11 lung cancer studies were downloaded from KM plot (KMplot.com). The accession numbers are as follows: TCGA [551], GSE50081 [552], GSE4573 [553], GSE37745[554], GSE31908 (unpublished), GSE3141 [555], GSE31210 [556][557], GSE30219 [558], GSE29013 [559], GSE19188 [560] and GSE14814 [561]. A median cut-off was applied to derive the univariate regression analysis of each gene as an independent predictor of overall survival. All studies used one of two microarray expression platforms: GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array or GPL96 [HG-U133A] Affymetrix Human Genome U133A Array. Raw expression values above and below the median were annotated as 0 and 1 respectively. “0”s and “1”s were compiled from each study. A total of 720 adenocarcinoma patients and 524 squamous cell carcinoma patients were used in the merged cohort. Biased arrays with two or more parameters that were outside the 95% range of all arrays were excluded from the analysis as quality control. Outliers were defined as a

parameter that is outside the 95% range of all arrays. Arrays with potential low-quality spike-in hybridization controls (bioB, BioC and BioD spikes) were also excluded.

2.23 *In vivo* Intra-peritoneal Mouse Model

$5x10^6$ Panc-1 cells (scramble control or *S100A10*-shRNA1) were suspended in PBS and intra-peritoneally injected into the lower right abdominal area of NOD-SCID mice. After 12 weeks post injection, tumors were collected, weighed, fixed with 10% formalin and embedded in paraffin for histological examination. The animal experiment studies were approved by Dalhousie Animal Ethics (protocol number 15-143) and housed at the Carlton Animal Care Facility (CACF).

2.24 Invasion Assay

$5x10^4$ Scramble control and *S100A10*-shRNA1 Panc-1 cells were seeded in serum-free media into the upper chamber of a trans-well Boyden chamber with 8µm pores that was coated with an artificial matrix, matrigel (BD Biosciences) (figure 62d). The bottom chamber contained 10% FBS as a chemoattractant. Plasminogen (0.5 µM) was added to the top chambers 5 hours after seeding. After 72 hours, the cells that traversed to the bottom of the 8 µm pore membrane were stained with hematoxylin and eosin and counted. Five fields of view were counted per membrane at 20X magnification.

2.25 Ras Activation Assay

Protein lysates from vehicle- and zarnestra-treated Panc-1 and BxPC-3 cells were incubated with a Raf-1 pulldown reagent linked to agarose beads as per manufacturer's

instructions (Millipore, 16117). Lysates were then separated on an SDS-polyacrylamide gel and immunoblotted using a RAS antibody (Millipore, Clone RAS10, 05-516).

2.26 MTS Assay

$1x10^4$ cells were seeded in a 96-well plate. Promega's CellTiter96 one solution reagent (Promega, G3582, 20 µl) was added to 100µL of the culture medium and incubated for 4hrs at 37°C after which the amount of soluble formazan was measured by recording the absorbance at 490nm using the SpectraM3 plate reader (Molecular Devices).

2.27 Annexin V and 7AAD Staining

Cells were incubated with 5µL of annexin V-FITC in 100 µl of binding buffer (10mM HEPES, 150 mM NaCl, 2.5 mM CaCl2 in PBS and pH adjusted to 7.4) for 15min at room temperature in the dark, centrifuged and washed 2X with PBS. Cells were then incubated with 7AAD for 5-10min at room temperature in the dark. Fluorescence on the FITC (FL-1) and PI (FL-3) channels was measured immediately using a flow cytometer. Live cells are negative for annexin V and 7AAD. Cells in early apoptosis are positive for Annexin V and negative for 7AAD. Late apoptotic cells are double positive.

2.28 Bisulfite conversion and pyrosequencing

As previously described 54,55, DNA methylation was analyzed by sodium bisulfite pyrosequencing on a PyroMark Q24 Advanced pyrosequencer using the DNA EpiTect Fast DNA Bisulfite Kit and PyroMark PCR Kit (Qiagen) as per manufacturer's instructions beginning with 500 ng template DNA. A custom assay covering the region immediately

upstream of the S100A10 gene transcription start site (TSS) was designed using PyroMark Assay Design software (v2.0; Qiagen) and validated to amplify a single PCR product (417 nt). Primers are listed in supplementary table S11. PCR conditions for both assays: 95°C, 15 min; (94°C, 30s; 56°C, 30s; 72°C, 30s) x 50 cycles; and 72°C, 10 min.

2.29 Statistical Analysis

All experiments were performed in triplicates in three independent experiments. All statistical analyses were performed using GraphPad Prism 5 software. Unless indicated in the figure legends, statistical significance was determined using the unpaired student t-test, paired t-test, one-way ANOVA or Z-ratio accordingly (see figure legends). A significance threshold of p-value < 0.05 was used ($p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***, $p < 0.0001$ ****) except for multiple comparisons tests (in ternary classification) (p-value < 0.017). For z-score transformation, a Z-ratio of 1.96 was considered equivalent to a p-value of 0.05.

CHAPTER 3: TGFβ1 and PI3K Regulate S100A10 and PAI-1 Expression to Modulate Plasminogen Activation in Cells Undergoing EMT.

3.1 Study rationale

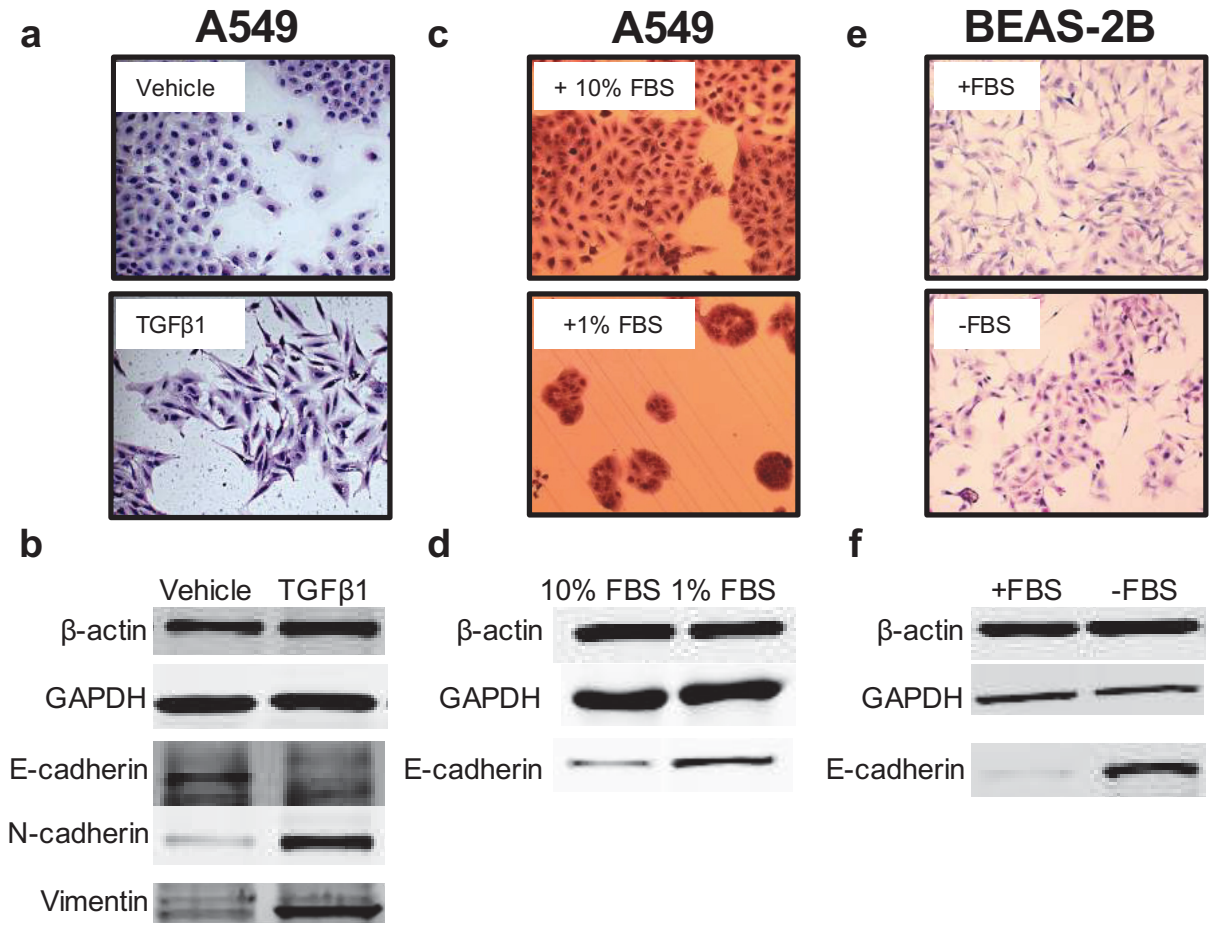
It is generally accepted that EMT contributes to cancer cell dissemination and escape into the circulation resulting in the formation of distant-site metastasis. The latter mandates cancer cells to undergo the reverse process of MET (mesenchymal to epithelial transition) to support metastatic growth [562]. An extensive body of research has demonstrated that EMT drives cellular migration and invasiveness *in vitro* and *in vivo* (reviewed in [563]). However, it has been assumed that EMT is often coupled with enhanced proteolytic activity particularly through the activation of MMPs. Eckert *et al*. demonstrated that Twist-induced EMT is associated with enhanced MMP activity at the surface of breast cancer cells that in turn enhances their invasiveness [107],[564]. Whether **cells undergoing EMT also possess an enhanced plasminogen activation** capacity has not been addressed. In addition, **the question of whether the driver of cancer cell dissemination depends on the degree to which cancer cell proteases are activated and/or the epithelial or mesenchymal state of the cell remains unanswered.** Here we decipher the mechanism of regulation of plasminogen activation in both epithelial and mesenchymal cells. Our findings show that S100A10, PAI-1 and uPAR are differentially modulated in epithelial and mesenchymal cells. The activation of plasminogen was partly dependent on surface levels of S100A10 and overall levels of uPAR and PAI-1 and less

dependent on the mesenchymal/epithelial status of cells. In addition, S100A10 was found to be regulated through canonical Smad4-dependent TGFβ1 signaling and repressed by FOXC2-mediated PI3K-mTOR signaling.

3.2 Establishment of 2D epithelial and mesenchymal *in vitro* cell models.

To assess the regulation of plasminogen activation in epithelial and mesenchymal cells, we utilized three 2D *in vitro* cell models; TGFβ1-induced EMT in A549 cells [565], serum withdrawal-induced generation of epithelial-like BEAS-2B [566] and A549 [567] cells. Based on morphology, A549 cells supplemented with 10% FBS appear to have an intermediate epithelial/mesenchymal phenotype (figure 20a, upper panel). TGFβ1 treatment induces a morphological transition into a fibroblast-like mesenchymal shape (figure 20a, lower panel) that can be blocked by the TGFβ1 receptor inhibition (ALK4/5/7 inhibitor, A83-01) (supplemental figure 1, lower right panel). Notably, A83-01 treatment reverts A549 cells into a highly epithelial-like round morphology (supplemental figure 1, lower left panel). A similar epithelial-like morphology was also achieved by culturing A549 cells [567] in 1% serum (figure 20c) and BEAS-2B cells [566] in the absence of serum (figure 20e). TGFβ1 induced the expression of EMT markers such as N-cadherin and vimentin and repressed E-cadherin expression in A549 cells (figure 20b). In contrast, serum withdrawal from A549 and BEAS-2B cells restored E-cadherin expression (figure 20d, 20f). Both N-cadherin and vimentin were not detectable in BEAS-2B cells (figure 20f) as previously reported [566][568].

Figure 20. TGFβ1 and serum withdrawal induce epithelial-like and mesenchymal-like phenotypes in A549 and BEAS-2B cells. Hematoxylin and eosin (H&E) staining of (a) vehicle (10 mM citric acid)-treated (top) and TGFβ1-treated (20 ng/ml) (bottom) A549 cells (96 hours), (c) A549 cells cultured in the presence of 10% (top) or 1% (bottom) FBS for 96 hours, and (e) serum-supplemented (+10% FBS) (top) BEAS-2B cells, serum-starved (-FBS) (bottom) BEAS-2B cells after 7 days of serum starvation. (b, d, f) Western blot analysis of β-actin, GAPDH, E-cadherin, N-cadherin and Vimentin in the three cell models. N-cadherin and Vimentin were not detectable in BEAS-2B cells.

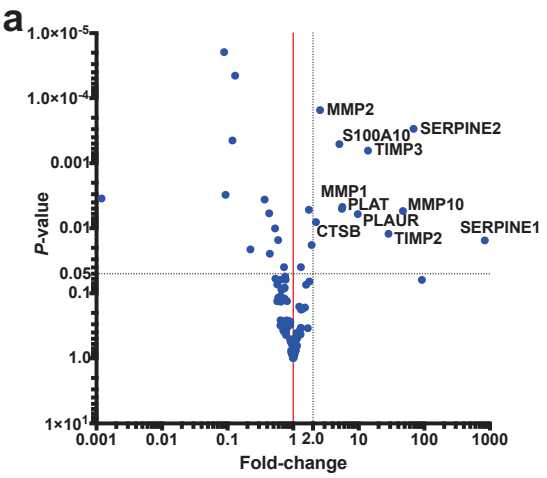

a
A549
Vehicle
TGFβ1
c
A549
+ 10% FBS
+1% FBS
e
BEAS-2B
+FBS
-FBS
b
Vehicle
TGFβ1
β-actin
GAPDH
E-cadherin
N-cadherin
Vimentin
d
10% FBS
1% FBS
β-actin
GAPDH
E-cadherin
f
+FBS
-FBS
β-actin
GAPDH
E-cadherin

3.3 S100A10 mRNA and protein expression is regulated by SMAD4-mediated TGFβ1 signaling in A549 cells.

To identify the components of the plasminogen activation system that might contribute to the ability of epithelial and mesenchymal cells to activate plasminogen, we examined the mRNA expression of a series of 130 putative upstream and downstream components of the extracellular protease regulatory components relevant to the plasminogen activation system (supplemental table 1) during TGFβ1-induced EMT in A549 cells treated with 5ng/ml TGFβ1 for 72 hours[569] (see methods). An overall upregulation of these components was observed in TGFβ1-treated A549 cells indicating their potential implications during EMT. A *p*-value of 0.05 and at least a two-fold difference were set as cut-offs which resulted in 11 significantly upregulated genes (*SERPINE1*, *SERPINE2*, *TIMP2*, *MMP10*, *PLAUR*, *TIMP3*, *PLAT*, *MMP1*, *S100A10*, *MMP2* and *CTSB*) (figure 21a). Interestingly, our analysis revealed that *S100A10* (S100A10) was the only plasminogen receptor to be significantly upregulated by TGFβ1 (5.06-fold increase) among all 11 characterized plasminogen receptors[195] (figure 21b). Since plasminogen binding to cell surface receptors is a rate-limiting step in the activation of plasminogen by plasminogen activators[570], we further interrogated the significance of this observation in the three models of epithelial and mesenchymal cells (figure 20). We first confirmed that TGFβ1 treatment increased mRNA expression of S100A10 (figure 21c). TGFβ1 also upregulated S100A10 protein expression (4.89-fold) in A549 cells (figure 21d) in a dose-dependent manner (supplemental figure 2b). Noteworthy, an upregulation of β-actin was observed in A549 cells in response to TGFβ1 treatment (figure 20b). Protein expression was therefore normalized relative to GAPDH under conditions

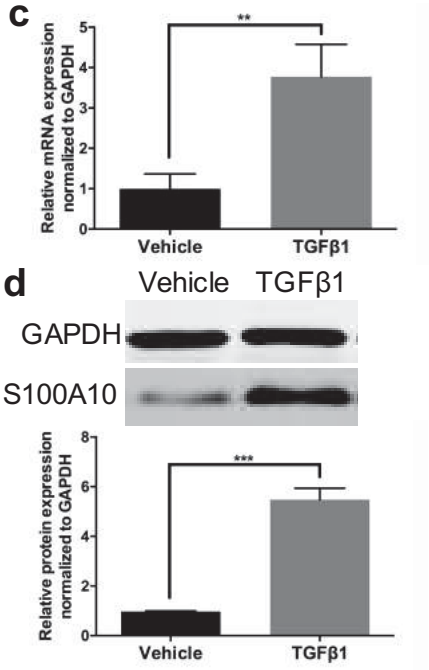
where A549 cells were treated with TGFβ1. To exclude the possibility that the observed increases in S100A10 were limited to A549 cells, we treated multiple cancer cell types that are known to undergo EMT in response to TGFβ1 treatment. The upregulation of S100A10 protein was observed in HMLE [571], MCF-7 [572], and Panc10.05 cells (supplemental figure 2c, 2d, 2e respectively).

Figure 21. TGFβ1 increases the expression of the plasminogen receptor S100A10 at the protein and mRNA levels in A549 cells. (a) Volcano plot showing the differential gene expression of 130 genes involved in the plasminogen activation process. (b) fold-change and p-value of S100A10 upregulation by TGFβ1 (5 ng/ml) in A549 cells after 72 hours. (c) RT-qPCR, (d) western blot analysis and quantification of S100A10 levels in vehicle-treated and TGFβ1-treated (96 hours) (20ng/ml) A549 cells.



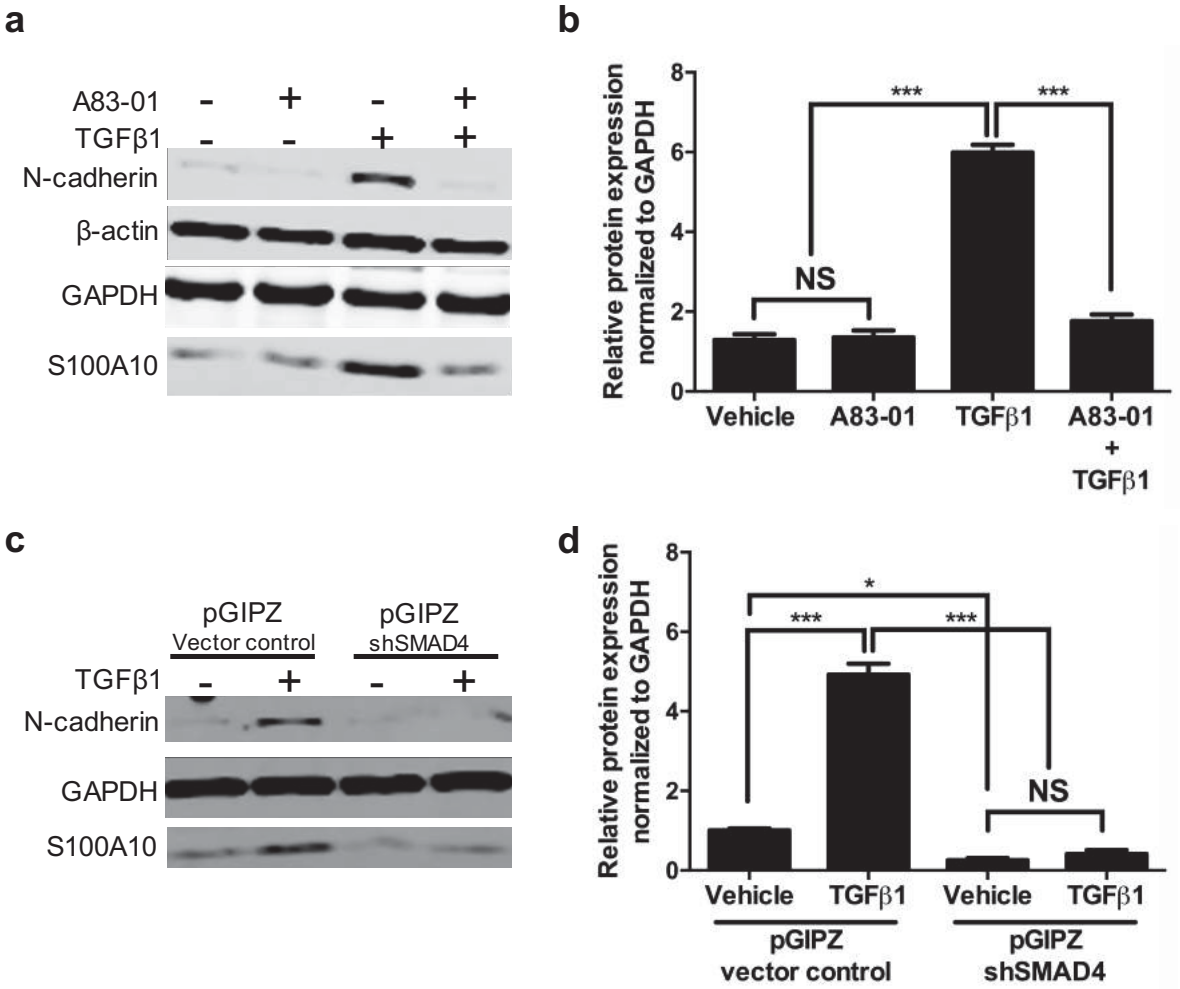
b

Gene	*P*-value	Fold-change
S100A10	0.0005	5.06



Next, we utilized A83-01 to inhibit TGFβR1-mediated EMT[573] in A549 cells (supplemental figure 1). The inhibition of TGFβ-receptor signaling in A549 cells is known to promote cell proliferation and inhibit TGFβ1-mediated apoptosis[574], migration[575] and invasion[576]. TGFβR1 inhibition decreased N-cadherin expression and importantly abrogated S100A10 upregulation after TGFβ1 treatment (figure 22a, 22b). Collectively, these results confirmed that the plasminogen receptor S100A10 is uniquely regulated by TGFβ1/ TGFβR1 signaling. Notably, and in contrast to Panc10.05 cells, TGFβ1 failed to upregulate S100A10 in the pancreatic cancer cell line BxPC-3 (supplemental figure 2f). The latter harbors a homozygous deletion in Smad4 and is therefore not responsive to TGFβ1 [577]. To assess the effect of canonical Smad-dependent TGFβ1 signaling on S100A10 expression, SMAD4 was depleted in A549 cells using short-hairpin RNA. Smad4-depleted cells treated with TGFβ1 failed to upregulate S100A10 (figure 22c, 22d) indicating that S100A10 regulation by TGFβ1 is dependent on Smad4. Similarly, Smad3 inhibition with the inhibitor SIS3 [578] achieved a similar reduction in S100A10 upregulation upon TGFβ1 treatment (supplemental figure 3a, 3b). In addition, we utilized bhFGF/H (basic human fibroblast growth factor constituted in heparin) treatment to prevent EMT-induced changes as demonstrated in A549 cells treated with TGFβ1[579]. bhFGF/H inhibited both N-cadherin and S100A10 upregulation by TGFβ1 in A549 cells in a dose-dependent manner (figure 22e).

Figure 22. S100A10 expression is driven by canonical SMAD4-dependent TGFβ1 signaling in A549 cells. Western blot analysis (a) and quantification of S100A10 protein levels (b) of A549 cells treated with 20 ng/ml TGFβ1 (96 hours) with and without the TGFβR1 inhibitor (A83-01, 25 μM). Western blot analysis (c) and S100A10 protein quantification (d) of TGFβ1-treated cells transfected with a stable pGIPZ shRNA knockdown construct targeting SMAD4. (e) Western blot analysis and quantification of protein lysates from vehicle-treated and TGFβ1-treated (20ng/ml) (96 hours) A549 cells in the presence of ascending concentrations of 0 to 200 ng/ml of bhFGF-1/H (basic human fibroblast growth factor-1 constituted in 100 ug/ml heparin) after 72 hours.


a
A83-01 - + - +
TGFβ1 - - + +
N-cadherin
β-actin
GAPDH
S100A10
b
Relative protein expression normalized to GAPDH
8
6
4
2
0
***
***
NS
Vehicle
A83-01
TGFβ1
A83-01 + TGFβ1
c
pGIPZ Vector control
pGIPZ shSMAD4
TGFβ1 - + - +
N-cadherin
GAPDH
S100A10
d
Relative protein expression normalized to GAPDH
8
6
4
2
0
*
***
***
NS
Vehicle
TGFβ1
Vehicle
TGFβ1
pGIPZ vector control
pGIPZ shSMAD4
e
hFGF-1/H (ng/ml) 0 12.5 25 50 100 200 0 12.5 25 50 100 200
TGFβ1 - - - - - - + + + + + +
N-cadherin
GAPDH
p-erk
p-S6K
S100A10
S100A10 quantification 1.00 0.97 0.86 0.98 0.91 1.04 3.23 2.31 1.91 1.63 1.82 1.63

3.4 S100A10 is a TGFβ1-responsive gene and not an EMT gene.

Concurrent treatment of TGFβ1 and A83-01 or SMAD4 depletion prevented A549 cells from undergoing EMT hence not allowing us to discern a TGFβ1-specific response or a global EMT effect on S100A10. To address this issue, we examined how S100A10 expression was affected in epithelial and mesenchymal cells independent of TGFβ1 using the serum-withdrawal models in A549 and BEAS-2B cell models (figure 20). Surprisingly, serum withdrawal, which induces an epithelial-like morphology [566][567], also upregulated S100A10 protein (figure 23a) and transcript (figure 23b) in A549 cells. Similar increases in S100A10 protein (figure 23c) and transcript (figure 23d) were also seen in BEAS-2B. Importantly, TGFβ1 treatment of serum-supplemented BEAS-2B cells, that are mesenchymal in nature, upregulated S100A10 protein expression (supplemental figure 3c). We were not able to examine the effect TGFβ1 treatment on BEAS-2B cells deprived of serum due to substantial cell death (data not shown). Collectively, these findings suggested that the effect on S100A10 is a TGFβ1-dependent response and is not necessarily linked to the epithelial or mesenchymal status of the cell.

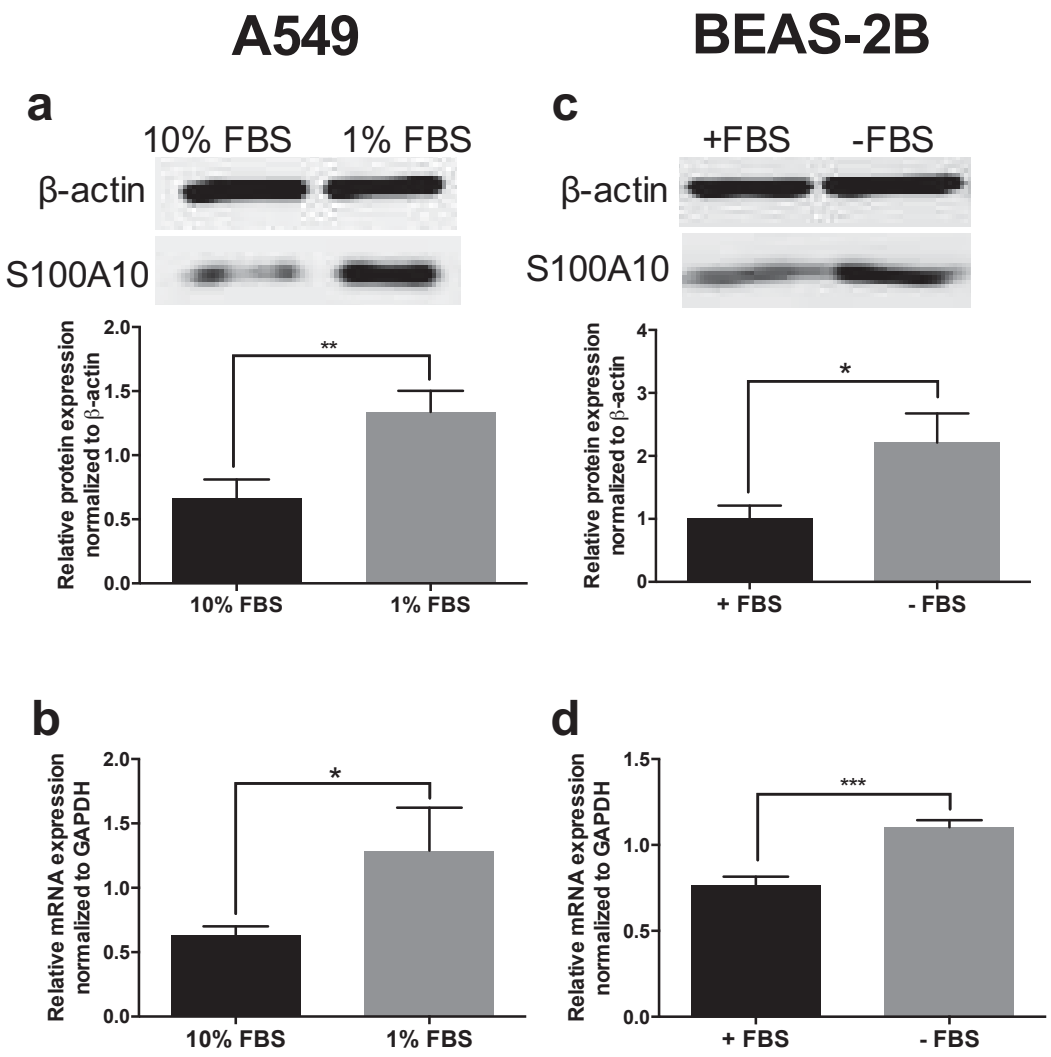
3.5 PI3kinase signaling represses S100A10 expression via FOXC2.

The serum withdrawal experiment with BEAS-2B and A549 cells not only uncoupled S100A10 expression from the epithelial/mesenchymal status of the cell, it also suggested the potential involvement of growth pathways in the regulation of S100A10 under EMT-inducing conditions. This is particularly relevant since TGFβ1, in addition to inducing EMT, inhibited cell growth as seen in A549 cells (supplemental figure 4a) and HMLE cells (supplemental figure 4b) concomitant with S100A10 upregulation. The effect

of serum withdrawal on A549 cells which increased S100A10 protein expression was exacerbated in the presence of TGFβ1 and abrogated by A83-01 (figure 24a).

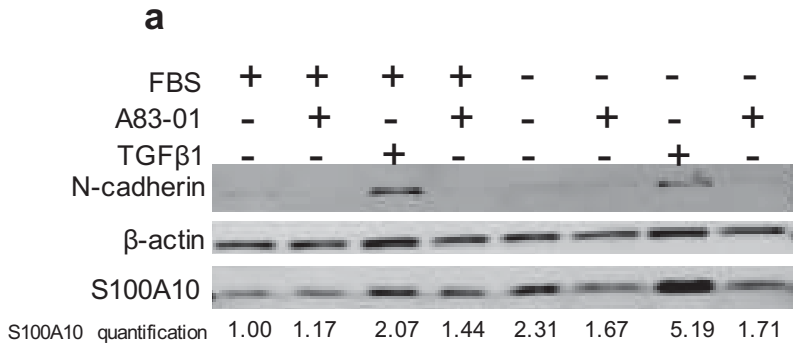
Serum growth factors are potent activators of receptor tyrosine kinases which trigger intracellular pro-growth signals[580]. In addition, the mechanism of action of the growth factor bhFGF is mediated through the activation of two major pathways namely MAPK/MEK/Erk and PI3K/Akt/mTOR. Inhibition of both pathways prevented the restoration of E-cadherin expression in response to bhFGF in A549 cells treated with TGFβ1[579]. To examine the involvement of pro-growth pathways such as the MAPK/MEK/Erk and PI3K signaling pathway on S100A10 expression and how it may affect TGFβ1-mediated upregulation of S100A10, we treated A549 cells with the MEK inhibitor U0126 and the PI3K inhibitor LY294002. Inhibition of MEK did not affect S100A10 expression in the absence or presence of TGFβ1 (supplemental figure 4c, 4d). In contrast, PI3K inhibition increased S100A10 protein expression, an effect that was then exacerbated in the presence of TGFβ1 (figure 24b, 24c). The upregulation upon PI3K inhibition was dose-dependent even in the presence of TGFβ1 (supplemental figure 5a). S100A10 upregulation was also achieved in A549 cells when treated with mTOR inhibitor rapamycin (supplemental figure 5b) implicating the PI3K/mTOR axis in regulating S100A10 in addition to the TGFβ1/Smad4 pathway. It should be noted that N-cadherin upregulation by TGFβ1 was inhibited by the concomitant inhibition of PI3K demonstrating a dependency of N-cadherin expression by both canonical Smad4-dependent TGFβ1 signaling (figure 22a, 22c) as well as PI3K signaling (supplemental figure 5a).

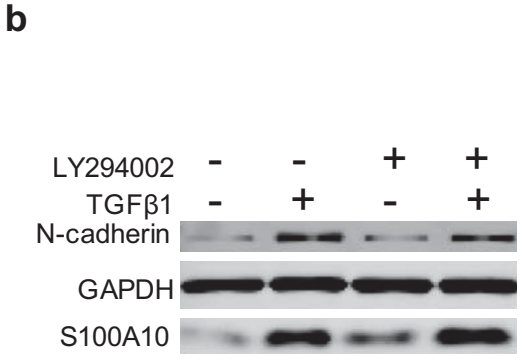
Figure 23. Serum deprivation promotes an epithelial-like phenotype and increases S100A10 protein and transcript levels. (a) Western blot analysis and (b) RT-qPCR of S100A10 in A549 cells supplemented with 10% serum (FBS) or 1% serum. (c) Western blot analysis and (d) RT-qPCR of S100A10 in BEAS-2B cells supplemented with 10% serum (+FBS) or no serum (-FBS).

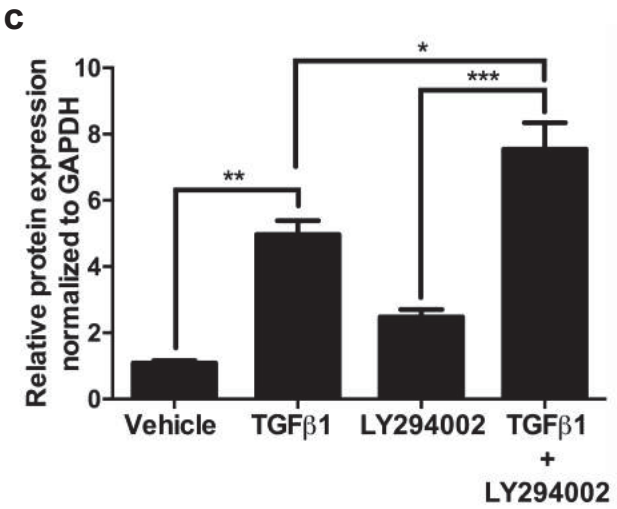

A549
BEAS-2B
a
10% FBS
1% FBS
β-actin
S100A10
Relative protein expression normalized to β-actin
2.0
1.5
1.0
0.5
0.0
**
10% FBS
1% FBS
c
+FBS
-FBS
β-actin
S100A10
Relative protein expression normalized to β-actin
4
3
2
1
0
*
+ FBS
- FBS
b
Relative mRNA expression normalized to GAPDH
2.0
1.5
1.0
0.5
0.0
*
10% FBS
1% FBS
d
Relative mRNA expression normalized to GAPDH
1.5
1.0
0.5
0.0
***
+ FBS
- FBS

A recent CHIP-chip analysis of the transcription factor *FOXC2* DNA binding sites revealed that the *S100A10* gene promoter contains the highly-conserved *de novo* motif (GCCAACAAAAACA, chr1: 150,219,126-150,220,276) [581]. FOXC2 has been implicated in PI3K in response to insulin[582][583]. Here we demonstrate that the inhibition of PI3K by LY294002 reduced FOXC2 expression [584] (figure 25d). The expression of FOXC2 increased phosphorylation of S6K (figure 25d) and partially rescued the growth of LY294002-treated cells (supplemental figure 5c) with no effect on TGFβ1-treated A549 cells. To verify whether FOXC2 regulates S100A10 expression via PI3K signaling, A549 cells were transfected with the pBabe-FOXC2 construct. FOXC2 expression caused a dramatic downregulation of S100A10 protein (figure 25a, 25b) and mRNA levels (figure 25c).

Figure 24. Serum starvation or PI3K inhibition have an additive effect on TGFβ1-induced increase of S100A10 in A549 cells. (a) Western blot analysis and S100A10 protein quantification of A549 cells treated with TGFβ1 and A83-01 for 96 hours in the presence/absence of serum. (d) Western blot analysis and (e) S100A10 quantification in A549 cells treated with the PI3K inhibitor LY294002 in the presence or absence of TGFβ1.


a
FBS + + + + - - - -
A83-01 - + - + - + - +
TGFβ1 - - + - - - + -
N-cadherin
β-actin
S100A10
S100A10 quantification 1.00 1.17 2.07 1.44 2.31 1.67 5.19 1.71

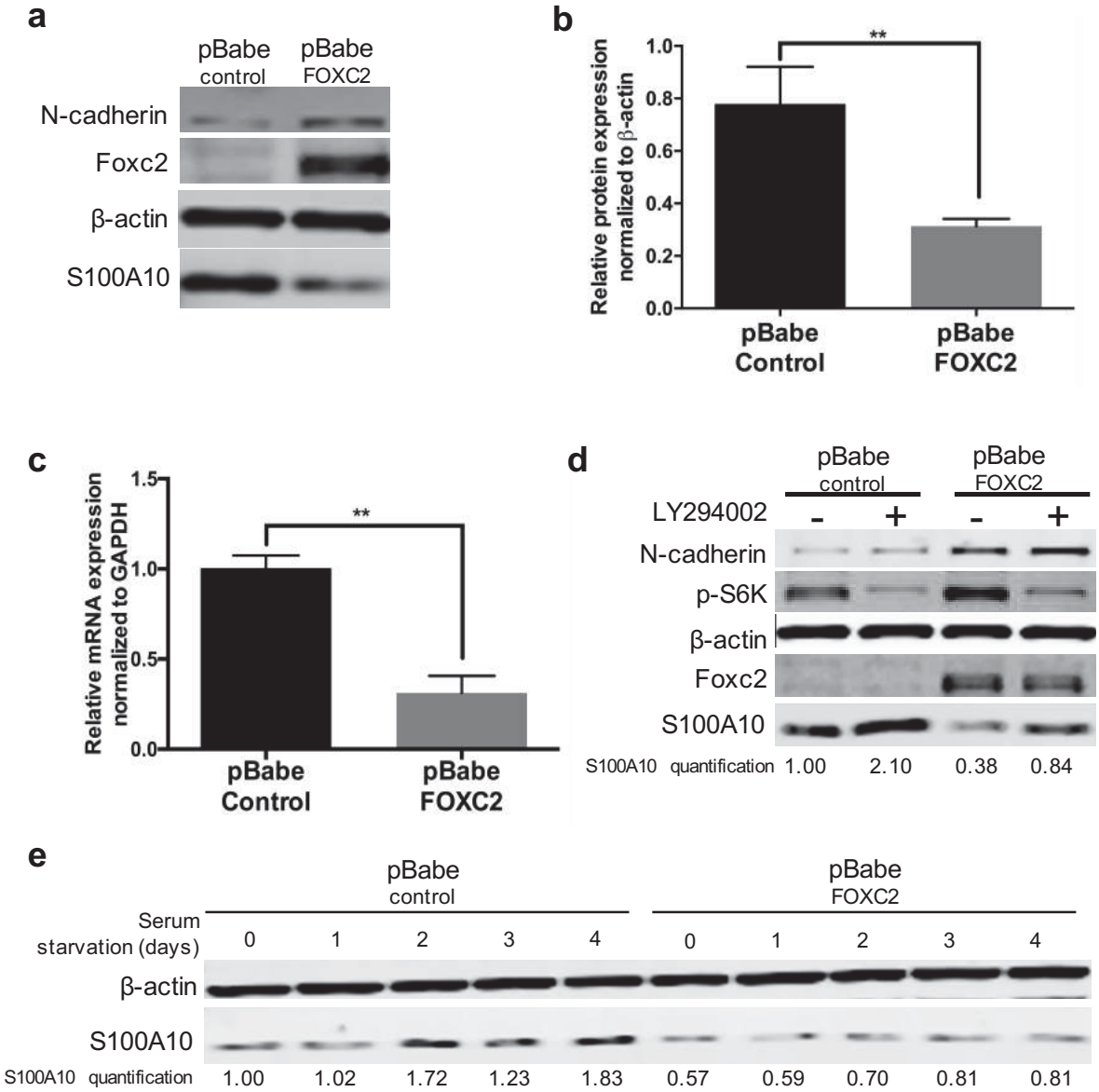

b
LY294002 - - + +
TGFβ1 - + - +
N-cadherin
GAPDH
S100A10


c
Relative protein expression normalized to GAPDH
0
2
4
6
8
10
*
***
**
Vehicle
TGFβ1
LY294002
TGFβ1 + LY294002

In contrast, knockdown of FOXC2 resulted in an increase in S100A10 expression (supplemental figure 6a). FOXC2 has been described as a crucial transcription factor for the development of lymphatic vessels during embryogenesis by promoting EMT. Under our conditions, FOXC2 increased N-cadherin (figure 25a, 25d) and decreased E-cadherin expression in A549 cells [584], consistent with EMT activation, arguably through non-canonical TGFβ1 signaling PI3K. This is also consistent with the fact that N-cadherin was downregulated upon inhibition of PI3K by LY294002 (supplemental figure 5a).

Since PI3K inhibition increases S100A10 expression, we examined whether the downstream inhibitory effect of FOXC2 on S100A10 can abrogate the S100A10 increase. Indeed, the expression of FOXC2 sustained the downregulation of S100A10 in the presence of LY294002 suggesting that PI3K signaling downregulates S100A10 through a FOXC2-dependent mechanism (figure 25d). Similarly, serum withdrawal that normally upregulates S100A10 failed to do so when FOXC2 was expressed (figure 25e). FOXC2 also maintained S100A10 downregulation in the presence of TGFβ1 (supplemental figure 6b). Together, these results indicate that S100A10 expression is positively modulated by canonical Smad-dependent TGFβ1 signaling and negatively by growth factor signaling pathways such as PI3K/mTOR via a FOXC2-dependent mechanism.

Figure 25. PI3K suppresses S100A10 expression through a FOXC2-mediated mechanism. Western blot analysis (a), S100A10 protein quantification (b) and S100A10 mRNA quantification (c) of pBabe-control and pBabe-FOXC2 A549 cells. Cells were transfected with pBabe vector to express FOXC2. Western blot of pBabe control and pBabe FOXC2 A549 cells treated with LY294002 (d) or serum starved for four consecutive days (e).


a
pBabe control
pBabe FOXC2
N-cadherin
Foxc2
β-actin
S100A10
b
Relative protein expression normalized to β-actin
1.0
0.8
0.6
0.4
0.2
0.0
**
pBabe Control
pBabe FOXC2
c
Relative mRNA expression normalized to GAPDH
1.5
1.0
0.5
0.0
**
pBabe Control
pBabe FOXC2
d
pBabe control
pBabe FOXC2
LY294002
-
+
-
+
N-cadherin
p-S6K
β-actin
Foxc2
S100A10
S100A10 quantification 1.00 2.10 0.38 0.84
e
pBabe control
pBabe FOXC2
Serum starvation (days)
0 1 2 3 4 0 1 2 3 4
β-actin
S100A10
S100A10 quantification 1.00 1.02 1.72 1.23 1.83 0.57 0.59 0.70 0.81 0.81

3.6 S100A10 serves as a plasminogen receptor at the surface of A549 cells.

Since S100A10 is a well-established plasminogen receptor [259], we examined how surface levels of S100A10 modulate plasminogen activation. We first compared total and surface S100A10 levels between BEAS-2B and A549 cells using flow cytometry. Both total (figure 26a, 26b) and surface (figure 26c) S100A10 protein expression were significantly higher in A549 cells compared to BEAS-2B cells. The difference in S100A10 expression was concomitant with differences in the ability of these cells to activate plasminogen (figure 26d) where A549 cells had a 10-fold higher capacity to activate plasminogen. ε-aminocaproic acid (ACA), a lysine analog, serves as a plasminogen activation inhibitor via inhibiting plasminogen binding to its receptors. ACA treatment completely abolished plasminogen activation indicating that plasminogen binding to plasminogen receptors is the rate limiting step under these conditions. In addition, we depleted S100A10 in both cell lines using a stable shRNA knockdown (figure 27a, 27c). The depletion reduced plasminogen activation by 45% at the cell surface of A549 cells compared to the scramble control (figure 27b). The remaining 55% was likely contributed by other plasminogen receptors (figure 27b). To avoid any compensation mechanisms upon stable shRNA knockdown, transient siRNA knockdown (supplemental figure 7a) of S100A10 in A549 cells was performed and resulted in a similar reduction in plasminogen activation (Supplemental figure 7b). In contrast, S100A10 depletion using shRNA (figure 27c) or siRNA (supplemental figure 7c) in BEAS-2B cells did not decrease plasminogen activation compared to the scramble control which could be partly attributed to the low baseline surface plasminogen activation rate (figure 27d, supplemental figure 7d). Additionally, ACA treatment did not completely abolish activation suggesting a low

expression of plasminogen receptors at the cell surface (figure 27d, supplemental figure 7d). The latter was concomitant with low surface expression of S100A10 (figure 26b, 26c). These findings suggest that S100A10 surface expression is crucial for maintaining the activation of plasminogen. However, whether any manipulations of S100A10 levels by TGFβ1 in A549 cells or by serum-withdrawal in A549 and BEAS-2B cells can affect plasminogen activation were yet to be addressed.

Figure 26. Total and surface S100A10 levels and significantly elevated in A549 compared to BEAS-2B cells concomitant with enhanced plasminogen activation. (a) western blot analysis and (b) quantification of total S100A10 protein and (c) flow cytometry of surface S100A10 levels in A549 and BEAS-2B cells. (d) Plasminogen activation assay of A549 and BEAS-2B cells in the presence of the lysine mimetic ε-aminocaproic acid (ACA) and protease inhibitor aprotinin (Ap).

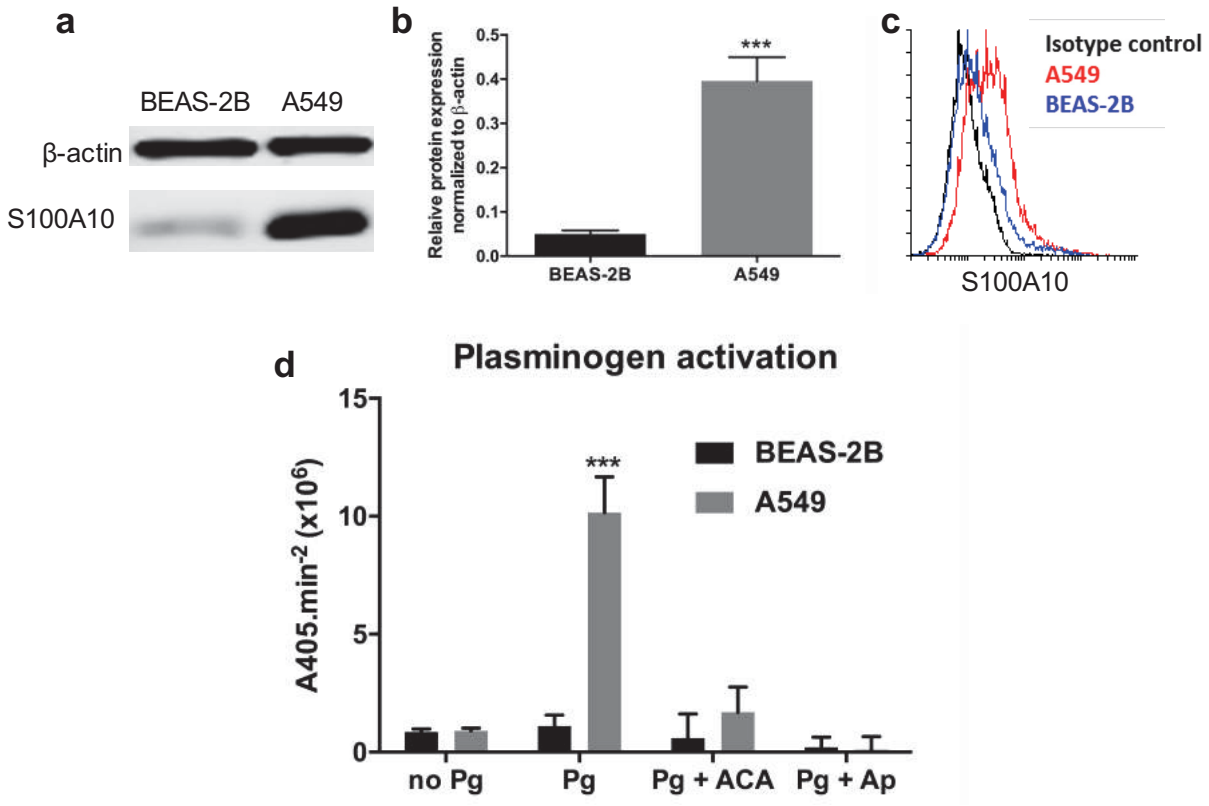
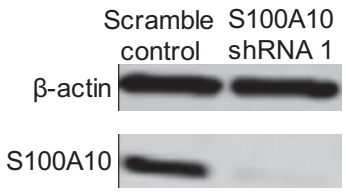
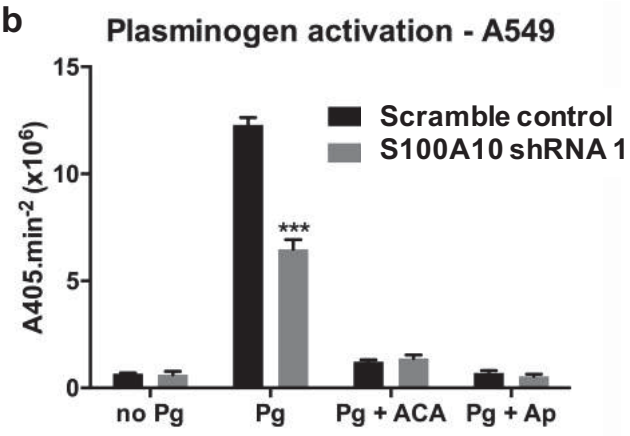
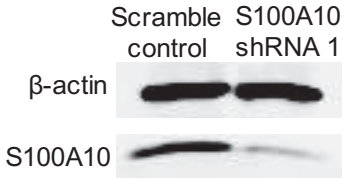
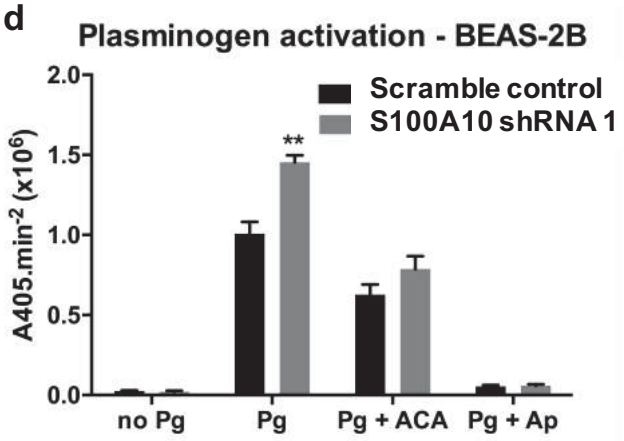

a
BEAS-2B
A549
β-actin
S100A10
b
Relaive protein expression normalized to β-actin
0.5
0.4
0.3
0.2
0.1
0.0
***
BEAS-2B
A549
c
Isotype control
A549
BEAS-2B
S100A10
d
Plasminogen activation
A405.min-2 (x10^6)
15
10
5
0
***
BEAS-2B
A549
no Pg
Pg
Pg + ACA
Pg + Ap

Figure 27. S100A10 depletion reduces plasminogen activation in A549 cells but not in BEAS-2B cells. (a) Western blot analysis of total S100A10 protein in scramble and S100A10-depleted (S100A10 shRNA 1) A549 cells. (b) Plasminogen activation assay of A549 scramble control and S100A10 shRNA 1 A549 cells in the presence of the lysine mimetic ε-aminocaproic acid (ACA) and protease inhibitor aprotinin (Ap). (c) western blot analysis of total S100A10 protein in scramble and S100A10-depleted (S100A10 shRNA 1) BEAS-2B cells. (d) Plasminogen activation assay of A549 scramble control and S100A10 shRNA 1 BEAS-2B cells.


a
Scramble control
S100A10 shRNA 1
β-actin
S100A10


b
Plasminogen activation - A549
Scramble control
S100A10 shRNA 1
A405.min-2 (x10^6)
15
10
5
0
***
no Pg
Pg
Pg + ACA
Pg + Ap

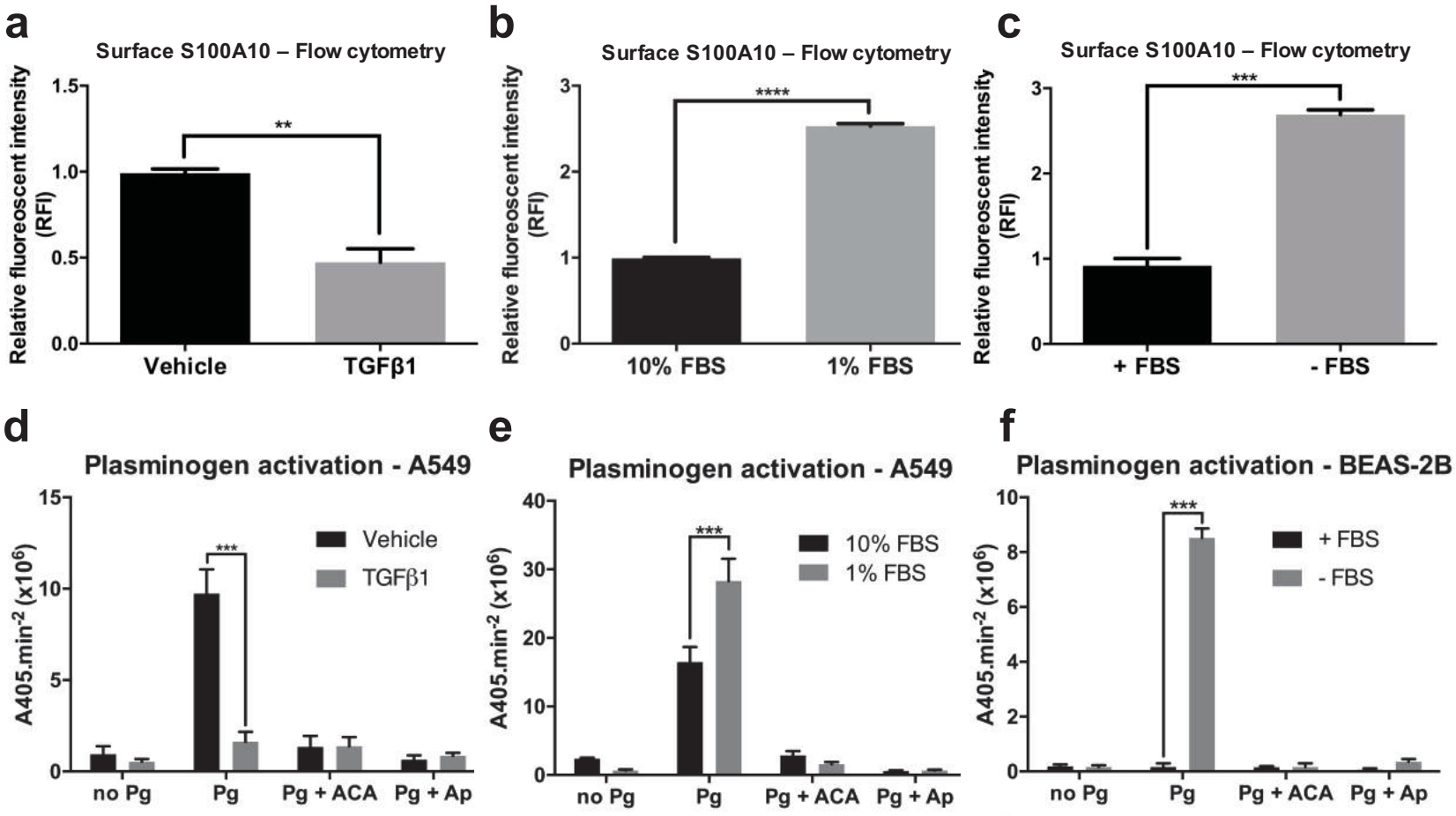

c
Scramble control
S100A10 shRNA 1
β-actin
S100A10


d
Plasminogen activation - BEAS-2B
Scramble control
S100A10 shRNA 1
A405.min-2 (x10^6)
2.0
1.5
1.0
0.5
0.0
**
no Pg
Pg
Pg + ACA
Pg + Ap

3.7 Mesenchymal cells downregulate S100A10 surface expression and demonstrate a low capacity to activate plasminogen.

Our results suggested that S100A10 is differentially expressed in response to TGFβ1 or serum withdrawal and is independent of the epithelial/mesenchymal phenotype of cells. We then examined how induction of epithelial- and mesenchymal-like phenotypes affects plasminogen activation in both A549 and BEAS-2B cells especially in terms of S100A10 surface expression. Surprisingly and despite the upregulation of total S100A10 levels upon TGFβ1 treatment of A549 cells (figure 21d), there was a decrease in S100A10 levels at the cell surface as demonstrated using flow cytometry (figure 28a, supplemental figure 8a, 8b) and surface biotinylation (supplemental figure 8d, 8e). Importantly, the decrease in S100A10 surface expression was concomitant with complete loss of plasminogen activation (figure 28d) which was predictably not affected by further S100A10 knockdown (supplemental figure 8f). In contrast, serum withdrawal of A549 cells that increased total S100A10 protein expression, resulted in an increase in surface expression of S100A10 (figure 26b) and concomitant increase in plasminogen activation (figure 26e). Similarly, the withdrawal of serum from restored/increased plasminogen activation at the cell surface of A549 (figure 28e) and BEAS-2B (figure 28f, supplemental figure 8g) cells, concomitant with increases in surface S100A10 levels (figure 28b, 28c respectively). Collectively, these results suggested that mesenchymal cells possess a low capacity to activate plasminogen, which is partly attributable to low surface S100A10 levels.

Figure 28. Plasminogen activation is partially dictated by the surface localization of plasminogen receptor S100A10 and not by the mesenchymal/epithelial state of A549 and BEAS-2B cells. Flow cytometry analysis/quantification of surface S100A10 expression and plasminogen activation upon TGFβ1 treatment in A549 cells (a, d), serum withdrawal in A549 cells (b, e) and serum withdrawal in BEAS-2B cells (c, f).


a
Surface S100A10 – Flow cytometry
Relative fluoreoscent intensity (RFI)
1.5
1.0
0.5
0.0
**
Vehicle
TGFβ1
b
Surface S100A10 – Flow cytometry
Relative fluoreoscent intensity (RFI)
3
2
1
0
****
10% FBS
1% FBS
c
Surface S100A10 – Flow cytometry
Relative fluoreoscent intensity (RFI)
3
2
1
0
***
+ FBS
- FBS
d
Plasminogen activation - A549
A405.min-2 (x10^6)
15
10
5
0
***
Vehicle
TGFβ1
no Pg
Pg
Pg + ACA
Pg + Ap
e
Plasminogen activation - A549
A405.min-2 (x10^6)
40
30
20
10
0
***
10% FBS
1% FBS
no Pg
Pg
Pg + ACA
Pg + Ap
f
Plasminogen activation - BEAS-2B
A405.min-2 (x10^6)
10
8
6
4
2
0
***
+ FBS
- FBS
no Pg
Pg
Pg + ACA
Pg + Ap

3.8 S100A10 and uPAR-mediated plasminogen activation is potentially masked by marked PAI-1 upregulation.

The low rate of plasminogen activation in TGFβ1-treated A549 cells and serum-supplemented BEAS-2B cells was unlikely to be entirely attributable to the decrease in S100A10 surface levels. Indeed, cells possess multiple plasminogen receptors[195] that contribute to plasminogen activation but are not necessarily regulated by the epithelial or mesenchymal state of cells. This is further supported by the fact that S100A10 depletion in A549 cells only resulted in a 45% decrease in plasminogen activation (figure 27b). In an attempt to understand the contribution of other components of the plasminogen activation system, we focused on the remaining significantly-upregulated genes (figure 21a) (supplemental table 1). *PLAUR* (uPAR, 9.64-fold) and *SERPINE1* (PAI-1, 835-fold) were of most interest considering their dramatic upregulation and direct involvement in binding and inhibiting the plasminogen activator uPA respectively. We first confirmed uPAR (figure 29a, 29b) and PAI-1 (figure 29a, 29c) upregulation in TGFβ1-treated A549 cells. In contrast, uPAR was upregulated (figure 29d, 29e) while PAI-1 was downregulated (figure 29d, 29f) in BEAS-2B cells upon withdrawal of serum, consistent with the increase in plasminogen activation). The dramatic upregulation of PAI-1 by TGFβ1 was inhibited by A83-01 treatment (figure 29g) and abrogated by Smad4 knockdown (figure 29h). PAI-1 upregulation was also concomitant with decrease in surface S100A10 levels, together contributing to the low rate of plasminogen activation on the surface of TGFβ1-treated A549 cells.

Since PAI-1 is a potent inhibitor of plasminogen activation, we assessed whether the inhibition of PAI-1 can rescue plasminogen activation in TGFβ1-treated A549 cells and serum-supplemented BEAS-2B. We treated these cells with the PAI-1 inhibitor tiplaxtinin (figure 30a). Only partial inhibition (45%) of PAI-1 was achieved with minimal cellular toxicity which might be attributed to plasminogen-independent functions of PAI-1 in cell survival[585]. Nonetheless, tiplaxtinin increased plasminogen activation in vehicle-treated cells and could restore some activation in TGFβ1-treated A549 cells (figure 30b). BEAS-2B cells treated with tiplaxtinin showed a similar but less dramatic increase in plasminogen activation (figure 30c). These results indicate that PAI-1 upregulation in mesenchymal cells greatly contributed to quenching global plasminogen activation.

Figure 29. S100A10, uPAR and PAI-1 are altered by TGFβ1 in A549 cells and serum withdrawal in BEAS-2B cells. (a) Western blot analysis and quantification of uPAR (b) and PAI-1 (c) in vehicle-treated and TGFβ1-treated A549 cells. (d) Western blot analysis and quantification of uPAR (e) and PAI-1 (f) in serum-supplemented and serum-starved BEAS-2B cells. Western blot analysis of PAI-1 in A549 cells either treated with A83-01 (g) or depleted of SMAD4 (h) in the presence or absence of TGFβ1.

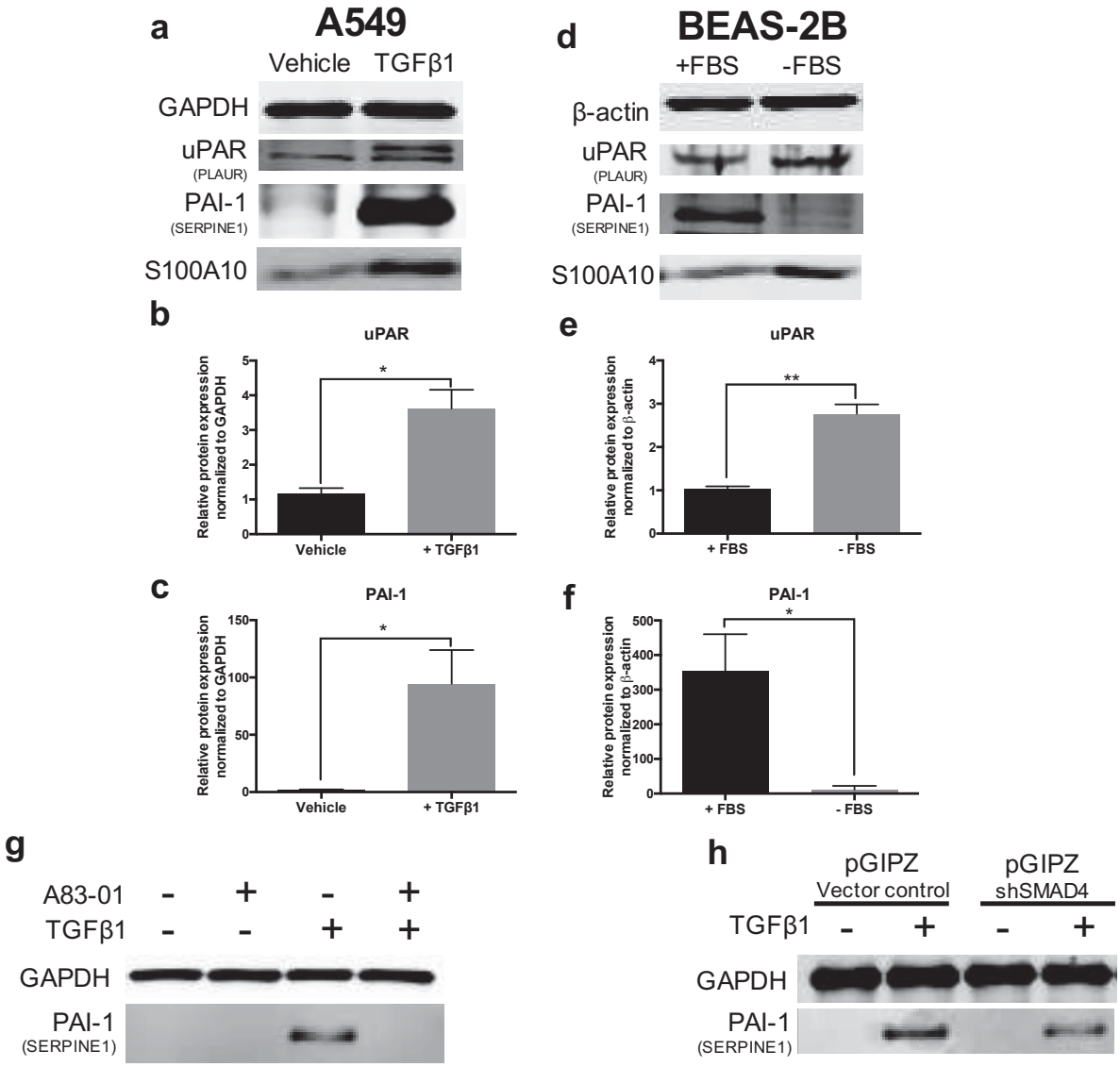
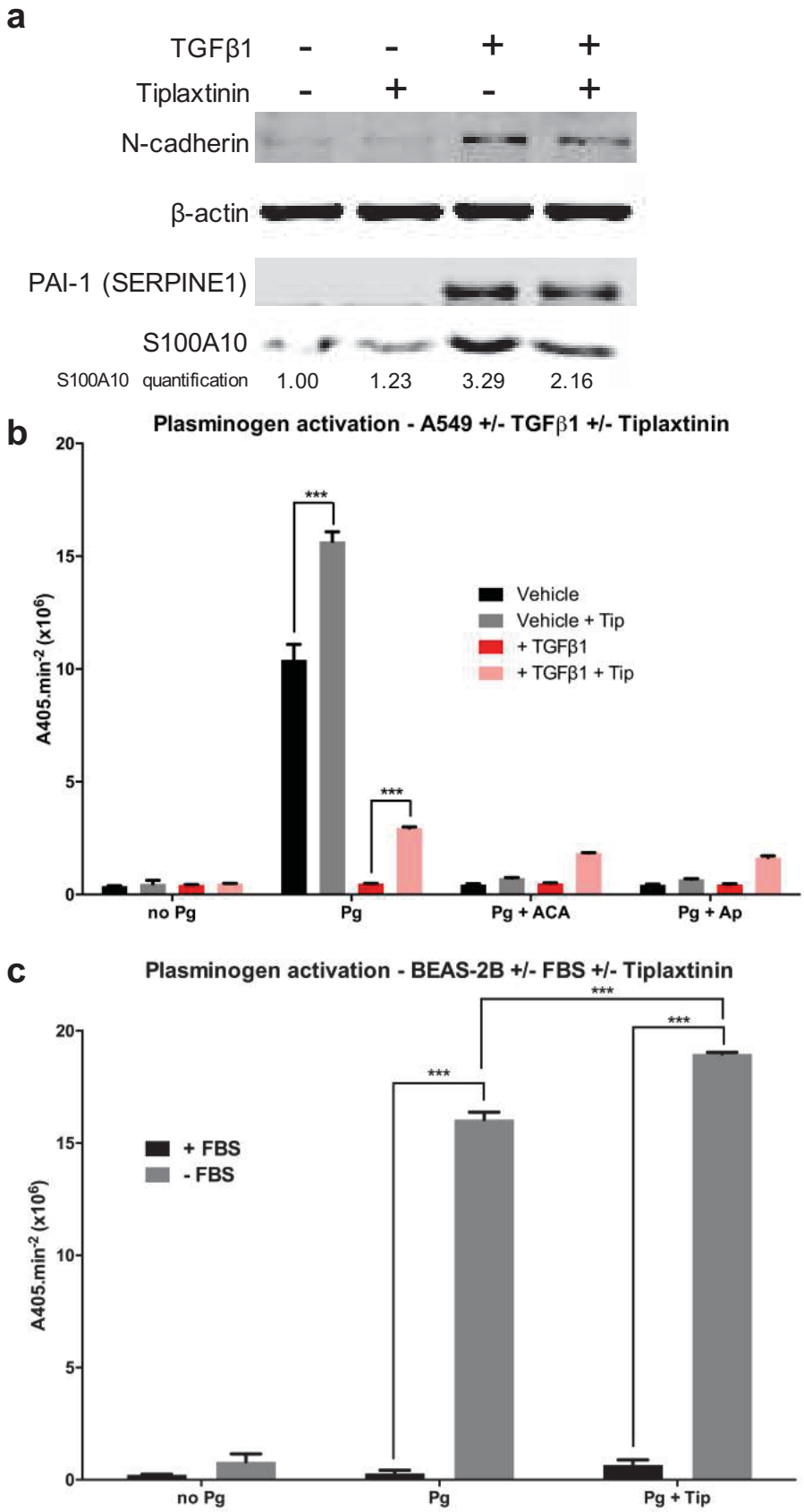

a
A549
Vehicle
TGFβ1
GAPDH
uPAR
(PLAUR)
PAI-1
(SERPINE1)
S100A10
d
BEAS-2B
+FBS
-FBS
β-actin
uPAR
(PLAUR)
PAI-1
(SERPINE1)
S100A10
b
uPAR
Relative protein expression normalized to GAPDH
Vehicle
+ TGFβ1
e
uPAR
Relative protein expression normalized to β-actin
+ FBS
- FBS
c
PAI-1
Relative protein expression normalized to GAPDH
Vehicle
+ TGFβ1
f
PAI-1
Relative protein expression normalized to β-actin
+ FBS
- FBS
g
A83-01
TGFβ1
GAPDH
PAI-1
(SERPINE1)
h
pGIPZ
Vector control
pGIPZ
shSMAD4
TGFβ1
GAPDH
PAI-1
(SERPINE1)

Figure 30. Partial Inhibition of PAI-1 restores plasminogen activation in TGFβ1-treated A549 cells and serum-supplemented BEAS-2B cells. (a) Western blot analysis of A549 cells treated with PAI-1 inhibitor tiplaxtinin (10 **μ**M) in the presence and absence of TGFβ1. Plasminogen activation assay of A549 cells in the presence of TGFβ1 (b) and BEAS-2B cells (c) treated with tiplaxtinin.


a
TGFβ1 - - + +
Tiplaxtinin - + - +
N-cadherin
β-actin
PAI-1 (SERPINE1)
S100A10
S100A10 quantification 1.00 1.23 3.29 2.16
b
Plasminogen activation - A549 +/- TGFβ1 +/- Tiplaxtinin
A405.min-2 (x10^6)
Vehicle
Vehicle + Tip
+ TGFβ1
+ TGFβ1 + Tip
***
no Pg
Pg
Pg + ACA
Pg + Ap
c
Plasminogen activation - BEAS-2B +/- FBS +/- Tiplaxtinin
A405.min-2 (x10^6)
+ FBS
- FBS
***
no Pg
Pg
Pg + Tip

CHAPTER 4: DISCUSSION of chapter 3

4.1 Discussion

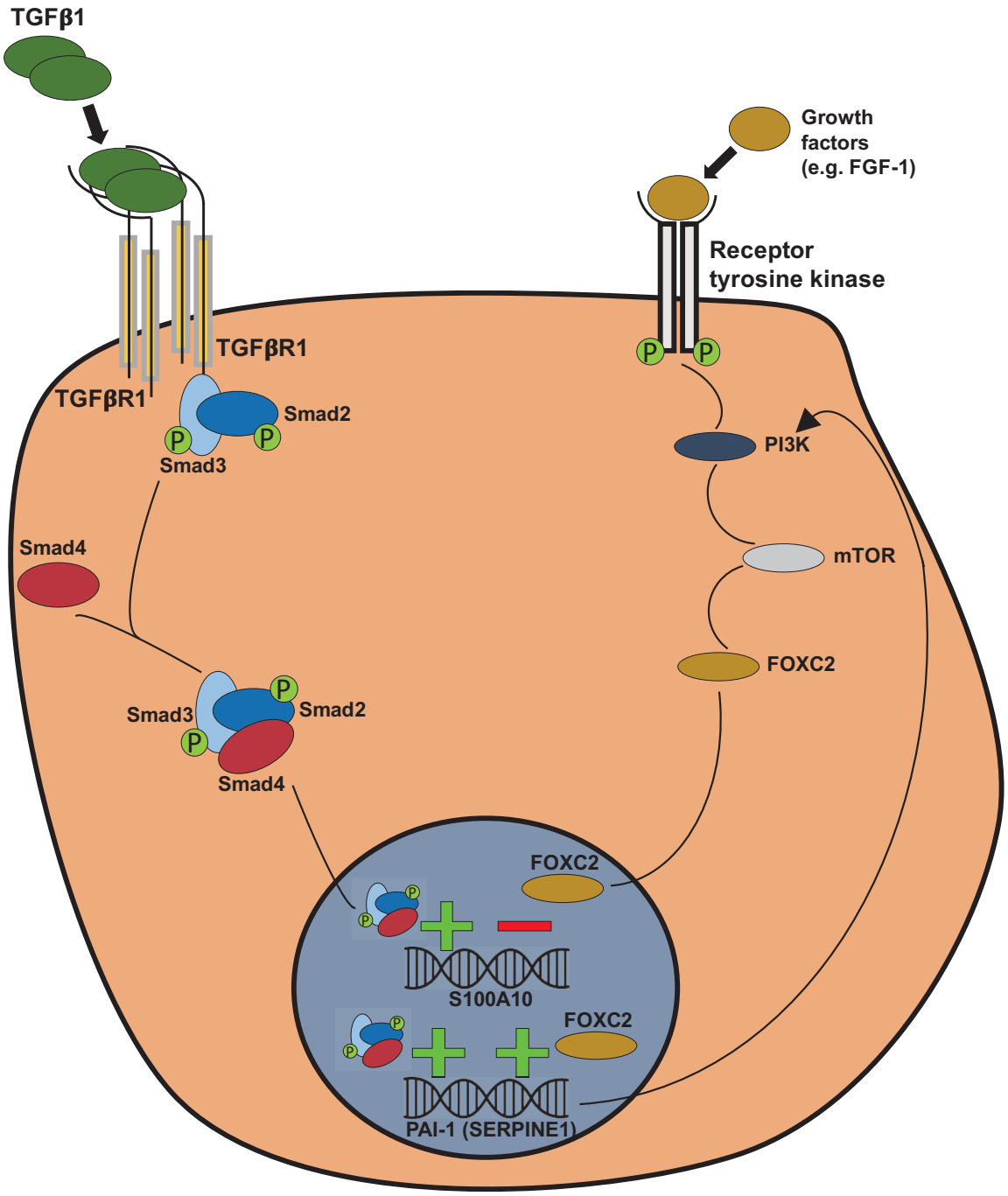
EMT and MET represent a continuum of cellular changes which provide cells with an ability to transition between epithelial and mesenchymal phenotypes. During malignancy, a select population of cancer cells, the origin of which remains elusive, can acquire the ability to undergo EMT and arguably metastasize [562]. Once cells reach a prospective metastasis site, they must implant and populate a clinically-distinguishable tumor site. A prerequisite step for cancer cells undergoing EMT is to degrade the underlying ECM and basement membrane. ECM degradation during EMT has been primarily linked to enhanced production of MMPs. For instance, Twist1 expression in HMLE cells increased MMP-dependent proteolysis [107]. Similarly, Snail1 expression in MCF-7, MDA-MB-231 [586] and MDCK [587] cells upregulated MMP9, MMP14 (MT1-MMP) and MMP15 (MT2-MMP) expression and enhanced matrix proteolysis. Meanwhile, the role of the serine protease plasmin in ECM proteolysis during EMT has never been addressed. This is important since most metalloproteinases are translated in their inactive pro-MMP form and require activation [207]. Plasmin is a potent physiologic activator of many pro-MMPs including MMP2 [588] and MMP9 [589] both of which are well-characterized drivers of cancer cell invasion [590]. Plasmin is also required for MMP2- and MMP9-dependent ECM degradation and cellular invasiveness [591]. Nonetheless, the role of plasmin and the proteins that regulate its production has never been addressed in cells undergoing EMT/MET.

The fact that mesenchymal cells are more likely to escape primary tumors does not necessitate that these same cells will give rise to metastatic growth. Indeed, the recent advent of mouse models that allow EMT lineage tracing of tumor cells has offered new insights into the role of EMT in metastasis *in vivo*. A 2015 report by Fischer *et al*. demonstrated that epithelial and not mesenchymal forms of cancer cells were largely responsible for lung metastases formation in breast cancer. Instead, EMT contributed to resistance to the chemotherapeutic agent cyclophosphamide [592]. Similarly, Zheng *et al*. reported that EMT induced by Twist and Snail transcription factors was dispensable for metastasis in a mouse model of pancreatic cancer [593]. A 2014 report also demonstrated that expression of E-cadherin, loss of which is considered a hallmark of EMT, has been shown to increase invasiveness of cancer cells *in vitro* [594]. These studies challenge previous notions that claim that mesenchymal cells in primary tumors are solely responsible for the dissemination process that initiates metastasis. EMT-dependency and metastasis have become matters of contention primarily due to their context-dependency.

Here we demonstrate that in cells undergoing TGFβ1-induced EMT, a select group of plasminogen activation proteins are differentially activated. For instance, S100A10 was the only differentially expressed plasminogen receptor that was regulated by TGFβ1 through a Smad4-dependent mechanism. Canonical TGFβ1 signaling involves the activation of Smad2 and/or Smad3 which will then form trimeric complex with Smad4 [595](figure 31). Smad4 is an integral part of canonical TGFβ1 signaling and is required for the induction of EMT. In fact, Smad4 deletion abrogates TGFβ1-induced upregulation of N-cadherin (figure 22c) and is associated with a decrease in Snail, CD31 and VE-cadherin expression and an increase in α-SMA and FSP1 expression [596]. For that

purpose, the dependency of S100A10 upregulation on the expression of wild-type Smad4 was manifested in the absence of a response in the pancreatic cell line BxPC-3 which harbors SMAD4 homozygous deep deletions. In addition, Ali *et al*. utilized mass spectrometry to demonstrate that reactivation of mutant Smad4 in HCT116 colorectal cancer cells upregulates a series of proteins including S100A2, FSP-1, S100A10 (p11) and S100A11 [597]. The question whether the S100A10 promoter or any intragenic sequences contain a SMAD4 binding locus is not known. However, a recent report by Kennedy *et al*. applied CHIP-seq genome-wide screen to identify sequences that are bound by SMAD4 only upon stimulation by TGFβ1 in A2780 ovarian cancer cells. The analysis demonstrated that SMAD4 bound the 3' distal region around 21.009kb of *S100A10* transcription start site[598] (supplemental figure 10, supplemental table 2).

Figure 31. S100A10 and PAI-1 are regulated by Smad4-dependent TGFβ1-mediated signaling and FOXC2-mediated PI3K signaling. The model illustrates that the treatment of epithelial cells with TGFβ1 increases S100A10 mRNA and protein levels through canonical Smad-dependent TGFβ1 signaling. S100A10 is also affected by the pro-growth PI3K pathway. Serum starvation, PI3K inhibition or mTOR inhibition upregulate S100A10 expression suggesting an inhibitory effect through this pathway. The transcription factor FOXC2, which is downstream of PI3K, mediates the repression of S100A10 expression.


TGFβ1
Growth factors (e.g. FGF-1)
Receptor tyrosine kinase
TGFβR1
TGFβR1
Smad2
Smad3
P
P
P
P
PI3K
mTOR
FOXC2
Smad4
Smad3
Smad2
Smad4
FOXC2
S100A10
FOXC2
PAI-1 (SERPINE1)

In addition to canonical Smad-mediated signaling, TGFβ1 activates non-canonical pathways through PI3K, MAPK and Rho-like GTPases. The pro-growth Akt/PI3K has been previously demonstrated to either negatively or positively complement the biological and morphological changes associated with EMT [345]. For instance, PI3K and Akt inhibited apoptosis induced by TGFβ1 via the interaction of Akt with Smad3 preventing Smad3 phosphorylation and its subsequent translocation to the nucleus in Hep3B and HEK293T cells [599][600]. mTOR inhibition alleviated the inhibitory effect of Akt on Smad3 activity [601]. S6K phosphorylation also hindered the inhibitory effect of TGFβ1 on cell growth [602]. In certain cell models including those described in this study, the cross-talk between TGFβ1 and PI3K signaling pathways produced antagonist effects. In A549 cells, the inhibition of PI3K/mTOR or the withdrawal of serum in the presence of TGFβ1 increased S100A10 expression partly due to direct Smad signaling as well as alleviating the inhibition of Smads by PI3K (figure 24a-24d). Evidently, the activation of PI3K by FGF-1 prevented the upregulation of S100A10 by TGFβ1 (figure 24e). This indicated that S100A10 is directly repressed by PI3K and induced by TGFβ1 or by alleviating PI3K-mediated inhibition of canonical TGFβ1/Smad signaling.

In other cell models, PI3K and TGFβ1 yield complementary effects. Indeed, the activation of PI3K by TGFβ1 can be mediated through Akt phosphorylation followed by activation of mTORC1 (mammalian TOR complex 1) and mTORC2 in the murine breast epithelial cell line NMuMG [603][521]. The latter represents a classic EMT model where the inhibition of PI3K hinders TGFβ1-induced EMT and mTORC1 was found to be important for cancer cell invasion and migration while mTORC2 was necessary for the transition from an epithelial to a mesenchymal phenotype [521]. TGFβ1-induced activation

of mTOR led to enhanced phosphorylation of S6K and 4E-BP1 in HaCAT keratinocytes and NMuMG cells. Inhibition of PI3K or inactivation of Akt abrogated TGFβ1-mediated activation of mTOR. Inhibition of mTOR also resulted in decrease in cellular migration and invasiveness but did not affect the acquisition of the mesenchymal phenotype which is likely induced via canonical smad signaling [603]. To that purpose and in our model systems, N-cadherin upregulation was primarily regulated by canonical Smad-dependent TGFβ1 signaling but was also regulated by PI3K signaling in A549 cells. Inhibition of PI3K by LY294002 (figure 24b) or serum withdrawal (figure 24a) reduced N-cadherin expression, an effect that was also achieved by TGFβR1 inhibition (figure 22a) or Smad4 depletion (figure 22c) in TGFβ1-treated cells.

Whether the dependency of TGFβ1-induced EMT on PI3K activation is a universal mechanism remains elusive and is highly context-specific [604]. Some earlier evidence suggested that the PI3K-dependency is present in systems where TGFβ1-mediated signaling was not reliant on Smads to downregulate E-cadherin and upregulate N-cadherin as seen in NMuMG cells [358]. In addition, treating NMuMG cells with TGFβ1 resulted in downregulation of S100A10 expression consistent with the PI3K dependency in this cell line (supplemental figure 9a, 9b). Notably, the modulation of S100A10 expression was not linked to N-cadherin expression indicating that S100A10 is a TGFβ1- and PI3K-regulated gene and not an “EMT gene”. This becomes more evident in BEAS-2B cells where serum withdrawal, known to diminish PI3K signaling, induced an epithelial-like morphology and increased S100A10 expression (figure 23c, 23d). These results are consistent with the idea of uncoupling EMT from S100A10 expression and vice versa.

The suppression of S100A10 by PI3K was likely mediated through a FOXC2-dependent mechanism (figure 25a, 25b) (figure 31). The transcription factor FOXC2 belongs to the forkhead-box family of transcription factors and is required for the maturation of the primary lymphatic plexus into collecting lymphatic vessels during embryonic development [581]. FOXC2 has also been implicated in oncogenic progression [605] and in promoting EMT and downregulating E-cadherin expression in breast cancer cells [606]. Yu *et al*. recently demonstrated that FOXC2 expression in A549 cells is driven by PI3K signaling and not by canonical TGFβ1 signaling [584] (figure 25d). In fact, FOXC2 overexpression in A549 cells treated with the anti-proliferative inhibitor LY294002 partially restored their growth capability (supplemental figure 5c) confirming that FOXC2 is indeed downstream of PI3K in A549 cells. The regulation of S100A10 by FOXC2 occurred at the transcriptional level where FOXC2 overexpression suppressed S100A10 mRNA levels (figure 25c). Whether FOXC2 can directly bind the *S100A10* gene promoter is yet to be addressed. Norrmén *et al*. utilized CHIP-chip analysis to generate a genome-wide map of FOXC2-binding sites. The FOXC2 motif *GCCAACAAAAACA* was present in the promoter region of the *S100A10* gene upstream of the transcription start site [581]. However, whether FOXC2 can directly bind upstream of the *S100A10* gene remains to be addressed.

