Dissertation

Characterization of cancer specific mechanisms in hepatobiliary carcinoma

Submitted by

Andrija Matak, mag. educ. biol. et chem.

For the Academic Degree of

Doctor of Philosophy

(Ph.D.)

at the

Medical University of Graz

Institute of Pathology

Under the Supervision of

Prof. Dr. Kurt ZATLOUKAL

2019

Declaration

I hereby declare that this thesis is my original work and that I have fully acknowledged by name all of those individuals and organizations that have contributed to the research for this thesis. Due acknowledgment has been made in the text to all other material used. Throughout this thesis and in all related publications I followed the guidelines of "Good Scientific Practice".

Parts of this thesis are based on the following original publication that is referred to in the text wherever required:

Matak A, Lahiri P, Ford E, Pabst D, Kashofer K, Stellas D, et al. Stochastic phenotype switching leads to intratumor heterogeneity in human liver cancer. Hepatology. 2018;68(3):933-48 and publication in preparation: Matak A, Kashofer K, Zoranovic T, Pichler M, Penninger J, Zatloukal K. Cathepsin F is underexpressed in invasive HCC.

All coauthors agreed to the use of data from the respective publications in the thesis.

The PhD work presented in this thesis was supported through the academic years 2008-2013 by the PhD programs MolMed and DK-MCD (FWF W1226) of the Medical University of Graz, and the Christian Doppler Laboratory for Biospecimen Research and Biobanking Technologies. Work in Dr. Thanos laboratory was supported by grants from the Greek Secretariat for Research and Technology (COOPERATION-NoisePlus and Excellence-Stochagen), Krauss-Maffei Wegmann offsets program and GENAU-MOBILITAS program (Project 836224) from the Austrian Federal Ministry of Science and Research (GEN-AU).

The study was performed in its entirety at the Institute of Pathology, Medical University of Graz and Biomedical Research Foundation, Academy of Athens, Athens, Greece.

I have received consent for all copyrighted materials such as figures/tables reused from previously reported studies in this thesis from the respective journals.

2019

Andrija Matak

Acknowledgments

I am thankful to Professor Kurt Zatloukal for allowing me to undertake a PhD program, guidance and support throughout the course, and the privilege to attend and present in several Conferences in the field of Cancer Biology. I am also very grateful to my second supervisor, Dr. Karl Kashofer, for valuable scientific discussions, guidance and support in difficult times. I would like to thank my colleagues Drs. Vineet Mahajan, and Dr. Aniket P. Nikam for their useful scientific discussions and their friendship. I especially thank Dr. Pooja Lahiri for her help in writing my first publication. I am thankful to Dr. Ethan Ford for his valuable contribution to the project and his assistance with analysis of next-generation sequencing data.

I am especially thankful to Dr. Tamara Zoranovic for her enduring support in project concerning analysis of cathepsin protein, and to Dr. Pavle Matak for his assistance with Western-blot detection in HCC samples.

I am grateful to Biobank and their members for their support in providing patient samples and database information used in this study. I thank Dr. H.A. Chapman for providing CTSF knockout mice for controls. I am thankful to Dr. Martin Pichler for providing clinicopathological data and analysis in CTSF expression in HCC.

I acknowledge the assistance Mrs. Andrea Lackner, Mr. Stefan Sauer, Mr. Robert Reich and Mr. Heimo Muller for their assistance in processing microphotographs and retrieving data from the Austrian National Cancer Registry. I am thankful to Mrs. Daniela Pabst and other members of Dr. Zatloukal's group for their technical support. I would also like to thank members of the Laboratory of Genomic analysis, Laboratory of Immunohistochemistry, Laboratory of Molecular Diagnostics for their Institute of Pathology for their support. I thank all the research staff from the core facilities at ZMF, Medical University Graz: Molecular Biology and Imaging core facility, including Dr. Beate Rinner for her help with FACS experiments and data analysis, Msc. Jasmin Strutz, and Dr. N. Dandachi for their assistance with pyrosequencing and data analysis.

I thank Professor H. Denk, Dr. M. Asslaber, Dr. A. Aigelsreiter and Dr. J. Haybaeck, Dr. K. Lackner, Dr. G. Gorkiewicz and Dr. Alfred Beham for their help and advice on the pathological evaluations.

Also, I thank Dr. Dimitris Thanos from BRFAA and Dr. Joseph Penninger from IMBA and members of their group for a fruitful scientific collaboration, good scientific discussions, and their friendship. I would especially like to thank Drs. Apostolos Klinakis

and Dimitris Stellas for their assistance in setting up xenografting experiments and analysis of the tumors.

I especially thank Dr. Simona Vuletic for her support in motivating me to finish my project. I thank Miss Terri Gordon, writer and editor for her assistance with proofreading. I also acknowledge the support of Mag. Karin Osibow for her continual support and logistic information during my PhD and Thesis submission.

Finally, I am deeply grateful to all my family and friends for support during my studies, especially in the final steps of this PhD voyage.

Table of Contents

Abbreviations

(alphabetical order)

aCGH Array-Comparative Genomic Hybridization

AEC, 3-Amino-9-Ethylcarbazole

AFP, Alpha Fetoprotein

ANOVA, Analysis Of Variance

ANPEP, Alanyl Aminopeptidase (Membrane)

APC, Allophycocyanin

ATM, Ataxia-Telangiesctasia Mutated

C-MET, Hepatocyte Growth Factor Receptor

CA19-9, Cancer Antigen 19-9

CAT-Tg DN4, Double Negative D4 Thymocytes From N-Terminal-Deleted Stabilized B-Catenin Mutant From The Proximal Lck Promoter Transgenic Mice

CC, Cholangiocarcinoma

CD117, KIT Proto-Oncogene Receptor Tyrosine Kinase

CDH1, Cadherin 1

CDH2, Cadherin 2

CDH3, Cadherin 3

CEA, Carcinoembryonic Antigen

Cga, Chromogranin A

CGN, Cingulin

CLDN3, Claudin 3

CLDN4, Claudin 4

CNA, Copy Number Alterations

CRB3, Crumbs 3, Cell Polarity Complex Component

CRP, C-Reactive Protein

CSC, Cancer Stem Cells

CT, Computed Tomography

CTSF, Cathepsin F

DAB, 3,3'Diaminobenzidine

DAPI, 4′,6-Diamidino-2-Phenylindole

DDR1, Discoidin Domain Receptor Tyrosine Kinase 1

DEG, Differentially Expressed Genes

DMEM, Dulbecco Modified Eagles Medium

DMR, Differentially Methylated Region

DMSO, Dimethyl Sulfoxide

DNA, Deoxyribonucleic Acid

DSP, Desmoplakin

E-Cadherin, Cadherin 1

ECL, Enhanced Chemiluminescence

ECM, Extracellular Matrix

EDTA, Ethylenediaminetetraacetic Acid

EMBL, European Molecular Biology Laboratory

EMT, Epithelial-Mesenchymal Transition

Epcam, Epithelial Cell Adhesion Molecule

ERK, Extracellular-Signal-Regulated Kinase

FBS, Fetal Bovine Serum

FELASA, Federation Of European Laboratory Animal Science Associations

FFPE Formalin Fixed Paraffin Embeded

FGFR2, Fibroblast Growth Factor Receptor 2

FITC, Fluorescein Isothiocyanate

FLP, Flippase Enzyme

FNH, Focal Nodular Hyperplasia

FSC, Forward Scatter

GFP, Green Fluorescent Protein

GGT, Gamma-Glutamyl Transferase

GJB3, Gap Junction Protein Beta 3

GJB4, Gap Junction Protein Beta 4

GJB6, Gap Junction Protein Beta 6

GO, Gene Ontology

GRN, Gene Regulatory Network

GSEA, Gene Set Enrichment Analysis

HB, Hepatitis B Virus

HCA, Hepatocellular Adenoma

HCC Hepatocellular Carcinoma

HCV, Hepatitis C Virus

HDL, High-Density Lipoproteins

Hep Par 1, Hepatocyte Paraffin 1 Antibody

HEPC, Human Epithelial Prostate Cells

IGV, Integrative Genomics Viewer

IH Score, Immunohistocemistry Score

IHC, Immunohistochemistry

IL13RA2, Interleukin 13 Receptor Subunit Alpha 2

ITH, Intratumor Heterogeneity

JUP, Junction Plakoglobin

LDL, Low-Density Lipoproteins

LSD, Least Significant Difference

MCAM, Melanoma Cell Adhesion Molecule

MCF-7, Michigan Cancer Foundation-7

MET, Mesenchymal-Epithelial Transition

MHC, Major Histocompatibility Complex

MMP11, Matrix Metallopeptidase 11

MMP2, Matrix Metallopeptidase 2

MSigDB, Molecular Signature Database

MUC-1, Mucin 1 Cell Surface Associated

NASH, Nonalcoholic Fatty Liver Disease

NCAM1, Neural Cell Adhesion Molecule 1

NEAA, Non-Essential Amino Acids

NL- Non-Neoplastic Liver

NOD, Non-Obese Diabetic

NTC, Non-Template Control

OCLN, Occludin

OS, Operating System

PBS, Phosphate Buffered Saline

PDPN, Podoplanin

PI, Propidium Iodide

PROCR, Protein C Receptor

PVDF, Polyvinylidene Difluoride

PVRL4, Nectin Cell Adhesion Molecule 4

qRT-PCR, Quantitative Reverse Transcriptase Polymerase Chain Reaction

Ras, Retrovirus-Associated DNA Sequences

RIPA, Radioimmunoprecipitation

RN, Ribonucleic Acid

RNA-FISH, RNA-Fluorescence In Situ Hybridization

RNAi, RNA Interference

RT, Room Temperature

SAHA, Suberoylanilide Hydroxamic Acid

SCID, Severe Combined Immune-Deficient

SD, Standard Deviation

SDS PAGE, Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis

SDS, Sodium Dodecyl Sulfate

SERPINE1, Serpin Family E Member 1

SSC, Side Scatter

TAD, Transactivation Domain

TBST, Tris Buffered Saline+

TGFBR1, Transforming Growth Factor Beta Receptor 1

TJP3, Tight Junction Protein 3

TLN2, Valin 2

TMA, Tissue Microarrays

TSA, Trichostatin A

TWIST1, Twist Family Bhlh Transcription Factor 1

UAS, Upstream Activating Sequence

VEGFC, Vascular Endothelial Growth Factor C

VIM, Vimentin

VSV-G, G Glycoprotein Of The Vesicular Stomatitis Virus

ZEB1, Zinc Finger E-Box Binding Homeobox 1

ZEB2, Zinc Finger E-Box Binding Homeobox 2

List of Figures

List of Tables

Abstract (in German)

Intratumorale-Heterogenität wird zunehmend als Hauptfaktor für die Diagnose und die personalisierte Behandlung von Krebs erkannt. Selbst in genetisch homogenen Zellpopulationen gibt es eine erhebliche phänotypische Vielfalt von Krebszellen, einschließlich spezifischer Zellen, Welch zum Fortbestehend der Erkrankung beitragen. Die nicht-genetische Heterogenität von Zellen ist häufig mit dem Erwerb maligner Eigenschaften verbunden, wie etwa krebsstammzellartigen Zuständen, Arzneimittelresistenz und metastatischen Eigenschaften, und solche Tumorzellen können irreversibel oder reversibel zwischen verschiedenen Zuständen wechseln. In dieser Arbeit identifizierte ich den stochastischen Phänotypwechsel als einen neuen Aspekt der nichtgenetischen Dynamik, welcher zur Tumorheterogenität beim sarkomatoiden Cholangiokarzinom beiträgt. Die Analyse von Einzelzell-abgeleiteten klonalen Populationen primärer Tumorzellkulturen unter Verwendung des Abstammungsmarkers Keratin-7 identifizierte verschiedene Arten von sich selbst vermehrenden Klonen, die entweder durch stabile oder instabile Expression des Abstammungsmarkers gekennzeichnet sind. Klonale Zellen sind entweder Keratin-7-positiv oder Keratin-7-negativ (stabile Phänotypen) oder verändern ständig die Expression des Abstammungsmarkers (instabiler Phänotyp). Diese Phänotypen persistierten in nachfolgenden Einzelzellpopulationen. Eine Methylierungsanalyse des Gesamtgenoms ergab eine unterschiedliche Methylierung des Keratin-7-Promotors, welche das Expressionsniveau regulierte. Die Transkriptomsequenzierung zeigte, dass der Wechsel zwischen Keratin-7-Phänotypen mit Veränderungen der Expression mehrerer Gene zusammenhängt, von denen einige mit dem epithelialen-mesenchymalen Transition zusammenhängen. In Xenotransplantationstests zeigte der Keratin-7-negative Phänotyp ein erhöhtes tumorigenes Potential, das mit dem Keratin-7-negativen Phänotyp der Metastasen des Primärtumors dieser Zellen korreliert.

Änderungen in mehreren Signalwegen, die sich gegenseitig koordinieren und zusammenarbeiten, sind für komplexe biologische Prozesse, wie das Zellwachstum und die Bildung von Metastasen bei Krebs erforderlich. Das bekannte RAS-Onkogen allein ist kein ausreichender Treiber für die Tumorentstehung, sondern benötigt kooperierende Faktoren. Es besteht die Notwendigkeit, neue Wege und Gene zu identifizieren, die am Krebswachstum und an der Metastasierung bei durch Ras hervorgerufenen Krebserkrankungen beteiligt sind. In meiner Arbeit, dass Cathepsin F (CTSF), eine

Cystatin-ähnliche Cysteinprotease mit einer gut etablierten Rolle bei physiologischen und pathologischen Prozessen, auch ein neues Tumorsuppressorgen beim hepatozellulären Karzinom ist. Ich habe gezeigt, dass CTSF beim hepatozellulären Karzinom im Vergleich zu nicht-neoplastischen Lebergeweben signifikant herunterreguliert ist. CTSF korreliert negativ mit erhöhtem Tumorgrad und Proliferationsmarker Ki-67 und p53. Patienten mit reseziertem CTSF-negativem HCC hatten im Vergleich zu CTSF-positivem HCC ein signifikant schlechteres 10-Jahres-Gesamtüberleben. Eine verringerte CTSF-Expression in HCC ist ein potenzieller Marker für das Progression, Invasivität und das schlechte Überlebensraten beim HCC.

Zusammenfassend habe zwei relevante Merkmale des Hepatobiliären Karzinoms in zwei verschiedenen Ansätzen untersucht. Zunächst beschreibe ich die nicht-genetische Heterogenität beim sarkomatoiden Cholangiokarzinom, die das derzeitige konzeptionelle Bild der intratumoralen Heterogenität in eine andere Richtung lenkt. Zweitens habe ich einen neuen Prognosemarker für Leberzellkrebs validiert, der klinische Relevanz impliziert, indem ich Leberzellpatienten in Kategorien mit guter und schlechter Prognose einteilte. Diese Arbeit trägt wesentlich zum Verständnis wichtiger krebsspezifischer Mechanismen bei, die beim Hepatobiliärkarzinom eine Rolle spielen.