Since S100A10 was the only plasminogen receptor to be differentially regulated by TGFβ1, we tested if the regulation of S100A10 under epithelial and mesenchymal states influenced plasminogen activation. The depletion of S100A10 in A549 cells resulted in marked decrease in plasminogen activation, which is likely justified by an adequate level of S100A10 expression at the cell surface (figure 27a). However, in the context of EMT,

the impact of S100A10 expression on plasminogen activation was not linked to the epithelial or mesenchymal state of the cell but rather to the surface expression of S100A10. Despite the upregulation of total S100A10 expression by TGFβ1, plasminogen activation was dramatically reduced (figure 28d), which is associated with lower surface S100A10 expression. Similarly, serum-supplemented BEAS-2B cells possessed a limited capability to activate plasminogen, which could be restored when an epithelial phenotype was induced (figure 28f). In addition, serum withdrawal of A549 cells increased plasminogen activation (figure 28e). Serum-starved A549 cells may represent a more epithelial state of A549 cells evident by E-cadherin expression (figure 20d). Dong Su *et al*. demonstrated that the epithelial-like morphology of A549 caused by serum withdrawal was mediated via c-src activation and subsequent upregulation of E-cadherin [567]. Our findings indicated the first association between the epithelial and mesenchymal state of cells and their differential capacity to activate plasminogen (figure 32).

An interesting observation emerged in which an increase in plasminogen activation occurred upon S100A10 knockdown (both shRNA and siRNA) in BEAS-2B cells (figure 27d, supplemental figure 7d). Although the lack of a decrease is potentially attributed to the low overall rate plasminogen activation at the cell surface, the observed increase may potentially implicate a novel process by which plasminogen activation is compensated for by other regulators of plasminogen. Although the concept of compensation among plasminogen receptors is novel, it may support the possibility that build-in redundancy and/or compensation is/are part of the rescue mechanisms by which cells and tissue systems maintain homeostasis despite a defect in one of these regulators. Evidence of such redundancy is seen mice lacking either tPA or uPA which do not display any of the major

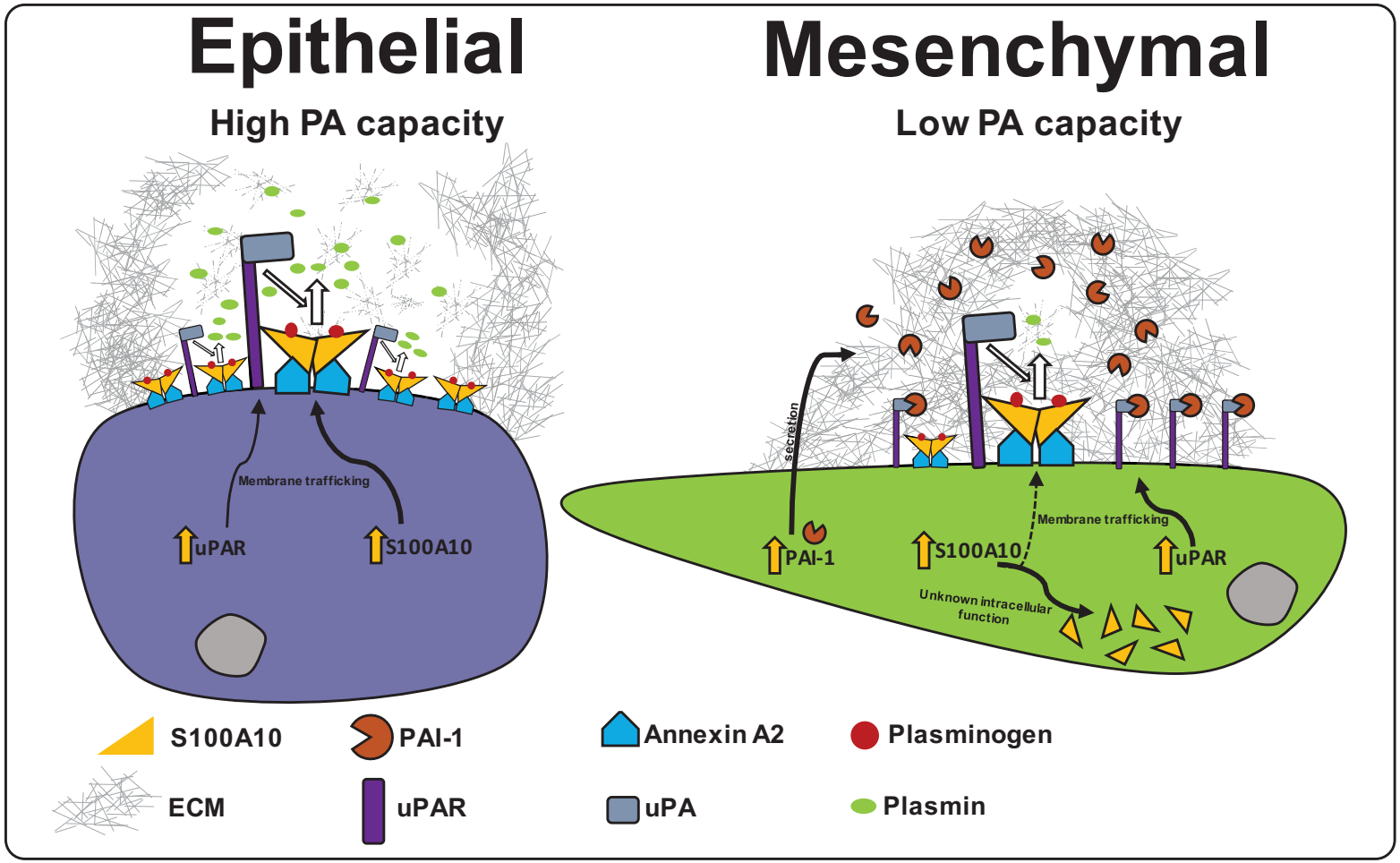
organ pathologies (e.g. tissue repair) seen in mice lacking both [607]. Additionally, uPA is normally expressed at low levels in the central nervous system and appears not to contribute to the physiological activation of plasminogen which is mostly driven by tPA [608].

The lack of plasminogen activation in mesenchymal A549 and BEAS-2B cells (figure 28d, 28f respectively) could not solely be explained by the low surface levels of S100A10 since S100A10 depletion only yielded a 45% decrease in plasminogen activation in A549 cells (figure 27b, 27d). This suggested the involvement of other components of the plasminogen activation system with focus on uPAR and PAI-1. Even though the expression of proteins involved in plasminogen activation have been reported, the interplay between these proteins has never been addressed particularly how they collectively contribute to plasminogen activation. We report that both uPAR and PAI-1 were markedly induced by TGFβ1 in A549 cells (figure 29a). PAI-1 was likely the major contributor to quenching plasminogen activation (figure 32) since its inhibition partially restored plasminogen activation in A549 (figure 30b) and BEAS-2B (figure 30c) cells. In contrast, uPAR and S100A10 upregulation coupled with PAI-1 downregulation contributed to the drastic increase in BEAS-2B cells upon serum withdrawal (figure 32). Interestingly, both uPAR signaling and PAI-1 expression have been shown to be required for activation of EMT in breast cancer cells [609] and fibroblasts [289] respectively. It is possible that TGFβ1-mediated activation of EMT was further compounded by the concurrent activation of PAI-1 and uPAR. In that context, S100A10 expression was downregulated when PAI-1 was inhibited (figure 30a). The plasminogen-independent function of PAI-1 in EMT could be explained by its interaction with LRP1 [585][610], through PAI-1-mediated activation of PI3K/Akt signaling [611] and/or activation of erk1/2 [612]. Zhang *et al*. showed that

transfection of lung mouse fibroblasts with PAI-1 siRNA inhibited the phosphorylation of erk whereas PAI-1 overexpression increased erk phosphorylation [612]. Interestingly, the PAI-1 inhibitor SK-216 did not alter phosphorylation of erk and Smad2 in A549 cells treated with TGF-β. However, SK-216 inhibited mRNA expression of the EMT-ATFs Slug and Snail.

Interestingly, FOXC2 which downregulates S100A10, was reported to be linked to higher plasma levels of PAI-1 and TGFβ1 during intravascular thrombosis [613]. In 2006, Fujita *et al*. demonstrated that FOXC2 binds upstream of SERPINE1 (PAI-1) in response to TGFβ1 (through Smads) or to insulin (through PI3K) in bovine and human endothelial cells [582][583](supplemental figure 5d). Rømer *et al*. demonstrated that PAI-1 protects murine fibrosarcoma cells from etoposide toxicity via activation of PI3K pathway[611] which can potentially contribute to S100A10 repression. Whether PAI-1 regulates S100A10 through PI3K during EMT remains elusive (figure 31).

Figure 32. Proposed model of plasminogen activation at the cell surface of epithelial and mesenchymal cells. Although mesenchymal cells upregulate total S100A10 expression, epithelial cells express higher surface levels of S100A10 compared to mesenchymal cells. The latter are likely shuttling S100A10 for an unknown intracellular function. Similarly, both uPAR and PAI-1 are also upregulated in mesenchymal cells. PAI-1 release hinders plasminogen activation into plasmin by inhibiting uPAR-bound uPA. The decrease of plasmin generation reduces extracellular matrix (ECM) degradation. Noteworthy, S100A10 is expressed on the cell surface as part of the annexin A2-S100A10 heterotetramer.


Epithelial
High PA capacity
Mesenchymal
Low PA capacity
Membrane trafficking
uPAR
S100A10
secretion
PAI-1
S100A10
Membrane trafficking
uPAR
Unknown intracellular function
S100A10
PAI-1
Annexin A2
Plasminogen
ECM
uPAR
uPA
Plasmin

4.2 Study limitations and future directions

4.2.1 3D vs 2D models to study EMT

The use of 3D models to study EMT would offer an additional insight to the behavior of cancer cells within a multi-dimensional microenvironment. The above study utilized 2D culture systems which first, lack the supportive matrix and second offer little insight into the localization of proteolysis particularly sites of plasminogen activation. Bidarra *et al* recently developed an optimized soft alginate hydrogel embedded with cell adhesive RGD peptide. This matrix formulation supported epithelial growth and promoted conversion into a mesenchymal-like morphology in the presence of TGFβ1 [614]. The addition of fluorescence protease substrates (e.g. gelatin) into the 3D matrix will allow measurement of protease activity at the cell surface and importantly enable the localization of proteins such as S100A10, uPAR and PAI-1 using subsequent confocal microscopy. A recent report showed that HEY ovarian cancer cells treated with TGFβ1 have distinct gene expression profiles when grown in 3D cultures compared to 2D cultures. Genes such as the E-cadherin regulator CCDC80 were downregulated while others such as aldo-keto reductase AKR1C1 were drastically upregulated in TGFβ1-treated 3D cultures compared to TGFβ1-treated 2D cultures. Gene ontology analysis of altered genes showed enhanced tumorigenicity, amino acid metabolism and activated stress responses (e.g. hypoxia and nutrient scarcity). Interestingly, further analysis of differential gene expression identified a epigenetic cluster of genes which suggested that changes in methylation profiles might be responsible for differences between 2D and 3D cultures [615]. Therefore, it is essential to

complement the performed 2D studies with 3D cultures and re-assess the impact of TGFβ1 on plasminogen activation.

4.2.2 Global perspective on E/M phenotypes

The three proposed models of epithelial and mesenchymal cells (figure 20) can offer further insight into the distinctive characteristics of each phenotype beyond plasminogen activation. More specifically, analysis of surface proteins using biotinylation followed by mass spectrometry will allow identification and quantification of all surface proteins [616]. These proteins will generate a list of differentially-expressed proteins of which top "hits" can be individually studied and functionally tested. In addition, plasminogen-related proteins can be studied accordingly.

4.2.3 Effect of other EMT-ATFs on S100A10

Our current study delineated a crucial role of smad4 as a mediator of TGFβ1-induced upregulation of S100A10. Whether smad4 directly binds the S100A10 promoter remains elusive. The use of EMSA (electrophoretic mobility shift assay) will allow identification of whether smad4 can bind the S100A10 promoter. In addition, if the smad4 binding is valid, what other transcription factor associates with smad4 is yet to be determined. The impact of EMT-ATFs that are downstream of smad signaling on S100A10 expression was not elucidated in this dissertation (figure 14). Whether factors such as Snail, Slug, Twist and ZEB1/2 affect S100A10 expression is yet to be addressed.

4.2.4 Inter-dependency of S100A10, uPAR and PAI-1

Despite concomitant regulation of S100A10, uPAR and PAI-1 in epithelial and mesenchymal cells (figure 29), the dependency among these three proteins is of great interest. Our attempt to inhibit PAI-1 achieved partial inhibition along with downregulation of S100A10 and N-cadherin. Further PAI-1 knockdown will ensure the extent of S100A10 and N-cadherin dependency on PAI-1 expression. Similarly, whether the knockdown of S100A10 or uPAR affects PA1-1 is yet to be addressed. This particularly relevant since both uPAR and PAI-1 have been shown to be required for TGFβ1-inuced EMT [609] [289].

CHAPTER 5: THE PLASMINOGEN ACTIVATION PATHWAY IS UNIQUE TO NON-SMALL CELL LUNG CANCER

5.1 Study rationale

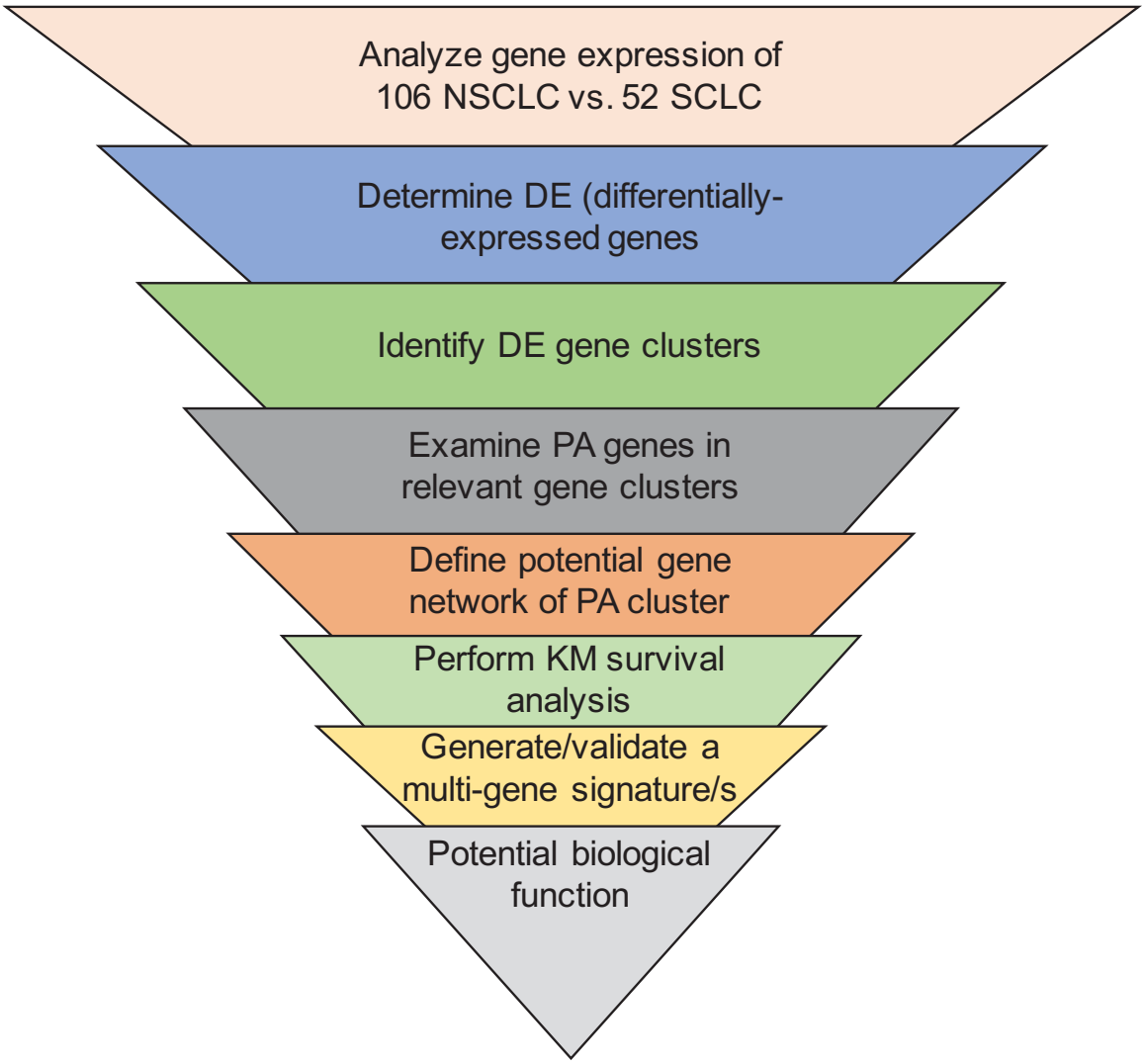
The involvement of various components of the plasminogen activation system in lung cancer cells (e.g. A549 and BEAS-2B) undergoing EMT prompted further investigation into their potential clinical implications. A549 and BEAS-2B cells respectively represent the adenocarcinoma and squamous cell-like subtypes of NSCLC (figure 19). Various early reports have implicated different components of the plasminogen system in determining NSCLC patient outcome particularly those related to uPA/uPAR-mediated activation of plasminogen and inhibition by PAI-1 [139]. In addition, a recent report also demonstrated that S100A10 expression (IHC) correlated with worsened prognosis, poor differentiation, higher TNM stage and increased severity and occurrence of intra-tumoral vascular invasion [617]. As discussed in the introduction, the distinction between SCLC and NSCLC is distinct as determined by its site of origin, histological morphology, biological behavior and risk factor correlations. However, differential gene expression between both lung cancer types have not been substantially addressed particularly in terms of the differential expression of components of the plasminogen activation system.

5.2 Developing a strategy to study PA genes in NSCLC

To assess the expression of genes that are part of the plasminogen activation system (henceforth referred to as PA genes), a multi-step hierarchical strategy was developed

(figure 33). First, z-scores of 16,215 genes from 52 SCLC and 106 NSCLC lung cell cancer lines from the CCLE (Cancer cell line encyclopedia) were downloaded from Cbioportal (figure 34). NSCLC cell lines exhibited 2,707 differentially-expressed (DE) compared to SCLC cell lines with at least a 2-fold change and an adjusted *p*-value of less than 0.01 (figure 35).

Figure 33. Schematic summary of the strategy used to generate outcome prediction models and gene signatures. Gene expression values were extracted from Cbioportal as z-scores. A total of 106 NSCLC and 52 SCLC cell lines were found based on the "histologic subtype" sorting criteria (see methods).


Analyze gene expression of
106 NSCLC vs. 52 SCLC
Determine DE (differentially-
expressed genes
Identify DE gene clusters
Examine PA genes in
relevant gene clusters
Define potential gene
network of PA cluster
Perform KM survival
analysis
Generate/validate a
multi-gene signature/s
Potential biological
function

However, only 26 out of the 130 PA genes were fit to the DE criteria as shown in figure 36 and supplemental table 3). Most of the DE PA genes (24/26) were upregulated in NSCLC while 2 genes were downregulated (*HMGB1*, *ADAM22*) (figure 36). To gain further insight into the co-expression profiles of the 26 upregulated genes, *k*-mean hierarchical clustering (up to 50 clusters) based on Euclidean distance was used to generate 8 distinct clusters (figures 37 to 44) (supplemental tables 4 to 11). Clusters 3, 4, 5, 6 and 8 contained PA genes (supplemental tables 6, 7, 8, 9 and 11). Cluster 3 contained 10/26 (38.46%) (*ANXA2*, *SERPINB6*, *PLAUR*, *S100A10*, *SERPINH1*, *CTSC*, *CTSL*, CTSZ, *PLAU* and *CTSA*), cluster 4 contained 2/26 (7.69%) (*CTSB* and *SERPINB8*), cluster 5 contained 1/26 (3.85%) (*ADAM22*) and cluster 6 contained 3/26 (11.54%) (*ADAM8*, *ADAM15* and *SERPINB5*) of the upregulated PA genes (supplemental tables 6, 7, 8, 9 and 11). Although Cluster 3 PA genes were overexpressed in NSCLC cell lines, further stratification into the three histological subtypes of NSCLC showed that these genes are uniformly expressed in adenocarcinoma, squamous cell carcinoma and large cell carcinoma. are expressed at similar levels among NSCLC subtypes (figure 45). Gene ontology analysis of cluster 3 revealed a variety of pathways (supplemental tables 12 and 13) which were then reduced and visualized using REVIGO (Reduce and Visualize Gene Ontology) [618]. These pathways included endocytosis, NF-κ_B signaling (e.g. *RELA*, *FADD*, *TRADD*, *TNFRSF1A*), protein hetero-oligo-dimerization (e.g. *STOM, CAV1, HMOX1, CLDN1, TGM2*), cell adhesion (e.g. *ITGA3*, *PDLIM5*, *ARHGAP18*, *TAGLN2, ANXA2*), GTPase signaling (*S100A10, ARHGAP18, CDC42EP1, RASA1*), and inhibition of apoptosis (e.g. *RELA, ANXA1, HMGA2, ANXA4, PLAUR, TNFRSF10D*). (figures 46 and 47)

Figure 34. Differentially-expressed genes in NSCLC vs. SCLC. The volcano plot shows the fold-change of all genes no change (1), downregulated (<1) and upregulated (>1)). A standard two-tailed t-test was performed using MeV. The raw p-value was then adjusted base on the Bonferroni test threshold to generate an adjusted p-value (see methods).

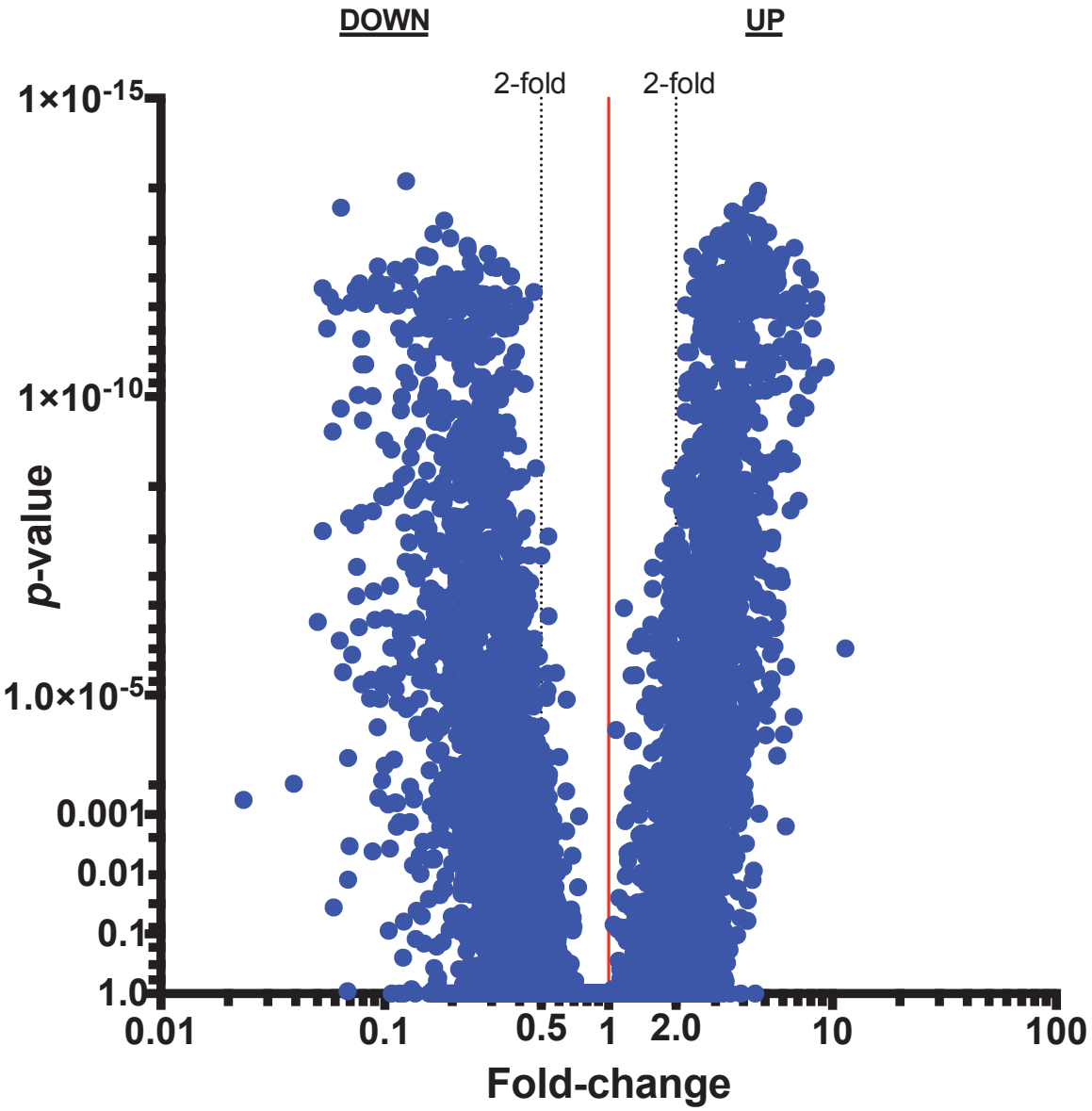

DOWN
UP
2-fold
2-fold
1×10-15
1×10-10
1.0×10-5
0.001
0.01
0.1
1.0
p-value
0.01
0.1
0.5
1
2.0
10
100
Fold-change

Figure 35. Differentially-expressed genes in NSCLC vs. SCLC with at least 2-fold difference and an adjusted *p*-value < 0.01. The volcano plot shows genes that showed at least a 2-fold change with a *p*-value less than 0.01.

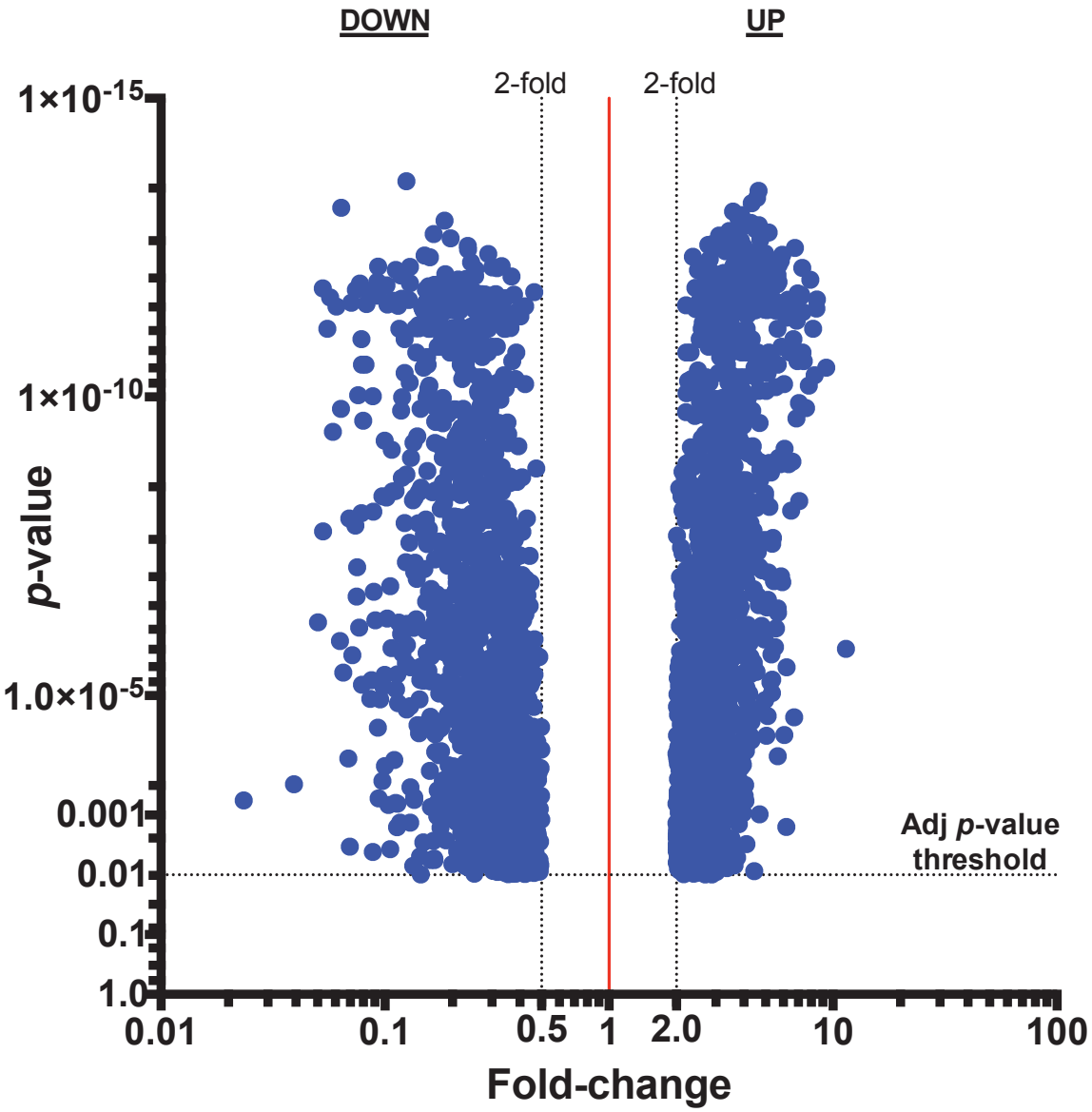
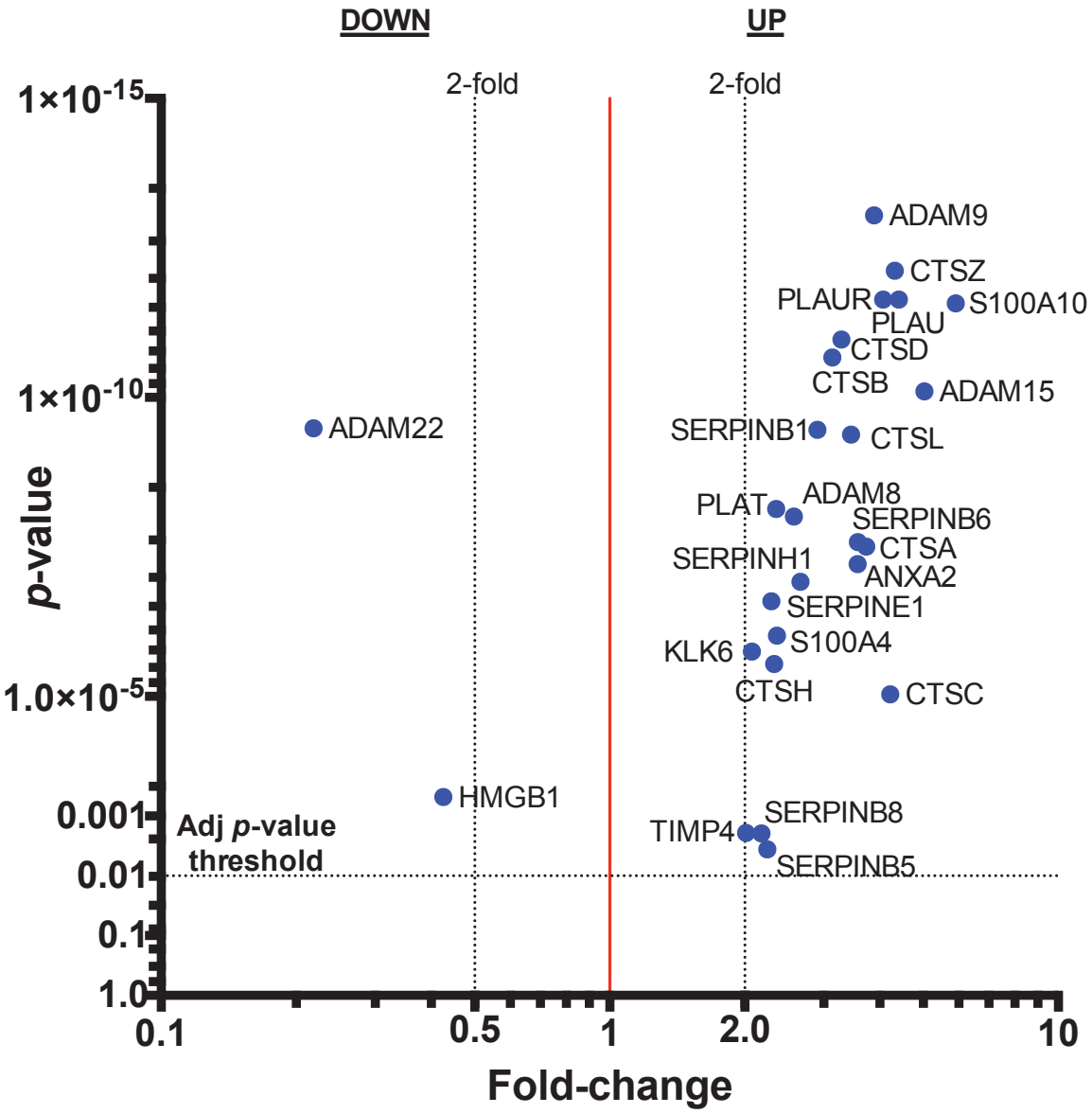

DOWN
UP
2-fold
2-fold
1×10-15
1×10-10
1.0×10-5
0.001
0.01
0.1
1.0
p-value
Adj p-value threshold
0.01
0.1
0.5
1
2.0
10
100
Fold-change

Figure 36. The 26 differentially-expressed PA genes in NSCLC vs. SCLC with at least 2-fold difference and an adjusted *p*-value < 0.01. The volcano plot shows PA (plasminogen activation) genes that showed at least a 2-fold change with a p-value less than 0.01.


DOWN
UP
2-fold
2-fold
1×10-15
1×10-10
1.0×10-5
0.001
0.01
0.1
1.0
p-value
ADAM9
CTSZ
PLAUR
S100A10
PLAU
CTSD
CTSB
ADAM15
ADAM22
SERPINB1
CTSL
PLAT
ADAM8
SERPINB6
CTSA
SERPINH1
ANXA2
SERPINE1
S100A4
KLK6
CTSH
CTSC
HMGB1
SERPINB8
TIMP4
SERPINB5
Adj p-value
threshold
0.1
0.5
1
2.0
10
Fold-change

Figures 37-44. Eight relevant gene clusters in NSCLC vs. SCLC. Up to 50 clusters were generated using MeV as heatmaps. eight heatmaps were significantly clustered between NSCLC and SCLC. Red and green color signify high and low z-scores respectively.

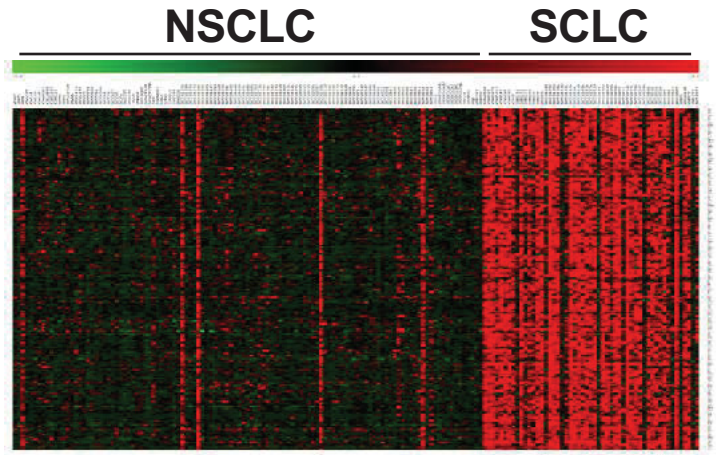


Figure 37

Cluster 1

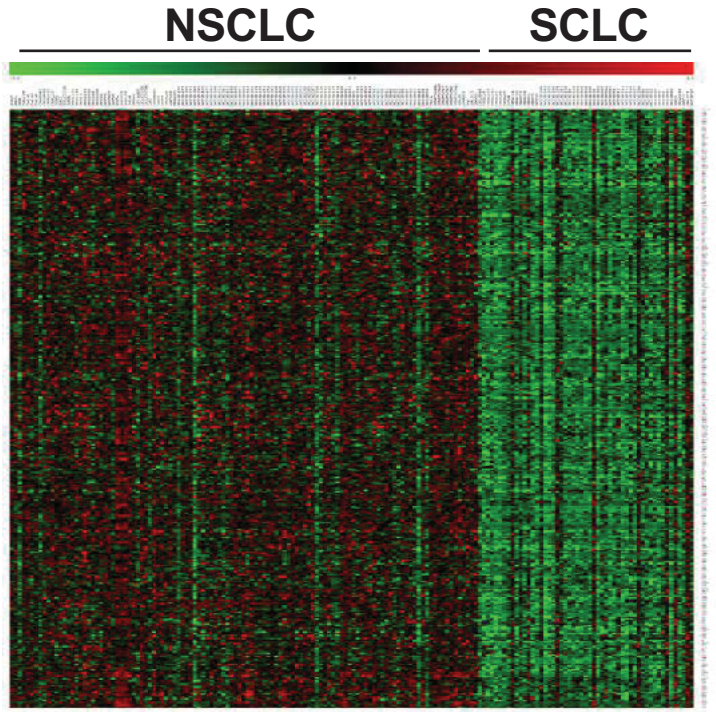
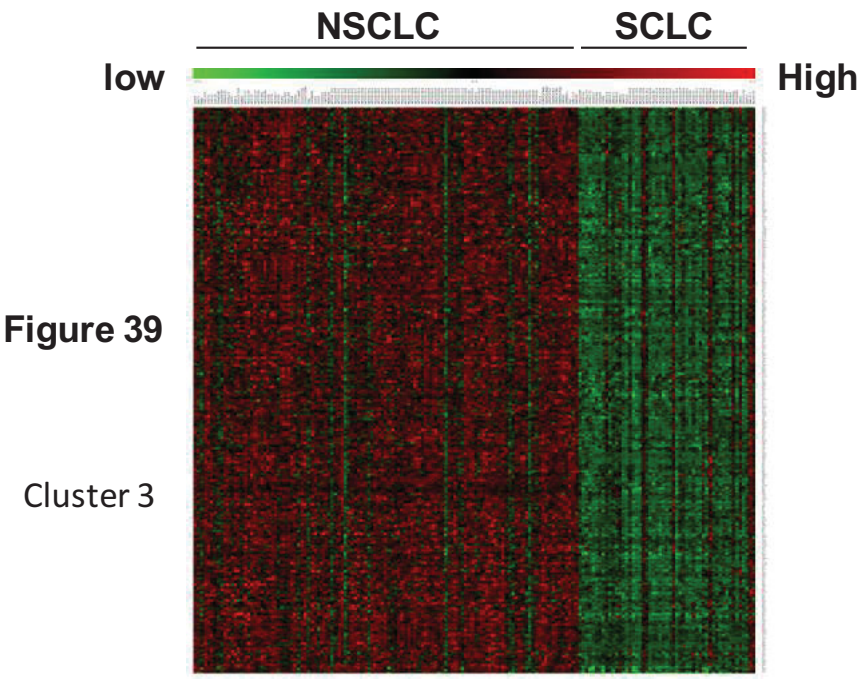


Figure 38

Cluster 2

Figure 37,Figure 38



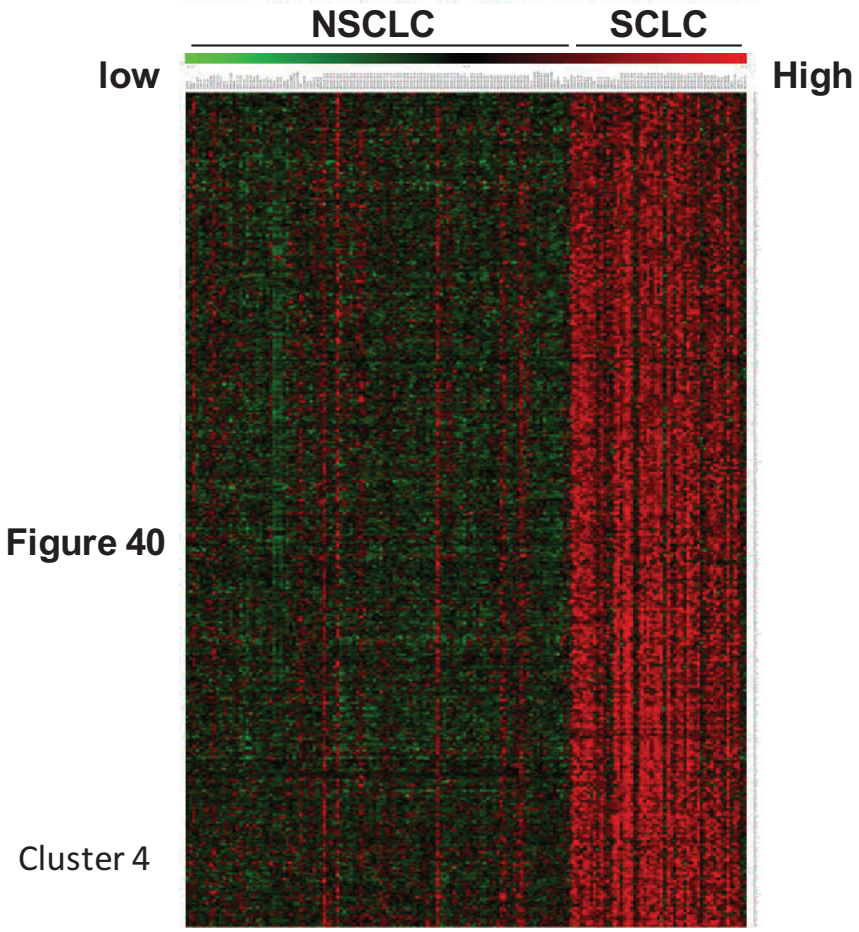


Figure 39, Figure 40

Figure 41,

Figure 42,

Figure 43

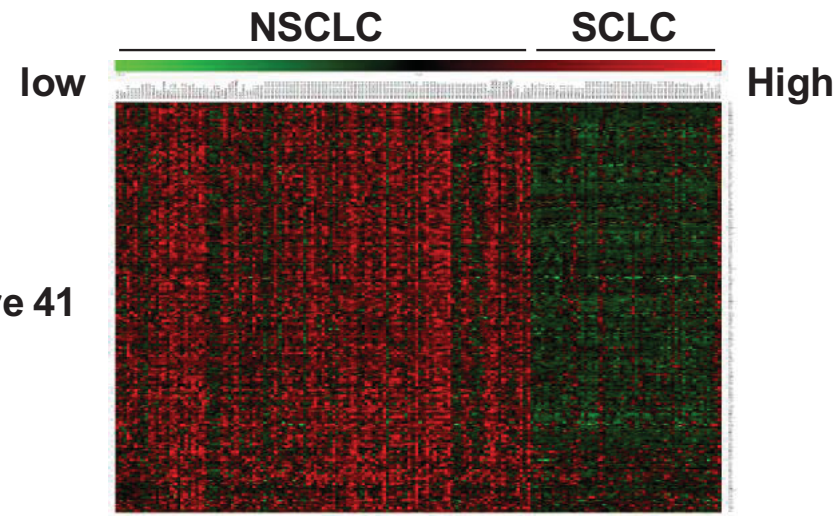


Figure 41

Cluster 5

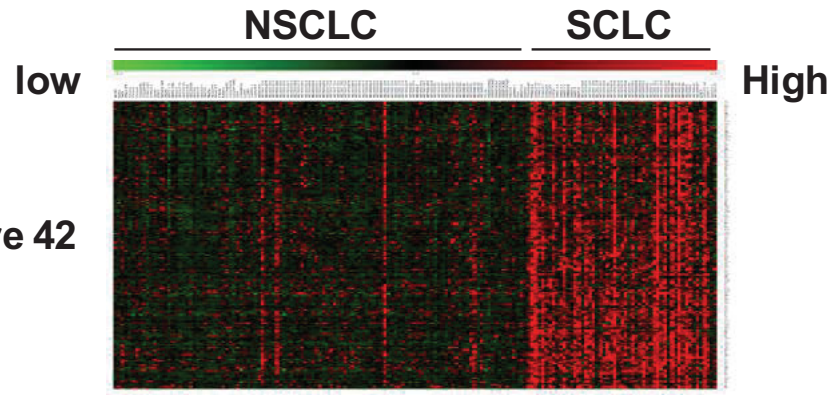


Figure 42

Cluster 6

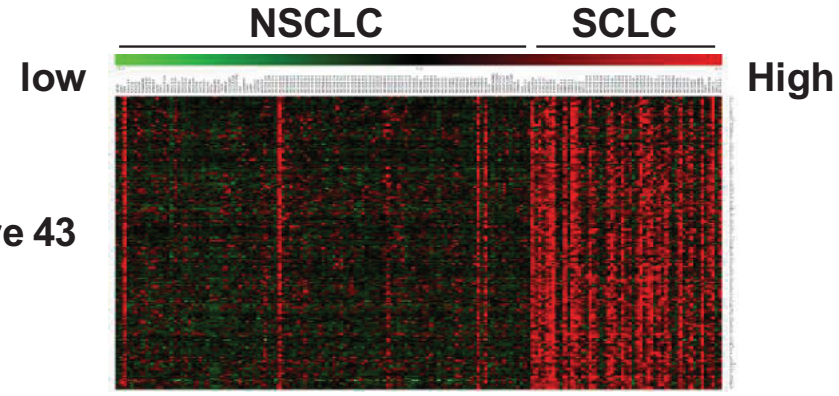


Figure 43

Cluster 7

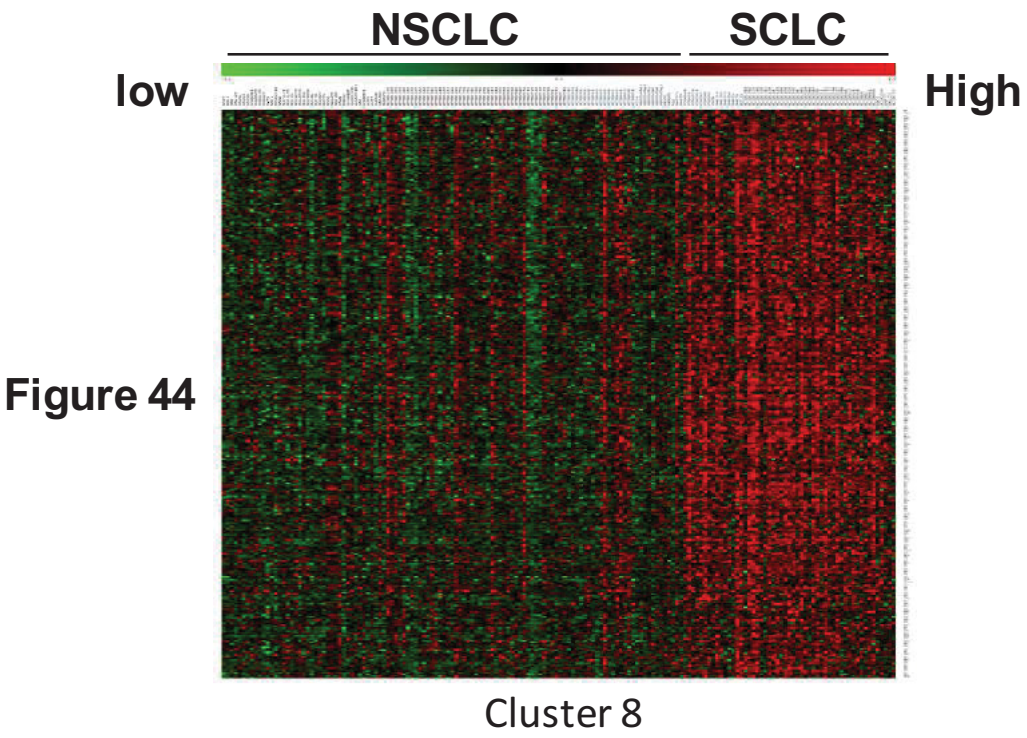
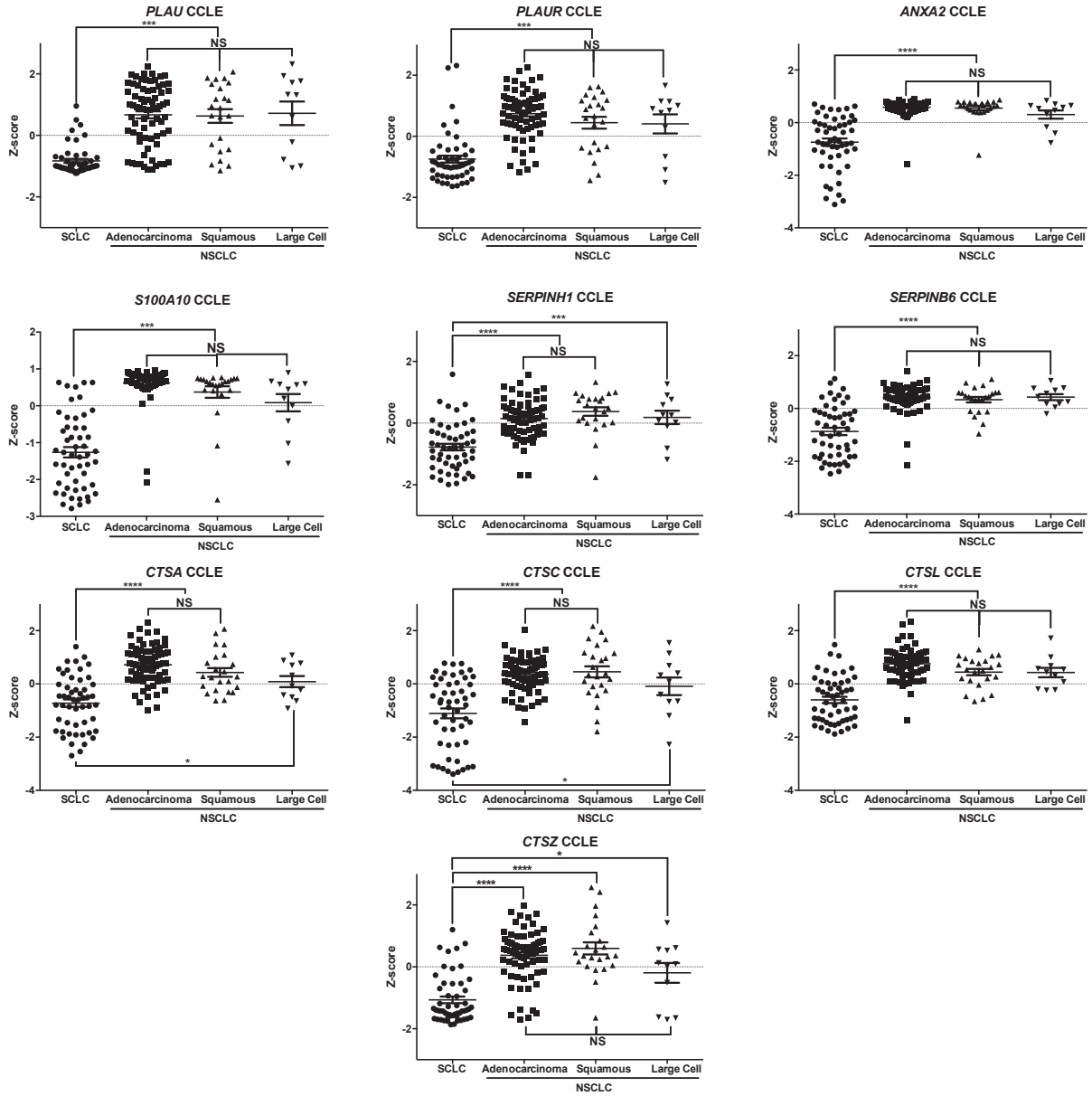


Figure 44

Figure 45. Subtype-specific expression of PA genes in cluster 50. NSCLC cell lines from the CCLE cohort were subdivided into the three histological subtypes, adenocarcinoma (n=69), squamous cell carcinoma (n=23) and large cell carcinoma (11). SCLC is predominantly small cell lung carcinoma of neuroendocrine origin (n=52). No further SCLC subtypes were included in the CCLE array.


PLAU CCLE
PLAUR CCLE
ANXA2 CCLE
S100A10 CCLE
SERPINH1 CCLE
SERPINB6 CCLE
CTSA CCLE
CTSC CCLE
CTSL CCLE
CTSZ CCLE
Z-score
SCLC
Adenocarcinoma
Squamous
Large Cell
NSCLC
NS

Figures 46 and 47. Gene ontology (GO) analysis of biological processes in cluster 3. All gene ontology annotations were obtained from the publicly available source Gene Ontology through http:// www.geneontology.org. The "biological process" of genes were considered for this experiment. Total listed genes were 386 out of 424 in cluster 3. The remaining genes are not linked to known pathways and biological processes.

Figure 46

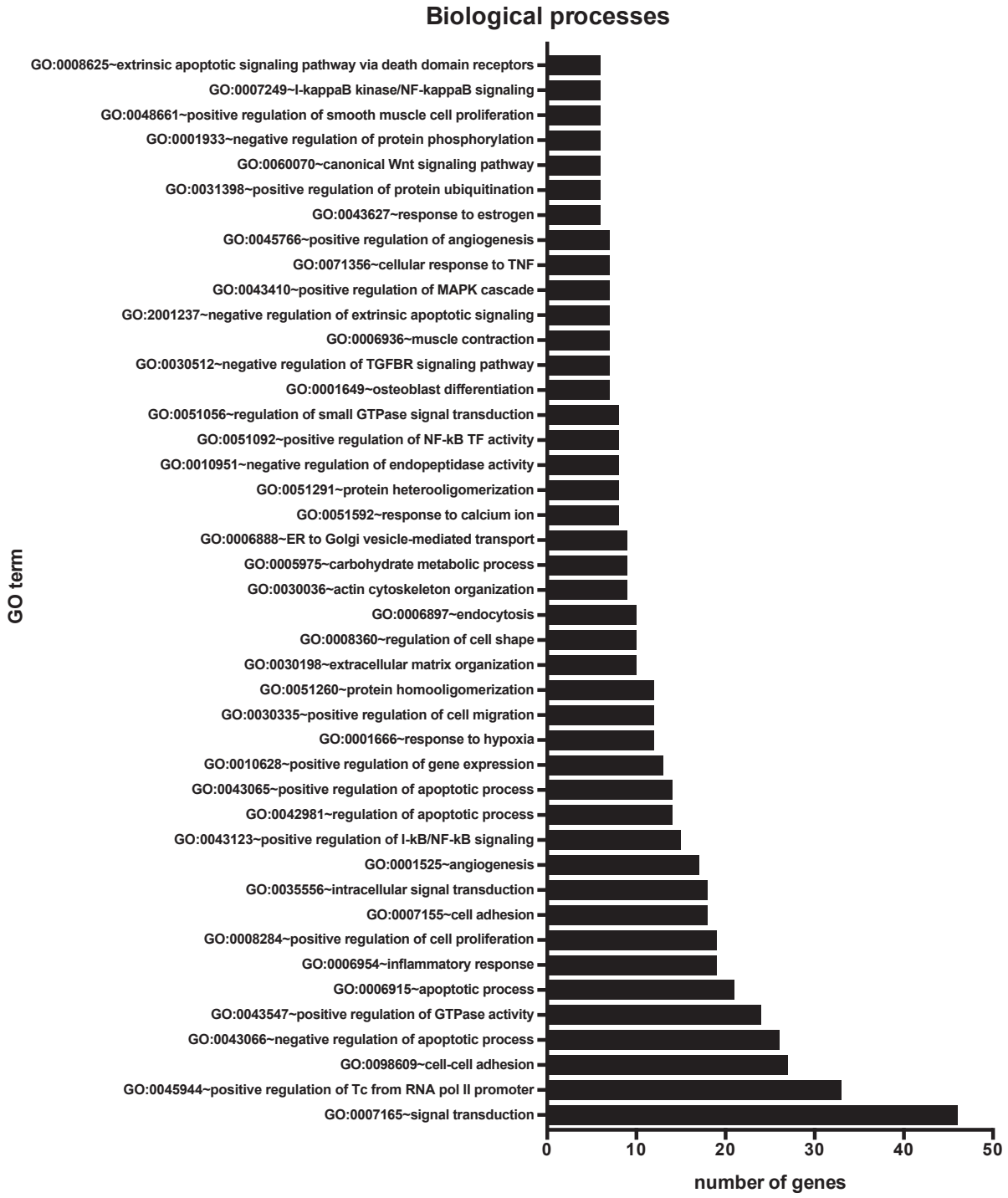


Figure 46

Figure 46

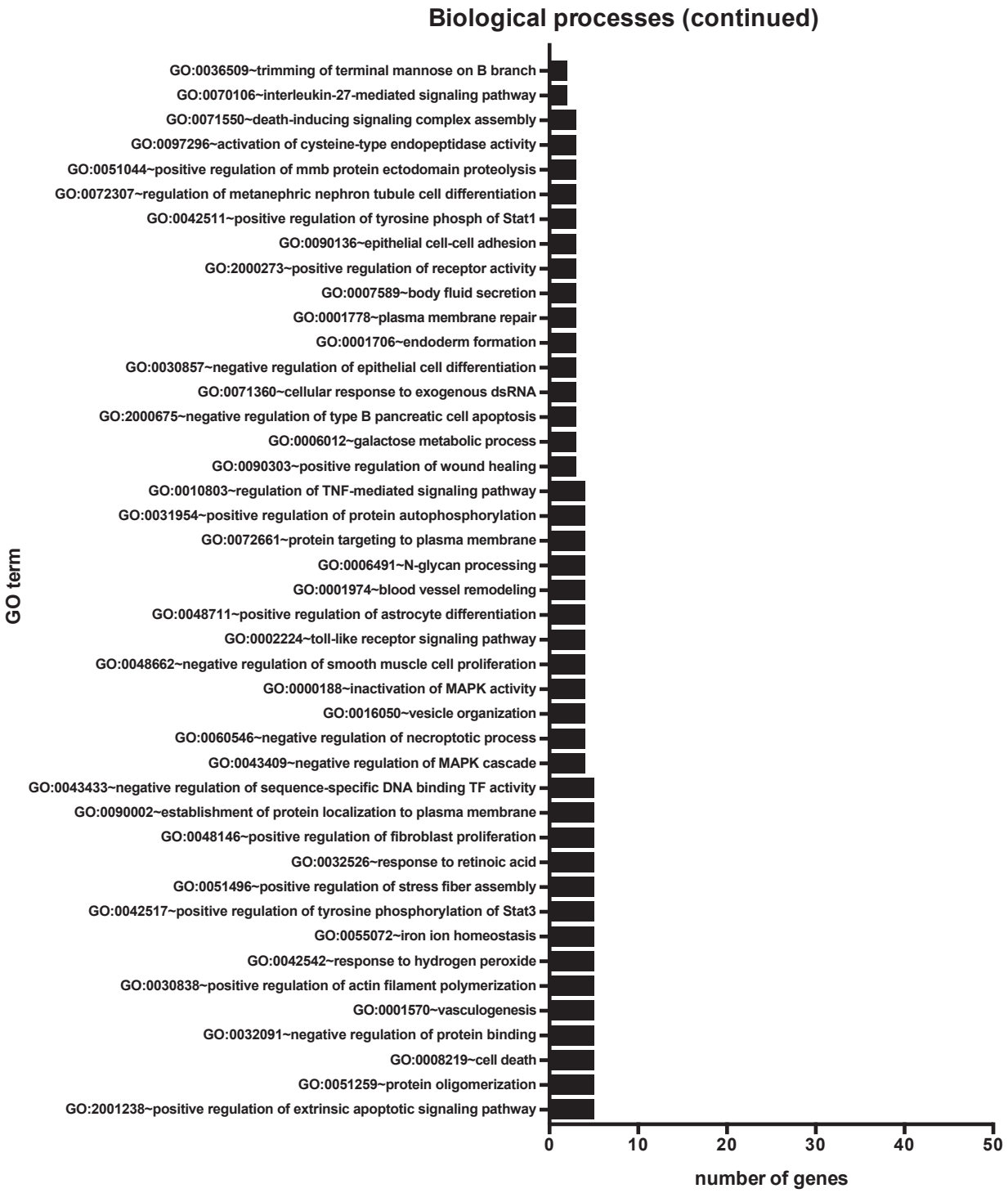


Figure 47

5.3 Select cluster 3 PA genes are predictive of overall survival in adenocarcinoma patients and not squamous cell carcinoma patients.

To assess a potential association between cluster 3 PA genes and patient outcome, a merged training cohort of 11 individual cohorts was utilized. The survival times and statuses of a total of 720 adenocarcinoma patients and 524 squamous cell carcinoma patients were examined (supplemental table 14) (see methods). A median cut-off was applied as an independent binary classifier to discern high- and low-risk patient groups. Kaplan Meier survival analysis showed that 6 genes were predictive of overall survival in adenocarcinoma patients. These gene are *PLAU* (HR:2.691, *p*-value<0.0001), *PLAUR* (HR:2.267, *p*-value<0.0001), *ANXA2* (HR:2.469, *p*-value<0.0001), *S100A10* (HR:1.914, *p*-value<0.0001), *SERPINH1* (HR:1.296, *p*-value=0.0286), *CTSA* (HR:1.612, *p*-value<0.0001) and *CTSC* (HR:0.6744, *p*-value=0.0009) (figure 48). In contrast, only ANXA2 (HR:1.371, *p*-value=0.0084) was predictive of overall survival in squamous cell carcinoma patients (figure 49). Collectively, these results demonstrated that these PA genes are potential predictive markers of overall survival in adenocarcinoma patients but not squamous cell carcinoma, even though they are expressed at similar levels.

5.4 A four-gene signature is a predictor of adenocarcinoma patient overall survival.

Using the 10 candidate prognostic genes, co-expression profiles were created based on Pearson correlations of gene expression in the CCLE NSCLC cell lines (supplemental table 15a) and the provisional TCGA adenocarcinoma patient cohort (n=517) (supplemental table 15b). Multiple comparisons (see methods) of gene associations revealed a strong correlation of expression between *S100A10*, *ANXA2*, *PLAUR* and *PLAU*

in CCLE and TCGA datasets (supplemental table 15). To validate the proposed prognostic association of the 4-gene signature, univariate analysis of survival was performed in the merged training cohort, TCGA provisional cohort and TCGA Nature 2014 cohort. By combining patients with low or high expression of these four genes, the signature achieved significance in the training cohort (HR:5.249, *p*-value<0.0001) (figure 50a) and both validation cohorts (HR:1.670, *p*-value=0.0222 and HR:2.503, *p*-value=0.0234) respectively) (figure 50b, 50c).

Figure 48. Kaplan Meier survival analysis of cluster 50 PA genes in adenocarcinoma patients. Survival statuses and times were collected from 9 out of the 11 patient cohorts. GSE4573 and TCGA cohorts were only squamous cell carcinoma cohorts. A median cut-off was applied to derive the univariate regression analysis of each gene as an independent predictor of overall survival. The survival times were directly extracted from KM plot (Kmplot.com). Biased arrays were excluded from the analysis as quality control (see methods).

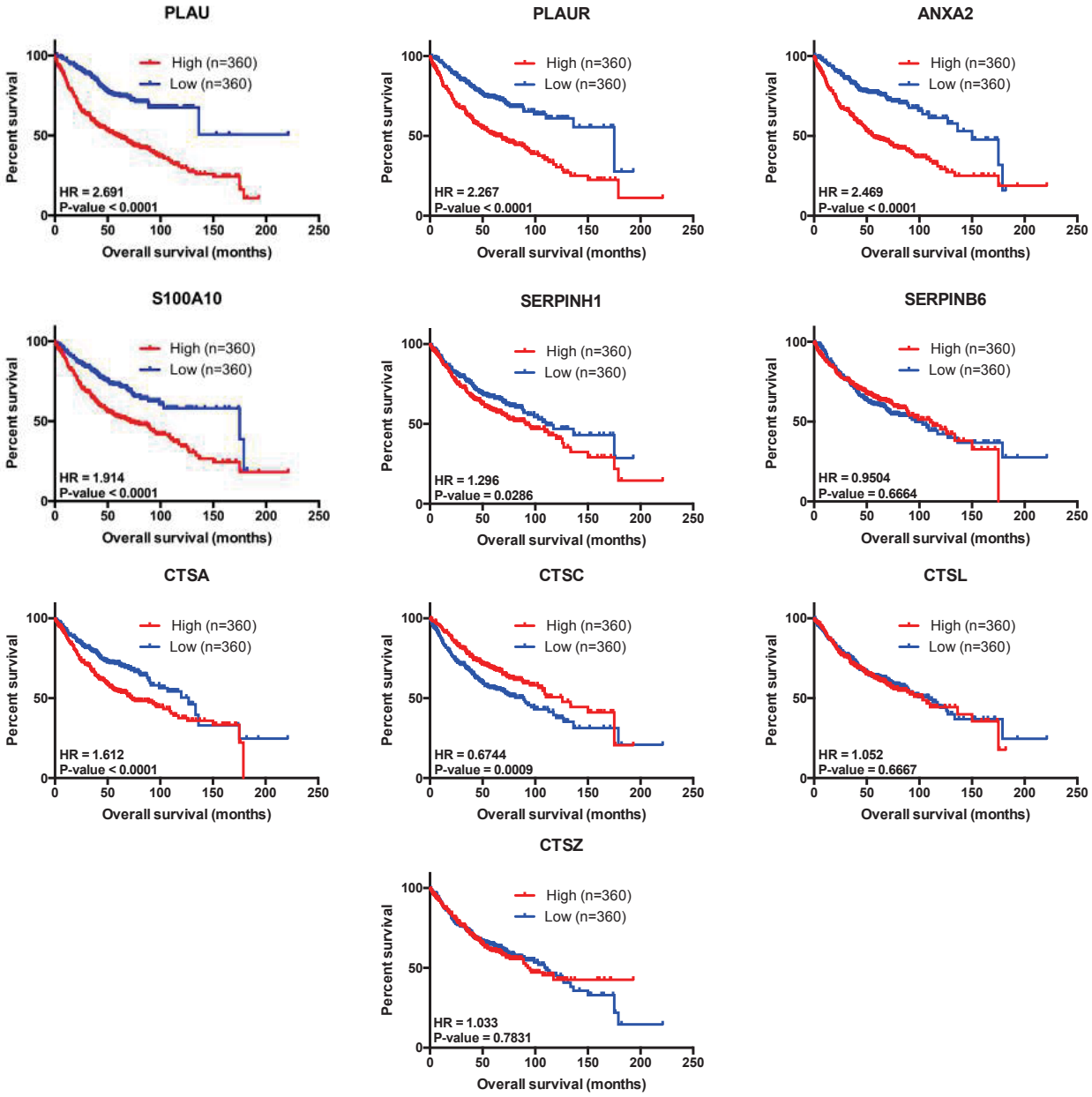

PLAU
High (n=360)
Low (n=360)
HR = 2.691
P-value < 0.0001
PLAUR
HR = 2.267
P-value < 0.0001
ANXA2
HR = 2.469
P-value < 0.0001
S100A10
HR = 1.914
P-value < 0.0001
SERPINH1
HR = 1.296
P-value = 0.0286
SERPINB6
HR = 0.9504
P-value = 0.6664
CTSA
HR = 1.612
P-value < 0.0001
CTSC
HR = 0.6744
P-value = 0.0009
CTSL
HR = 1.052
P-value = 0.6667
CTSZ
HR = 1.033
P-value = 0.7831
Percent survival
Overall survival (months)

Figure 49. Kaplan Meier survival analysis of cluster 50 PA genes in squamous cell carcinoma patients. Survival statuses and times were collected from 9 out of the 11 patient cohorts. GSE31908 and GSE31210 cohorts were only adenocarcinoma cohorts. A median cut-off was applied to derive the univariate regression analysis of each gene as an independent predictor of overall survival. The survival times were directly extracted from KM plot (Kmplot.com). Biased arrays were excluded from the analysis as quality control (see methods).

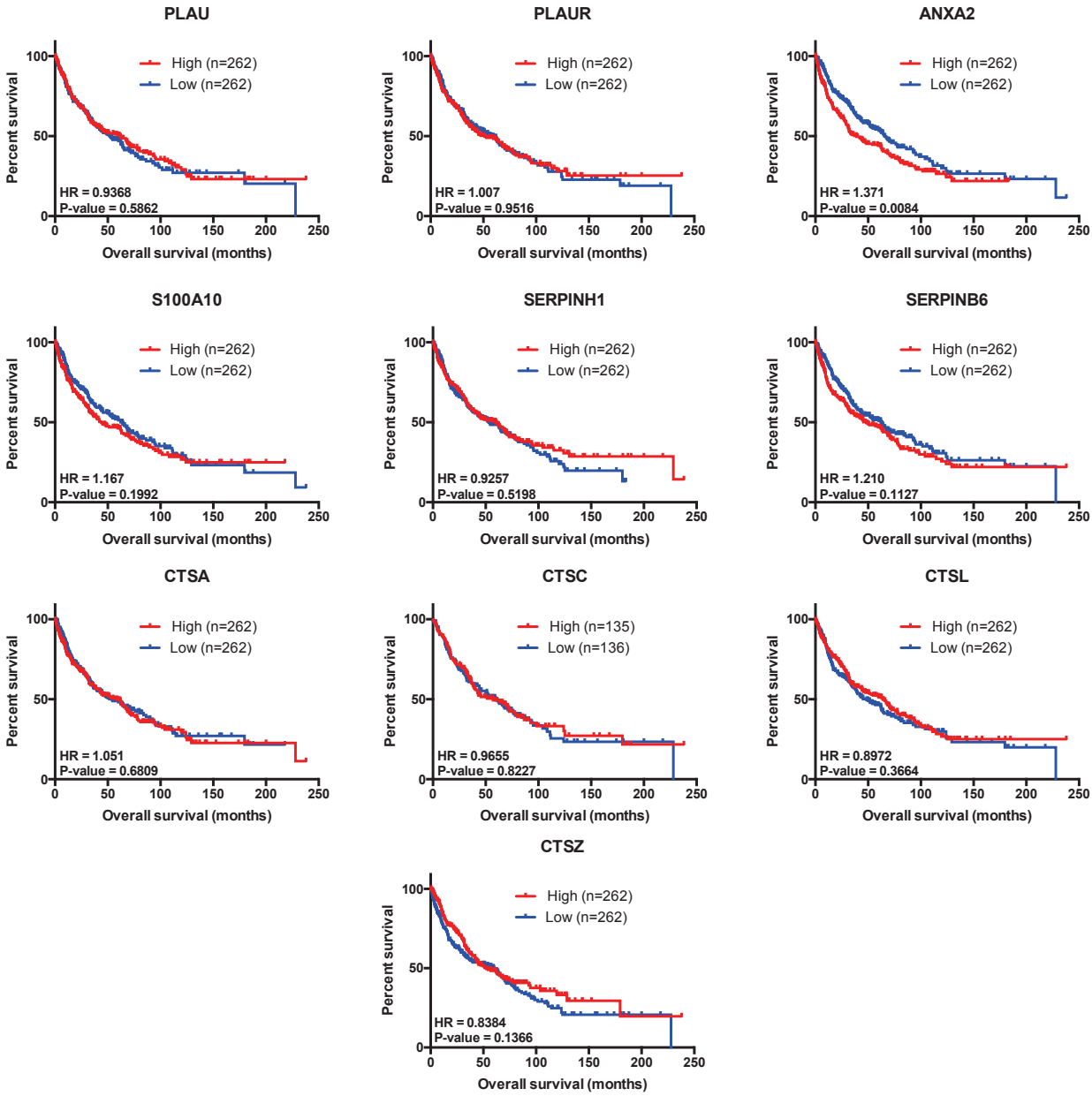
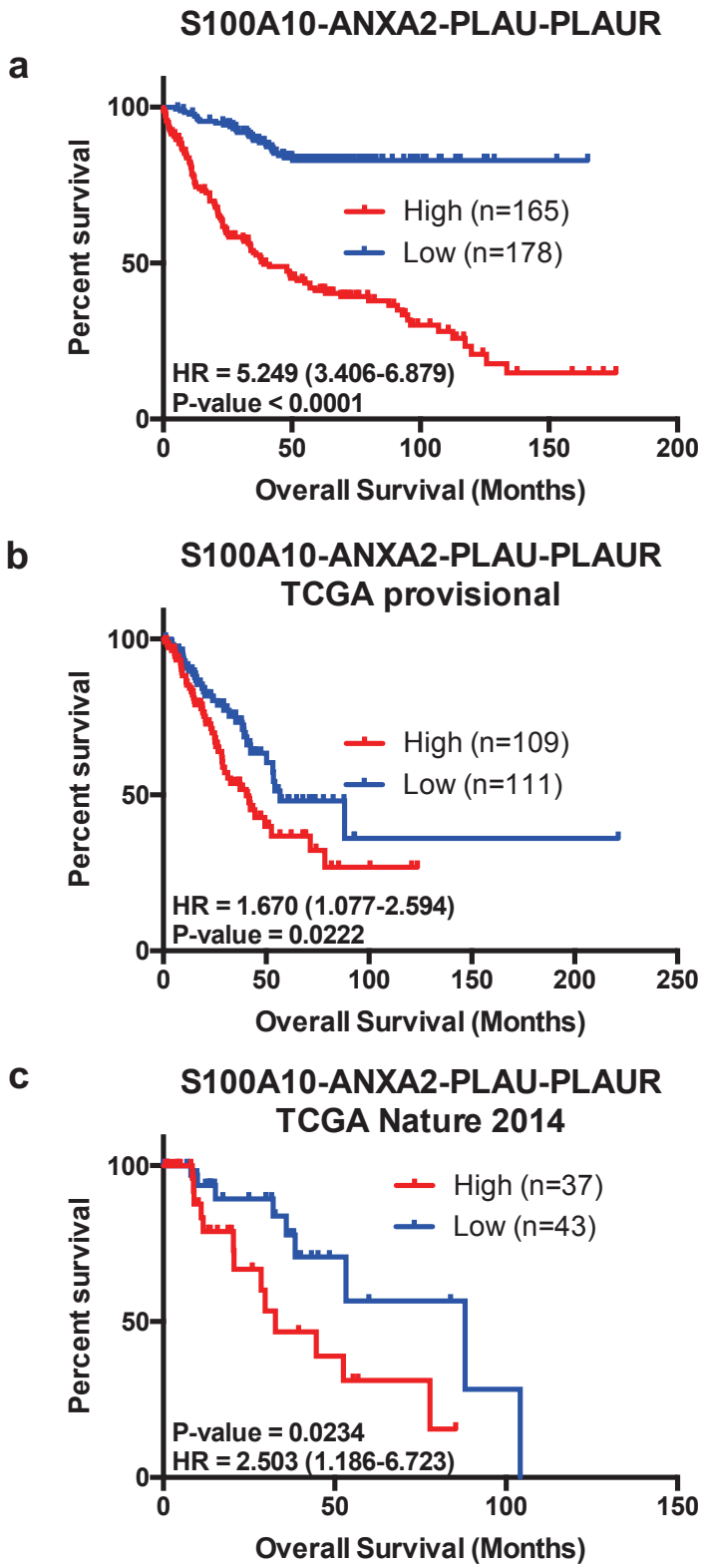

PLAU
High (n=262)
Low (n=262)
HR = 0.9368
P-value = 0.5862
PLAUR
High (n=262)
Low (n=262)
HR = 1.007
P-value = 0.9516
ANXA2
High (n=262)
Low (n=262)
HR = 1.371
P-value = 0.0084
S100A10
High (n=262)
Low (n=262)
HR = 1.167
P-value = 0.1992
SERPINH1
High (n=262)
Low (n=262)
HR = 0.9257
P-value = 0.5198
SERPINB6
High (n=262)
Low (n=262)
HR = 1.210
P-value = 0.1127
CTSA
High (n=262)
Low (n=262)
HR = 1.051
P-value = 0.6809
CTSC
High (n=135)
Low (n=136)
HR = 0.9655
P-value = 0.8227
CTSL
High (n=262)
Low (n=262)
HR = 0.8972
P-value = 0.3664
CTSZ
High (n=262)
Low (n=262)
HR = 0.8384
P-value = 0.1366
Percent survival
Overall survival (months)
0
50
100
150
200
250

Figure 50. Kaplan Meier survival analysis of the S100A10-ANXA2-PLAU-PLAUR gene signature in the multi-cohort discovery studies and two validation studies. Kaplan Meier survival analysis of the 9 cohorts in the multi-cohort discovery set (a), TCGA lung adenocarcinoma (provisional) (b) and TCGA lung adenocarcinoma (Nature, 2014) (c). Low and high expression were determined as patients with below and above (respectively) median expression for each individual gene.


a
S100A10-ANXA2-PLAU-PLAUR
Percent survival
100
50
0
High (n=165)
Low (n=178)
HR = 5.249 (3.406-6.879)
P-value < 0.0001
0
50
100
150
200
Overall Survival (Months)
b
S100A10-ANXA2-PLAU-PLAUR
TCGA provisional
Percent survival
100
50
0
High (n=109)
Low (n=111)
HR = 1.670 (1.077-2.594)
P-value = 0.0222
0
50
100
150
200
250
Overall Survival (Months)
c
S100A10-ANXA2-PLAU-PLAUR
TCGA Nature 2014
Percent survival
100
50
0
High (n=37)
Low (n=43)
P-value = 0.0234
HR = 2.503 (1.186-6.723)
0
50
100
150
Overall Survival (Months)

5.5 *S100A10*, *ANXA2* and *PLAUR* are predictive of chemotherapeutic response in adenocarcinoma patients

The ability to predict patient outcome is closely linked to the patient response to treatment regimen. To examine whether the four PA genes were also predictive of adenocarcinoma patient outcome in the context of chemotherapy, patients that received adjuvant chemotherapy were included in the analysis (GSE29013; n=19 (treatment unspecified) and GSE14814; n=17 treated with (ACT: adjuvant cisplatin/vinorelbine)). Only PLAUR (HR:4.585, *p*-value=0.0111) (figure 51a), ANXA2 (HR:7.331, *p*-value<0.0001) (figure 51c) and S100A10 (HR:7.331, *p*-value<0.0001) (figure 51d) showed a strong correlation with chemotherapeutic response. Patients who received adjuvant chemotherapy and who had high expression of these three genes are at a higher risk of death (i.e. no response to therapy) compared to lower expression (low-risk group) (figure X51). In addition, a 100% concordance was present between the high and low-risk patients based on *ANXA2* and *S100A10* expression. This is further supported by the high Pearson correlation coefficient of these two genes in the CCLE NSCLC cell lines and the TCGA provisional adenocarcinoma patient cohort (supplemental table 15a, 15b).

5.6 *S100A10* is upregulated by various chemotherapeutic agents and may contribute to cisplatin resistance.

In attempt to understand the contribution of PA genes (specifically S100A10) to respond to chemotherapies, A549 cells were treated with various chemotherapeutics. S100A10 was upregulated by cisplatin in a dose-dependent manner (figure 52a). To discern if this response is specific to cisplatin, A549 cells were treated with three other

chemotherapeutic agents, doxorubicin (100nM), methotrexate (1μM) and paclitaxel (10nM). S100A10 was upregulated by all three agents suggesting that S100A10 is responsive to the common pathways activated by these agents (figure 52b). To understand whether S100A10 can promote drug resistance, scramble control and S100A10 shRNA A549 cells were treated with 5μM cisplatin and stained with apoptosis markers. Interestingly, cells depleted of S100A10 were more likely to be in early apoptosis than scramble control cells (figure 52c). This suggested that S100A10 is a chemotherapy-responsive gene that could potentially contribute to drug resistance.

Figure 51. Kaplan Meier survival analysis of the individual four genes in patients who received chemotherapy. Chemotherapy clinical data was only available for 36 patients, 19 of which are from the GSE29013 cohort and 17 from the GSE14814 cohort. A median cut-off was applied to identify high (n=18) and low risk (n=18) individuals with high and low expression of S100A10.