Abstract

Intratumor heterogeneity is increasingly recognized as a major factor affecting diagnosis and personalized treatment of cancer. Even in genetically homogeneous cell populations, there is a substantial amount of phenotypic diversity of cancer cells, including specific cells that sustain cancers. Non-genetic heterogeneity of cells is often associated with the acquisition of malignant properties, such as cancer stem-like states, drug resistance, and metastatic properties, and such tumor cells can transition between different states irreversibly or reversibly. In this work, I identified stochastic phenotype switching as a novel aspect of non-genetic dynamics contributing to tumor heterogeneity in sarcomatoid cholangiocarcinoma. Analysis of single cell-derived clonal populations of primary tumor cell cultures using the lineage marker keratin-7 identified different types of self-propagating clones characterized by either stable or unstable expression of the lineage marker. Clonal cells are either keratin-7 positive or keratin-7 negative (stable phenotypes), or constantly changing the expression of the lineage marker (unstable phenotype). These phenotypes persisted in subsequent single cell-derived populations. Whole-genome methylation analysis revealed differential methylation of the keratin-7 promoter regulating its expression levels. Transcriptome sequencing showed that switching between keratin-7 phenotypes was linked to changes in the expression of several genes, some of which were related to the epithelial-mesenchymal transition. In xenotransplantation assays, keratin-7 negative phenotype showed increased tumorigenic potential, which correlates with the keratin-7 negative phenotype of the metastases of the primary tumor of these cells.

Changes in multiple signaling pathways that mutually coordinate and cooperate are required for complex biological processes in cancer, like cell growth and formation of metastases. The well-known RAS oncogene alone is not a sufficient driver for tumorigenesis but needs cooperating factors. There is a need to identify novel pathways and genes that are involved in cancer growth and metastasis in Ras-driven cancers. Here I show that Cathepsin F (CTSF), a cystatin-like cysteine protease with a well-established role in physiological and pathological processes is also a novel tumor suppressor gene in hepatocellular carcinoma. I have shown that CTSF is significantly downregulated in hepatocellular carcinoma compared to non-neoplastic liver tissues. CTSF negatively correlates with increased tumor grade and proliferation marker Ki-67, and p53. Patients with resected CTSF-negative HCC had significantly poorer 10-year overall survival

compared to CTSF-positive HCC. Reduced CTSF expression in HCC is a potential marker of HCC progression, invasiveness, and poor survival.

Collectively, I studied two relevant features in hepatobiliary carcinoma using two different approaches. First, I describe non-genetic heterogeneity in sarcomatoid cholangiocarcinoma giving another twist to the current conceptual picture of intratumor heterogeneity. Second, I validated a new prognosis marker in hepatocellular cancer, which implies clinical relevance by segregating hepatocellular patients into categories of good and poor prognosis patients. This body of work significantly advances the understanding of important cancer-specific mechanisms involved in hepatobiliary carcinoma.

1 Introduction

1.1 Hepatobiliary cancers

Hepatobiliary cancers are defined as primary malignancies that develop in the liver: hepatocellular cancers (HCC) and biliary tract cancers. Among the hepatobiliary cancers, 15 % to 20 % is accounted for cholangiocarcinoma (CC) (1). Based on anatomical site, CC can be classified as an intrahepatic, perihilar, and distal subtype (2). According to the Global Cancer Statistics of 2018, the worldwide incidence of hepatobiliary malignancy is estimated to be 841,080 of new liver cancer cases, and 781,631 of liver cancer-related deaths for both sexes and all ages (3).

1.1.1 HCC

HCC is the most common form of primary liver cancer (4). HCC typically develops on a background of chronic liver disease, predominantly caused by hepatitis B virus (HBV), hepatitis C virus (HCV), alcohol abuse, or non-alcoholic steatohepatitis (NASH) (5). HCC is a heterogeneous disease both morphologically and molecularly (6). A major effort has been taken over the last 10 years to understand molecular heterogeneity between tumors in different patients through the development of mutation signatures and molecular classifications (7). The Genomic and transcriptional analysis of HCC made possible to group patients into defined molecular classes of HCC, characterized by specific constellation of gene mutations and expression signatures. Molecular classes of HCC are also different in terms of activated signaling pathways, structural DNA alterations, and clinicopathological features including survival of patients (8). To increase the survival perspectives of HCC patients, improve the diagnosis and treatment, we require discovery of new survival and prognostic markers.

1.1.2 CC

CC is a tumor entity originating from the intrahepatic and extrahepatic biliary tracts. It is considered the second most common primary liver tumor with rising global incidence and mortality rates and it accounts for 15 % to 20 % of primary hepatobiliary malignancies (1). Sarcomatoid cholangiocarcinomas (sCC) are an unusual and uncommon variant of

cholangioncarcinoma characterized by coinciding epithelial and mesenchymal differentiated cells in regionally isolated and transitional tumor compartments. It is considered a prototype for elucidating cellular, histological, and molecular intratumor heterogeneity. Presences of sarcomatoid component in tumors is generally associated with more aggressive tumor pathology and distant organ metastases leading to early patient demise after initial diagnosis (9-13). Recent findings established a divergent, monoclonal progression model of sarcomatoid carcinomas, where unstable epithelial cells undergo trans-differentiation epithelial-mesenchymal transition (EMT) process toward different degrees of mesenchymal cell differentiation (14, 15). Cholangiocarcinoma with sarcomatoid differentiation is, therefore, particularly suitable for studying stochastic genetic and epigenetic contributors involved in intratumor heterogeneity.

1.2 Intratumor heterogeneity (ITH)

Cancer is a complex disease displaying pronounced diversity between patients having the same type of cancer (inter-tumor heterogeneity) and within single cancer (intratumor heterogeneity, ITH). ITH is a phenomenon that influences many aspects of cancer research including a basic understanding of tumor biology, development of techniques used to study tumors and different strategies for the treatment of patients. ITH is a prominent feature and a poorly understood phenomenon of most cancers. It accounts for extensive tumor sub-classification, malignant progression, ambiguous prognosis, treatment failure and disease recurrence (development of therapeutic resistance) (16-18). ITH may seriously hinder realistic opportunities for managing tumors through personalized medicine approaches, which rely on reliable prognostic or predictive biomarker strategies, and comprehensive tumor sampling (19, 20). Heterogeneity within a tumor may include diversity in cell morphology and biological behavior relating to proliferation, metastatic potential, and responsiveness to target therapies. ITH has become particularly important in the context of personalized medicine where it can markedly affect malignant progression, the efficacy of diagnostic and prognostic biomarker analysis, sample size bias, and treatment failure associated with the selective response of cancer cell subpopulations (21, 22). Insights into the mutational and gene expression heterogeneity in different tumor regions are considered to provide a gateway for understanding the biological impact of genetic and phenotypic diversity within a primary tumor and its metastases (20, 23).

To appropriately develop the concept of ITH we need to consider the source of ITH: Genetic heterogeneity and non-genetic (epigenetic) heterogeneity. The main focus of this thesis is based on the general hypothesis that stochastic phenotype dynamics, as an alternative source of ITH drives non-genetic tumor heterogeneity.

1.3 Origins of intratumor heterogeneity

1.3.1 Somatic clonal evolution model

To assemble the comprehensive picture of the complexity of tumor development it is necessary to take into account several layers of heterogeneity. Only after the completion of the human genome project, by large scale sequencing of coding and non-coding aberrations, was it possible to grasp the genetic complexity underlying different types of cancer. Genetic heterogeneity is generated by genomic instability that is a pervasive property of tumors in general, and is fundamental to the process of tumor progression (24). Genetic alterations are widespread in cancers with an estimated frequency of at least 11,000 genomic alterations in the colon carcinoma clone, even though these mutations are positioned in non-coding regions (25). A model that accounts for the genetic heterogeneity and tumor development in an individual patient, was first proposed by Peter Nowell in 1976 in his landmark paper on clonal evolution: "The clonal evolution of tumor cell populations" (26) where he describes cancer development using the same principles as Darwinian evolution. Increased genomic instability generates heritable mutations or genotypes, which are subclonaly selected based on the fitness advantage under selective environment pressures (26, 27).

As a consequence of equal chance that each cell in a tumor can gain advantageous mutation, (such as for growth potential), it is conceivable that more than one clonal population can be present in clonal outgrowth and outcompeting clones with less fitness resulting in a series of clonal expansions or contractions, resulting in complex subclonal architecture inside the tumor (28). It has also been reported that minor clones can cooperate thus increasing the survival of the entire tumour, which is the main argument for looking at heterogeneous tumors as complex societies (29, 30). In addition, clonal growth can change in time and space as different resource become available in different areas inside the tumor. Certain areas may select for hypoxia resistant clones, while fast proliferating clones may be located in nutrient-dense regions of the tumor. Spatial and

temporal clonal heterogeneity can change as a consequence of therapy (31, 32). This implies that it is common that all cancers, through complex evolutionary dynamics have complex clonal architecture. Also, diversification of cancer cells such as spatial constraints and differences in nutritional resources may further cause the preservation of clonally heterogeneous tumors (33) In summary, tumors represent a complex dynamic ecological system where heterogeneity is not only a substrate for evolution, but can also promote, or even be a requirement for, continued tumor development and progression (34).

1.3.2 Cancer stem cells (CSC)

CSC model proposes irrespective of the identity of the cells, that cellular heterogeneity is formed by developmental hierarchy resulting in various cellular phenotypes.

Based on the functional criteria of CSC of asymmetric and symmetric cell division, disbalance of these differentiation programs can be seen in early stage cancers, characterized by asymmetric divisions, while late-stage cancers are comprised of CSC mostly undergoing symmetric division, resulting in a majority of undifferentiated self-renewing cells (35). As with normal stem cells, CSC can differentiate through epigenetic reprogramming into non-tumorigenic cells, resulting in a majority being of cells in the tumor as well as a minor subpopulation of highly tumorigenic cells that sustains the growth of a primary tumor and are responsible for tumor initiation and growth of metastasis (36). Today there is convincing data that supports the hypothesis that variety of cancers follow a cancer stem cell model, including breast cancers (37), colon cancers (38), liver (39), head and neck (40), prostate (41), brain cancers (42), leukemias (43, 44) and lung cancer (45).

It is logical to assume that initial mutation occurs in a normal stem cell/progenitor cell compartment as CSC simply re-uses its normal self-renewing programs for tissue regeneration (46). A cell of origin can be a source for intratumor heterogeneity (47). This is specifically exemplified in a rare liver cancer subtype classified as hepatocellular-cholangiocarcinoma, which is formed by coexisting malignant hepatocyte and cholangiocyte cell lineages, consistent with the notion that the cell of origin could have been a bipotential liver progenitor cell (48, 49) However, it is now clear that CSC can originate by oncogene transformation or by dysregulated signaling pathways in differentiated cells (50, 51).

CSC are difficult to study, especially due to questionable reliability of the CSC markers, which are important for isolation and characterisation of CSC. For instance, CSC markers may vary in cancer subtypes or between patients in the same cancer subtype (52).

Another issue with stem cell markers is the stability of CSC state. An intriguing study in human melanoma cells showed that a stem cell marker, initially thought to be expressed in slow-cycling stem-like subpopulations and marking cells that are the most competent in sustaining tumor growth, JARID1B histone demethylase, is dynamic and frequently appears in cells that are initially negative (53). Since almost all cells can acquire features of CSC, such as drug-resistance or can switch between sensitive and non-sensitive subtypes, targeting any subpopulation in tumor is irrelevant since all cells are relevant and represent CSC (54). It was experimentally proven in melanoma cell lines and lung cancer cell lines that epigenetic reprogramming is responsible for the transient and reversible acquisition of drug tolerance and has no relation to classical stem cell models with stable cellular hierarchies (55, 56).

Another confounding aspect of CSC is that many studies have proven that stemness features can be acquired by non-CSC cells (57). This bidirectional interconversion may be caused by both intrinsic and extrinsic factors such as tumor microenvironment (57, 58). A hierarchical organization of tumors may be directly influenced by the type of driver mutations and differentiated status of the targeted cells and is often supported by factors or cell entities from surrounding tumor microenvironment (52, 59).

It is important to note that differentiation hierarchy is not the only source of tumour heterogeneity as genetically uniform cells can display variability in phenotypes that are a result of the stochastic processes in gene expression and signaling pathways (60, 61).

Clonal evolution and tumorigenic stem cell differentiation can both complement or independently contribute to intratumor heterogeneity. Essentially every cell in the tumor that is capable of propagation long term and continues to accumulate mutations is likely to increase its competitive advantage compared with other self-propagating cells (28). This would make CSC represent units of selection (62, 63).

1.3.3 Influence of microenvironment

Tumor microenvironment presents an independent but also complementary source that directly or indirectly influences phenotypic variability of cancer cells by providing different micro-environments, each with its own selective pressures (64, 65). Chronic

inflammation, hypoxia, and stromal cell type composition can all influence cancer cell phenotypic heterogeneity. Chronic inflammation increases the risk of cancer occurrence, and liver and colon cancers can develop in chronic, unresolved inflammation in cirrhosis and inflammatory bowel disease, respectively (66). It is well known that hypoxia directly influences the malignant phenotype of cancer cells and hypoxic conditions are often found in the center of solid tumor and exert selective pressure under which aggressive cells may escape to the tumor edge. Hypoxic conditions also play a role in immune responses, metastasis, and therapeutic responses (67). It has been shown that both chronic inflammation and hypoxia can direct affect activity and levels of DNA and histone modifying enzymes (68, 69). Pertinent to the role of tissue microenvironment influencing aggressive tumor behaviour, extracellular matrix (ECM) structure and stiffness can directly regulate features of cell growth by mechanically influencing epigenetic modulators involved with transcriptional regulators of growth-promoting genes (70).

1.4 Intratumor heterogeneity at the epigenetic level

Epigenetics is a branch of biology that studies interaction between DNA sequence and its products. Epigenetic information is heritable. This means that epigenetic information is maintained during cell division and propagates stable gene expression states (71).
Epigenetics has typically been defined as rigid system that is storing cellular information. Recent discoveries have shown that epigenetic information can be extensively influenced by the microenvironment and is able to shape different gene expression states, especially in aging, and affects susceptibility to disease (72). DNA methylation is the most prevalent mechanism that is altered in cancers, and it represents an important contributing factor to the processes that underlie tumor heterogeneity (73). Human genome in sparingly methylated; however, specific regions of the genome, called CpG islands, are markedly methylated. These areas are associated with the start sites of actively transcribed genes (71), and approximately 60-90 % of cytosines are methylated at these sites (74). One of the most frequent mechanisms of gene inactivation, particularly tumor suppressor gene promotors, is hypermethylation of CpG islands (75). Correspondingly, hypomethylation of oncogenes has been linked with activation of an oncogenes in major cancer types, and affects chromosomal stability (76, 77). Loss of epigenetic stability is a unifying feature across divergent tumor types, and it may play an important role in tumor progression and

occurrence of tumor (78, 79). It has become clear that epigenetic and genetic defects may influence each other and cooperatively be involved in processes of tumorigenesis, resulting in the generation of specific cancer phenotypes (80). In an aberrant cancer setting, stable and reproducible epigenetic mechanisms may be compromised, enabling sufficient epigenetic plasticity that contributes to dynamic phenotype heterogeneity (78). Recent reports show that stochastic DNA methylation variability is confined to large hypomethylated chromosomal regions that involve genes implicated in tissue differentiation, epigenetic reprogramming, and cancer occurrence (81, 82). These regions of the genome contain one of the most variably methylated regions in the tumor genome and the most variably expressed genes regulating functions that are relevant to cancer cells (82). Also, variability in methylation in precancerous cells is a predictor of cancer risk, which further establishes a causal link between epigenetic alterations and cancer (83, 84). Global dynamic changes in DNA methylation are involved in stabilizing phenotype switches. These changes include either stochastic variability or stable patterns in DNA methylation, including hybrid states of increased and decreased epigenetic plasticity (78, 85).