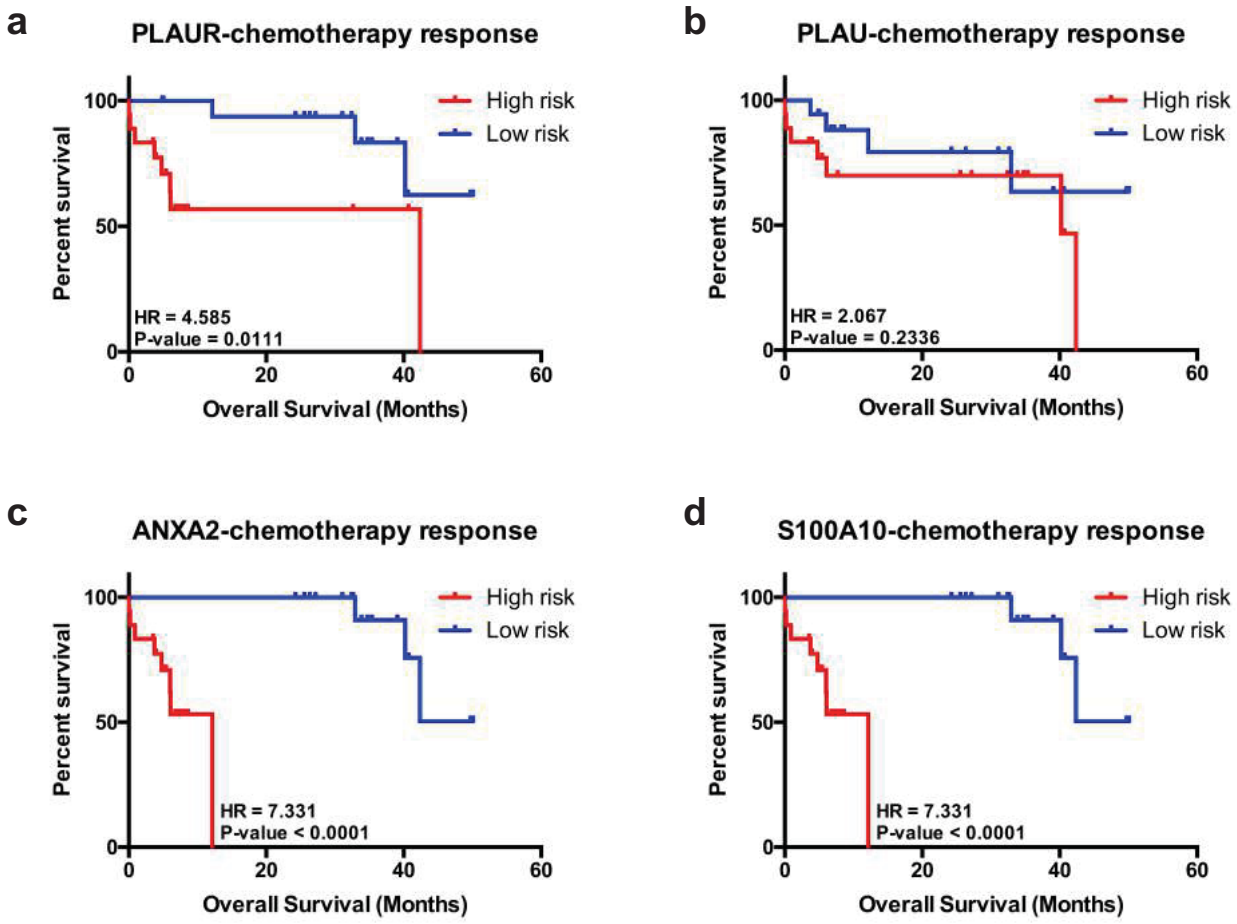
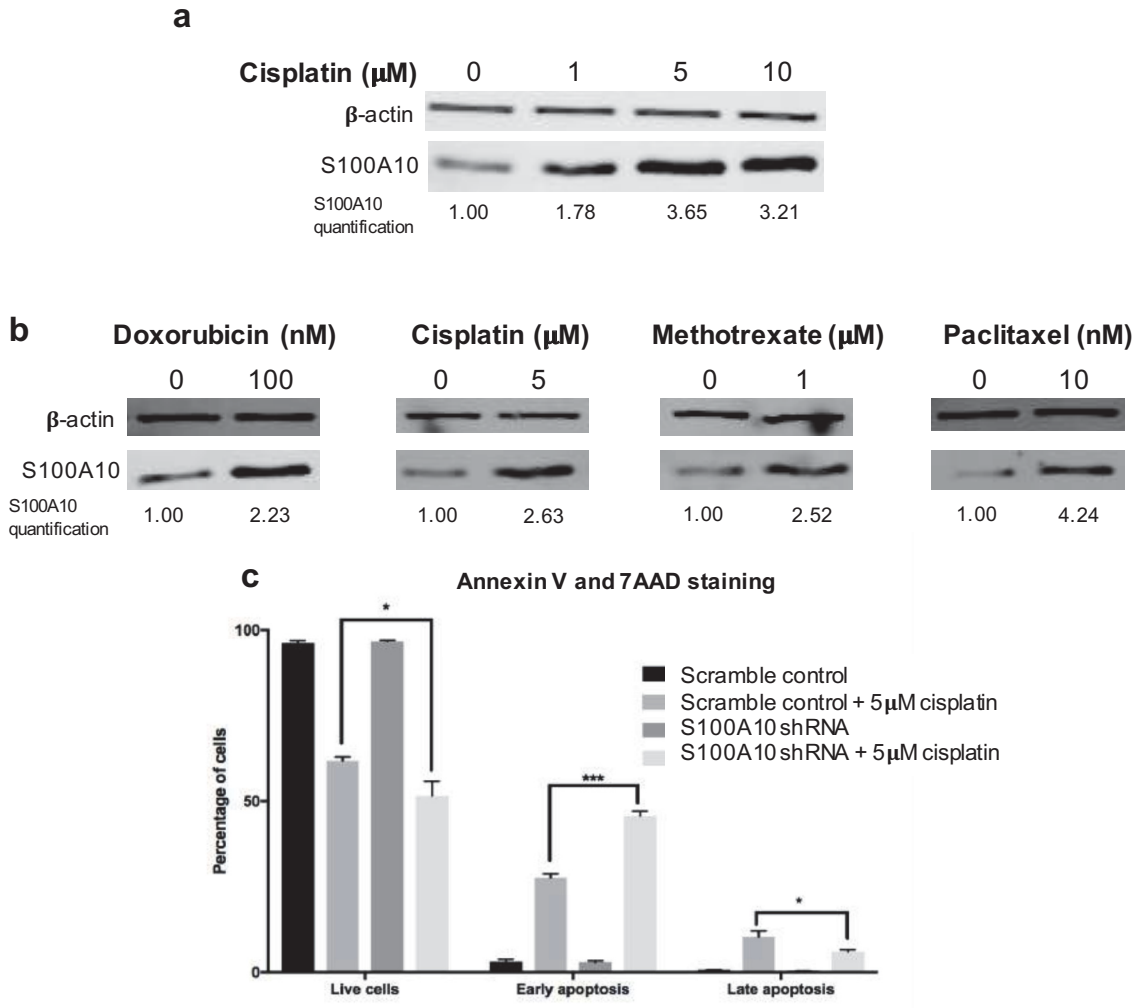

a
PLAUR-chemotherapy response
High risk
Low risk
Percent survival
100
50
0
HR = 4.585
P-value = 0.0111
0
20
40
60
Overall Survival (Months)
b
PLAU-chemotherapy response
High risk
Low risk
Percent survival
100
50
0
HR = 2.067
P-value = 0.2336
0
20
40
60
Overall Survival (Months)
c
ANXA2-chemotherapy response
High risk
Low risk
Percent survival
100
50
0
HR = 7.331
P-value < 0.0001
0
20
40
60
Overall Survival (Months)
d
S100A10-chemotherapy response
High risk
Low risk
Percent survival
100
50
0
HR = 7.331
P-value < 0.0001
0
20
40
60
Overall Survival (Months)

Figure 52. S100A10 is responsive to various chemotherapeutic agents. (a) A549 cells were treated with ascending concentrations (0 to 10 μM cisplatin) for 72 hours. (b) A549 cells were treated with sub-cytotoxic doses of four chemotherapeutic agents: 100nM doxorubicin, 5 μM cisplatin, 1 μM methotrexate and 10 nM paclitaxel. (c) A549 scramble control and S100A10 shRNA were treated with 5 μM cisplatin for 72 hours after which cells were stained with Annexin V and 7AAD.


a
Cisplatin (μM) 0 1 5 10
β-actin
S100A10
S100A10 quantification 1.00 1.78 3.65 3.21
b
Doxorubicin (nM)
0 100
β-actin
S100A10
S100A10 quantification 1.00 2.23
Cisplatin (μM)
0 5
1.00 2.63
Methotrexate (μM)
0 1
1.00 2.52
Paclitaxel (nM)
0 10
1.00 4.24
c
Annexin V and 7AAD staining
Percentage of cells
100
50
0
Scramble control
Scramble control + 5μM cisplatin
S100A10 shRNA
S100A10 shRNA + 5μM cisplatin
*
***
*
Live cells
Early apoptosis
Late apoptosis

CHAPTER 6: DISCUSSION of chapter 5

6.1 Discussion

SCLC and NSCLC are two histologically different cancers in which SCLC arises from neuroendocrine cells while NSCLC arises from epithelial cells [619]. However, the transcriptomic landscapes of both cancers have not been sufficiently addressed. Although such comparisons may not necessarily alter the histological distinction at the diagnosis stage, it will however identify novel and unique markers of both NSCLC or SCLC. In the current work, we have identified DE genes (e.g. *TCF4, LIPH, ARHGAP27, ELAVL1, EPHA2*) that are of potential interest for further biomarker analyses but were not explored in this dissertation due to hypothesis-driven bias (addressed in 6.2.1) (figure 35). Twenty-six PA genes were however identified as DE in NSCLC compared to SCLC most of which were upregulated suggesting a global upregulation of PA genes. Only two genes (*ADAM22* and *HMGB1*) were downregulated in NSCLC (or upregulated in SCLC) (figure 36). A literature search revealed that no associations of these two genes with SCLC have been previously made rendering this observation novel. A recent meta-analysis of HMGB1 mRNA expression studies showed that HMGB1 was upregulated in NSCLC tumors compared to normal tissues [620]. However, the question of whether HMGB1 expression in SCLC tumors is markedly different than that in NSCLC tumors is yet to be addressed.

Unsupervised hierarchical clustering revealed that a short list of 10 PA genes that were clustered (cluster 3) (figure 39) (supplemental table 6). This multi-step top-down approach allowed the identification of a select list of candidate survival predictors (*ANXA2, S100A10, PLAUR, PLAU, SERPINH1, CTSA* and *CTSC*). The implications of *PLAU*,

PLAUR, *S100A10* and *ANXA2* have been previously addressed in the literature particularly in the context of correlation with clinical features of NSCLC patients. For instance, higher stromal tissue levels of uPA have been linked to poor outcome, increased tumor size, lymph node involvement and advanced staging in NSCLC [621], consistent with our survival analysis of uPA (figure 48). Elevated levels of cleaved and intact uPA have also been linked to poor prognosis [617][622][623]. Interestingly, when measured using ELISA, uPA levels did not correlate with outcome [624] suggesting potential inconsistencies and variations in methods used for measurement. Serum and tumor levels of uPAR levels also correlated with poorer outcome and likelihood of metastasis in NSCLC patients [625][626][179] which is consistent with our survival analysis of uPAR(figure 48). The cleaved form of uPAR in serum was also indicative of increased tumor-associated uPA and, together (i.e. uPA and uPAR) offered a higher predictive power than either alone [179]. S100A10 expression (IHC) correlated with poor prognosis, poor differentiation, higher TNM stage and severity of intra-tumoral vascular invasion [617]. In addition, higher expression of *ANXA2* has been linked to poor prognosis in all NSCLC patients [627] as well as adenocarcinoma and squamous cell carcinoma patients [628]. The prognostic roles of the remaining genes (*CTSA, CTSC* and *SERPINH1*) in NSCLC are novel observations that will require future studies for validation.

PAI-1 (SERPINE1) levels in tumors (IHC) correlated with survival, lymph node positivity and stage in squamous cell carcinoma with significant correlations in adenocarcinoma. Increased levels and secretion of PAI-1 has also been recently linked to enhanced radio-resistance of lung NSCLC cell lines [629]. . Interestingly, PAI-1 and uPA (*PLAU*) serum levels were found to be predictive of disease in lymph-node negative triple

negative breast cancer patients [630]. The use of uPA/PAI-1 levels as a biomarker in breast cancer has been approved in level-of-evidence 1 studies [631]. Such association between uPA and PAI-1 was not seen in our NSCLC analysis (supplemental table 6) which could be attributed to either differences in cancer models or poor association between mRNA levels and protein (in serum) levels of uPA and PAI-1. In addition, although these studies addressed the combinatorial benefit of using two genes/proteins as predictors of outcome, these approaches were based on predisposed notion of the function of the two PA genes.

Ultimately, four genes (*PLAU, PLAUR, ANXA2* and *S100A10*) showed high Pearson correlation coefficients (supplemental table 15) which prompted further examination of a potential gene signature. All four genes were individually and collectively predictive in two independent adenocarcinoma patient cohorts (figure 48 and figure 50a, 50b, 50c). Interestingly, these genes were not predictive (except *ANXA2*) in squamous cell carcinoma patients (figure 49) even though they were expressed at similar levels in the CCLE NSCLC cell lines (figure 45). This is particularly important for two reasons: first, a subtype-specific gene signature can be developed regardless of levels of expression across various subtypes, and second, the absence of a correlation with the squamous cell carcinoma patients serves as an internal negative control for the univariate analysis. The prognostic values of *PLAUR, PLAU* and *S100A10* in SCLC have not been addressed in the literature.

The expression of three (*PLAUR, ANXA2* and *S100A10*) of the four genes also correlated with response to chemotherapy (figure 51a, 51c, 51d). *PLAUR* expression was shown to reduce cisplatin sensitivity in mesothelioma cells [632] and SCLC [633] but not

in NSCLC or adenocarcinoma. Hence, the role of *PLAUR* in drug resistance in NSCLC cell lines and tumors is yet to further addressed. *ANXA2* expression was recently linked to cisplatin resistance in NSCLC cell lines (A549, H460 and H1650) [634]. Similarly, forced expression of S100A10 in COLO-320 colorectal cancer cells increased their resistance to oxaliplatin, a platinum-based therapy [635][636]. Nymoen *et al*. also showed that higher *S100A10* mRNA expression correlated with increased chemo-resistance in ovarian serous carcinoma patients [637]. We showed that treatment of A549 cells with multiple chemotherapies increased S100A10 protein expression (figure 52b). Knockdown of S100A10 in A549 cells increased the number of cells in early apoptosis suggesting a role of S100A10 in drug sensitivity (figure 52c). In that context, various reports demonstrated that TGFβ1 promotes drug resistance across multiple cell lines and tumor types [638] [639]. Hence, the responsiveness of S100A10 to TGFβ1, serum withdrawal, and/or PI3K inhibition further indicate its involvement in drug resistance, through a mechanism that is yet to be addressed.

The Kaplan Meier survival analysis showed that patients with elevated expression of these genes (*PLAUR, ANXA2* and *S100A10*) predicted a shorter survival in NSCLC patients who received chemotherapy in an adjuvant setting (figure 51). Since mRNA measurements were made right after surgical resection (i.e. prior to chemotherapy administration), it suggested that intrinsic higher levels of *PLAUR*, *ANXA2* and *S100A10* predicted overall survival although their levels may then be affected the chemotherapeutic agent itself. This will potentially lead to increased resistance or positive selection of cell populations that express higher levels of these genes.

6.2 Study limitations and future directions

6.2.1 Biased assessment of PA genes

The above study performed a comprehensive analysis of DE genes in NSCLC vs. SCLC cell lines. Although the initial analysis was an unbiased comparison of DE genes (figure 35), further assessment of the 130 PA genes (figure 36) added a biased layer driven by the proposed hypothesis. This could potentially hinder the identification of the most robust DE prognostic markers in preference of examining DE PA prognostic markers. Many of the derived prognostic PA markers are novel and will require further validation.

6.2.2 The impracticality of multivariate regression modeling

The above study examined 11 merged cohorts of lung cancer patients with various degrees of clinical data availability. Although this permitted univariate analysis of overall survival based on each predictor (gene expression), multivariate regression analysis was not applicable due to the absence of complete annotated clinical covariates of each individual cohort (e.g. stage, grade, lymph node involvement, etc.).

6.2.3 *In vivo* drug resistance

Examination of the predictive potential of *S100A10*, *PLAUR* and *ANXA2* showed promising involvement in a drug resistance mechanism (figure 51). Exploration of *S100A10* only demonstrated that it is involved in protecting cells against apoptosis (figure 52). Similar examination of *PLAUR* and *ANXA2* is required to discern their potential involvement. Ultimately, the knockdown of these genes in lung tumors *in vivo* will

recapitulate their role in protection against apoptosis when mice are challenged with a chemotherapeutic agent.

CHAPTER 7: The Plasminogen Receptor S100A10 is Predictive of Patient Survival and a Driver of Tumorigenesis in Pancreatic Ductal Adenocarcinoma

7.1 Study rationale

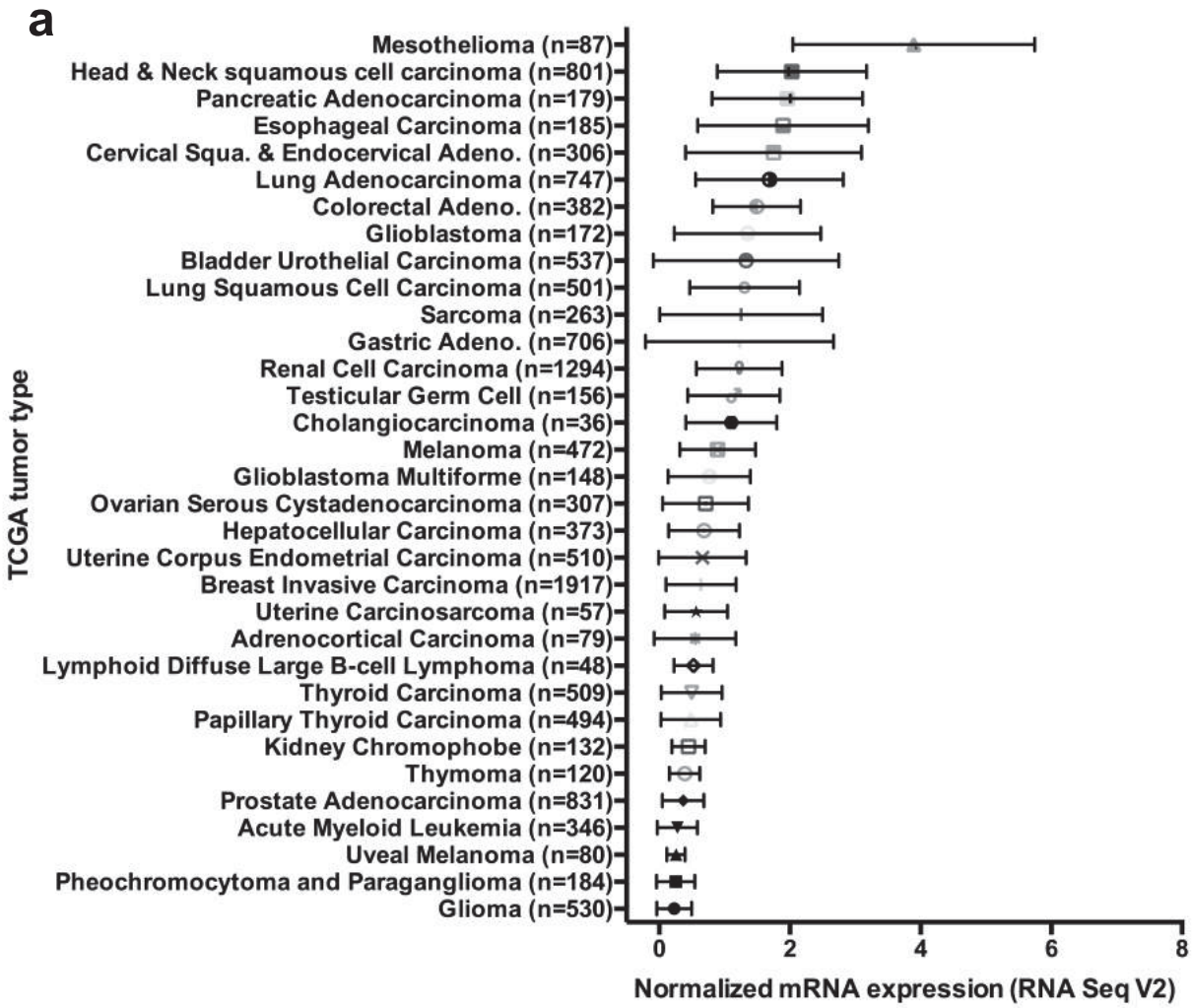
The gold standard for predicting PDAC patient outcome is TNM (tumor, node, metastasis) staging which performs adequately in late stage (stage III and IV) patients, in which their tumors are usually not resectable. However, the prognostic performance of TNM staging is below par in early stage (stage I and II) resectable patients [468]. The consequence of this poor performance is a tendency to undertreat patients who have a high risk of recurrent disease and over-treating patients who are at low risk of recurrence. carbohydrate antigen 19-9 (CA 19-9), a long-established marker of pancreatic cancer has shown performance inconsistencies. For instance, 10% of clinically-diagnosed patients do not express the (CA 19-9). Furthermore, its levels are heavily influenced by confounding medical conditions such as cystic fibrosis, liver cirrhosis, inflammatory bowel disease and others [640]. Hence, there is a lack of reliable clinical markers that can identify patients with a high risk of recurrence or metastatic disease. Novel biomarkers are therefore needed to help identify high and low risk patient subgroups and help shape their treatment modalities accordingly. To address these issues, we herein use systematic clinical and functional validation methods to describe a novel biomarker, *S100A10*, and demonstrate its efficacy in predicting PDAC patient outcome. The upcoming series of experiments have two objectives: first, to establish if S100A10 is involved in the progression of PDAC in

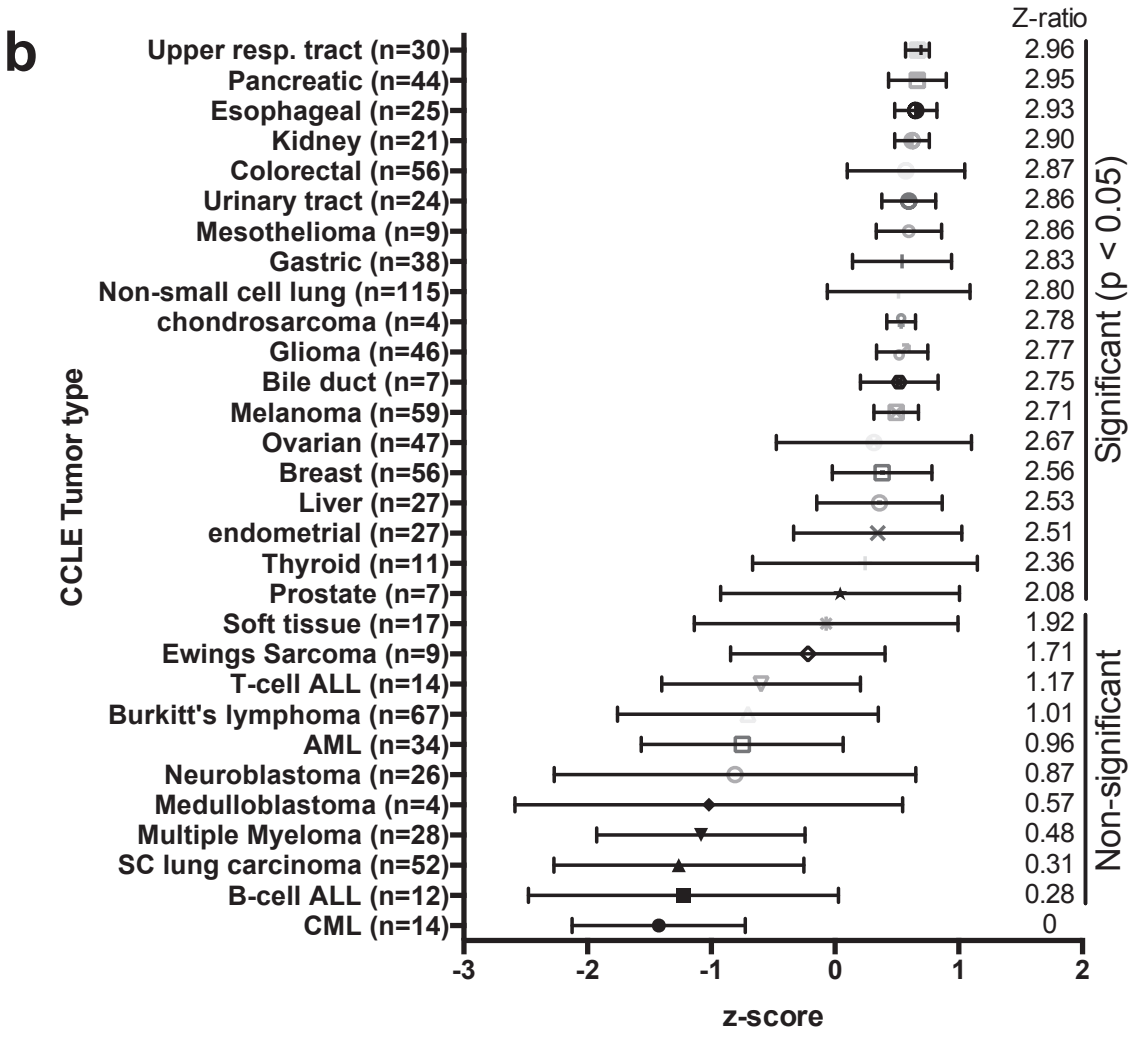
cell and mouse models and therefore might represent a targetable protein for treatment of PDAC patients, and second, to investigate the potential use of *S100A10* as a predictive biomarker. Here, we demonstrate that the protease-activating function of S100A10 regulates PDAC cell invasion *in vitro* and that it also mediates tumor growth in *in vivo* mouse models. We also demonstrate, for the first time, that *S100A10* mRNA and protein are overexpressed in pancreatic tumors and that *S100A10* mRNA and methylation status are prognostic indicators of overall survival and recurrence-free survival in PDAC patients.

7.2 S100A10 mRNA is highly expressed in pancreatic tumors and cell lines.

To assess the relative expression levels of the *S100A10* gene in cancer, we examined *S100A10* mRNA levels (RNA Seq V2) across all 33 cancer types in the Genomic Data Commons (GDC) portal of the National Cancer Institute (NCI) [641](see methods to normalization). *S100A10* mRNA expression in PDAC (n=179) was the third highest (mean = 1.959, C.I. 1.789-2.129) after Mesothelioma (n=87) (mean = 3.895, C.I. 3.501-4.290) and Head and Neck squamous cell carcinoma (n=801) (mean = 2.030, C.I 1.951-2.109) (figure 53a). We also examined *S100A10* mRNA levels (microarray z-scores) across all 930 human cancer cell lines listed in the CCLE (Cancer Cell Line Encyclopedia) from the Broad Institute (GSE36133) [642]. *S100A10* was highly expressed in many cancer cell lines including upper respiratory tract (n=30) (mean = 0.6671 C.I. 0.6314-0.7029), pancreatic (n=44) (mean = 0.6657, C.I. 0.5948-0.7366) and esophageal (n=25) (mean = 0.6542, C.I. 0.5838-0.7245) cancer cell lines (figure 53b). These results established that *S100A10* mRNA is highly expressed in many cancer types including pancreatic tumors and cell lines suggesting a possible role for *S100A10* in PDAC.

Figure 53. S100A10 mRNA is over-expressed in pancreatic TCGA tumors and CCLE cell lines. (a) S100A10 REVs (RNA Seq V2 RSEM) were extracted from Cbioportal and normalized by dividing by the mean REV of the 33 TCGA tumor types. (b) Z-scores of S100A10 of the 930 CCLE cell lines were extracted from Cbioportal. Z-ratios were used to determine significance with respect to CML (control). A z-ratio of 1.96 is equivalent to a p-value of 0.05.


a
Mesothelioma (n=87)
Head & Neck squamous cell carcinoma (n=801)
Pancreatic Adenocarcinoma (n=179)
Esophageal Carcinoma (n=185)
Cervical Squa. & Endocervical Adeno. (n=306)
Lung Adenocarcinoma (n=747)
Colorectal Adeno. (n=382)
Glioblastoma (n=172)
Bladder Urothelial Carcinoma (n=537)
Lung Squamous Cell Carcinoma (n=501)
Sarcoma (n=263)
Gastric Adeno. (n=706)
Renal Cell Carcinoma (n=1294)
Testicular Germ Cell (n=156)
Cholangiocarcinoma (n=36)
Melanoma (n=472)
Glioblastoma Multiforme (n=148)
Ovarian Serous Cystadenocarcinoma (n=307)
Hepatocellular Carcinoma (n=373)
Uterine Corpus Endometrial Carcinoma (n=510)
Breast Invasive Carcinoma (n=1917)
Uterine Carcinosarcoma (n=57)
Adrenocortical Carcinoma (n=79)
Lymphoid Diffuse Large B-cell Lymphoma (n=48)
Thyroid Carcinoma (n=509)
Papillary Thyroid Carcinoma (n=494)
Kidney Chromophobe (n=132)
Thymoma (n=120)
Prostate Adenocarcinoma (n=831)
Acute Myeloid Leukemia (n=346)
Uveal Melanoma (n=80)
Pheochromocytoma and Paraganglioma (n=184)
Glioma (n=530)
TCGA tumor type
0
2
4
6
8
Normalized mRNA expression (RNA Seq V2)


b
Z-ratio
Upper resp. tract (n=30) 2.96
Pancreatic (n=44) 2.95
Esophageal (n=25) 2.93
Kidney (n=21) 2.90
Colorectal (n=56) 2.87
Urinary tract (n=24) 2.86
Mesothelioma (n=9) 2.86
Gastric (n=38) 2.83
Non-small cell lung (n=115) 2.80
chondrosarcoma (n=4) 2.78
Glioma (n=46) 2.77
Bile duct (n=7) 2.75
Melanoma (n=59) 2.71
Ovarian (n=47) 2.67
Breast (n=56) 2.56
Liver (n=27) 2.53
endometrial (n=27) 2.51
Thyroid (n=11) 2.36
Prostate (n=7) 2.08
Soft tissue (n=17) 1.92
Ewings Sarcoma (n=9) 1.71
T-cell ALL (n=14) 1.17
Burkitt's lymphoma (n=67) 1.01
AML (n=34) 0.96
Neuroblastoma (n=26) 0.87
Medulloblastoma (n=4) 0.57
Multiple Myeloma (n=28) 0.48
SC lung carcinoma (n=52) 0.31
B-cell ALL (n=12) 0.28
CML (n=14) 0
Significant (p < 0.05)
Non-significant
CCLE Tumor type
-3
-2
-1
0
1
2
z-score

7.3 S100A10 is highly expressed in pancreatic tumors compared to adjacent non-ductal stroma and normal ducts.

After establishing that *S100A10* mRNA is highly expressed in pancreatic tumors and cell lines, we focused on studying its relevance in this cancer. For that purpose, we compared *S100A10* mRNA expression in normal and tumor samples from previously published DNA microarray and RNA Seq expression datasets. A consistent upregulation of *S100A10* mRNA was observed in pancreatic tumors compared to normal tissues of unmatched (figure 54a-54d, supplemental figure 11a, 11b) and matched (figure 54c, 54e, supplemental figure 11c) patients.

To gain further insight into *S100A10* expression in pancreatic tumors beyond mRNA levels, we examined protein expression in archived human pancreatic tumors using immunohistochemistry (IHC). The additional benefit of IHC is the ability to discern the type of tissue that is producing the S100A10 protein signal. Consistent with our mRNA analysis, S100A10 protein expression was also upregulated in cancerous tissues compared to nearby normal ducts (supplemental figure 12a) which could also be visualized within a single duct containing both normal and neoplastic ductal epithelia (supplemental figure 12b, 12c). We then constructed tissue microarrays (TMAs) to examine S100A10 protein expression of the entire PDAC patient cohort. Control, pre-cancerous lesions (PanINs), and cancerous lesions (PDAC) were selected from each tumor sample and assembled on TMA blocks which were then stained with an anti-S100A10 antibody. The quantification of protein expression on digitized slides was performed using the IHC profiler plugin in ImageJ as described in Verghese *et al*. [549]. Color deconvolution allowed the separation

of the DAB brown-colored stain from the Meyer's hematoxylin stain. The intensity and coverage of the DAB stain was then quantified by ImageJ (figure 55a) (see methods). Six different regions of each sample were quantified and were assigned to a three-tier score system: high positive (H-score > 200), low positive (100 < H-score < 200) and negative/weak (H-score < 100). Weak/negative staining was observed in 0% (0/88) of PDAC (cancerous lesions/ducts), 66.67% (38/57) of PanINs, 94.94% (75/79) of normal ducts adjacent to PDAC, 87.50% (49/56) of normal duct adjacent to PanINs, 100% (88/88) of PDAC non-ductal stroma, and 100% (63/63) of non-ductal PanINs stroma. Low positive staining was observed in 34.09% (30/88) of PDAC, 33.33% (19/57) of PanINs, 5.06% (4/79) of normal ducts adjacent to PDAC and 12.50% (7/56) of normal duct adjacent to PanINs. Importantly, we observed that high positive staining was exclusive to PDAC at 65.91% (58/88) (figure 55c). Collectively, the protein levels of S100A10 revealed a similar trend of upregulation in tumor tissue compared to normal tissue as observed at the mRNA level. Additionally, the immunohistochemistry results demonstrated that S100A10 protein is overexpressed in carcinoma (PDAC) regions compared to PanINs, normal ducts and non-ductal stroma.

Figure 54. S100A10 mRNA is overexpressed in pancreatic tumors compared to normal pancreatic tissue. Gene expression from six publically available gene expression datasets from Oncomine (a-c, e) and gene expression omnibus (GEO) (d, f) were extracted from the normalized data on Oncomine (a-c, e) and GEOR (d, f). The datasets compare gene expression in normal vs. tumor from pancreatic cancer patients. Badea et al. and Balasenthil et al. represent matched samples of pancreatic tumors and corresponding adjacent normal tissue. Significance was determined using unpaired (a-d) or paired (e, f) t-tests. Significance was determined based on a p-value of 0.05. Data are represented as means ± SD.

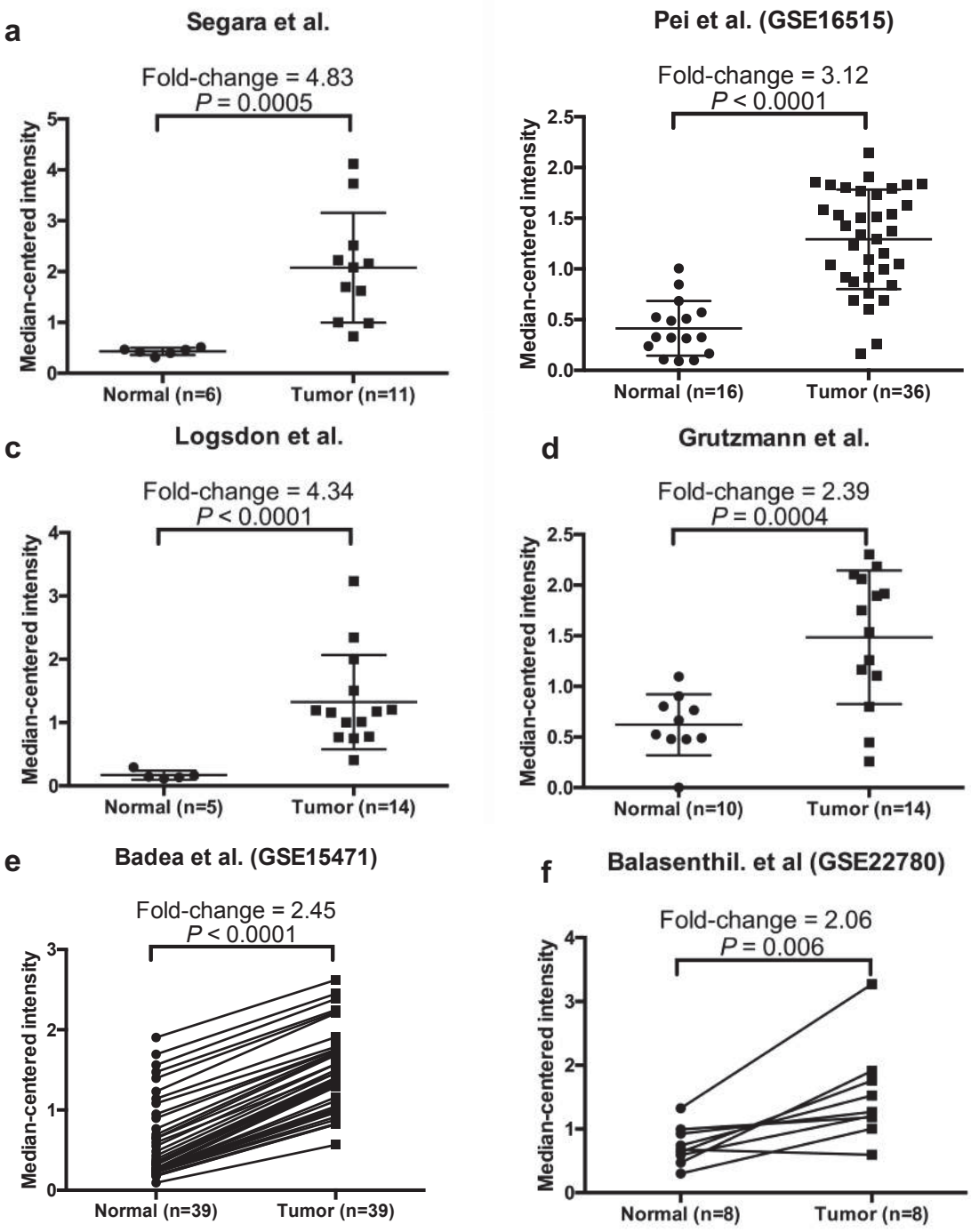
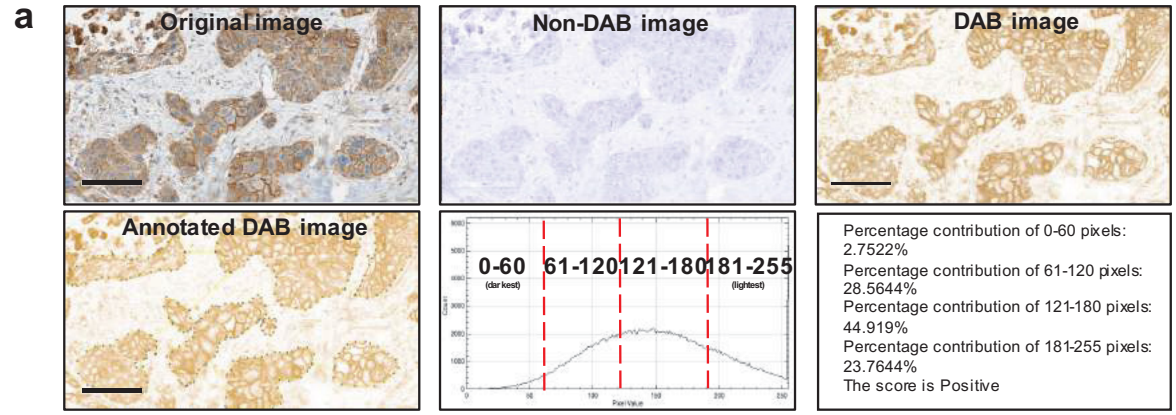
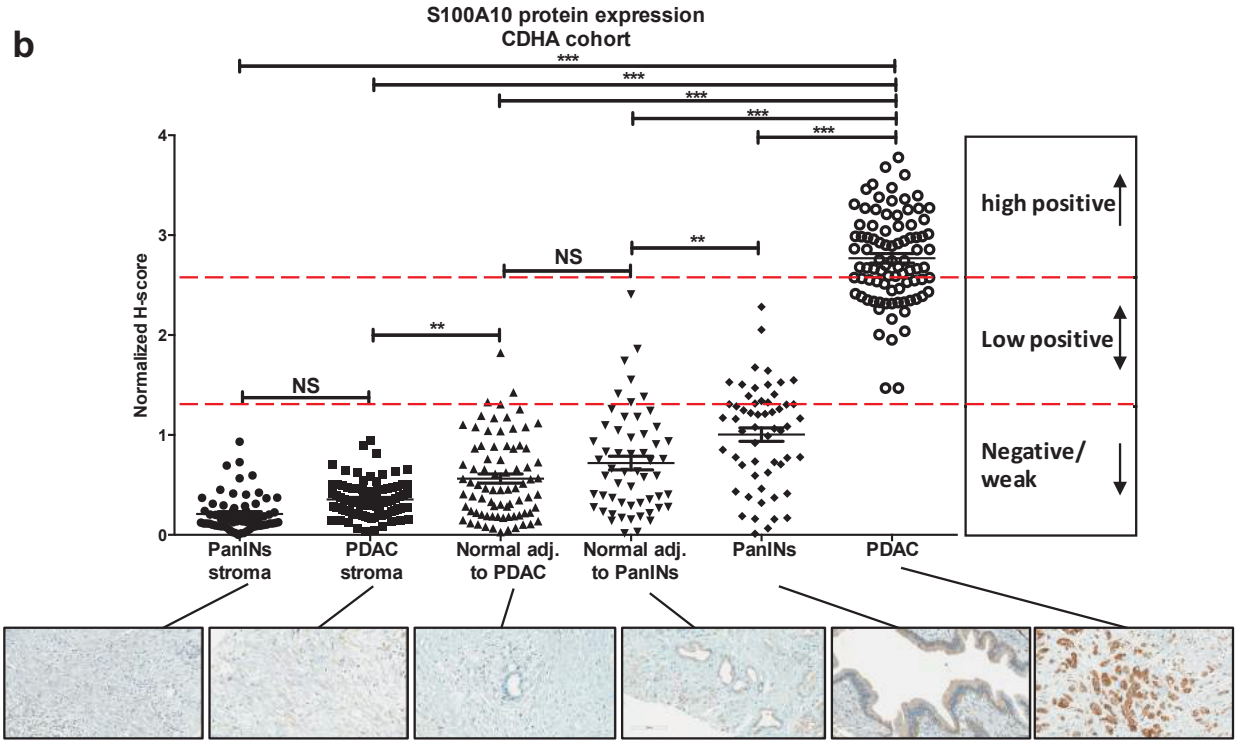

a
Segara et al.
Fold-change = 4.83
P = 0.0005
Median-centered intensity
Normal (n=6)
Tumor (n=11)
Pei et al. (GSE16515)
Fold-change = 3.12
P < 0.0001
Median-centered intensity
Normal (n=16)
Tumor (n=36)
c
Logsdon et al.
Fold-change = 4.34
P < 0.0001
Median-centered intensity
Normal (n=5)
Tumor (n=14)
d
Grutzmann et al.
Fold-change = 2.39
P = 0.0004
Median-centered intensity
Normal (n=10)
Tumor (n=14)
e
Badea et al. (GSE15471)
Fold-change = 2.45
P < 0.0001
Median-centered intensity
Normal (n=39)
Tumor (n=39)
f
Balasenthil. et al (GSE22780)
Fold-change = 2.06
P = 0.006
Median-centered intensity
Normal (n=8)
Tumor (n=8)

Figure 55. S100A10 protein overexpressed in pancreatic carcinoma (PDAC) lesions compared to pre-cancerous lesions, stroma and normal tissue. (a) ImageJ IHC profiler plugin was used to quantify S100A10 protein expression in 89 patients of the CDHA cohort. Briefly, images were color deconvoluted to expose the brown DAB stain. An area of interest was manually selected and quantified based on pixel intensity and the percentage contribution of each pixel sub-category (0-60, 61-120, 121-180, 181-255). (b) The graph demonstrates the S100A10 protein expression quantified by ImageJ in six different regions of patient cores. Each H-score was divided by the mean H-score of all measurements to yield a mean-normalized H-score ± SEM. Significance was determined using one-way ANOVA of unmatched samples (non-paired). Scale bars, 100 **μ**m.


a
Original image
Non-DAB image
DAB image
Annotated DAB image
0-60
(darkest)
61-120
121-180
181-255
(lightest)
Percentage contribution of 0-60 pixels: 2.7522%
Percentage contribution of 61-120 pixels: 28.5644%
Percentage contribution of 121-180 pixels: 44.919%
Percentage contribution of 181-255 pixels: 23.7644%
The score is Positive


b
S100A10 protein expression
CDHA cohort
Normalized H-score
0
1
2
3
4
***
**
NS
high positive
Low positive
Negative/ weak
PanINs stroma
PDAC stroma
Normal adj. to PDAC
Normal adj. to PanINs
PanINs
PDAC

7.4 S100A10 mRNA expression and copy number are predictive of overall and recurrence-free survival in PDAC patients.

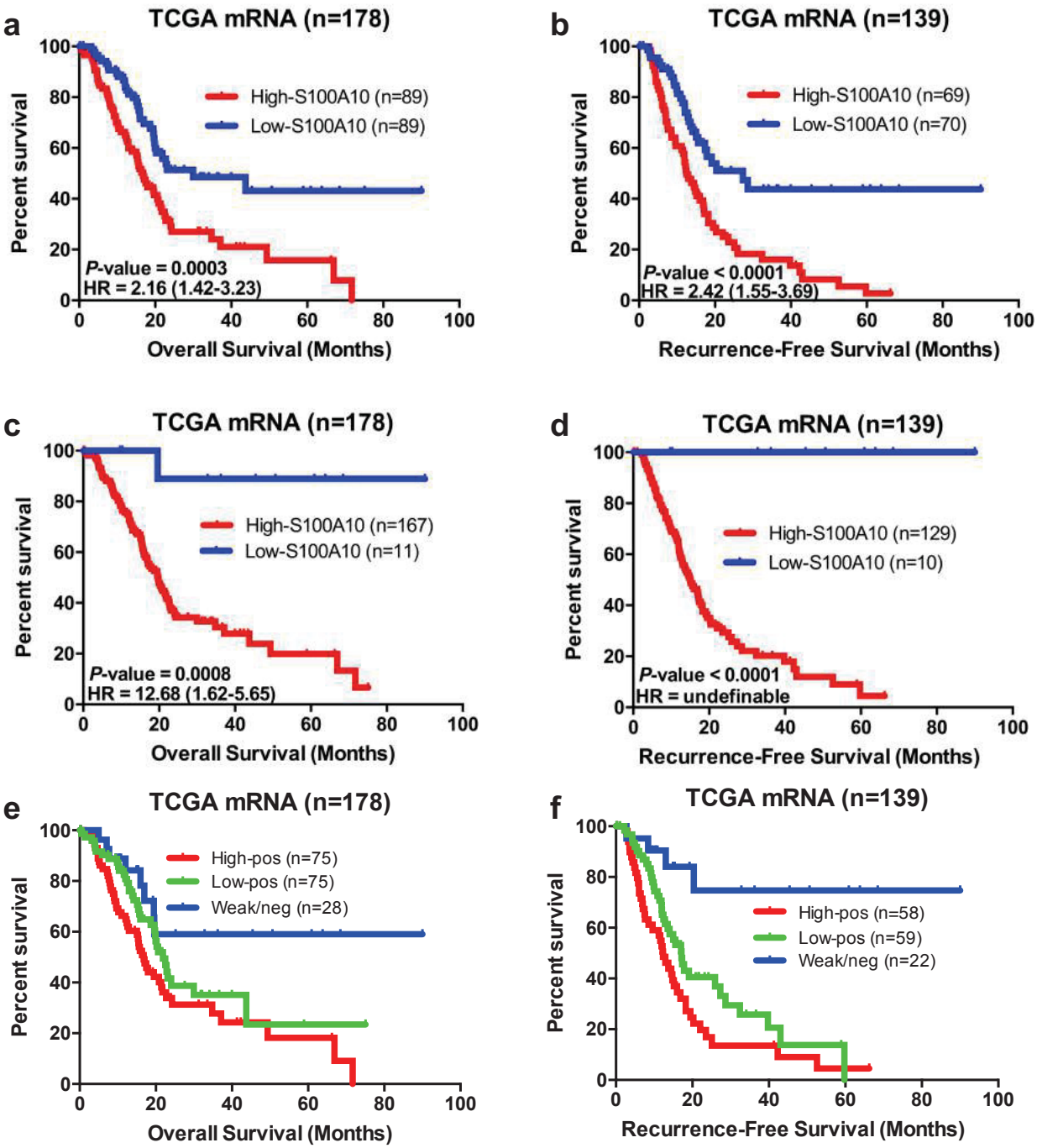
Having established S100A10 upregulation in PDAC, we examined the potential clinical significance of S100A10 in the prognosis of the TCGA provisional PDAC patient cohort. The latter contains genomic profiles of up to 178 PDAC patients with clinical data, RNA-Seq V2 expression data, HM450 methylation data and GISTIC copy number alterations. To assess the prognostic value of *S100A10* mRNA expression, Kaplan Meier survival analysis was performed on patients using three cut-off classifiers (median cut-off, optimal cut-off and ternary cut-off) (supplemental figure 13a-13c). A median cut-off (raw expression value (REV) > or < median) (supplemental figure 13a) revealed that *S100A10* mRNA expression is predictive of both overall survival (OS; HR=2.16, *p*-value=0.0003, n=178) and recurrence-free survival (RFS; HR=2.42, *p*-value<0.0001, n=139) (figure 56a, 56b). High-*S100A10* mRNA levels also predicted poorer long-term survival and patients were more likely to recur over the 90-month follow-up period. In addition, one-, three- and five-year survivals in low-*S100A10* patients (e.g. 1yr OS:69.66%, 1-yr RFS: 58.57%) were significantly higher than that in high-*S100A10* patients (e.g. 1-yr OS: 59.55%, 1-yr RFS: 49.28%) (Supplemental table 17).

Although a median cut-off resulted in a strong correlation between OS and RFS and *S100A10* mRNA expression, we attempted to utilize a more optimal cut-off that would allow a strict binary classification of high and low expressors (supplemental figure 13b). The cut-off finder tool previously described by Budczies *et al*. identified a new binary classifier with a high-risk group (93.82%) with high expression of *S100A10* mRNA

(REV>3790.9211) and low-risk group (6.18%) with considerably low expression of *S100A10* mRNA (REV<=3790.9211) (figure 56c) [643]. The low-risk group had a favorable long-term OS. Applying the same REV cut-off to the RFS data revealed a similar trend where low-risk patients were unlikely to develop recurrent disease compared to high risk patients (figure 56d). To further test the prognostic performance of S100A10 and bypass the conservative and biased approach of optimal cut-offs, we developed a ternary classifier based on the frequency distribution of REVs in the TCGA cohort (supplemental figure 13c) (see methods). The ternary classification identified three subgroups of patients; a weak/neg group with a favorable OS and RFS outcomes (*p*-values of 0.0039 and <0.0001 compared to high pos) and two largely indifferent groups (low-pos and high pos) with less favorable outcomes (figure 56e, 56f) (supplemental table 18).

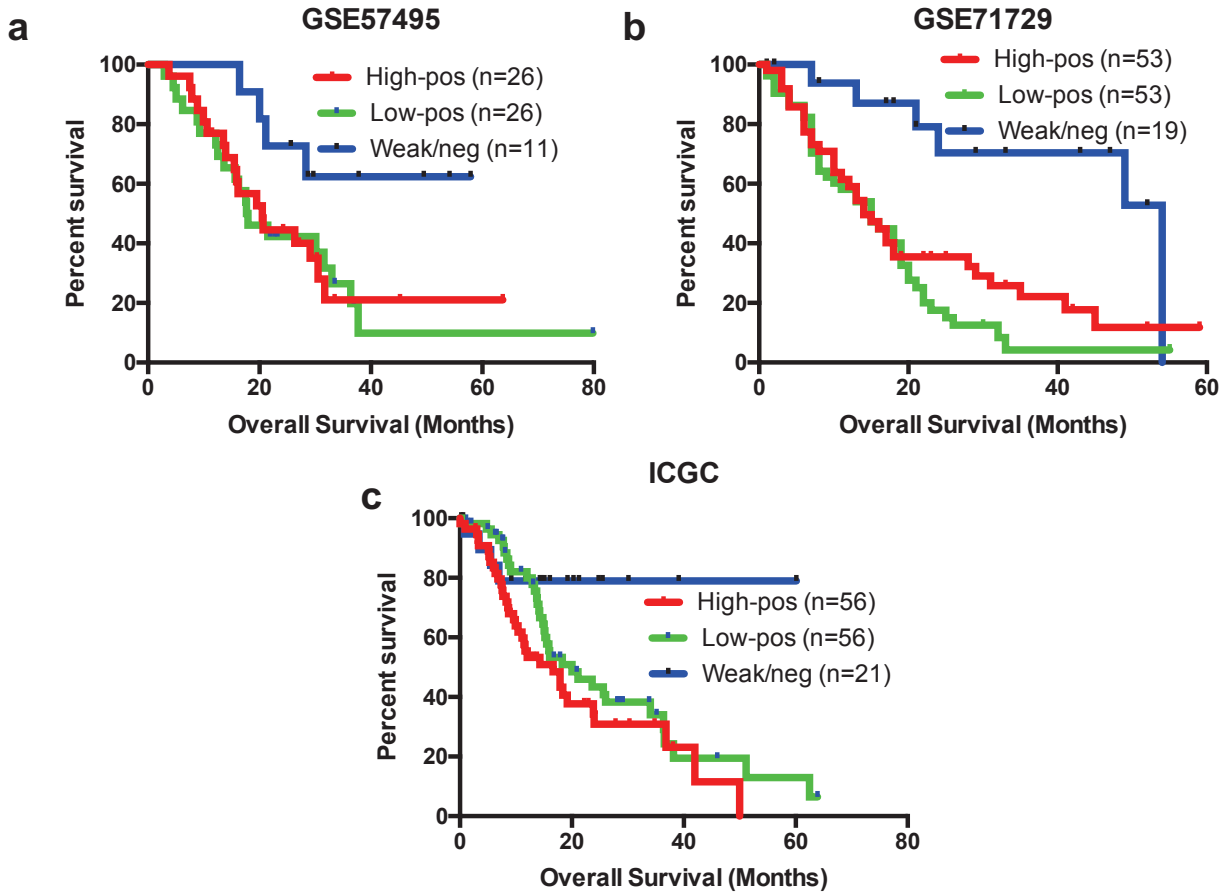
To confirm the existence of the low-risk group (weak/neg), we applied the same ternary classifier to three additional independent PDAC studies: Chen *et al*. (GSE57495, n=63) [644], Moffitt *et al*. (GSE71729, n=125)[532] and ICGC (international cancer genome consortium, n=133) [645]. Kaplan Meier survival curves revealed a similar trend of survivability to that seen in the TCGA PDAC cohort. An equivalent low-risk group with favorable OS emerged in Chen *et al*. (figure 57a, *p*-value = 0.0402), in Moffitt *et al*. (figure 57b, *p*-value = 0.0026) and in ICGC (figure 57c, *p*-value = 0.0073) cohorts when compared to the high-pos group (supplemental table 18). Collectively, these survival analyses showed that low expression of *S100A10* mRNA can serve as a strong predictor of favorable short- and long-term survival in PDAC patients.

Figure 56. S100A10 mRNA expression is predictive of overall and recurrence-free survival in TCGA PDAC patient cohort. Kaplan Meier (KM) plots of overall survival (n=178) (a, c, e) and recurrence-free survival (n=139) (b, d, f) of patients based on their S100A10 mRNA expression. A three-tier method of classification was used; A median cut-off (a, b), best cut-off (c, d), and a ternary cut-off (e, f) (see supplemental figure 13). Optimal cut-offs were determined using the cut off finder database (http://molpath.charite.de/cutoff/) Budczies et al. (2012), PLoS ONE 7(12): e51862. In summary, patients with low levels of S100A10 mRNA had a better overall and recurrence-free survival than those with high S100A10 mRNA levels.


a
TCGA mRNA (n=178)
Percent survival
High-S100A10 (n=89)
Low-S100A10 (n=89)
P-value = 0.0003
HR = 2.16 (1.42-3.23)
Overall Survival (Months)
b
TCGA mRNA (n=139)
Percent survival
High-S100A10 (n=69)
Low-S100A10 (n=70)
P-value < 0.0001
HR = 2.42 (1.55-3.69)
Recurrence-Free Survival (Months)
c
TCGA mRNA (n=178)
Percent survival
High-S100A10 (n=167)
Low-S100A10 (n=11)
P-value = 0.0008
HR = 12.68 (1.62-5.65)
Overall Survival (Months)
d
TCGA mRNA (n=139)
Percent survival
High-S100A10 (n=129)
Low-S100A10 (n=10)
P-value < 0.0001
HR = undefinable
Recurrence-Free Survival (Months)
e
TCGA mRNA (n=178)
Percent survival
High-pos (n=75)
Low-pos (n=75)
Weak/neg (n=28)
Overall Survival (Months)
f
TCGA mRNA (n=139)
Percent survival
High-pos (n=58)
Low-pos (n=59)
Weak/neg (n=22)
Recurrence-Free Survival (Months)

Since the Kaplan Meier analysis of *S100A10* mRNA expression correlated with OS and RFS of PDAC patients within the TCGA cohort, we decided to examine whether *S100A10* gene copy number showed similar correlations. The rationale was driven by the fact that *S100A10* mRNA expression significantly correlated with the copy number score (supplemental figure 14a) and status (supplemental figure 14b) in these patients. Higher *S100A10* copy number score correlated with poorer OS (HR=1.816, *p*-value = 0.0357, n=176) (supplemental figure 14c) and RFS (HR=1.691, *p*-value = 0.0190, n=139) (supplemental figure 14d). Short-term OS and RFS after one, three and five years post diagnosis also correlated with *S100A10* copy number score (supplemental table 17). In an attempt to complement the copy number score-based stratification, patients were also stratified based on *S100A10* copy number status (i.e. deletion, diploid, gain, or amplification). Patients with *S100A10* amplifications had a noticeably shorter OS and RFS compared to patients with S100A10 deletions (supplemental figure 14e, 14f) respectively). The usage of mRNA levels as a predictive marker is supported by the fact that *S100A10* copy number also possessed similar predictive potential in the same cohort.

Figure 57. S100A10 mRNA expression is predictive of overall survival in three independent PDAC patient cohorts. Kaplan Meier (KM) plots of overall survival in two independent cohorts of pancreatic cancer patients by Chen et al. (GSE57495, 2015) (top left), Moffitt et al. (GSE71729, 2015) (top right) and ICGC (bottom). The Ternary cut-off was applied to classify the high-pos, low-pos and weak/neg subgroups. P-values were adjusted to the Bonferroni-corrected threshold. Adjusted p-value is p-value/K = 0.017 where K=3 and represents the number of comparisons made.


a
GSE57495
High-pos (n=26)
Low-pos (n=26)
Weak/neg (n=11)
Percent survival
Overall Survival (Months)
b
GSE71729
High-pos (n=53)
Low-pos (n=53)
Weak/neg (n=19)
Percent survival
Overall Survival (Months)
c
ICGC
High-pos (n=56)
Low-pos (n=56)
Weak/neg (n=21)
Percent survival
Overall Survival (Months)

7.5 S100A10 mRNA and lymph node positivity are linked predictors of overall and recurrence-free survival.

To understand the relationship between *S100A10* mRNA and other clinical covariates, we applied univariate and multivariate regression models. Single variable analysis using the Wald test showed that five variables were predictive of OS: *S100A10* mRNA (HR=1.79, C.I. 1.30-2.46, *p*-value=0.00038), age (HR=1.03, C.I. 1.01-1.05, *p*-value=0.008), grade II (HR=2.00, C.I. 1.07-5.08, *p*-value=0.041), grade III (HR=2.55, C.I. 1.26-5.14, *p*-value 0.009) lymph node positivity (HR=2.09, C.I. 1.24-3.51, *p*-value=0.005) and stage II (HR=2.33, C.I. 1.07-5.08, *p*-value=0.03). Although age as a single variable was a significant predictor of OS, the hazard ratio was marginal (table 11). The likelihood ratio test for all five variables revealed that *S100A10* mRNA (*p*-value=0.0001), age (*p*-value=0.007) and lymph node positivity (*p*-value=0.003) were significant but not tumor grade (*p*-value=0.111). In contrast, multivariate regression fitting re-confirmed the prognostic significance of *S100A10* mRNA (HR=1.59, C.I. 1.07-2.35), lymph node positivity (HR=2.17, C.I. 1.09-4.35) and age (HR=1.02, C.I. 1.001-1.044) (table 12). An ANOVA test of these variables validated their predictive power (p-values 0.007, 0.003 and 0.034 respectively). A final model using these three variables was then derived which shows that for every exponential unit increase (Y=*e*X, where *e=2.718*) in *S100A10* mRNA REV, the likelihood of dying is 1.54 higher (C.I. 1.07-2.21, *p*-value=0.02). Similarly, being lymph node positive increase risk of death by 1.93 times (C.I. 1.15-3.24, *p*-value=0.01). The effect of age on this model is minor although statistically significant. The risk of death is 2.97 times higher in lymph node-positive patient with one unit increase in *S100A10*

mRNA (i.e. REV=Y) compared to a lymph node-negative patient with lower *S100A10* mRNA (REV=X) (supplemental table 19).

Univariate and multivariate regression models of RFS functions were also generated. The single variable analysis using the Wald test showed that *S100A10* mRNA (HR=2.12, C.I. 1.52-2.94, p-value=7.89e-06), grade II (HR=2.14, C.I. 1.08-4.23, *p*-value 0.029), grade III (HR=3.29, C.I. 1.61-6.71, *p*-value=0.001) and lymph node positivity (HR=1.79, C.I. 1.10-2.94, *p*-value=0.018) were predictive of RFS (table 13). The likelihood ratio test rendered S100A10 mRNA (*p*-value=8.97e-07), grade (*p*-value =0.0043), lymph node positivity (p-value 0.0143) as the only significant variables. Subsequent multivariate analysis revealed that only *S100A10* mRNA (HR=1.71, C.I. 1.12-2.61) and lymph node positivity (HR=1.96, C.I. 1.00-3.84) were the only significant predictors of RFS (table 14). ANOVA tests showed of the above variables showed that only *S100A10* mRNA and lymph node positivity were the only two significant predictors of RFS (p-values 0.0003 and 0.02 respectively. Thus, a final two-variable model was derived which predicts that the likelihood of recurrence is 1.89 times higher for every unit increase in S100A10 mRNA. The recurrence rate also increases by 1.54 times in lymph node-positive patients. Consequently, a lymph node-positive patient with one unit increase in S100A10 mRNA is 2.91 times more likely to recur than a lymph node-negative patient with lower S100A10 mRNA (supplemental table 19). These results established that S100A10 mRNA and lymph node status are linked co-variates and are strong predictors of OS and RFS in PDAC patients.

Table 11. Univariate cox regression analysis of overall survival (OS) of the TCGA PDAC cohort. Abbreviations are as follows: Coef: beta coefficients, exp(coef): exponential of the coefficient, se(coef): standard error of the coefficient, z: Z statistics to test coefficient =0, Pr(> |z|): P-value based on the Wald test to test coefficient =0, Exp(-coef): exponential of the negative coefficient, Lower .95 and upper .95: the lower and upper limits for the 95% CI for exp(coef). Univariate regression models were fitted to the overall survival (OS) of the TCGA PDAC patient cohort. The variables/predictors are: S100A10 mRNA (RNA Seq V2 RSEM), gender, race, age, grade, tumor dimension, stage, metastasis, smoking and alcohol consumption.

Univariate analysis									
Variable	coef	exp(coef)	se(coef)	z	Pr(> |z|)	exp(-coef)	lower .95	upper .95	Significant (Y/N)
S100A10 mRNA	0.58	1.79	0.16	3.55	0.00038	0.56	1.3	2.46	Y
Gender									
Female (n=80)	0	1	-	-	-	-	-	-	
Male (n=97)	-0.22	0.81	0.21	-1.03	0.3	1.24	0.53	1.21	N
Race									
White (n=156)	0	1	-	-	-	-	-	-	
Asian (n=11)	-0.23	0.79	0.46	-0.5	0.62	1.26	0.32	1.97	N
Black/African American (n=6)	-0.04	0.96	0.51	-0.08	0.94	1.04	0.35	2.62	N
age	0.03	1.03	0.01	2.65	0.008	0.97	1.01	1.05	Y
Grade									
Grade I (n=31)	0	1	-	-	-	-	-	-	
Grade II (n=95)	0.69	2	0.34	2.04	0.04	0.5	1.03	3.88	Y
Grade III (n=47)	0.93	2.55	0.36	2.61	0.01	0.39	1.26	5.14	Y
Grade IV (n=2)	0.51	1.67	1.05	0.49	0.62	0.6	0.21	13.05	N
Tumor dimension (n=164)	0	1	0.06	0.09	0.93	1	0.9	1.12	N
Lymph node involvement									
N0 (negative, n=49)	0	1	-	-	-	-	-	-	
N1 (positive, n=123)	0.74	2.09	0.26	2.78	0.01	0.48	1.24	3.51	Y
Metastasis									
M0 (no mets, n=79)	0	1	-	-	-	-	-	-	
M1 (mets, n=4)	-0.07	0.94	0.73	-0.09	0.93	1.07	0.23	3.88	N
Stage									
Stage I (n=21)	0	1	-	-	-	-	-	-	
Stage II (n=146)	0.85	2.33	0.40	2.12	0.03	0.43	1.07	5.08	Y
Stage III (n=3)	0.23	1.26	1.07	0.22	0.83	0.79	0.15	10.32	N
Stage IV (n=5)	0.77	2.15	0.81	0.95	0.34	0.46	0.44	10.51	N
Smoking (n=56)	0	1	0.01	-0.05	0.96	1	0.98	1.02	N
alcohol consumption (Yes, n=101**/No,n=**64**)**	-0.1	0.91	0.22	-0.44	0.66	1.1	0.58	1.41	N

.

Table 12. Multivariate cox regression analysis of overall survival (OS) of the TCGA PDAC cohort. The fitted multivariate model for predicting OS included all variables except smoking history and alcohol consumption due to high number of missing values on these two variables. A semi-parametric proportional hazard regression model was fitted to identify variables that are predictors of survival time.

Multivariate analysis									
Variable	coef	exp(coef)	se(coef)	z	Pr(> |z|)	exp(-coef)	lower .95	upper .95	Significant (Y/N)
S100A10 mRNA	0.46	1.58	0.2	2.27	0.02	0.63	1.07	2.35	Y
Gender									
Female (n=80)	0	1	-	-	-	-	-	-	
Male (n=97)	-0.35	0.71	0.23	-1.5	0.13	1.42	0.45	1.11	N
Race									
White (n=156)	0	1	-	-	-	-	-	-	
Asian (n=11)	-1.81	0.16	1.61	-1.12	0.26	6.09	0.01	3.83	N
Black/African American (n=6)	0.51	1.66	0.54	0.93	0.35	0.6	0.57	4.8	N
age	0.02	1.02	0.01	1.33	0.18	0.99	0.99	1.04	N
Grade									
Grade I (n=31)	0	1	-	-	-	-	-	-	
Grade II (n=95)	0.29	1.34	0.39	0.74	0.46	0.75	0.62	2.87	N
Grade III (n=47)	0.4	1.49	0.41	0.98	0.33	0.67	0.67	3.29	N
Grade IV (n=2)	-0.17	0.85	1.07	-0.16	0.88	1.18	0.1	6.91	N
Tumor dimension (n=164)	0.06	1.06	0.07	0.8	0.42	0.94	0.92	1.22	N
Lymph node involvement									
N0 (negative, n=49)	0	1	-	-	-	-	-	-	
N1 (positive, n=123)	0.78	2.18	0.36	2.18	0.03	0.46	1.08	4.39	Y
Metastasis									
M0 (no mets, n=79)	0	1	-	-	-	-	-	-	
M1 (mets, n=4)	-0.09	0.92	1.27	-0.07	0.95	1.09	0.08	11.14	N
Stage									
Stage I (n=21)	0	1	-	-	-	-	-	-	
Stage II (n=146)	-0.43	0.65	0.56	-0.77	0.44	1.54	0.22	1.96	N
Stage III (n=3)	-0.49	0.61	1.09	-0.45	0.65	1.63	0.07	5.23	N
Stage IV (n=5)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	

Table 13. Univariate cox regression analysis of Recurrence-free survival (RFS) of the TCGA PDAC cohort. Univariate regression models were fitted to the recurrence-free survival (RFS) of the TCGA PDAC patient cohort. The variables/predictors are: S100A10 mRNA (RNA Seq V2 RSEM), gender, race, age, grade, tumor dimension, stage, metastasis, smoking and alcohol consumption.

Univariate analysis									
Variable	coef	exp(coef)	se(coef)	z	Pr(> |z|)	exp(-coef)	lower .95	upper .95	Significant (Y/N)
S100A10 mRNA	0.75	2.12	0.17	4.47	0	0.47	1.52	2.94	Y
Gender									
Female (n=63)	0	1							
Male (n=78)	-0.17	0.84	0.22	-0.77	0.44	1.18	0.55	1.3	N
Race									
White (n=123)	0	1	-	-	-	-	-	-	N
Asian (n=8)	-0.02	0.98	0.46	-0.05	0.96	1.02	0.39	2.43	N
Black/African American (n=5)	0.26	1.3	0.52	0.51	0.61	0.77	0.47	3.57	N
age	0.02	1.02	0.01	1.8	0.07	0.98	1	1.04	N
Grade									
Grade I (n=28)	0	1	-	-	-	-	-	-	
Grade II (n=72)	0.76	2.14	0.35	2.19	0.029	0.47	1.08	4.23	Y
Grade III (n=37)	1.19	3.29	0.36	3.27	0.001	0.3	1.61	6.71	Y
Grade IV (n=2)	0.35	1.42	1.05	0.33	0.74	0.71	0.18	11.11	N
Tumor dimension (n=127)	-0.02	0.98	0.06	-0.33	0.74	1.02	0.87	1.1	N
Lymph node involvement									
N0 (negative, n=43)	0	1	-	-	-	-	-	-	
N1 (positive, n=95)	0.59	1.80	0.25	2.36	0.018	0.56	1.10	0.59	Y
Metastasis									
M0 (no mets, n=71)	0	1	-	-	-	-	-	-	N
M1 (mets, n=3)	-0.13	0.88	0.72	-0.18	0.86	1.14	0.21	3.63	N
Stage									
Stage I (n=21)	0	1	-	-	-	-	-	-	
Stage II (n=114)	1.02	2.77	0.40	2.53	0.01	0.36	1.26	6.12	Y
Stage III (n=4)	0.88	2.41	1.08	0.82	0.41	0.42	0.29	19.82	N
Stage IV (n=3)	1.03	2.80	0.81	1.27	0.21	0.36	0.57	13.73	N
Smoking (n=43)	0	1	0.01	0.05	0.96	1	0.98	1.02	N
alcohol consumption (Yes, n=82/**No,**n=49**)**	-0.2	0.81	0.24	-0.85	0.4	1.23	0.51	1.31	N

Table 14. Multivariate cox regression analysis of Recurrence-free survival (RFS) of the TCGA PDAC cohort. The fitted multivariate model for predicting RFS included all variables except smoking history and alcohol consumption due to high number of missing values on these two variables. A semi-parametric proportional hazard regression model was fitted to identify variables that are predictors of survival time.

Multivariate analysis									
Variable	coef	exp(coef)	se(coef)	z	Pr(> |z|)	exp(-coef)	lower .95	upper .95	Significant (Y/N)
S100A10 mRNA	0.54	1.71	0.22	2.49	0.01	0.58	1.12	2.61	Y
Gender									
female	0	1	-	-	-	-	-	-	
male	-0.27	0.76	0.26	-1.03	0.3	1.31	0.46	1.27	N
Race									
White	0	1	-	-	-	-	-	-	
Asian	-16.43	0	2767.98	-0.01	1	0.39	0	Inf	N
Black/African american	0.89	2.44	0.64	1.39	0.16	0.41	0.7	8.54	N
age	0.01	1.01	0.01	0.48	0.63	0.99	0.98	1.03	N
Grade									
Grade I (n=32)	0	1							
Grade II (n=97)	0	1	0.42	0	1	1	0.44	2.25	N
Grade III (n=50)	0.34	1.41	0.44	0.79	0.43	0.71	0.6	3.31	N
Grade IV (n=5)	-0.73	0.48	1.08	-0.68	0.49	2.09	0.06	3.94	N
Tumor dimension	0.03	1.03	0.08	0.35	0.73	0.97	0.88	1.2	N
Lymph node involvement									
N1 (positive, n=129)	0	1	-	-	-	-	-	-	
N0 (negative, n=50)	-0.68	0.51	0.34	-1.97	0.05	1.97	0.26	1	Y
Metastasis									
M0 (no mets, n=84)	0	1	-	-	-	-	-	-	
M1 (no mets, n=5)	-1.01	0.37	1.3	-0.78	0.44	2.74	0.03	4.65	N
Stage									
Stage I (n=21)	0	1	-	-	-	-	-	-	
Stage II (n=114)	-0.46	0.63	1.1	-0.42	0.67	1.59	0.07	5.44	N
Stage III (n=4)	0.46	1.59	1.10	0.42	0.68	0.63	0.18	13.7	N
Stage IV (n=3)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	

7.6 S100A10 methylation status is predictive of overall and recurrence-free survival in PDAC patients.

The availability of HM450 methylation data of the TCGA cohort enabled us to address the methylation status of the *S100A10* gene and importantly its correlation with *S100A10* mRNA. Fifteen probes mapped to the *S100A10* gene and promoter regions as illustrated in figure 58a. Although the *S100A10* gene is encoded on the negative strand (-), four probes mapped to the opposite positive (+) strand. Five probes were mapped to TSS1500 (region between 200bp and 1500bp upstream of transcription start site (T_cSS)), three to TSS200 (200bp upstream of T_cSS) and seven probes to the 5'UTR (5' untranslated region) (figure 58a). We also identified all the CpG sites corresponding to each probe (supplemental table 20). Since mRNA and protein levels were significantly higher in PDAC tumors compared to normal tissue, we examined the HM450 β values in both normal (n=9) and tumor (n=85) tissues of the TCGA cohort [646]. Six probes met the criteria of 1) being differentially hypo-methylated in tumor tissue compared to normal tissue and 2) negatively correlated with *S100A10* mRNA expression (figure 58b). The remaining probes were not hypo-methylated in tumors and/or did not negatively correlate with mRNA expression (supplemental figure 15). The third criterion was to discern which of the six probes was predictive of patient survival in PDAC cohorts. Kaplan Meier survival analysis using the ternary classifier showed that high β values of the probes cg13249591 and cg13445177 predicted that a low-risk group of patient with high-methylation score of S100A10 had favorable OS (figure 59a and 59b respectively) and RFS (figure 59c and 59d respectively) compared to the groups with moderate and low methylation scores (supplemental table 21). Similar trends in predicting OS and RFS were seen using the

median and optimal cut-offs (supplemental figure 18). The OS and RFS curves of the remaining four probes are shown in supplemental figures 16 and 17. Noteworthy, under the optimal cut-off conditions, there was an 81.82% (9/11) patient concordance in the low-risk groups and 98.8% (165/167) in the high-risk groups between mRNA and cg13445177 methylation assessments of OS (figure 56c) (figure 59c). Meanwhile, RFS assessment revealed 90% (9/10) and 99.22% (128/129) concordances in the low-risk and high-risk groups respectively (figure 56d) (figure 59d). In addition, the low and intermediate groups were also largely indifferent in terms of OS and RFS (supplemental table 11) (figure 59). We then assessed both probes in the ICGC methylation dataset using the same ternary classifier which also yielded similar OS pattern (figure 59e, 59d). To ensure that the high β values in the patient subgroup with high methylation scores were not due to global increase in methylation by the *de novo* methyl transferases [647], we compared the mRNA expression of these DNMTs with β values of the two probes. No positive correlation was observed between the two probes and mRNA expression of *DNMT1*, *DNMT3A* or *DNMT3B* (supplemental figure 19a, 19b).

Figure 58. Differentially-methylated CpG sites negatively correlate with S100A10 mRNA expression. (a) Schematic illustration of the human S100A10 gene based on UCSC (University of California San Diego) RefSeq. The genomic distance is approximate but is not drawn to scale. T_cSS: transcription start site, T_LSS: translation start site, TSS1500: region between 200bp and 1500bp upstream of T_cSS, TSS200: region 200bp upstream of T_cSS, 5'UTR: 5' untranslated region. (b) The β values of each probe were assessed in 85 PDA tumors and 9 normal tissues. The raw data was extracted from MethHC (http://methhc.mbc.nctu.edu.tw/php/index.php) which was described by Huang et al. (2015). Nucleic Acids Res. (database issue): D856-61. Raw β values of individual probes were extracted from Maplab Wanderer (http://maplab.imppc.org/wanderer/) (Villanueva et al. 2015); Epigenetics Chromatin. 8:22 (eCollection 2015) and plotted against RNA Seq (RSEM) expression values of S100A10 in matched patients. Pearson correlation was used to generate correlation graphs of β values and S100A10 mRNA expression. β values for the probe cg06786599 were absent for normal samples and no significant correlation (*p*-value = 0.1023) between S100A10 tumor mRNA and cg06786599 β values was found. Cg06786599 was then excluded from further analysis.

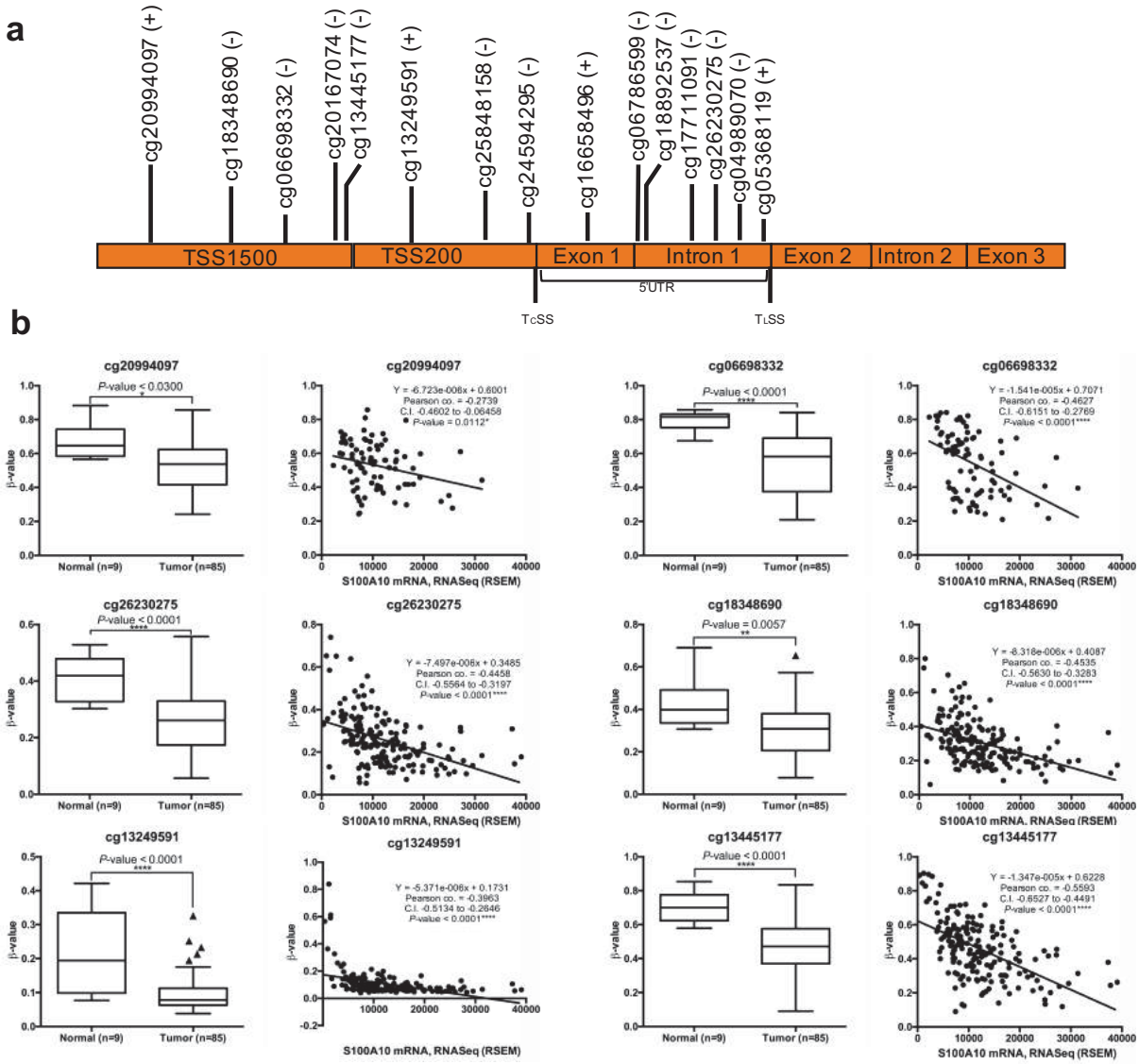
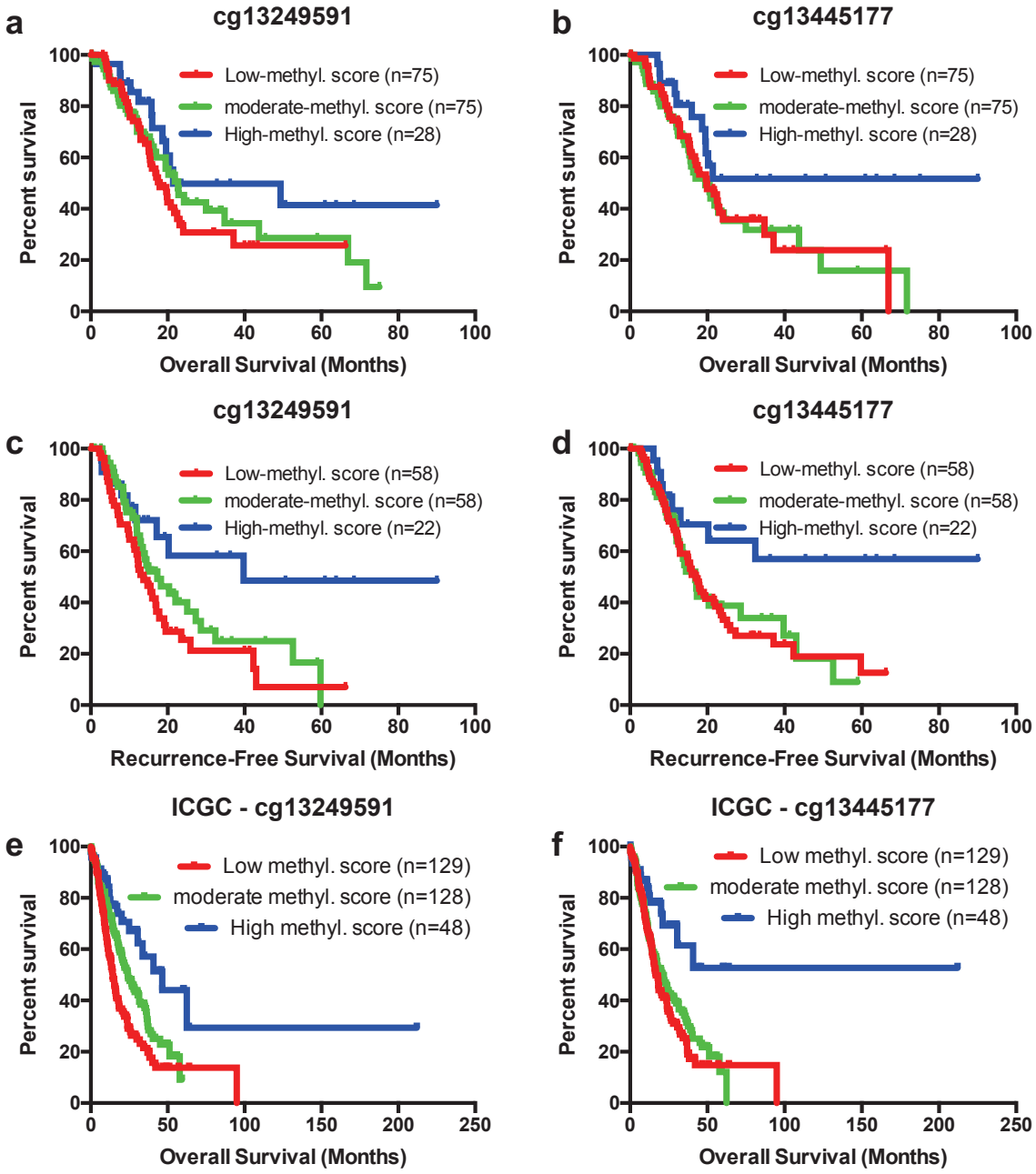

a
cg20994097 (+)
cg18348690 (-)
cg06698332 (-)
cg20167074 (-)
cg13445177 (-)
cg13249591 (+)
cg25848158 (-)
cg24594295 (-)
cg16658496 (+)
cg06786599 (-)
cg18892537 (-)
cg17711091 (-)
cg26230275 (-)
cg04989070 (-)
cg05368119 (+)
TSS1500
TSS200
Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
5UTR
TcSS
TLSS
b
cg20994097
P-value < 0.0300
Normal (n=9)
Tumor (n=85)
β-value
Y = -6.723e-006x + 0.6001
Pearson co. = -0.2739
C.I. -0.4602 to -0.06458
P-value = 0.0112*
S100A10 mRNA, RNASeq (RSEM)
cg06698332
P-value < 0.0001
Y = -1.541e-005x + 0.7071
Pearson co. = -0.4627
C.I. -0.6151 to -0.2769
P-value < 0.0001****
cg26230275
P-value < 0.0001
Y = -7.497e-006x + 0.3485
Pearson co. = -0.4458
C.I. -0.5564 to -0.3197
P-value < 0.0001****
cg18348690
P-value = 0.0057
Y = -8.318e-006x + 0.4087
Pearson co. = -0.4535
C.I. -0.5630 to -0.3283
P-value < 0.0001****
cg13249591
P-value < 0.0001
Y = -5.371e-006x + 0.1731
Pearson co. = -0.3963
C.I. -0.5134 to -0.2646
P-value < 0.0001****
cg13445177
P-value < 0.0001
Y = -1.347e-005x + 0.6228
Pearson co. = -0.5593
C.I. -0.6527 to -0.4491
P-value < 0.0001****

Figure 59. CpG islands corresponding to probes cg13249591 and cg13445177 are predictors of patient survival in the TCGA and ICGC PDAC cohorts. Kaplan Meier (KM) plots of (a, b) overall survival (n=178) and (c, d) recurrence-free survival (n=139) based on β values of the (a, c) cg13249591 and (b, d) cg13445177 CpG sites. OS in the ICGC cohort was assessed based on the β values of both probes (e, f). The same three-tier method of classification was used (see supplemental figure 13); Data where a ternary cut-off was used is shown above. Raw β values of individual probes were extracted from Maplab Wanderer (Villanueva et al. 2015). Epigenetics Chromatin. 8:22 (eCollection 2015) matched with OS and RFS of TCGA PDAC patients. Statistical analysis was performed using Bonferroni-corrected p-values (see methods). The p-values are listed in supplemental table 21.


a
cg13249591
Low-methyl. score (n=75)
moderate-methyl. score (n=75)
High-methyl. score (n=28)
Percent survival
Overall Survival (Months)
b
cg13445177
Low-methyl. score (n=75)
moderate-methyl. score (n=75)
High-methyl. score (n=28)
Percent survival
Overall Survival (Months)
c
cg13249591
Low-methyl. score (n=58)
moderate-methyl. score (n=58)
High-methyl. score (n=22)
Percent survival
Recurrence-Free Survival (Months)
d
cg13445177
Low-methyl. score (n=58)
moderate-methyl. score (n=58)
High-methyl. score (n=22)
Percent survival
Recurrence-Free Survival (Months)
e
ICGC - cg13249591
Low methyl. score (n=129)
moderate methyl. score (n=128)
High methyl. score (n=48)
Percent survival
Overall Survival (Months)
f
ICGC - cg13445177
Low methyl. score (n=129)
moderate methyl. score (n=128)
High methyl. score (n=48)
Percent survival
Overall Survival (Months)

7.7 S100A10 expression is regulated by methylation in PDAC cell lines

To validate that *S100A10* is regulated by DNA methylation *in cellulo*, we first compared *S100A10* mRNA expression in the CCLE cell lines. A negative correlation between *S100A10* mRNA (RNA Seq V2 RSEM) and DNA methylation was observed across all cell lines (Pearson correlation coefficient = -0.581) (figure 60a) including pancreatic cell lines (supplemental figure 20a). We then compared *S100A10* mRNA and protein levels and promoter methylation in three cell lines that are representative of expression/methylation spectrum (Panc 10.05, Panc-1 and AsPC-1). Panc10.05 cells had the lowest *S100A10* mRNA (Fig. 60b) and protein expression (Fig. 60c) followed by Panc-1 and AsPC-1 cells. To examine whether the *S100A10* promoter region was differentially-methylated in the three-cell line panel, we performed bisulfite conversion followed by pyrosequencing of a 377-nucleotide promoter region containing 24 CpG sites (Fig. 60d) (supplemental figure 20b). Consistent with the mRNA levels, global DNA methylation of that region was the highest in Panc 10.05 cells followed by Panc-1 and AsPC-1 cells (figure 60e). Notably, AsPC-1 cells had considerably higher mRNA and protein levels and significantly low DNA methylation. To address effect of DNA demethylation on S100A10 expression, all three cell lines were treated with the DNA de-methylating agent decitabine. *S100A10* mRNA and protein levels were dramatically upregulated in Panc 10.05 (figure. 61a, 61d) and to a lesser extent in Panc-1 cells (figure 60b, 60e). In contrast, no increase was observed in the AsPC-1 cell line (figure 61c, 61f). Despite the differential response in *S100A10* mRNA, the overall methylation of the promoter region was further decreased in all three cell lines in response to decitabine (Fig. 61g, 61h, 61i). Such decrease was also seen across the individual CpG sites examined (Fig. 5j, 5k, 5l). Notably, the cg13445177

and cg13249591 probes mapped CpG sites 6 and 7 and sites 9 and 10 respectively. Only CpG-9 was differentially de-methylated across all three cell lines indicating that this site (in addition to others) was likely responsible in sustaining low *S100A10* mRNA in PDAC patients. Collectively, these results indicated that *S100A10* expression is regulated through hypomethylation at specific CpG sites.