Also, a studies in human B-cell cultures showed that hypomethylated blocks with hypervariable DNA methylation regions and associated gene expression occur early in malignant transformation (82). Moreover, using single-molecule RNA-fluorescence *in situ* hybridization (RNA-FISH) in embryonic stem cells, DNA methylation was implicated as a key factor in mediating stochastic-state switching in single cells (86). Overall, stochastic DNA methylation variation is an inherent characteristic of the phenotypic variability implicated in fuelling tumor cell heterogeneity, where cellular diversity favors the selection of the fittest cancer clones influenced by the changing tumor environment (87-89).

Finally, a fascinating review by Feinberg and colleagues based on their 10-year-old tumor progenitor gene model, summarized in detail epigenetic dynamics in cancer development by defining new functional classes of genes called: epigenetic mediators, epigenetic modifiers, and epigenetic modulators (90). The concept of tumor progenitor genes highlights the necessity of deregulated epigenome, which occurs even before mutations in the normal tissues, and alters normal differentiation pathways that, with mutations, later on, lead to tumor development (91). Mutations, aging, and environmental factors affect epigenetic modulators and activity of epigenetic modifiers, which regulate the expression of epigenetic mediators, which shape the epigenetic landscape and create a feedback loop with modulator and modifiers that all alter the 3D nuclear architecture in the nucleus,

which ultimately shifts the phenotype towards cancer stem-like states and phenotypic plasticity (90, 92).

1.5 Stochasticity in biochemical reactions: Noise

Stochastic gene expression is essential for many biological processes. Phenotype heterogeneity may also derive from the stochastic nature of biological processes ("noise") (93, 94), which are manifested as stochastic gene expression, stochastic differential methylation, variant lineage commitment, cell survival, and drug resistance (55, 85, 95, 96).

The 2000s were the beginning of an increased interest in studying and analyzing stochastic gene expression. After overcoming technical hurdles in single-cell analytics, and development of models for studying synthetic gene networks, it was possible to study variations in gene expression in a cell under constant environment (97, 98). It was also made possible to quantify noise across population (99). Embryonic and adult stem cells show a substantial degree of heterogeneity of fluctuating transcription factors regulated by epigenetic regulatory machinery (100). In one of the earliest reports on the topic of gene expression in stem cells, Efroni *et al*. published a article showing that transcriptional activity differs between undifferentiated stem-like states compared to cells that are more prone to differentiation. Further, reports followed showing that variability or stochastic expression of a panel of stemness genes, in embryonal stem cells is intrinsically linked with the pluripotent state (101). Interestingly, gene expression variability is dynamic in embryonal stem cells and is associated with the openness of the chromatin landscape (102). One of the advantages of this variability is its potential to choose from several developmental options during development and differentiation (96).

Heterogeneity in clonal cell populations can arise from noise in gene expression due to the random molecular events inherent to the process of transcription (94, 98). This contributes to the transcriptomic heterogeneity, which is likely regulated by multi-gene epigenetic regulatory networks. For example, enzymes involved in remodeling nucleosomes, such as histone H3K4 demethylases, have a crucial role in organizing and controlling gene expression variability (103). In addition, changes in active and inactive chromatin states caused by chromatin remodeling factors can lead to increased transcriptomic heterogeneity

(104). Therefore, it is important to understand how transcriptome variability caused by non-genetic heterogeneity impacts fitness of cancer cells. Majority of work has so far defined transcriptome heterogeneity as a mechanism for cells to increase their survival, particularly in developing drug resistance, metastatic recurrence, and viral latency (53, 105-107). Since transcriptome heterogeneity is clearly manifested in phenotypic heterogeneity, targeting epigenetic regulators of this heterogeneity, in combination with existing targeted therapy may change the course of tumor development (108).

Although stochasticity of gene expression is the best understood as a phenomenon that generates cellular variability in a genetically uniform cell population, stochastic processes also influence, rate of protein production, protein localization, cytoskeletal rearrangements, and formation of protein complexes, thus contributing to the cell-to-cell variability of the proteome (109). As a consequence, proteome variability results in variable drug responses in equally stimulated cells, as well as protein interaction that occur within the protein complexes (110-112).

1.6 Cancer cell plasticity

1.6.1 Epithelial-mesenchymal transitions (EMT)

Phenotypic plasticity can be defined as dynamic changes in phenotypes of the same genotype caused by intrinsic or extrinsic factors (58). In the field of cancer cell plasticity, one of the important pillars is EMT, which is directly associated with generating tumor heterogeneity. EMT is a highly organized program that imparts heritable phenotype changes through epigenetic mechanisms without new genetic alterations (113), and it has been associated with tumor initiation, malignant progression, tumor stemness, tumor cell migration, intravasation of tumor cells into the bloodstream, and resistance to therapy (114). Through EMT, epithelial cells lose their epithelial features such as apicobasal polarity, become migratory, and start expressing markers that characterize mesenchymal cells (115). EMT causes cells to gain properties of stem cells and transition to an undifferentiated or stem-like state, which correlates with acquisition of a stem-like gene expression program, and increased potential for self- renewal (116, 117). In the majority of cases, the metastases are the cause of cancer patient death (118). EMT's role in metastatic cascade has been extensively studied in the past years, from tumor invasion, intravasation,

extravasation and, colonization (119-122). New insight into existence and stability of EMT hybrid cells, which express both epithelial and mesenchymal markers, have seeded doubt into the classical view of EMT and its reverse process, mesenchymal to epithelial transition (MET), as indispensable for the completion of metastatic cascade (123-125). To understanding the process of EMT, it is important to define what mechanisms govern the transition of cells through different hybrid states. Research in this field shows that hybrid states commonly have similar tumor-propagating potential, but differ in cell plasticity and invasiveness. Surprisingly, this is not evident in cells that undergo MET, shifting weight of the sole process to other mechanisms that increase cell stemness (126). It is critically important to understand the mechanisms that give tumor cell tumor initiation capability from the rest of the tumor cells, and how cellular plasticity contributes to this heterogeneity.

1.6.2 Hybrid states/Transitional states

EMT has been traditional viewed as a binary process where epithelial cells lose their epithelial markers (E-cadherin), and acquire mesenchymal markers such as vimentin. However, this assumption was challenged by studies that revealed dynamic, gradual flow between stable and hybrid states that expressed both epithelial and mesenchymal markers with intermediate transcriptional and epigenetic features (125). In addition, these partial EMT cells are associated with more aggressive tumor features (124). When we assume parallels of EMT process and developmental processes found in the differentiation of cell, and especially when considering cell dynamics, it is not surprising to expect that EMT will be a multistep process instead of the bistable switch of individual cells from one state to another (127).

Studies on co-expression of epithelial and mesenchymal markers documented in various cell lines, xenografts, and primary tumors indicated increased cell survival, invasion, and migration. In addition, hybrid EMT cells were usually shown to be localized at the forefront of the tumor. Genome-wide transcriptional analysis coupled with bioinformatic analysis of the chromatin landscaped revealed unprecedented identification of epigenetic molecular mechanisms, transcription factors, and activation of signaling pathways that are responsible for maintaining gene expression states of distinct EMT hybrid cells in the process of EMT (126). A similar analysis from the same group also revealed the importance of lineage states of cancer cells in controlling EMT (128).

Recently, several research studies and reviews have elegantly described the existence of different EMT states and the role they play during tumorigenesis and metastasis formation. Of particular importance is the identification of genes associated with each EMT state that belong to a gene regulatory networks governed by the epigenetic constraints. This increased understanding of the mechanisms controlling different EMT hybrid states and their transition is undoubtedly important for understanding tumor heterogeneity.

1.6.3 Epithelial-mesenchymal transition (EMT) and cancer stem cells

Studies that uncovered direct connections between EMT and tumor stemness showed that genetically induced EMT transcription factors (Twist1 and Snail1) in mammary epithelial cells increased transplantation potential into serial recipients (116, 117). It is worth noting that cancer cell stemness is not always dependent on the process of EMT, and that cancer cells can metastasize without the involvement of EMT process (129, 130).

Many reports have indicated that the process of mesenchymal-epithelial transformation is required for colonization of metastatic cells (131, 132). However, there is an ongoing debate over whether MET is required for metastasis (131, 133). The proclivity of epithelial cancer to metastasize tends to be strongly correlated with their differentiation status, where poorly differentiated tumors show increased potential for metastasis compared to well-differentiated, “stationary” tumors (134, 135). Tumors with mesenchymal-like phenotypes are associated with enriched CSC phenotypes and are probably directly responsible for invasion and metastasis leading to an early demise of the patient following the initial diagnosis. It is believed that such plasticity occurs as a result of the EMT-related histogenetic mechanisms involving a series of well-defined transcription factors and signaling pathways. The EMT-related cellular plasticity is actively involved with metastatic spread of solid tumors during tumor progression (136, 137), and is also associated with the “dynamic stemness” model in which non-tumorigenic cells spontaneously acquire CSC characteristics through the EMT-like processes (138-140). EMT is critical for morphogenesis by interconverting epithelial cell types into cells with mesenchymal properties (115, 141). EMT programs activated in carcinoma cells endow cellular traits associated with high-grade malignancy, including the ability to complete steps of the metastatic cascade (137, 142). Certain epithelial cells that pass through an EMT acquire the self-renewing trait associated with either normal tissue stem cells (SCs) and cancer stem cells (CSCs) (116, 117).

1.7 Stochastic phenotype switching

During normal development or cancer progression, cells undergo phenotypic changes such as differentiation of cells, self-renewal of stem cells, immune response after inflammation, or epithelial to mesenchymal transition (EMT). Non-genetic phenotypic heterogeneity is a common phenomenon that is recognized in many genetically homogeneous cell populations, which are non-necessarily affected by extrinsic influences from the microenvironment (17). Interestingly, various phenotypes can be prominently different in specific cell populations, e.g. non-malignant cells can stochastically convert to neoplastic subclones, or cells of malignant phenotype in culture can revert to non-malignant and stable phenotype (143, 144). A switch between cell phenotypes (states) is driven by genome-wide gene expression changes that follow a specific design. These specific phenotypes are controlled by the genome-wide gene expression changes that are coordinated by the gene regulatory network (GRN). The gene expression control is preset in the information encoded in the genome (145). Each network of genes that are mutually influenced (directly or indirectly) can display many gene expression configurations, or attractor states, in the coordinated network system. This multistability presumes that there is robustness in the system since the cell can be characterized by stable phenotypes. However, it also assumes the presence of fluctuations between attractor states induced by stochastic gene expression at the molecular level, or from a directed signaling input (146). Such non-genetically derived variability can directly generate cells with malignant traits and can be subject to "mutation-less" Darwinian selection (89, 111). This is not to say that non-genetic dynamics undermines the role of selection in a genetically uniform cell population, but rather it increases the survival capacity of the original non-genetic variant and facilitates the natural selection process. Conversely, some mutations may increase non-genetic variability by shifting the coordinates of the GRN and creating new attractor states, or by increasing the gene expression noise by changing the sequence of regulatory proteins complexes that control transcription (147, 148). Stochastic phenotype heterogeneity can offer survival mechanisms by increasing the range of adaptation options under dynamic selective pressure in the microenvironment, and has a constructive role in the process of differentiation process (149). Probably the most well-studied example in how non-genetic stochastic phenotype heterogeneity is exploited is in the emergence of drug resistance (150), which is an alternative mechanism to the clonal mutation model, where acquired

resistance is gained through genetic mutations (28). Causes of stochastic phenotype heterogeneity are still unclear. What is known so far is that phenotype heterogeneity is determined by stochastic molecular interactions at the level of gene promoter, transcription, and translation, which all affects gene product dynamics, both between isogenic cells and temporally (93, 94, 148, 151). In summary, the coexistence of multiple stable steady-states, or 'attractors', in gene regulatory networks, coupled with stochastic gene expression, provides a framework for understanding cancer cell phenotype dynamics, and allows formulation of alternative mutation-less concepts of tumor progression (89).

1.8 Identification of novel tumor suppressor genes in HCC

1.8.1 Drosophila melanogaster genetic screens

Drosophila melanogaster represents a pivotal model system in which many pathways in developmental and signaling networks were originally discovered (152). Of the conserved signaling pathways, Ras and Notch were among the most extensive studied (153, 154). Study of molecular mechanisms of tumor development in the *Drosophila melanogaster* model has proven to be a reliable cancer model for studying oncogenes and cooperative tumorigenesis processes from cellular to tissue levels (155-157). Metastatic invasion is a complex process characterized by joint dysregulation of several signaling pathways (158). Considering that majority of human deaths occur due to metastatic dissemination, it is of prime importance to understand the molecular mechanisms that govern tumor growth and metastasis, and the role of tumor microenvironment in the these processes (159). The *Drosophila melanogaster* model is particularly suitable for studying human disease-causing genes because of its conserved genomic sequences, and epithelial tissue organization. It has been estimated that *Drosophila melanogaster* has approximately 75 % of the shared human disease genes (160). It also offers a valuable opportunity to genetically deconstruct cancer-relevant signaling networks in the context of the whole animal, important for identification of new biomarkers and new therapeutic strategies for a variety of epithelial cancers (161). Some of the most important and prevalent activating mutations are found in Ras genes (*Ha-RAS*, *N-RAS*, and *K-RAS),* of which *K-RAS*, is the major oncogenic driver of cell proliferation in human cancer (162). Mutation Ras^{V12} is

present in more than 30 % of human cancers (163, 164). Nevertheless, Ras-activating mutations are unable to progress to tumor growth due to senescence (165).

Recently, an elegant *Drosophila melanogaster* RNAi systematic reverse genetic screens were used to identify more than 6,500 conserved genes involved in tumor suppression of Ras^{V12}-induced epithelial tumors. The authors expressed *dRasV12, GFP* and *UAS-RNAi* constructs in eye imaginal disc of fly larvae using the *eyeless (ey)- FLP; actin flip-out* system (166), which allowed monitoring GFP-labelled hyperplasia and invasion in transparent third instar larvae (161). Furthermore, the authors identified conserved human gene orthologs of the fly hits, which were significantly downregulated in thousands of human cancer. Top candidate tumor suppressor genes were further validated in secondary *Drosophila* screen, and low expression of some of the hits significantly correlated with the mutation status of $KRas^{G12}$ and unfavorable prognosis in pancreatic cancer. Using this methodology, a recent study identified known and new cancer genes that cooperate with Ras oncogene in the context of invasive cancer phenotype (161).