Figure 60. S100A10 mRNA and protein expression is regulated by methylation in PDAC cell lines. S100A10 mRNA and protein expression negatively correlated with promoter methylation in PDAC cell lines. The relationship between S100A10 methylation and mRNA expression in 831 CCLE cell lines. mRNA expression (RNA Seq V2 RSEM) and methylation (RRBS β values) were extracted from the broad institute CCLE portal (https://portals.broadinstitute.org/ccle). S100A10 mRNA (RT-qPCR) (B) and protein expression (C) in three PDAC representative cell lines: Panc 10.05, Panc-1 and AsPC-1. (D) S100A10 promoter construct for bisulfite and pyrosequencing covering 24 CpG dinucleotides. (E) Global methylation of the 24 CpGs in the S100A10 promoter. The graph represents the averages of percentages of all 24 sites in each cell line. Significance was determined using one-way ANOVA. Data are represented as mean ± SD.

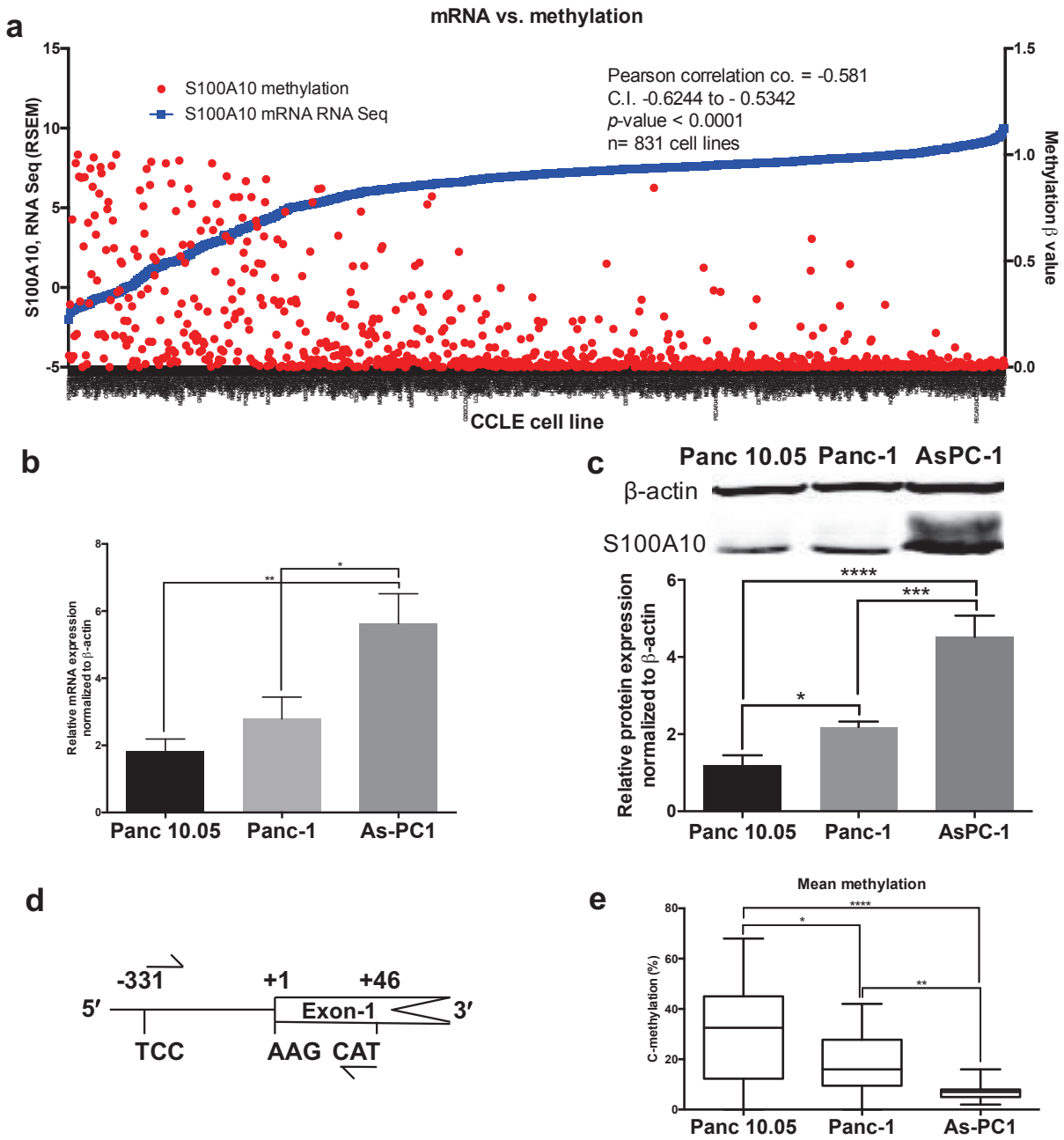
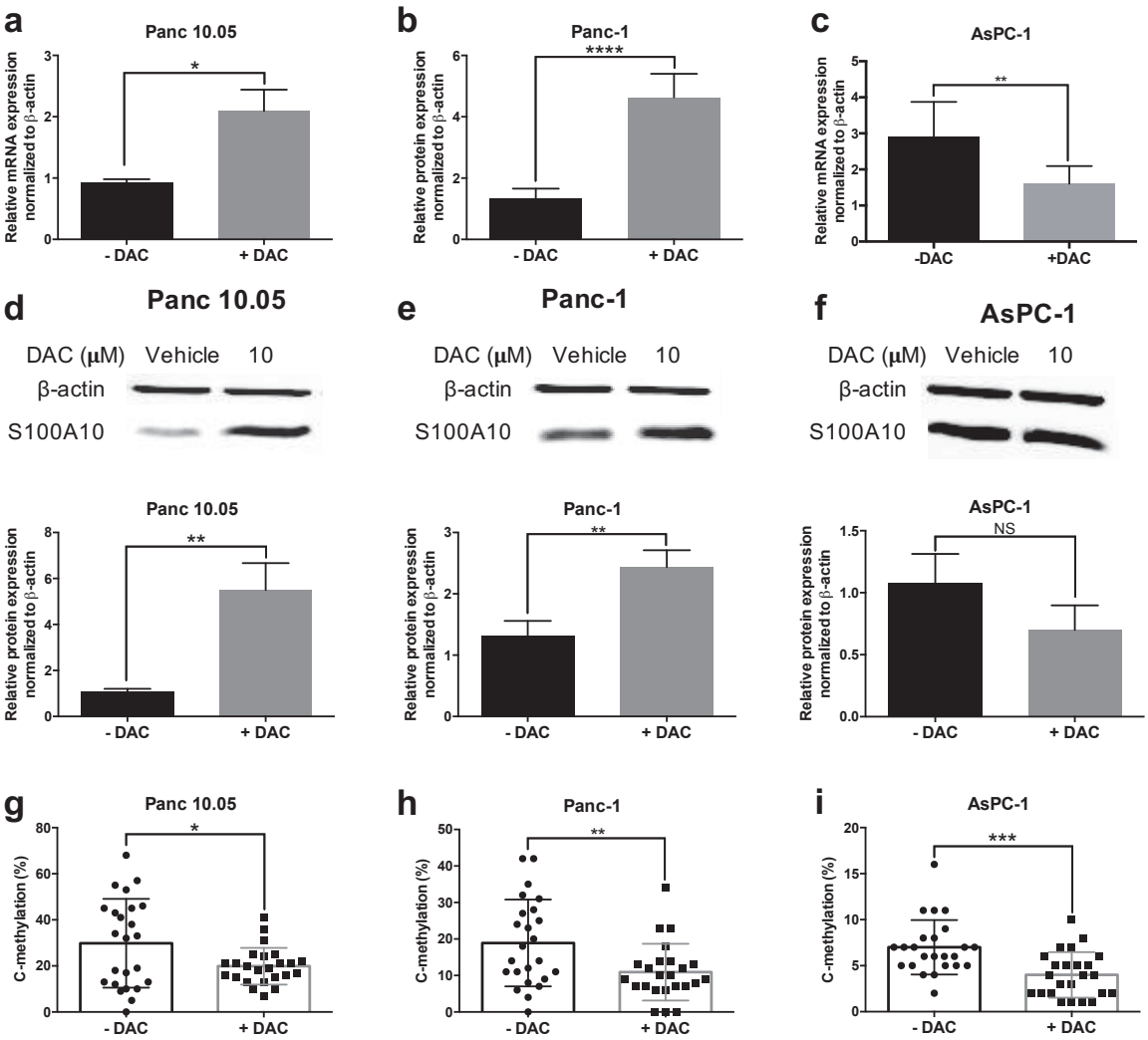
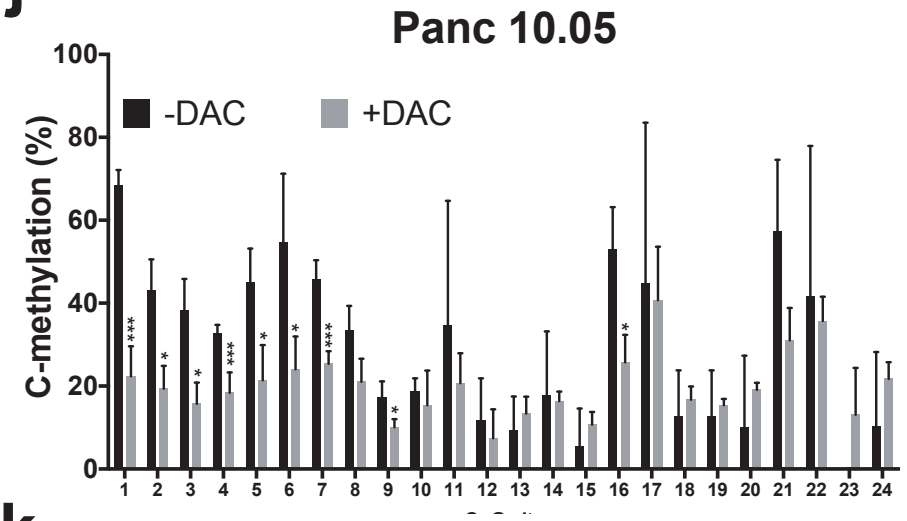

a
mRNA vs. methylation
S100A10 methylation
S100A10 mRNA RNA Seq
Pearson correlation co. = -0.581
C.I. -0.6244 to - 0.5342
p-value < 0.0001
n= 831 cell lines
S100A10, RNA Seq (RSEM)
Methylation β value
CCLE cell line
b
Relative mRNA expression normalized to β-actin
Panc 10.05
Panc-1
As-PC1
c
Panc 10.05 Panc-1 AsPC-1
β-actin
S100A10
Relative protein expression normalized to β-actin
Panc 10.05
Panc-1
AsPC-1
d
-331
+1
+46
5′
Exon-1
3′
TCC
AAG CAT
e
Mean methylation
C-methylation (%)
Panc 10.05
Panc-1
As-PC1

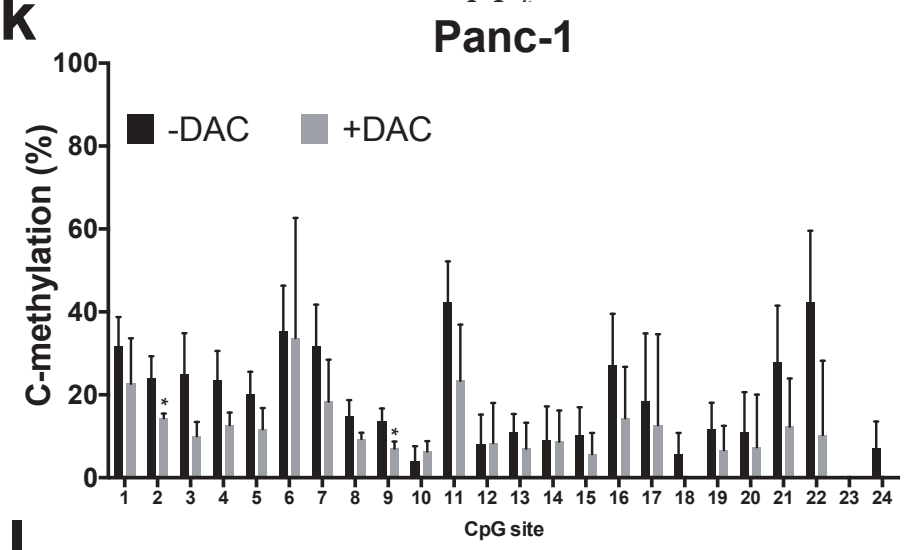
Figure 61. S100A10 mRNA expression is regulated by differential CpG site methylation. S100A10 mRNA (a, b, c) and protein (d, e, f) changes in Panc 10.05 (a, d), Panc-1 (b, e) and AsPC-1 (c, f) in response to 10μM decitabine (DAC) for 72 hours. Global and CpG-specific methylation of the 24 CpGs in the S100A10 promoter in Panc 10.05 (g, j), Panc-1 (H, K) and AsPC-1 (i, l). Graphs g-i represent the averages of percentages of all 24 sites in each cell line. Graphs j-l represent the percentage methylated of cytosines of a specific CpG site within each sample. Significance was determined using unpaired t-tests. Data are represented as mean ± SD.


a
Panc 10.05
Relative mRNA expression normalized to β-actin
*
- DAC
+ DAC
b
Panc-1
Relative protein expression normalized to β-actin
****
- DAC
+ DAC
c
AsPC-1
Relative mRNA expression normalized to β-actin
**
-DAC
+DAC
d
Panc 10.05
DAC (μM) Vehicle 10
β-actin
S100A10
e
Panc-1
DAC (μM) Vehicle 10
β-actin
S100A10
f
AsPC-1
DAC (μM) Vehicle 10
β-actin
S100A10
Panc 10.05
Relative protein expression normalized to β-actin
**
- DAC
+ DAC
Panc-1
Relative protein expression normalized to β-actin
**
- DAC
+ DAC
AsPC-1
Relative protein expression normalized to β-actin
NS
- DAC
+ DAC
g
Panc 10.05
C-methylation (%)
*
- DAC
+ DAC
h
Panc-1
C-methylation (%)
**
- DAC
+ DAC
i
AsPC-1
C-methylation (%)
***
- DAC
+ DAC

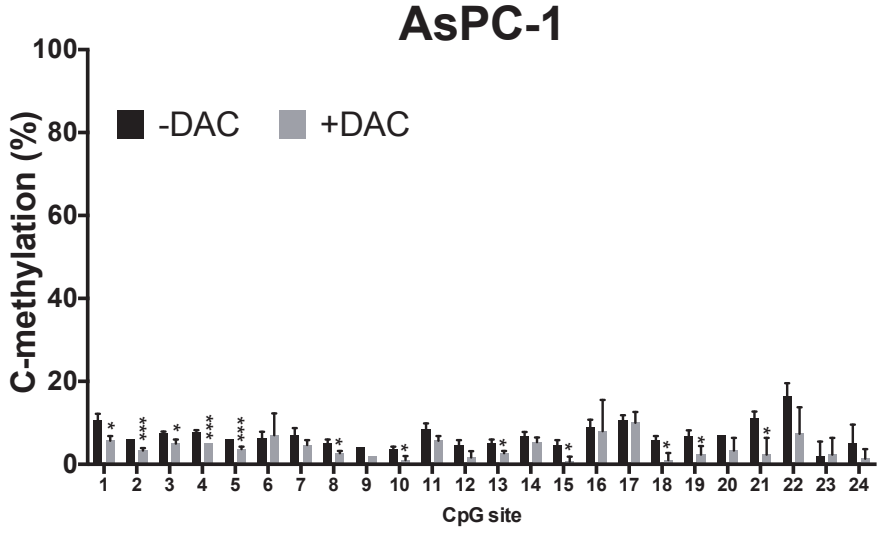
j


Panc 10.05
-DAC
+DAC
C-methylation (%)
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

k


Panc-1
-DAC
+DAC
C-methylation (%)
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
CpG site

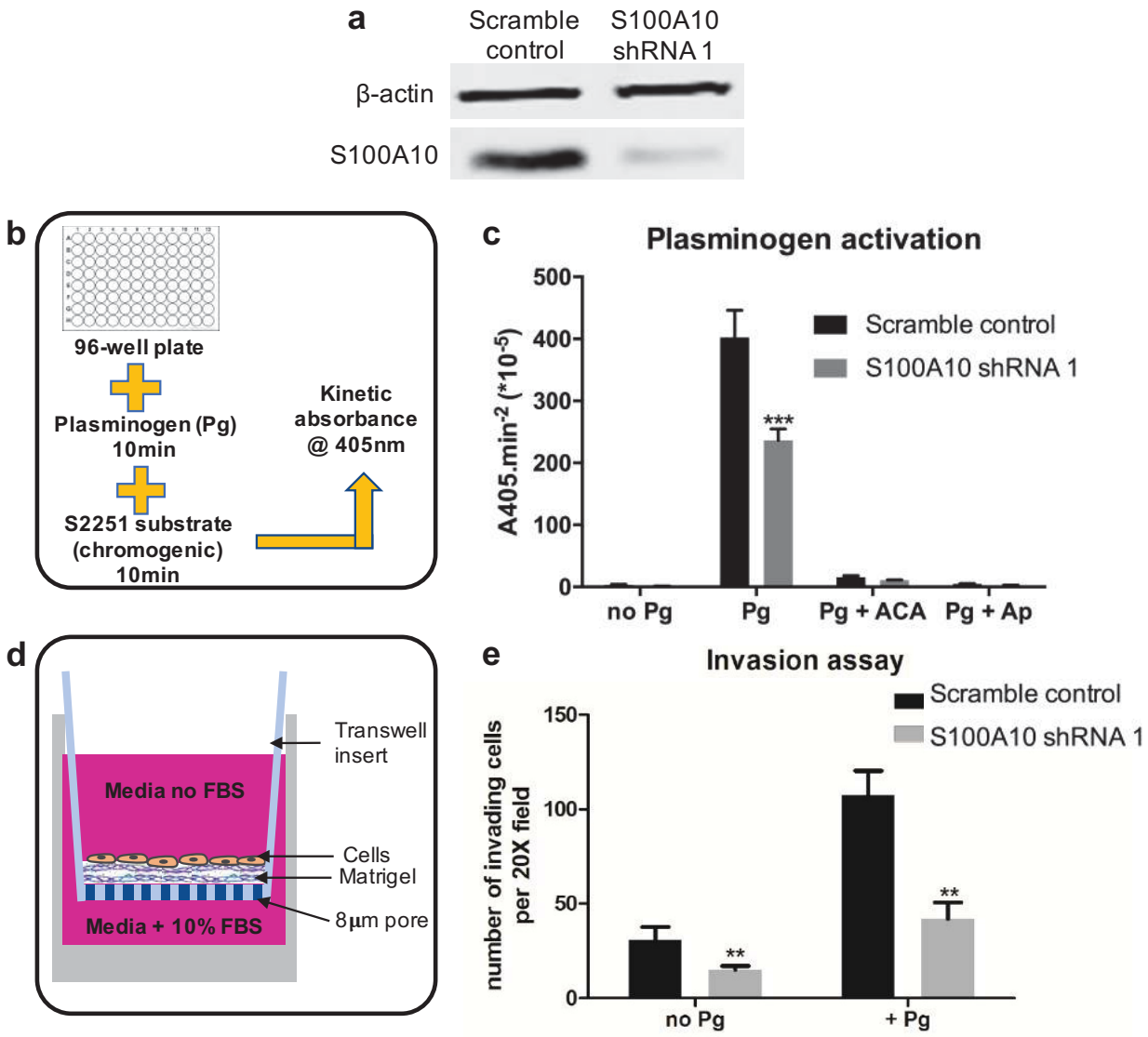
l


AsPC-1
-DAC
+DAC
C-methylation (%)
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
CpG site

7.8 S100A10 acts as a plasminogen receptor at the surface of pancreatic cancer cells and contributes to cancer cell invasion.

Considering the predictive role of *S100A10* mRNA expression and methylation status as classifiers of patient outcome and its epigenetic regulation, the cellular mechanism by which S100A10 protein, as a plasminogen receptor, may contribute to the underlying pathology of PDAC remains elusive. Our laboratory has extensively studied the functional plasminogen-activating aspect of S100A10 [234][235][238][259][237][233][241]. However, whether S100A10, plays a role at the surface of pancreatic cancer cells has never been addressed. The depletion of S100A10 using short-hairpin sequences (shRNA) (figure 61a) in Panc-1 cells resulted in a 50% reduction of plasminogen activation (figure 61b, 61c). ε-aminocaproic acid (ACA) is a lysine analog that prevents plasminogen interaction with the carboxyl-terminal lysine of plasminogen receptors and hence is a well-established inhibitor of plasminogen activation. The dramatic effect of ACA on plasminogen activation indicates that plasminogen activation is primarily driven by plasminogen receptors of which S100A10 accounts for 50% of that activation at the surface of Panc-1 cells. Aprotinin (Ap) is a serine protease pan-inhibitor, which quenches the generated plasmin confirming the ability of these cells to generate plasmin (figure 61c). Subsequent assessment of cancer cell invasion using the well-established Boyden chamber method (figure 61d) revealed that S100A10 depletion reduced the ability of Panc-1 cells to pass through the ECM-dense matrigel even in the presence of exogenous plasminogen (+Pg) compared to scramble control cells (figure 61e). These findings inferred the role of S100A10 as a recognized plasminogen receptor and a mediator of plasminogen-dependent invasiveness of pancreatic cancer cells.

Figure 62. S100A10 depletion in Panc-1 cells reduces plasminogen activation and cellular invasiveness *in vitro*. (a) Western blot analysis of scramble control and S100A10-depleted (S100A10 shRNA1) Panc-1 cells. (b) Schematic representation of the plasminogen assay; cells were incubated with 0.5μM plasminogen and plasmin activity was measured as the absorbance of the chromogenic plasmin substrate (S2251) at a wavelength of 405nm. (c) $5x10^3$ cells of scramble control and S100A10 shRNA1 Panc-1 cells were seeded into 96-well plates. Plasminogen activation (per $1x10^5$ cells) was then calculated under the following conditions: no plasminogen, with plasminogen, with the lysine analog ε-aminocaproic acid (ACA, 100mM) and the serine protease Aprotinin (Ap 2.2μM). (d) Schematic representation of the matrigel boyden chamber used for the invasion assay. The assay assesses the ability of cells to invade through a Matrigel barrier (substitute for ECM) in response to a chemoattractant (10% FBS). (e) Invasion assay of scramble control and S100A10 shRNA 1 Panc-1 cells in the presence/absence of Pg. The results are represented as the number of invading cells per one field of view at 20X magnification.

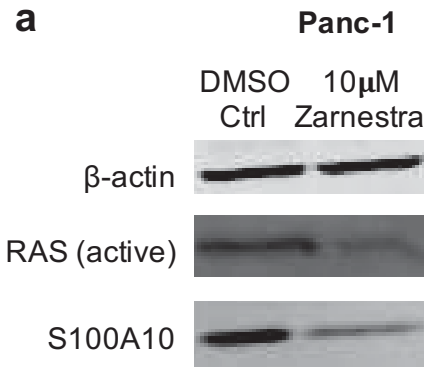

a
Scramble control
S100A10 shRNA 1
β-actin
S100A10
b
96-well plate
Plasminogen (Pg) 10min
S2251 substrate (chromogenic) 10min
Kinetic absorbance @ 405nm
c
Plasminogen activation
A405.min-2 (*10-5)
500
400
300
200
100
0
Scramble control
S100A10 shRNA 1
***
no Pg
Pg
Pg + ACA
Pg + Ap
d
Transwell insert
Media no FBS
Cells
Matrigel
8µm pore
Media + 10% FBS
e
Invasion assay
Scramble control
S100A10 shRNA 1
number of invading cells per 20X field
150
100
50
0
**
**
no Pg
+ Pg

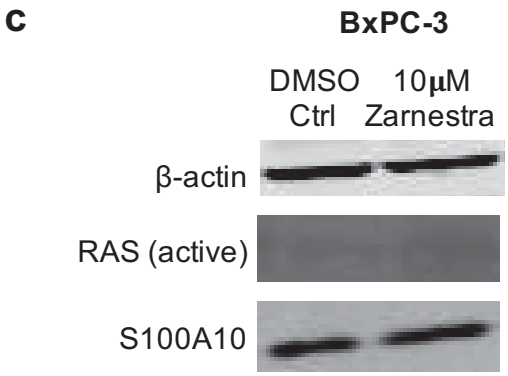
7.9 S100A10 expression is regulated by oncogenic *KRAS*^{G12D} in pancreatic cancer cells.

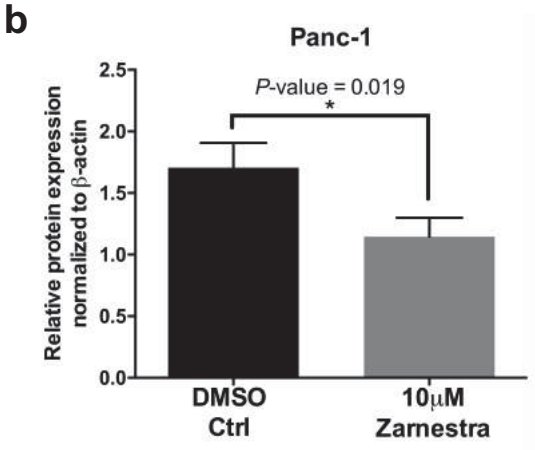
KRAS mutations are ubiquitous in PDAC with over 95% penetrance [461]. The mutation is the earliest genetic alteration and is found as early as low-grade PanIN-A lesions [479]. We have previously demonstrated that RAS proteins, particularly HRAS, upregulate S100A10 expression in HEK293 cells [233]. Considering the direct involvement of oncogenic KRAS activity in PDAC pathobiology and the role of S100A10 in cellular proteolytic activity and invasiveness, we examined whether S100A10 is regulated via KRAS signaling. To address this issue, we utilized three cell lines representing three forms of *KRAS* expression, Bx-PC3 (Wild type-*KRAS*), Panc-1 (mutant *KRAS*, *KRAS*^{G12D}) and iKRAS (inducible *KRAS*^{G12D}). Treating BxPC-3 and Panc-1 cells with the farnesyltransferase inhibitor tipifarnib (Zarnestra) decreased S100A10 protein expression in the mutant-*KRAS* cell line Panc-1 (figure 62a, 62b) but not in the wild type-*KRAS* cell line BxPC3 (figure 62c, 63d). Only Panc-1 cells responded to the inhibition which is consistent with the fact that active RAS (RAS-GTP) was only expressed in Panc-1 and not in Bx-PC3 cells. Similarly, ectopic expression of oncogenic *KRAS*^{G12D} in *KRAS*-wildtype Bx-PC3 (figure 62e) and HEK293 (figure 62f) cells also upregulated S100A10 protein expression. The iKRAS mouse cell line possesses a doxycycline-inducible *KRAS*^{G12D} construct (figure 63a). Addition of 1 µg/ml of doxycycline induced *KRAS* expression and a concomitant two-fold increase in S100A10 protein expression which was inhibited by Zarnestra (figure 63b, 63c). *KRAS* induction dramatically increased plasminogen activation which was concomitant with S100A10 upregulation while Zarnestra treatment abolished this activation (figure 63d). Considering the regulation of

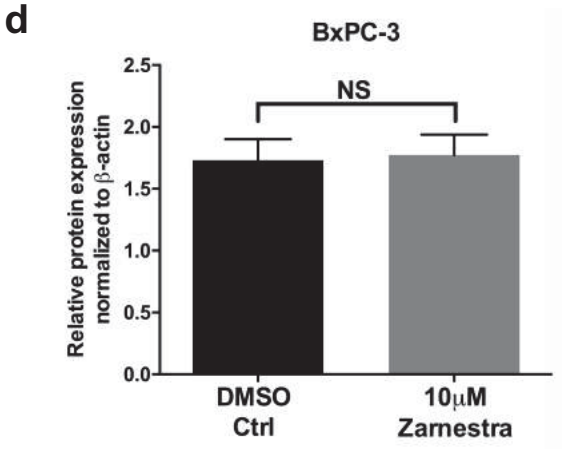
S100A10 by methylation, we treated non-induced and induced cells with decitabine. Results revealed a potentially independent effect of KRAS induction and promoter demethylation since the increase in S100A10 was higher in the presence of doxycycline and decitabine compared to either alone (figure 63e). These results indicated that oncogenic *KRAS* regulates S100A10 which in turn drives the activation of plasminogen.

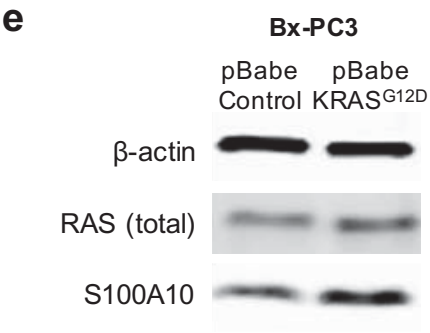
Figure 63. S100A10 expression is regulated by oncogenic KRAS^{G12D} in pancreatic cancer cells. Western blot analysis of S100A10, active RAS, and β-actin in Panc-1 (a) and BxPC-3 (c) treated with 10**μ**M of the farnesyltransferase inhibitor Zarnestra for 48 hours. A Raf-pulldown was performed to measure RAS activity. Quantification of S100A10 protein expression normalized to β-actin in DMSO- and Zarnestra-treated Panc-1 (b) and BxPC-3 (d). Western blot analysis of S100A10 protein in BxPC-3 (e) and HEK293 (f) cells which were transfected with the pBabe Control and pBabe KRAS^{G12D} vectors.


a
Panc-1
DMSO Ctrl
10μM Zarnestra
β-actin
RAS (active)
S100A10


c
BxPC-3
DMSO Ctrl
10μM Zarnestra
β-actin
RAS (active)
S100A10


b
Panc-1
P-value = 0.019
*
Relative protein expression normalized to β-actin
2.5
2.0
1.5
1.0
0.5
0.0
DMSO Ctrl
10μM Zarnestra


d
BxPC-3
NS
Relative protein expression normalized to β-actin
2.5
2.0
1.5
1.0
0.5
0.0
DMSO Ctrl
10μM Zarnestra


e
Bx-PC3
pBabe Control
pBabe KRASG12D
β-actin
RAS (total)
S100A10

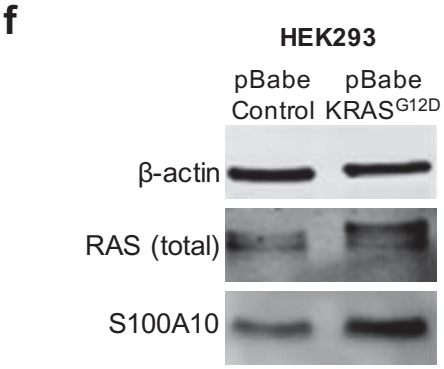
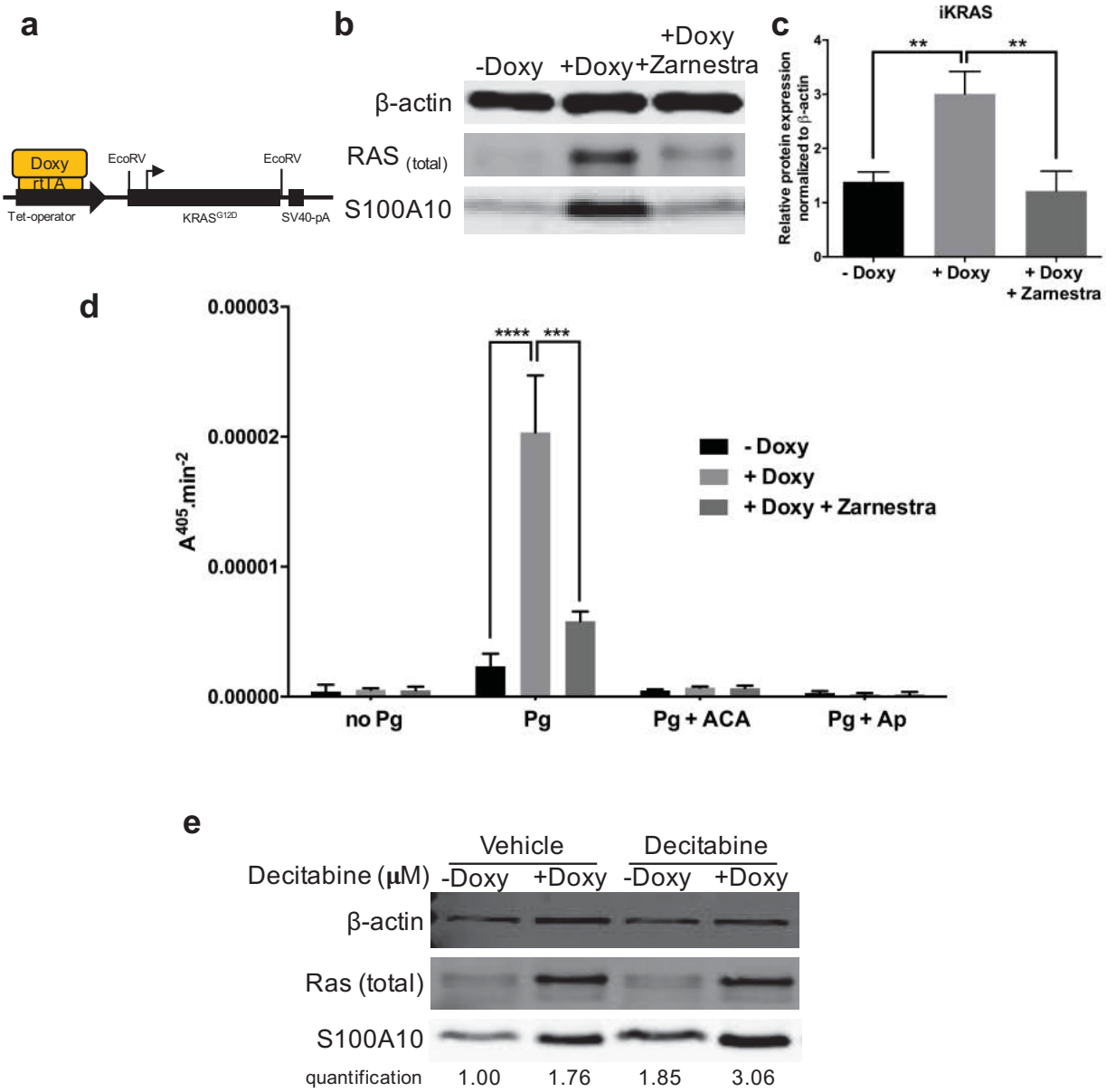

f
HEK293
pBabe Control
pBabe KRASG12D
β-actin
RAS (total)
S100A10

Figure 64. Inducible expression of KRAS^{G12D} upregulates S100A10 protein expression and plasminogen activation. (a) Genomic construct setup of the mouse iKRAS pancreatic cancer cells. rtTA is a reverse tetracycline trans-activator and is required for doxycycline-inducible expression of KRAS^{G12D}. Western blot analysis (b) and quantification (c) of S100A10 protein in iKRAS cells in the absence (-Doxy) or presence (+Doxy) of 1µg/ml doxycycline and Zarnestra (10µM) for 4 days. (d) Plasminogen activation assay of IKRAS cells treated with Doxycycline and Zarnestra). (e) Western blot analysis of iKRAS cells treated with Doxycycline in the presence/absence of 10µM decitabine for 72 hours.

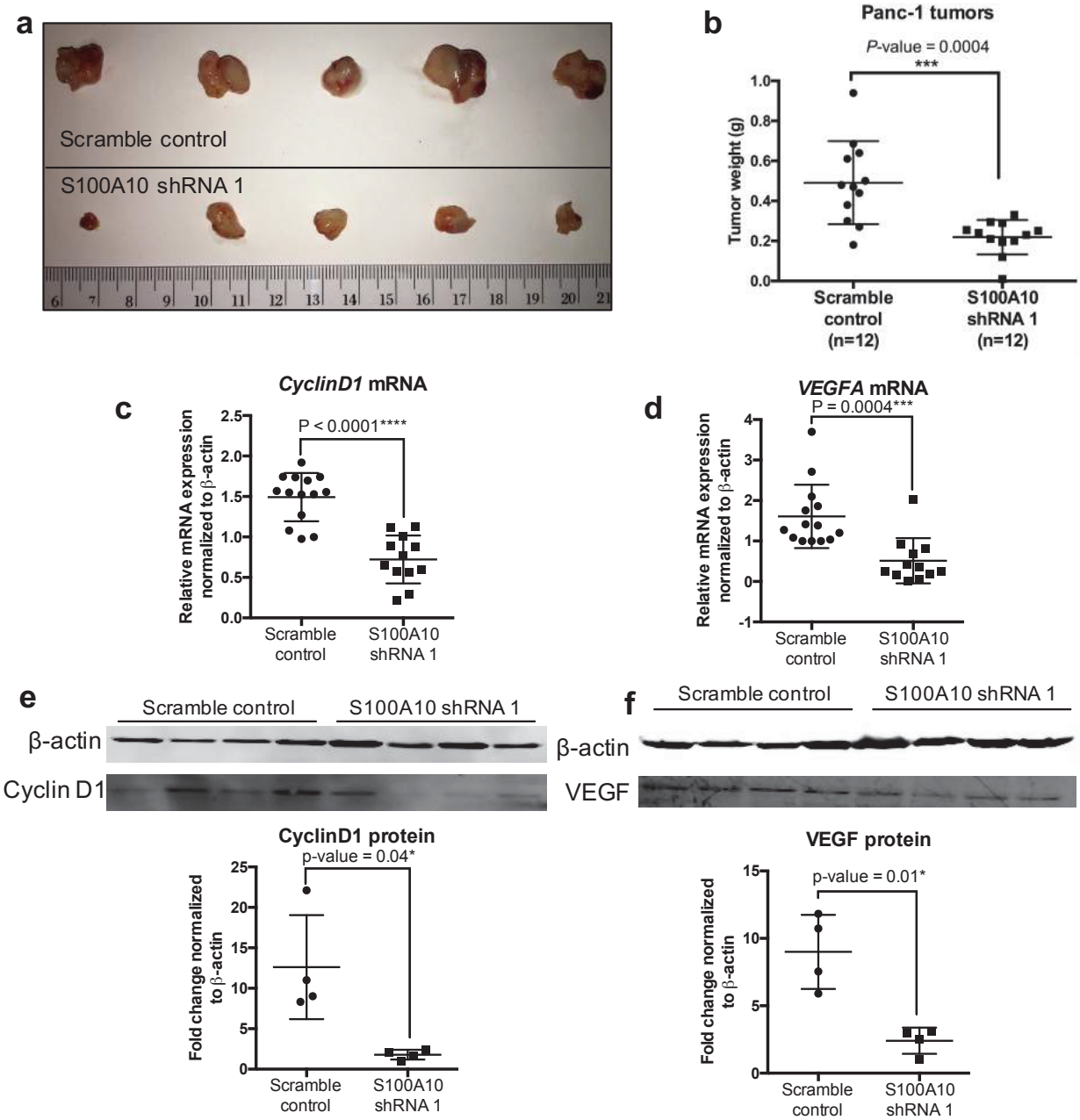

a
Doxy
rtTA
EcoRV
EcoRV
Tet-operator
KRAS^{G12D}
SV40-pA
b
-Doxy
+Doxy
+Doxy
+Zarnestra
β-actin
RAS (total)
S100A10
c
iKRAS
Relative protein expression normalized to β-actin
**
**
- Doxy
+ Doxy
+ Doxy
+ Zarnestra
d
$A^{405}.min^{-2}$
0.00003
0.00002
0.00001
0.00000
****
***
- Doxy
+ Doxy
+ Doxy + Zarnestra
no Pg
Pg
Pg + ACA
Pg + Ap

e
Vehicle
Decitabine
Decitabine (μM)
-Doxy
+Doxy
-Doxy
+Doxy
β-actin
Ras (total)
S100A10
quantification
1.00
1.76
1.85
3.06

7.10 S100A10 is important for growth of pancreatic tumors.

To address whether S100A10 is implicated in *in vivo* PDAC tumorigenesis, we utilized a well-established intra-peritoneal model of PDAC. It has been demonstrated by Schwarz *et al*. that the intraperitoneal injection of Panc-1 cells into NOD/SCID (immune-deficient) mice results in spontaneous homing of the Panc-1 cells to the pancreas. This quasi-orthotopic tumor development model shares many characteristics with human PDAC [648]. After 12 weeks post intraperitoneal injection, juxta-pancreatic tumors were extracted and weighed. Results showed that tumors formed by *S100A10*-depleted Panc-1 cells (0.4913g, C.I. 0.3595g-0.6230g, n = 12) were 2.24-fold smaller than tumors formed by scramble control cells (0.2188g, C.I. 0.1644g-0.2731g, n = 12) (figure 64a, 64b). In an attempt to understand the differences in tumor size, we examined the expression of several genes involved in apoptosis (*BAD*, *BAX* and *PUMA*), cell proliferation (*CCND1*), metastasis (*MMP9*, *CDH1*, *CDH2* and *VIM*) and angiogenesis (*VEGF*) using RT-qPCR (supplemental figure 21). The results showed that mRNA levels of cyclin D1 (*CCND1*) (0.7219 +/- 0.08553, n=12) and VEGF (0.5118 +/- 0.1614, n=12) were significantly lower in S100A10-shRNA 1 tumors compared to Cyclin D1 (1.492 +/- 0.07961, n=14) and VEGF (1.608 +/- 0.2094, n=14) in scramble control tumors (figures 64c and 64d respectively). The downregulation of cyclin D1 and VEGF was also confirmed at the protein level (figure 64e, 64f respectively).

Figure 65. S100A10 knockdown in Panc-1 cells reduces primary tumor size *in vivo*. $5x10^6$ Panc-1 cells scramble control and S100A10 shRNA 1 Panc-1 cells were injected intra-peritoneally into 24 NOD/SCID mice (12 mice each group). (a) Representative image of extracted tumors from the scramble control and S100A10 shRNA groups (5 mice each). (b) 12 tumors from each group were collected and their weights were compared. RT-qPCR (c, d) and western blots (e, f) of Cyclin D1 (c, e) and VEGF (d, f).


a
Scramble control
S100A10 shRNA 1
b
Panc-1 tumors
P-value = 0.0004
***
Tumor weight (g)
Scramble control (n=12)
S100A10 shRNA 1 (n=12)
c
CyclinD1 mRNA
P < 0.0001****
Relative mRNA expression normalized to β-actin
Scramble control
S100A10 shRNA 1
d
VEGFA mRNA
P = 0.0004***
Relative mRNA expression normalized to β-actin
Scramble control
S100A10 shRNA 1
e
Scramble control
S100A10 shRNA 1
β-actin
Cyclin D1
CyclinD1 protein
p-value = 0.04*
Fold change normalized to β-actin
Scramble control
S100A10 shRNA 1
f
Scramble control
S100A10 shRNA 1
β-actin
VEGF
VEGF protein
p-value = 0.01*
Fold change normalized to β-actin
Scramble control
S100A10 shRNA 1

CHAPTER 8: DISCUSSION of chapter 7

8.1 Discussion

Cancer advancement into metastasis is increasingly being attributed to aberrant expression of surface proteins that drive cancer invasion [649]. These proteins are typically overexpressed by tumors and offer a unique opportunity for marker identification and potential therapeutic targeting. During the early days of DNA microarrays, Iacobuzio-Donahue *et al*. identified the gene encoding the plasminogen receptor *S100A10* as one of the top upregulated genes in pancreatic tumors and cell lines compared to their normal counterparts [650]. Many later studies aimed to further analyze differential gene expression using DNA microarrays and more recently RNA-Seq [546][547][544][540][650][543][651]. We analyzed these studies and demonstrated that *S100A10* mRNA is highly expressed in pancreatic tumors and cell lines (figure 53) and is upregulated in virtually all pancreatic tumor tissues compared to matched and unmatched normal tissues (figure 54, supplemental figure 11). The question whether S100A10 protein was also upregulated was first addressed by a study by Sitek *et al*. which utilized mass spectrometry to identify 31 proteins (includes S100A10) that were overexpressed in pancreatic tumors [652]. We herein performed an extensive automated quantification method of stained tissue microarrays (TMAs) from 88 PDAC patients. The expression of S100A10 was found to be markedly low in pancreatic non-ductal stroma and normal tissue with no significant difference whether the normal ducts or non-ductal stroma were adjacent to PanINs or PDAC. There was however a significant but modest increase in expression in PanINs compared to normal ducts which was then exacerbated when these tumors

developed into PDAC (figure 55). This presents the possibility that S100A10 upregulation by pancreatic tumors is a late event that appears to be unique to PDAC.

In addition to assessing S100A10 expression in pancreatic tissues, we addressed the novel predictive value of S100A10 in PDAC. *S100A10* mRNA expression and methylation status were found to be predictive of long-term overall survival and recurrence-free survival in multiple patient cohorts (TCGA, ICGC, Moffit *et al*. cohort and Chen *et al*. cohort). We have developed a reliable ternary classification method through which we identified a low risk group of patients with very low *S100A10* mRNA levels or high *S100A10* methylation score. These patients had significantly longer survival and a lower probability of their cancers recurring. These results delineated, for the first time, the predictive role of S100A10 in PDAC. These finding are supported by other studies that addressed the predictive potential of S100A10 in various cancer models. Shang *et al*. revealed a correlation between positive S100A10 protein expression and poor tumor differentiation, disease stage and poor overall survival in colorectal cancer patients [653]. Li *et al*. demonstrated that, although S100A10 expression did not correlate with long term survival in gastric cancer patients, it did however correlate with lymph node positivity [654] which is consistent with our multi-model fitting of OS and RFS (supplemental table 19). Domoto *et al*. showed that S100A10 is an independent marker of survival in renal cell carcinoma while showing no correlation to tumor grade or stage of renal cell carcinoma patients [655]. High *S100A10* mRNA and protein expression also predicted poorer overall survival in serous ovarian carcinoma [656]. These studies establish S100A10 as a robust pan-cancer biomarker of patient survivability and tumor progression.

The clinical significance of S100A10 in PDAC patients can be partly explained by its plasminogen-dependent role in *in vitro* cancer cell proteolytic activity and invasiveness. As mentioned, plasminogen receptors are essential for the binding and the subsequent activation of the pro-protease plasminogen into the active protease plasmin [657][658][659][199]. Treatment with the lysine analog ε-aminocaproic acid, which competes with plasminogen for receptor binding, completely abrogated plasminogen activation in Panc-1 cells (figure 62c). Consistent with its well-established role as a receptor for plasminogen, S100A10 depletion reduced plasminogen activation which led to significant decrease in invasion of Panc-1 cells (figure 62e) (figure 66). Noteworthy, the significant reduction in invasion upon S100A10 depletion in the absence of plasminogen (-Pg) could be attributed to the plasminogen present in serum [660]. This highlights the importance of plasminogen receptors, in general, in activating plasminogen in the presence of endogenous levels of plasminogen activators.

Oncogenic *KRAS* is a known driver of PDAC tumorigenesis which is attributed to a constitutively active form unable to hydrolyze GTP [661] (figure 16, figure 17). Studies in the early 1990s demonstrated that KRAS increased levels of total [662] and receptor-bound tPA and uPA [663] delineating the potential implication of the plasminogen activation system in KRAS-mediated oncogenesis. Whether possible aberrant regulation of plasminogen receptors is implicated in PDAC has never been addressed. We demonstrated that S100A10 protein expression was driven by oncogenic $KRAS^{G12D}$ contributing to the enhancement of plasminogen activation in pancreatic cancer cells (figures 63, figure 64, figure 66). This is supported by our recent findings which showed that S100A10 is driven by the RAS family of proteins in HEK293 cells via the RalGDS

signaling arm. S100A10 enhanced Ras-mediated plasminogen activation and was important for plasminogen-dependent Ras-induced invasion of HEK293 cells [664]. Notably, the ACA treatment of iKRAS cells abolished plasminogen activation in induced and non-induced cells. Since ACA blocks the interaction of plasminogen with plasminogen receptors but does not block the direct interaction of plasminogen with uPA or tPA, it is likely that the interaction of plasminogen with plasminogen receptors is the rate-limiting step in plasminogen activation by pancreatic cells. Furthermore, since uPA and tPA alone have a limited capacity to activate plasminogen in absence of a plasminogen receptor in cell-free *in vitro* conditions, it is likely that the oncogenic activation of plasminogen receptors is also the rate-limiting step in plasminogen activation and plasminogen-mediated invasion [195]. In addition, we have previously demonstrated that S100A10 colocalized with uPAR at the cell surface of HT1080 fibrosarcoma [234] and colo222 [235] colorectal cancer cells to drive plasminogen activation. S100A10 is also capable of protecting plasmin from inactivation by α2-antiplasmin [665][666][667]. Collectively, these studies strongly indicate that S100A10 is a central player in facilitating uPA-mediated cleavage of plasminogen in *KRAS*-transformed cancer cells.

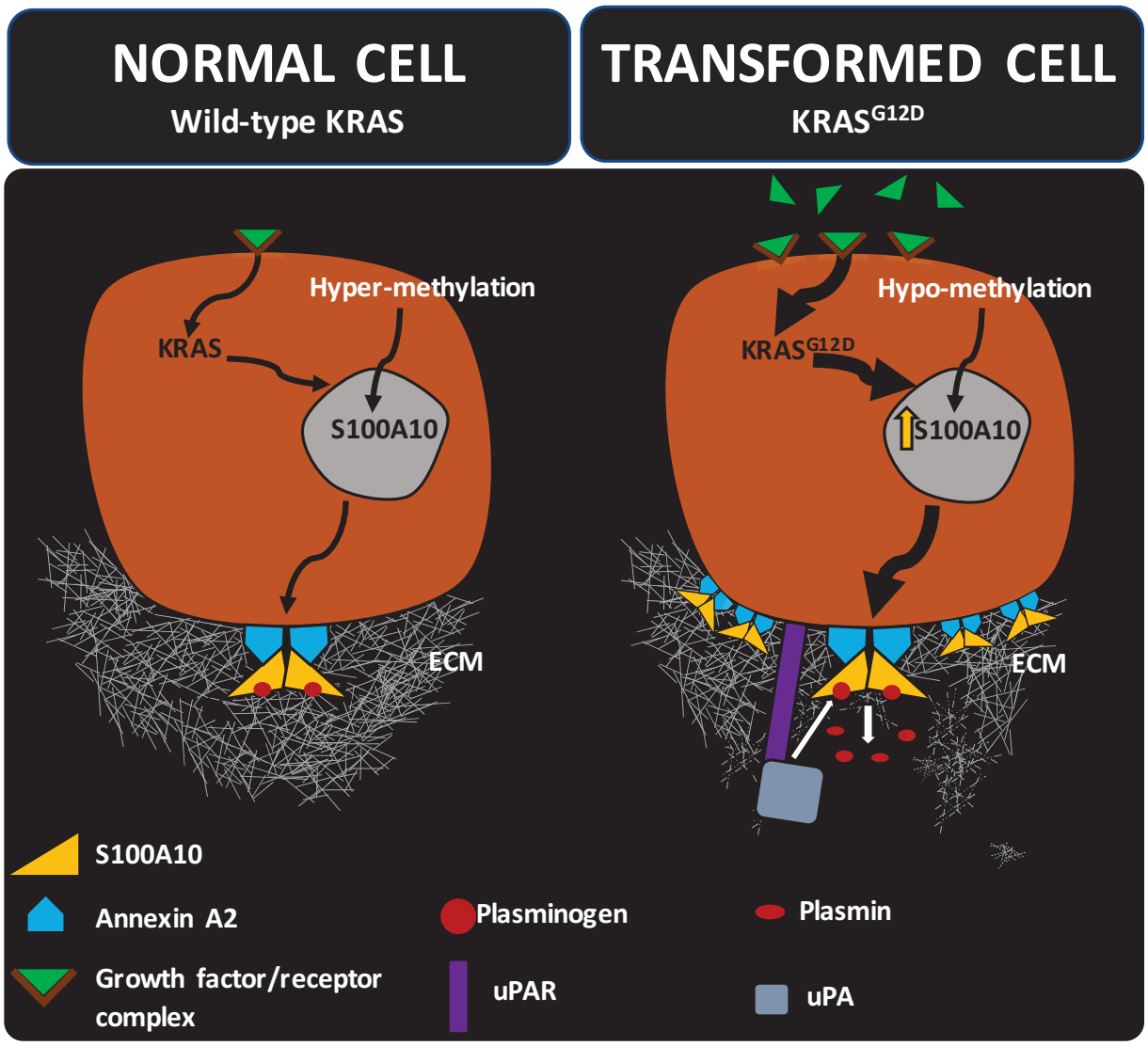
Epigenetic modulation of *S100A10* gene expression adds a layer of complexity to its regulation by *KRAS* (figure 66). We have demonstrated that methylation of the ~400bp promoter region of *S100A10* modulates its expression. Previous reports examining the 1q21 S100 genes revealed that regions upstream of the proximal 400 bp region were differentially methylated. The -600 to -745 region and -400 to -652 region were both found to be hyper-methylated in human pituitary tumors [668] and in medulloblastoma [669]. It should be noted that although the transcription start site of exon 1 of S100A10 appears to

be essential for gene regulation, the 97-amino acid protein constitutes only exons 2 and 3. CpG islands often occur within gene promoters and their methylation is linked to modulation of transcription. A potential CpG island spans the proximal promoter region, the untranslated region of exon 1 and part of intron 1 [670]. This CpG island matches the stringent measures defined by Takai and Jones which necessitates that a region is considered a CpG island if it is longer than 500bp with a G+C content equal to or greater than 55% and observed/expected CpG ratio is 0.65 or higher [671] (supplemental figure 20c). The cg13249591 probe maps to the 5' region of this CpG island while the cg13445177 maps to its south shore. The cg13249591 contains two CpG sites whose methylation status was predictive of PDAC patient OS and RFS and was significantly-demethylated in all three cell lines in response to decitabine.

Considering the role of S100A10 in pancreatic cancer cell invasion *in vitro*, we addressed the role of S100A10 during *in vivo* tumorigenesis. The growth of Panc-1 tumors in immunocompromised NOD/SCID mice was hindered upon depletion of S100A10 compared to the scramble control (figure 64a, 64b). This indicates that S100A10 depletion in these cells is sufficient to reduce tumor growth in the absence of tumor-promoting immune cells. It should be noted that S100A10-depleted Panc-1 cells have similar proliferation rates *in vitro* (supplemental figure 22) which suggests that the *in vivo* effects are likely mediated by the micro-environmental interactions with tumor cells. Our previous findings show that LLC (Lewis Lung Carcinoma) cells yield dramatically smaller tumors in S100A10-null mice compared to wild-type mice and that both tumoral microenvironment and tumor-associated macrophages were essential for sustaining tumor growth [238]. These results indicate that both tumor cell and stromal cell S100A10 are both

implicated in tumorigenesis. It remains unclear whether the reduced tumor growth is due to the plasminogen-dependent function of S100A10 or a novel intracellular function related to apoptosis or proliferation. The latter is supported by evidence showing significant reduction in expression of VEGF and Cyclin D1 (figure 64e, 64f). Shan *et al*. recently demonstrated that miR-590-5P directly binds 3' UTR of S100A10 to inhibit its expression which was associated with downregulation of cyclin D1 in HepG2 hepatocellular carcinoma cells [672]. In addition, Phipps *et al*. presented that S100A10 deficient mice form a poorly vascularized environment for wild-type S100A10 LLC cells based on CD31 staining [238]. It is hence possible that tumor cell VEGF is required for adequate angiogenesis to occur. Collectively, these studies and our findings indicate that S100A10 potentially contributes to tumor cell proliferation via sustenance of cyclin D1 levels and to angiogenesis by maintaining VEGF production to ensure blood vessel development.

Figure 66. Schematic representation of KRAS^{G12D}- and methylation-mediated regulation of S100A10-dependent plasminogen activation. Both oncogenic KRAS and DNA demethylation induced S100A10 upregulation which in turn contributed to increased plasminogen activation and plasminogen-dependent invasion. A heterotetrameric complex is formed of two annexin A2 subunits and 2 subunits of S100A10 (dimer). KRAS is also capable of upregulating uPA and uPAR whose localization is induced by S100A10 binding to plasminogen. The latter is activated into plasmin which cleaves extracellular matrix (ECM) proteins and destabilizes its structure allowing pancreatic cancer cell advancement.


NORMAL CELL
Wild-type KRAS
TRANSFORMED CELL
KRAS^{G12D}
Hyper-methylation
KRAS
S100A10
ECM
Hypo-methylation
KRAS^{G12D}
S100A10
ECM
S100A10
Annexin A2
Growth factor/receptor complex
Plasminogen
uPAR
Plasmin
uPA

8.2 Study limitations and future directions

8.2.1 S100A10 as a PDAC biomarker and its level-of-evidence

The majority of PDAC patients (92%) are diagnosed with locally advanced or metastatic disease [464]. At that point, surgery is rarely curable and often not recommended to avoid post-operative complications. Patients eligible for surgical resection will receive adjuvant chemotherapy with or without radiation which results in a 15-30% chance of surviving to five years [466][673]. The development of clinical tools for early detection and risk prediction is key for improving patient outcome and quality of life. Biomarker discovery represents a direct translational path to clinical applications.

S100A10 expression has been linked to prediction of patient outcome in PDAC (figure 56, figure 57), non-small cell lung adenocarcinoma (figure 48) [674], renal cell carcinoma [655], colorectal cancer [653] and ovarian cancer [637], [656]. The above studies and the proposed dissertation are retrospective studies that examined the prognostic value of S100A10 in archived samples. The next logical step is further validate S100A10 mRNA levels in other retrospective cohorts and establish a method of measurement (e.g. RT-qPCR-based test on biopsies) and the adequate cut-offs for identifying the low, intermediate and high risk groups. Once established, the proposed test must be assessed in prospective samples in a randomized clinical trial where pre-established guidelines are in place. This will achieve a level-of-evidence 1 which requires multiple retrospective studies and at least one prospective trial testing the biomarker performance of S100A10. PAI-1 and uPA have achieved level-of-evidence 1 as biomarkers in lymph node negative breast cancer patients [631].

8.2.2 Is S100A10 protein expression predictive in the TCGA PDAC patients?

The promising results of S100A10 mRNA correlation with outcome of PDAC patients in the TCGA provisional dataset have promoted us to examine whether S100A10 protein levels can also predict patient survivability. To answer this question, we performed correlation analyses between S100A10 protein expression (as quantified by ImageJ) and OS of the CDHA cohort.

Since all PDAC regions scored as positive/strong, a new score-based dichotomous approach was needed. As a result, a H-score of 200 was used to distinguish a low positive group (n=30) and a high-positive group (n=58). Kaplan Meier analysis of long-term OS showed no correlation with S100A10 protein expression in PDAC lesions (supplementary figure 23a). A H-score of 100 was used to stratify expression in PanINs (weak/negative vs. strong positive). Similarly, there was no correlation between S100A10 protein expression in PanINs and OS in the CDHA cohort (supplementary figure 23b). No correlations were found between OS and S100A10 protein expression in the PDAC stroma, PanIN stroma, normal adjacent to PDAC or normal adjacent to PanIN (data not shown). However, we also assessed short-term survival of the same cohort based on the above cut-offs. S100A10 protein expression in both PanINs and PDAC lesions correlated with one-year OS but not three- or five-year OS. The chance of being alive after one year after diagnosis was higher in low-positive PDAC lesions (70%) and negative/weak PanINs (73.68%) compared to high-positive PDAC lesions (46.55%) and positive/strong precancerous lesions (55.26%) (figure 6c, 6d) (supplementary figure 23c, 23d). The ability of S100A10 expression to predict three- and five-year OS was modest.

At first glance, these results suggested that S100A10 protein was not predictive of OS. However, the clinical data from this cohort raised some concerns. First, the survival curve of this cohort is very steep which makes it difficult for any biomarker to identify low- and high-risk groups with substantial survival advantage. Second, we also performed univariate survival analysis on the remaining clinical co-variates. Only two covariates showed correlation with OS: margin involvement (HR=1.659, C.I. 1.132 to 2.709, *p*-value=0.0146) and poor differentiation (HR=6.343, C.I. 2.234 to 9.580, *p*-value<0.0001) both of which were not available in the TCGA cohort. Third, lymph node involvement, which was predictive of OS in the TCGA cohort, was not predictive of OS in the CDHA cohort (HR=0.8266, C.I. 0.5149to 1.300, *p*-value=0.3991) (data not shown).

8.2.3 S100A10's role in metastasis

Our *in vivo* experimentation was limited to intra-peritoneal injection of Panc-1 cells as means to measure primary tumor growth. Whether S100A10 plays a role in metastasis is yet to be deciphered. In that context, Scramble control and S100A10 shRNA 1 cells are to be injected into the tail vein of NOD-SCID mice. Liver, lungs and spleen will be collected at 12 weeks post injection, fixed with 10% formalin and embedded in paraffin. The results are expected beyond the time frame available for the completion of the dissertation.

8.2.4 Transgenic PDAC model

Orthotopic mouse models described above have clear disadvantages in recapitulating human PDAC. These disadvantages include the inability to study the impact

of tumor microenvironment and immune surveillance in a context where S100A10 is depleted in either tumor cells and/or stromal cells. For that purpose and to bypass such limitations, we have acquired a spontaneous PDAC model that was dubbed iKRAS. The iKRAS mice are transgenic mice that exhibits pancreas-specific and Doxycycline-inducible expression of $KRAS^{G12D}$ and conditional TP53 null alleles [675]. This model is the gold standard murine model for human pancreatic cancer; it utilizes doxycycline to induce PDAC in mice with high frequency. iKRAS mice have been crossed with S100A10 -/- mice in attempt to derive the desired iKRAS S100A10 -/- mice. These mice will ultimately permit studying the effect of S100A10 depletion on tumor growth, metastasis as well as on immune and stromal cell profiles within the tumor microenvironment.

CHAPTER 9: SUMMARIES

9.1 Chapter 3 summary

1. Various models of epithelial-like and mesenchymal-like cells were established in 2D cultures to study plasminogen activation *in vitro*.
2. S100A10 mRNA and protein expression is regulated by SMAD4-mediated TGFβ1 signaling in A549 cells.
3. S100A10 is a TGFβ1-responsive gene and not an EMT gene.
4. PI3kinase signaling represses S100A10 expression via FOXC2.
5. S100A10 serves as a plasminogen receptor at the surface of A549 cells.
6. Mesenchymal cells downregulate S100A10 surface expression and demonstrate a low capacity to activate plasminogen.
7. S100A10 and uPAR-mediated plasminogen activation is potentially masked by marked PAI-1 upregulation.

9.2 Chapter 5 summary

1. A multi-step strategy was developed to study PA genes in NSCLC
2. Cluster 3 PA genes are predictive of overall survival in adenocarcinoma patients and not squamous cell carcinoma patients.
3. A four-gene signature (*S100A10*, *ANXA2*, *PLAUR* and *PLAU*) is a strong predictor of adenocarcinoma patient overall survival
4. *S100A10*, *ANXA2* and *PLAUR* are predictive of chemotherapeutic response in adenocarcinoma patients

5. S100A10 is upregulated by various chemotherapeutic agents and may contribute to cisplatin resistance.

9.3 Chapter 7 summary

1. S100A10 mRNA is highly expressed in pancreatic tumors and cell lines.
2. S100A10 is highly expressed in pancreatic tumors compared to adjacent non-ductal stroma and normal ducts.
3. S100A10 mRNA expression and copy number are predictive of overall and recurrence-free survival in PDAC patients.
4. S100A10 mRNA and lymph node positivity are linked predictors of overall and recurrence-free survival.
5. S100A10 methylation status is predictive of overall and recurrence-free survival in PDAC patients.
6. S100A10 expression is regulated by methylation at several CpG sites.
7. S100A10 acts as a plasminogen receptor at the surface of pancreatic cancer cells and contributes to cancer cell invasion *in vitro*.
8. S100A10 expression is regulated by oncogenic $KRAS^{G12D}$ in pancreatic cancer cells.
9. S100A10 is important for growth of pancreatic tumors.

CHAPTER 10: Conclusions

10.1 S100A10: one of the best studied plasminogen receptors in cancer.

This work further establishes S100A10 as a plasminogen receptor and a *bona fide* contributor to tumorigenesis. A summary of this work and previous literature indicates that S100A10 responds to various stimuli: oncogene activation (e.g. HRAS, KRAS, PML-RARα), growth factors (e.g. TGFβ1, TGFα [676], FGF-1, EGF[677], BDNF[678]), interferons (e.g. IFN-γ [679]), synthetic compounds (e.g. cisplatin, paclitaxel, dexamethasone), transcription factors (e.g. SMAD4, FOXC2) and other signaling molecules (e.g. thrombin[680], retinoic acid[681]). The diversity of these stimuli renders S100A10 as a highly inducible gene through which it serves both known and potentially novel functions. Functions of S100A10 beyond binding plasminogen are under current investigation in the Waisman laboratory. This will be an arduous task for two reasons. First, the intrinsic plasticity of S100 proteins to bind various interactors (addressed in 1.6.7) renders deciphering a new intracellular function difficult. In fact, attempts to detect these interactors in the cancer models described above were unsuccessful (e.g. serotonin receptor expression in PDAC upon S100A10 depletion, data not shown). Second, the S100A10 promoter contains consensus sequences for *de novo* DNA binding proteins (e.g. AP-1, SP-1, SP-2, ATF and NFkB) [233] delineating both complexity and promiscuity of expression. This is further supported by its relatively ubiquitous expression in most cells and tissues. This work offers new insights into potential intracellular function/s of S100A10 that might involve drug resistance possibly through its contribution to the autophagic response via interaction with ULK1 [264].

10.2 S100A10 mRNA vs. protein.

A strength of this work is the that both mRNA and protein changes in S100A10 were addressed as means to understand the biological and contextual implications of these changes. However, the relationship between said metrics was not directly addressed. This is partly due to the fact that the both mRNA and protein levels were concomitantly altered. Indeed, the dynamics of S100A10 mRNA and protein expression are of complex nature. For instance, S100A10 protein expression is highly dependent on annexin A2 (section 1.6.6.1); hence any alterations in annexin A2 may affect S100A10 protein expression [248]. In contrast, any potential effects on S100A10 protein may not manifest if insufficient amounts of annexinA2 are present in the cell. In addition, the lack of concordance between the predictive value of S100A10 mRNA and protein could be explained by the fact that changes in mRNA expression does not always result in corresponding changes in protein expression. Kosti *et al*. described a modest correlation between mRNA and protein levels in normal pancreatic tissue (spearman correlation factor r=0.360) which was noticeably higher than that seen in the TCGA PDAC cohort (r=0.095) [682].

10.3 Plasminogen activation genes as clinical markers.

The dissertation represents the first attempt to utilize hierarchical clustering of genes involved in protease networks to identify differentially-expressed genes and derive a gene signature using a systematic top-down strategy. This strategy is unique as it incorporated key genes that are potentially involved in proteolytic networks and was not limited to genes with known functions. The reemergence of the *PLAUR-PLAU-ANXA2-S100A10* signature was partly serendipitous since these four genes were found to be highly

co-expressed. However, such association is not surprising because the four proteins are directly involved in the binding and activation of plasminogen. It is plausible that the co-expression of these genes is an evolutionarily conserved process that serves to form a hub where plasminogen is promptly activated.

10.4 Uncoupling S100A10 from EMT.

A major component of this dissertation was the realization that S100A10 was a TGFβ1 responsive gene and not an EMT gene. Such observation expands beyond S100A10 and is highly relevant when it comes to complex processes (e.g. apoptosis, proliferation etc.) that involve various interconnected signaling pathway. It is crucial to discern between the "responsiveness of a gene" and the "requirement of a gene" for a specific pathway. In this case, S100A10 was responsive to TGFβ1 but was not required for TGFβ1 signaling (S100A10 knockdown did not affect EMT, data not shown). In contrast, S100A10 knockdown resulted in increased apoptosis in A549 cells only when treated with the chemotherapeutic agent cisplatin but was also responsive to cisplatin treatment. Here, S100A10 is both a cisplatin-responsive gene and a gene important for cisplatin resistance.

10.5 There is a need to study both total and localized expression of any protein.

Another relevant observation in this dissertation was that of S100A10's expression and localization. The modulation of S100A10 was an example where examining both the total and localized expression is essential to make conclusions on the functionality of a protein and more importantly the impact it has on a particular phenotype (plasminogen activation during EMT in this case). We showed that although S100A10 total levels were

higher in TGFβ1-treated A549 cells, its surface expression was drastically lower than that in vehicle-treated cells. This had major translatable implications on plasminogen activation since S100A10 is a well-established plasminogen receptor. Furthermore, the amount of S100A10 on the cell surface will likely dictate the extent of the loss in plasminogen activation when S100A10 is depleted. This is a realization that is often under-studied when comparing various cell lines.

10.6 Culture methods "matter".

As demonstrated in this work, the culture condition of A549 and BEAS-2B cells greatly affected their E/M statuses (figure 20). The ramification of such observation expands beyond the effect on plasminogen activation into other areas of research where the E/M state of a cell can alter the experimental outcome. More specifically, the presence of serum appears to promote a mesenchymal-like phenotype that can be more resistant to otherwise cytotoxic doses of a chemotherapeutic agent thus promoting drug resistance [683].

10.7 Mesenchymal cells have a limited capacity to activate plasminogen in 2D cultures.

This dissertation utilized a new approach to study the capacity of epithelial and mesenchymal cancer cells to activate plasminogen in 2D cultures. Although different components of the plasminogen activation system (e.g. PAI-1[684], uPAR[685]) were previously shown to be altered under EMT-inducing conditions, the consequential effect on plasminogen activation was never addressed. Here, the role of S100A10, PLAUR and

PAI-1 was studied in terms of driving the activation of plasminogen at the cell surface. In addition, this is the first time where the impact of these proteins on plasminogen activation in epithelial vs. mesenchymal context has been addressed. The novelty of the S100A10 observation promoted further investigation into the downstream signaling pathways by which TGFβ1 and other growth factors regulate its expression. An obvious challenge emerged from the substantial cross-talk between pathways particularly the canonical smad-dependent TGFβ1 signaling and PI3K signaling, both of which are known to affect the epithelial/mesenchymal characteristics of cells.

10.8 Plasminogen activation in 2D *in vitro* cultures and EMT-dependent invasion and metastasis *in vivo*: A bit of a stretch?

The initial working hypothesis was that mesenchymal cells will have enhanced capabilities to activate plasminogen based on 1) the role of EMT in cancer cell metastasis, 2) role of plasmin in invasion and 3) the previous involvement of proteins (e.g. S100A10, PLAUR) in tumor growth and metastasis. Such a linear result would have rendered our conclusions more streamlined and some extrapolations (although not demonstrated) can be made regarding the potential effect of enhanced plasminogen activation on *in vivo* tumorigenesis. To our surprise, mesenchymal cells did not have the postulated effect on plasminogen activation. Hence, any remarks regarding potential *in vivo* implications were not made particularly since 2D culture systems were utilized which do not mimic the 3D microenvironment *in vivo*. The above argument becomes more complex in light of recent publications showing that EMT is not a prerequisite for invasion and metastasis (discussed later in 10.10 and 10.11) [686][687].

10.9 A true MET model

The proposed experiment considered the TGFβ1 treatment of A549 cells as a well-established model of EMT. While that is true, there are no MET models where an epithelial-like cell is induced without the direct inhibition of smad-dependent TGFβ1 signaling. While the FGF/H model activated an epithelial-like phenotype in A549 cells through the activation of PI3K and MAPK/Erk pathways, these pathways are also known to inhibit smad signaling. Shimbori *et al*. demonstrated that FGF-1 reduces phosphorylation of smad2 to attenuate TGFβ1-induced EMT [688]. For that reason, we refrained from using the term MET and resorted to using "epithelial-like" and "mesenchymal-like". It would however be advantageous to develop a model where MET can be induced independently from smad signaling.

10.10 Context dependency and EMT dispensability

The role of EMT in cancer metastasis remains a contentious topic. Various studies have reported that the activation of the development program driven by EMT plays a fundamental role in cancer cell dissemination and metastasis [562]. However, numerous reports have addressed that EMT is dispensable for dissemination and metastasis in spontaneous transgenic mouse models of cancer. Instead, a role of EMT in promoting chemo-resistance *in vivo* emerged in models of breast cancer [686] and pancreatic cancer [687] in two seminal articles by Fischer *et al.* and Zheng *et al.* respectively. The Zheng *et al* study utilized transgenic mice where Snail or Twist1 were genetically deleted. This resulted in a claimed reduction in EMT as evident by the decreased expression of the mesenchymal marker α-SMA. A lineage tracing model which tracks the E/M state of cells

was used *in vivo* to demonstrate that only epithelial cells which never underwent EMT, were responsible for PDAC metastatic growth and repopulation. This effect was not affected by Snail and Twist1 deletion [687]. Similarly, Fischer *et al* utilized a Cre recombinase-based lineage tracing model where the expression of the EMT markers FSP-1 or vimentin will induce RFP expression, indicating the occurrence of EMT. Similar to the Zheng *et al*. study, no EMT was observed since no RFP cells were detected in the lungs indicating that epithelial cells in the primary tumor never underwent EMT prior to metastasizing to the lungs [686].

Although these studies do not necessarily nullify the previous findings accumulated over the past two decades, they do bring into attention that the context or model is potentially more relevant in determining the role of EMT than EMT itself. Neito *et al*. recently addressed the context-dependency issue and suggested that EMT-independent events such as the role of fibroblasts in pulling cancer cells out of the primary tumors contributes to cancer cell dissemination. The fibroblast effect is dependent on both E- and N-cadherin expression [689].

10.11 The backlash to EMT dispensability

The dispensable nature of EMT in the breast and pancreatic cancer models was recently challenged by two concurrent reviews [690][691] which addressed the methodology and conclusions in the previous Zheng *et al* and Fischer *et al* studies. Aiello *et al* questioned the usage of α-SMA as a *bona fide* EMT marker. In fact, Aiello's response demonstrated that α-SMA is not a reliable EMT marker as its expression was rarely induced in the same transgenic PDAC model. In addition, Snail or Twist1 genetic deletion is not

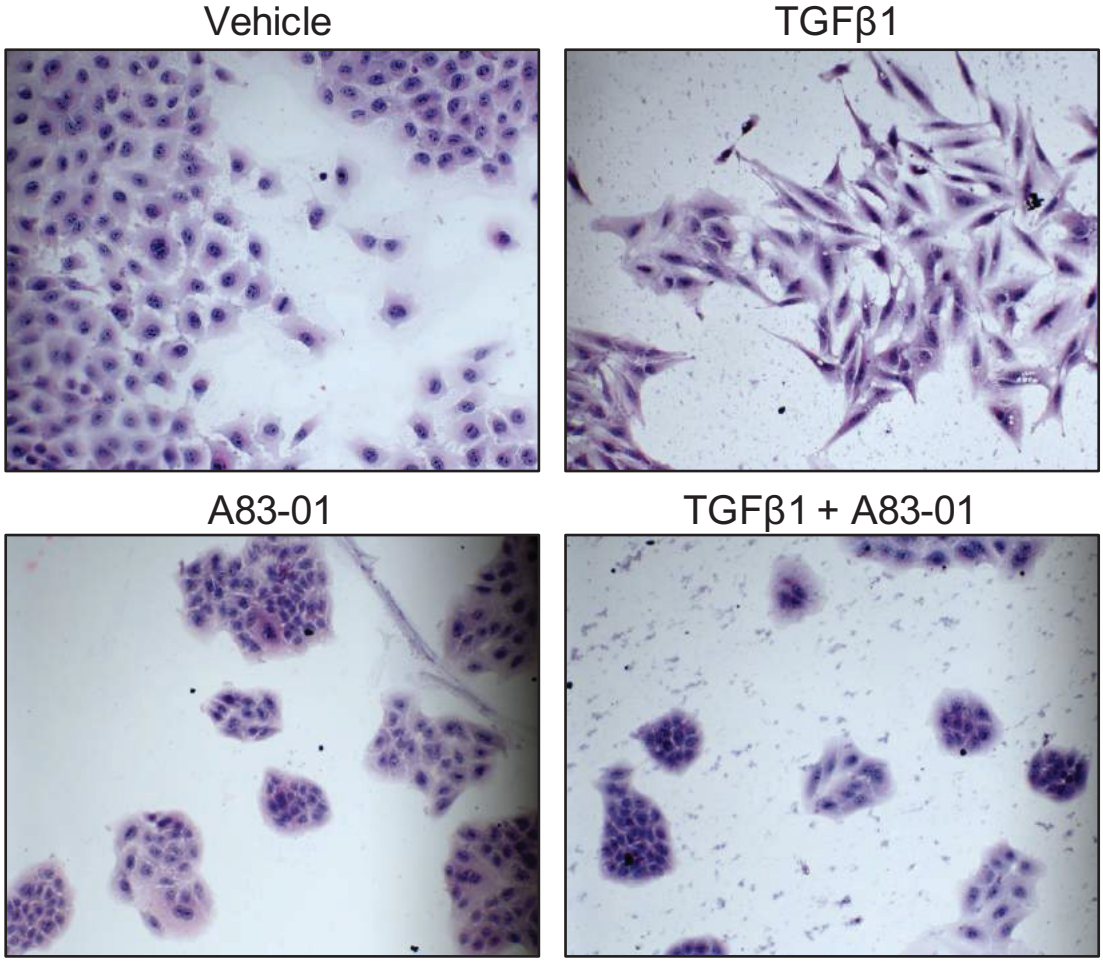
necessarily sufficient to attenuate EMT. In fact, poorly-differentiated tumor regions (which represent EMT) were not affected by Snail or Twist1 depletion suggesting that the assumption that EMT was suppressed by the genetic depletion of either of these two factors was inaccurate. Instead, Aiello suggested that the occurrence of EMT was in fact still plausible independently of the absence of these transcription factors [690]. Xe *et al* raised similar concerns regarding the use of FSP1 and vimentin as "gate-keeping" EMT markers. Although FSP-1 is required for EMT activation *in vitro* in renal proximal tubular epithelial cells [692], Xu *et al* eluded to the fact that extending that assertion to malignant mammary epithelial cells was largely inaccurate. FSP-1 knockout mice undergo normal embryogenesis and are viable and fertile which undermines the necessity of FSP-1 for EMT. Vimentin was also found to be expressed in tumor-associated fibroblasts which were still epithelial as indicated by the absence of vimentin-induced RFP-positive cells [691]. At the date of publication of this dissertation, the controversial role of EMT in metastasis has not been solved and additonal *in vivo* models of EMT are still requried to discern methodoligacal inconsisencies from biolgocial differences. Once resovled, the *in vivo* role of palsminogen activation genes can then be addressed.

10.12 Plasminogen activation and cancer: A "revived" association

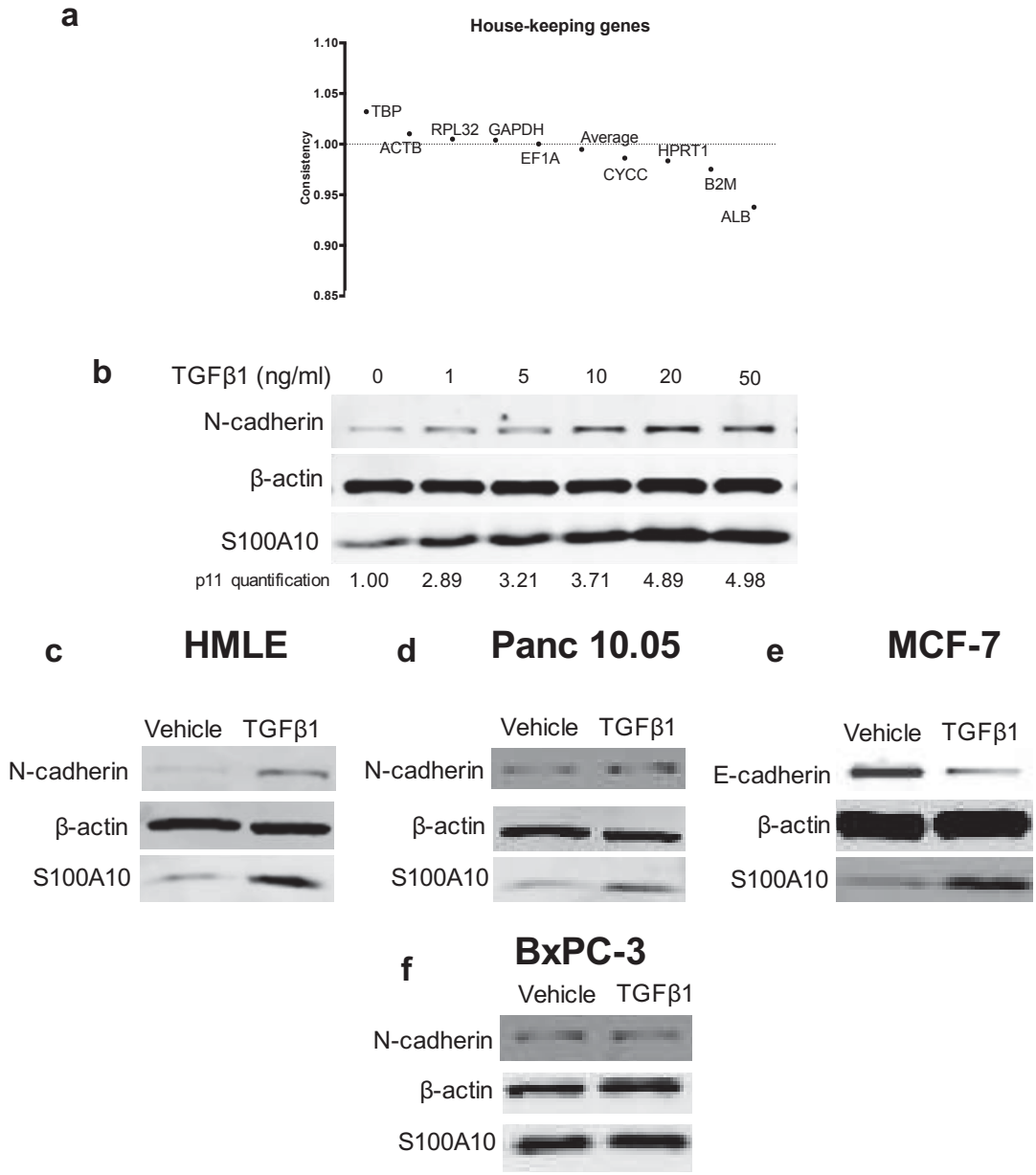
The golden era of studying plasminogen activation in cancer is often attributed to the 1980s. A fundamental goal of this dissertation was an attempt to revive the golden era by generating novel associations between plasminogen activation genes and key cancer processes (e.g. EMT, KRAS signaling, methylation etc.) especially in the current era of "big cancer data". The dissertation suggested that plasminogen activation is a biologically relevant process that must be addressed in future studies of EMT and metastasis.

APPENDIX A: SUPPLEMENTAL FIGURES I

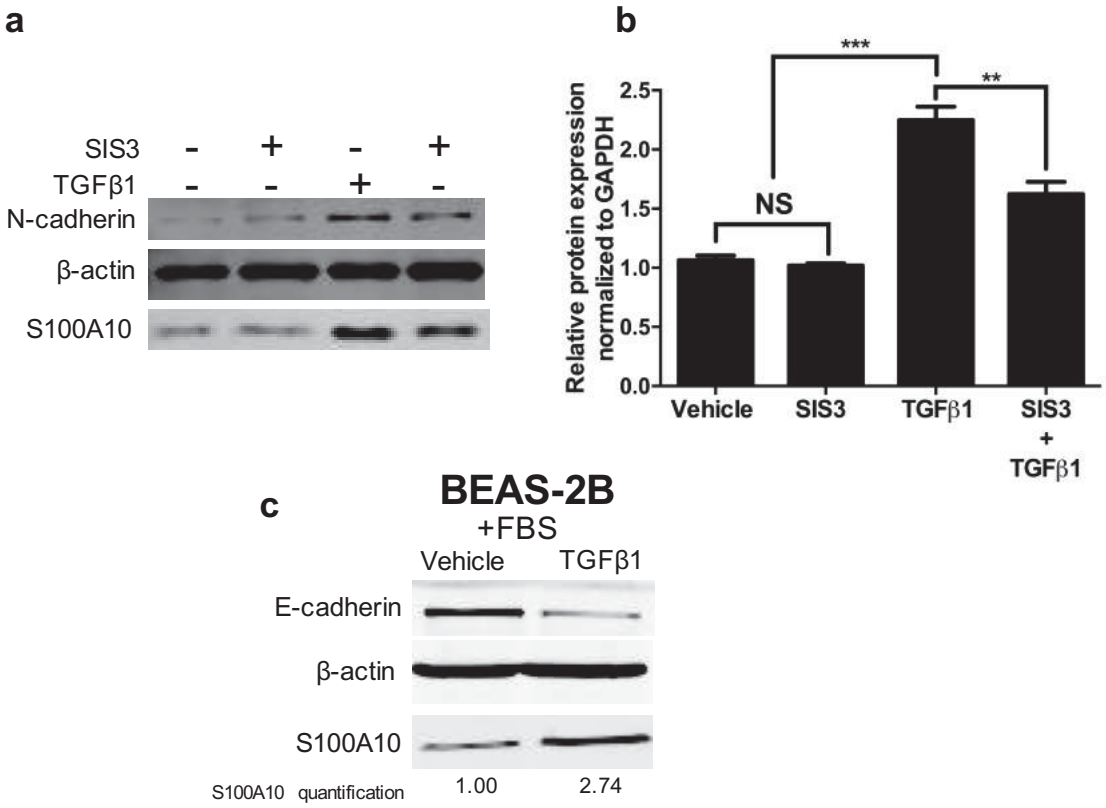
Supplemental Figure - 1. TGFβ1 treatment of A549 cells induces EMT that can be reversed by TGFβR1 inhibition. TGFβ1 induces a morphological change in A549 cells to become fibroblast-like mesenchymal cells (upper right panel) compared to vehicle-treated A549 cells (upper left panel). This change can be inhibited by the TGFβR1 inhibitor (A83-01) in vehicle-treated and TGFβ1-treated A549 cells. A83-01 generates a epithelial-like phenotype that appears to be more epithelial than the vehicle-treated cells.


Vehicle
TGFβ1
A83-01
TGFβ1 + A83-01

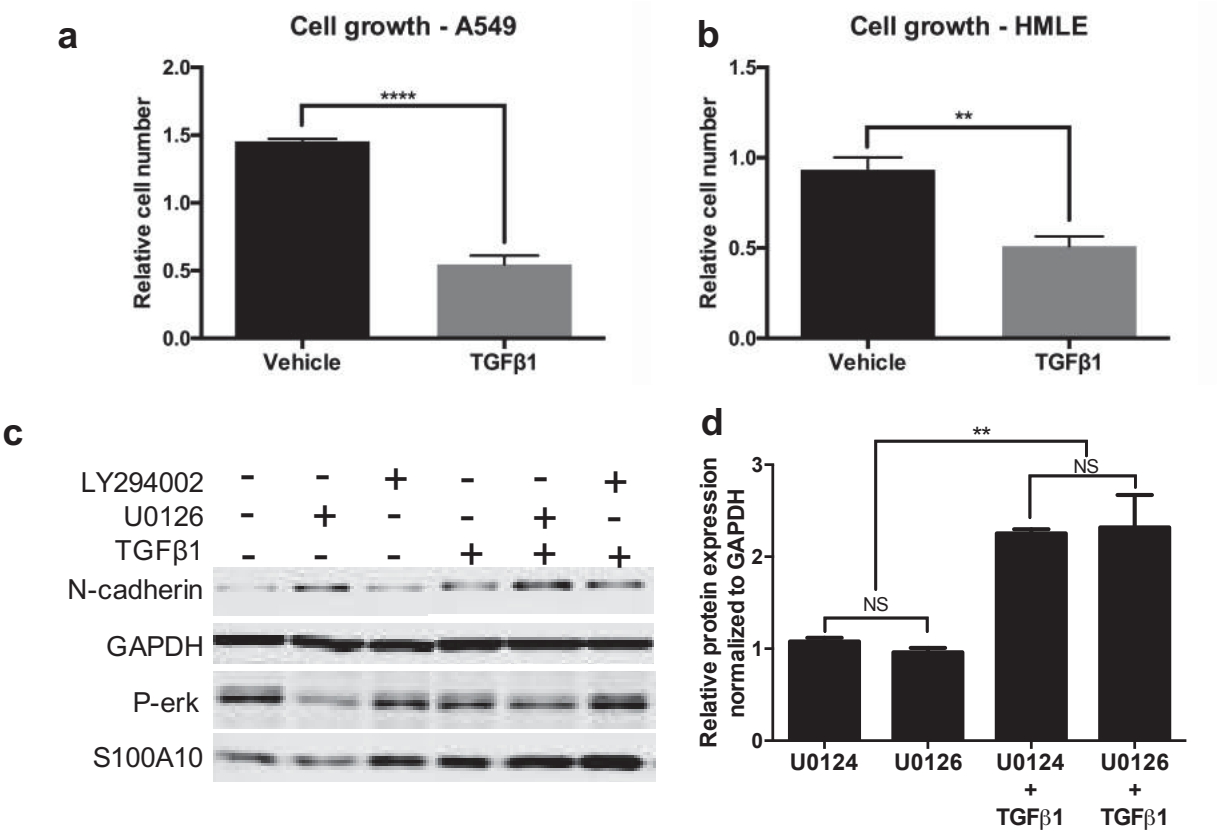
Supplemental Figure - 2. The effect of TGFβ1 treatment on S100A10 and other plasminogen receptors in multiple cancer cell lines. (a) Identification of the least variable house-keeping gene based on consistency of expression between untreated and treated samples. A value of indicates no change in expression between untreated and treated samples. (b) Western blot analysis of S100A10 in A549 cells treated with a increasing concentrations of TGFβ1 (0, 1, 5, 10, 20 and 50ng/ml). Western blot analysis of S100A10 in HMLE (c), Panc 10.05 (d), MCF-7 (e) and BxPC-3 (f) treated with 20ng/ml TGFβ1 for 8, 3, 4 and 4 consecutive days respectively.


a
House-keeping genes
Consistency
1.10
1.05
1.00
0.95
0.90
0.85
TBP
ACTB
RPL32
GAPDH
EF1A
Average
CYCC
HPRT1
B2M
ALB
b
TGFβ1 (ng/ml)
0
1
5
10
20
50
N-cadherin
β-actin
S100A10
p11 quantification
1.00
2.89
3.21
3.71
4.89
4.98
c
HMLE
Vehicle
TGFβ1
N-cadherin
β-actin
S100A10
d
Panc 10.05
Vehicle
TGFβ1
N-cadherin
β-actin
S100A10
e
MCF-7
Vehicle
TGFβ1
E-cadherin
β-actin
S100A10
f
BxPC-3
Vehicle
TGFβ1
N-cadherin
β-actin
S100A10

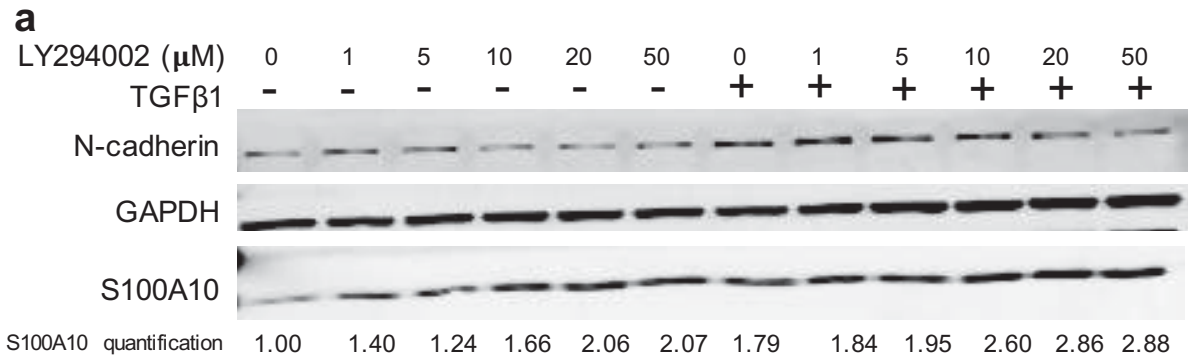
Supplemental Figure- 3. SIS3 treatment of TGFβ1-treated A549 cells abrogates S100A10 upregulation. (a) western blot analysis and quantification of S100A10 in A549 cells which were treated with the Smad3 inhibitor SIS3 (10**μ**M) in the presence or absence of TGFβ1. (c) TGFβ1 (20ng/ml) treatment of serum-supplemented BEAS-2B cells for 72 hours.

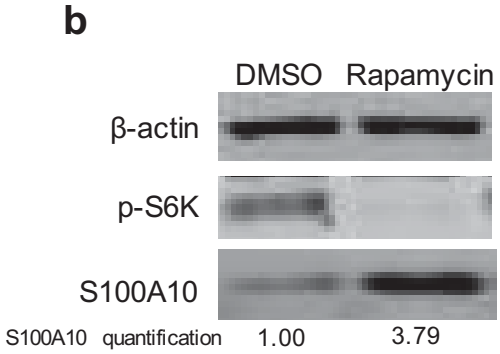

a
SIS3 - + - +
TGFβ1 - - + -
N-cadherin
β-actin
S100A10
b
***
**
NS
Relative protein expression normalized to GAPDH
2.5
2.0
1.5
1.0
0.5
0.0
Vehicle
SIS3
TGFβ1
SIS3 + TGFβ1
c
BEAS-2B
+FBS
Vehicle
TGFβ1
E-cadherin
β-actin
S100A10
S100A10 quantification
1.00
2.74

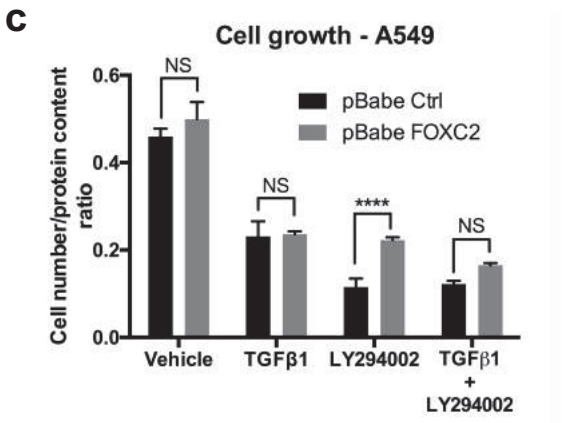
Supplemental Figure - 4. TGFβ1 suppresses the growth of A549 and HMLE cells *in vitro*. A549 (a) and HMLE (b) cells were counted after 4 days of vehicle or TGFβ1 treatment using the Trypan blue dye. (c) Western blot analysis and (d) S100A10 protein quantification in cells treated with the MEK inhibitor U0126 (and its negative control U0124) and PI3K inhibitor LY294002 in the presence or absence of TGFβ1.

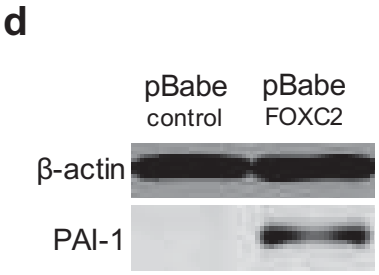

a
Cell growth - A549
Relative cell number
2.0
1.5
1.0
0.5
0.0
****
Vehicle
TGFβ1
b
Cell growth - HMLE
Relative cell number
1.5
1.0
0.5
0.0
**
Vehicle
TGFβ1
c
LY294002 - - + - - +
U0126 - + - - + -
TGFβ1 - - - + + +
N-cadherin
GAPDH
P-erk
S100A10
d
Relative protein expression normalized to GAPDH
3
2
1
0
**
NS
NS
U0124
U0126
U0124 + TGFβ1
U0126 + TGFβ1

Supplemental Figure - 5. LY294002 and rapamycin treatment of TGFβ1-treated A549 cells further increase S100A10 expression. (a) western blot analysis of S100A10 in A549 cells treated with increasing doses of LY294002 (0, 1, 5, 10, 20 and 50μM) in the presence of absence of TGFβ1. (b) Western blot analysis of A549 cells treated with DMSO or rapamycin (10μM) for 48 hours. (c) Quantification of cell growth in A549 pBabe ctrl and pBabe FOXC2 cells treated with TGFβ1 and/or LY294002 after 72 hours (d) western blot of PAI-1 in A549 pBabe ctrl and pBabe FOXC2.

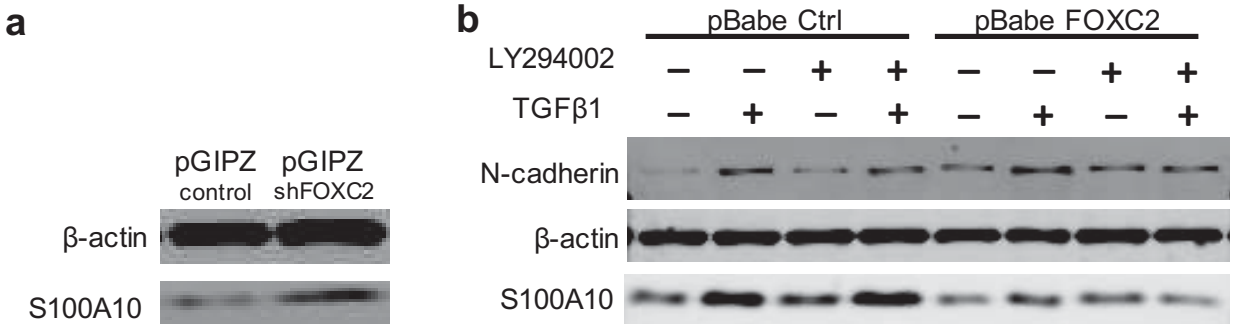

a
LY294002 (μM) 0 1 5 10 20 50 0 1 5 10 20 50
TGFβ1 - - - - - - + + + + + +
N-cadherin
GAPDH
S100A10
S100A10 quantification 1.00 1.40 1.24 1.66 2.06 2.07 1.79 1.84 1.95 2.60 2.86 2.88


b
DMSO Rapamycin
β-actin
p-S6K
S100A10
S100A10 quantification 1.00 3.79

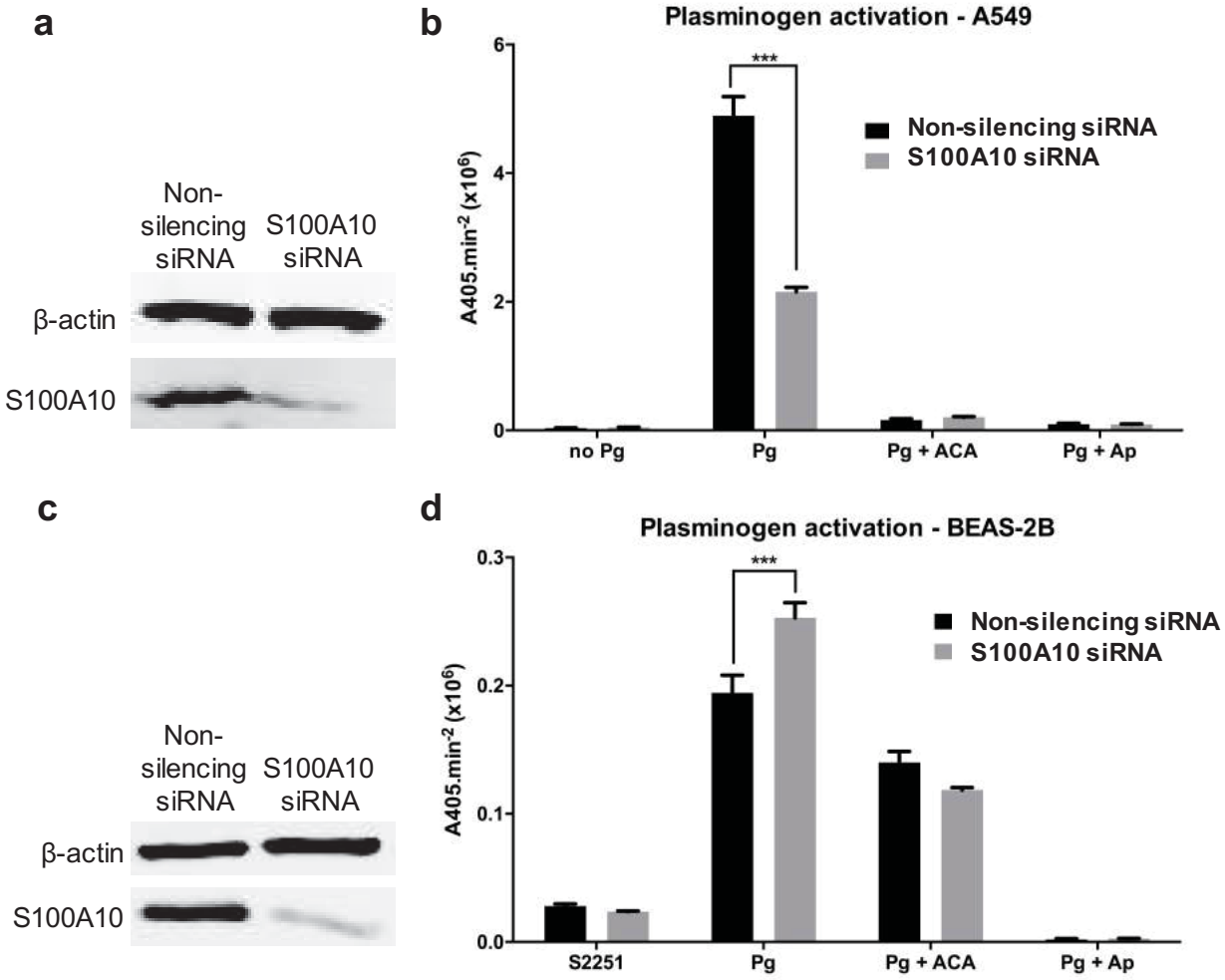

c
Cell growth - A549
Cell number/protein content ratio
0.6
0.4
0.2
0.0
NS
NS
****
NS
pBabe Ctrl
pBabe FOXC2
Vehicle
TGFβ1
LY294002
TGFβ1 + LY294002


d
pBabe control
pBabe FOXC2
β-actin
PAI-1

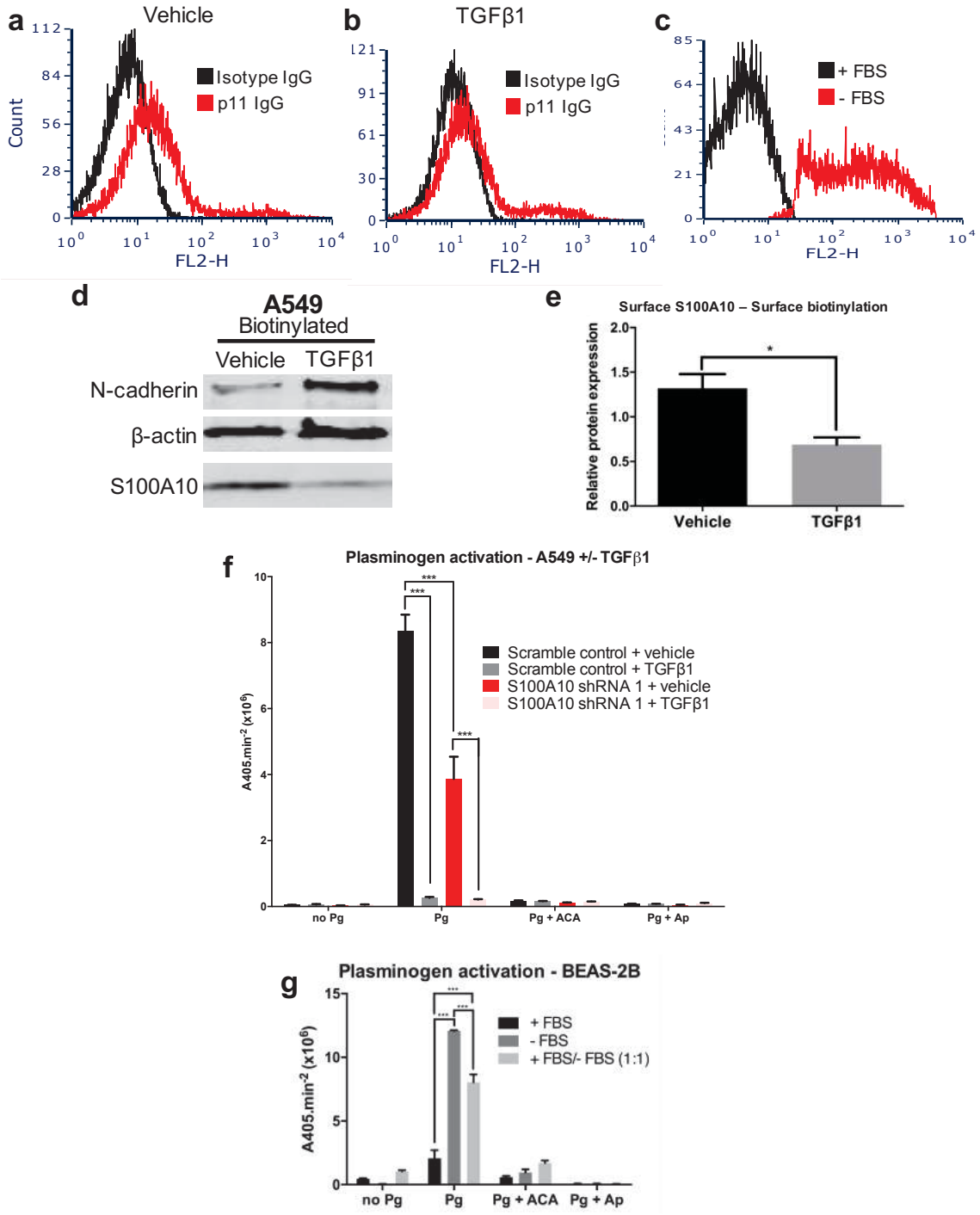
Supplemental Figure - 6. FOXC2 represses S100A10 expression despite the addition of TGFβ1 and LY294002. (a) western blot analysis of S100A10 in A549 cells stably transfected with pGIPZ control or pGIPZ shFOXC2. (b) western blot analysis of A549 cells stably transfected with pBabe control or pBabe FOXC2 treated with LY294002 and TGFβ1.


a
pGIPZ control
pGIPZ shFOXC2
β-actin
S100A10
b
pBabe Ctrl
pBabe FOXC2
LY294002
TGFβ1
N-cadherin
β-actin
S100A10

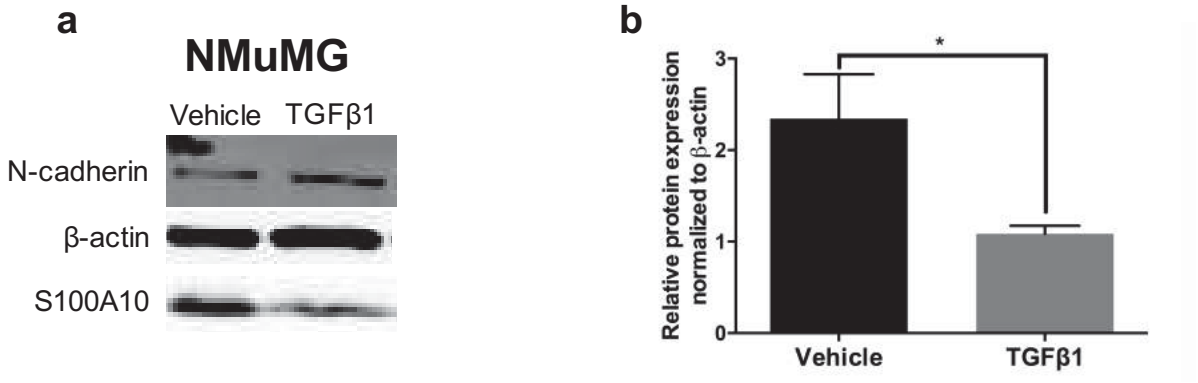
Supplemental Figure - 7. Effect of S100A10 siRNA depletion on plasminogen activation on the surface of A549 and BEAS-2B cells. Western blot analysis of A549 (a) and BEAS-2B (c) cells which were transiently transfected with non-silencing siRNA or S100A10 siRNA. Plasminogen activation assay of A549 (b) and BEAS-2B (d) cells transfected with non-silencing siRNA and S100A10 siRNA.


a
Non-silencing siRNA
S100A10 siRNA
β-actin
S100A10
b
Plasminogen activation - A549
***
Non-silencing siRNA
S100A10 siRNA
A405.min-2 (x10^6)
6
4
2
0
no Pg
Pg
Pg + ACA
Pg + Ap
c
Non-silencing siRNA
S100A10 siRNA
β-actin
S100A10
d
Plasminogen activation - BEAS-2B
***
Non-silencing siRNA
S100A10 siRNA
A405.min-2 (x10^6)
0.3
0.2
0.1
0.0
S2251
Pg
Pg + ACA
Pg + Ap

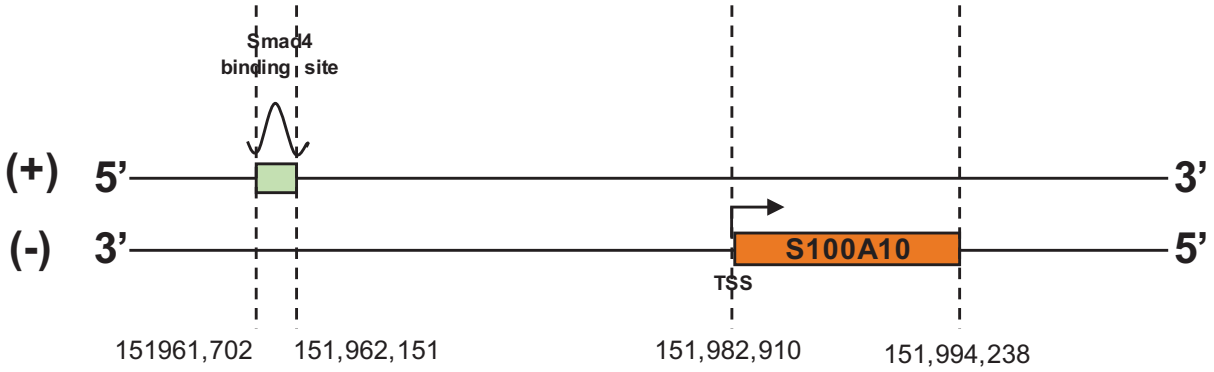
Supplemental Figure- 8. TGFβ1 treatment and serum supplementation abolishes plasminogen activation in A549 and BEAS-2B cells respectively partly due reduced cell surface expression of S100A10. Flow cytometry histogram of S100A10 surface expression (FL-2) of vehicle-treated (a) and TGFβ1-treated (b) A549 cells and serum-deprived BEAS-2B (c). Western blot analysis (d) and quantification (e) of S100A10 expression in biotinylated lysates from vehicle- or TGFβ1-treated A549 cells. (f) Plasminogen activation of A549 cells with scramble control or S100A10 shRNA 1 and treated with vehicle or TGFβ1. (g) Plasminogen activation of BEAS-2B cells in the presence or absence serum (FBS) or a 1:1 ratio of serum-free and serum-supplemented media.


a
Vehicle
Count
Isotype IgG
p11 IgG
FL2-H
b
TGFβ1
Isotype IgG
p11 IgG
FL2-H
c
+ FBS
- FBS
FL2-H
d
A549
Biotinylated
Vehicle
TGFβ1
N-cadherin
β-actin
S100A10
e
Surface S100A10 – Surface biotinylation
Relative protein expression
Vehicle
TGFβ1
f
Plasminogen activation - A549 +/- TGFβ1
A405.min-2 (x10^6)
Scramble control + vehicle
Scramble control + TGFβ1
S100A10 shRNA 1 + vehicle
S100A10 shRNA 1 + TGFβ1
no Pg
Pg
Pg + ACA
Pg + Ap
g
Plasminogen activation - BEAS-2B
A405.min-2 (x10^6)
+ FBS
- FBS
+ FBS/- FBS (1:1)
no Pg
Pg
Pg + ACA
Pg + Ap

Supplemental Figure - 9. The effect of TGFβ1 treatment on S100A10 in NMuMG cells. Western blot analysis (a) and quantification (b) of S100A10 in NMuMG treated with 20ng/ml TGFβ1 for three consecutive days.


a
NMuMG
Vehicle
TGFβ1
N-cadherin
β-actin
S100A10
b
Relative protein expression normalized to β-actin
3
2
1
0
*
Vehicle
TGFβ1

Supplemental Figure - 10. Schematic representation of the proposed SMAD4 binding site with respect to S100A10 gene. The distance between gene and the binding site is drawn to scale and the annotations are based on the GRCh37.p13 assembly.


Smad4
binding site
(+)
5'
3'
(-)
3'
5'
S100A10
TSS
151961,702
151,962,151
151,982,910
151,994,238

APPENDIX B: SUPPLEMENTAL TABLES I

Supplemental Table -- 1. Gene expression analysis of 130 components of the plasminogen activation system in response to TGFβ1 treatment in A549 cells. These components include plasminogen activators (PLAU, PLAT), plasminogen activator receptors (PLAUR), plasminogen activator inhibitors (e.g. SERPINE1), plasminogen receptors (e.g. ENO1, HMGB1, RUVBL1, S100A10), MMPs, MMP inhibitors (TIMPs) and kallikreins (KLKs). The expression data of vehicle-treated and TGFβ1-treated (72-hour time point) cells was obtained from the gene expression omnibus (GEO; access code GSE17708) (Sartor et al. 2010). The expression values were first normalized against the expression house-keeping genre EF1A then against a sample with the lowest normalized expression value. The cut-off for the adjusted p-value was 0.05.

Gene	Significant ?	P value	Mean1	Mean2	Difference	SE of difference	t ratio	df
ACTB		0.0655417	1.06665	1.88338	-0.816729	0.32444	2.51735	4
ANXA2		0.656983	140.376	151.382	-11.006	22.9796	0.478947	4
CTSA		0.132142	1886.19	1074.68	811.511	429.955	1.88743	4
CTSB	*	0.00804488	2572.34	5718.74	-3146.41	642.144	4.89985	4
CTSC		0.260814	9360.06	7389.73	1970.32	1505.76	1.30853	4
CTSD	*	4.50158E-05	16731.6	2174.22	14557.3	765.378	19.0198	4
CTSE		0.369556	193396	243732	-50336.3	49829.3	1.01017	4
CTSF		0.134388	117314	75885	41429.3	22121.1	1.87285	4
CTSG		0.412099	267066	208433	58633.1	64100.8	0.914702	4
CTSH		0.116163	72219.2	42703.8	29515.3	14760.2	1.99965	4
CTSK	*	0.00518956	53672.3	92980.4	-39308.1	7094.97	5.54028	4
CTSL		0.0620989	303.94	192.236	111.704	43.4954	2.56817	4
CTSO		0.0850339	197375	130735	66640.8	29262.3	2.27736	4
CTSS	*	0.00305958	96691.7	9001.16	87690.5	13700	6.40077	4
CTSV		0.917622	52691.4	54063.5	-1372.08	12460.5	0.110114	4
CTSW		0.637492	291693	332518	-40825	80193.1	0.509083	4
CTSZ		0.111297	194861	125575	69286.4	34010.5	2.0372	4
ENO1		0.0814024	3106.09	2283.33	822.756	355.092	2.31702	4
HIST1H2BE		0.757777	4258.15	4655.26	-397.108	1202.41	0.330259	4
HMGB1		0.344148	116.294	193.952	-77.6581	72.4519	1.07186	4
ITGA2B		0.598057	323544	343787	-20242.3	35401.1	0.571799	4
ITGAM		0.0890787	290975	193003	97972.1	43828.2	2.23537	4
ITGB2		0.493036	278371	319138	-40767.2	54098.9	0.753569	4
ITGB3		0.407994	239912	270401	-30488.4	33012.5	0.92354	4
KRT8	*	0.0245433	9.70558	4.25953	5.44606	1.54903	3.51578	4
MMP1	*	0.00469768	66499.3	373558	-307059	53920.4	5.69467	4
MMP10	*	0.00541865	9846.51	462807	-452961	82742.1	5.47437	4
MMP11		0.269333	380369	275004	105365	82234.3	1.28127	4
MMP12		0.45227	297901	339352	-41450.5	49829.2	0.831851	4
MMP13		0.339444	535358	697862	-162503	149950	1.08372	4
MMP14		0.482827	195934	210831	-14897	19279.4	0.772689	4
MMP15		0.945628	289955	284156	5799.18	79904.7	0.0725762	4
MMP16		0.827978	409330	397835	11494.8	49562.4	0.231925	4
MMP17		0.361487	332274	258496	73778.8	71675.7	1.02934	4
MMP19		0.055105	266018	200290	65728	24505.6	2.68216	4
MMP2	*	0.000152514	112405	288096	-175692	12581	13.9648	4
MMP20		0.853672	307686	296120	11566	58808.6	0.196672	4
MMP21		0.839517	425599	409984	15615.2	72274.9	0.216053	4
MMP23A/23B		0.259226	515981	331526	184454	140408	1.3137	4
MMP24	*	0.00589544	194384	83372.1	111012	20759.6	5.34752	4
MMP24-AS1		0.119836	42212.9	31079.9	11133.1	5644.36	1.97243	4
MMP25		0.305215	230548	207671	22877.1	19472.2	1.17486	4
MMP26		0.311743	421357	268669	152687	131994	1.15677	4
MMP27		0.610768	225605	244811	-19206.6	34838.1	0.551309	4
MMP28		0.159793	228389	281823	-53433.3	30993.1	1.72404	4
MMP3	*	0.0395219	245540	321228	-75687.8	25139.3	3.01073	4
MMP7	*	0.00348635	199413	238.3	199175	32237.4	6.17838	4
MMP8		0.264265	367913	313873	54039.8	41652.9	1.29739	4
MMP9		0.173403	125029	174036	-49007.8	29623.9	1.65433	4
PLAT	*	0.00501131	59076.7	327208	-268131	47931.2	5.59408	4
PLAU		0.0622362	114.606	10509	-10394.4	4050.67	2.56608	4
PLAUR	*	0.00604463	2644.46	25480.7	-22836.3	4300.3	5.31039	4
PLG		0.761346	390284	362913	27371.1	84173.3	0.325175	4
PLGRKT		0.0596164	16554.7	8881.24	7673.5	2943.58	2.60686	4
RUVBL1		0.0613583	6309.61	4839.27	1470.35	570.007	2.57952	4
S100A10	*	0.000508743	83.6273	423.151	-339.523	33.091	10.2603	4
S100A4	*	0.021085	30.8968	6.87697	24.0199	6.51579	3.68641	4

Gene	Significant ?	P value	Mean1	Mean2	Difference	SE of difference	t ratio	df
SERPIN A1		0.977307	72287.8	73351.7	-1063.87	35154.1	0.0302631	4
SERPIN A10		0.437131	408445	316902	91543.4	106157	0.862344	4
SERPIN A12		0.177996	209476	275862	-66386.1	40675.4	1.6321	4
SERPIN A2		0.542926	410246	443578	-33332.2	50188.3	0.664143	4
SERPIN A3		0.0729967	358348	204801	153547	63527	2.41704	4
SERPIN A4		0.0833094	279671	206226	73444.7	31988.8	2.29595	4
SERPIN A5		0.42862	215158	277864	-62706.4	71268.4	0.879863	4
SERPIN A6	*	0.000446666	399969	47314.7	352655	33237.6	10.6101	4
SERPIN A7		0.30982	322074	227708	94365.8	81205.2	1.16207	4
SERPIN A9	*	0.0180716	64698.5	123471	-58772.9	15206.8	3.8649	4
SERPIN B1	*	1.95894E-05	61986.9	5475.71	56511.2	2409.43	23.4542	4
SERPIN B10		0.856401	312501	326968	-14467.9	74983.7	0.192947	4
SERPIN B11		0.505211	364216	324753	39463.4	53974.3	0.731153	4
SERPIN B12		0.985868	283861	281805	2056.6	109138	0.018844	4
SERPIN B13		0.272733	425399	384133	41266.5	32477.2	1.27063	4
SERPIN B2		0.820274	494131	461659	32471.7	133870	0.242562	4
SERPIN B3	*	0.0151873	352309	207454	144855	35566	4.07285	4
SERPIN B3/B4	*	0.010058	246752	130380	116372	25317.9	4.59642	4
SERPIN B4		0.126558	449017	311405	137611	71490.6	1.92489	4
SERPIN B5		0.866615	448047	466398	-18351.3	102503	0.179032	4
SERPIN B6		0.659616	145559	161584	-16024.8	33742.6	0.474914	4
SERPIN B7	*	0.00361606	285737	104691	181046	29595.5	6.11735	4
SERPIN B8		0.336268	121859	98346.9	23511.7	21534.5	1.09181	4
SERPIN B9		0.271465	59337.3	43647.1	15690.2	12310	1.27459	4
SERPIN C1		0.557411	339460	310797	28663.4	44836.6	0.639287	4
SERPIN D 1		0.909029	412118	419110	-6992.77	57474.1	0.121668	4
SERPIN E1	*	0.015365	225.648	188463	-188237	46378.5	4.05871	4
SERPIN E2	*	0.000295171	409.58	28035.5	-27625.9	2341.13	11.8003	4
SERPIN F1		0.131914	376471	302248	74223	39293.7	1.88893	4
SERPIN F2	*	0.0396057	222236	161020	61216.6	20347.3	3.00858	4
SERPIN G 1		0.340708	387056	508577	-121521	112465	1.08052	4
SERPIN H 1		0.165028	9334.56	14130.3	-4795.74	2826.82	1.69651	4
SERPIN I1		0.0738383	237296	373100	-135804	56433.3	2.40645	4
SERPIN I2		0.61928	319009	307210	11798.8	21941.4	0.53774	4
TIMP1		0.386609	102.239	74.0368	28.2023	29.0497	0.970828	4
TIMP2	*	0.0121503	968.717	27537	-26568.2	6106.64	4.35071	4
TIMP3	*	0.000638519	17834.8	246146	-228311	23597.2	9.67534	4
TIMP4		0.369286	73150.2	51614.7	21535.5	21305.2	1.01081	4
KLK1		0.982762	141113	140693	420.128	18277.1	0.0229866	4
KLK2		0.637777	362861	396745	-33883.9	66617	0.508638	4
KLK3		0.918952	290340	292330	-1990.5	18374.7	0.108329	4
KLK4		0.747751	180908	171971	8937.03	25934.6	0.344599	4
KLK5		0.160476	100580	81143.5	19436.2	11297.6	1.72039	4
KLK6		0.0988511	139809	177246	-37437.4	17476.7	2.14213	4
KLK7		0.969859	201993	200993	999.453	24861	0.0402016	4
KLK8		0.544346	194907	219726	-24818.8	37508.5	0.661685	4
KLK9		0.309813	358755	296180	62575.3	53847.5	1.16208	4
KLK10		0.141366	539383	356994	182390	99712.3	1.82916	4
KLK11		0.28774	262501	355155	-92654.5	75629.4	1.22511	4
KLK12		0.529046	166905	158212	8692.7	12627.4	0.688401	4
KLK13		0.426468	309352	367068	-57716.3	65265	0.884338	4
KLK14	*	0.00230252	313009	202912	110097	15935.8	6.90876	4
KLK15	*	0.0178079	284197	351997	-67799.7	17464.2	3.88221	4

Supplemental Table -- 2. SMAD4 proposed binding location at the 3' distal region of S100A10. (a) The genome-wide CHIP (chromatin-immunoprecipitation) was performed by Kennedy et al to identify Smad4 binding sites in response to TGFB1 treatment. (b) The location of the Smad4 peak was determined based on the Kennedy et al annotation as well as the GRCh38.p7 and p13 assemblies.

a

RefSeq_ID	Gene Symbol	Gene_ID	Chromosome	Gene Strand	Gene Region	P-value	Binding Pattern
NM_002966	S100A10	6281	chr1	-	3` Distal	8.91E-17	Binding stimulated only with TGFβ1

b

Genome annotation	Gene location	Peak Start	Peak End	Gene Start	Gene End
Kennedy et al.	50,233,338-150,222,009	150,200,751	150,201,250	150,233,338	150,222,009
GRCh38.p7 (GCF_000001405.33)	151,982,910-151,994,238	151,934,128	151,934,627	151,966,714	151,955,386
GRCh37.p13 (GCF_000001405.25)	151,955,386-151,966,714	151,961,702	151,962,151	151,994,238	151,982,910

APPENDIX C: SUPPLEMENTAL FIGURES II

NONE

APPENDIX D: SUPPLEMENTAL TABLES II

Supplemental Table -- 3. 26 differentially-expressed PA genes in NSCLC vs. SCLC with at least 2-fold difference and a p-value < 0.01. The table shows the absolute t-value, degrees of freedom, raw and adjusted p-values, FDR (false-discovery rate) as well as SCLC and NSCLC mean z-scores, SD (standard deviation) of these means and fold-change.

Gene	Absolute t value	degrees of freedom	raw p value	adj p value	FDR	SCLC mean	NSCLC mean	fold-change	SCLC SD	NSCLC SD
ADAM9	12.44011	83	5.5522E-18	8.99871E-14	2.19569E-16	0.443677248	1.723626498	3.884865644	1.991658483	1.758786179
CTSZ	10.791589	111	4.72491E-17	7.65787E-13	1.86852E-15	0.343740867	1.484767508	4.319438427	2.173178842	2.4206732 4
PLAU	12.955367	164	1.42076E-16	2.30269E-12	5.61856E-15	0.436188495	1.922955701	4.408542915	1.6044240 1	2.736901935
PLAUR	10.082667	103	1.42748E-16	2.31359E-12	5.64515E-15	0.474049136	1.928387767	4.067906933	2.270231707	2.364398605
S100A10	11.832253	66	1.64942E-16	2.67329E-12	6.52282E-15	0.283079104	1.672502236	5.908250418	2.747417007	1.780743725
CTSD	9.370407	117	6.66E-16	1.08E-11	2.63E-14	0.410296644	1.34842858	3.286472365	2.037672347	2.354922834
CTSB	9.520729	98	1.33E-15	2.16E-11	4.80E-14	0.333622188	1.046268024	3.136086453	2.049855433	2.053929525
ADAM15	9.003679	117	4.88E-15	7.91E-11	1.56E-13	0.434286815	2.183917014	5.028743542	2.726077874	3.370766613
ADAM22	9.677279	68	2.04E-14	3.30E-10	5.64E-13	3.411046719	0.746216744	0.218764739	2.854130947	1.887761055
SERPINB1	8.890039	103	2.15E-14	3.48E-10	5.93E-13	0.468914584	1.36221275	2.905033872	2.025555735	2.1036626 4
CTSL	9.457141	73	2.60E-14	4.19E-10	7.07E-13	0.547816194	1.889182697	3.448570371	2.353663617	1.807188387
PLAT	7.890197	159	4.59E-13	7.32E-09	9.39E-12	0.503944821	1.184377535	2.350212736	1.447555063	2.781158034
ADAM8	8.037784	124	6.14E-13	9.79E-09	1.21E-11	0.532147485	1.370239059	2.574923487	1.1970231 3	3.432272575
SERPINB6	8.8303175	61	1.66E-12	2.64E-08	2.98E-11	0.419889583	1.499799406	3.571890009	2.695671935	1.5909948 3
CTSA	8.390075	76	1.97E-12	3.12E-08	3.47E-11	0.484856746	1.807631455	3.72817635	2.760014132	2.098631992
ANXA2	8.753867	57	3.91E-12	6.16E-08	6.47E-11	0.475320721	1.697303474	3.570859417	2.768813807	1.448815607
SERPINH1	7.9631667	83	7.70E-12	1.21E-07	1.19E-10	0.46096784	1.22618724	2.660027737	2.173075184	1.887557188
SERPINE1	7.250604	161	1.64E-11	2.57E-07	2.36E-10	0.538855663	1.234690116	2.291318807	1.659574759	2.631711776
S100A4	7.1447678	127	6.24E-11	9.65E-07	7.89E-10	0.420333935	0.991467897	2.358762436	1.910983059	2.351914083
KLK6	7.0509887	122	1.16E-10	1.78E-06	1.38E-09	0.566236049	1.174014028	2.073365039	1.156694918	2.967326562
CTSH	6.9042516	133	1.88E-10	2.88E-06	2.12E-09	0.488241968	1.135520709	2.325733516	1.904789112	2.4464566 2
CTSC	7.218834	67	6.14E-10	9.28E-06	6.08E-09	0.329805596	1.391840609	4.220184933	3.803886092	2.208493039
HMGB1	6.014186	95	3.35E-08	4.83E-04	2.31E-07	1.150514945	0.489407795	0.425381519	2.359134159	2.300065498
TIMP4	5.5155873	161	1.36E-07	0.001914104	8.22E-07	0.641260685	1.290780327	2.01287925	1.565022702	3.273597064
SERPINB8	5.6532497	104	1.39E-07	0.001955517	8.38E-07	0.417048023	0.908860665	2.179271006	2.242441214	2.365472877
SERPINB5	5.4447436	126	2.61E-07	0.003628267	1.48E-06	0.951663065	2.136126623	2.244624912	2.232139695	2.867982153

Supplemental Tables 4 to 11. The genes in each of the ten relevant clusters that are differentially-expressed in SCLC and NSCLC. PA genes are highlighted in red.

Cluster 1 (168 genes)			
DLL3	FAM222A	MTURN	SOX11
HFM1	FBLL1	MYCL	SPHKAP
HOXD10	FGF12	MYO3A	SRRM4
NKAIN2	FGF14	MYT1	ST18
STXBP5L	FNDC5	NAPB	ST8SIA3
SEPT3	FOXG1	NECAB1	ST8SIA5
ACTL6B	FRAS1	NELL1	SVOP
ADCY1	FUT9	NOL4	SYT1
ADD2	FXYD6	NOVA1	SYT14
AMN1	GABRB3	NRCAM	SYT4
ANKRD65	GAD2	NRSN1	TAGLN3
AP3B2	GADD45G	NRXN1	TCERG1L
ASCL1	GDAP1	ONECUT2	TEKT2
ASXL3	GHRH	PAK7	TIGD3
ATCAY	GNAO1	PCDH8	TMEM108
BAI3	GNAZ	PCDHA9	TMEM178A
BSN	GNG4	PCLO	TMEM198
BTBD17	GNGT1	PCP4	TMEM74
CACNA1A	GPR12	PCSK1	TMOD2
CACNA1B	GPR98	PCSK2	TRIM9
CADPS	GRIK3	PEX5L	TRIT1
CALCB	GRIP1	PGAP1	TUBB2B
CAMK2B	GRM8	PGBD5	UBE2QL1
CAMK2N2	HCN3	PHYHIPL	UNC13A
CBFA2T2	HEPACAM2	PLEKHG4B	UNC80
CCDC177	HES6	POU4F1	ZBTB8B
CCDC181	HOXD11	PPM1E	ZMAT4
CDK5R2	HPCAL4	PROX1	ZNF334
CECR6	IGSF21	RAB39A	
CHGA	INA	RAB3C	
CHGB	INSM1	RALYL	
CNTNAP2	ISL1	RAPGEF4	
CPLX1	JAKMIP2	RHBDL3	
CRB1	KALRN	RIC3	
CRMP1	KCNA1	RIMBP2	
CRTAC1	KCNB2	RIMS2	
CRYBA2	KCNC1	RIPPLY2	
CSRNP3	KCND2	RIPPLY3	
DAPL1	KCNH7	RMST	
DDX25	KCNH8	RNF183	
DGKB	KCNK3	RPRM	
DIRAS2	KCNMB2	RTN1	
DLL1	KCNT2	RUNDC3A	
DLX6	KIF19	RUNX1T1	
DNALI1	KIF1A	SBK1	
DOK6	KIF5C	SCAMP5	
DPP10	KSR2	SCG3	
DPYSL5	LHFPL3	SCN3B	
DSCAM	LHFPL4	SETBP1	
DTNA	LINC01018	SEZ6	
DUSP26	LOH12CR2	SEZ6L	
DYNC1I1	LRFN5	SH3GL2	
EFR3B	MAATS1	SHC2	
ELAVL3	MAP6	SIX6	
ELAVL4	MAPT	SNAP25	
ESRRG	MARK1	SNAP91	
FAM105A	MFSD2A	SNCAIP	
FAM184A	MIAT	SOGA3	
FAM19A1	MICALCL	SOWAHA	
FAM211A	MTMR7	SOX1	

Supplemental table 4

supplemental Table -- 4,

Cluster 2 (n=626)										
PCDH17	PI4KA	AGL	MARCKSL1	CYCS	LRRC8B	CAMSAP1	SS18L1	PGS1	RADIL	SALL3
HIRIP3	RHOT1	DHFRL1	SOX4	SBNO1	SSBP3	EML5	EPC2	MDC1	IPO9	ATP6V0E2-AS1
SFT2D3	RCOR3	PAXIP1-AS1	SP4	IL17RB	CADM1	CELSR3	GCK	MATR3	C11ORF30	NELL2
PASK	KMT2E	EPB41	STXBP1	CHD7	AUTS2	ENO2	B3GAT1	DMTF1	FAM155A	C1ORF109
RNFT2	PARP6	KDM1B	LOC730101	IQCC	ZNF337	AMER2	ZNF775	CDKN2A	VASH2	DHX40
C2ORF68	MAN1C1	DKFZP586I1420	CACNB2	PELI2	HECTD4	HNRNPA3	C17ORF100	ZNF664	JAM3	HSDL1
ZNF124	ZNF398	RSBN1L	ZXDC	KIAA1147	ANK2	SORCS3	MGC57346	BEST3	CHD6	AKAP5
FZD3	KLHL14	ZNF652	TMEM180	KIAA1211	HUNK	CEP68	TMEM151B	LINC00526	EFCAB7	RNASEH2A
H3F3B	SIPA1L2	PAXIP1	ZSCAN16	USP37	SOCS7	TARDBP	MT3	ELAVL1	CDK5R1	GPR19
KDM1A	TMEM181	PDZRN3	TTC25	RAP2A	HSD11B2	ABCA5	ASTN1	AK9	PHF14	NOTCH4
ESCO1	FBRSL1	MCF2L2	KISS1R	RBBP8	NMNAT3	ZNF620	ZBTB21	RNF182	MANEAL	BARX1
TATDN3	PTPRD	NFASC	ATG4D	MRPS14	MLLT4	LCORL	LUC7L3	DDHD2	ZNF662	ERI3
GRIA1	ZNF77	SYNE2	TRIM24	BCL11B	CD200	FAM19A5	ATP8A2	IFT81	RBM12B	ICA1L
ID2	RGS12	VPS8	PYGO2	MRPS26	MSI1	NSG1	PHIP	HSF2	ATL1	FGF9
ILDR2	ZNF704	ADNP	L3MBTL4	SLIT1	NASP	GNB1	FANCL	ID4	ELOVL2	SEC22A
BCAS3	MCUR1	KCNH3	RPS6KA5	SOX12	PHOSPHO2	EPHA10	YEATS2	GATS	DZANK1	TRIM36
MAPK10	GZF1	HEY1	ACVR2B	TAL2	OPA1	C3ORF70	C21ORF59	CTXN1	SCG5	RNF168
TMX4	SP2	INTS7	ZNF516	TERF1	SUV420H1	XKR6	TULP4	SMAD9	RASL11B	TTLL7
RANBP2	TROVE2	MYRIP	TMEM170B	TTC3	PGAM2	HMX2	ZFP14	WHSC1L1	ALMS1	MUC15
XCL1	UGT8	TENM4	ABCC5	ZNF3	POU3F2	ZDHHC21	FLJ37453	DCUN1D2	LINC00626	NPTXR
SLC18A2	ZKSCAN1	ERC2	DCAF7	ZNF74	PPM1A	ARG2	XRCC5	ENAH	TET1	MTERFD3
MZF1	ZNF133	NARF	RTN3	ZNF669	GID8	LOC441204	FAXC	SHD	FZD9	ACYP1
CACNB3	ZNF195	AGO1	RNF157	BRD3	UBR7	NDUFA5	KRTAP3-3	ZNF287	CHST9	DNAL1
LOC81691	SECISBP2	GATC	TADA1	PLA2G12A	RAVER2	PIPOX	DCLK1	SALL2	FAM161A	LHX2
USP30	KIAA1683	KANSL1	PAQR4	ENKD1	ZNF821	REV1	TMEM169	RAB2B	FRS3	INO80E
SHANK3	C1ORF21	GRM3	KRT40	ST6GAL2	CAND1	TERF2IP	SV2A	DLK1	FAT3	LRRC49
RAB36	PUS3	GSTA4	COL2A1	CBX2	PCDHB4	NEURL1B	NFYC-AS1	LIN52	TMEM132D	
GAB2	SH3BGRL2	LOC338799	C22ORF39	TMTC4	GBA2	IFT122	HNRNPR	CEP41	SOCS4	
CDH8	TMEM246	FREM2	LYSMD4	PIK3R3	RBBP4	MDM1	TLK2	CASP8AP2	ANKRD6	
CDKN2D	HINT2	IGFBP2	ZNF785	IKBKAP	KLHL12	SMIM8	FAM19A4	KLHL32	EPB41L3	
RAB40B	INSM2	CRIP3	FAM117B	UNK	SCN1A	VANGL2	ZNF483	LINGO2	CAMTA1	
SLC27A3	DISP1	MUT	ZNF385B	H1FX	KIAA0895L	TBC1D24	DLX5	DPP6	VAX2	
C12ORF57	NCOA1	NRTN	DDX5	MTMR4	SNRPE	MIB1	SYT11	EPHA7	CECR2	
ZNF554	DCAF5	SCAPER	C6ORF118	SESTD1	SPAST	PTPRO	FYN	STOX1	FAM221A	
CCDC173	ACPL2	PLA2G3	DUSP8	ZNF764	TAF4	ZNF250	TCTEX1D2	SDK1	LSAMP	
TTC32	CYTH2	PARD6A	IMMP1L	NRXN3	C12ORF73	RBP1	ARMC8	KIAA1324L	TPPP3	
UBR3	ACVR2A	LCMT1	EIF4A2	PUM1	GKAP1	RFX3	UBXN7	ZNF776	CCDC40	
COX7A2	RNF144A	ZDHHC13	RDM1	NOS1AP	MED25	SMARCE1	COPG2	ZNF660	TBCCD1	
HINT3	SLC35E2	POU2F1	BPTF	TOMM20	RPAIN	RBM4B	GNL1	HNRNPU	IFT80	
FGFBP3	ARNT2	ING3	RTKN2	LOC100128288	MEGF11	PPM1D	MIR4697HG	ZNF678	ELOVL4	
LOC148709	KBTBD11	PPIA	GPC2	CDKN1B	C9ORF24	DCHS1	YPEL1	CCDC39	SEMA6D	
RNF187	PAN2	MBTD1	LPHN1	HNRNPA0	KIAA1737	VAPB	HDAC2	ENHO	PRKRA	
FBXO15	LINC01003	KLHL24	NCOA6	GLCCI1	MADD	H2AFV	HSBP1	KIT	CCNA1	
SP8	RCAN2	TMEM206	POGZ	WDR17	HERC2	SLC4A8	ZNF713	CRNKL1	BTF3L4	
NLGN1	ACAA2	PCMTD2	GSE1	SLC16A10	BAZ1B	RALGPS1	DRAXIN	H2AFY2	MEX3A	
GPATCH8	WASF3	OGDHL	PACS2	COLCA2	SAMD1	TOX	ZNF706	PKIA	MLLT11	
FBXO9	FAM71E1	SDR39U1	ASTN2	SLAIN1	PPP1R3E	LOC100294362	KIDINS220	PTPRZ1	CSTF3	
STAU2	GABARAPL2	CTNNBIP1	UBN2	KRBA2	C2CD5	ZBTB18	HACE1	ZBED5	CNTN4	
MORN3	C1QTNF3	GATAD2B	ZNF324	ZFP3	FAM20B	POLR3F	TXNDC16	STK33	TSPYL4	
IGSF10	CLK2	SLC12A5	ATRNL1	PARP1	RBM8A	RNPS1	SYN2	CCDC34	KIAA1467	
GRK4	AK7	KLHL42	NBEA	GLT1D1	EIF1	AP3M2	TCF12	RIMS3	MEAF6	
OLA1	SLC25A29	TNRC6C	GDPD1	DHPS	IVNS1ABP	DIDO1	TIA1	CPA2	WRB	
BAZ2B	ZNF786	TP53INP2	NKIRAS2	DPYSL3	SLITRK1	KLHDC3	TIAM1	FAM13C	FSD1L	
HOXD4	UBXN2B	KMT2C	HMGCS1	ARID2	COL9A2	IRAK1BP1	UBE2N	SARM1	FBXL16	
MMAB	DHRS13	RRAGD	HOXD8	ZNF48	C10ORF82	LSM14B	VEZF1	TMEFF2	GPR6	
PLEKHM3	PDIK1L	BCL7A	LNP1	FANCC	CRHR1-IT1	PLD5	BEND5	MYEF2	ENY2	
LINC00938	RASGEF1B	MOAP1	ARF5	FOXJ3	ILF3-AS1	C9ORF72	ASRGL1	STMN3	ERC2-IT1	
C12ORF76	RUNDC3B	ACTR6	NCAM1	RUFY3	CTNNA2	CCDC171	UCK1	ZNF853	CPT1C	
DNAJC27	ANKRD46	CCDC14	C18ORF56	KLHDC10	PPM1L	AZI1	KRTAP3-2	SOBP	MDH1B	
PDE7A	THAP5	EFCAB6	BCL11A	LPHN3	JAKMIP1	EBF3	ZNF397	RIC8B	LONRF2	

Supplemental table 5

Supplemental Table -- 5

Cluster 3 (424 genes)							
ANXA2	CYR61	PTPN14	PCBD2	CCDC85B	LMNA	LGALS3	GAPDH
SERPINB6	IL1RAP	PXN	DUSP11	KDELR3	SMAD3	LY6E	TMBIM6
PLAUR	IL6ST	PYGL	ACTRT3	RAB31	MOV10	HSBP1L1	VASP
S100A10	IL15	SNX6	MT4	CDC42EP1	ASL	NPAS2	TCIRG1
SERPINH1	IRAK2	RAP2B	SPPL2A	PLCD3	MTRR	SLC25A37	
CTSC	JAK1	RARS	AJUBA	AHNAK2	PCDH7	PKM	
CTSL	KPNA4	RBMS1	PPFIBP1	SDSL	TMED7	PPARG	
CTSZ	RHOC	RBMS2	HELZ2	LACTB	PLD1	RNF213	
PLAU	LAMB2	CCND1	CAV1	TRIM6	TMED9	RARG	
CTSA	LPP	RPE	CAV2	CLTB	TOR4A	RASA1	
SEC23A	MICB	RRAS	RGS20	GALM	SLC35F6	SH2D4A	
PROCR	MITF	CLIP1	ADAM9	ADK	HERC6	RBKS	
GAS2L1	MOCS2	SIL1	FADD	SETD9	RNLS	SLC4A2	
MAN1A2	ABCC1	MPP5	INPP4B	CD109	EDEM2	WNT3	
MAN1B1	MYO1C	INF2	IQGAP1	ADORA2B	TMEM30A	RHBDF2	
L3HYPDH	NBN	BLVRB	SPHK1	CSNK1A1	GPR126	OBFC1	
OSBPL10	NEDD4	SHC1	SQSTM1	B3GNTL1	SQRDL	TMEM133	
SLC31A2	NPC1	FNDC3B	BCL10	CTBS	RALB	S100A16	
CPNE8	NT5E	SPATA20	RPS6KA4	CTGF	RELA	ADRB2	
RMDN2	GPX8	NABP1	STBD1	CTNNA1	BCL3	EEF1D	
PRELID2	TMED5	SIPA1	MCFD2	CCNYL1	AVPI1	EPHB4	
SLFN5	KLHL5	SLC22A4	PAPSS2	SH3RF2	S100A11	EFR3A	
DUSP1	ANGPTL4	SOAT1	DDX60L	ITPRIPL2	ZFYVE21	GALE	
DUSP3	DDX47	SSFA2	MYADM	SGMS2	ARHGAP10	BCL9L	
EDN1	CRIM1	SYPL1	TMSB10	HBEGF	TFPI2	PACSIN3	
EMP1	RNF181	TCF7L2	PTPLA	DUSP5	THSD4	CARD10	
STOM	TNFRSF12A	TFPI	STK17A	EPHA2	LRRC8E	ITGA3	
EXT1	TMEM138	TGFBI	FAM114A1	MLKL	SLC35F5	TM4SF1	
ELL2	PDGFA	TGM2	TRIP11	ELK3	SFXN3	DHX32	
SAMD4A	CINP	THBS1	TRIP10	REEP3	C11ORF68	ESYT2	
MESDC2	PFKP	TK1	TRIP4	FHL2	TAGLN2	ARAP3	
DNAJC13	C11ORF24	TNFRSF1A	VAMP3	PHLDA1	BCAR3	UBE2H	
FOSL2	PELO	TLCD2	HOMER3	DKK1	HPS3	ARHGAP29	
ZBTB38	DHX29	VEGFC	IL27RA	FLNB	CARD6	SH3D19	
GALNT2	CROT	WFS1	CHST3	FAF2	IL17RC	ALDH3B1	
ATL3	PPIC	SLC30A1	BAG3	DNMBP	YBX3	OSBPL3	
SPATS2L	IMPAD1	AHNAK	VPS26A	SEC11A	BHLHE40	IL18	
RAB11FIP5	LPCAT2	C2ORF49	BRE	TMEM245	PDXK	MYO1E	
TSPAN17	C14ORF119	MAPKAP1	PRDX6	FUCA2	TRADD	S100A6	
BIN1	PLEKHB2	MBOAT7	RIN1	PROSER2	TNFRSF10D	CSTB	
PTRF	ATG16L1	USB1	FEZ2	SGMS1	IER3	EPAS1	
TRIM59	TMEM248	KLHL36	CD58	WWTR1	DPP9	B4GALT1	
N6AMT1	RALGPS2	SHCBP1	CD151	MYOF	SLC16A5	RHOF	
SNX8	LEPREL1	DOCK5	TRIM14	COQ2	SLC16A3	RNASE4	
SERTAD3	UEVLD	WWC2	KIAA0196	NEAT1	ARHGAP18	AGRN	
SERTAD1	TBC1D2	UXS1	MVP	C14ORF182	SLK	LTBR	
MDFIC	GALNT10	ZC3H12A	FSTL3	RNF149	BRE-AS1	GPRC5A	
ANXA1	ZDHHC7	SPSB1	MYL12B	GYG1	ARPC1B	CXCL1	
ANXA2P1	PACS1	COQ10B	YAP1	HEXB	BET1	ZNHIT6	
ANXA2P2	H2AFJ	FOSL1	CDC42EP2	ANXA4	MTMR11	TTC27	
ANXA2P3	BCAP29	HMGA2	SEMA3C	EHD4	SLC35D2	SSSCA1	
CFH	MAPK9	CALU	IFI44	ANXA7	PRSS23	ABTB2	
HFE	PCDHGB5	YIPF5	NPC2	IGFBP3	MGAT4B	DYSF	
EHD2	PRNP	TRIM7	PDLIM5	ITGB5	IFFO2	CD9	
HMOX1	POLE4	ACTN4	HEXIM1	KIFC3	CLCF1	ALDH1A3	
HRH1	AGTRAP	CAPG	TGOLN2	C15ORF52	LRP10	PRKCE	
BIRC3	ZNFX1	CAST	IGF2BP2	RNF207	GGCX	UNC93B1	
IFIT1	KIAA1191	RILP	PLK2	LASP1	FHOD1	CLDN1	
C15ORF38	SMAGP	CASP4	NEK6	LIF	TMOD3	SMURF1	
IGFBP4	NCEH1	TM2D2	RAB32	LIMK1	KCNN4	TKT	

Supplemental table 6

Supplemental Table -- 6

Cluster 4 (367 genes)						
ADAM10	CNN2	GLUD1	MED15	RHBDF1	CASP8	SAR1B
CTSB	FCRLB	GM2A	TMBIM4	CSRNP1	ZCCHC9	SNX14
SERPINB8	COMT	GNG5	ERAP1	ELOVL1	FLYWCH1	COPB1
EHD1	KLF6	GOLGA4	PFDN1	SLC30A5	RNF135	RAB18
KIN	VTI1A	SUMF1	PFN1	TTC23	ZDHHC16	SLFN12
HMGXB3	CACUL1	USP25	PLOD1	BMP1	ORAI1	PDLIM7
HMOX2	KCTD11	GMPPA	PML	SLC12A4	SPRYD3	IKBIP
SIAE	TICAM1	GYS1	PMM2	SP100	PHYKPL	
XYLT2	MOB3C	NRBF2	FXYD5	TRIM21	TRIM5	
TRIP6	SLC38A9	HADHB	C8ORF58	STAT6	MICALL1	
TTC1	DGKA	ANXA11	NANS	ZFP36L1	ITPRIP	
MFSD5	SPRED1	HLA-E	FBXW5	STK10	AP3B1	
SAMD8	DCTD	MR1	DNAJB12	STX4	MKNK1	
PAXIP1-AS2	DDOST	HPS1	WBP1L	SURF4	USO1	
ARPC2	DECR1	HSPA4	TMEM214	BTD	RTCA	
PITRM1	PAPD4	NDST1	RETSAT	TAP1	NUMB	
LYSMD3	PPP1R18	ICAM3	CMTM6	TAP2	SNX3	
ETHE1	DR1	IDE	NADSYN1	TGFB1	GBF1	
PLA2G15	HIGD2A	IFI16	MOB1A	TGFBR2	RIPK2	
EOGT	EFNA4	IFI35	BLOC1S4	TK2	SNAP23	
ACADVL	RILPL2	SP110	DRAM1	TPM4	TNFRSF10B	
MAP2K3	ELF1	IFIT3	C19ORF66	NR1H2	DYRK4	
PARP3	TVP23C	FAS	SLC35C1	UROS	SUCLG2	
MICA	EMP3	IL4R	PI4K2A	VCL	HDAC3	
SCAMP2	EPB41L2	IL15RA	HIF1AN	XRCC4	FCHSD1	
ACTR3	ERCC2	INPP5A	NECAP2	TRIM25	TRIM41	
ACTR2	ERF	ITGA5	PRKAG1	ZNF217	NMI	
PLIN3	ETFB	GSTK1	TMEM184C	ZYX	RFT1	
MFSD10	FAH	SFT2D2	EMC3	LUZP1	CCDC102A	
CDK7	CYB561A3	ARF4	SAR1A	IFRD2	ORAI3	
CALCOCO2	OAF	RHOG	B2M	MAPKAPK3	TRIP12	
TNIP1	FER	LGALS1	TM9SF3	SLMAP	SP140L	
GNB2L1	ENDOD1	LIPA	GPR108	TMEM109	QKI	
TRIM38	TBC1D9B	LNPEP	PSMB9	TNIP2	GSTO1	
CRTAP	WAPAL	M6PR	PSMB10	PLEKHF1	STX8	
CIB1	FLII	MAN2A1	AVEN	TMEM43	PIGB	
ATG7	FAM175B	MBNL1	ERGIC1	FYCO1	SEC24C	
ARL6IP5	FKBP15	MGAT1	CLK4	LRRK1	IKBKE	
RPP38	MAN2B2	MPG	SLAIN2	COLGALT1	CLINT1	
IFITM2	JADE2	EIF2AK4	VPS18	GSDMD	EDEM1	
ERLIN1	TRAM1	MYD88	ZBTB4	HPS6	CD97	
MYL12A	DNPEP	MYH9	POLD4	PTCD2	KIAA0141	
NFAT5	PPP1R15A	NAB2	CTDSP1	RIN3	EFCAB14	
YME1L1	BCL2L13	NFKB1	TRAPPC1	DHDDS	WDR1	
KIF1C	EHBP1L1	NFKB2	PCTP	DNAJB14	NAT1	
TOB2	EML3	NOTCH2	OSTC	UBTD1	LATS2	
SEC24A	SNX33	P4HA1	RAC2	ALPK1	ATP6V0E1	
FAM114A2	DHRS7B	PBX2	RAP1A	RUFY1	CD44	
TUBGCP2	IBTK	ASCC1	RELB	CCDC6	B4GALT7	
PNPLA6	HERC4	EXOSC1	REST	NCOA4	KIAA1033	
LMAN2	TCTN3	MRPS16	RGS10	NDEL1	SMARCAL1	
RER1	GBE1	LAP3	RNH1	TRIM8	MITD1	
SEC23IP	GBP3	MECR	SPCS3	SLC25A28	STAM2	
CMTM7	AMFR	PHF11	RPS14	ACOX3	ADCY7	
CHP1	GHITM	NAGPA	RSU1	ADPGK	PQLC3	
ECD	PRELID1	SHISA5	ATXN1	SH3BGRL3	NPHP3	
SFT2D1	GLB1	ZBTB7A	TMBIM1	BCL2L12	GADD45B	
CLIC1	APOBEC3C	TAOK3	MAP2K4	TMEM120A	DERL2	
WHAMM	MAT2B	DDX41	MMS19	CASP7	ERBB2IP	
JOSD2	GLRX	HDAC7	PDLIM2	UTP15	MED7	

Supplemental table 7

Supplemental Table -- 7

Cluster 5 (199 genes)			
ADAM22	KCNJ3	RNF165	DISP2
HEY2	KCNJ4	OPRK1	BRINP1
GDAP1L1	KCNQ2	ATP6V1G2	GRIA2
FSD1	LMO1	GPR88	TPPP
GALNT16	MYH8	SNTG2	COLCA1
DDX24	NEFM	SLC6A15	PODXL2
BRSK2	NEFL	PPP2R2B	MAGI2
ATP2B2	NEUROD2	LRRC7	REEP1
BARHL1	ATP1A3	NGB	MYT1L
NEUROD4	NHLH2	CDH22	NRXN2
SCN4B	PDE2A	SLC6A3	FLRT1
SLN	PGF	MLLT4-AS1	BHLHE22
SNCB	POU3F1	CACNA2D1	INHBE
TP73	SUSD4	MAP6D1	UNC5A
CAMKV	ENOX1	PDRG1	CALM1
KIRREL2	FEZF2	RPS6KL1	TSHR
DOC2A	RELN	CBLN1	LPL
NTNG2	TMEM63C	GNG8	EYA2
SHF	PTCHD2	RAPGEF5	SEMA6A
USP2	PPP4R4	STMN2	
RASL10B	RAB3A	MRAP2	
CACNA2D2	C14ORF93	TUBA3FP	
TTYH2	NEUROD6	CYYR1	
TMCC2	CDH24	UGT3A1	
OLFM1	SLC26A10	CRABP1	
KPTN	SLC8A2	TCP10L	
CHRNA1	SSTR2	ELAVL2	
CHRNA3	NKAIN1	APLP1	
ADCYAP1R1	RND2	KLC1	
FOXN4	SYN3	EML6	
TMEM132E	FAM57B	PDE1A	
ANKRD13B	STON2	LPPR1	
RIMS4	EMILIN3	PRPH	
C16ORF92	GABBR2	SLC17A6	
SYT6	KIAA0226	HRASLS	
CLVS1	PPP1R17	SLC7A14	
OTUD7A	C1QL1	CLSTN2	
DRD2	PHF21B	NDRG4	
DAND5	GPRIN1	C2ORF40	
EPHA8	OLIG1	REC8	
FAM181B	CNTN1	RAMP2	
ARC	SYNPR	CCM2L	
WSCD1	CHODL	COCH	
ACSL6	CTNND2	AMPH	
MAPK8IP2	IGSF11	KCNB1	
GABRA1	DACH1	C14ORF23	
GABRB2	KANK4	LY6H	
KIF26A	CABP7	CDKAL1	
RCOR2	DTX3	C14ORF132	
ZDHHC22	MAST1	RGS16	
NPW	SYNGR4	PPFIA2	
GRIA4	SULT4A1	LIN7A	
GRIK5	GAP43	MAPK8IP1	
GRM2	GAS2	CDH12	
HLF	LOC283731	CKB	
HPCA	HOXD3	ANO5	
SMTNL2	HOXD13	NMNAT2	
PAQR9	RPRML	TMEM145	
PLCXD3	NEUROD1	SIX3	
C1ORF95	NPTX2	DPF1	

Supplemental table 8

Supplemental Table -- 8

Cluster 6 (285 genes)				
ADAM8	B3GNT6	MST1R	TFAP2C	NAMPT
MMP13	EFNA1	MUC4	TGFA	AGR2
SERPINB5	EGFR	MYO10	THBD	NAPSA
ADAM15	SDR16C5	NFKBIA	TPBG	MERTK
SERPINA1	KRT78	SLC22A18AS	TPD52L2	DNAH5
SLFN13	AHR	P2RY2	PHLDA2	GPR116
SFN	FLJ23867	P2RY6	FAM160A1	CD274
SLCO4A1	LIPH	PAWR	UPP1	CEBPD
IL20RB	ERBB2	ACP6	EZR	DNTTIP1
OTUB2	EREG	STYXL1	VRK2	SEC61G
PTGES	EVPL	PEX13	WNT7B	ERGIC2
SPAG4	F2RL1	PLSCR1	MLPH	TCN2
PPP1R1C	F3	GPR87	C1ORF116	NOL3
KLC3	EFEMP1	GSAP	ZBED2	ELFN2
GPX3	GPR115	ERRFI1	CALB2	ACOT4
KRT4	KDM2A	HCG4	C3ORF52	CPM
ASPH	WWC1	ANLN	STEAP4	CTTN
BCL2L1	FOLR1	RIN2	MAFK	HDGF
CDH3	CD2AP	PON2	MICALL2	NSUN2
SH2D3A	RAB38	RBM47	TMC5	PON3
MPZL2	FRK	LY6K	ARHGEF5	RYR1
CNKSR1	ABCA4	RHBDL2	EPHX3	TAX1BP1
CCNO	EPHX4	MOCOS	TMC7	DUS4L
B3GNT3	ZNF718	MRGBP	TMEM156	LACTB2
SEMA3A	TES	TMEM144	FBXL18	STK31
NDRG1	STEAP2	FERMT1	ARL14	HOPX
BAIAP2	NKX2-8	FLVCR2	HSD3B7	DNAH11
SEMA4B	PLEK2	FGD6	FER1L4	HSPA6
GIPC1	MYEOV	MAP2K1	CCDC68	SLAMF7
PPP1R13L	GPR110	GSDMC	DNAJC5	ARHGAP5-AS1
SERINC3	STEAP1	CCL28	SHARPIN	SLC34A2
HIBADH	TNFRSF21	PSG3	CAPN2	VSTM2L
PLA2G16	WFDC10B	SLC2A4RG	PITPNM3	AVL9
PKP3	KCTD21	PMEPA1	ITCH	C1GALT1
GALNT6	ZNF707	CEACAM19	SYT16	DNAH2
EXOC3	GPR39	CCDC47	SLC41A2	C9ORF84
DTX2	TMPRSS11E	PTGS2	C15ORF48	CCT5
TLCD1	CXCL2	PTHLH	AGPAT9	TIPARP
CATSPER1	RHOD	PTK6	ALG10	CLDN12
GSTO2	ERO1L	RAB27B	FAM83A	TP53I3
ERP27	HIST1H2BD	NTN4	PPAP2C	TRIO
CANT1	ANXA3	SAV1	VAMP8	SLC35F3
USP43	APLP2	S100A2	SCEL	C1ORF27
MISP	AMIGO2	S100A13	TNFRSF10A	TOP1
AP1S3	IL8	SDC1	NRP1	DHRS3
CLDN23	ITGA2	SDC4	SYS1	
CRABP2	EIF6	SECTM1	TMEM41A	
WFDC3	ITGB6	PRSS22	ZNF622	
KRT80	KRT7	C19ORF33	ZFAND2A	
C16ORF89	SFTA2	TNS3	TRIM47	
CST6	LAMA3	EPS8L2	SYT12	
ATP8B3	LAMA5	FAM129B	OSMR	
LOC152225	LAMB3	CDCP1	SCGB3A1	
CYP1B1	LAMC2	SLC2A1	DMKN	
CD55	LRP5	SLCO2A1	KLF4	
DAP	LTBP3	STAT4	MUC16	
TMEM92	TACSTD2	STK3	CDA	
DSC2	CD46	SVIL	GOLGA5	
DSG2	MET	BTC	TNFSF15	
DVL1	MGST1	KLF5	FGFBP1	

Supplemental table 9

Supplemental Table -- 9

Cluster 7 (354 genes)					
OSBPL6	KCTD7	SGCB	MTFR1L	EFHC1	TPTE
ZNF829	ZNF610	TAF11	PHTF2	SLC10A4	ZNF426
ZNF568	ZNF320	SUMO1	ZNF71	VASH1	PPIE
TUBA1A	ZNF780B	ZNF14	RLF	WDR47	ZNF572
AKIRIN1	ZNF100	ZNF26	SNX16	MTOR	PHF13
WASF1	ZNF675	ZNF43	TUBB2A	NEGR1	C1ORF213
AGPAT4	AGO3	ZNF45	ZFP37	GPR137C	ZIC2
SALL1	ZNF585A	ZNF135	ZNF10	HCFC2	GTF2H5
NKX3-2	EML1	ZSCAN9	ZNF84	WDR19	WNK1
ZNF471	C2ORF69	ZNF227	ZNF184	SCN8A	MSANTD4
ABI2	ERCC3	ZNF230	IFT74	EXO5	PAQR3
C1ORF216	MED19	C1ORF50	KIAA1841	DDHD1	KIFAP3
MACF1	FKBP1B	ZYG11B	MAP1LC3A	HS2ST1	ANKRD45
SCCPDH	DZIP1	SNIP1	C16ORF45	ULK2	ZNF222
ADPRHL2	PHLPP2	ZNF430	ZMYM4	KIAA0355	KATNAL1
ZNF416	MAST2	PBX4	COQ10A	CCDC104	STRADA
DMAP1	ZCCHC11	PRR3	KDM4A	CSMD2	LPIN2
ELMO2	ZNF345	ZNF611	CDKN2C	FAM161B	ZNF547
KIAA0754	KIAA1429	ZNF512	SLMO1	TRIM37	RSPH4A
ZNF91	HINFP	ZNF527	SF3A3	FMN2	ZNF85
ZNF655	STK36	ZNF594	TBCB	RRAGC	FAM218A
FAM118B	C11ORF31	ZNF607	SSX2IP	CASD1	BMPR2
ZNF329	ZNF615	SYNJ1	ZNF684	LOC728392	E2F3
CCDC15	ZFP82	C12ORF65	OSCP1	C2ORF44	MTF2
CSRNP2	HKR1	ZNF235	ZNF565	CCDC3	ANKRD12
ZNF93	MICU3	ZNF254	KANSL1L	ZNF439	ZNF549
WDR54	GNL2	DZIP3	ZFP1	DGKI	FAM66C
ZNF528	DNAJA1	CKAP5	FAM171B	LINC00662	SEC61A2
ZNF559	ZNF680	RAMP2-AS1	ADAMTS18	CCDC23	ZKSCAN7
PGBD1	IPP	MAD2L2	DLX1	PLEKHO1	CACHD1
ZNF382	KIF3C	PHTF1	DLX2	CNTLN	CENPBD1P1
ZNF566	KPNA5	KIF3A	EXTL2	SH3BGR	ZNF134
ZNF251	SKIDA1	EID2B	KIAA1009	ZNF436	ZNF605
ZNF682	STMN1	TTC7B	USP33	PER3	ZNF417
TMEM67	KMT2A	TTL	SATB2	HMGN3	CCDC92
UBTD2	NFYC	CCDC112	ZDHHC17	ACVR2B-AS1	PPP1R21
TMEM44	PBX3	ZNF540	CNRIP1	CORO2B	LYRM2
SLIT2	IFT52	LCA5	ZNF493	EID2	SOS2
ORMDL1	PDE6D	THAP8	GPX7	FAM219A	EPM2AIP1
MED17	HSPB11	A1BG	IMPA1	FOXE3	ZNF625
APPBP2	SPATA6	CENPV	ARL3	PLCB4	ZNF347
ZNF211	ZNF562	DNAJC18	MTF1	MED29	ZFR2
KHDRBS1	TRNAU1AP	STX2	NDUFS5	ANKRD7	FAM89B
ZNF273	TMEM39B	RBM24	ZDHHC2	ZNF136	ZNF667
BTG3	CEP192	ZFP30	CUTA	BBS10	SYNRG
PDAP1	CDCA8	CAND2	TM6SF1	MPDZ	ZNF709
LRRC37B	LRRC40	TTC28	DCAF16	EXOC5	BEND7
SPATA33	CCDC88A	SYCE2	PTBP2	CSRP2	GPR161
ZNF573	TENM3	DENND2A	SMAP1	ZNF260	MBD5
ZFP28	SAYSD1	DPY19L2	RPA2	SERP2	ZNF253
C11ORF84	PRKAR2B	FBXO43	ZFP69B	PDE6B	RBPJ
ZFP90	KDM3A	THYN1	CCDC30	PRMT6	ZNF83
ZNF420	MAPK7	ZFP69	ZNF708	FAM196A	PHACTR4
ZNF583	LRRN1	ZNF793	ALG9	S100PBP	GPATCH2L
ZNF738	KIAA1586	TAS2R14	SLC25A33	SPOP	
ZNF681	ZNF529	ZNF571	HSD17B6	LOC100272217	
ZNF569	RFX4	ZCCHC17	AKT3	KIF1B	
ZNF570	FAM229B	AHI1	CDH2	UTP11L	
C1ORF52	CLSPN	PNMAL1	CHN1	PRKD1	
ZNF362	PCNXL4	LRRC36	DCTN3	SHOX2	

Supplemental table 10

Supplemental table -- 10

Cluster 8 (284 genes)			
MMP26	VWA5B2	SCGN	KCNH2
KLK12	SYT7	SIX2	NEB
KLK11	IFT140	PTP4A3	LMO3
SYT13	MTSS1	ADCYAP1	KCTD16
TMEM176B	KLHL41	TSHZ2	KLHDC9
GRIN2C	NPM2	CNTNAP5	RIBC2
CBLN2	CACFD1	FBLN7	LINC00574
FAM134B	MGAT4A	C2ORF15	UNC13B
CGA	FAM83F	RASSF6	LOC283070
OR51E1	CCDC151	SLC29A4	JPH1
C8ORF22	C1ORF194	MCF2L	RALGAPA2
CALY	SPATA17	CADM2	CCKBR
RET	ACVR1C	QPCT	CPEB3
RIT2	PPP1R36	GLS2	SDK2
VGF	CNKSR3	SLCO3A1	ZSWIM5
CALCA	DDC	RIMKLA	TRPM8
OR51E2	TMEM61	HOXD1	RAB3IP
SCIN	KHDRBS2	HOXD9	TMEM230
SV2B	C8ORF47	ICA1	SHISA2
CPLX2	AMER3	IGFBP5	ANKH
POU6F2	ABCA3	NKX2-2	LRRC10B
SGSM1	PLCB1	NPPA	GCH1
C9ORF135	MORN5	NPTX1	C2ORF70
C2CD4A	LINC00957	ATP6V0B	CXXC4
LOC145837	SERGEF	SMPD3	
SLC38A11	SEZ6L2	PKIB	
C8ORF56	RGS17	ERO1LB	
COL22A1	GPX2	CAMK1D	
ZBTB7C	TFCP2L1	MKL2	
GABRG2	HABP2	PTPRN2	
GALNT8	HOXB5	PLEKHB1	
IGSF22	C19ORF45	PRR15L	
FOXA2	KCNJ6	MS4A8	
GLDN	LFNG	ESPN	
BMP8A	FAM174B	ABLIM2	
KCNF1	11-Mar	SEC11C	
NTHL1	TSPAN11	ADAMTSL2	
NTS	HMP19	DNAJC6	
PAH	FAM3B	HCN4	
KCNK10	PON1	CHN2	
MOV10L1	LGI2	PIFO	
RNF186	KIAA1244	PRUNE2	
TMEM176A	KIAA1324	CCDC67	
PRMT8	RAB3B	EXTL3	
DNAJC12	SCN2A	ZNF396	
PTPRN	SMOC2	FRMD3	
RGS7	HS3ST6	TOX3	
NDST4	SMYD3	PRR18	
SLC18A1	GAREM	GRP	
SLCO1A2	SPTB	SLC35D3	
AACS	ABCC8	INPPL1	
SST	SCG2	CCDC178	
TFF3	NARS2	SCN3A	
CLDN5	TMEM163	BMP8B	
CA8	RTBDN	CRISP2	
CACNA1D	C8ORF12	C11ORF49	
TRAPPC9	CADPS2	BAALC	
NCALD	LZTS3	DYDC2	
NR0B2	KIAA0087	PROM1	
B3GALT2	CDH7	TDH	

Supplemental table 11

Supplemental table -- 11

Supplemental Tables 12 and 13. Gene ontology (GO) results of biological processes (BP) in cluster 3. The table lists the GO term describing the BP involved, the percentage of genes present in the cluster and linked to a BP compared to all genes that are linked to the same BP, p-value, list of genes and the fold-enrichment of each biological process.