1.8.2 Cathepsin F (CTSF) involvement in cancer

Human CTSF is a papain-like cysteine proteases with the MEROPS classification Clan CA, family C1, subfamily A (C01.018). CTSF is a small monomeric protein with endopeptidase activity. A key structural difference between CTSF and other family C1 proteases is a long pro-peptide sequence (251 residues) and a cystatin-like domain found within. Mouse and *Drosophila* homologs of CTSF are identified and share the same structural domains (signaling sequence, pro-domain, and active domain) (167). Genome localization of CTSF is found on chromosome 11 at 11q13 close to the evolutionarily related cathepsin W (168). Interestingly, the increased rate of loss of heterozygosity in this region is found in progressive stages of breast carcinoma, from early pre-neoplastic lesion to invasive breast ductal carcinoma (169). CTSF expression profiling by Northern blot in a range of human tissues and cell lines revealed a ubiquitous expression pattern of a single transcript (approximately 2.1 kb). High levels were detected in normal tissues such as prostate, testis, ovary, heart, liver, brain, pancreas, and skeletal muscle. Low levels were observed in the small intestine, colon, spleen, lung and kidney, while no transcripts were found in thymus and in the leukocyte cell population. In addition to the wide range of expressing tissues, high variability in expression levels suggests that CTSF may be involved in specific cellular processes in these tissues (170). CTSF is also expressed by

macrophages in atherosclerotic lesions. *In vitro* studies have shown that CTSF is secreted in its active form from cultured monocyte-derived macrophages where it is involved in degrading apoB-100, leading to the aggregation and fusion of low density lipoprotein (LDL) particles (171). CTSF has a regulatory role in the processing of invariant chain that is associated with (MHC) class II in macrophages (172). It is also involved in the hydrolysis of high-density lipoprotein (HDL) particles, and impairs their ability to remove cholesterol from macrophage foam cells (173). CTSF participates in the degradation of the extracellular matrix of the advanced atherosclerotic plaques, and AngII enhances secretion of CTSF from human monocyte-derived macrophages in atherosclerotic lesions (174). Recently, a gene expression microarray study on corneas from patients suffering from diabetic retinopathy identified CTSF as one of the up-regulated proteinases, suggesting its involvement in disrupting the integrity of the basement membrane. Functional aspects of these results were, in addition, confirmed in *ex vivo* model of cultured corneas, where adenoviral transduction of CTSF and MMP-10 in normal corneas reproduced the diabetic corneas (175). Specific inactivation of murine CTSF resulted in a lysosomal storage defect and progressive neurological features in mice (176), while mutations in CTSF are involved in Kufs disease Type B , an adult form of neuronal ceroid-lipofuscinosis (177). Finally, nuclear localization and modulation of activation markers in hepatic stellate cells suggests a new role of CTSF as a nuclear transcription factor (178). Unlike involvement of other cathepsins in cancer, studies trying to elucidate specific role of CTSF have not produced conclusive results. However, recent studies have identified CTSF alterations in specific cancers (161, 179, 180). Northern blot analysis showed upregulated levels of CTSF RNA in cancer cell cultures of colon, lung, melanoma, leukemia, and in HeLa cell line (170, 181). In addition, in a series of cervical cancer cell lines and tissue samples, CTSF mRNA was reported to be overexpressed. However, their *in situ* hybridization experiment on cervical tissue microarrays revealed inconclusive expression of CTSF tissue RNA (182). CTSF was also identified as a part of the degradome gene expression signature that was upregulated in long-term adapted metformin MCF-7 cell lines (183). Several published studies have directly or indirectly associated CTSF with tumor suppression properties. CTSF mRNA was decreased in an osteosarcoma cell line with high metastatic potential, compared to low metastatic cell line derivatives, while CTSF protein was not detected in both cases (184). Also, CTSF was downregulated in pediatric brain tumors (185). In a human proteome microarray study of serum autoantibodies, Yang *et al*. reported that CTSF and three additional serum biomarkers served as prognostic and survival markers in gastric

cancer patients (179). Yi *et al*. reported the first tumor-suppressive function of CTSF in gastric tumor samples, and its functional effects on proliferation and apoptosis in gastric carcinoma cell line (180). In addition, microarray analysis in the same study showed that CTSF downregulation was linked to ATM-signaling, p53- signaling, transcriptional regulation, and other cancer-related cellular processes. Authors also reported an association between clinicopathological features and CTSF mRNA expression in 44 gastric cancer patients (180). Tumor suppression in the context of genotoxic stress or oncogene-induced hyperproliferation is achieved by p53-activated cellular functions, including apoptotic cell death, cell cycle arrest, and cellular senescence (186). Accumulating evidence suggests that CTSF expression is closely associated with cell senescence. For example, CTSF was identified as a novel senescence maker in terminally senescent human epithelial prostate cells (HEPC) (187). CTSF was used as a marker of senescence in developing mouse thymocites in which oncogenic beta-catenin induced senescence (188). CTSF was upregulated but not indispensable in functional p53 cells upon activation of programmed cell death as a consequence of DNA damage (189). Analysis of p53 transactivation domain (TAD) mutants in *HrasV12* mouse embryonal fibroblasts (MEFs), identified CTSF as an additional p53-dependent target gene, involved in tumor suppression activity and downregulated in human and mouse cancers (190).

Finally, most conclusive proof that CTSF is involved in tumor suppressive processes came from an *in vivo* shRNA screens where knockout of CTSF in inactive p53-regulated apoptosis genes leads to development of lymphoma/leukemia. This study also showed that p53 driven tumor suppressor processes overlap and depend on oncogenic drivers in pre-neoplastic cells. (191).

2 Stochastic Phenotype Switching Leads to Intratumor Heterogeneity in Human Liver Cancer

2.1 Aims of the study

Tumor heterogeneity, both in morphological and functional aspects, is a common feature arising in hepatobiliary carcinoma, and it is caused either by genetic, epigenetic, or stochastic changes of gene expression. Our limited understanding of tumor heterogeneity presents multiple challenges in diagnosis and treatment of these cancers, especially in the context of personalized medicine.

Our first aim was to use immunohistochemistry (IHC) and gene expression profiling to characterize different regions within the sCC sample. In order to extend our analysis of intratumor heterogeneity, we established a primary cell culture, which enabled us to identify mechanisms through which the cellular variability arises. Clonal analysis of primary tumor cell cultures identified different types of self-propagating subclones characterized by stable (keratin-7 positive or keratin-7 negative) phenotypes and an unstable phenotype consisting of mixtures of keratin-7 positive and negative cells. Keratin-7, a member of the intermediate filament family, is normally expressed in association with keratin-19 in epithelia of bile ducts, and regulated by hypermethylation in breast and prostate cancer cell lines (192, 193). The specific aims in this study were:

1. Characterize newly-established cell line that recapitulates cellular phenotypes involved in sCC;
2. Validate genetic uniformity between different subclones derived from primary tumor tissue;
3. Use next-generation sequencing to fully characterize subclones with different keratin-7 phenotypes;
4. Demonstrate if switching between phenotypes is linked to changes in gene expression related to epithelial-mesenchymal transition;
5. Investigate whether DNA methylation is involved in the dynamics of keratin-7 expression;
6. Use xenotransplantation assay with different subclones to demonstrate that phenotype switching increases tumorigenicity and characterize resulting mouse tumors.

Our intention was to expand the knowledge base of the cancer heterogeneity-inducing mechanisms. We hope that our findings on stochastic phenotype instability in clonal populations using a differentiation marker keratin-7 would further shed insights on the involvement of trans-differentiation processes in tumor spread and distant-site colonization.

2.2 Methods

Please note that parts of this chapter have been published in (194).

2.2.1 Ethics statement

Primary tumor cell culture was established from fresh tissue originating from a surgically resected sarcomatoid cholangiocarcinoma of the liver. The study was approved by the research ethics committee of the Medical University of Graz (12-159 ex 01/02 and EK20-119).