Term	Count	% within BP	P-Value	Genes	List Total	Fold Enrichment
GO:0007165~signal transduction	46	11.00478469	3.83E-04	CXCL1, S100A6, SIPA1, PPARG, IL15, ELK3, GPRC5A, PXN, IQGAP1, TNFRSF1A, RALB, SHC1, AGRN, TRIP10, RASA1, CSNK1A1, RAP2B, LTBR, EPAS1, LIMK1, SPHK1, ANXA1, S100A11, FADD, PRKCE, HMGA2, FLNB, ANXA4, PLAUR, TRADD, VEGFC, NPC1, TNFRSF10D, BRE, HBEGF, RIN1, INPP4B, EXT1, ARAP3, EEF1D, FEZ2, PLAU, NEK6, IGFBP4, BCAR3, ARHGAP10	386	1.723616857
GO:0045944~positive regulation of Tc from RNA polII promoter	33	7.894736842	0.028274536	FOSL2, HELZ2, IL18, MITF, EDN1, HEXB, PPARG, FSTL3, ELK3, TCF7L2, ZBTB38, LIF, TNFRSF1A, NPAS2, SQSTM1, BCL3, ZC3H12A, BCL9L, AGRN, YAP1, FOSL1, SERTAD1, CYR61, RARG, EPAS1, RELA, SMAD3, FADD, WWTR1, HMGA2, ADRB2, RPS6KA4, FHOD1	386	1.4633899
GO:0098609~cell-cell adhesion	27	6.459330144	7.82E-10	CAST, PDLIM5, ARHGAP18, TAGLN2, ESYT2, GPRC5A, IQGAP1, PKM, CDC42EP1, SLK, BAG3, AHNAK, EHD4, PPFIBP1, S100A11, PFKP, FLNB, VASP, EPHA2, ANXA2, DHX29, LASP1, PRDX6, RARS, CAPG, TMOD3, EEF1D	386	4.334206451
GO:0043066~negative regulation of apoptotic process	26	6.220095694	5.38E-05	IER3, IL6ST, YBX3, FHL2, SQSTM1, BAG3, TGM2, BCL3, THBS1, ANGPTL4, CYR61, RARG, RELA, TMBIM6, SPHK1, ANXA1, SMAD3, BIRC3, HMGA2, ANXA4, PLAUR, DUSP1, PLK2, TNFRSF10D, PRNP, ARHGAP10	386	2.485862324
GO:0043547~positive regulation of GTPase activity	24	5.741626794	0.005845742	CAV2, RALGPS2, PDGFA, SIPA1, S100A10, ARHGAP18, ARHGAP29, DOCK5, IQGAP1, DNMBP, CDC42EP2, RGS20, CDC42EP1, RIN1, HBEGF, JAK1, SHC1, AGRN, ARAP3, TRIP10, RASA1, BCAR3, TBC1D2, ARHGAP10	386	1.847897657
GO:0006915~apoptotic process	21	5.023923445	0.038180105	BCL10, IER3, LTBR, FADD, SGMS1, STK17A, BIRC3, PRKCE, CARD6, TRADD, CASP4, SLK, SQSTM1, BCAP29, BRE, RALB, ZC3H12A, CTSC, IGFBP3, NEK6, PHLDA1	386	1.611207062
GO:0006954~inflammatory response	19	4.545454545	0.003026945	CXCL1, IRAK2, LTBR, RELA, IL18, SPHK1, ANXA1, SGMS1, IL15, EPHA2, TNFRSF1A, HRH1, CASP4, RPS6KA4, TNFRSF10D, IL1RAP, ZC3H12A, THBS1, IGFBP4	386	2.180868662
GO:0008284~positive regulation of cell proliferation	19	4.545454545	0.022164336	TCIRG1, RARG, IL6ST, PDGFA, RELA, EDN1, IL15, WWTR1, LIF, VEGFC, CTGF, CLCF1, HBEGF, SHC1, YAP1, THBS1, SLC35F6, FOSL1, SERTAD1	386	1.773710779
GO:0007155~cell adhesion	18	4.306220096	0.036081531	B4GALT1, LPP, PPFIBP1, ITGB5, ITGA3, PRKCE, CTNNA1, CD151, EPHB4, PXN, CD9, LAMB2, CTGF, CD58, TGFBI, THBS1, CYR61, ADAM9	386	1.705983948
GO:0035556~intracellular signal transduction	18	4.306220096	0.011981542	CXCL1, SPSB1, SIPA1, SPHK1, EDN1, ARHGAP29, STK17A, PRKCE, DNMBP, RPS6KA4, DUSP1, SQSTM1, CTGF, HMOX1, PLCD3, JAK1, SHC1, RASA1	386	1.943043752
GO:0001525~angiogenesis	17	4.066985646	5.88E-05	CAV1, EPAS1, TNFRSF12A, PDGFA, IL18, ELK3, EPHB4, RNF213, ANXA2, VEGFC, CTGF, HMOX1, TGFBI, PLCD3, ZC3H12A, SHC1, ANGPTL4	386	3.316340993
GO:0043123~positive regulation of I-kB/NF-kB signaling	15	3.588516746	2.09E-05	BCL10, LTBR, RELA, FADD, BIRC3, PRKCE, TRADD, AJUBA, TNFRSF1A, PLK2, HMOX1, TGM2, RHOC, EEF1D, NEK6	386	4.053036398
GO:0042981~regulation of apoptotic process	14	3.349282297	0.001309068	BCL10, ACTN4, MITF, FADD, BIRC3, CARD6, CARD10, TNFRSF1A, CASP4, DUSP1, SLK, TNFRSF10D, BCL3, IGFBP3	386	2.859325209
GO:0043065~positive regulation of apoptotic process	14	3.349282297	0.021456971	RARG, TNFRSF12A, FADD, STK17A, HMGA2, TRADD, DUSP1, SQSTM1, ALDH1A3, HMOX1, TGM2, BIN1, IGFBP3, FOSL1	386	2.030120898
GO:0010628~positive regulation of gene expression	13	3.110047847	0.018137859	CAV1, WNT3, CTGF, TRIM6, MITF, HFE, SMAD3, MAPK9, ZC3H12A, ITGA3, RNF207, SGMS1, HMGA2	386	2.158525491
GO:0001666~response to hypoxia	12	2.870813397	0.002094363	AJUBA, PKM, VEGFC, CAV1, ACTN4, EPAS1, HMOX1, SMAD3, THBS1, PLAU, AGTRAP, ANGPTL4	386	3.035064466
GO:0030335~positive regulation of cell migration	12	2.870813397	0.003528156	MYO1C, ACTN4, PDGFA, EDN1, SPHK1, SMAD3, HBEGF, SEMA3C, THBS1, MYADM, PLAU, CYR61	386	2.837125479
GO:0051260~protein homooligomerization	12	2.870813397	0.002619378	STOM, BCL10, CAV1, HMOX1, ATL3, CLDN1, TGM2, ATG16L1, PRNP, RNF213, EHD4, ANGPTL4	386	2.949328181
GO:0030198~extracellular matrix organization	10	2.392344498	0.037318223	B4GALT1, LAMB2, PDGFA, TGFBI, BCL3, ITGB5, ITGA3, AGRN, THBS1, CYR61	386	2.219519932

Supplemental table -- 12

Term	Count	% within BP	P-Value	Genes	List Total	Fold Enrichment
GO:0008360~regulation of cell shape	10	2.392344498	0.005012067	CSNK1A1, ANXA7, CDC42EP2, CDC42EP1, HEXB, ANXA1, ARHGAP18, MYL12B, ARAP3, RASA1	386	3.107327905
GO:0006897~endocytosis	10	2.392344498	0.004783733	NPC1, SNX6, PACSIN3, SNX8, MYO1E, RIN1, TRIP10, ESYT2, BIN1, EHD2	386	3.129682782
GO:0030036~actin cytoskeleton organization	9	2.153110048	0.010223445	CXCL1, INF2, CDC42EP2, PDGFA, LIMK1, TMOD3, TMSB10, TRIP10, FLNB	386	3.011717816
GO:0005975~carbohydrate metabolic process	9	2.153110048	0.0475414	MGAT4B, B4GALT1, GALM, RPE, UEVLD, PYGL, HEXB, CHST3, FUCA2	386	2.250134
GO:0006888~ER to Golgi vesicle-mediated transport	9	2.153110048	0.031356128	SEC23A, TMED7, CTSZ, MCFD2, ATL3, BET1, BCAP29, CTSC, TMED9	386	2.447020725
GO:0051592~response to calcium ion	8	1.913875598	3.47E-04	ANXA7, CAV1, CCND1, S100A16, DUSP1, NEDD4, THBS1, ADAM9	386	6.000357334
GO:0051291~protein heterooligomerization	8	1.913875598	8.40E-04	BCL10, PCBD2, SQSTM1, CLDN1, FADD, BIRC3, CTNNA1, TRADD	386	5.194339185
GO:0010951~negative regulation of endopeptidase activity	8	1.913875598	0.021474902	CAST, SERPINB6, CD109, TFPI, CSTB, SERPINH1, TFPI2, CRIM1	386	2.876204342
GO:0051092~positive regulation of NF-kB TF activity	8	1.913875598	0.033656441	IRAK2, BCL10, RPS6KA4, RELA, IL1RAP, SPHK1, TRIM14, TRADD	386	2.616697183
GO:0051056~regulation of small GTPase signal transduction	8	1.913875598	0.034842855	SIPA1, ARHGAP18, RHOC, ARHGAP29, ARAP3, TRIP10, RHOF, ARHGAP10	386	2.597169592
GO:0030512~negative regulation of TGFBR signaling pathway	7	1.674641148	0.003417623	CAV2, CAV1, SNX6, CD109, SMAD3, BCL9L, SMURF1	386	4.758095855
GO:0001649~osteoblast differentiation	7	1.674641148	0.032493223	DNAJC13, BCAP29, FHL2, WWTR1, IGFBP3, EPHA2, CYR61	386	2.928058988
GO:0006936~muscle contraction	7	1.674641148	0.03659252	DYSF, TMOD3, ITGB5, MYL12B, STBD1, MYOF, PXN	386	2.845963876
GO:2001237~negative regulation of extrinsic apoptotic signaling	7	1.674641148	2.07E-04	LGALS3, RELA, LMNA, YAP1, SGMS1, THBS1, TCF7L2	386	3.013635124
GO:0043410~positive regulation of MAPK cascade	7	1.674641148	0.010710167	LIF, CAV2, ADRB2, PDGFA, PRKCE, IGFBP3, IGFBP4	386	3.759483145
GO:0071356~cellular response to TNF	7	1.674641148	0.041009462	RELA, CD58, EDN1, YBX3, ZC3H12A, SGMS1, THBS1	386	2.768346679
GO:0045766~positive regulation of angiogenesis	7	1.674641148	0.049091369	VEGFC, HMOX1, SPHK1, RRAS, ZC3H12A, THBS1, ANGPTL4	386	2.64798378
GO:0043627~response to estrogen	6	1.435406699	0.016576588	ARPC1B, CAV1, CCND1, HMOX1, PPARG, CTNNA1	386	4.015623754
GO:0031398~positive regulation of protein ubiquitination	6	1.435406699	0.015585623	BCL10, CAV1, ADRB2, WFS1, SPHK1, BIRC3	386	4.078367876
GO:0060070~canonical Wnt signaling pathway	6	1.435406699	0.041930851	CCND1, WNT3, RARG, SMAD3, BCL9L, TCF7L2	386	3.144765591
GO:0001933~negative regulation of protein phosphorylation	6	1.435406699	0.012852042	CD109, ZC3H12A, WWTR1, PRNP, IGFBP3, MYADM	386	4.278943345
GO:0048661~positive regulation of smooth muscle cell proliferation	6	1.435406699	0.012018291	IL18, HMOX1, EDN1, TGM2, HBEGF, THBS1	386	4.350259067
GO:0007249~I-kappaB kinase/NF kappaB signaling	6	1.435406699	0.012018291	IRAK2, BCL10, TNFRSF1A, BCL3, BIRC3, TRADD	386	4.350259067

Supplemental figure 12 continued.

Term	Count	%	P-Value	Genes	List Total	Fold Enrichment
GO:0008625~extrinsic apoptotic signaling pathway via death domain receptors	6	1.435406699	0.001662449	TNFRSF1A, DDX47, TNFRSF10D, BAG3, FADD, TRADD	386	6.868830106
GO:2001238~positive regulation of extrinsic apoptotic signaling pathway	5	1.196172249	0.00272744	BCL10, CAV1, LTBR, TNFRSF12A, FADD	386	8.365882822
GO:0051259~protein oligomerization	5	1.196172249	0.043821456	BCL10, CAV2, ZC3H12A, ZNHIT6, AHNAK	386	3.750223334
GO:0008219~cell death	5	1.196172249	0.011886348	BCL10, RGS20, FOSL2, HMOX1, EMP1	386	5.577255215
GO:0032091~negative regulation of protein binding	5	1.196172249	0.04152084	CAV1, IFIT1, DKK1, TMBIM6, RALB	386	3.816016726
GO:0001570~vasculogenesis	5	1.196172249	0.039290532	CAV1, MYO1E, YAP1, RASA1, EPHA2	386	3.884159882
GO:0030838~positive regulation of actin filament polymerization	5	1.196172249	0.019352922	CDC42EP2, CDC42EP1, MYO1C, PRKCE, VASP	386	4.833621186
GO:0042542~response to hydrogen peroxide	5	1.196172249	0.029191474	DUSP1, HMOX1, KPNA4, FOSL1, ADAM9	386	4.26495987
GO:0055072~iron ion homeostasis	5	1.196172249	0.004652498	EPAS1, HMOX1, SFXN3, HFE, SLC25A37	386	7.250431779
GO:0042517~positive regulation of tyrosine phosphorylation of Stat3	5	1.196172249	0.010859193	LIF, CLCF1, IL6ST, IL18, IL15	386	5.724025089
GO:0051496~positive regulation of stressfiber assembly	5	1.196172249	0.015334055	LIMK1, CTGF, SMAD3, S100A10, FHOD1	386	5.178879842
GO:0032526~response to retinoic acid	5	1.196172249	0.01412267	MICB, RARG, DKK1, DUSP1, PPARG	386	5.305193985
GO:0048146~positive regulation of fibroblast proliferation	5	1.196172249	0.035040779	S100A6, FOSL2, PDGFA, SPHK1, ANXA2	386	4.028017655
GO:0090002~establishment of protein localization to plasma membrane	5	1.196172249	0.015334055	TNFRSF1A, MPP5, EFR3A, S100A10, TSPAN17	386	5.178879842
GO:0043433~negative regulation of sequence-specific DNA binding TF activity	5	1.196172249	0.048633325	WFS1, HMOX1, BHLHE40, PRNP, TCF7L2	386	3.625215889
GO:0043409~negative regulation of MAPK cascade	4	0.956937799	0.003606638	CAV1, DUSP3, DUSP1, RNF149	386	12.42931162
GO:0060546~negative regulation of necroptotic process	4	0.956937799	6.15E-04	CAV1, YBX3, FADD, BIRC3	386	21.75129534
GO:0016050~vesicle organization	4	0.956937799	0.025634194	CAV2, CAV1, SNX6, SNX8	386	6.214655811
GO:0000188~inactivation of MAPK activity	4	0.956937799	0.018924507	DUSP5, CAV1, DUSP3, DUSP1	386	6.960414508
GO:0048662~negative regulation of smooth muscle cell proliferation	4	0.956937799	0.0281183	HMOX1, PPARG, IL15, IGFBP3	386	6.000357334
GO:0002224~toll-like receptor signaling pathway	4	0.956937799	0.023273738	IRAK2, CTSL, BCL10, UNC93B1	386	6.444828248

Supplemental table -- 13

Term	Count	%	P-Value	Genes	List Total	Fold Enrichment
GO:0048711~positive regulation of astrocyte differentiation	4	0.956937799	0.001720603	LIF, CLCF1, IL6ST, BIN1	386	15.81912388
GO:0001974~blood vessel remodeling	4	0.956937799	0.036307139	LIF, EPAS1, TGM2, SEMA3C	386	5.437823834
GO:0006491~N-glycan processing	4	0.956937799	0.010203557	MGAT4B, MAN1A2, MAN1B1, EDEM2	386	8.700518135
GO:0072661~protein targeting to plasma membrane	4	0.956937799	0.021037168	PACS1, SMURF1, MYADM, ANXA2	386	6.692706257
GO:0031954~positive regulation of protein autophosphorylation	4	0.956937799	0.011704972	RAP2B, VEGFC, NBN, PDGFA	386	8.286207747
GO:0010803~regulation of TNF-mediated signaling pathway	4	0.956937799	0.030725611	TNFRSF1A, SPHK1, BIRC3, TRADD	386	5.800345423
GO:0090303~positive regulation of wound healing	3	0.717703349	0.039777285	ANXA1, HBEGF, PRKCE	386	9.321983716
GO:0006012~galactose metabolic process	3	0.717703349	0.016966637	B4GALT1, GALM, GALE	386	14.50086356
GO:2000675~negative regulation of type B pancreatic cell apoptosis	3	0.717703349	0.007398766	CAST, WFS1, TCF7L2	386	21.75129534
GO:0030857~negative regulation of epithelial cell differentiation	3	0.717703349	0.025149953	CAV1, CCND1, YAP1	386	11.86434291
GO:0071360~cellular response to exogenous dsRNA	3	0.717703349	0.03461005	CAV1, IFIT1, RALB	386	10.03905939
GO:0001706~endoderm formation	3	0.717703349	0.029728876	DUSP5, DKK1, DUSP1	386	10.87564767
GO:0001778~plasma membrane repair	3	0.717703349	0.016966637	DYSF, AHNAK2, MYOF	386	14.50086356
GO:0007589~body fluid secretion	3	0.717703349	0.016966637	EDN1, SLC22A4, ANXA2	386	14.50086356
GO:2000273~positive regulation of receptor activity	3	0.717703349	0.03461005	HFE, PRKCE, ANXA2	386	10.03905939
GO:0090136~epithelial cell-cell adhesion	3	0.717703349	0.016966637	ITGB5, CTNNA1, KIFC3	386	14.50086356
GO:0042511~positive regulation of tyrosine phosph of Stat1	3	0.717703349	0.025149953	LIF, TNFRSF1A, IL6ST	386	11.86434291
GO:0072307~regulation of metanephric nephron tubule cell differentiation	3	0.717703349	0.007398766	LIF, YAP1, WWTR1	386	21.75129534
GO:0051044~positive regulation of mmb protein ectodomain proteolysis	3	0.717703349	0.045214971	PACSIN3, SH3D19, ADAM9	386	8.700518135
GO:0097296~activation of cysteine-type endopeptidase activity	3	0.717703349	0.03461005	SMAD3, FADD, TRADD	386	10.03905939
GO:0071550~death-inducing signaling complex assembly	3	0.717703349	0.013397693	TNFRSF1A, FADD, TRADD	386	16.3134715
GO:0070106~interleukin-27-mediated signaling pathway	2	0.4784689	0.045330829	IL27RA, IL6ST	386	43.50259067
GO:0036509~trimming of terminal mannose on B branch	2	0.4784689	0.045330829	MAN1B1, EDEM2	386	43.50259067

supplemental table 13 continued.

Supplemental Table -- 14. Summary of the eleven patient cohorts used for survival analysis. The table lists the total number of patients and the number of adenocarcinoma and squamous cell carcinoma in each cohort. The overall survival times were directly downloaded from Kmplot.com.

Study	total number of patients in cohort	Adenocarcinoma	Squamous	Reference
TCGA	133	0	71	Cancer Genome Atlas Research Network. 2012
GSE50081	181	127	42	Der SD, et al. *2014*
GSE4573	130	0	130	Raponi M, et al. 2006
GSE37745	196	106	66	Bolting J et al. 2013
GSE31908	40	20	0	unpublished
GSE3141	111	58	53	Bild AH, et al. 2006
GSE31210	246	226	0	Okayama H, et al. 2012 Yamauchi M, et al. 2012
GSE30219	307	85	61	Rousseaux S, et al. 2013
GSE29013	55	30	25	Xie Y, et al. 2011
GSE19188	157	41	24	Hou J, et al. 2010
GSE14814	90	27	52	Zhu CQ, et al. 2010
Total	1646	720	524	

Supplemental Table -- 15. Pearson and Spearman correlation coefficients of cluster 50 PA genes in CCLE NSCLC cell lines and TCGA adenocarcinoma patient cohort. (a) Correlation analyses of gene z-scores for CCLE was performed using GraphPad. (b) Correlation analyses of TCGA adenocarcinoma provisional cohort (n=517) was performed on Cbioportal. The latter also calculates a logs odd ratio. P-values were adjusted to the Bonferroni-corrected threshold. Adjusted p-value is p-value/K = 0.005 where K=10 and represents the number of comparisons made (10 comparisons).

a

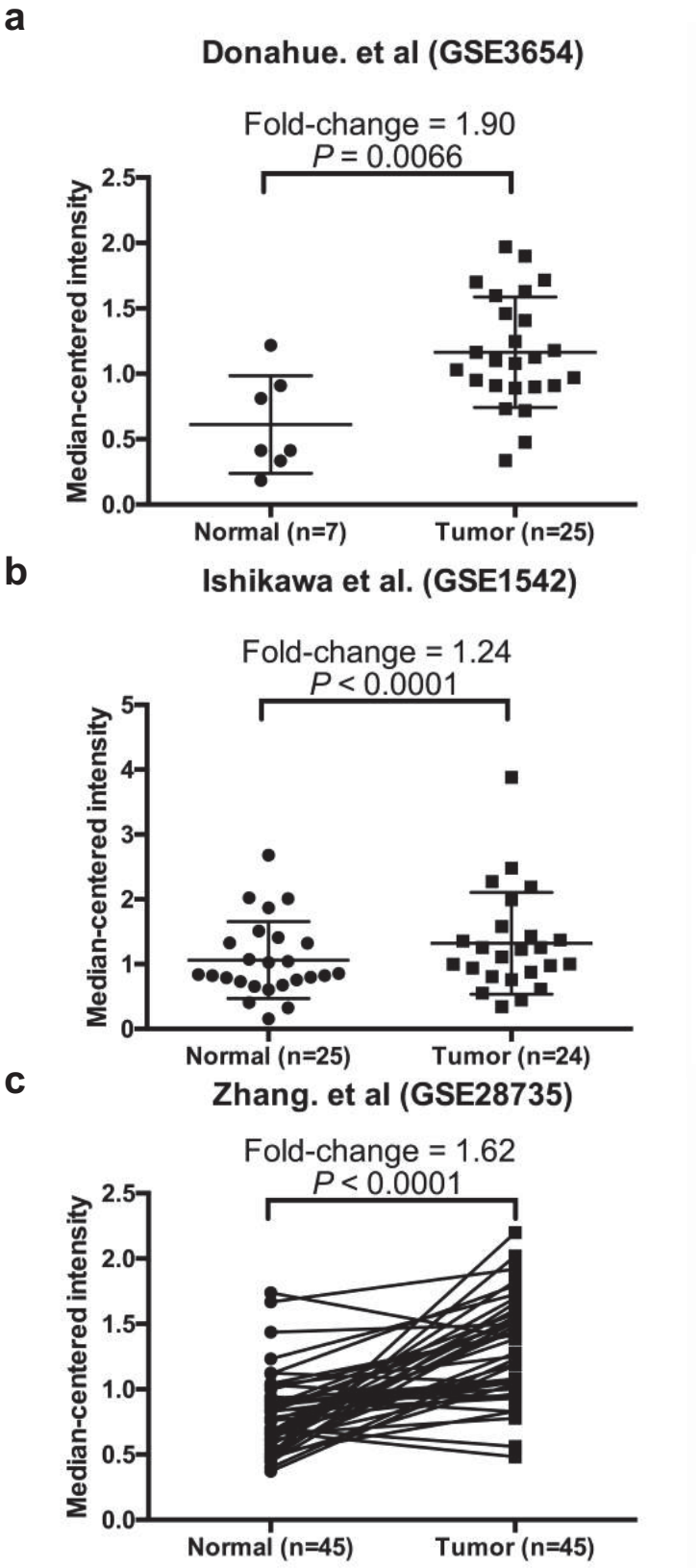
CCLE cell line discovery				
Gene A	Gene B	p-value	Pearson correlation coefficient	Spearman correlation coefficient
ANXA2	CTSL	4.56266E-05	0.3705988	0.1261895
ANXA2	CTSZ	5.80101E-07	0.4461888	0.2395644
SERPINB6	ANXA2	0.001749433	0.2887636	0.1233962
PLAU	ANXA2	1.01303E-06	0.4374958	0.4219372
PLAUR	CTSC	0.004944281	0.2604008	0.2104711
PLAUR	CTSZ	0.000105097	0.3537699	0.2153713
PLAUR	ANXA2	7.64965E-08	0.4758945	0.3535548
PLAUR	SERPINB6	0.000302974	0.33096	0.3427839
PLAUR	PLAU	1.04237E-12	0.6025786	0.5514096
S100A10	CTSL	1.63621E-06	0.4298222	0.2384439
S100A10	CTSZ	3.24005E-05	0.3772422	0.1437544
S100A10	ANXA2	2.58052E-20	0.7289339	0.4193719
S100A10	PLAU	0.000379338	0.3258825	0.1665315
S100A10	PLAUR	3.6341E-07	0.4532998	0.1263395
S100A10	CTSA	0.004606715	0.2624281	0.318062
SERPINH1	ANXA2	1.16028E-06	0.4353427	0.276951
SERPINH1	PLAU	0.00202308	0.2849679	0.2595247
SERPINH1	PLAUR	0.000209655	0.3390921	0.2975854
SERPINH1	S100A10	0.000578625	0.3160997	0.07758226

b

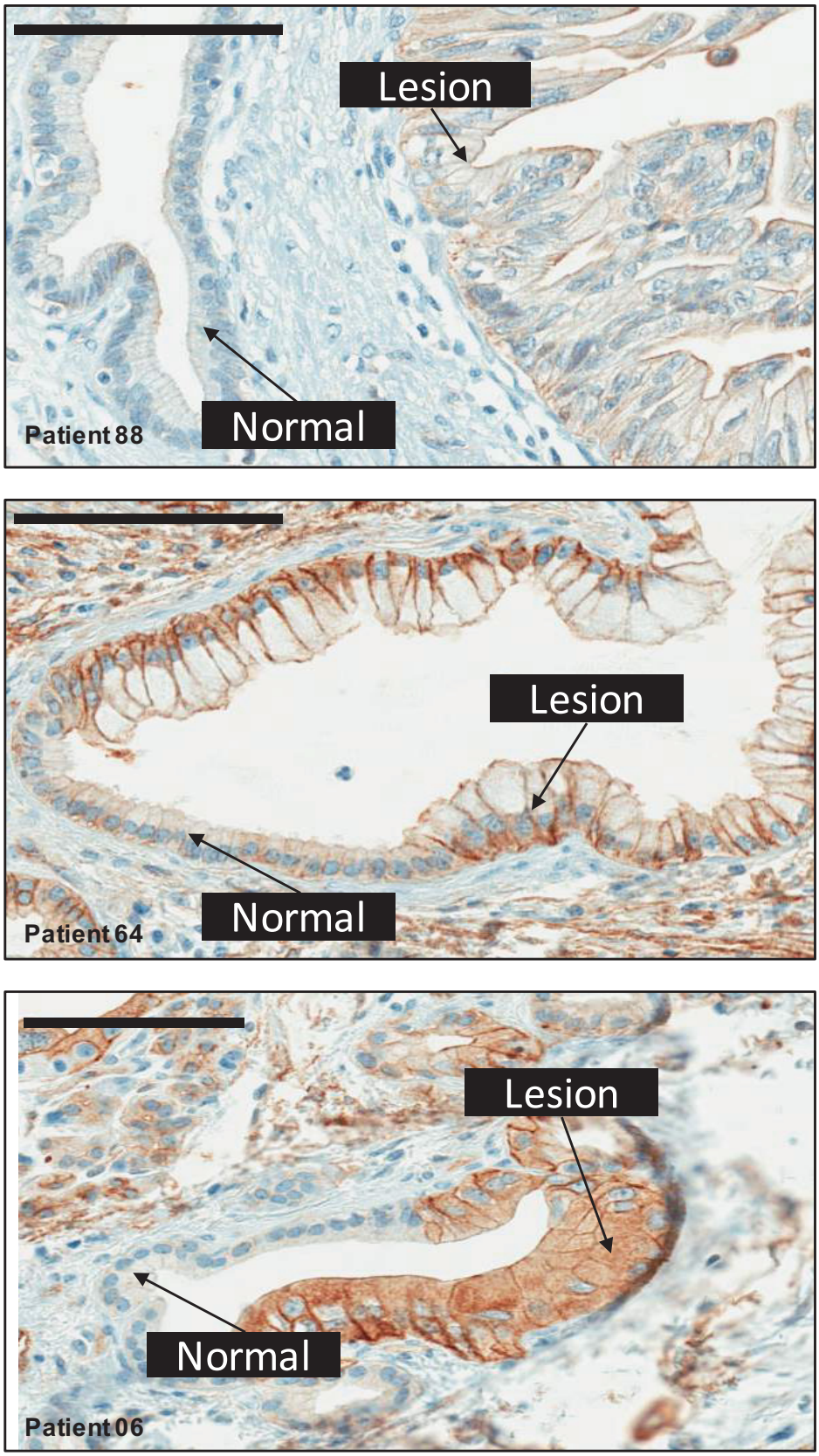
TCGA adenocarcinoma validation cohort (n=517)					
Gene A	Gene B	P-value	Log odds ratio	Pearson correlation coefficient	Spearman correlation coefficient
S100A10	ANXA2	< 0.001	2.593	0.482	0.576
S100A10	PLAU	< 0.001	2.124	0.362	0.438
PLAU	PLAUR	< 0.001	2.307	0.393	0.614
S100A10	PLAUR	0.002	1.625	0.384	0.483
ANXA2	PLAU	0.016	1.994	0.265	0.388
ANXA2	PLAUR	0.026	1.786	0.282	0.439
CTSC	CTSA	0.005	1.323	0.090	0.168
CTSC	S100A10	0.005	1.323	0.150	0.162
CTSC	ANXA2	0.012	1.773	0.231	0.228
ANXA2	SERPINB6	0.018	1.938	0.435	0.384
CTSL	S100A10	0.027	1.491	0.218	0.328

APPENDIX C: SUPPLEMENTAL FIGURES III

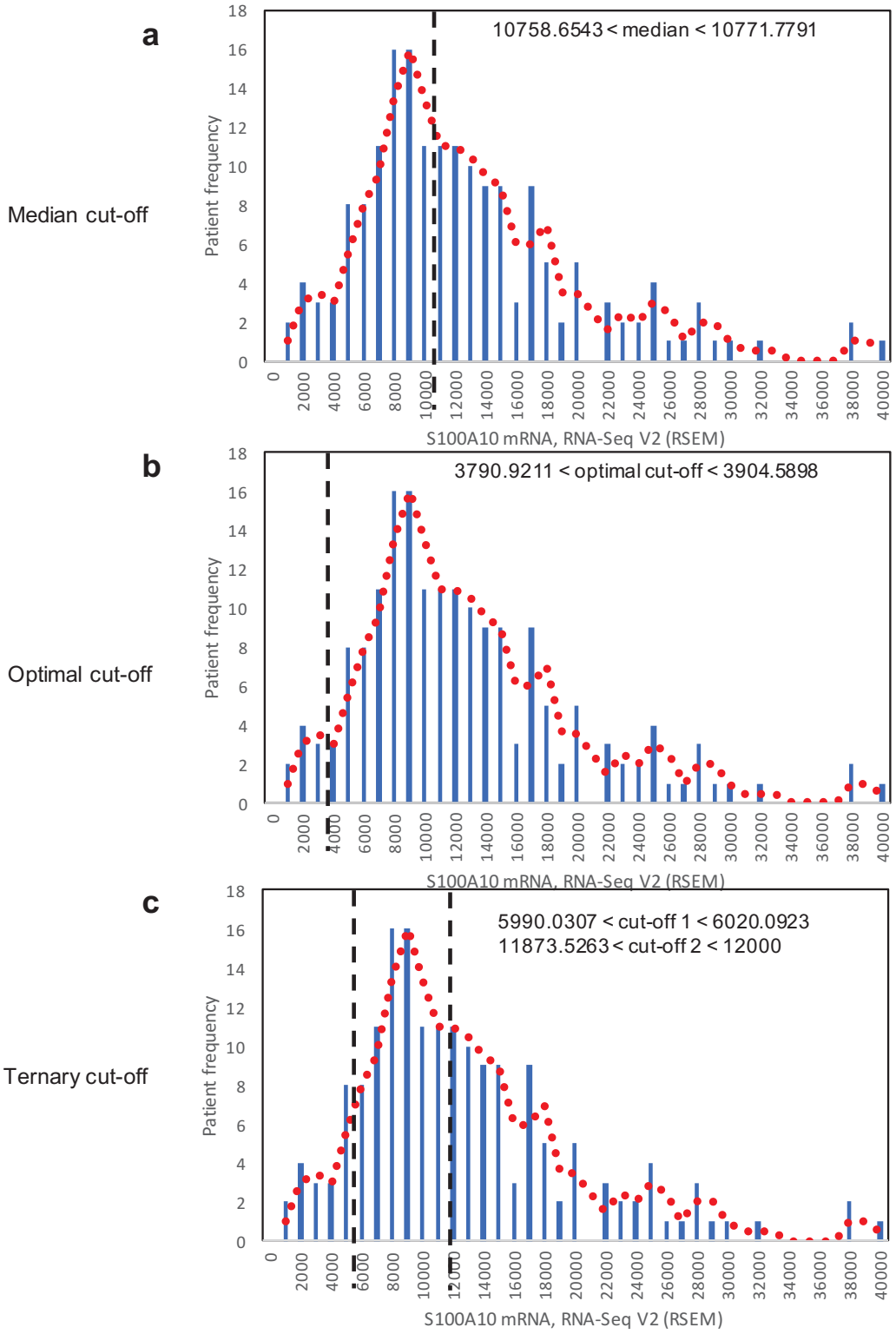
Supplemental Figure - 11. S100A10 mRNA is overexpressed in pancreatic tumors compared to normal pancreatic tissue. Gene expression from an additional three publically available gene expression datasets from Oncomine (a-c, e) extracted from the normalized data on Oncomine. The datasets compare gene expression in normal vs. tumor from pancreatic cancer patients. Zheng et al. represents matched samples of pancreatic tumors and corresponding adjacent normal tissue.


a
Donahue. et al (GSE3654)
Fold-change = 1.90
P = 0.0066
Median-centered intensity
2.5
2.0
1.5
1.0
0.5
0.0
Normal (n=7)
Tumor (n=25)
b
Ishikawa et al. (GSE1542)
Fold-change = 1.24
P < 0.0001
Median-centered intensity
5
4
3
2
1
0
Normal (n=25)
Tumor (n=24)
c
Zhang. et al (GSE28735)
Fold-change = 1.62
P < 0.0001
Median-centered intensity
2.5
2.0
1.5
1.0
0.5
0.0
Normal (n=45)
Tumor (n=45)

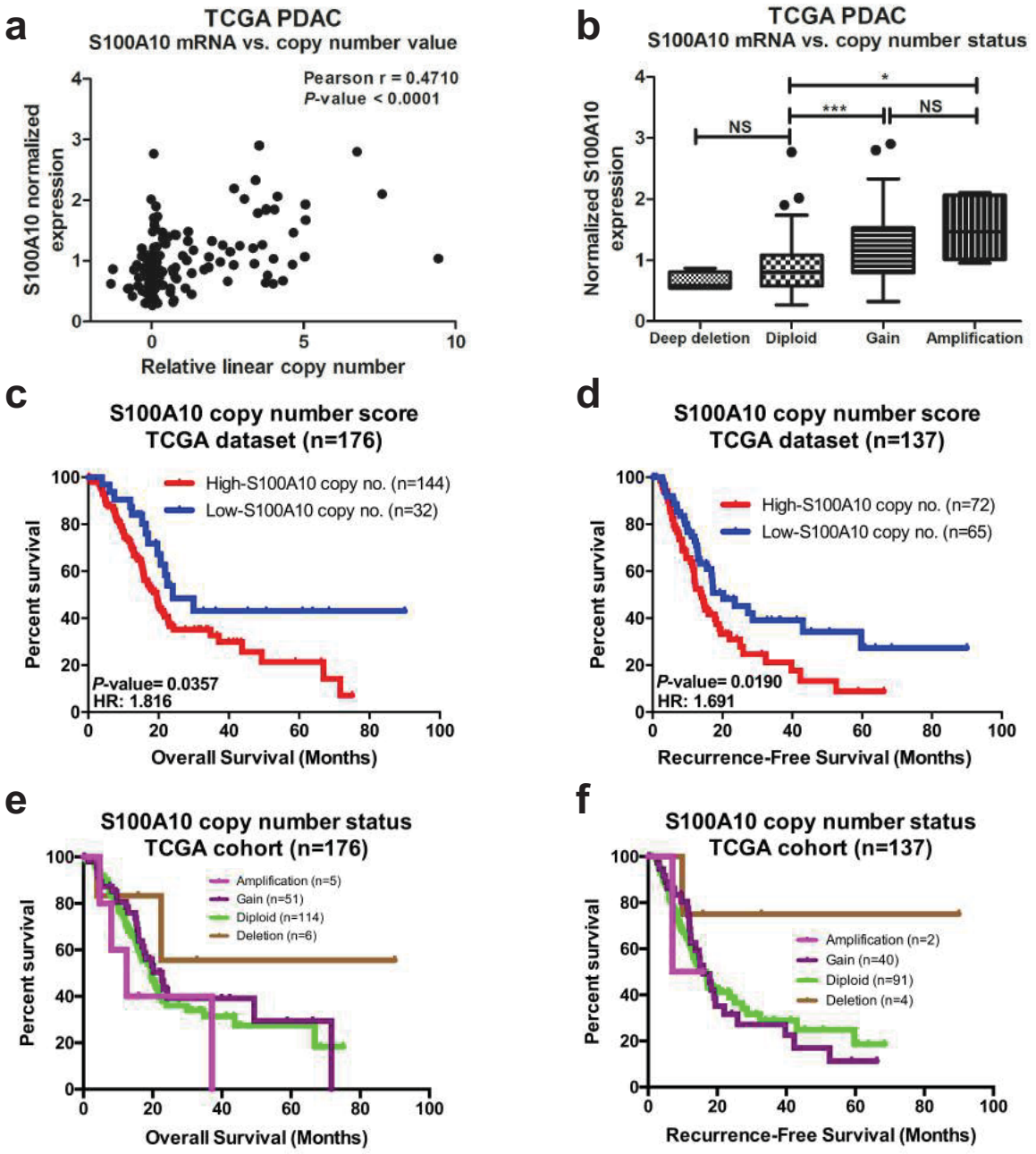
Supplemental Figure - 12. **Representative images of S100A10 staining in normal ducts and cancerous lesions.** Images represent three patient samples showing the upregulation of S100A10 (IHC) in tumor ducts/lesions compared to normal ducts. Scale bars, 100 μm.


Lesion
Normal
Patient 88
Lesion
Normal
Patient 64
Lesion
Normal
Patient 06

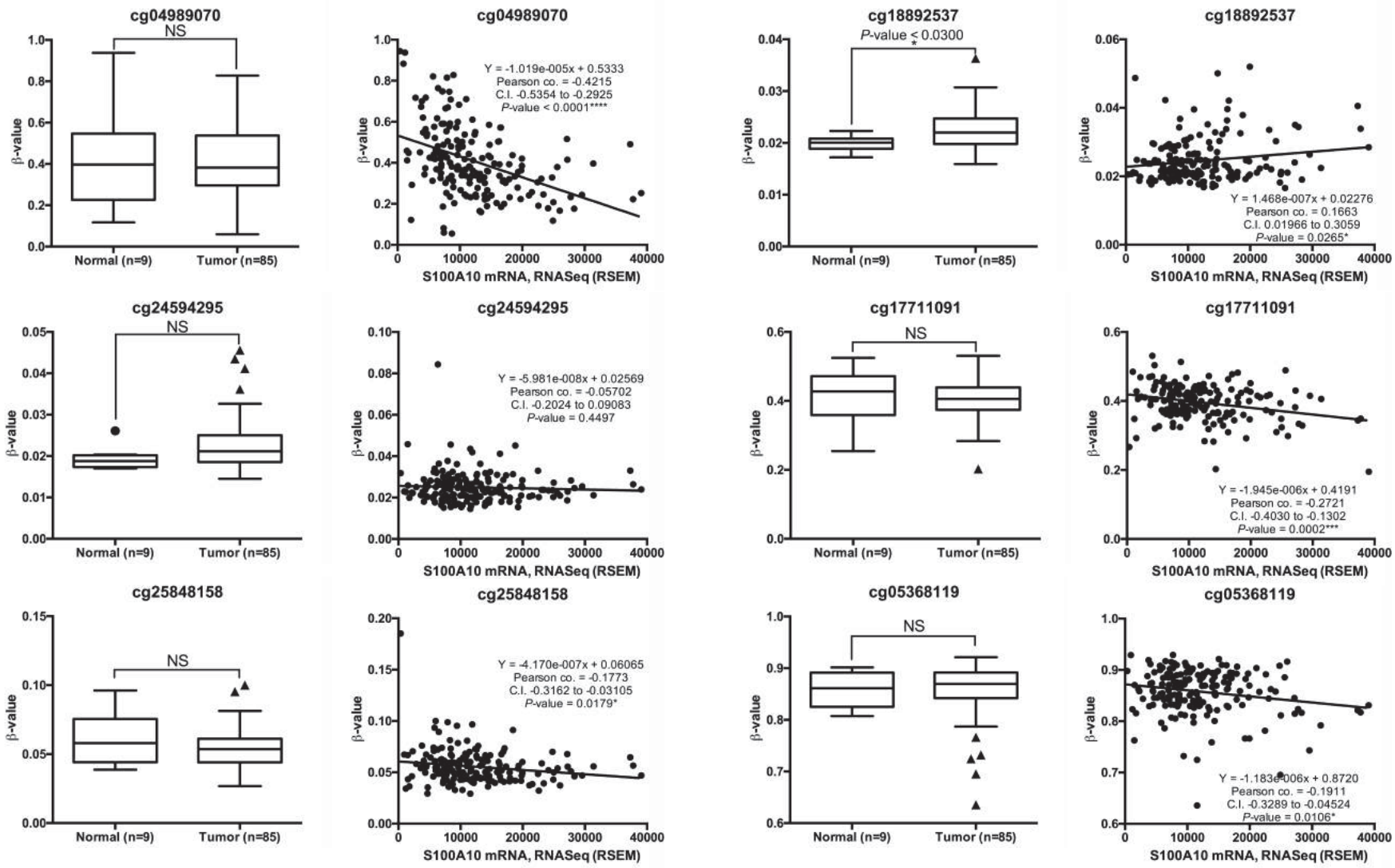
Supplemental Figure - 13. Identification of the three-tier cut-off system of S100A10 mRNA based on patient frequency. The three cut-off system is based on the median expression value (a), optimal expression value (b), or a ternary expression classifier (c). Optimal cut-offs were extrapolated from cut off finder (http://molpath.charite.de/cutoff/).


a
Median cut-off
10758.6543 < median < 10771.7791
Patient frequency
S100A10 mRNA, RNA-Seq V2 (RSEM)
b
Optimal cut-off
3790.9211 < optimal cut-off < 3904.5898
Patient frequency
S100A10 mRNA, RNA-Seq V2 (RSEM)
c
Ternary cut-off
5990.0307 < cut-off 1 < 6020.0923
11873.5263 < cut-off 2 < 12000
Patient frequency
S100A10 mRNA, RNA-Seq V2 (RSEM)

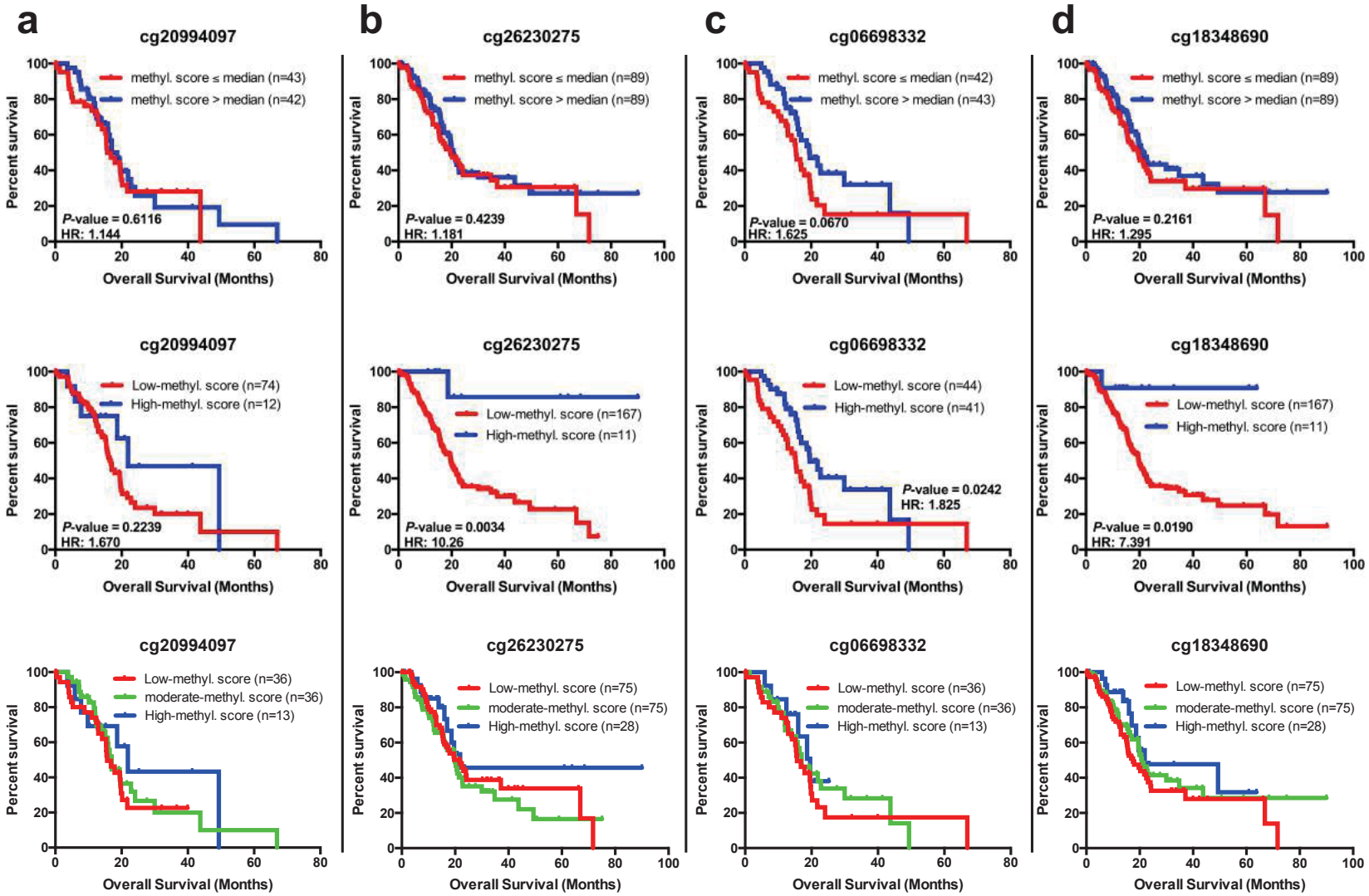
Supplemental Figure - 14. Correlation of S100A10 mRNA expression, linear copy number and copy number status with overall and recurrence-free survival. Pearson correlation analysis of S100A10 mRNA (expression values normalized to average) with (a) relative linear copy number and (b) copy number status. Kaplan Meier analysis of overall survival of TCGA PDAC patients in relation to S100A10 copy number score based on a optimal cut-off of (c) OS and (d) RFS. Kaplan Meier analysis of (e) OS and (f) RFS based on copy number status of S100A10. Gain and amplification are based on the Cbioportal definition where gain represents a low-level increase in copy number while amplification represents a high-level of increase.


a
TCGA PDAC
S100A10 mRNA vs. copy number value
Pearson r = 0.4710
P-value < 0.0001
S100A10 normalized expression
Relative linear copy number
b
TCGA PDAC
S100A10 mRNA vs. copy number status
Normalized S100A10 expression
NS
***
*
NS
Deep deletion
Diploid
Gain
Amplification
c
S100A10 copy number score
TCGA dataset (n=176)
High-S100A10 copy no. (n=144)
Low-S100A10 copy no. (n=32)
Percent survival
P-value= 0.0357
HR: 1.816
Overall Survival (Months)
d
S100A10 copy number score
TCGA dataset (n=137)
High-S100A10 copy no. (n=72)
Low-S100A10 copy no. (n=65)
Percent survival
P-value= 0.0190
HR: 1.691
Recurrence-Free Survival (Months)
e
S100A10 copy number status
TCGA cohort (n=176)
Amplification (n=5)
Gain (n=51)
Diploid (n=114)
Deletion (n=6)
Percent survival
Overall Survival (Months)
f
S100A10 copy number status
TCGA cohort (n=137)
Amplification (n=2)
Gain (n=40)
Diploid (n=91)
Deletion (n=4)
Percent survival
Recurrence-Free Survival (Months)

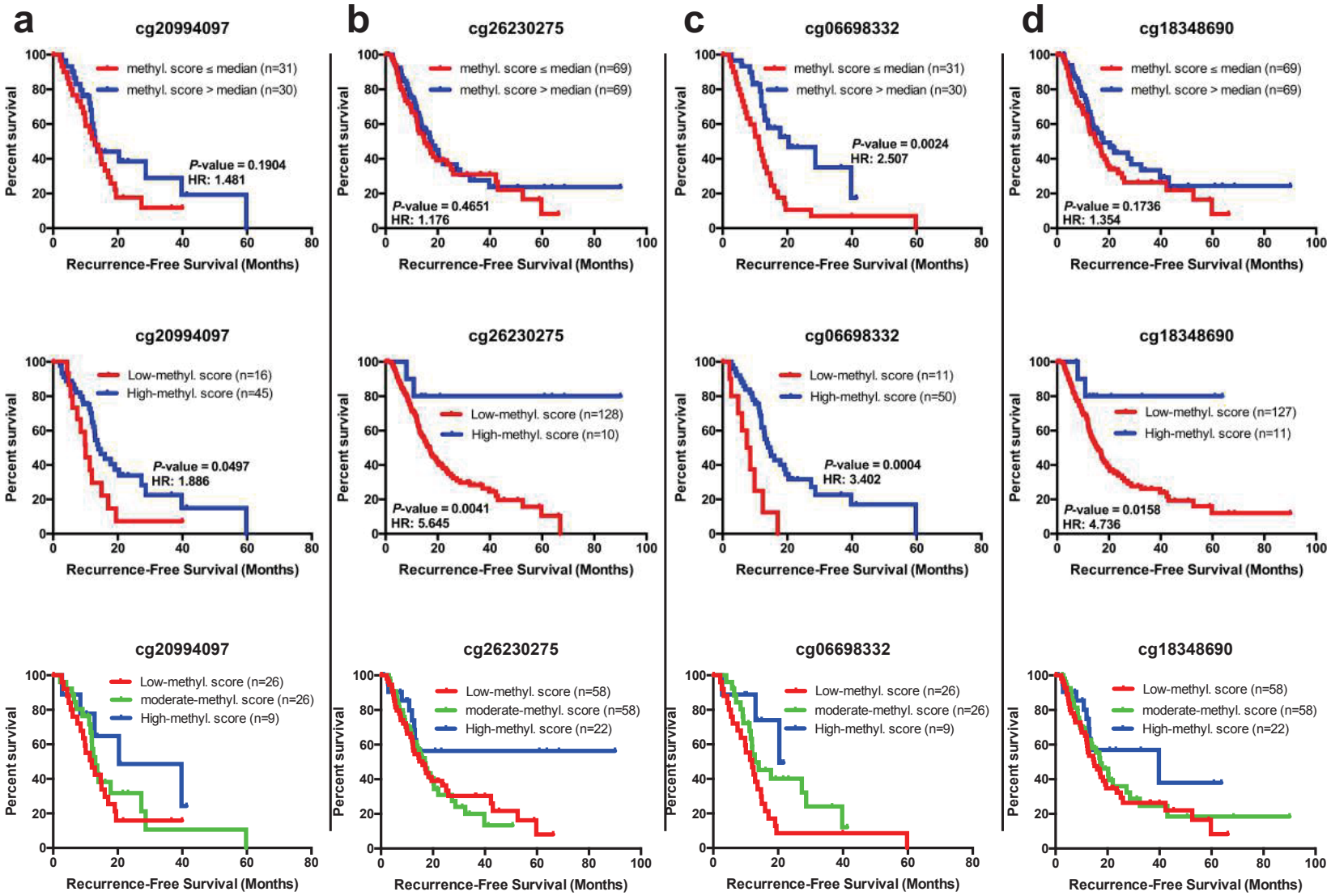
Supplementary Figure- 15. The β values of probes (CpG sites) that were not differentially-methylated and/or did not negatively correlate with S100A10 mRNA expression. The β values of each probe (CpG site) were assessed in 85 PDA tumors and 9 normal tissues. The raw data was extracted from MethHC (http://methhc.mbc.nctu.edu.tw/php/index.php) which was described by Huang et al. (2015). Nucleic Acids Res. (database issue):D856-61. Raw β values of individual probes were extracted from Maplab (http://maplab.imppc.org/wanderer/) (Villanueva et al. 2015). Epigenetics Chromatin. 8:22 (eCollection 2015). and plotted against RNASeq (RSEM) expression values of S100A10 of matched patients. used to generate correlation graphs of β values and S100A10 mRNA expression.


cg04989070
NS
β-value
Normal (n=9)
Tumor (n=85)
cg04989070
Y = -1.019e-005x + 0.5333
Pearson co. = -0.4215
C.I. -0.5354 to -0.2925
P-value < 0.0001****
S100A10 mRNA, RNASeq (RSEM)
cg18892537
P-value < 0.0300*
cg18892537
Y = 1.468e-007x + 0.02276
Pearson co. = 0.1663
C.I. 0.01966 to 0.3059
P-value = 0.0265*
cg24594295
NS
cg24594295
Y = -5.981e-008x + 0.02569
Pearson co. = -0.05702
C.I. -0.2024 to 0.09083
P-value = 0.4497
cg17711091
NS
cg17711091
Y = -1.945e-006x + 0.4191
Pearson co. = -0.2721
C.I. -0.4030 to -0.1302
P-value = 0.0002***
cg25848158
NS
cg25848158
Y = -4.170e-007x + 0.06065
Pearson co. = -0.1773
C.I. -0.3162 to -0.03105
P-value = 0.0179*
cg05368119
NS
cg05368119
Y = -1.183e-006x + 0.8720
Pearson co. = -0.1911
C.I. -0.3289 to -0.04524
P-value = 0.0106*

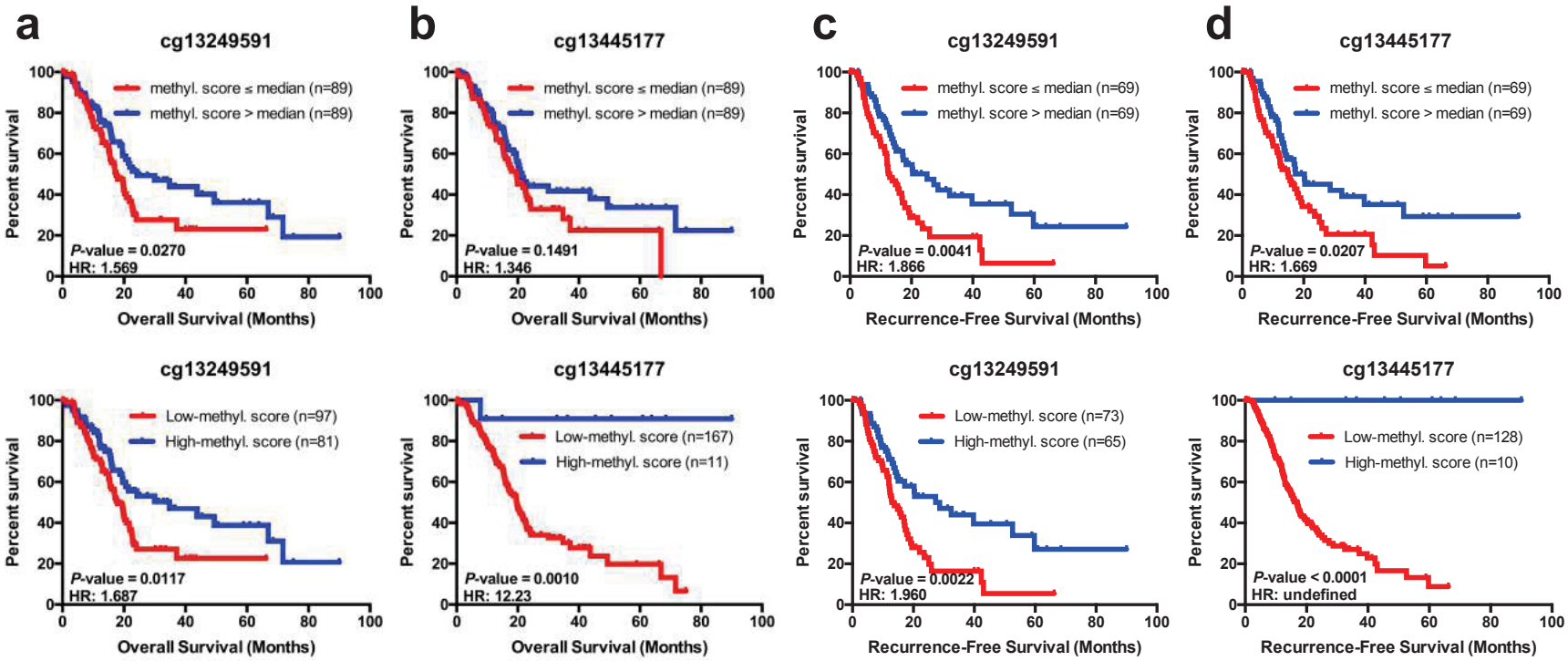
Supplemental Figure - 16. Kaplan Meier survival analyses of OS based on β values of the remaining four probes in the TCGA PDAC cohort. Kaplan Meier (KM) plots of overall survival (n=178) based on β values of the (a) cg20994097, (b) cg26230275, (c) cg06698332 and (d) cg18348690. The same three-tier method of classification was used; A median cut-off (top), best cut-off (middle), and a ternary cut-off (bottom). Raw β values of individual probes were extracted from Maplab Wanderer (Villanueva et al. 2015). Epigenetics Chromatin. 8:22 (eCollection 2015) matched with OS of TCGA PDAC patients. β values for probes cg20994097 and cg06698332 were available for 85 patients only.


a
cg20994097
methyl. score ≤ median (n=43)
methyl. score > median (n=42)
P-value = 0.6116
HR: 1.144
Percent survival
Overall Survival (Months)
cg20994097
Low-methyl. score (n=74)
High-methyl. score (n=12)
P-value = 0.2239
HR: 1.670
cg20994097
Low-methyl. score (n=36)
moderate-methyl. score (n=36)
High-methyl. score (n=13)
b
cg26230275
methyl. score ≤ median (n=89)
methyl. score > median (n=89)
P-value = 0.4239
HR: 1.181
cg26230275
Low-methyl. score (n=167)
High-methyl. score (n=11)
P-value = 0.0034
HR: 10.26
cg26230275
Low-methyl. score (n=75)
moderate-methyl. score (n=75)
High-methyl. score (n=28)
c
cg06698332
methyl. score ≤ median (n=42)
methyl. score > median (n=43)
P-value = 0.0670
HR: 1.625
cg06698332
Low-methyl. score (n=44)
High-methyl. score (n=41)
P-value = 0.0242
HR: 1.825
cg06698332
Low-methyl. score (n=36)
moderate-methyl. score (n=36)
High-methyl. score (n=13)
d
cg18348690
methyl. score ≤ median (n=89)
methyl. score > median (n=89)
P-value = 0.2161
HR: 1.295
cg18348690
Low-methyl. score (n=167)
High-methyl. score (n=11)
P-value = 0.0190
HR: 7.391
cg18348690
Low-methyl. score (n=75)
moderate-methyl. score (n=75)
High-methyl. score (n=28)

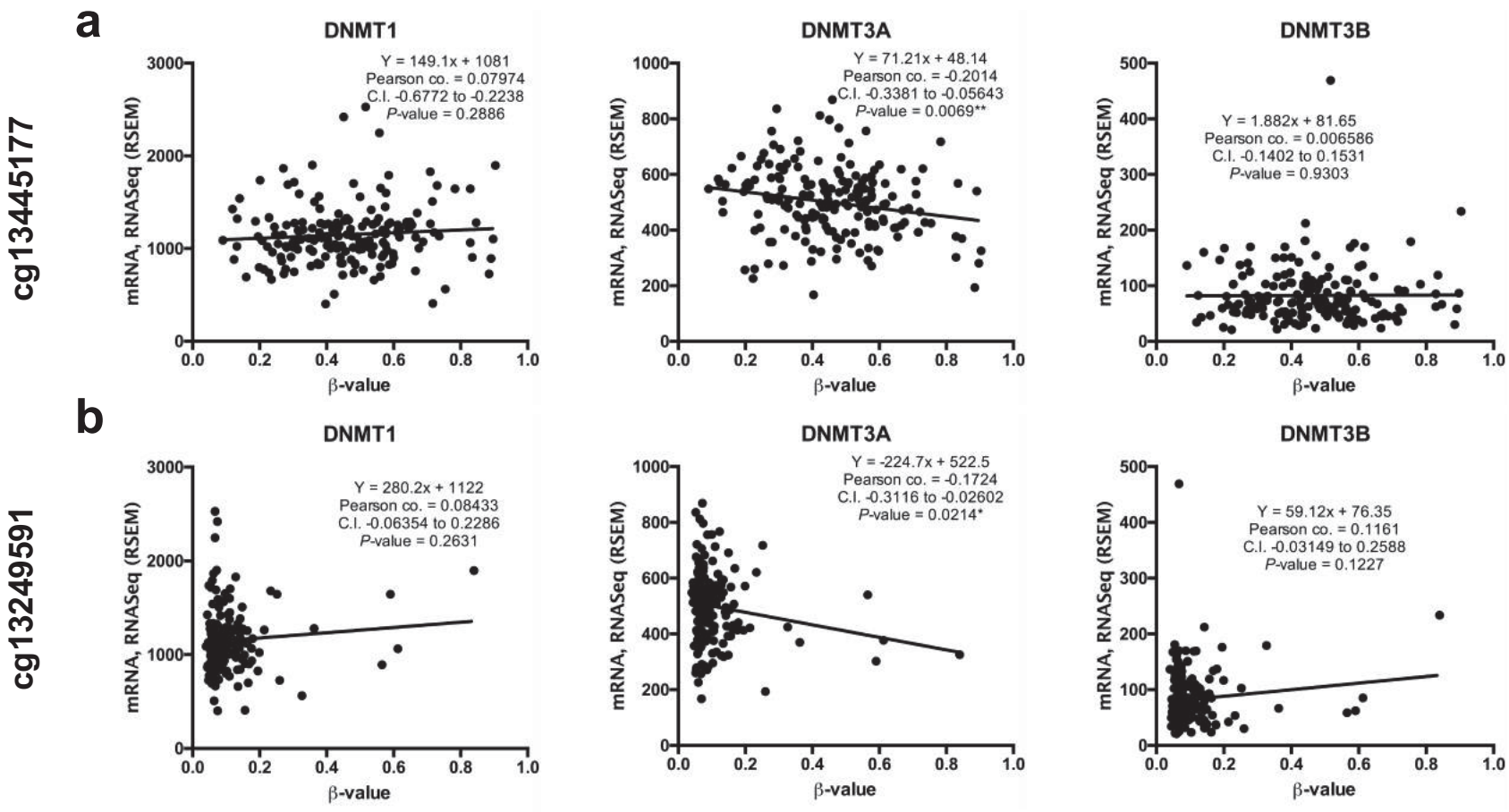
Supplemental Figure - 17. Kaplan Meier survival analyses of RFS based on β values of the remaining four probes in the TCGA PDAC cohort. Kaplan Meier (KM) plots of recurrence-free survival (n=138) based on β values of the (a) cg20994097, (b) cg26230275, (c) cg06698332 and (d) cg18348690. The same three-tier method of classification was used; A median cut-off (top), best cut-off (middle), and a ternary cut-off (bottom). Raw β values of individual probes were extracted from Maplab Wanderer (Villanueva et al. 2015). Epigenetics Chromatin. 8:22 (eCollection 2015) matched with RFS of TCGA PDAC patients. β values for probes cg20994097 and cg06698332 were available for 61 patients only.


a
cg20994097
methyl. score ≤ median (n=31)
methyl. score > median (n=30)
P-value = 0.1904
HR: 1.481
Percent survival
Recurrence-Free Survival (Months)
cg20994097
Low-methyl. score (n=16)
High-methyl. score (n=45)
P-value = 0.0497
HR: 1.886
cg20994097
Low-methyl. score (n=26)
moderate-methyl. score (n=26)
High-methyl. score (n=9)
b
cg26230275
methyl. score ≤ median (n=69)
methyl. score > median (n=69)
P-value = 0.4651
HR: 1.176
cg26230275
Low-methyl. score (n=128)
High-methyl. score (n=10)
P-value = 0.0041
HR: 5.645
cg26230275
Low-methyl. score (n=58)
moderate-methyl. score (n=58)
High-methyl. score (n=22)
c
cg06698332
methyl. score ≤ median (n=31)
methyl. score > median (n=30)
P-value = 0.0024
HR: 2.507
cg06698332
Low-methyl. score (n=11)
High-methyl. score (n=50)
P-value = 0.0004
HR: 3.402
cg06698332
Low-methyl. score (n=26)
moderate-methyl. score (n=26)
High-methyl. score (n=9)
d
cg18348690
methyl. score ≤ median (n=69)
methyl. score > median (n=69)
P-value = 0.1736
HR: 1.354
cg18348690
Low-methyl. score (n=127)
High-methyl. score (n=11)
P-value = 0.0158
HR: 4.736
cg18348690
Low-methyl. score (n=58)
moderate-methyl. score (n=58)
High-methyl. score (n=22)

Supplementary Figure - 18. Kaplan Meier analyses of CpG islands corresponding to probes cg13249591 and cg13445177 using median and optimal cut-offs. Kaplan Meier (KM) plots of (a, b) overall survival (n=178) and (c, d) recurrence-free survival (n=139) based on β values of the (a, c) cg13249591 and (b, d) cg13445177 CpG sites.

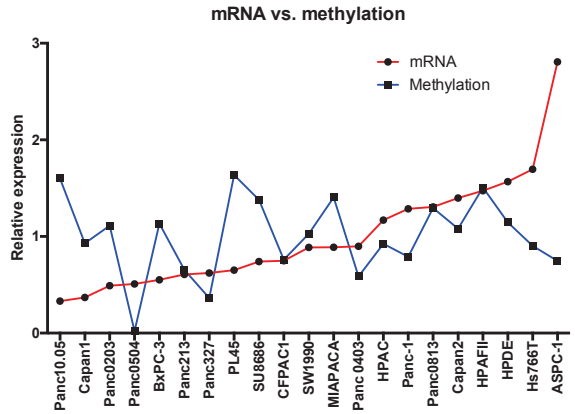

a
cg13249591
methyl. score ≤ median (n=89)
methyl. score > median (n=89)
P-value = 0.0270
HR: 1.569
Percent survival
Overall Survival (Months)
b
cg13445177
methyl. score ≤ median (n=89)
methyl. score > median (n=89)
P-value = 0.1491
HR: 1.346
Percent survival
Overall Survival (Months)
c
cg13249591
methyl. score ≤ median (n=69)
methyl. score > median (n=69)
P-value = 0.0041
HR: 1.866
Percent survival
Recurrence-Free Survival (Months)
d
cg13445177
methyl. score ≤ median (n=69)
methyl. score > median (n=69)
P-value = 0.0207
HR: 1.669
Percent survival
Recurrence-Free Survival (Months)
cg13249591
Low-methyl. score (n=97)
High-methyl. score (n=81)
P-value = 0.0117
HR: 1.687
Percent survival
Overall Survival (Months)
cg13445177
Low-methyl. score (n=167)
High-methyl. score (n=11)
P-value = 0.0010
HR: 12.23
Percent survival
Overall Survival (Months)
cg13249591
Low-methyl. score (n=73)
High-methyl. score (n=65)
P-value = 0.0022
HR: 1.960
Percent survival
Recurrence-Free Survival (Months)
cg13445177
Low-methyl. score (n=128)
High-methyl. score (n=10)
P-value < 0.0001
HR: undefined
Percent survival
Recurrence-Free Survival (Months)

Supplementary Figure - 19. The β values of the probes cg13445177 and cg13249591 do not positively correlate with mRNA expression of de novo methyltransferases. Raw β values of cg13445177 and cg13249591 were extracted from Maplab and plotted against RNA Seq (RSEM) expression values of the de novo methyltransferases DNMT1, DNMT3B and DNMT3A.


a
cg13445177
DNMT1
Y = 149.1x + 1081
Pearson co. = 0.07974
C.I. -0.6772 to -0.2238
P-value = 0.2886
DNMT3A
Y = 71.21x + 48.14
Pearson co. = -0.2014
C.I. -0.3381 to -0.05643
P-value = 0.0069**
DNMT3B
Y = 1.882x + 81.65
Pearson co. = 0.006586
C.I. -0.1402 to 0.1531
P-value = 0.9303
b
cg13249591
DNMT1
Y = 280.2x + 1122
Pearson co. = 0.08433
C.I. -0.06354 to 0.2286
P-value = 0.2631
DNMT3A
Y = -224.7x + 522.5
Pearson co. = -0.1724
C.I. -0.3116 to -0.02602
P-value = 0.0214*
DNMT3B
Y = 59.12x + 76.35
Pearson co. = 0.1161
C.I. -0.03149 to 0.2588
P-value = 0.1227
mRNA, RNASeq (RSEM)
β-value

Supplemental Figure - 20. S100A10 promoter methylation. (A) Analysis of relative mRNA and methylation scores of the 21 PDAC cell lines in CCLE. The CCLE expression values were normalized to the average of the mRNA and methylation scores respectively to allow single-axis plot. (B) The 377-nucleotide promoter region of S100A10 used for pyrosequencing. The sequence highlights the sequenced CpG sites as well the location of HM450 methylation probes (as highlighted). The beginning of exon1 is underlined. (C) Promoter CpG island analysis using EMBOSS CpGplot tool from the EMBL-EBI database: (https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/). The CpG island criteria set by Takai and Jones (2002) were used. These include: 1) minimum length of an island is 500bp. 2) Minimum observed/expected is the minimum average observed to expected ratio of C plus G to CpG in a set of 10 windows that are required before a CpG island is reported. The threshold value is 0.65. 3) Minimum percentage is minimum average percentage of G plus C a set of 10 windows that are required before a CpG island is reported. The threshold value is 0.55.

a



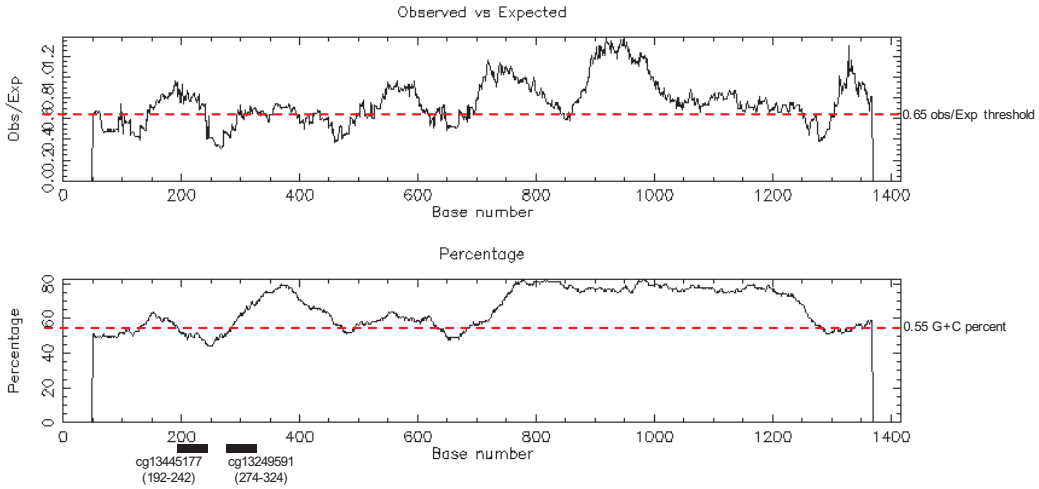
b

Analyzed Sequence (377 nt)

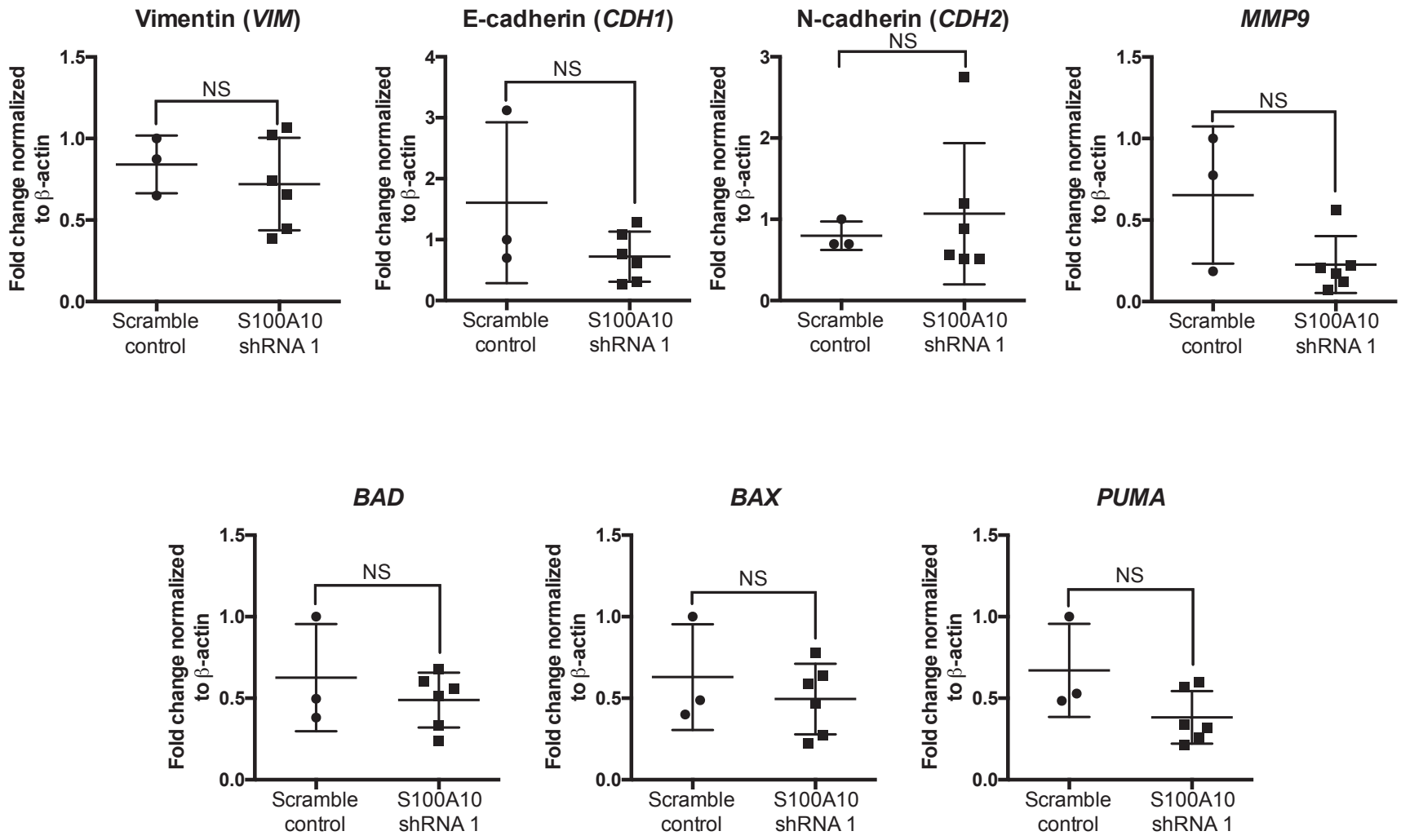
TC[1]**CG**CC[2]**CG**CCT[3]**CG**GCCTCCCAAAGTGTTGGGATTACAGG[4]**CG**TGAGCCAC[5]**CG**[6]**CG**CC[7]**CG**GCCAGTTTTTAACAC
TATTAGCCACACTGAAACTGAACTATTGATCAAGTGA[8]**CG**CCACACAAAGGGGTAAATCCCCTGTTCAACAAAGGG
TTTGTGA[9]**CG**CCCCTGGGTGCTGACAAGCCAAAC[10]**CG**CACCCTCCCTG[11]**CG**GCACCT[12]**CG**[13]**CG**GGC[14]**CG**GTGGGG[15]
CGGGAAGCC[16]**CG**GCTTCTGGGGAGGTGC[17]**CG**CCCCTCCACTGG[18]**CG**CAGGC[19]**CG**C[20]**CG**AGACCCCCAGA[21]**CG**GAC
CTCCTAGGGCTAATCTGATAGTGCCTCTGAGGT[22]**CG**ATAGGACTCCA[23]**CG**TGCCACTCCCTGCAGGGTCATCCAGCA
AGTAATTCCTAGACC[24]**CG**

HM450 Array Probe: cg20167074; cg13445177; cg13249591; cg25848158; cg24594295

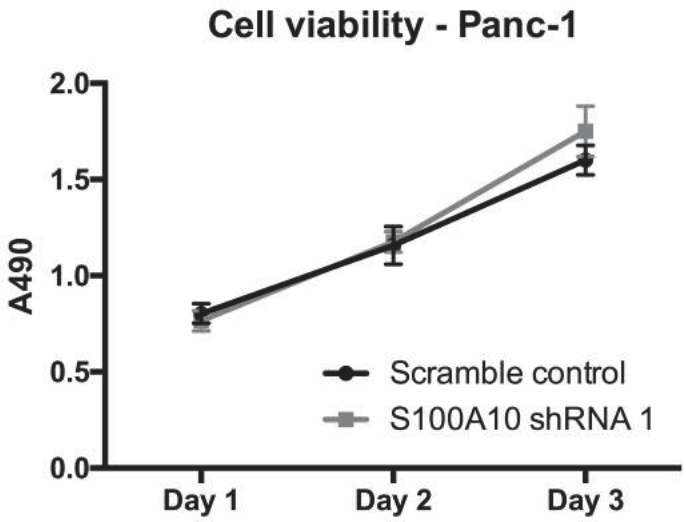
c



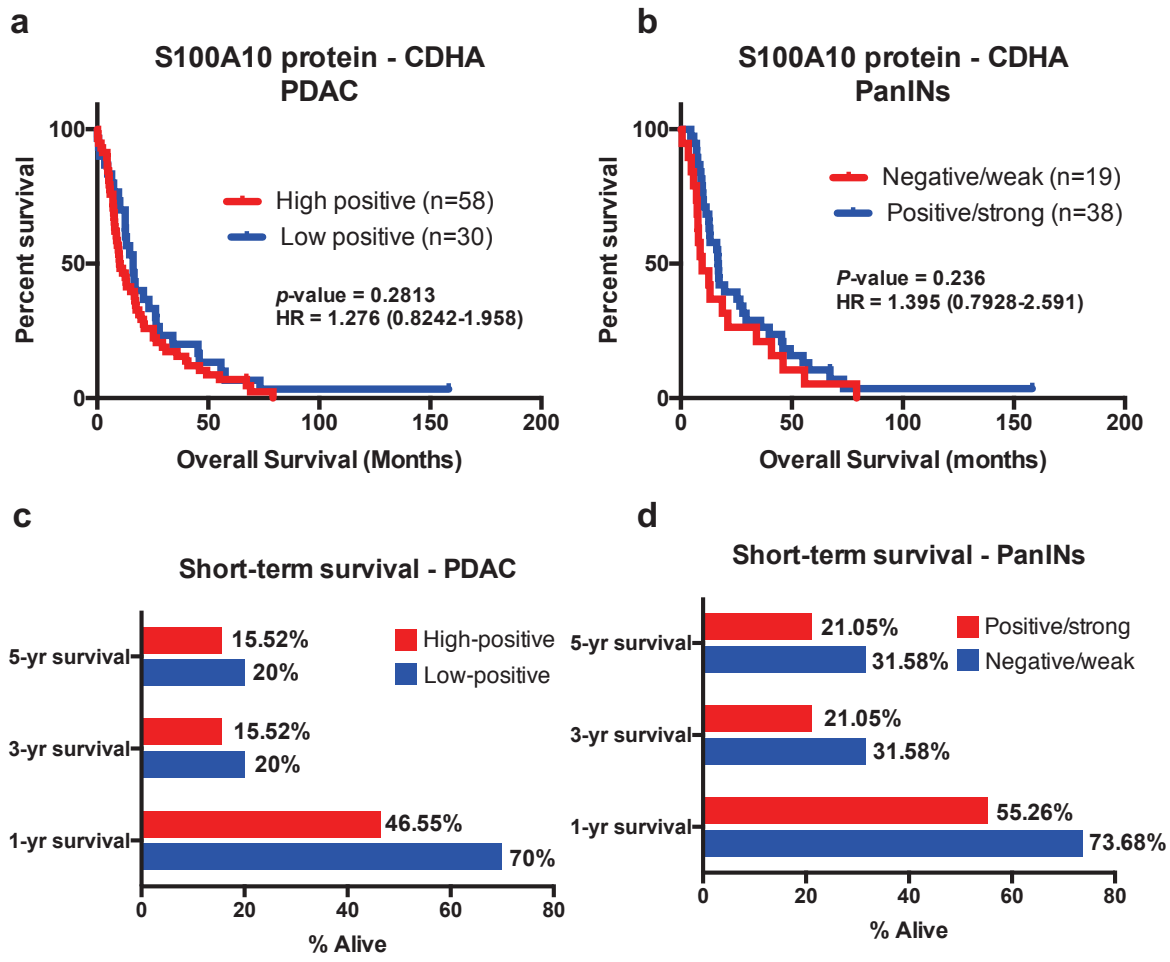
Supplemental Figure - 21. RT-qPCR of several genes in scramble control and S100A10-shRNA 1 Panc-1 tumors. These genes were not significantly altered by S100A10 depletion.


Vimentin (VIM)
E-cadherin (CDH1)
N-cadherin (CDH2)
MMP9
BAD
BAX
PUMA
Fold change normalized to β-actin
NS
Scramble control
S100A10 shRNA 1
0.0
0.5
1.0
1.5
0
1
2
3
4

Supplemental Figure - 22. Assessment of short-term cell viability of scramble control and S100A10 shRNA1 Panc-1 cells. Cells were equally seeded into a 96-well plate and cell viability was measured every day for three consecutive days. The absorbance of the MTS reagent at 490nm is plotted for each time point.


Cell viability - Panc-1
A490
2.0
1.5
1.0
0.5
0.0
Day 1
Day 2
Day 3
Scramble control
S100A10 shRNA 1

Supplemental Figure - 23. Overall survival estimators in CDHA PDAC patients based on S100A10 protein expression. Kaplan Meier analysis of OS of the CDHA cohort based on S100A10 protein expression in the PDAC (a) and PanINs (b) regions. Short-term survival (1-, 3- and 5-year) of CDHA patients based on S100A10 expression in PDAC (c) and PanINs (d).


a
S100A10 protein - CDHA
PDAC
Percent survival
100
50
0
High positive (n=58)
Low positive (n=30)
p-value = 0.2813
HR = 1.276 (0.8242-1.958)
0
50
100
150
200
Overall Survival (Months)
b
S100A10 protein - CDHA
PanINs
Percent survival
100
50
0
Negative/weak (n=19)
Positive/strong (n=38)
P-value = 0.236
HR = 1.395 (0.7928-2.591)
0
50
100
150
200
Overall Survival (months)
c
Short-term survival - PDAC
5-yr survival
3-yr survival
1-yr survival
15.52%
20%
15.52%
20%
46.55%
70%
High-positive
Low-positive
0
20
40
60
80
% Alive
d
Short-term survival - PanINs
5-yr survival
3-yr survival
1-yr survival
21.05%
31.58%
21.05%
31.58%
55.26%
73.68%
Positive/strong
Negative/weak
0
20
40
60
80
% Alive

APPENDIX C: SUPPLEMENTAL TABLES III

Supplemental Table -- 16. Calculation scheme of the H-score. The score represents both the intensity and number of DAB-positive pixels in stained tissue microarrays.

Score	Staining intensity
0	Negative
1	Low positive
2	Positive
3	High Positive

H-score =

0 x percentage contribution of negative pixels
1 x percentage contribution of low positive pixels
2 x percentage contribution of positive pixels
3 x percentage contribution of highly positive pixels

Supplemental Table -- 17. Higher S100A10 mRNA, higher copy number and low-methylation scores correlate with lower short-term survival. The percentages are calculated as the likelihoods of being alive or recurrence-free. Briefly, the percentage of patients alive (y-axis value on KM survival curve) was noted after one, three and five years of the duration of follow-up in the CDHA cohort.

S100A10 genomic profile	Overall Survival (OS)			Recurrence-free Survival (RFS)		
	1-yr survival	3-yr survival	5-yr survival	1-yr survival	3-yr survival	5-yr survival
Low mRNA	69.66% (62/89)	13.48% (12/89)	5.62% (5/89)	58.57% (41/70)	14.29% (10/70)	5.71% (4/70)
High mRNA	59.55% (53/89)	8.99% (8/89)	3.37% (3/89)	49.28% (34/69)	10.14% (7/69)	1.45% (1/69)
High copy number	65.91% (58/88)	10.23% (9/88)	3.41% (3/88)	48.57% (34/70)	8.570% (6/70)	1.43% (1/70)
Low copy number	62.5% (55/88)	12.5% (11/88)	5.68% (5/88)	57.97% (40/69)	15.94% (11/69)	5.80% (4/69)
High methylation score	65.17% (58/89)	6.74% (6/89)	2.25% (2/89)	55.71% (39/70)	17.14% (12/70)	5.71% (4/70)
Low methylation score	64.04% (57/89)	15.73% (14/89)	6.74% (6/89)	52.17% (36/69)	7.25% (5/69)	1.45% (1/69)

Supplemental Table -- 18. Multiple comparisons of OS and RFS with the mRNA Ternary classifier. Multiple comparisons of S100A10 mRNA survival functions were performed on the TCGA, Chen et al. (GSE57495, n=63), Moffitt et al. (GSE71729, n=125) and ICGC (international cancer genome consortium, n=133). P-values were adjusted to the Bonferroni-corrected threshold. Adjusted p-value is p-value/K = 0.017 where K=3 and represents the number of comparisons made.

S100A10 mRNA									
Study	weak neg vs. high-pos			weak neg vs. low-pos			low-pos vs. high pos		
	HR	p-value	C.I.	HR	p-value	C.I.	HR	p-value	C.I.
TCGA OS	2.84	0.0039	1.310 to 4.000	1.855	0.1056	0.8974 to 3.317	1.457	0.0856	0.9503 to 2.236
TCGA RFS	6.668	< 0.0001	2.067 to 6.833	4.742	0.0008	1.640 to 5.919	1.597	0.0353	1.036 to 2.545
GSE57495 OS	2.929	0.0402	1.047 to 5.788	3.28	0.02	1.180 to 6.058	0.9491	0.8716	0.5021 to 1.792
GSE71729 OS	3.365	0.0026	1.454 to 5.182	4.373	< 0.0001	1.926 to 6.163	0.7575	0.2055	0.4734 to 1.158
ICGC OS	3.715	0.0073	1.296 to 5.112	2.692	0.0495	1.007 to 4.458	1.429	0.1297	0.9028 to 2.350

Supplemental Table -- 19. Final three-variable and two-variable models of OS and RFS in the TCGA PDAC cohort. These models calculate hazard ratios based on the most significant variables in predicting OS and RFS.

OS								
Variable	coef	exp(coef)	se(coef)	z	Pr(> |z|)	exp(-coef)	lower .95	upper .95
S100A10 mRNA	0.43	1.54	0.19	2.31	0.02	0.65	1.07	2.21
Lymph node (N1)	0.66	1.93	0.27	2.48	0.01	1.93	1.15	3.24
age	0.02	1.02	0.01	2.08	0.04	0.98	1	1.04

RFS								
Variable	coef	exp(coef)	se(coef)	z	Pr(> |z|)	exp(-coef)	lower .95	upper .95
S100A10 mRNA	0.64	1.89	0.19	3.44	0	0.53	1.32	2.72
Lymph node (N1)	0.43	1.54	0.25	1.71	0.09	0.65	0.94	2.53

Supplemental Table -- 20. The location and target sequence of 15 methylation probes associated with S100A10. All 15 probe sites were extracted from the Illumina Human Methylation450 v1.2(https://support.illumina.com) and CpG sites identified in the genomic sequence complementary to each probe. T_cSS: transcription start site, T_LSS: translation start site, TSS1500: region between 200bp and 1500bp upstream of T_cSS, TSS200: region 200bp upstream of T_cSS, 5'UTR: 5' untranslated region.

Name	Strand	UCSC Ref	Genomic sequence (5'-3')
cg04989070	-	5'UTR	**CG**AGAAAATAGCAAGTGTTAGAAGAGAAGGAGCACAGTCATGTCATTCTG
cg05368119	+	5'UTR/Body	**CG**TGTTCCATTTGAGATGGCATTTTGGTGTGGTC**CG**TTGAAGCCTATTAA
cg06698332	-	TSS1500	**CG**AGGAGTTGGTAAGCATCCCCTAGGAAACACTTAGGTTTTCTCTAAATT
cg06786599	-	5'UTR	**CGCG**CCCCTCCTGGGTAGTCCCCAGGCC**CG**GACCTGCTGCC**CG**GGGAAAA
cg13249591	+	TSS200	**CG**GTTTGGCTTGTCAGCACCCAGGGG**CG**TCACAAACCCTTTGTTGAACAG
cg13445177	-	TSS1500	**CG**CC**CG**GCCAGTTTTTAACACTATTAGCCACACTGAAACTGAACTATTGA
cg17711091	-	5'UTR	**CG**AATCCCTCCTACA**CG**CCCCTGCCCTGGCTGG**CG**TGTTTCTTGGTGGGT
cg18348690	-	TSS1500	**CG**GAAAGTAATAGCTGAAATCCAAGTTGGGTTTTCCTGGCAACAGCCAAT
cg18892537	-	5'UTR	**CG**GCTGGTGGGGAATC**CG**CTGCTCAGTGCTC**CG**GGCCACACCCAAA**CG**AG
cg20167074	-	TSS1500	**CGCG**CC**CG**GCCAGTTTTTAACACTATTAGCCACACTGAAACTGAACTATT
cg20994097	+	TSS1500	**CG**TTAGATAAGCAAACAATTAAAATCCAGTGTGCTTAGTGAATGGTAGAG
cg24594295	-	TSS200	**CG**GACCTCCTAGGGCTAATCTGATAGTGCCTCTGAGGT**CG**ATAGGACTCC
cg25848158	-	TSS200	**CG**C**CG**AGACCCCCAGA**CG**GACCTCCTAGGGCTAATCTGATAGTGCCTCTG
cg26230275	-	5'UTR	**CG**CTCCAAGCCAGGCCAGCACAGGGGAGCCCTAAGCCAGATTCTGGGATG
cg16658496	+	Exon1;5'UTR	**CG**CCCTGCCCT**CG**CTCC**CG**GACC**CG**CCT**CG**CAGAGGCCT**CG**CC**CG**CCCCA

Supplemental Table -- 21. Multiple comparisons of OS and RFS using the mRNA Ternary classifier. Multiple comparisons of S100A10 methylation survival functions were performed on the TCGA and ICGC patient cohorts. P-values were adjusted to the Bonferroni-corrected threshold. Adjusted p-value is p-value/K = 0.017 where K=3 and represents the number of comparisons made.

S100A10 methylation OS TCGA									
	high vs. low			high vs. intermediate			low vs. intermediate		
Probe	HR	p-value	C.I.	HR	p-value	C.I.	HR	p-value	C.I.
cg13249591	1.702	0.094	0.9280 to 2.876	1.53	0.1936	0.8210 to 2.697	1.203	0.4002	0.7814 to 1.878
cg13445177	1.848	0.0613	0.9831 to 3.100	2.041	0.0298	1.076 to 3.403	0.9533	0.8276	0.6155 to 1.473
cg18348690	1.668	0.1338	0.8744 to 2.861	1.399	0.3442	0.7173 to 2.599	1.268	0.2842	0.8221 to 1.954
cg26230275	1.472	0.2664	0.7660 to 2.660	1.704	0.1244	0.8797 to 2.936	0.8486	0.4559	0.5488 to 1.308
cg20994097	1.577	0.2641	0.7358 to 3.350	1.321	0.5103	0.5955 to 2.866	1.154	0.6117	0.6591 to 2.050
cg06698332	1.535	0.3362	0.6716 to 3.250	1.182	0.708	0.5018 to 2.789	1.179	0.5575	0.6773 to 2.076

S100A10 methylation RFS TCGA									
	high vs. low			high vs. intermediate			low vs. intermediate		
Probe	HR	p-value	C.I.	HR	p-value	C.I.	HR	p-value	C.I.
cg13249591	2.441	0.011	1.209 to 3.927	2	0.054	1.004 to 3.561	1.372	0.1774	0.8667 to 2.189
cg13445177	2.375	0.0155	1.154 to 3.486	2.25	0.0317	1.082 to 3.977	0.9959	0.9849	0.6469 to 1.533
cg18348690	1.768	0.1156	0.8873 to 3.070	1.494	0.2827	0.7393 to 2.827	1.164	0.5211	0.7330 to 1.848
cg26230275	2.009	0.0636	0.9748 to 3.452	1.835	0.1133	0.8850 to 3.281	0.9221	0.7231	0.5796 to 1.455
cg20994097	2.053	0.1329	0.8360 to 4.379	1.683	0.2819	0.6841 to 3.954	1.256	0.4758	0.6690 to 2.392
cg06698332	3.723	0.0187	1.206 to 6.208	1.787	0.3323	0.6048 to 4.713	1.68	0.0936	0.9248 to 3.218

S100A10 methylation OS ICGC									
	high vs. low			high vs. intermediate			low vs. intermediate		
Probe	HR	p-value	C.I.	HR	p-value	C.I.	HR	p-value	C.I.
cg13249591	2.63	< 0.0001	1.510 to 3.331	1.862	0.0134	1.133 to 2.699	1.521	0.0052	1.144 to 2.097
cg13445177	2.666	0.0006	1.393 to 3.298	2.372	0.0025	1.282 to 3.134	1.167	0.302	0.8706 to 1.573

Supplemental Table -- 22. List of human primer sequences used in RT-qPCR and pyrosequencing as well as dsDNA oligo used for S100A10 shRNA. b represents biotinylated primers. ”Seq” is used for the pyrosequencing step along with the biotinylated reverse primer.

Gene	Primer	Sequence
ACTB (β-Actin)	Forward	CTTCCAGCCTTCCTTCCTGG
	Reverse	CTGTGTTGGCGTACAGGTCCT
S100A10	Forward	CCCTCTGGCTGTGGACAAAA
	Reverse	CGACCCTTTGGGACAACTCT
VEGFA	Forward	CTTGCCTTGCTGCTCTACCT
	Reverse	GCAGTAGCTGCGCTGTGATAGA
CCND1 (Cyclin D1)	Forward	GATGCCAACCTCCTCAACGA
	Reverse	GGAAGCGGTCCAGGTAGTTC
VIM (Vimentin)	Forward	TCTACGAGGAGGAGATGCGG
	Reverse	GGTCAAGACGTGCCAGAGAC
CDH1 (E-Cadherin)	Forward	TCATGAGTGTCCCCCGGTAT
	Reverse	TCTTGAAGCGATTGCCCCAT
CDH2 (N-Cadherin)	Forward	AGCCTGACACTGTGGAGCCT
	Reverse	TCAGCGTGGATGGGTCTTTC
MMP9	Forward	CAAGGACCGGTTTATTTGGC
	Reverse	ATTCCCTGCGAAGAACACAGC
BAD	Forward	GGTTCTGAGGGGAGACTGAGGT
	Reverse	ACTCGGCTCAAACTCTGGGA
BAX	Forward	CAGGGGCCCTTTTGCTTCAG
	Reverse	TAGAAAAGGGCGACAACCCG
PUMA	Forward	CTCCTCTCGGTGCTCCTTCA
	Reverse	CTCTCTCTAAACCTATGCAATGGGA
Primers for bisulfite conversion and pyrosequencing		
S100A10	Forward	TGGTTAAGTTGGTGTTGAATTT
	Reverse[b]	AATAACCCTACAAAAAATAACA[b]
	Forward[Seq]	AATTTTTGATTTGAGGTGA
p-SUPER shRNA		
S100A10	dsDNA oligo	5'-GAT CCC CGT GGG CTT CCA GAG CTT CTT TCA AGA GAA GAA GCT CTG GAA GCC CAC TTT TTA-3'
		5'-AGC TTA AAA AGT GGG CTT CCA GAG CTT CTT CTC TTG AAA GAA GCT CTG GAA GCC CAC GGG-3'

Collaborators who contributed to his work:

- **David Waisman** (PhD, Departments of Pathology and Biochemistry and Molecular Biology, Dalhousie University) contributed to experimental design and to editing the dissertation manuscript.
- **Dr. Weei-Yuan Huang** (MD/PhD, Department of Pathology, CDHA) annotated pancreatic tissue samples (normal, PanINs and PDAC) and provided input on experimental design.
- **Dr. Hong Gu** (PhD, Department of Mathematics and Statistics, Dalhousie University) performed univariate and multivariate regression analyses of overall survival and recurrence-free survival.
- **Andra Sterea** (BSc, Biology and Neuroscience, Dalhousie University) performed RT-qPCR analysis.
- **Dr. Andrea Uzans** (MD, Dalhousie Medical School, Dalhousie University) constructed TMAs and helped annotate pancreatic tissue samples.
- **Henry Liptay** (BSc, Biology and Neuroscience, Dalhousie University) performed Zarnestra treatments of Panc-1 and Bx-PC3 cells.
- **Dr. Ian Weaver** (PhD, Department of Psychology and Neuroscience, Department of Psychiatry, Brain Repair Centre, Dalhousie University) conceptualized and performed DNA methylation experiments.
- **Gloria Rodrigues** (BSc, Department of Psychology and Neuroscience) performed DNA methylation experiments.

REFERENCES

[1] S. Statistics Canada, “The 10 leading causes of death,” *The 10 leading causes of death*, 2012. .

[2] American Cancer Society, “No Title,” *Cancer Facts & Figures 2017*, 2017. .

[3] N. C. National Cancer Institute, “No Title,” *Cancer Statistics*, 2016. .

[4] P. C. Nowell, “The clonal evolution of tumor cell populations.,” *Science*, vol. 194, no. 4260, pp. 23–8, 1976.

[5] D. Hanahan and R. A. Weinberg, “Review Hallmarks of Cancer : The Next Generation,” *Cell*, vol. 144, no. 5, pp. 646–674, 2011.

[6] D. Hanahan and R. A. Weinberg, “The hallmarks of cancer.,” *Cell*, vol. 100, no. 1, pp. 57–70, 2000.

[7] R. A. Weinberg, “Coming full circle - From endless complexity to simplicity and back again,” *Cell*, vol. 157, no. 1. pp. 267–271, 2014.

[8] American Society of Clinical Oncology, “Guidelines, Tools and Resources,” *Guidelines, Tools and Resources*, 2017. .

[9] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman, “Survival Analysis Part I: Basic concepts and first analyses,” *British Journal of Cancer*, vol. 89, no. 2. pp. 232–238, 2003.

[10] et al. Gaynon PS, Steinherz PG, Bleyer WA, Ablin AR, Albo VC, Finklestein JZ,

Grossman NJ, Novak LJ, Pyesmany AF, Reaman GH, “Improved therapy for children with acute lymphoblastic leukemia and unfavorable presenting features: a follow-up report of the Childrens Cancer Group Study CCG-106,” *J Clin Oncol*, vol. 11, no. 11, pp. 2234–42, 1993.

[11] J. A. O’Shaughnessy *et al*., “Commentary concerning demonstration of safety and efficacy of investigational anticancer agents in clinical trials,” *J. Clin. Oncol*., vol. 9, no. 12, pp. 2225–2232, 1991.

[12] Early Breast Cancer Trialists’ Collaborative Group, “Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy,” *Lancet*, vol. 339, no. 1–15, pp. 71–85, 1992.

[13] K. Donovan, R. W. Sanson-Fisher, and S. Redman, “Measuring quality of life in cancer patients.,” *J. Clin. Oncol*., vol. 7, no. 7, pp. 959–68, 1989.

[14] J. V. McGowan, R. Chung, A. Maulik, I. Piotrowska, J. M. Walker, and D. M. Yellon, “Anthracycline Chemotherapy and Cardiotoxicity,” *Cardiovasc. Drugs Ther*., vol. 31, no. 1, pp. 63–75, 2017.

[15] et al Moinpour CM, Feigl P, Metch B, “Quality of life end points in cancer clinical trials: Review and recommendations,” *J Natl Cancer Inst*, vol. 81, pp. 485–495, 1989.

[16] T. J. Smith, B. E. Hillner, and C. E. Desch, “Efficacy and cost-effectiveness of cancer treatment: Rational allocation of resources based on decision analysis,” *J. Natl. Cancer Inst*., vol. 85, no. 18, pp. 1460–1474, 1993.

[17] M. Tamburini, “Health-related quality of life measures in cancer,” *Ann. Oncol.*, vol. 12, no. Suppl. 3, pp. S7–S10, 2001.

[18] D. Osoba, “Health-related quality of life and cancer clinical trials,” *Therapeutic Advances in Medical Oncology*, vol. 3, no. 2. pp. 57–71, 2011.

[19] S. G. Nayfield, P. A. Ganz, C. M. Moinpour, D. F. Cella, and B. J. Hailey, “Report from a National Cancer Institute (USA) workshop on quality of life assessment in cancer clinical trials,” *Qual. Life Res.*, vol. 1, no. 3, pp. 203–210, 1992.

[20] Y. Lu *et al.*, “New-onset type 2 diabetes, elevated HbA1c, anti-diabetic medications, and risk of pancreatic cancer,” *Br. J. Cancer*, vol. 113, no. 11, pp. 1607–1614, 2015.

[21] S. K. Das *et al.*, “Gene Therapies for Cancer: Strategies, Challenges and Successes.,” *J. Cell. Physiol.*, vol. 230, no. 2, pp. 259–71, 2015.

[22] E. R. Mardis and R. K. Wilson, “Cancer genome sequencing: A review,” *Hum. Mol. Genet.*, vol. 18, no. R2, 2009.

[23] I. H. G. S. Consortium, “Initial sequencing and analysis of the human genome.,” *Nature*, vol. 409, pp. 860–921, 2001.

[24] G. M. Grant *et al.*, “Microarrays in Cancer Research,” in *Anticancer Research*, 2004, vol. 24, no. 2 A, pp. 441–448.

[25] Z. Wang, M. Gerstein, and M. Snyder, “RNA-Seq: A revolutionary tool for transcriptomics,” *Nature Reviews Genetics*, vol. 10, no. 1. pp. 57–63, 2009.

[26] A. T. Maia, S.-J. Sammut, A. Jacinta-Fernandes, and S.-F. Chin, "Big data in cancer genomics," *Curr. Opin. Syst. Biol.*, 2017.

[27] J. D. Campbell *et al.*, "The Case for a Pre-Cancer Genome Atlas (PCGA)," *Cancer Prevention Research*, vol. 9, no. 2. pp. 119–124, 2016.

[28] A. D. Goldberg, C. D. Allis, and E. Bernstein, "Epigenetics: A Landscape Takes Shape," *Cell*, vol. 128, no. 4. pp. 635–638, 2007.

[29] C. Waddington, "Genetic Assimilation of the Bithorax Phenotype," *Evolution (N. Y).*, vol. 10, no. 1, pp. 1–13, 1956.

[30] A. P. Feinberg and B. Tycko, "The history of cancer epigenetics," *Nat. Rev. Cancer*, vol. 4, no. 2, pp. 143–153, 2004.

[31] A. P. Feinberg and B. Vogelstein, "Hypomethylation distinguishes genes of some human cancers from their normal counterparts," *Nature*, vol. 301, no. 5895, pp. 89–92, 1983.

[32] M. A. Gama-sosa *et al.*, "The 5-methylcytosine content of DNA from human tumors," *Nucleic Acids Res.*, vol. 11, no. 19, pp. 6883–6894, 1983.

[33] N. Sato *et al.*, "Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma," *Cancer Res.*, vol. 63, no. 14, pp. 4158–4166, 2003.

[34] N. Nakamura and K. Takenaga, "Hypomethylation of the metastasis-associated S100A4 gene correlates with gene activation in human colon adenocarcinoma cell

lines.," *Clin. Exp. Metastasis*, vol. 16, no. 5, pp. 471–9, 1998.

[35] C. J. Piyathilake *et al.*, "Race- and age-dependent alterations in global methylation of DNA in squamous cell carcinoma of the lung (United States)," *Cancer Causes Control*, vol. 14, no. 1, pp. 37–42, 2003.

[36] Y. Oshimo *et al.*, "Promoter methylation of cyclin D2 gene in gastric carcinoma.," *Int. J. Oncol.*, vol. 23, no. 6, pp. 1663–1670, 2003.

[37] S. B. Baylin and P. A. Jones, "A decade of exploring the cancer epigenome-biological and translational implications," *Nature Reviews Cancer*, vol. 11, no. 10. pp. 726–734, 2011.

[38] H. Wu and Y. Zhang, "Mechanisms and functions of Tet protein- mediated 5-methylcytosine oxidation," *Genes Dev.*, vol. 25, pp. 2436–2452, 2011.

[39] M. Tan *et al.*, "Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification," *Cell*, vol. 146, no. 6, pp. 1016–1028, 2011.

[40] L. Z. Strichman-Almashanu *et al.*, "A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes," *Genome Res.*, vol. 12, no. 4, pp. 543–554, 2002.

[41] K. D. Robertson, "DNA methylation and human disease," *Nature Reviews Genetics*, vol. 6, no. 8. pp. 597–610, 2005.

[42] T. J. Ley *et al.*, "DNMT3A mutations in acute myeloid leukemia," *N Engl J Med*, vol. 363, no. 25, pp. 2424–2433, 2010.

[43] E. Li, T. H. Bestor, and R. Jaenisch, “Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.,” *Cell*, vol. 69, no. 6, pp. 915–926, 1992.

[44] M. Okano, D. W. Bell, D. A. Haber, and E. Li, “DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development,” *Cell*, vol. 99, no. 3, pp. 247–257, 1999.

[45] R. J. Klose and A. P. Bird, “Genomic DNA methylation: The mark and its mediators,” *Trends in Biochemical Sciences*, vol. 31, no. 2. pp. 89–97, 2006.

[46] V. G. ALLFREY, R. FAULKNER, and A. E. MIRSKY, “Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 51, no. 1938, pp. 786–94, 1964.

[47] T. Kouzarides, “Chromatin Modifications and Their Function,” *Cell*, vol. 128, no. 4. pp. 693–705, 2007.

[48] Z. Wang *et al.*, “Combinatorial patterns of histone acetylations and methylations in the human genome,” *Nat. Genet.*, vol. 40, no. 7, pp. 897–903, 2008.

[49] S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, and D. J. Patel, “How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers,” *Nature Structural and Molecular Biology*, vol. 14, no. 11. pp. 1025–1040, 2007.

[50] N. Avvakumov and J. Côté, “The MYST family of histone acetyltransferases and

their intimate links to cancer," *Oncogene*, vol. 26, no. 37. pp. 5395–5407, 2007.

[51] L. Pasqualucci *et al*., "Inactivating mutations of acetyltransferase genes in B-cell lymphoma," *Nature*, vol. 471, no. 7337, pp. 189–196, 2011.

[52] J. Wang *et al*., "Conditional MLL-CBP targets GMP and models therapy-related myeloproliferative disease," *EMBO J*., vol. 24, no. 2, pp. 368–381, 2005.

[53] A. J. Bannister and T. Kouzarides, "Regulation of chromatin by histone modifications," *Cell Research*, vol. 21, no. 3. pp. 381–395, 2011.

[54] M. Bantscheff *et al*., "Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes," *Nat. Biotechnol*., vol. 29, no. 3, pp. 255–268, 2011.

[55] F. Grignani *et al*., "Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia.," *Nature*, vol. 391, no. 6669, pp. 815–8, 1998.

[56] R. W. Johnstone and J. D. Licht, "Histone deacetylase inhibitors in cancer therapy: Is transcription the primary target?," *Cancer Cell*, vol. 4, no. 1. pp. 13–18, 2003.

[57] E. A. Olsen *et al*., "Phase IIB multicenter trial of vorinostat in patients with persistent, progressive, or treatment refractory cutaneous t-cell lymphoma," *J. Clin. Oncol*., vol. 25, no. 21, pp. 3109–3115, 2007.

[58] F. Crea, P. L. Clermont, A. Parolia, Y. Wang, and C. D. Helgason, "The non-coding transcriptome as a dynamic regulator of cancer metastasis," *Cancer and*

Metastasis Reviews, vol. 33, no. 1. pp. 1–16, 2014.

[59] B. M. Berger, B. Levin, and R. J. Hilsden, "Multitarget stool DNA for colorectal cancer screening: A review and commentary on the United States Preventive Services Draft Guidelines," *World J. Gastrointest. Oncol.*, vol. 8, no. 5, p. 450, 2016.

[60] B. Molnár, K. Tóth, B. K. Barták, and Z. Tulassay, "Plasma methylated septin 9: A colorectal cancer screening marker," *Expert Review of Molecular Diagnostics*, vol. 15, no. 2. pp. 171–184, 2015.

[61] I. J. Jacobs and U. Menon, "Progress and challenges in screening for early detection of ovarian cancer.," *Mol. Cell. Proteomics*, vol. 3, no. 4, pp. 355–66, 2004.

[62] S. J. Skates *et al.*, "Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers.," *Cancer*, vol. 76, no. 10 Suppl, pp. 2004–2010, 1995.

[63] S. J. Skates, D. K. Pauler, and I. J. Jacobs, "Screening based on the risk of cancer calculation from bayesian hierarchical changepoint and mixture models of longitudinal markers," *J. Am. Stat. Assoc.*, vol. 96, no. 454, pp. 429–439, 2001.

[64] I. JACOBS, D. ORAM, J. FAIRBANKS, J. TURNER, C. FROST, and J. G. GRUDZINSKAS, "A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer," *BJOG An Int. J. Obstet. Gynaecol.*, vol. 97, no. 10, pp. 922–929, 1990.

[65] P. Tangkijvanich *et al*., "Clinical characteristics and prognosis of hepatocellular carcinoma: analysis based on serum alpha-fetoprotein levels.," *J. Clin. Gastroenterol*., vol. 31, no. 4, pp. 302–308, 2000.

[66] A. J. Hanazaki K, Kajikawa S, Koide N, Adachi W, "Prognostic factors after hepatic resection for hepatocellular carcinoma with hepatitis C viral infection: univariate and multivariate analysis," *Am J Gastroenterol*, vol. 96, no. 4, pp. 1243–1250, 2001.

[67] N. Personeni *et al*., "Usefulness of alpha-fetoprotein response in patients treated with sorafenib for advanced hepatocellular carcinoma," *J. Hepatol*., vol. 57, no. 1, pp. 101–107, 2012.

[68] R. J. De Haas *et al*., "Tumor marker evolution: Comparison with imaging for assessment of response to chemotherapy in patients with colorectal liver metastases," *Ann. Surg. Oncol*., vol. 17, no. 4, pp. 1010–1023, 2010.

[69] T. Eisen *et al*., "Sorafenib in advanced melanoma: A Phase II randomised discontinuation trial analysis," *Br. J. Cancer*, vol. 95, no. 5, pp. 581–586, 2006.

[70] K. T. Flaherty *et al*., "Inhibition of Mutated, Activated BRAF in Metastatic Melanoma," *N. Engl. J. Med*., vol. 363, no. 9, pp. 809–819, 2010.

[71] R. Kefford *et al*., "Phase I/II study of GSK2118436, a selective inhibitor of oncogenic mutant BRAF kinase, in patients with metastatic melanoma and other solid tumors.," *J Clin Oncol*, vol. 28:15s, p. (suppl; abstr 8503), 2010.

[72] M. R. Cooperberg *et al*., "The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy.," *J. Urol*., vol. 173, no. 6, pp. 1938–42, 2005.

[73] S. H. Garg A, Ahmed S, Sinha A, "Tumor markers – Its advantages and limitations in diagnosis of oral cancer," *Univ J Dent Sci*, vol. 1, pp. 42–45, 2015.

[74] M. Matsumoto *et al*., "Predicting tumor metastasis in patients with oral cancer by means of the proliferation marker Ki67.," *J. Oral Sci*., vol. 41, no. 2, pp. 53–56, 1999.

[75] J. L. Wilder *et al*., "Clinical implications of a rising serum CA-125 within the normal range in patients with epithelial ovarian cancer: A preliminary investigation," in *Gynecologic Oncology*, 2003, vol. 89, no. 2, pp. 233–235.

[76] C. M. Mery, A. Duarte-Rojo, F. Paz-Pineda, E. Gomez, and G. Robles-Diaz, "[Does cholestasis change the clinical usefulness of CA 19-9 in pacreatobiliary cancer?]," *Rev. Investig. Clin*., vol. 53, no. 6, pp. 511–517, 2001.

[77] A. Frena, "SPan-1 and exocrine pancreatic carcinoma. The clinical role of a new tumor marker," *Int. J. Biol. Markers*, vol. 16, no. 3, pp. 189–197, 2001.

[78] R. Harris and L. S. Kinsinger, "Principles of screening for cancer," in *Oncology: An Evidence-Based Approach*, 2006, pp. 161–176.

[79] V. Sanz-Moreno and C. J. Marshall, "The plasticity of cytoskeletal dynamics

underlying neoplastic cell migration," *Current Opinion in Cell Biology*, vol. 22, no. 5. pp. 690–696, 2010.

[80] D. A. Lauffenburger and A. F. Horwitz, "Cell migration: A physically integrated molecular process," *Cell*, vol. 84, no. 3. pp. 359–369, 1996.

[81] P. Friedl and K. Wolf, "Plasticity of cell migration: A multiscale tuning model," *Journal of Cell Biology*, vol. 188, no. 1. pp. 11–19, 2010.

[82] A. J. Ridley, "Rho GTPases and cell migration," *J Cell Sci*, vol. 114, no. Pt 15, pp. 2713–2722, 2001.

[83] A. G. Clark and D. M. Vignjevic, "Modes of cancer cell invasion and the role of the microenvironment," *Current Opinion in Cell Biology*, vol. 36. pp. 13–22, 2015.

[84] W. G. Jiang, R. P. Bryce, D. F. Horrobin, and R. E. Mansel, "Regulation of tight junction permeability and occludin expression by polyunsaturated fatty acids," *Biochem. Biophys. Res. Commun.*, vol. 244, no. 2, pp. 414–420, 1998.

[85] T. A. Martin, R. E. Mansel, and W. G. Jiang, "Antagonistic effect of NK4 on HGF/SF induced changes in the transendothelial resistance (TER) and paracellular permeability of human vascular endothelial cells," *J. Cell. Physiol.*, vol. 192, no. 3, pp. 268–275, 2002.

[86] T. A. Martin, R. E. Mansel, and W. G. Jiang, "Loss of occludin leads to the progression of human breast cancer," *Int. J. Mol. Med.*, vol. 26, no. 5, pp. 723–734, 2010.

[87] T. A. Martin, G. Watkins, R. E. Mansel, and W. G. Jiang, "Loss of tight junction plaque molecules in breast cancer tissues is associated with a poor prognosis in patients with breast cancer," *Eur. J. Cancer*, vol. 40, no. 18, pp. 2717–2725, 2004.

[88] A. J. Knights, A. P. W. Funnell, M. Crossley, and R. C. M. Pearson, "Holding Tight: Cell Junctions and Cancer Spread.," *Trends Cancer Res.*, vol. 8, pp. 61–69, 2012.

[89] S. Yonemura, M. Itoh, A. Nagafuchi, and S. Tsukita, "Cell-to-cell adherens junction formation and actin filament organization: similarities and differences between non-polarized fibroblasts and polarized epithelial cells.," *J. Cell Sci.*, vol. 108, pp. 127–142, 1995.

[90] A. Hartsock and W. J. Nelson, "Adherens and tight junctions: Structure, function and connections to the actin cytoskeleton," *Biochimica et Biophysica Acta - Biomembranes*, vol. 1778, no. 3. pp. 660–669, 2008.

[91] M. Peifer, S. Berg, and A. B. Reynolds, "A repeating amino acid motif shared by proteins with diverse cellular roles," *Cell*, vol. 76, no. 5. pp. 789–791, 1994.

[92] M. R. Groves and D. Barford, "Topological characteristics of helical repeat proteins," *Current Opinion in Structural Biology*, vol. 9, no. 3. pp. 383–389, 1999.

[93] D. Garrod and M. Chidgey, "Desmosome structure, composition and function," *Biochim. Biophys. Acta - Biomembr.*, vol. 1778, no. 3, pp. 572–587, 2008.

[94] S. T. Kundu *et al.*, "Plakophilin3 downregulation leads to a decrease in cell

adhesion and promotes metastasis," *Int. J. Cancer*, vol. 123, no. 10, pp. 2303–2314, 2008.

[95] R. L. Dusek and L. D. Attardi, "Desmosomes: New perpetrators in tumour suppression," *Nat. Rev. Cancer*, vol. 11, no. 5, pp. 317–323, 2011.

[96] K. A. Stauffer, "The gap junction proteins ??1-connexin (connexin-32) and ??2-connexin (connexin-26) can form heteromeric hemichannels," *J. Biol. Chem.*, vol. 270, no. 12, pp. 6768–6772, 1995.

[97] G. Kanaporis, G. Mese, L. Valiuniene, T. W. White, P. R. Brink, and V. Valiunas, "Gap junction channels exhibit connexin-specific permeability to cyclic nucleotides.," *J. Gen. Physiol.*, vol. 131, no. 4, pp. 293–305, 2008.

[98] Q. L. Liang, B. R. Wang, G. Q. Chen, G. H. Li, and Y. Y. Xu, "Clinical significance of vascular endothelial growth factor and connexin43 for predicting pancreatic cancer clinicopathologic parameters," *Med. Oncol.*, vol. 27, no. 4, pp. 1164–1170, 2010.

[99] G. Benko, B. Spajić, A. Demirović, G. Štimac, B. Kruŝlin, and D. Tomas, "Prognostic value of connexin43 expression in patients with clinically localized prostate cancer," *Prostate Cancer Prostatic Dis.*, vol. 14, no. 1, pp. 90–95, 2011.

[100] S. Sirnes *et al.*, "Connexin43 acts as a colorectal cancer tumor suppressor and predicts disease outcome," *Int. J. Cancer*, vol. 131, no. 3, pp. 570–581, 2012.

[101] I. Teleki *et al.*, "Correlations of differentially expressed gap junction connexins

cx26, cx30, cx32, cx43 and cx46 with breast cancer progression and prognosis," *PLoS One*, vol. 9, no. 11, 2014.

[102] G. Du *et al*., "Thrombocytosis and immunohistochemical expression of connexin 43 at diagnosis predict survival in advanced non-small-cell lung cancer treated with cisplatin-based chemotherapy," *Cancer Chemother. Pharmacol*., vol. 71, no. 4, pp. 893–904, 2013.

[103] A. Estecha *et al*., "Moesin orchestrates cortical polarity of melanoma tumour cells to initiate 3D invasion," *J. Cell Sci*., vol. 122, no. 19, pp. 3492–3501, 2009.

[104] R. Poincloux *et al*., "Contractility of the cell rear drives invasion of breast tumor cells in 3D Matrigel," *Proc. Natl. Acad. Sci*., vol. 108, no. 5, pp. 1943–1948, 2011.

[105] K. Wolf *et al*., "Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion," *Nat. Cell Biol*., vol. 9, no. 8, pp. 893–904, 2007.

[106] P. Friedl and S. Alexander, "Cancer invasion and the microenvironment: Plasticity and reciprocity," *Cell*, vol. 147, no. 5. pp. 992–1009, 2011.

[107] M. A. Eckert *et al*., "Twist1-Induced Invadopodia Formation Promotes Tumor Metastasis," *Cancer Cell*, vol. 19, no. 3, pp. 372–386, 2011.

[108] D. A. Murphy and S. A. Courtneidge, "The 'ins' and 'outs' of podosomes and invadopodia: Characteristics, formation and function," *Nature Reviews Molecular Cell Biology*, vol. 12, no. 7. pp. 413–426, 2011.

[109] P. M. Kulesa and L. S. Gammill, “Neural crest migration: Patterns, phases and signals,” *Developmental Biology*, vol. 344, no. 2. pp. 566–568, 2010.

[110] E. T. Roussos *et al*., “Mena invasive (MenaINV) promotes multicellular streaming motility and transendothelial migration in a mouse model of breast cancer,” *J. Cell Sci*., vol. 124, no. 13, pp. 2120–2131, 2011.

[111] C. Gaggioli *et al*., “Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells,” *Nat. Cell Biol*., vol. 9, no. 12, pp. 1392–1400, 2007.

[112] M. Iguchi, T., Aishima, S., Taketomi, A., Nishihara, Y., Fujita, N., Sanefuji, K., Maehara, Y., and Tsuneyoshi, “Extracapsular penetration is a new prognostic factor in human hepatocellular carcinoma,” *Am. J. Surg. Pathol*, vol. 32, pp. 1675–1682, 2008.

[113] M. Ishizaki *et al*., “The formation of capsule and septum in human hepatocellular carcinoma,” *Virchows Arch*., vol. 438, no. 6, pp. 574–580, 2001.

[114] B. Weigelin, G.-J. Bakker, and P. Friedl, “Third harmonic generation microscopy of cells and tissue organization.,” *J. Cell Sci*., vol. c, pp. 245–255, 2016.

[115] O. Ilina, G. J. Bakker, A. Vasaturo, R. M. Hofmann, and P. Friedl, “Two-photon laser-generated microtracks in 3D collagen lattices: Principles of MMP-dependent and -independent collective cancer cell invasion,” *Phys. Biol*., vol. 8, no. 1, 2011.

[116] S. D. Mason and J. A. Joyce, “Proteolytic networks in cancer,” *Trends in Cell*

Biology, vol. 21, no. 4. pp. 228–237, 2011.

[117] F. Sabeh, X. Y. Li, T. L. Saunders, R. G. Rowe, and S. J. Weiss, "Secreted versus membrane-anchored collagenases: Relative roles in fibroblast-dependent collagenolysis and invasion," *J. Biol. Chem.*, vol. 284, no. 34, pp. 23001–23011, 2009.

[118] H. A. Kenny, S. Kaur, L. M. Coussens, and E. Lengyel, "The initial steps of ovarian cancer cell metastasis are mediated by MMP-2 cleavage of vitronectin and fibronectin," *J. Clin. Invest.*, vol. 118, no. 4, pp. 1367–1379, 2008.

[119] N. E. Sounni *et al.*, "Stromal regulation of vessel stability by MMP14 and TGFbeta.," *Dis. Model. Mech.*, vol. 3, no. 5–6, pp. 317–332, 2010.

[120] L. J. Talbot, S. D. Bhattacharya, and P. C. Kuo, "Epithelial-mesenchymal transition, the tumor microenvironment, and metastatic behavior of epithelial malignancies," *International Journal of Biochemistry and Molecular Biology*, vol. 3, no. 2. pp. 117–136, 2012.

[121] P. Mehlen and A. Puisieux, "Metastasis: A question of life or death," *Nature Reviews Cancer*, vol. 6, no. 6. pp. 449–458, 2006.

[122] M. Mark Taketo, "Reflections on the spread of metastasis to cancer prevention," *Cancer Prevention Research*, vol. 4, no. 3. pp. 324–328, 2011.

[123] H. Wikman, R. Vessella, and K. Pantel, "Cancer micrometastasis and tumour dormancy," *APMIS*, vol. 116, no. 7–8. pp. 754–770, 2008.

[124] J. Grossmann, “Molecular mechanisms of ‘detachment-induced apoptosis--Anoikis’.,” *Apoptosis*, vol. 7, no. 3, pp. 247–260, 2002.

[125] S. P. H. Chiang, R. M. Cabrera, and J. E. Segall, “Tumor cell intravasation,” *Am. J. Physiol. - Cell Physiol.*, vol. 311, no. 1, pp. C1–C14, 2016.

[126] J. Folkman and Y. Shing, “Angiogenesis,” *Journal of Biological Chemistry*, vol. 267, no. 16. pp. 10931–10934, 1992.

[127] J. Folkman, “Fighting cancer by attacking its blood supply.,” *Scientific American*, vol. 275, no. 3. pp. 150–154, 1996.

[128] M. Ono, H. Torisu, J. Fukushi, a Nishie, and M. Kuwano, “Biological implications of macrophage infiltration in human tumor angiogenesis.,” *Cancer Chemother. Pharmacol.*, vol. 43 Suppl, no. September 1998, pp. S69-71, 1999.

[129] P. C. Brooks, “Cell adhesion molecules in angiogenesis,” *Cancer and Metastasis Reviews*, vol. 15, no. 2. pp. 187–194, 1996.

[130] N. Jouve *et al.*, “CD146 mediates VEGF-induced melanoma cell extravasation through FAK activation,” *Int. J. Cancer*, vol. 137, no. 1, pp. 50–60, 2015.

[131] Y. Kienast *et al.*, “Real-time imaging reveals the single steps of brain metastasis formation,” *Nat. Med.*, vol. 16, no. 1, pp. 116–122, 2010.

[132] K. Stoletov *et al.*, “Visualizing extravasation dynamics of metastatic tumor cells,” *J. Cell Sci.*, vol. 123, no. 13, pp. 2332–2341, 2010.

[133] B. Strilic and S. Offermanns, “Intravascular Survival and Extravasation of Tumor Cells,” *Cancer Cell*, vol. 32, no. 3. pp. 282–293, 2017.

[134] M. Forsgren, B. Råden, M. Israelsson, K. Larsson, and L. O. Hedén, “Molecular cloning and characterization of a full-length cDNA clone for human plasminogen.,” *FEBS Lett*., vol. 213, no. 2, pp. 254–60, 1987.

[135] R. H. P. Law *et al*., “The X-ray Crystal Structure of Full-Length Human Plasminogen,” *Cell Rep*., vol. 1, no. 3, pp. 185–190, 2012.

[136] E. F. Plow, J. Felez, and L. A. Miles, “Cellular regulation of fibrinolysis,” in *Thrombosis and Haemostasis*, 1991, vol. 66, no. 1, pp. 32–36.

[137] K. a Hajjar, P. C. Harpel, E. a Jaffe, and R. L. Nachman, “Binding of plasminogen to cultured human endothelial cells.,” *J. Biol. Chem*., vol. 261, no. 25, pp. 11656–11662, 1986.

[138] L. A. Miles, C. M. Dahlberg, and E. F. Plow, “The cell-binding domains of plasminogen and their function in plasma,” *J. Biol. Chem*., vol. 263, no. 24, pp. 11928–11934, 1988.

[139] P. a Andreasen, R. Egelund, and H. H. Petersen, “The plasminogen activation system in tumor growth, invasion, and metastasis.,” *Cell. Mol. Life Sci*., vol. 57, no. 1, pp. 25–40, 2000.

[140] W. B, “Primary Structure of Peptides Released during Activation of Human Plasrninogen by Urokinase,” *Eur. J. Biochem*, vol. 36, pp. 1–9, 1973.

[141] F. J. Castellino and B. N. Violand, "The fibrinolytic system-Basic considerations," *Prog. Cardiovasc. Dis.*, vol. 21, no. 4, pp. 241–254, 1979.

[142] and M. E. N. Michael B. Boffa, Wei Wang, Laszlo Bajzar, "Plasma and Recombinant Thrombin-activable Fibrinolysis Inhibitor (TAFI) and Activated TAFI Compared with Respect to Glycosylation, Thrombin/Thrombomodulin-dependent Activation, Thermal Stability, and Enzymatic Properties," *) J. Biol. Chem*, vol. 273, no. 4, pp. 2127–2135, 1998.

[143] D. Holvoet, P., Lijnen, H.R. & Collen, "A monoclonal antibody specific for lys-plasminogen," *J. Biol. Chem.*, vol. 260, pp. 12106–12111, 1985.

[144] J. C. Fredenburgh and M. E. Nesheim, "Lys-plasminogen is a significant intermediate in the activation of Glu- plasminogen during fibrinolysis in vitro," *J. Biol. Chem.*, vol. 267, no. 36, pp. 26150–26156, 1992.

[145] B. Wiman and D. Collen, "Molecular mechanism of physiological fibrinolysis," *Nature*, vol. 272, no. 5653, pp. 549–550, 1978.

[146] Y. Gong, S. O. Kim, J. Felez, D. K. Grella, F. J. Castellino, and L. A. Miles, "Conversion of Glu-Plasminogen to Lys-Plasminogen Is Necessary for Optimal Stimulation of Plasminogen Activation on the Endothelial Cell Surface," *J. Biol. Chem.*, vol. 276, no. 22, pp. 19078–19083, 2001.

[147] J. M. Edelberg, J. J. Enghild, S. V. Pizzo, and M. Gonzalez-Gronow, "Neonatal plasminogen displays altered cell surface binding and activation kinetics: Correlation with increased glycosylation of the protein," *J. Clin. Invest.*, vol. 86,

no. 1, pp. 107–112, 1990.

[148] S. W. Hall, S. R. Vandenberg, and S. L. Gonias, “Plasminogen carbohydrate side chains in receptor binding and enzyme activation: A study of C6 glioma cells and primary cultures of rat hepatocytes,” *J. Cell. Biochem.*, vol. 43, no. 3, pp. 213–227, 1990.

[149] R. Aisina, L. Mukhametova, K. Gershkovich, and S. Varfolomeyev, “The role of carbohydrate side chains of plasminogen in its activation by staphylokinase,” *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1725, no. 3, pp. 370–376, 2005.

[150] A. Fischer, “No Title,” *Arch. Anat. Entwicklungsmech. Org*, vol. 104, pp. 210–219, 1925.

[151] Y. A. DeClerck, S. Imren, A. M. P. Montgomery, B. M. Mueller, R. A. Reisfeld, and W. E. Laug, “Proteases and protease inhibitors in tumor progression.,” *Adv. Exp. Med. Biol.*, vol. 425, pp. 89–97, 1997.

[152] H. M. Zhou, a Nichols, P. Meda, and J. D. Vassalli, “Urokinase-type plasminogen activator and its receptor synergize to promote pathogenic proteolysis.,” *EMBO J.*, vol. 19, no. 17, pp. 4817–4826, 2000.

[153] H. Zhai *et al.*, “Annexin A2 Promotes Glioma Cell Invasion and Tumor Progression,” *Neurobiol. Dis.*, vol. 31, no. 40, pp. 14346–60, 2011.

[154] G. MURPHY *et al.*, “Mechanisms for pro matrix metalloproteinase activation,” *APMIS*, vol. 107, no. 1–6, pp. 38–44, 1999.

[155] Y. S. Kim and T. H. Joh, “Matrix metalloproteinases, new insights into the understanding of neurodegenerative disorders,” *Biomol. Ther.*, vol. 20, no. 2, pp. 133–143, 2012.

[156] M. A. A. Parry, X. C. Zhang, and W. Bode, “Molecular mechanisms of plasminogen activation: Bacterial cofactors provide clues,” *Trends in Biochemical Sciences*, vol. 25, no. 2. pp. 53–59, 2000.

[157] E. S. Ford, “Lippincott Williams & Wilkins,” *Lippincott Williams & Wilkins*, vol. 13, no. 5, pp. 561–568, 2015.

[158] V. Ellis, M. F. Scully, and V. V. Kakkar, “Plasminogen activation initiated by single-chain urokinase-type plasminogen activator. Potentiation by U937 monocytes,” *J. Biol. Chem.*, vol. 264, no. 4, pp. 2185–2188, 1989.

[159] F. Blasi and P. Carmeliet, “uPAR: A versatile signalling orchestrator,” *Nature Reviews Molecular Cell Biology*, vol. 3, no. 12. pp. 932–943, 2002.

[160] P. LC, “Kinetics of reciprocal pro-urokinase/plasminogen activation-- stimulation by a template formed by the urokinase receptor bound to poly(D-lysine).,” *Eur. J. Biochem*, vol. 245, no. 2, pp. 316–323, 1997.

[161] S.-H. Kwak *et al.*, “The kringle domain of urokinase-type plasminogen activator potentiates LPS-induced neutrophil activation through interaction with {alpha}V{beta}3 integrins.,” *J. Leukoc. Biol.*, vol. 78, no. 4, pp. 937–45, 2005.

[162] T. Tarui *et al.*, “Direct interaction of the kringle domain of urokinase-type

plasminogen activator (uPA) and integrin αvβ3 induces signal transduction and enhances plasminogen activation," *Thromb. Haemost.*, vol. 95, no. 3, pp. 524–534, 2006.

[163] M. R. Mastronicola, M. P. Stoppelli, A. Migliaccio, F. Auricchio, and F. Blasi, "Serine phosphorylation of biosynthetic pro-urokinase from human tumor cells," *FEBS Lett.*, vol. 266, no. 1–2, pp. 109–114, 1990.

[164] B. S. et al Rabbani SA, Mazar AP, "Structural requirements for the growth factor activity of the aminoterminal domain of urokinase," *J Biol Chem*, vol. 267, no. 1415, p. 1416, 1992.

[165] K. List *et al.*, "Plasminogen-independent initiation of the pro-urokinase activation cascade in vivo. Activation of pro-urokinase by glandular kallikrein (mGK-6) in plasminogen-deficient mice," *Biochemistry*, vol. 39, no. 3, pp. 508–515, 2000.

[166] H. W. Smith and C. J. Marshall, "Regulation of cell signalling by uPAR," *Nature Reviews Molecular Cell Biology*, vol. 11, no. 1. pp. 23–36, 2010.

[167] Z. C. Mauro MA, Murphy K, Thomson K, *Image-Guided Intervention*. 2008.

[168] E. G. Levin, U. Marzec, J. Anderson, and L. A. Harker, "Thrombin stimulates tissue plasminogen activator release from cultured human endothelial cells.," *J. Clin. Invest.*, vol. 74, no. 6, pp. 1988–95, 1984.

[169] Y. P. Wu *et al.*, "The tissue plasminogen activator (tPA)/plasmin extracellular proteolytic system regulates seizure-induced hippocampal mossy fiber outgrowth

through a proteoglycan substrate," *J. Cell Biol.*, vol. 148, no. 6, pp. 1295–1304, 2000.

[170] G. H. Janicke F, Schmitt M, "Clinical relevance of the urokinase-type and tissue-type plasminogen activators and of their type 1 inhibitor in breast cancer," *Semin Thromb Hemost*, vol. 17, pp. 303–312, 1991.

[171] M. A. Kwaan HC, Lo R, "On the production of plasma fibrinolytic activity within veins," *Clin Sci*, vol. 16, no. 2, pp. 241–253, 1957.

[172] P. Carmeliet *et al.*, "Physiological consequences of loss of plasminogen activator gene function in mice.," *Nature*, vol. 368, no. 6470, pp. 419–24, 1994.

[173] P. Tabrizi *et al.*, "Tissue plasminogen activator (tPA) deficiency exacerbates cerebrovascular fibrin deposition and brain injury in a murine stroke model studies in tPA-deficient mice and wild-type mice on a matched genetic background," *Arterioscler. Thromb. Vasc. Biol.*, vol. 19, no. 11, pp. 2801–2806, 1999.

[174] F. Siddiqi, T. M. Odrljin, P. J. Fay, C. Cox, and C. W. Francis, "Binding of tissue-plasminogen activator to fibrin: effect of ultrasound.," *Blood*, vol. 91, no. 6, pp. 2019–2025, 1998.

[175] S. S. Husain, "Fibrin affinity of urokinase-type plasminogen activator. Evidence that Zn2+mediates strong and specific interaction of single-chain urokinase with fibrin," *J. Biol. Chem.*, vol. 268, no. 12, pp. 8574–8579, 1993.

[176] H. E. Runge MS, Bode C, Matsueda GR, "Antibody-enhanced thrombolysis:

targeting of tissue plasminogen activator in vivo," *Proc Natl Acad Sci U S A*, vol. 84, no. 21, pp. 7659–7662, 1987.

[177] G. Kassam *et al*., "The p11 subunit of the annexin II tetramer plays a key role in the stimulation of t-PA-dependent plasminogen activation," *Biochemistry*, vol. 37, no. 48, pp. 16958–16966, 1998.

[178] C. Thelwell and C. Longstaff, "The regulation by fibrinogen and fibrin of tissue plasminogen activator kinetics and inhibition by plasminogen activator inhibitor 1.," *J. Thromb. Haemost*., vol. 5, no. 4, pp. 804–11, 2007.

[179] M. G. Rasch, I. K. Lund, C. E. Almasi, and G. Hoyer-Hansen, "Intact and cleaved uPAR forms: diagnostic and prognostic value in cancer.," *Front. Biosci*., vol. 13, pp. 6752–62, 2008.

[180] R. H. P. Law *et al*., "An overview of the serpin superfamily," *Genome Biology*, vol. 7, no. 5. 2006.

[181] S. Janciauskiene, "Conformational properties of serine proteinase inhibitors (serpins) confer multiple pathophysiological roles," *Biochimica et Biophysica Acta - Molecular Basis of Disease*, vol. 1535, no. 3. pp. 221–235, 2001.

[182] J. Boncela, P. Przygodzka, I. Papiewska-Pajak, E. Wyroba, M. Osinska, and C. S. Cierniewski, "Plasminogen activator inhibitor type 1 interacts with alpha3 subunit of proteasome and modulates its activity," *J.Biol.Chem*., vol. 286, no. 1083–351X (Electronic), pp. 6820–6831, 2011.

[183] J. A. Van Mourik, D. A. Lawrence, and D. J. Loskutoff, “Purification of an inhibitor of plasminogen activator (antiactivator) synthesized by endothelials cells,” *J. Biol. Chem.*, vol. 259, no. 23, pp. 14914–14921, 1984.

[184] K. D. Howard EW, “Characterization of the receptor for protease nexin-I:protease complexes on human fibroblasts,” *J Cell Physiol.*, vol. 131, no. 2, pp. 276–283, 1987.

[185] R. L. Medcalf and S. J. Stasinopoulos, “The undecided serpin: The ins and outs of plasminogen activator inhibitor type 2,” in *FEBS Journal*, 2005, vol. 272, no. 19, pp. 4858–4867.

[186] M. Moroi and N. Aoki, “Isolation and characterization of alpha2-plasmin inhibitor from human plasma. {A} novel proteinase inhibitor which inhibits activator-induced clot lysis.,” *J Biol Chem*, vol. 251, no. 19, pp. 5956–5965, 1976.

[187] G. A. Hastings *et al.*, “Neuroserpin, a brain-associated inhibitor of tissue plasminogen activator is localized primarily in neurons. Implications for the regulation of motor learning and neuronal survival,” *J. Biol. Chem.*, vol. 272, no. 52, pp. 33062–33067, 1997.

[188] T. Y. Takada A, “Interaction of plasmin with α2-macroglobulin and α2-antiplasmin in the presence and absence of tranexamic acid,” *Thromb. Res.*, vol. 18, no. 1–2, pp. 1–15, 1980.

[189] K. E. & V. J. Belin D, Wohlwend A, Schleuning WD, “Facultative polypeptide translocation allows a single mRNA to encode the secreted and cytosolic forms of

plasminogen activators inhibitor 2," *EMBO J*, vol. 8, pp. 3287–3294, 1989.

[190] Colman RW, *Hemostasis and thrombosis: basic principles and clinical practice*. 2006.

[191] J. Félez, L. A. Miles, P. Fábregas, M. Jardí, E. F. Plow, and R. H. Lijnen, "Characterization of cellular binding sites and interactive regions within reactants required for enhancement of plasminogen activation by tPA on the surface of leukocytic cells," *Thromb. Haemost.*, vol. 76, no. 4, pp. 577–584, 1996.

[192] L. A. Miles and E. F. Plow, "Receptor mediated binding of the fibrinolytic components, plasminogen and urokinase, to peripheral blood cells," *Thromb. Haemost.*, vol. 58, no. 3, pp. 936–942, 1987.

[193] H. Bai, N. Baik, W. B. Kiosses, S. Krajewski, L. A. Miles, and R. J. Parmer, "The novel plasminogen receptor, plasminogen receptorKT(Plg-RKT), regulates catecholamine release," *J. Biol. Chem.*, vol. 286, no. 38, pp. 33125–33133, 2011.

[194] A. S. Chauhan *et al.*, "Moonlighting glycolytic protein glyceraldehyde-3-phosphate dehydrogenase (GAPDH): An evolutionarily conserved plasminogen receptor on mammalian cells," *FASEB J.*, vol. 31, no. 6, pp. 2638–2648, 2017.

[195] L. A. Miles, E. F. Plow, D. M. Waisman, and R. J. Parmer, "Plasminogen receptors," *Journal of Biomedicine and Biotechnology*, vol. 2012, 2012.

[196] E. F. Plow, L. Doeuvre, and R. Das, "So many plasminogen receptors: Why?," *Journal of Biomedicine and Biotechnology*, vol. 2012. 2012.

[197] L. a Miles, S. B. Hawley, N. Baik, N. M. Andronicos, F. J. Castellino, and R. J. Parmer, "Plasminogen receptors: the sine qua non of cell surface plasminogen activation.," *Front. Biosci.*, vol. 10, no. 33, pp. 1754–1762, 2005.

[198] R. Das, E. Pluskota, and E. F. Plow, "Plasminogen and Its Receptors as Regulators of Cardiovascular Inflammatory Responses," *Trends in Cardiovascular Medicine*, vol. 20, no. 4. pp. 120–124, 2010.

[199] A. Godier and B. J. Hunt, "Plasminogen receptors and their role in the pathogenesis of inflammatory, autoimmune and malignant disease," *Journal of Thrombosis and Haemostasis*, vol. 11, no. 1. pp. 26–34, 2013.

[200] T. A. Hembrough, L. Li, and S. L. Gonias, "Cell-surface cytokeratin 8 is the major plasminogen receptor on breast cancer cells and is required for the accelerated activation of cell- associated plasminogen by tissue-type plasminogen activator," *J. Biol. Chem.*, vol. 271, no. 41, pp. 25684–25691, 1996.

[201] M. Wygrecka *et al.*, "Enolase-1 promotes plasminogen-mediated recruitment of monocytes to the acutely inflamed lung," *Blood*, vol. 113, no. 22, pp. 5588–5598, 2009.

[202] L. A. Miles *et al.*, "The plasminogen receptor, Plg-RKT, and macrophage function," *Journal of Biomedicine and Biotechnology*, vol. 2012. 2012.

[203] S. T. Sundström B, "A two-site enzyme-linked immunosorbent assay for cytokeratin 8," *Int J Cancer*, vol. 46, no. 4, pp. 604–607, 1990.

[204] T. Fillies *et al*., “Cytokeratin 8/18 expression indicates a poor prognosis in squamous cell carcinomas of the oral cavity,” *BMC Cancer*, vol. 6, 2006.

[205] A. Page-McCaw, A. J. Ewald, and Z. Werb, “Matrix metalloproteinases and the regulation of tissue remodelling,” *Nature Reviews Molecular Cell Biology*, vol. 8, no. 3. pp. 221–233, 2007.

[206] W. C. & L.-B. Y. Parks WC, “Matrix metalloproteinases as modulators of inflammation and innate immunity,” *Nat Rev Immunol*, vol. 4, pp. 617–629, 2004.

[207] H. Nagase, R. Visse, and G. Murphy, “Structure and function of matrix metalloproteinases and TIMPs,” *Cardiovascular Research*, vol. 69, no. 3. pp. 562–573, 2006.

[208] A. M. Weaver, “Invadopodia: Specialized cell structures for cancer invasion,” *Clinical and Experimental Metastasis*, vol. 23, no. 2. pp. 97–105, 2006.

[209] M. Suzuki, G. Raab, M. A. Moses, C. A. Fernandez, and M. Klagsbrun, “Matrix metalloproteinase-3 releases active heparin-binding EGF-like growth factor by cleavage at a specific juxtamembrane site,” *J. Biol. Chem*., vol. 272, no. 50, pp. 31730–31737, 1997.

[210] T. Maretzky *et al*., “ADAM10 mediates E-cadherin shedding and regulates epithelial cell-cell adhesion, migration, and -catenin translocation,” *Proc. Natl. Acad. Sci*., vol. 102, no. 26, pp. 9182–9187, 2005.

[211] G. Murphy, “The ADAMs: Signalling scissors in the tumour microenvironment,”

Nature Reviews Cancer, vol. 8, no. 12. pp. 929–941, 2008.

[212] N. Rocks *et al*., "Emerging roles of ADAM and ADAMTS metalloproteinases in cancer," *Biochimie*, vol. 90, no. 2. pp. 369–379, 2008.

[213] K. Brew and H. Nagase, "The tissue inhibitors of metalloproteinases (TIMPs): An ancient family with structural and functional diversity," *Biochimica et Biophysica Acta - Molecular Cell Research*, vol. 1803, no. 1. pp. 55–71, 2010.

[214] A. Amour *et al*., "The in vitro activity of ADAM-10 is inhibited by TIMP-1 and TIMP-3," *FEBS Lett*., vol. 473, no. 3, pp. 275–279, 2000.

[215] H. Zhao *et al*., "Differential Inhibition of Membrane Type 3 (MT3)-Matrix Metalloproteinase (MMP) and MT1-MMP by Tissue Inhibitor of Metalloproteinase (TIMP)-2 and TIMP-3 Regulates Pro-MMP-2 Activation," *J. Biol. Chem*., vol. 279, no. 10, pp. 8592–8601, 2004.

[216] M. Fonović and B. Turk, "Cysteine cathepsins and extracellular matrix degradation," *Biochimica et Biophysica Acta - General Subjects*, vol. 1840, no. 8. pp. 2560–2570, 2014.

[217] B. Sobotič *et al*., "Proteomic Identification of Cysteine Cathepsin Substrates Shed from the Surface of Cancer Cells," *Mol. Cell. Proteomics*, vol. 14, no. 8, pp. 2213–2228, 2015.

[218] L. Akkari *et al*., "Distinct functions of macrophage-derived and cancer cell-derived cathepsin Z combine to promote tumor malignancy via interactions with the

extracellular matrix," *Genes Dev*., vol. 28, no. 19, pp. 2134–2150, 2014.

[219] O. Vasiljeva *et al*., "Tumor cell-derived and macrophage-derived cathepsin B promotes progression and lung metastasis of mammary cancer," *Cancer Res*., vol. 66, no. 10, pp. 5242–5250, 2006.

[220] J. A. Joyce *et al*., "Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis," *Cancer Cell*, vol. 5, no. 5, pp. 443–453, 2004.

[221] L. Sevenich and J. A. Joyce, "Pericellular proteolysis in cancer," *Genes and Development*, vol. 28, no. 21. pp. 2331–2347, 2014.

[222] C. a Borgoño, I. P. Michael, and E. P. Diamandis, "Human tissue kallikreins: physiologic roles and applications in cancer.," *Mol. Cancer Res*., vol. 2, no. 5, pp. 257–280, 2004.

[223] R. Donato, "S100: A multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles," *International Journal of Biochemistry and Cell Biology*, vol. 33, no. 7. pp. 637–668, 2001.

[224] A. C. Rintala-Dempsey, A. Rezvanpour, and G. S. Shaw, "S100-annexin complexes--structural insights," *FEBS J*., vol. 275, no. 20, pp. 4956–4966, 2008.

[225] R. H. Kretsinger and C. E. Nockolds, "Carp muscle calcium-binding protein. II. Structure determination and general description.," *J. Biol. Chem*., vol. 248, no. 9, pp. 3313–3326, 1973.

[226] M. C. Schaub and C. W. Heizmann, “Calcium, troponin, calmodulin, S100 proteins: From myocardial basics to new therapeutic strategies,” *Biochem. Biophys. Res. Commun.*, vol. 369, no. 1, pp. 247–264, 2008.

[227] I. Marenholz, C. W. Heizmann, and G. Fritz, “S100 proteins in mouse and man: From evolution to function and pathology (including an update of the nomenclature),” *Biochemical and Biophysical Research Communications*, vol. 322, no. 4. pp. 1111–1122, 2004.

[228] Y. Zhou, W. Yang, M. Kirberger, H. W. Lee, G. Ayalasomayajula, and J. J. Yang, “Prediction of EF-hand calcium-binding proteins and analysis of bacterial EF-hand proteins,” *Proteins Struct. Funct. Genet.*, vol. 65, no. 3, pp. 643–655, 2006.

[229] S. E. Permyakov, R. G. Ismailov, B. Xue, A. I. Denesyuk, V. N. Uversky, and E. a Permyakov, “Intrinsic disorder in S100 proteins.,” *Mol. Biosyst.*, vol. 7, no. 7, pp. 2164–80, 2011.

[230] S. Réty *et al.*, “The crystal structure of a complex of p11 with the annexin II N-terminal peptide,” *Nat. Struct. Biol.*, vol. 6, no. 1, pp. 89–95, 1999.

[231] M. T. Christen, P. Frank, J. Schaller, and M. Llinás, “Human plasminogen kringle 3: Solution structure, functional insights, phylogenetic landscape,” *Biochemistry*, vol. 49, no. 33, pp. 7131–7150, 2010.

[232] G. Kassam *et al.*, “The role of annexin II tetramer in the activation of plasminogen,” *J. Biol. Chem.*, vol. 273, no. 8, pp. 4790–4799, 1998.

[233] P. A. Madureira, P. A. O'Connell, A. P. Surette, V. A. Miller, and D. M. Waisman, "The biochemistry and regulation of S100A10: A multifunctional plasminogen receptor involved in oncogenesis," *Journal of Biomedicine and Biotechnology*, vol. 2012. 2012.

[234] K.-S. Choi, D. K. Fogg, C.-S. Yoon, and D. M. Waisman, "p11 regulates extracellular plasmin production and invasiveness of HT1080 fibrosarcoma cells.," *FASEB J.*, vol. 17, no. 2, pp. 235–46, 2003.

[235] L. Zhang, D. K. Fogg, and D. M. Waisman, "RNA Interference-mediated Silencing of the S100A10 Gene Attenuates Plasmin Generation and Invasiveness of Colo 222 Colorectal Cancer Cells," *J. Biol. Chem.*, vol. 279, no. 3, pp. 2053–2062, 2004.

[236] P. A. O'Connell, A. P. Surette, R. S. Liwski, P. Svenningsson, and D. M. Waisman, "S100A10 regulates plasminogen-dependent macrophage invasion," *Blood*, vol. 116, no. 7, pp. 1136–1146, 2010.

[237] P. a O'Connell, P. a Madureira, J. N. Berman, R. S. Liwski, and D. M. Waisman, "Regulation of S100A10 by the PML-RAR-alpha oncoprotein.," *Blood*, vol. 117, no. 15, pp. 4095–4105, 2011.

[238] K. D. Phipps, A. P. Surette, P. A. O'Connell, and D. M. Waisman, "Plasminogen receptor S100A10 is essential for the migration of tumor-promoting macrophages into tumor sites," *Cancer Res.*, vol. 71, no. 21, pp. 6676–6683, 2011.

[239] P. A. Madureira, A. P. Surette, K. D. Phipps, M. A. S. Taboski, V. A. Miller, and

D. M. Waisman, “The role of the annexin A2 heterotetramer in vascular fibrinolysis,” *Blood*, vol. 118, no. 18. pp. 4789–4797, 2011.

[240] A. P. Surette, P. A. Madureira, K. D. Phipps, V. A. Miller, P. Svenningsson, and D. M. Waisman, “Regulation of fibrinolysis by S100A10 in vivo,” *Blood*, vol. 118, no. 11, pp. 3172–3181, 2011.

[241] M. Bydoun and D. M. Waisman, “On the contribution of S100A10 and annexin A2 to plasminogen activation and oncogenesis: an enduring ambiguity.,” *Future Oncol.*, vol. 10, no. 15, pp. 2469–79, 2014.

[242] A. Bharadwaj, M. Bydoun, R. Holloway, and D. Waisman, “Annexin A2 heterotetramer: Structure and function,” *International Journal of Molecular Sciences*, vol. 14, no. 3. pp. 6259–6305, 2013.

[243] R. G. Sitrin *et al.*, “Lipid Raft Compartmentalization of Urokinase Receptor Signaling in Human Neutrophils,” *Am. J. Respir. Cell Mol. Biol.*, vol. 30, no. 2, pp. 233–241, 2004.

[244] V. Gerke and S. E. Moss, “Annexins: From Structure to Function,” *Physiol. Rev.*, vol. 82, no. 2, pp. 331–371, 2002.

[245] C. Longstaff, C. Thelwell, S. C. Williams, M. M. C. G. Silva, L. Szabó, and K. Kolev, “The interplay between tissue plasminogen activator domains and fibrin structures in the regulation of fibrinolysis: Kinetic and microscopic studies,” *Blood*, vol. 117, no. 2, pp. 661–668, 2011.

[246] D. K. Fogg *et al*., "The p11 subunit of annexin II heterotetramer is regulated by basic carboxypeptidase," *Biochemistry*, vol. 41, no. 15, pp. 4953–4961, 2002.

[247] S. Oliferenko *et al*., "Analysis of CD44-containing lipid rafts: Recruitment of annexin II and stabilization by the actin cytoskeleton," *J. Cell Biol*., vol. 146, no. 4, pp. 843–854, 1999.

[248] K. L. He *et al*., "Endothelial cell annexin A2 regulates polyubiquitination and degradation of its binding partner S100A10/p11," *J. Biol. Chem*., vol. 283, no. 28, pp. 19192–19200, 2008.

[249] Y. Hou *et al*., "Annexin A2 regulates the levels of plasmin, S100A10 and fascin in L5178Y cells," *Cancer Invest*., vol. 26, no. 8, pp. 809–815, 2008.

[250] A. Puisieux, J. Ji, and M. Ozturk, "Annexin II up-regulates cellular levels of p11 protein by a post-translational mechanisms," *Biochem J*, vol. 313 (Pt 1, pp. 51–55, 1996.

[251] A. Chetcuti *et al*., "Loss of Annexin II heavy and light chains in prostate cancer and its precursors," *Cancer Res*., vol. 61, no. 17, pp. 6331–6334, 2001.

[252] N. Zobiack, U. Rescher, C. Ludwig, D. Zeuschner, and V. Gerke, "The annexin 2/S100A10 complex controls the distribution of transferrin receptor-containing recycling endosomes.," *Mol. Biol. Cell*, vol. 14, no. 12, pp. 4896–908, 2003.

[253] T. J. MacLeod, M. Kwon, N. R. Filipenko, and D. M. Waisman, "Phospholipid-associated annexin A2-S100A10 heterotetramer and its subunits. Characterization

of the interaction with tissue plasminogen activator, plasminogen, and plasmin," *J. Biol. Chem.*, vol. 278, no. 28, pp. 25577–25584, 2003.

[254] E. Kube, T. Becker, K. Weber, and V. Gerke, "Protein-protein interaction studied by site-directed mutagenesis. Characterization of the annexin II-binding site on p11, a member of the S100 protein family," *J. Biol. Chem.*, vol. 267, no. 20, pp. 14175–14182, 1992.

[255] T. Becker, K. Weber, and N. Johnsson, "Protein-protein recognition via short amphiphilic helices; a mutational analysis of the binding site of annexin II for p11.," *EMBO J.*, vol. 9, no. 13, pp. 4207–13, 1990.

[256] N. Johnsson, G. Marriott, and K. Weber, "P36, the Major Cytoplasmic Substrate of Src Tyrosine Protein Kinase, Binds To Its P11 Regulatory Subunit Via a Short Amino-Terminal Amphiphatic Helix.," *EMBO J.*, vol. 7, no. 8, pp. 2435–2442, 1988.

[257] S. L. Fitzpatrick, G. Kassam, K. S. Choi, H. M. Kang, D. K. Fogg, and D. M. Waisman, "Regulation of plasmin activity by annexin II tetramer," *Biochemistry*, vol. 39, no. 5, pp. 1021–1028, 2000.

[258] K. Hajjar, "Annexin II A Mediator of the Plasmin /Plasminogen Activator System," *Trends Cardiovasc. Med.*, vol. 9, no. 5, pp. 128–138, 1999.

[259] M. Kwon, T. J. MacLeod, Y. Zhang, and D. M. Waisman, "S100A10, annexin A2, and annexin a2 heterotetramer as candidate plasminogen receptors.," *Front. Biosci.*, vol. 10, pp. 300–25, 2005.

[260] X. Yang, N. C. Popescu, and D. B. Zimonjic, “DLC1 interaction with S100A10 mediates inhibition of in vitro cell invasion and tumorigenicity of lung cancer cells through a rhogap-independent mechanism,” *Cancer Res*., vol. 71, no. 8, pp. 2916–2925, 2011.

[261] F. Hessner *et al*., “CC chemokine receptor 10 cell surface presentation in melanocytes is regulated by the novel interaction partner S100A10,” *Sci. Rep*., vol. 6, 2016.

[262] D. I. Jarmin *et al*., “Cutting edge: identification of the orphan receptor G-protein-coupled receptor 2 as CCR10, a specific receptor for the chemokine ESkine.,” *J. Immunol*., vol. 164, no. 7, pp. 3460–4, 2000.

[263] T. Chehab *et al*., “A novel Munc13-4/S100A10/annexin A2 complex promotes Weibel–Palade body exocytosis in endothelial cells,” *Mol. Biol. Cell*, vol. 28, no. 12, pp. 1688–1700, 2017.

[264] Y. Da Chen *et al*., “S100A10 Regulates ULK1 Localization to ER–Mitochondria Contact Sites in IFN-γ-Triggered Autophagy,” *J. Mol. Biol*., vol. 429, no. 1, pp. 142–157, 2017.

[265] R. C. Russell *et al*., “ULK1 induces autophagy by phosphorylating Beclin-1 and activating VPS34 lipid kinase,” *Nat. Cell Biol*., vol. 15, no. 7, pp. 741–750, 2013.

[266] D. Papinski and C. Kraft, “Atg1 kinase organizes autophagosome formation by phosphorylating Atg9,” *Autophagy*, vol. 10, no. 7, pp. 1338–1340, 2014.

[267] J. L. Hankins *et al.*, "Ceramide 1-phosphate mediates endothelial cell invasion via the annexin a2-p11 heterotetrameric protein complex," *J. Biol. Chem.*, vol. 288, no. 27, pp. 19726–19738, 2013.

[268] G. Ozorowski, S. Milton, and H. Luecke, "Structure of a C-terminal AHNAK peptide in a 1:2:2 complex with S100A10 and an acetylated N-terminal peptide of annexin A2," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 69, no. 1, pp. 92–104, 2013.

[269] A. W. Woodham *et al.*, "The S100A10 subunit of the annexin A2 heterotetramer facilitates L2-mediated human papillomavirus infection," *PLoS One*, vol. 7, no. 8, 2012.

[270] P. Svenningsson *et al.*, "Alterations in 5-HT1B receptor function by p11 in depression-like states," *Science (80-.).*, vol. 311, no. 5757, pp. 77–80, 2006.

[271] L. Maroteaux, F. Saudou, N. Amlaiky, U. Boschert, J. L. Plassat, and R. Hen, "Mouse 5HT1B serotonin receptor: cloning, functional expression, and localization in motor control centers.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 7, pp. 3020–4, 1992.

[272] P. Meschin *et al.*, "P11 modulates calcium handling through 5-HT4R pathway in rat ventricular cardiomyocytes," *Cell Calcium*, vol. 58, no. 6, pp. 549–557, 2015.

[273] K. W. Lee *et al.*, "Alteration by p11 of mGluR5 localization regulates depression-like behaviors," *Mol. Psychiatry*, vol. 20, no. 12, pp. 1546–1556, 2015.

[274] K. Okuse *et al*., "Annexin II light chain regulates sensory neuron-specific sodium channel expression," *Nature*, vol. 417, no. 6889, pp. 653–656, 2002.

[275] S. F. J. Van de Graaf *et al*., "Functional expression of the epithelial Ca2+ channels (TRPV5 and TRPV6) requires association of the S100A10-annexin 2 complex," *EMBO J*., vol. 22, no. 7, pp. 1478–1487, 2003.

[276] L. A. Borthwick, A. Neal, L. Hobson, V. Gerke, L. Robson, and R. Muimo, "The annexin 2-S100A10 complex and its association with TRPV6 is regulated by cAMP/PKA/CnA in airway and gut epithelia," *Cell Calcium*, vol. 44, no. 2, pp. 147–157, 2008.

[277] C. Girard, N. Tinel, C. Terrenoire, G. Romey, M. Lazdunski, and M. Borsotto, "p11, an annexin II subunit, an auxiliary protein associated with the background K+channel, TASK-1," *EMBO J*., vol. 21, no. 17, pp. 4439–4448, 2002.

[278] R. R. Markwald, T. P. Fitzharris, and F. J. Manasek, "Structural development of endocardial cushions," *Am. J. Anat*., vol. 148, no. 1, pp. 85–119, 1977.

[279] R. Kalluri and R. A. Weinberg, "The basics of epithelial-mesenchymal transition," *Journal of Clinical Investigation*, vol. 119, no. 6. pp. 1420–1428, 2009.

[280] L. Vićovac and J. D. Aplin, "Epithelial-mesenchymal transition during trophoblast differentiation," *Cells Tissues Organs*, vol. 156, no. 3, pp. 202–216, 1996.

[281] J. P. Thiery and J. P. Sleeman, "Complex networks orchestrate epithelial-mesenchymal transitions," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 2.

pp. 131–142, 2006.

[282] J. L. Duband and J. P. Thiery, "Appearance and distribution of fibronectin during chick embryo gastrulation and neurulation," *Dev. Biol.*, vol. 94, no. 2, pp. 337–350, 1982.

[283] A. K. Knecht and M. Bronner-Fraser, "Induction of the neural crest: a multigene process.," *Nat. Rev. Genet.*, vol. 3, no. 6, pp. 453–461, 2002.

[284] T. Sauka-Spengler and M. Bronner-Fraser, "A gene regulatory network orchestrates neural crest formation," *Nature Reviews Molecular Cell Biology*, vol. 9, no. 7. pp. 557–568, 2008.

[285] F. Strutz *et al.*, "Identification and characterization of a fibroblast marker: FSP1," *J. Cell Biol.*, vol. 130, no. 2, pp. 393–405, 1995.

[286] M. Zeisberg *et al.*, "BMP-7 counteracts TGF-beta1-induced epithelial-to-mesenchymal transition and reverses chronic renal injury.," *Nat. Med.*, vol. 9, no. 7, pp. 964–968, 2003.

[287] N. E. Okada H, Strutz F, Danoff TM, Kalluri R, "Possible mechanisms of renal fibrosis," *Contrib Nephrol*, vol. 118, pp. 147–154, 1996.

[288] P. J. Sime, Z. Xing, F. L. Graham, K. G. Csaky, and J. Gauldie, "Adenovector-mediated gene transfer of active transforming growth factor- β1 induces prolonged severe fibrosis in rat lung," *J. Clin. Invest.*, vol. 100, no. 4, pp. 768–776, 1997.

[289] K. Omori *et al.*, "Inhibition of Plasminogen Activator Inhibitor-1 Attenuates

Transforming Growth Factor-β-Dependent Epithelial Mesenchymal Transition and Differentiation of Fibroblasts to Myofibroblasts.," *PLoS One*, vol. 11, no. 2, p. e0148969, 2016.

[290] T. Brabletz, "To differentiate or not — routes towards metastasis," *Nat. Rev. Cancer*, vol. 12, no. 6, pp. 425–436, 2012.

[291] M. Zeisberg and E. G. Neilson, "Biomarkers for epithelial-mesenchymal transitions," *Journal of Clinical Investigation*, vol. 119, no. 6. pp. 1429–1437, 2009.

[292] I. J. Fidler and G. Poste, "The 'seed and soil' hypothesis revisited," *The Lancet Oncology*, vol. 9, no. 8. p. 808, 2008.

[293] K. R. Mansouri A, Spurr N, Goodfellow PN, "Molecular cloning and chromosomal localization of the gene encoding the human cell adhesion molecule ovomorulin," *Differentation*, vol. 38, pp. 67–71, 1998.

[294] M. J. Humphries and P. Newham, "The structure of cell-adhesion molecules.," *Trends Cell Biol.*, vol. 8, no. February, pp. 78–83, 1998.

[295] A. S. Yap, C. M. Niessen, and B. M. Gumbiner, "The juxtamembrane region of the cadherin cytoplasmic tail supports lateral clustering, adhesive strengthening, and interaction with p120(ctn)," *J. Cell Biol.*, vol. 141, no. 3, pp. 779–789, 1998.

[296] M. Takeichi, "Morphogenetic roles of classic cadherins," *Curr. Opin. Cell Biol.*, vol. 7, no. 5, pp. 619–627, 1995.

[297] W. J. Gallin, B. C. Sorkin, G. M. Edelman, and B. A. Cunningham, "Sequence analysis of a cDNA clone encoding the liver cell adhesion molecule, L-CAM," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 84, no. 9, pp. 2808–2812, 1987.

[298] A. I. M. Barth, I. S. Näthke, and W. James Nelson, "Cadherins, catenins and APC protein: Interplay between cytoskeletal complexes and signaling pathways," *Current Opinion in Cell Biology*, vol. 9, no. 5. pp. 683–690, 1997.

[299] H. Peinado, D. Olmeda, and A. Cano, "Snail, ZEB and bHLH factors in tumour progression: An alliance against the epithelial phenotype?," *Nature Reviews Cancer*, vol. 7, no. 6. pp. 415–428, 2007.

[300] H. Peinado, F. Portillo, and A. Cano, "Transcriptional regulation of cadherins during development and carcinogenesis," *International Journal of Developmental Biology*, vol. 48, no. 5–6. pp. 365–375, 2004.

[301] G. Berx, K. F. Becker, H. Höfler, and F. Van Roy, "Mutations of the human E-cadherin (CDH1) gene," *Human Mutation*, vol. 12, no. 4. pp. 226–237, 1998.

[302] J. A. Efstathiou *et al.*, "Mutated epithelial cadherin is associated with increased tumorigenicity and loss of adhesion and of responsiveness to the motogenic trefoil factor 2 in colon carcinoma cells.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 5, pp. 2316–21, 1999.

[303] V. Bolos, H. Peinado, M. A. Perez-Moreno, M. F. Fraga, M. Esteller, and A. Cano, "The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47

repressors," *J. Cell Sci.*, vol. 129, no. 6, pp. 1283–1283, 2016.

[304] C. Castro Alves *et al.*, "Slug is overexpressed in gastric carcinomas and may act synergistically with SIPI and Snail in the down-regulation of E-cadherin," *J. Pathol.*, vol. 211, no. 5, pp. 507–515, 2007.

[305] P. Coopman and A. Djiane, "Adherens Junction and E-Cadherin complex regulation by epithelial polarity," *Cellular and Molecular Life Sciences*, vol. 73, no. 18. pp. 3535–3553, 2016.

[306] L. Li, S. A. L. Bennett, and L. Wang, "Role of E-cadherin and other cell adhesion molecules in survival and differentiation of human pluripotent stem cells.," *Cell Adh. Migr.*, vol. 6, no. 1, pp. 59–70, 2012.

[307] C. Gamallo *et al.*, "Correlation of E-cadherin expression with differentiation grade and histological type in breast carcinoma.," *Am. J. Pathol.*, vol. 142, no. 4, pp. 987–93, 1993.

[308] T. T. Onder, P. B. Gupta, S. A. Mani, J. Yang, E. S. Lander, and R. A. Weinberg, "Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways," *Cancer Res.*, vol. 68, no. 10, pp. 3645–3654, 2008.

[309] S.-L. Luo, Y.-G. Xie, Z. Li, J.-H. Ma, and X. Xu, "E-cadherin expression and prognosis of oral cancer: a meta-analysis.," *Tumour Biol.*, vol. 35, no. 6, pp. 5533–7, 2014.

[310] B. M. Gumbiner, "Regulation of cadherin-mediated adhesion in morphogenesis,"

Nature Reviews Molecular Cell Biology, vol. 6, no. 8. pp. 622–634, 2005.

[311] M. Takeichi, "The cadherins: cell-cell adhesion molecules controlling animal morphogenesis.," *Development*, vol. 102, no. 4, pp. 639–655, 1988.

[312] G. M. Edelman, W. J. Gallin, A. Delouvée, B. A. Cunningham, and J. P. Thiery, "Early epochal maps of two different cell adhesion molecules.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 80, no. 14, pp. 4384–8, 1983.

[313] M. Takeichi, S. Nakagawa, S. Aono, T. Usui, and T. Uemura, "Patterning of cell assemblies regulated by adhesion receptors of the cadherin superfamily.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 355, no. 1399, pp. 885–890, 2000.

[314] H. Oda, S. Tsukita, and M. Takeichi, "Dynamic behavior of the cadherin-based cell-cell adhesion system during Drosophila gastrulation," *Dev. Biol.*, vol. 203, no. 2, pp. 435–450, 1998.

[315] A. B. Reynolds, "p120-catenin: past and present," *Biochim. Biophys. Acta*, vol. 1773, pp. 2–7, 2007.

[316] M. T. Nieman, R. S. Prudoff, K. R. Johnson, and M. J. Wheelock, "N-cadherin promotes motility in human breast cancer cells regardless of their E-cadherin expression.," *J. Cell Biol.*, vol. 147, no. 3, pp. 631–644, 1999.

[317] M. Maeda *et al.*, "Expression of inappropriate cadherins by epithelial tumor cells promotes endocytosis and degradation of E-cadherin via competition for p120ctn," *Oncogene*, vol. 25, no. 33, pp. 4595–4604, 2006.

[318] P. Doherty, G. Williams, and E. J. Williams, "CAMs and axonal growth: A critical evaluation of the role of calcium and the MAPK cascade," *Molecular and Cellular Neurosciences*, vol. 16, no. 4. pp. 283–295, 2000.

[319] S. D. Skaper, S. E. Moore, and F. S. Walsh, "Cell signalling cascades regulating neuronal growth-promoting and inhibitory cues," *Progress in Neurobiology*, vol. 65, no. 6. pp. 593–608, 2001.

[320] C. S. Theisen, J. K. Wahl, K. R. Johnson, and M. J. Wheelock, "NHERF links the N-cadherin/catenin complex to the platelet-derived growth factor receptor to modulate the actin cytoskeleton and regulate cell motility.," *Mol. Biol. Cell*, vol. 18, no. 4, pp. 1220–32, 2007.

[321] M. G. Lampugnani *et al.*, "VE-Cadherin Regulates Endothelial Actin Activating Rac and Increasing Membrane Association of Tiam," *Mol. Biol. Cell*, vol. 13, no. April, pp. 1175–1189, 2002.

[322] P. Kouklis, M. Konstantoulaki, and A. B. Malik, "VE-cadherin-induced Cdc42 signaling regulates formation of membrane protrusions in endothelial cells," *J. Biol. Chem.*, vol. 278, no. 18, pp. 16230–16236, 2003.

[323] L. Levy and C. S. Hill, "Alterations in components of the TGF-β superfamily signaling pathways in human cancer," *Cytokine and Growth Factor Reviews*, vol. 17, no. 1–2. pp. 41–58, 2006.

[324] J. P. Thiery, H. Acloque, R. Y. J. Huang, and M. A. Nieto, "Epithelial-Mesenchymal Transitions in Development and Disease," *Cell*, vol. 139, no. 5. pp.

871–890, 2009.

[325] T. Brabletz, “EMT and MET in Metastasis: Where Are the Cancer Stem Cells?,” *Cancer Cell*, vol. 22, no. 6. pp. 699–701, 2012.

[326] F. Liu, C. Pouponnot, and J. Massagué, “Dual role of the Smad4/DPC4 tumor suppressor in TGFb-inducible transcriptional complexes,” *Genes Dev*., vol. 11, pp. 3157–3167, 1997.

[327] M. Deckers *et al*., “The tumor suppressor Smad4 is required for transforming growth factor β-induced epithelial to mesenchymal transition and bone metastasis of breast cancer cells,” *Cancer Res*., vol. 66, no. 4, pp. 2202–2209, 2006.

[328] R. Derynck and Y. E. Zhang, “Smad-dependent and Smad-independent pathways in TGF-?? family signalling,” *Nature*, vol. 425, no. 6958. pp. 577–584, 2003.

[329] K. Eppert *et al*., “MADR2 maps to 18q21 and encodes a TGFβ-regulated MAD-related protein that is functionally mutated in colorectal carcinoma,” *Cell*, vol. 86, no. 4, pp. 543–552, 1996.

[330] Y. Zhang, X. H. Feng, R. Y. Wu, and R. Derynck, “Receptor-associated Mad homologues synergize as effectors of the TGF-?? response,” *Nature*, vol. 383, no. 6596, pp. 168–172, 1996.

[331] R. Derynck, R. J. Akhurst, and a Balmain, “TGF-beta signaling in tumor suppression and cancer progression.,” *Nat. Genet*., vol. 29, no. 2, pp. 117–29, 2001.

[332] T. Vincent *et al*., “A SNAIL1-SMAD3/4 transcriptional repressor complex promotes TGF-beta mediated epithelial-mesenchymal transition.,” *Nat. Cell Biol*., vol. 11, no. 8, pp. 943–50, 2009.

[333] E. Sánchez-Tilló *et al*., “EMT-activating transcription factors in cancer: Beyond EMT and tumor invasiveness,” *Cellular and Molecular Life Sciences*, vol. 69, no. 20. pp. 3429–3456, 2012.

[334] S. a Hahn *et al*., “DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1.,” *Science (80-.)*., vol. 271, no. 5247, pp. 350–353, 1996.

[335] M. Schutte *et al*., “DPC4 gene in various tumor types,” *Cancer Res*., vol. 56, no. 11, pp. 2527–2530, 1996.

[336] X. Xu *et al*., “Haploid loss of the tumor suppressor Smad4/Dpc4 initiates gastric polyposis and cancer in mice.,” *Oncogene*, vol. 19, no. 15, pp. 1868–74, 2000.

[337] W. Wick *et al*., “Evidence for a novel tumor suppressor gene on chromosome 15 associated with progression to a metastatic stage in breast cancer,” *Oncogene*, vol. 12, no. 5, pp. 973–978, 1996.

[338] K. S. Ravichandran, “Signaling via Shc family adapter proteins,” *Oncogene*, vol. 20, no. 44 REV. ISS. 5. pp. 6322–6330, 2001.

[339] J. Zavadil *et al*., “Genetic programs of epithelial cell plasticity directed by transforming growth factor-beta.,” *Proc. Natl. Acad. Sci. U. S. A*., vol. 98, no. 12, pp. 6686–91, 2001.

[340] K. M. Mulder and S. L. Morris, “Activation of p21ras by transforming growth factor ?? in epithelial cells,” *J. Biol. Chem.*, vol. 267, no. 8, pp. 5029–5031, 1992.

[341] I. Mucsi, K. L. Skorecki, and H. J. Goldberg, “Extracellular signal-regulated kinase and the small GTP-binding protein, Rac, contribute to the effects of transforming growth factor-beta1 on gene expression.,” *J. Biol. Chem.*, vol. 271, no. 28, pp. 16567–72, 1996.

[342] M. T. Hartsough and K. M. Mulder, “Transforming growth factor ?? activation of p44mapk in proliferating cultures of epithelial cells,” *J. Biol. Chem.*, vol. 270, no. 13, pp. 7117–7124, 1995.

[343] R. S. Frey and K. M. Mulder, “TGFbeta regulation of mitogen-activated protein kinases in human breast cancer cells.,” *Cancer Lett.*, vol. 117, no. 1, pp. 41–50, 1997.

[344] J. J. Northey *et al.*, “Signaling through ShcA is required for transforming growth factor beta- and Neu/ErbB-2-induced breast cancer cell motility and invasion.,” *Mol. Cell. Biol.*, vol. 28, no. 10, pp. 3162–76, 2008.

[345] a V Bakin, a K. Tomlinson, N. a Bhowmick, H. L. Moses, and C. L. Arteaga, “Phosphatidylinositol 3-kinase function is required for transforming growth factor beta-mediated epithelial to mesenchymal transition and cell migration.,” *J. Biol. Chem.*, vol. 275, no. 47, pp. 36803–10, 2000.

[346] I. Shin, a V Bakin, U. Rodeck, a Brunet, and C. L. Arteaga, “Transforming growth factor beta enhances epithelial cell survival via Akt-dependent regulation

of FKHRL1.," *Mol. Biol. Cell*, vol. 12, no. 11, pp. 3328–39, 2001.

[347] F. Viñals and J. Pouysségur, "Transforming growth factor beta1 (TGF-beta1) promotes endothelial cell survival during in vitro angiogenesis via an autocrine mechanism implicating TGF-alpha signaling.," *Molecular and cellular biology*, vol. 21, no. 21. pp. 7218–30, 2001.

[348] M. C. Wilkes *et al.*, "Transforming growth factor-β activation of phosphatidylinositol 3-kinase is independent of Smad2 and Smad3 and regulates fibroblast responses via p21-activated kinase-2," *Cancer Res.*, vol. 65, no. 22, pp. 10431–10440, 2005.

[349] Y. Y. Jae, I. Shin, and C. L. Arteaga, "Type I transforming growth factor β receptor binds to and activates phosphatidylinositol 3-kinase," *J. Biol. Chem.*, vol. 280, no. 11, pp. 10870–10876, 2005.

[350] S. Lamouille and R. Derynck, "Cell size and invasion in TGF-beta-induced epithelial to mesenchymal transition is regulated by activation of the mTOR pathway.," *J. Cell Biol.*, vol. 178, no. 3, pp. 437–51, 2007.

[351] H. Valderrama-Carvajal *et al.*, "Activin/TGF-β induce apoptosis through Smad-dependent expression of the lipid phosphatase SHIP," *Nat. Cell Biol.*, vol. 4, no. 12, pp. 963–969, 2002.

[352] C. R. Weston and R. J. Davis, "The JNK signal transduction pathway," *Current Opinion in Cell Biology*, vol. 19, no. 2. pp. 142–149, 2007.

[353] M. E. Engel, M. a McDonnell, B. K. Law, and H. L. Moses, “Interdependent SMAD and JNK signaling in transforming growth factor-beta-mediated transcription.,” *J. Biol. Chem.*, vol. 274, no. 52, pp. 37413–37420, 1999.

[354] B. a Hocevar, T. l Brown, and P. h Howe, “TGF-beta induces fibronectin synthesis through a c-Jun N-terminal kinase-dependent, Smad4-independent pathway,” *EMBO J.*, vol. 18, no. 5, pp. 1345–1356, 1999.

[355] et al Edlund S, Bu S, Schuster N, “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*, vol. 14, pp. 529–544, 2003.

[356] L. Yu, M. C. Hébert, and Y. E. Zhang, “TGF-β receptor-activated p38 MAP kinase mediates Smad-independent TGF-beta responses.,” *EMBO J.*, vol. 21, no. 14, pp. 3749–59, 2002.

[357] A. V Bakin, C. Rinehart, A. K. Tomlinson, and C. L. Arteaga, “p38 mitogen-activated protein kinase is required for TGFbeta-mediated fibroblastic transdifferentiation and cell migration.,” *J. Cell Sci.*, vol. 115, no. Pt 15, pp. 3193–3206, 2002.

[358] N. a Bhowmick *et al.*, “Transforming growth factor-beta1 mediates epithelial to mesenchymal transdifferentiation through a RhoA-dependent mechanism.,” *Mol. Biol. Cell*, vol. 12, no. 1, pp. 27–36, 2001.

[359] S. Edlund, M. Landström, C.-H. Heldin, and P. Aspenström, “Transforming

growth factor-beta-induced mobilization of actin cytoskeleton requires signaling by small GTPases Cdc42 and RhoA.," *Mol. Biol. Cell*, vol. 13, no. 3, pp. 902–14, 2002.

[360] N. A. Bhowmick *et al.*, "Transforming Growth Factor- 1 Mediates Epithelial to Mesenchymal Transdifferentiation through a RhoA-dependent Mechanism," *Mol. Biol. Cell*, vol. 12, no. 1, pp. 27–36, 2001.

[361] X. Shen, J. Li, P. P. Hu, D. Waddell, J. Zhang, and X. F. Wang, "The activity of guanine exchange factor NET1 is essential for transforming growth factor-beta-mediated stress fiber formation.," *J. Biol. Chem.*, vol. 276, no. 18, pp. 15362–15368, 2001.

[362] B. Calabrese, M. S. Wilson, and S. Halpain, "Development and regulation of dendritic spine synapses.," *Physiology (Bethesda).*, vol. 21, pp. 38–47, 2006.

[363] M. Barrios-Rodiles *et al.*, "High-throughput mapping of a dynamic signaling network in mammalian cells," *Science (80-.).*, vol. 307, no. 5715, pp. 1621–1625, 2005.

[364] M. C. Wilkes, S. J. Murphy, N. Garamszegi, and E. B. Leof, "Cell-type-specific activation of PAK2 by transforming growth factor beta independent of Smad2 and Smad3.," *Mol. Cell. Biol.*, vol. 23, no. 23, pp. 8878–89, 2003.

[365] H. Murakoshi, H. Wang, and R. Yasuda, "Local, persistent activation of Rho GTPases during plasticity of single dendritic spines," *Nature*, vol. 472, no. 7341, pp. 100–106, 2011.

[366] B. Ozdamar, R. Bose, M. Barrios-Rodiles, H. R. Wang, Y. Zhang, and J. L. Wrana, "Regulation of the polarity protein Par6 by TGFβ receptors controls epithelial cell plasticity," *Science (80-.)*., vol. 307, no. 5715, pp. 1603–1609, 2005.

[367] L. Gao, G. Joberty, and I. G. Macara, "Assembly of epithelial tight junctions is negatively regulated by Par6," *Curr. Biol*., vol. 12, no. 3, pp. 221–225, 2002.

[368] W. Xu, Z. Yang, and N. Lu, "A new role for the PI3K/Akt signaling pathway in the epithelial-mesenchymal transition," *Cell Adh. Migr*., vol. 9, no. 4, pp. 317–324, 2015.

[369] R. Cianfrocca, L. Rosanò, F. Spinella, V. Di Castro, P. G. Natali, and A. Bagnato, "Beta-arrestin-1 mediates the endothelin-1-induced activation of Akt and integrin-linked kinase.," *Can. J. Physiol. Pharmacol*., vol. 88, no. 8, pp. 796–801, 2010.

[370] K. K. Ganguly, S. Pal, S. Moulik, and A. Chatterjee, "Integrins and metastasis," *Cell Adhesion and Migration*, vol. 7, no. 3. pp. 251–261, 2013.

[371] G. X. Jiang, L. P. Cao, P. C. Kang, X. Y. Zhong, T. Y. Lin, and Y. F. Cui, "Interleukin-6 induces epithelial-mesenchymal transition in human intrahepatic biliary epithelial cells," *Mol. Med. Rep*., vol. 13, no. 2, pp. 1563–1569, 2016.

[372] N. J. Sullivan *et al*., "Interleukin-6 induces an epithelial-mesenchymal transition phenotype in human breast cancer cells," *Oncogene*, vol. 28, no. 33, pp. 2940–2947, 2009.

[373] S. O. Lee *et al*., "IL-6 promotes growth and epithelial-mesenchymal transition of

CD133 + cells of non-small cell lung cancer," *Oncotarget*, vol. 7, no. 6, pp. 6626–6638, 2015.

[374] L. Y. Ye *et al*., "Hypoxia-induced epithelial-to-mesenchymal transition in hepatocellular carcinoma induces an immunosuppressive tumor microenvironment to promote metastasis," *Cancer Res*., vol. 76, no. 4, pp. 818–830, 2016.

[375] L. Zhang *et al*., "Hypoxia induces epithelial-mesenchymal transition via activation of SNAI1 by hypoxia-inducible factor -1alpha in hepatocellular carcinoma," *BMC Cancer*, vol. 13, p. 108, 2013.

[376] M.-H. Yang *et al*., "Direct regulation of TWIST by HIF-1alpha promotes metastasis.," *Nat. Cell Biol*., vol. 10, no. 3, pp. 295–305, 2008.

[377] Y.-G. Jiang *et al*., "Role of Wnt/beta-catenin signaling pathway in epithelial-mesenchymal transition of human prostate cancer induced by hypoxia-inducible factor-1alpha.," *Int. J. Urol*., vol. 14, no. 11, pp. 1034–9, 2007.

[378] M. ME, "The Wnt/β-catenin Signaling Pathway in Epithelial Mesenchymal Transition," *J. Postdr. Res*., vol. 2, no. 7, 2014.

[379] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA. Cancer J. Clin*., vol. 67, no. 1, pp. 7–30, 2017.

[380] J. Ferlay *et al*., "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *Int. J. Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.

[381] A. Giangreco, K. R. Groot, and S. M. Janes, "Lung cancer and lung stem cells: Strange bedfellows?," *American Journal of Respiratory and Critical Care Medicine*, vol. 175, no. 6. pp. 547–553, 2007.

[382] Y. W. S. Cheng, Y. Gao, X. Dong, Y. Lu, Q. An, T. Tong, "Molecular and cytogenetic alterations in early stage of carcinogenesis of human lung," *Cancer Lett*, vol. 162, pp. S5–S10, 2001.

[383] R. Peto, A. D. Lopez, J. Boreham, M. Thun, and C. Heath, "Mortality from Smoking in Developed Countries 1950-2000. Indirect Estimates from National Vital Statistics," *Am. J. Epidemiol. Copyr.*, vol. 143, no. 51, pp. 658–4995, 1996.

[384] S. Sun, J. H. Schiller, and A. F. Gazdar, "Lung cancer in never smokers - A different disease," *Nature Reviews Cancer*, vol. 7, no. 10. pp. 778–790, 2007.

[385] P. Saracci, R., & Boffetta, *Interactions of tobacco smoking and other causes of lung cancer*. 1994.

[386] S. J. Hwang, L. S. C. Cheng, G. Lozano, C. I. Amos, X. Gu, and L. C. Strong, "Lung cancer risk in germline p53 mutation carriers: Association between an inherited cancer predisposition, cigarette smoking, and cancer risk," *Hum. Genet.*, vol. 113, no. 3, pp. 238–243, 2003.

[387] R. E. Sanders BM, Jay M, Draper GJ, "Non-ocular cancer in relatives of retinoblastoma patients," *Br J Cancer*, vol. 60, pp. 238–365, 1989.

[388] D. W. Bell *et al.*, "Inherited susceptibility to lung cancer may be associated with

the T790M drug resistance mutation in EGFR," *Nat. Genet.*, vol. 37, no. 12, pp. 1315–1316, 2005.

[389] J. E. Bailey-Wilson *et al.*, "A major lung cancer susceptibility locus maps to chromosome 6q23-25.," *Am. J. Hum. Genet.*, vol. 75, no. 3, pp. 460–74, 2004.

[390] C. I. Amos *et al.*, "Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1," *Nat. Genet.*, vol. 40, no. 5, pp. 616–622, 2008.

[391] D. C.-L. Lam *et al.*, "Expression of nicotinic acetylcholine receptor subunit genes in non-small-cell lung cancer reveals differences between smokers and nonsmokers.," *Cancer Res.*, vol. 67, no. 10, pp. 4638–4647, 2007.

[392] A. Spira *et al.*, "Effects of cigarette smoke on the human airway epithelial cell transcriptome.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 27, pp. 10143–8, 2004.

[393] X. Tang *et al.*, "EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients," *Cancer Res.*, vol. 65, no. 17, pp. 7568–7572, 2005.

[394] M. Guo *et al.*, "Promoter hypermethylation of resected bronchial margins: A field defect of changes?," *Clin. Cancer Res.*, vol. 10, no. 15, pp. 5131–5136, 2004.

[395] I. W. Park *et al.*, "Multiple clonal abnormalities in the bronchial epithelium of patients with lung cancer.," *J. Natl. Cancer Inst.*, vol. 91, no. 21, pp. 1863–8,

1999.

[396] M. Sato, D. S. Shames, A. F. Gazdar, and J. D. Minna, "A translational view of the molecular pathogenesis of lung cancer," *Journal of Thoracic Oncology*, vol. 2, no. 4. pp. 327–343, 2007.

[397] I. I. Wistuba, L. Mao, and A. F. Gazdar, "Smoking molecular damage in bronchial epithelium," *Oncogene*, vol. 21, no. 6, pp. 7298–7306, 2002.

[398] S. K. Jin, W. K. Jin, J. Han, M. S. Young, J. Park, and D. H. Kim, "Cohypermethylation of p16 and FHIT promoters as a prognostic factor of recurrence in surgically resected stage I non-small cell lung cancer," *Cancer Res.*, vol. 66, no. 8, pp. 4049–4054, 2006.

[399] M. V Brock *et al.*, "DNA methylation markers and early recurrence in stage I lung cancer.," *N. Engl. J. Med.*, vol. 358, no. 11, pp. 1118–28, 2008.

[400] J. P. Van Meerbeeck, D. A. Fennell, and D. K. M. De Ruysscher, "Small-cell lung cancer," *Lancet*, vol. 378, no. 9804, pp. 1741–1755, 2011.

[401] W. D. Travis, E. Brambilla, H. K. Müller-hermelink, and C. C. Harris, "World Health Organization Classification of Tumours WHO Classification. Pathology & Genetics Tumours of the Lung , Pleura , Thymus and Heart," in *IARC/Press*, 2004, pp. 9–122.

[402] W. D. Travis *et al.*, "International association for the study of lung cancer/american thoracic society/european respiratory society international

multidisciplinary classification of lung adenocarcinoma.," *J. Thorac. Oncol.*, vol. 6, no. 2, pp. 244–85, 2011.

[403] T. Xu, B., Thong, N., Tan, D., & Khoury, "Expression of thyroid transcription factor-1 in colorectal carcinoma," *Appl. Immunohistochem. Mol. Morphol.*, vol. 18, pp. 244–249, 2010.

[404] A. G. Nicholson *et al.*, "Refining the diagnosis and EGFR status of non-small cell lung carcinoma in biopsy and cytologic material, using a panel of Mucin staining, TTF-1, cytokeratin 5/6, and P63, and EGFR mutation analysis," *J. Thorac. Oncol.*, vol. 5, no. 4, pp. 436–441, 2010.

[405] P. S. Loo, S. C. Thomas, M. C. Nicolson, M. N. Fyfe, and K. M. Kerr, "Subtyping of undifferentiated non-small cell carcinomas in bronchial biopsy specimens," *J. Thorac. Oncol.*, vol. 5, no. 4, pp. 442–447, 2010.

[406] W. D. Travis, E. Brambilla, A. P. Burke, A. Marx, and A. G. Nicholson, *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart. Fourth edition.* 2015.

[407] W. H. Westra, "Early glandular neoplasia of the lung," *Respiratory Research*, vol. 1, no. 1. pp. 163–169, 2000.

[408] M. L. Sos *et al.*, "PTEN loss contributes to erlotinib resistance in EGFR-mutant lung cancer by activation of akt and EGFR," *Cancer Res.*, vol. 69, no. 8, pp. 3256–3261, 2009.

[409] C. Zhou *et al.*, "Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study.," *Lancet. Oncol.*, vol. 12, no. 8, pp. 735–42, 2011.

[410] R. Rosell *et al.*, "Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): A multicentre, open-label, randomised phase 3 trial," *Lancet Oncol.*, vol. 13, no. 3, pp. 239–246, 2012.

[411] T. S. Mok *et al.*, "Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma," *N. Engl. J. Med.*, vol. 361, no. 10, pp. 947–957, 2009.

[412] S. Kobayashi *et al.*, "*EGFR* Mutation and Resistance of Non–Small-Cell Lung Cancer to Gefitinib," *N. Engl. J. Med.*, vol. 352, no. 8, pp. 786–792, 2005.

[413] W. Pao *et al.*, "Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain," *PLoS Med.*, vol. 2, pp. 0225–0235, 2005.

[414] A. J. Gonzales *et al.*, "Antitumor activity and pharmacokinetic properties of PF-00299804, a second-generation irreversible pan-erbB receptor tyrosine kinase inhibitor," *Mol. Cancer Ther.*, vol. 7, no. 7, pp. 1880–1889, 2008.

[415] J. A. Engelman *et al.*, "PF00299804, an irreversible pan-ERBB inhibitor, is effective in lung cancer models with EGFR and ERBB2 mutations that are resistant to gefitinib," *Cancer Res.*, vol. 67, no. 24, pp. 11924–11932, 2007.

[416] S. S. Ramalingam *et al*., "Randomized phase II study of dacomitinib (PF-00299804), an irreversible pan-human epidermal growth factor receptor inhibitor, versus erlotinib in patients with advanced non-small-cell lung cancer.," *J. Clin. Oncol*., vol. 30, no. 27, pp. 3337–44, 2012.

[417] C. Brzezniak, C. A. Carter, and G. Giaccone, "Dacomitinib, a new therapy for the treatment of non-small cell lung cancer," *Expert Opin. Pharmacother*., vol. 14, no. 2, pp. 247–253, 2013.

[418] M. A. Nelson, J. Wymer, and N. Clements, "Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers," *Cancer Lett*., vol. 103, no. 1, pp. 115–121, 1996.

[419] T. Mitsudomi, J. Viallet, J. L. Mulshine, R. I. Linnoila, J. D. Minna, and A. F. Gazdar, "Mutations of ras genes distinguish a subset of non-small-cell lung cancer cell lines from small-cell lung cancer cell lines.," *Oncogene*, vol. 6, no. 8, pp. 1353–62, 1991.

[420] S. a Ahrendt *et al*., "Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung.," *Cancer*, vol. 92, no. 6, pp. 1525–1530, 2001.

[421] M. C. Garassino *et al*., "Different types of K-Ras mutations could affect drug sensitivity and tumour behaviour in non-small-cell lung cancer," *Annals of Oncology*, vol. 22, no. 1. pp. 235–237, 2011.

[422] M. Choi, H. Bien, A. Mofunanya, and S. Powers, "Challenges in Ras therapeutics

in pancreatic cancer," *Seminars in Cancer Biology*, 2017.

[423] F. McCormick, "KRAS as a therapeutic target," *Clinical Cancer Research*, vol. 21, no. 8. pp. 1797–1801, 2015.

[424] P. Keohavong, M. A. A. DeMichele, A. C. Melacrinos, R. J. Landreneau, R. J. Weyant, and J. M. Siegfried, "Detection of K-ras mutations in lung carcinomas: Relationship to prognosis," *Clin. Cancer Res.*, vol. 2, no. 2, pp. 411–418, 1996.

[425] G. Gnanasampanthan, H. Elsaleh, K. McCaul, and B. Iacopetta, "Ki-ras mutation type and the survival benefit from adjuvant chemotherapy in Dukes' C colorectal cancer," *J. Pathol.*, vol. 195, no. 5, pp. 543–548, 2001.

[426] D. A. Eberhard *et al.*, "Mutations in the epidermal growth factor receptor and in KRAS are predictive and prognostic indicators in patients with non-small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib," *J. Clin. Oncol.*, vol. 23, no. 25, pp. 5900–5909, 2005.

[427] M. Soda *et al.*, "Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer," *Nature*, vol. 448, no. 7153, pp. 561–566, 2007.

[428] J. P. Koivunen *et al.*, "EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer," *Clin. Cancer Res.*, vol. 14, no. 13, pp. 4275–4283, 2008.

[429] A. T. Shaw *et al.*, "Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK," *J. Clin. Oncol.*, vol. 27, no. 26, pp. 4247–4253, 2009.

[430] A. T. Shaw *et al*., "Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: A retrospective analysis," *Lancet Oncol*., vol. 12, no. 11, pp. 1004–1012, 2011.

[431] K. F. To *et al*., "Detection of ALK rearrangement by immunohistochemistry in lung adenocarcinoma and the identification of a novel EML4-ALK variant," *J. Thorac. Oncol*., vol. 8, no. 7, pp. 883–891, 2013.

[432] T. Sasaki, S. J. S. Rodig, L. L. R. Chirieac, and P. a Jänne, "The biology and treatment of EML4-ALK non-small cell lung cancer," *Eur. J. Cancer*, vol. 46, no. 10, pp. 1773–1780, 2010.

[433] T. Kohno *et al*., "KIF5B-RET fusions in lung adenocarcinoma," *Nat. Med*., vol. 18, no. 3, pp. 375–377, 2012.

[434] D. Matsubara and Y. Kanai, "Identification of CCDC6-RET Fusion in the Human Lung Adenocarcinome Cell Line, LC-2/ad," *J. Thorac. …*, vol. 7, no. 12, pp. 1872–1876, 2012.

[435] S. Dacic, "Molecular genetic testing for lung adenocarcinomas: A practical approach to clinically relevant mutations and translocations," *Journal of Clinical Pathology*, vol. 66, no. 10. pp. 870–874, 2013.

[436] A. Drilon *et al*., "Response to cabozantinib in patients with RET fusion-positive lung adenocarcinomas," *Cancer Discov*., vol. 3, no. 6, pp. 630–635, 2013.

[437] J. F. Gainor and A. T. Shaw, "Novel targets in non-small cell lung cancer: ROS1

and RET fusions.," *Oncologist*, vol. 18, no. 7, pp. 865–75, 2013.

[438] K. Bergethon *et al*., "ROS1 rearrangements define a unique molecular class of lung cancers," *J. Clin. Oncol*., vol. 30, no. 8, pp. 863–870, 2012.

[439] V. M. Rimkunas *et al*., "Analysis of receptor tyrosine kinase ROS1-positive tumors in non-small cell lung cancer: Identification of a FIG-ROS1 fusion," *Clin. Cancer Res*., vol. 18, no. 16, pp. 4449–4457, 2012.

[440] A. T. Shaw *et al*., "Crizotinib in ROS1-Rearranged Non-Small-Cell Lung Cancer.," *N. Engl. J. Med*., vol. 371, no. 21, pp. 1963–1971, 2014.

[441] C. Birchmeier, W. Birchmeier, E. Gherardi, and G. F. Vande Woude, "Met, metastasis, motility and more," *Nature Reviews Molecular Cell Biology*, vol. 4, no. 12. pp. 915–925, 2003.

[442] F. Cappuzzo *et al*., "Increased MET gene copy number negatively affects survival of surgically resected non-small-cell lung cancer patients.," *J. Clin. Oncol*., vol. 27, no. 10, pp. 1667–74, 2009.

[443] L. Toschi and F. Cappuzzo, "Clinical implications of MET gene copy number in lung cancer.," *Future Oncol*., vol. 6, no. 2, pp. 239–47, 2010.

[444] J. Bean *et al*., "MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib," *Proc. Natl. Acad. Sci*., vol. 104, no. 52, pp. 20932–20937, 2007.

[445] R. Katayama *et al*., "Cytotoxic activity of tivantinib (ARQ 197) is not due solely

to c-MET inhibition," *Cancer Res*., vol. 73, no. 10, pp. 3087–3096, 2013.

[446] H. Jin *et al*., "MetMAb, the one-armed 5d5 anti-c-met antibody, inhibits orthotopic pancreatic tumor growth and improves survival," *Cancer Res*., vol. 68, no. 11, pp. 4360–4368, 2008.

[447] D. R. Spigel *et al*., "Randomized phase II trial of Onartuzumab in combination with erlotinib in patients with advanced non-small-cell lung cancer," *J. Clin. Oncol*., vol. 31, no. 32, pp. 4105–4114, 2013.

[448] Y. Samuels *et al*., "High Frequency of Mutations of the PIK3CA Gene in Human Cancers," *Science (80-.)*., vol. 304, no. 5670, p. 554, 2004.

[449] J. E. Chaft *et al*., "Coexistence of PIK3CA and Other Oncogene Mutations in Lung Adenocarcinoma-Rationale for Comprehensive Mutation Profiling," *Mol. Cancer Ther*., vol. 11, no. 2, pp. 485–491, 2012.

[450] H. Lee *et al*., "A novel imidazopyridine PI3K inhibitor with anticancer activity in non-small cell lung cancer cells," *Oncol. Rep*., vol. 30, no. 2, pp. 863–869, 2013.

[451] R. B. Campbell, F. Liu, and A. H. Ross, "Allosteric Activation of PTEN Phosphatase by Phosphatidylinositol 4,5-Bisphosphate," *J. Biol. Chem*., vol. 278, no. 36, pp. 33617–33620, 2003.

[452] C. J. Marsit *et al*., "PTEN expression in non-small-cell lung cancer: Evaluating its relation to tumor characteristics, allelic loss, and epigenetic alteration," *Hum. Pathol*., vol. 36, no. 7, pp. 768–776, 2005.

[453] H. Takeda *et al.*, "Vandetanib is effective in EGFR-mutant lung cancer cells with PTEN deficiency," *Exp. Cell Res.*, vol. 319, no. 4, pp. 417–423, 2013.

[454] S. Dearden, J. Stevens, Y. L. Wu, and D. Blowers, "Mutation incidence and coincidence in non small-cell lung cancer: Meta-analyses by ethnicity and histology (mutMap)," *Ann. Oncol.*, vol. 24, no. 9, pp. 2371–2376, 2013.

[455] M. Katoh and H. Nakagama, "FGF Receptors: Cancer Biology and Therapeutics," *Med. Res. Rev.*, vol. 34, no. 2, pp. 280–300, 2014.

[456] A. Dutt *et al.*, "Inhibitor-sensitive fgfr1 amplification in human non-small cell lung cancer," *PLoS One*, vol. 6, no. 6, 2011.

[457] T. N. Tran *et al.*, "Fibroblast growth factor receptor 1 (FGFR1) copy number is an independent prognostic factor in non-small cell lung cancer.," *Lung Cancer*, vol. 81, no. 3, pp. 462–7, 2013.

[458] R. S. Heist *et al.*, "FGFR1 amplification in squamous cell carcinoma of the lung.," *J. Thorac. Oncol.*, vol. 7, no. 12, pp. 1775–80, 2012.

[459] M. Ren *et al.*, "Novel FGFR inhibitor ponatinib suppresses the growth of non-small cell lung cancer cells overexpressing FGFR1," *Oncol. Rep.*, vol. 29, no. 6, pp. 2181–2190, 2013.

[460] P. R. Gavine *et al.*, "AZD4547: An orally bioavailable, potent, and selective inhibitor of the fibroblast growth factor receptor tyrosine kinase family," *Cancer Res.*, vol. 72, no. 8, pp. 2045–2056, 2012.

[461] M. Hidalgo, “Pancreatic cancer,” *N Engl J Med*, vol. 362, no. 17, pp. 1605–1617, 2010.

[462] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2016,” *CA Cancer J Clin*, vol. 66, no. 1, pp. 7–30, 2016.

[463] S. Yachida and C. a Iacobuzio-Donahue, “The pathology and genetics of metastatic pancreatic cancer.,” *Arch. Pathol. Lab. Med.*, vol. 133, no. 3, pp. 413–22, 2009.

[464] et al Sohn TA, Yeo CJ, Cameron JL, “Resected adenocarcinoma of the pancreas — 616 patients: results, outcomes, and prognostic indicators,” *J Gastrointest Surg*, vol. 4, pp. 567–79, 2000.

[465] J. M. Winter *et al.*, “Survival after resection of pancreatic adenocarcinoma: Results from a single institution over three decades,” *Ann. Surg. Oncol.*, vol. 19, no. 1, pp. 169–175, 2012.

[466] C. R. Ferrone *et al.*, “Pancreatic adenocarcinoma: The actual 5-year survivors,” *J. Gastrointest. Surg.*, vol. 12, no. 4, pp. 701–706, 2008.

[467] T. Schnelldorfer *et al.*, “Long-term survival after pancreatoduodenectomy for pancreatic adenocarcinoma is cure possible?,” *Ann. Surg.*, vol. 247, no. 3, pp. 456–462, 2008.

[468] J. Helm *et al.*, “Histologic characteristics enhance predictive value of American joint committee on cancer staging in resectable pancreas cancer,” *Cancer*, vol. 115,

no. 18, pp. 4080–4089, 2009.

[469] J. A. Brosnan and C. A. Iacobuzio-Donahue, "A new branch on the tree: Next-generation sequencing in the study of cancer evolution," *Seminars in Cell and Developmental Biology*, vol. 23, no. 2. pp. 237–242, 2012.

[470] S. Yachida *et al*., "Distant metastasis occurs late during the genetic evolution of pancreatic cancer," *Nature*, vol. 467, no. 7319, pp. 1114–1117, 2010.

[471] A. Makohon-Moore and C. A. Iacobuzio-Donahue, "Pancreatic cancer biology and genetics from an evolutionary perspective," *Nature Reviews Cancer*, vol. 16, no. 9. pp. 553–565, 2016.

[472] P. M. Altrock, L. L. Liu, and F. Michor, "The mathematics of cancer: Integrating quantitative models," *Nature Reviews Cancer*, vol. 15, no. 12. pp. 730–745, 2015.

[473] I. Bozic *et al*., "Accumulation of driver and passenger mutations during tumor progression," *Proc. Natl. Acad. Sci*., vol. 107, no. 43, pp. 18545–18550, 2010.

[474] W. M. Klein, R. H. Hruban, A. J. P. Klein-Szanto, and R. E. Wilentz, "Direct correlation between proliferative activity and dysplasia in pancreatic intraepithelial neoplasia (panIN): Additional evidence for a recently proposed model of progression," *Mod. Pathol*., vol. 15, no. 4, pp. 441–447, 2002.

[475] G. M. Petersen *et al*., "A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33," *Nat. Genet*., vol. 42, no. 3, pp. 224–228, 2010.

[476] C. Tomasetti and B. Vogelstein, “Variation in cancer risk among tissues can be explained by the number of stem cell divisions,” *Science (80-.)*., vol. 347, no. 6217, pp. 78–81, 2015.

[477] S. Wu, S. Powers, W. Zhu, and Y. A. Hannun, “Substantial contribution of extrinsic risk factors to cancer development,” *Nature*, vol. 529, no. 7584, pp. 43–47, 2016.

[478] J. W. Shay, W. E. Wright, and H. Werbin, “Defining the molecular mechanisms of human cell immortalization.,” *Biochim. Biophys. Acta*, vol. 1072, no. 1, pp. 1–7, 1991.

[479] M. Kanda *et al*., “Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia,” *Gastroenterology*, vol. 142, no. 4, 2012.

[480] G. M. Petersen, “Familial pancreatic cancer,” *Seminars in Oncology*, vol. 43, no. 5. pp. 548–553, 2016.

[481] B. M. Wolpin *et al*., “Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer,” *Nat. Genet*., vol. 46, no. 9, pp. 994–1000, 2014.

[482] E. J. Childs *et al*., “Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer,” *Nat. Genet*., vol. 47, no. 8, pp. 911–916, 2015.

[483] N. A. Willis *et al*., “BRCA1 controls homologous recombination at Tus/Ter-stalled

mammalian replication forks," *Nature*, vol. 510, no. 7506, pp. 556–559, 2014.

[484] K. W. Kinzler and B. Vogelstein, "Gatekeepers and caretakers," *Nature*, vol. 386, no. 6627, pp. 761–763, 1997.

[485] F. U. Weiss, "Pancreatic cancer risk in hereditary pancreatitis," *Frontiers in Physiology*, vol. 5 FEB. 2014.

[486] W. Kimura, "How Many Millimeters do Atypical Epithelia of the Pancreas Spread Intraductally before Beginning to Infiltrate?," *Hepatogastroenterology*., vol. 50, no. 54, pp. 2218–2224, 2003.

[487] R. H. Hruban, M. Goggins, J. Parsons, and S. E. Kern, "Progression model for pancreatic cancer.," *Clin. Cancer Res*., vol. 6, no. 8, pp. 2969–72, 2000.

[488] R. E. Wilentz *et al*., "Inactivation of the p16 (INK4A) tumor-suppressor gene in pancreatic duct lesions: Loss of intranuclear expression," *Cancer Res*., vol. 58, no. 20, pp. 4740–4744, 1998.

[489] J. A. DiGiuseppe *et al*., "Overexpression of p53 protein in adenocarcinoma of the pancreas," *Am. J. Clin. Pathol*., vol. 101, no. 6, pp. 684–688, 1994.

[490] N. Waddell *et al*., "Whole genomes redefine the mutational landscape of pancreatic cancer," *Nature*, vol. 518, no. 7540, pp. 495–501, 2015.

[491] G. O. Ceyhan and H. Friess, "Pancreatic disease in 2014 : L Pancreatic fibrosis and standard diagnostics," *Nature Reviews Gastroenterology and Hepatology*, vol. 12, no. 1. pp. 38–40, 2015.

[492] G. C. Gurtner, S. Werner, Y. Barrandon, and M. T. Longaker, “Wound repair and regeneration,” *Nature*, vol. 453, no. 7193. pp. 314–321, 2008.

[493] H. F. Dvorak, “Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing.,” *N. Engl. J. Med.*, vol. 315, no. 26, pp. 1650–9, 1986.

[494] A. Masamune and T. Shimosegawa, “Pancreatic stellate cells: A dynamic player of the intercellular communication in pancreatic cancer,” *Clinics and Research in Hepatology and Gastroenterology*, vol. 39. pp. S98–S103, 2015.

[495] J. Taeger *et al.*, “Targeting FGFR/PDGFR/VEGFR impairs tumor growth, angiogenesis, and metastasis by effects on tumor cells, endothelial cells, and pericytes in pancreatic cancer.,” *Mol. Cancer Ther.*, vol. 10, no. 11, pp. 2157–67, 2011.

[496] M. H. Sherman *et al.*, “Vitamin D receptor-mediated stromal reprogramming suppresses pancreatitis and enhances pancreatic cancer therapy,” *Cell*, vol. 159, no. 1, pp. 80–93, 2014.

[497] K. P. Olive *et al.*, “Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer.,” *Science*, vol. 324, no. 5933, pp. 1457–61, 2009.

[498] P. P. Provenzano, C. Cuevas, A. E. Chang, V. K. Goel, D. D. Von Hoff, and S. R. Hingorani, “Enzymatic Targeting of the Stroma Ablates Physical Barriers to Treatment of Pancreatic Ductal Adenocarcinoma,” *Cancer Cell*, vol. 21, no. 3, pp.

418–429, 2012.

[499] M. A. Jacobetz *et al*., "Hyaluronan impairs vascular function and drug delivery in a mouse model of pancreatic cancer," *Gut*, vol. 62, no. 1, pp. 112–120, 2013.

[500] A. D. Rhim *et al*., "Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma," *Cancer Cell*, vol. 25, no. 6, pp. 735–747, 2014.

[501] B. C. Özdemir *et al*., "Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival," *Cancer Cell*, vol. 25, no. 6, pp. 719–734, 2014.

[502] I. M. Stromnes, K. E. DelGiorno, P. D. Greenberg, and S. R. Hingorani, "Stromal re-engineering to treat pancreas cancer," *Carcinogenesis*, vol. 35, no. 7, 2014.

[503] A. V. Suarez and N. D. Tsutsui, "The evolutionary consequences of biological invasions," *Molecular Ecology*, vol. 17, no. 1. pp. 351–360, 2008.

[504] C. E. Clark, S. R. Hingorani, R. Mick, C. Combs, D. A. Tuveson, and R. H. Vonderheide, "Dynamics of the immune reaction to pancreatic cancer from inception to invasion," *Cancer Res*., vol. 67, no. 19, pp. 9518–9527, 2007.

[505] R. H. Vonderheide and L. J. Bayne, "Inflammatory networks and immune surveillance of pancreatic carcinoma," *Current Opinion in Immunology*, vol. 25, no. 2. pp. 200–205, 2013.

[506] C. Feig *et al*., "Targeting CXCL12 from FAP-expressing carcinoma- associated fibroblasts synergizes with anti – PD-L1 immunotherapy in pancreatic cancer,"

Proc Natl Acad Sci U S A, vol. 110, no. 50, pp. 20212–20217, 2013.

[507] G. L. Beatty *et al*., "CD40 agonists alter tumor stroma and show efficacy against pancreatic carcinoma in mice and humans," *Science (80-.)*., vol. 331, no. 6024, pp. 1612–1616, 2011.

[508] A. D. Rhim *et al*., "EMT and dissemination precede pancreatic tumor formation," *Cell*, vol. 148, no. 1–2, pp. 349–361, 2012.

[509] I. J. Fidler, "Metastasis: Quantitative analysis of distribution and fate of tumor emboli labeled with 125l-5-lodo-2'-deoxyuridine," *J. Natl. Cancer Inst*., vol. 45, no. 4, pp. 773–782, 1970.

[510] S. Yachida *et al*., "Clinical significance of the genetic landscape of pancreatic cancer and implications for identification of potential long-term survivors," *Clin. Cancer Res*., vol. 18, no. 22, pp. 6339–6347, 2012.

[511] R. Maddipati and B. Z. Stanger, "Pancreatic cancer metastases harbor evidence of polyclonality," *Cancer Discov*., vol. 5, no. 10, pp. 1086–1097, 2015.

[512] P. J. Campbell *et al*., "The patterns and dynamics of genomic instability in metastatic pancreatic cancer," *Nature*, vol. 467, no. 7319, pp. 1109–1113, 2010.

[513] A. K. Witkiewicz *et al*., "Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets," *Nat. Commun*., vol. 6, 2015.

[514] Y. Pylayeva-Gupta, E. Grabocka, and D. Bar-Sagi, "RAS oncogenes: Weaving a tumorigenic web," *Nature Reviews Cancer*, vol. 11, no. 11. pp. 761–774, 2011.

[515] P. Lito, M. Solomon, L. S. Li, R. Hansen, and N. Rosen, "Cancer therapeutics: Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism," *Science (80-.).*, vol. 351, no. 6273, pp. 604–608, 2016.

[516] L. G. Aoude, K. A. W. Wadt, A. L. Pritchard, and N. K. Hayward, "Genetics of familial melanoma: 20 years after CDKN2A," *Pigment Cell and Melanoma Research*, vol. 28, no. 2. pp. 148–160, 2015.

[517] B. Vogelstein, D. Lane, and A. J. Levine, "Surfing the p53 network," *Nature*, vol. 408, no. 6810, pp. 307–310, 2000.

[518] K. S. Redston MS, Caldas C, Seymour AB, Hruban RH, da Costa L, Yeo CJ, "p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions," *Cancer Res.*, vol. 54, no. 11, pp. 3025–3033, 1994.

[519] U. Herbig, W. A. Jobling, B. P. C. Chen, D. J. Chen, and J. M. Sedivy, "Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21CIP1, but not p16INK4a," *Mol. Cell*, vol. 14, no. 4, pp. 501–513, 2004.

[520] N. J. Roberts *et al.*, "ATM mutations in patients with hereditary pancreatic cancer," *Cancer Discov.*, vol. 2, no. 1, pp. 41–46, 2012.

[521] S. Lamouille, E. Connolly, J. W. Smyth, R. J. Akhurst, and R. Derynck, "TGF- -induced activation of mTOR complex 2 drives epithelial-mesenchymal transition and cell invasion," *J. Cell Sci.*, vol. 125, no. 5, pp. 1259–1273, 2012.

[522] S. Jones *et al*., "Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.," *Science*, vol. 321, no. 5897, pp. 1801–6, 2008.

[523] D. Saha, P. K. Datta, and R. D. Beauchamp, "Oncogenic Ras Represses Transforming Growth Factor-β/Smad Signaling by Degrading Tumor Suppressor Smad4," *J. Biol. Chem*., vol. 276, no. 31, pp. 29531–29537, 2001.

[524] M. Kretzschmar, J. Doody, I. Timokhina, and J. Massagué, "A mechanism of repression of TGFbeta/ Smad signaling by oncogenic Ras.," *Genes Dev*., vol. 13, no. 7, pp. 804–16, 1999.

[525] M. Cordenonsi, S. Dupont, S. Maretto, A. Insinga, C. Imbriano, and S. Piccolo, "Links between tumor suppressors: p53 is required for TGF-β gene responses by cooperating with Smads," *Cell*, vol. 113, no. 3, pp. 301–314, 2003.

[526] M. Cordenonsi *et al*., "Integration of TGF-?? and Ras/MAPK signaling through p53 phosphorylation," *Science (80-.)*., vol. 315, no. 5813, pp. 840–843, 2007.

[527] M. Adorno *et al*., "A Mutant-p53/Smad Complex Opposes p63 to Empower TGFβ-Induced Metastasis," *Cell*, vol. 137, no. 1, pp. 87–98, 2009.

[528] M. Yamamoto, T. Wakatsuki, A. Hada, and A. Ryo, "Use of serial analysis of gene expression (SAGE) technology," *J. Immunol. Methods*, vol. 250, no. 1–2, pp. 45–66, 2001.

[529] E. A. Collisson *et al*., "Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy," *Nat. Med*., vol. 17, no. 4, pp. 500–503, 2011.

[530] M. F. B. Nielsen, M. B. Mortensen, and S. Detlefsen, "Key players in pancreatic cancer-stroma interaction: Cancer-associated fibroblasts, endothelial and inflammatory cells," *World Journal of Gastroenterology*, vol. 22, no. 9. pp. 2678–2700, 2016.

[531] P. Grippo, H. Munshi, and Z. Rasheed, "Pathology of pancreatic stroma in PDAC," *Pancreat. Cancer Tumor Microenviron.*, vol. 661, no. 2, pp. 1–10, 2012.

[532] R. A. Moffitt *et al.*, "Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma," *Nat. Genet.*, vol. 47, no. 10, pp. 1168–1178, 2015.

[533] P. Bailey *et al.*, "Genomic analyses identify molecular subtypes of pancreatic cancer," *Nature*, vol. 531, no. 7592, pp. 47–52, 2016.

[534] P. J. Stephens *et al.*, "Massive genomic rearrangement acquired in a single catastrophic event during cancer development," *Cell*, vol. 144, no. 1, pp. 27–40, 2011.

[535] C. A. Griffin *et al.*, "Consistent Chromosome Abnormalities in Adenocarcinoma of the Pancreas," *Cancer Res.*, vol. 55, no. 11, pp. 2394–2399, 1995.

[536] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.," *Methods*, vol. 25, no. 4, pp. 402–8, 2001.

[537] F. Sahraneshin Samani, J. K. Moore, P. Khosravani, and M. Ebrahimi, "Features

of free software packages in flow cytometry: A comparison between four non-commercial software sources," *Cytotechnology*, vol. 66, no. 4, pp. 555–559, 2014.

[538] B. Li and C. N. Dewey, "RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome," *BMC Bioinformatics*, vol. 12, 2011.

[539] C. Cheadle, M. P. Vawter, W. J. Freed, and K. G. Becker, "Analysis of microarray data using Z score transformation.," *J. Mol. Diagn.*, vol. 5, no. 2, pp. 73–81, 2003.

[540] H. Pei *et al.*, "FKBP51 Affects Cancer Cell Response to Chemotherapy by Negatively Regulating Akt," *Cancer Cell*, vol. 16, no. 3, pp. 259–266, 2009.

[541] S. Balasenthil *et al.*, "A migration signature and plasma biomarker panel for pancreatic adenocarcinoma," *Cancer Prev. Res.*, vol. 4, no. 1, pp. 137–149, 2011.

[542] T. R. Donahue *et al.*, "Integrative survival-based molecular profiling of human pancreatic cancer.," *Clin. Cancer Res.*, vol. 18, no. 5, pp. 1352–63, 2012.

[543] M. Ishikawa *et al.*, "Experimental trial for diagnosis of pancreatic ductal carcinoma based on gene expression profiles of pancreatic ductal cells," *Cancer Sci.*, vol. 96, no. 7, pp. 387–393, 2005.

[544] L. Badea, V. Herlea, S. O. Dima, T. Dumitrascu, and I. Popescu, "Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia," *Hepatogastroenterology.*, vol. 55, no. 88, pp. 2016–2027, 2008.

[545] G. Zhang *et al.*, "DPEP1 inhibits tumor cell invasiveness, enhances

chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma," *PLoS One*, vol. 7, no. 2, 2012.

[546] D. Segara *et al*., "Expression of HOXB2, a retinoic acid signaling target in pancreatic cancer and pancreatic intraepithelial neoplasia," *Clin. Cancer Res*., vol. 11, no. 9, pp. 3587–3596, 2005.

[547] C. D. Logsdon *et al*., "Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer," *Cancer Res*., vol. 63, no. 10, pp. 2649–2657, 2003.

[548] D. R. Rhodes *et al*., "Oncomine 3.0: Genes, Pathways, and Networks in a Collection of 18,000 Cancer Gene Expression Profiles," *Neoplasia*, vol. 9, no. 2, pp. 166–180, 2007.

[549] F. Varghese, A. B. Bukhari, R. Malhotra, and A. De, "IHC profiler: An open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples," *PLoS One*, vol. 9, no. 5, 2014.

[550] K. S. McCarty *et al*., "Use of a monoclonal anti-estrogen receptor antibody in the immunohistochemical evaluation of human tumors," *Cancer Res*., vol. 46, no. 8 SUPPL., 1986.

[551] TCGA, "Comprehensive genomic characterization of squamous cell lung cancers.," *Nature*, vol. 489, no. 7417, pp. 519–25, 2012.

[552] S. D. Der *et al.*, "Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients," *J. Thorac. Oncol.*, vol. 9, no. 1, pp. 59–64, 2014.

[553] M. Raponi *et al.*, "Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung," *Cancer Res.*, vol. 66, no. 15, pp. 7466–7472, 2006.

[554] J. Botling *et al.*, "Biomarker discovery in non-small cell lung cancer: Integrating gene expression profiling, meta-analysis, and tissue microarray validation," *Clin. Cancer Res.*, vol. 19, no. 1, pp. 194–204, 2013.

[555] A. H. Bild *et al.*, "Oncogenic pathway signatures in human cancers as a guide to targeted therapies," *Nature*, vol. 439, no. 7074, pp. 353–357, 2006.

[556] H. Okayama *et al.*, "Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas," *Cancer Res.*, vol. 72, no. 1, pp. 100–111, 2012.

[557] M. Yamauchi *et al.*, "Epidermal Growth Factor Receptor Tyrosine Kinase Defines Critical Prognostic Genes of Stage I Lung Adenocarcinoma," *PLoS One*, vol. 7, no. 9, 2012.

[558] S. Rousseaux *et al.*, "Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers," *Sci. Transl. Med.*, vol. 5, no. 186, 2013.

[559] Y. Xie *et al.*, "Robust gene expression signature from formalin-fixed paraffin-

embedded samples predicts prognosis of non-small-cell lung cancer patients.," *Clin. Cancer Res.*, vol. 17, no. 17, pp. 5705–14, 2011.

[560] J. Hou *et al.*, "Gene Expression-Based Classification of Non-Small Cell Lung Carcinomas and Survival Prediction," *PLoS One*, vol. 5, no. 4, p. e10312, 2010.

[561] C.-Q. Zhu *et al.*, "Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer.," *J. Clin. Oncol.*, vol. 28, no. 29, pp. 4417–4424, 2010.

[562] M. A. Nieto, "Epithelial Plasticity: A Common Theme in Embryonic and Cancer Cells," *Science (80-.).*, vol. 342, no. 6159, pp. 1234850–1234850, 2013.

[563] N. Tiwari, A. Gheldof, M. Tatari, and G. Christofori, "EMT as the ultimate survival mechanism of cancer cells," *Seminars in Cancer Biology*, vol. 22, no. 3. pp. 194–207, 2012.

[564] M. A. Eckert and J. Yang, "Targeting invadopodia to block breast cancer metastasis," *Oncotarget*, vol. 2, no. 7, pp. 562–568, 2011.

[565] M. Kawata *et al.*, "TGF-β-induced epithelial-mesenchymal transition of A549 lung adenocarcinoma cells is enhanced by pro-inflammatory cytokines derived from RAW 264.7 macrophage cells," *J. Biochem.*, vol. 151, no. 2, pp. 205–216, 2012.

[566] F. Zhao and W. T. Klimecki, "Culture conditions profoundly impact phenotype in BEAS-2B, a human pulmonary epithelial model," *J. Appl. Toxicol.*, vol. 35, no. 8, pp. 945–951, 2015.

[567] S. Dong *et al.*, "Serum starvation regulates E-cadherin upregulation via activation of c-Src in non-small-cell lung cancer A549 cells.," *Am. J. Physiol. Cell Physiol.*, vol. 307, no. 9, pp. C893-9, 2014.

[568] Y. Kondo *et al.*, "Induction of epithelial-mesenchymal transition by flagellin in cultured lung epithelial cells," *AJP Lung Cell. Mol. Physiol.*, vol. 303, no. 12, pp. L1057–L1069, 2012.

[569] M. A. Sartor *et al.*, "ConceptGen: A gene set enrichment and gene set relation mapping tool," *Bioinformatics*, vol. 26, no. 4, pp. 456–463, 2009.

[570] L. A. Miles and E. F. Plow, "Binding and activation of plasminogen on the platelet surface," *J. Biol. Chem.*, vol. 260, no. 7, pp. 4303–4311, 1985.

[571] S. A. Mani *et al.*, "The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells," *Cell*, vol. 133, no. 4, pp. 704–715, 2008.

[572] Z.-D. Lv *et al.*, "Transforming growth factor-β 1 enhances the invasiveness of breast cancer cells by inducing a Smad2-dependent epithelial-to-mesenchymal transition.," *Oncol. Rep.*, vol. 29, no. 1, pp. 219–25, 2013.

[573] M. Tojo *et al.*, "The ALK-5 inhibitor A-83-01 inhibits Smad signaling and epithelial-to-mesenchymal transition by transforming growth factor-β," *Cancer Sci.*, vol. 96, no. 11, pp. 791–800, 2005.

[574] S. Gurung, J. A. Werkmeister, and C. E. Gargett, "Inhibition of Transforming Growth Factor-β Receptor signaling promotes culture expansion of

undifferentiated human Endometrial Mesenchymal Stem/stromal Cells," *Sci. Rep.*, vol. 5, no. 1, p. 15042, 2015.

[575] G. Desnoyers, L. D. Frost, L. Courteau, M. L. Wall, and S. M. Lewis, "Decreased eIF3e Expression Can Mediate Epithelial-to-Mesenchymal Transition through Activation of the TGFβ Signaling Pathway.," *Mol. Cancer Res.*, vol. 13, no. 10, pp. 1421–30, 2015.

[576] Y. Wu *et al.*, "A83-01 inhibits TGF-β-induced upregulation of Wnt3 and epithelial to mesenchymal transition in HER2-overexpressing breast cancer cells," *Breast Cancer Res. Treat.*, vol. 163, no. 3, pp. 449–460, 2017.

[577] O. Le Gendre, A. Sookdeo, S.-A. Duliepre, M. Utter, M. Frias, and D. A. Foster, "Suppression of AKT Phosphorylation Restores Rapamycin-Based Synthetic Lethality in SMAD4-Defective Pancreatic Cancer Cells," *Mol. Cancer Res.*, vol. 11, no. 5, pp. 474–481, 2013.

[578] M. Jinnin, "Characterization of SIS3, a Novel Specific Inhibitor of Smad3, and Its Effect on Transforming Growth Factor-beta1-Induced Extracellular Matrix Expression," *Mol. Pharmacol.*, vol. 69, no. 2, pp. 597–607, 2005.

[579] C. Ramos and C. Becerril, "FGF-1 reverts epithelial-mesenchymal transition induced by TGF-β1 through MAPK/ERK kinase pathway," *Am. J. …*, vol. 229, no. 2, pp. 222–231, 2010.

[580] B. M. Slomovitz and R. L. Coleman, "The PI3K/AKT/mTOR pathway as a therapeutic target in endometrial cancer," *Clinical Cancer Research*, vol. 18, no.

21. pp. 5856–5864, 2012.

[581] C. Norrmén *et al*., "FOXC2 controls formation and maturation of lymphatic collecting vessels through cooperation with NFATc1," *J. Cell Biol*., vol. 185, no. 3, pp. 439–457, 2009.

[582] H. Fujita, M. Kang, M. Eren, L. a Gleaves, D. E. Vaughan, and T. Kume, "Foxc2 is a common mediator of insulin and transforming growth factor beta signaling to regulate plasminogen activator inhibitor type I gene expression.," *Circ. Res*., vol. 98, no. 5, pp. 626–34, 2006.

[583] N. R. Madamanchi and M. S. Runge, "Five easy pieces: The obesity paradigm," *Circulation Research*, vol. 98, no. 5, pp. 576–578, 2006.

[584] Y.-H. Yu *et al*., "MiR-520h-mediated FOXC2 regulation is critical for inhibition of lung cancer progression by resveratrol," *Oncogene*, vol. 32, no. 4, pp. 431–443, 2013.

[585] R. D. Balsara and V. A. Ploplis, "Plasminogen activator inhibitor-1: The double-edged sword in apoptosis," *Thrombosis and Haemostasis*, vol. 100, no. 6. pp. 1029–1036, 2008.

[586] I. Ota, X. Y. Li, Y. Hu, and S. J. Weiss, "Induction of a MT1-MMP and MT2-MMP-dependent basement membrane transmigration program in cancer cells by Snail1," *Proc. Natl. Acad. Sci. U. S. A*., vol. 106, no. 48, pp. 20318–20323, 2009.

[587] D. Olmeda, M. Jordá, H. Peinado, Á. Fabra, and A. Cano, "Snail silencing

effectively suppresses tumour growth and invasiveness," *Oncogene*, vol. 26, no. 13, pp. 1862–1874, 2007.

[588] S. Monea, K. Lehti, J. Keski-Oja, and P. Mignatti, "Plasmin activates pro-matrix metalloproteinase-2 with a membrane-type 1 matrix metalloproteinase-dependent mechanism," *J. Cell. Physiol.*, vol. 192, no. 2, pp. 160–170, 2002.

[589] N. Ramos-DeSimone, E. Hahn-Dantona, J. Sipley, H. Nagase, D. L. French, and J. P. Quigley, "Activation of matrix metalloproteinase-9 (MMP-9) via a converging plasmin/stromelysin-1 cascade enhances tumor cell invasion," *J. Biol. Chem.*, vol. 274, no. 19, pp. 13066–13076, 1999.

[590] M. W. Roomi, J. C. Monterrey, T. Kalinovsky, M. Rath, and A. Niedzwiecki, "Patterns of MMP-2 and MMP-9 expression in human cancer cell lines," *Oncol. Rep.*, vol. 21, no. 5, pp. 1323–1333, 2009.

[591] E. N. Baramova *et al.*, "Involvement of PA/plasmin system in the processing of pro-MMP-9 and in the second step of pro-MMP-2 activation," *FEBS Lett.*, vol. 405, no. 2, pp. 157–162, 1997.

[592] K. R. Fischer *et al.*, "Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance," *Nature*, vol. 527, no. 7579, pp. 472–476, 2015.

[593] X. Zheng *et al.*, "Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer," *Nature*, vol. 527, no. 7579, pp. 525–530, 2015.

[594] X. Liu, H. Huang, N. Remmers, and M. A. Hollingsworth, "Loss of E-cadherin and epithelial to mesenchymal transition is not required for cell motility in tissues or for metastasis," *Tissue Barriers*, vol. 2, no. 4, p. e969112, 2014.

[595] S. Lamouille, J. Xu, and R. Derynck, "Molecular mechanisms of epithelial–mesenchymal transition," *Nat. Rev. Mol. Cell Biol.*, vol. 15, no. 3, pp. 178–196, 2014.

[596] D. Medici, S. Potenta, and R. Kalluri, "Transforming growth factor-β2 promotes Snail-mediated endothelial-mesenchymal transition through convergence of Smad-dependent and Smad-independent signalling.," *Biochem. J.*, vol. 437, no. 3, pp. 515–20, 2011.

[597] N. A. Ali, M. J. McKay, and M. P. Molloy, "Proteomics of Smad4 regulated transforming growth factor-beta signalling in colon cancer cells.," *Mol. Biosyst.*, vol. 6, no. 11, pp. 2332–2338, 2010.

[598] B. A. Kennedy *et al.*, "Chip-Seq defined Genome-Wide map of TGFβ/SMAD4 targets: Implications with clinical outcome of ovarian cancer," *PLoS One*, vol. 6, no. 7, 2011.

[599] A. R. Conery, Y. Cao, E. A. Thompson, C. M. Townsend, T. C. Ko, and K. Luo, "Akt interacts directly with Smad3 to regulate the sensitivity to TGF-beta induced apoptosis.," *Nat. Cell Biol.*, vol. 6, no. 4, pp. 366–372, 2004.

[600] I. Remy, A. Montmarquette, and S. W. Michnick, "PKB/Akt modulates TGF-β signalling through a direct interaction with Smad3," *Nat. Cell Biol.*, vol. 6, no. 4,

pp. 358–365, 2004.

[601] K. Song, H. Wang, T. L. Krebs, and D. Danielpour, “Novel roles of Akt and mTOR in suppressing TGF-beta/ALK5-mediated Smad3 activation,” *EMBO J.*, vol. 25, no. 1, pp. 58–69, 2006.

[602] C. Petritsch, H. Beug, A. Balmain, and M. Oft, “TGF-beta inhibits p70 S6 kinase via protein phosphatase 2A to induce G(1) arrest,” *Genes Dev*, vol. 14, no. 24, pp. 3093–3101, 2000.

[603] S. Lamouille and R. Derynck, “Cell size and invasion in TGF-??-induced epithelial to mesenchymal transition is regulated by activation of the mTOR pathway,” *J. Cell Biol.*, vol. 178, no. 3, pp. 437–451, 2007.

[604] L. Larue and A. Bellacosa, “Epithelial–mesenchymal transition in development and cancer: role of phosphatidylinositol 3′ kinase/AKT pathways,” *Oncogene*, vol. 24, no. 50, pp. 7443–7454, 2005.

[605] Y. M. Cui *et al.*, “FOXC2 promotes colorectal cancer metastasis by directly targeting MET,” *Oncogene*, vol. 34, no. 33, pp. 4379–4390, 2015.

[606] B. G. Hollier *et al.*, “FOXC2 expression links epithelial-mesenchymal transition and stem cell properties in breast cancer,” *Cancer Res.*, vol. 73, no. 6, pp. 1981–1992, 2013.

[607] K. Shanmukhappa, G. E. Sabla, J. L. Degen, and J. A. Bezerra, “Urokinase-type plasminogen activator supports liver repair independent of its cellular receptor,”

BMC Gastroenterol., vol. 6, 2006.

[608] L. R. Lund *et al.*, "Plasminogen activation independent of uPA and tPA maintains wound healing in gene-deficient mice," *EMBO J.*, vol. 25, no. 12, pp. 2686–2697, 2006.

[609] M. Jo, R. D. Lester, V. Montel, B. Eastman, S. Takimoto, and S. L. Gonias, "Reversibility of epithelial-mesenchymal transition (EMT) induced in breast cancer cells by activation of urokinase receptor-dependent cell signaling," *J. Biol. Chem.*, vol. 284, no. 34, pp. 22825–22833, 2009.

[610] P. G. W. Gettins and K. Dolmer, "The high affinity binding site on plasminogen activator inhibitor-1 (PAI-1) for the low density lipoprotein receptorrelated protein (LRP1) Is composed of four basic residues," *J. Biol. Chem.*, vol. 291, no. 2, pp. 800–812, 2016.

[611] M. U. Romer, L. Larsen, H. Offenberg, N. Brunner, and U. A. Lademann, "Plasminogen activator inhibitor 1 protects fibrosarcoma cells from etoposide-induced apoptosis through activation of the PI3K/Akt cell survival pathway.," *Neoplasia*, vol. 10, no. 10, pp. 1083–1091, 2008.

[612] Y. P. Zhang *et al.*, "Plasminogen activator inhibitor-1 promotes the proliferation and inhibits the apoptosis of pulmonary fibroblasts by Ca2 +signaling," *Thromb. Res.*, vol. 131, no. 1, pp. 64–71, 2013.

[613] A. Cederberg, L. M. Gronning, B. Ahrén, K. Taskén, P. Carlsson, and S. Enerbäck, "FOXC2 is a winged helix gene that counteracts obesity, hypertriglyceridemia, and

diet-induced insulin resistance," *Cell*, vol. 106, no. 5, pp. 563–573, 2001.

[614] S. J. Bidarra, P. Oliveira, S. Rocha, D. P. Saraiva, C. Oliveira, and C. C. Barrias, "A 3D in vitro model to explore the inter-conversion between epithelial and mesenchymal states during EMT and its reversion," *Sci. Rep.*, vol. 6, 2016.

[615] T. Paullin *et al.*, "Spheroid growth in ovarian cancer alters transcriptome responses for stress pathways and epigenetic responses," *PLoS One*, vol. 12, no. 8, p. e0182930, 2017.

[616] S. Elschenbroich, Y. Kim, J. A. Medin, and T. Kislinger, "Isolation of cell surface proteins for mass spectrometry-based proteomics," *Expert Review of Proteomics*, vol. 7, no. 1. pp. 141–154, 2010.

[617] et al Oka T, Ishida T, Nishino T, "Immunohistochemical evidence of urokinase-type plasminogen activator in primary and metastatic tumors of pulmonary adenocarcinoma," *Cancer Res.*, vol. 51, pp. 3522–3523, 1991.

[618] F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc, "Revigo summarizes and visualizes long lists of gene ontology terms," *PLoS One*, vol. 6, no. 7, 2011.

[619] K. D. Sutherland and A. Berns, "Cell of origin of lung cancer," *Molecular Oncology*, vol. 4, no. 5. pp. 397–403, 2010.

[620] A. Feng, Z. Tu, and B. Yin, "The effect of HMGB1 on the clinicopathological and prognostic features of non-small cell lung cancer.," *Oncotarget*, vol. 7, no. 15, pp. 20507–19, 2016.

[621] I. Bolon *et al*., "Expression of c-ets-1, collagenase 1, and urokinase-type plasminogen activator genes in lung carcinomas.," *Am. J. Pathol*., vol. 147, no. 5, pp. 1298–310, 1995.

[622] M. J. Duffy, "The urokinase plasminogen activator system: role in malignancy.," *Curr. Pharm. Des*., vol. 10, no. 1, pp. 39–49, 2004.

[623] P. Thummarati *et al*., "High level of urokinase plasminogen activator contributes to cholangiocarcinoma invasion and metastasis," *World J. Gastroenterol*., vol. 18, no. 3, pp. 244–250, 2012.

[624] et al Pedersen H, Grondahl-Hansen J, Francis D, "Urokinase and plasminogen activator inhibitor type-1 in pulmonary adenocarcinoma," *Cancer Res*., vol. 54, pp. 120–123, 1994.

[625] C. E. Almasi, G. Høyer-Hansen, I. J. Christensen, and H. Pappot, "Prognostic significance of urokinase plasminogen activator receptor and its cleaved forms in blood from patients with non-small cell lung cancer.," *APMIS*, vol. 117, no. 10, pp. 755–61, 2009.

[626] B. Jacobsen and M. Ploug, "The urokinase receptor and its structural homologue C4.4A in human cancer: expression, prognosis and pharmacological inhibition.," *Curr. Med. Chem*., vol. 15, no. 25, pp. 2559–2573, 2008.

[627] C. Y. Wang *et al*., "Annexin A2 silencing induces G2 arrest of non-small cell lung cancer cells through p53-dependent and -independent mechanisms," *J. Biol. Chem*., vol. 287, no. 39, pp. 32512–32524, 2012.

[628] W. Y. Cui JW, “Expression and function of Annexin II in lung cancer tissue,” *Asian Pac J Trop Med*, vol. 6, no. 2, pp. 150–152, 2013.

[629] J. Kang, W. Kim, T. Kwon, H. Youn, J. S. Kim, and B. Youn, “Plasminogen activator inhibitor-1 enhances radioresistance and aggressiveness of non-small cell lung cancer cells.,” *Oncotarget*, 2016.

[630] I. Zemzoum *et al*., “Invasion factors uPA/PAI-1 and HER2 status provide independent and complementary information on patient outcome in node-negative breast cancer,” *J. Clin. Oncol*., vol. 21, no. 6, pp. 1022–1028, 2003.

[631] M. J. Duffy, P. M. McGowan, N. Harbeck, C. Thomssen, and M. Schmitt, “UPA and PAI-1 as biomarkers in breast cancer: Validated for clinical use in level-of-evidence-1 studies,” *Breast Cancer Research*, vol. 16, no. 4. 2014.

[632] T. S. Wang S, Jiang L, Han Y, Chew SH, Ohara Y, Akatsuka S, Weng L, Kawaguchi K, Fukui T, Sekido Y, Yokoi K, “Urokinase-type plasminogen activator receptor promotes proliferation and invasion with reduced cisplatin sensitivity in malignant mesothelioma,” *Oncotarget*, vol. 7, no. 43, pp. 69565–69578, 2016.

[633] M. Gutova *et al*., “Identification of uPAR-positive chemoresistant cells in small cell lung cancer,” *PLoS One*, vol. 2, no. 2, 2007.

[634] X. Feng, H. Liu, Z. Zhang, Y. Gu, H. Qiu, and Z. He, “Annexin A2 contributes to cisplatin resistance by activation of JNK-p53 pathway in non-small cell lung cancer cells,” *J. Exp. Clin. Cancer Res*., vol. 36, no. 1, 2017.

[635] S. Suzuki and Y. Tanigawara, “Forced expression of S100A10 reduces sensitivity to oxaliplatin in colorectal cancer cells,” *Proteome Sci.*, vol. 12, no. 1, p. 26, 2014.

[636] S. Suzuki, Y. Yamayoshi, A. Nishimuta, and Y. Tanigawara, “S100A10 protein expression is associated with oxaliplatin sensitivity in human colorectal cancer cells,” *Proteome Sci.*, vol. 9, no. 1, p. 76, 2011.

[637] D. A. Nymoen *et al.*, “Expression and clinical role of chemoresponse-associated genes in ovarian serous carcinoma,” *Gynecol. Oncol.*, vol. 139, no. 1, pp. 30–39, 2015.

[638] D. Brunen, S. M. Willems, U. Kellner, R. Midgley, I. Simon, and R. Bernards, “TGF-??: An emerging player in drug resistance,” *Cell Cycle*, vol. 12, no. 18, pp. 2960–2968, 2013.

[639] N. Oshimori, D. Oristian, and E. Fuchs, “TGF-?? Promotes Heterogeneity and Drug Resistance in Squamous Cell Carcinoma,” *Cell*, vol. 160, no. 5, pp. 963–976, 2015.

[640] K. S. Goonetilleke and A. K. Siriwardena, “Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer,” *Eur. J. Surg. Oncol.*, vol. 33, no. 3, pp. 266–270, 2007.

[641] R. L. Grossman *et al.*, “Toward a Shared Vision for Cancer Genomic Data,” *N. Engl. J. Med.*, vol. 375, no. 12, pp. 1109–1112, 2016.

[642] J. Barretina *et al.*, “The Cancer Cell Line Encyclopedia enables predictive

modelling of anticancer drug sensitivity," *Nature*, vol. 483, no. 7391, pp. 603–607, 2012.

[643] J. Budczies *et al*., "Cutoff Finder: A Comprehensive and Straightforward Web Application Enabling Rapid Biomarker Cutoff Optimization," *PLoS One*, vol. 7, no. 12, 2012.

[644] D. T. Chen *et al*., "Prognostic fifteen-gene signature for early stage pancreatic ductal adenocarcinoma," *PLoS One*, vol. 10, no. 8, 2015.

[645] J. Zhang *et al*., "International cancer genome consortium data portal-a one-stop shop for cancer genomics data," *Database*, vol. 2011, 2011.

[646] W.-Y. Huang *et al*., "MethHC: a database of DNA methylation and gene expression in human cancer.," *Nucleic Acids Res*., vol. 43, pp. 856–61, 2015.

[647] B. Jin and K. D. Robertson, "DNA methyltransferases, DNA damage repair, and cancer," *Advances in Experimental Medicine and Biology*, vol. 754. pp. 3–29, 2013.

[648] R. E. Schwarz, T. M. McCarty, E. A. Peralta, D. J. Diamond, and J. D. I. Ellenhorn, "An orthotopic in vivo model of human pancreatic cancer," *Surgery*, vol. 126, no. 3, pp. 562–567, 1999.

[649] S. Larkin and C. Aukim-Hastie, "Proteomic evaluation of cancer cells: identification of cell surface proteins.," *Methods Mol. Biol*., vol. 731, pp. 395–405, 2011.

[650] C. a Iacobuzio-Donahue *et al*., "Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays.," *Am. J. Pathol*., vol. 162, no. 4, pp. 1151–1162, 2003.

[651] R. Grützmann *et al*., "Gene Expression Profiling of Microdissected Pancreatic Ductal Carcinomas Using High-Density DNA Microarrays," *Neoplasia*, vol. 6, no. 5, pp. 611–622, 2004.

[652] B. Sitek *et al*., "Analysis of the pancreatic tumor progression by a quantitative proteomic approach and immunhistochemical validation," *J. Proteome Res*., vol. 8, no. 4, pp. 1647–1656, 2009.

[653] J. Shang *et al*., "S100A10 as a novel biomarker in colorectal cancer," *Tumor Biol*., vol. 34, no. 6, pp. 3785–3790, 2013.

[654] L. Z. Chunmei Li, Yanhong Hou, Jing Zhang, "The expressions and roles of S100A6 and S100A10 in gastric cancer," *Biomed. Res*., vol. 28, no. 5, pp. 2131–2138, 2017.

[655] T. Domoto *et al*., "Evaluation of S100A10, annexin II and B-FABP expression as markers for renal cell carcinoma," *Cancer Sci*., vol. 98, no. 1, pp. 77–82, 2007.

[656] N. A. Lokman, C. E. Pyragius, A. Ruszkiewicz, M. K. Oehler, and C. Ricciardelli, "Annexin A2 and S100A10 are independent predictors of serous ovarian cancer outcome.," *Transl. Res*., vol. 171, p. 83–95.e2, 2016.

[657] L. Miles and R. Parmer, "Plasminogen receptors: The first quarter century,"

Seminars in Thrombosis and Hemostasis, vol. 39, no. 4. pp. 329–337, 2013.

[658] M. Didiasova, L. Wujak, M. Wygrecka, and D. Zakrzewicz, "From plasminogen to plasmin: Role of plasminogen receptors in human cancer," *International Journal of Molecular Sciences*, vol. 15, no. 11. pp. 21229–21252, 2014.

[659] P. Ceruti, M. Principe, M. Capello, P. Cappello, and F. Novelli, "Three are better than one: Plasminogen receptors as cancer theranostic targets," *Experimental Hematology and Oncology*, vol. 2, no. 1. 2013.

[660] D. J. Loskutoff, "Effects of acidified fetal bovine serum on the fibrinolytic activity and growth of cells in culture," *J. Cell. Physiol.*, vol. 96, no. 3, pp. 361–369, 1978.

[661] K. Scheffzek *et al.*, "The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic ras mutants," *Science (80-.).*, vol. 277, no. 5324, pp. 333–338, 1997.

[662] L. Buø, G. I. Meling, T. S. Karlsrud, H. T. Johansen, and A. O. Aasen, "Antigen levels of urokinase plasminogen activator and its receptor at the tumor-host interface of colorectal adenocarcinomas are related to tumor aggressiveness," *Hum. Pathol.*, vol. 26, no. 10, pp. 1133–1138, 1995.

[663] J. Jankun, V. M. Maher, and J. J. McCormick, "Malignant transformation of human fibroblasts correlates with increased activity of receptor-bound plasminogen activator.," *Cancer Res.*, vol. 51, no. 4, pp. 1221–6, 1991.

[664] W. D. Madureira PA, Bharadwaj AG, Bydoun M, Garant K, O'Connell P, Lee P,

"Cell surface protease activation during RAS transformation: Critical role of the plasminogen receptor, S100A10," *Oncotarget*, vol. 7, no. 30, pp. 47720–47737, 2016.

[665] M. Ploug and V. Ellis, "Structure-function relationships in the receptor for urokinase-type plasminogen activator Comparison to other members of the Ly-6 family and snake venom α-neurotoxins," *FEBS Letters*, vol. 349, no. 2. pp. 163–168, 1994.

[666] E. F. Plow, T. Herren, a Redlitz, L. a Miles, and J. L. Hoover-Plow, "The cell biology of the plasminogen system.," *FASEB J.*, vol. 9, pp. 939–945, 1995.

[667] Q. Huai *et al.*, "Structure of human urokinase plasminogen activator in complex with its receptor," *Science (80-.).*, vol. 311, no. 5761, pp. 656–659, 2006.

[668] K. J. Dudley, K. Revill, P. Whitby, R. N. Clayton, and W. E. Farrell, "Genome-wide analysis in a murine Dnmt1 knockdown model identifies epigenetically silenced genes in primary human pituitary tumors," *Mol Cancer Res*, vol. 6, no. 10, pp. 1567–1574, 2008.

[669] J. C. Lindsey, M. E. Lusher, J. A. Anderton, R. J. Gilbertson, D. W. Ellison, and S. C. Clifford, "Epigenetic deregulation of multiple S100 gene family members by differential hypomethylation and hypermethylation events in medulloblastoma," *Br. J. Cancer*, vol. 97, no. 2, pp. 267–274, 2007.

[670] P. Rice, L. Longden, and A. Bleasby, "EMBOSS: The European Molecular Biology Open Software Suite," *Trends Genet.*, vol. 16, no. 6, pp. 276–277, 2000.

[671] D. Takai and P. a Jones, “Comprehensive analysis of CpG islands in human chromosomes 21 and 22.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 6, pp. 3740–5, 2002.

[672] X. Shan *et al.*, “MiR-590-5p inhibits growth of HepG2 cells via decrease of S100A10 expression and inhibition of the wnt pathway,” *Int. J. Mol. Sci.*, vol. 14, no. 4, pp. 8556–8569, 2013.

[673] C. R. Ferrone *et al.*, “Pancreatic ductal adenocarcinoma: long-term survival does not equal cure.,” *Surgery*, vol. 152, no. 3 Suppl 1, pp. S43-9, 2012.

[674] M. N. Katono K, Sato Y, Jiang SX, Kobayashi M, Saito K, Nagashio R, Ryuge S, Satoh Y, Saegusa M, “Clinicopathological Significance of S100A10 Expression in Lung Adenocarcinomas,” *Asian Pac J Cancer Prev.*, vol. 17, no. 1, pp. 289–294, 2016.

[675] H. Ying *et al.*, “Oncogenic kras maintains pancreatic tumors through regulation of anabolic glucose metabolism,” *Cell*, vol. 149, no. 3, pp. 656–670, 2012.

[676] and T. S. S. Akiba, R. Hatazawa, K. Ono, M. Hayama, H. Matsui, “ransforming growth factor-α stimulates prostaglandin generation through cytosolic phospholipase A2 under the control of p11 in rat gastric epithelial cells,” *Br. J. Pharmacol.*, vol. 131, no. 5, pp. 1004–1010, 2000.

[677] X. L. Huang *et al.*, “Epidermal growth factor induces p11 gene and protein expression and down-regulates calcium ionophore-induced arachidonic acid release in human epithelial cells,” *J. Biol. Chem.*, vol. 277, no. 41, pp. 38431–

38440, 2002.

[678] J. L. Warner-Schmidt *et al*., "A role for p11 in the antidepressant action of brain-derived neurotrophic factor," *Biol. Psychiatry*, vol. 68, no. 6, pp. 528–535, 2010.

[679] X. -l. Huang, "Interferon-gamma Induces p11 Gene and Protein Expression in Human Epithelial Cells through Interferon-gamma -activated Sequences in the p11 Promoter," *J. Biol. Chem*., vol. 278, no. 11, pp. 9298–9308, 2003.

[680] E. A. Peterson, M. R. Sutherland, M. E. Nesheim, and E. L. Pryzdial, "Thrombin induces endothelial cell-surface exposure of the plasminogen receptor annexin 2," *J. Cell Sci*., vol. 116, no. Pt 12, p. 2399–408., 2003.

[681] M. T. Gladwin *et al*., "Retinoic acid reduces p11 protein levels in bronchial epithelial cells by a posttranslational mechanism.," *Am. J. Physiol. Lung Cell. Mol. Physiol*., vol. 279, no. 6, pp. L1103-9, 2000.

[682] I. Kosti, N. Jain, D. Aran, A. J. Butte, and M. Sirota, "Cross-tissue Analysis of Gene and Protein Expression in Normal and Cancer Tissues," *Sci. Rep*., vol. 6, 2016.

[683] Y. Shang, X. Cai, and D. Fan, "Roles of epithelial-mesenchymal transition in cancer drug resistance.," *Curr. Cancer Drug Targets*, vol. 13, no. 9, pp. 915–29, 2013.

[684] C. Dong, S. Zhu, T. Wang, W. Yoon, and P. J. Goldschmidt-Clermont, "Upregulation of PAI-1 is mediated through TGF-β/Smad pathway in transplant

arteriopathy," *J. Hear. Lung Transplant.*, vol. 21, no. 9, pp. 999–1008, 2002.

[685] Q. Wang, Y. Wang, Y. Zhang, Y. Zhang, and W. Xiao, "The role of uPAR in epithelial-mesenchymal transition in small airway epithelium of patients with chronic obstructive pulmonary disease," *Respir. Res.*, vol. 14, no. 1, 2013.

[686] K. R. Fischer *et al.*, "Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance," *Nature*, vol. 527, no. 7579, pp. 472–476, 2015.

[687] X. Zheng *et al.*, "Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer," *Nature*, vol. 527, no. 7579, pp. 525–530, 2015.

[688] C. Shimbori *et al.*, "Fibroblast growth factor-1 attenuates TGF-β1-induced lung fibrosis," *J. Pathol.*, vol. 240, no. 2, pp. 197–210, 2016.

[689] M. A. Nieto, "Context-specific roles of EMT programmes in cancer cell dissemination," *Nature Cell Biology*, vol. 19, no. 5. pp. 416–418, 2017.

[690] N. M. Aiello, T. Brabletz, Y. Kang, M. A. Nieto, R. A. Weinberg, and B. Z. Stanger, "Upholding a role for EMT in pancreatic cancer metastasis," *Nature*, vol. 547, no. 7661, pp. E7–E8, 2017.

[691] X. Ye *et al.*, "Upholding a role for EMT in breast cancer metastasis," *Nature*, vol. 547, no. 7661, pp. E1–E6, 2017.

[692] H. Okada, T. M. Danoff, R. Kalluri, and E. G. Neilson, "Early role of Fsp1 in

epithelial-mesenchymal transformation.," *Am. J. Physiol.*, vol. 273, no. 4 Pt 2, pp. F563–F574, 1997.