2.2.2 Primary tumor cell culture

Primary tumor cell culture was established from fresh tissue originating from a surgically resected sarcomatoid cholangiocarcinoma of the liver. The study was approved by the research ethics committee of the Medical University of Graz (12-159 ex 01/02 and EK20-119). Tumor tissue was mechanically cut into small tissue slices which were placed into culture dishes in Dulbecco Modified Eagles Medium (DMEM) supplemented with 20 % FBS (Fetal Bovine Serum) and 1 % Penicillin/Streptomycin and cultured at 37°C in 5 % CO_2 for 1 month. Outgrowing cells were detached by mechanical scraping and transferring to 6-well plates where they were cultured for two months. From the original five dishes with outgrowing cells, only one dish contained cells that could be propagated for further passages. From this dish, clone C with epithelial morphology was isolated and maintained for 23 passages with regular freezing of cell aliquots between passages. Clone C (passage 16) was used for single-cell sorting (FACS-Aria, BD Biosciences) into 96-well plates. After 24 hours, plates were checked for wells containing single cells. After two weeks, expanded colonies (>50 cells) were either sub-cultured into 24-well plates, or directly stained in the wells with anti-keratin-7 antibody (Table 2-3), or used for further single-cell sorting. In total, we determined the keratin-7 phenotype in 1043 single cell-derived

subclones, comprising the three different keratin-7 expression phenotypes (i.e., heterogeneous K7het; positive K7pos; negative K7neg). In parallel, we selected 47 subclones (f1 and f2 subclones) and propagated them separately for further clonal studies.

2.2.3 Cell culture

All cells utilized in these studies were grown in DMEM (Sigma), supplemented with penicillin (Sigma-Aldrich, 100,000 units/L), L-Glutamine (Gibco, 2mM), and non-essential amino acids (NEAA; Gibco, 0.1mM). All manipulations were performed using aseptic technique. All harvesting of adherent cells was performed using trypsin (Sigma). Cell numbers and viability were determined with hemocytometer and tryptan blue staining, respectively.

2.2.4 Cell proliferation, migration and cell cycle analysis

Growth curves were measured by crystal violet to estimate cell proliferation. Cells were seeded at 10, 000 cells per well in 24-well plates. At designated time points (48 hours), cells were washed with phosphate buffered saline (PBS), fixed with 4 % formalin, followed by washing with PBS, and staining with 0.1 % crystal violet solution for 20 minutes at room temperature (RT). Cells were washed three times with water, and air-dried for five minutes. Absorbance was measured at 590 nm with Synergy4 (BioTek Instruments, Winooski, U.S.A.) All measurements were carried out in triplicates using three independent subclones belonging to each clonal keratin-7 expression phenotypes.

Measurement of cell migration was performed on monolayer cells by creating a “scratch” and taking images at the beginning and at defined time intervals during cell migration until the scratch was closed, as described (195).

Cell cycle of different subclones was measured by staining with propidium iodide (PI) and measured with flow cytometry. Approximately 0.5x106 cells were harvested by trypsinization, followed by washing with PBS and incubated at RT for 20 minutes with 50 µg/ml PI in 0.1 % Na-acetate with 0.1 % Triton X-100. Cells were evaluated using FACSCalibur (Becton-Dickinson) flow cytometer and analyzed with CellQuest Pro software (BD Biosciences).

2.2.5 FACS analysis and FACS-Aria Single-cell sorting

For characterization of the cellular composition of Clone C cultures, passage 10 cells were treated with trypsin/EDTA (Gibco/Thermo Fisher), washed, and re-suspended in PBS. Cells ($5x10^5$) were then labeled either with unconjugated antibodies directed to CD68, CD3, Ve-Cadherin or EpCam in combination with FITC-secondary antibody or APC-conjugated antibody directed to CD19 for 20 minutes at RT. Fluorescence antibody-stained cells were re-suspended in washing buffer and analyzed using s CytoFlex S flow cytometer (Beckman Coulter). For FACS analysis (FACSCalibur; BD Biosciences) and FACS sorting (FACS Aria; BD Biosciences), single-cell suspensions were fixed by Cytofix-Cytoperm kit (BD Biosciences) for 15 minutes at 4°C, followed by staining with anti-keratin-7 antibody (DAKO) and Alexa Fluor 488–conjugated secondary antibodies (Invitrogen). Double staining was performed with anti-keratin-7 antibody and Alexa Fluor 594-conjugated secondary antibodies (Invitrogen, Table 2-3) and fluorescein-conjugated human IL-13 R alpha 2 antibody (R&D Systems).

2.2.6 Lentiviral production

For lentivirus preparation, we used the second generation packaging system (196). Briefly, lentivectors were generated by calcium chloride transfection of green fluorescent protein (GPF) containing lentiviral vector, VSV-G expressing construct pMD.2 and the second generation pCMVΔR8.9 packaging plasmid in 293T cells. After transfection, supernatants were collected after 48 hours and passed through 0.45 μM syringe filter. K7pos subclones were infected with 5 ml filtered virus with 5 ml fresh media in which we added 10 μl of 8 mg/ml polybrene. Cells were incubated overnight, washed with PBS and incubated in fresh media for an additional 48 hours after which they were ready for GFP visualization.

2.2.7 RNA extraction

RNA extraction from cell cultures were performed using RNeasy Mini Kit (Quiagen) or with Trizol reagent (Thermo Fisher Scientific). Briefly, cells were washed in phosphate buffered saline (PBS) and resuspended in Trizol reagent. Tissue samples (50-100 mg) were homogenized using glass homogenizer and added to 1ml of Trizol. Phase separation was achieved by adding 200μl of chloroform per 1ml of Trizol. Samples were vortexed vigorously and spun at 12,000x g for 15 minutes at 4ºC. The clear top phase was removed and RNA further precipitated by addition of isopropanol (1:2 parts). Samples were left at

RT for up to 10 minutes and then further spun at 12,000x g for further 15 minutes at 4ºC. Isopropanol was then removed and RNA pellet washed in 1ml of 75 % ethanol and further centrifuged at 6,000x g for ten minutes. The ethanol was then removed, the pellet was air dried and resuspended in nuclease-free water.

Total RNA from formalin-fixed paraffin-embedded (FFPE) tissues blocks were extracted from five to six independent areas by needle biopsies that corresponded to different tumor components (Carc = carcinoma, Met = metastasis, and Sarc = sarcomatoid). Isolated areas were morphologically verified under the microscope. Carc components contained >66 % epithelial structures, Sarc components contained <33 % epithelial structures, while Met components had no visible epithelial structures (Figure 2-1A-D). RNA was isolated using RNeasy FFPE Kit (Qiagen). For quality control of isolated RNA, concentrations and purity were measured with Nanodrop 1000 (Thermo Fisher Scientific), and RNA integrity was analyzed by spectrophotometry and different amplicon lengths using quantitative reverse transcriptase polymerase chain reaction (qRT-PCR), as described (197).

2.2.8 cDNA synthesis

cDNA synthesis from extracted total RNA was performed using the High Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems). The following components were aliquoted per reaction mix: 2 µl 10x RT buffer, 0.8 µl 25x dNTPs (100mM), 2 µl 10x RT random primers, 1 µl RNase inhibitor (20U/µl), 1µL MultiScribe RT (50U/µL), 4.2 µl nuclease-free water to make up to 10 µl. The mixture was mixed, briefly spun and pipetted into a 0.2 ml Eppendorf tube. 2 µg RNA in nuclease-free water (to bring total volume to 10µl) was added. The reaction mix was incubated at 25ºC for ten minutes, followed by incubation at 37ºC for 120 minutes, and finished with denaturation at 85 ºC for five seconds. The samples were then stored at 4ºC or -20 ºC for-long term storage.

2.2.9 Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

qRT-PCR was carried out using 7900HT Fast Real-Time Sequence Detection System (Applied Biosystems), where cDNA was amplified with Power SYBR® Green PCR Master Mix (Applied Biosystems), per manufacturer's protocol. Transcripts of genes of interest were amplified with the sequence-specific primers and computed relative to

internal control (GAPDH) in each sample. qRT-PCR were analyzed by the $2^{-\Delta\Delta Ct}$ method (198). List of primers used for qRT-PCR are shown in Table 2-1.

2.2.10 RNA sequencing (RNA-seq)

For the cell clones, 500 ng of RNA was used for Illumina RNA-seq library construction using Illumina TruSeq™ RNA Sample Preparation Kit v2 according to the manufacturer's instructions, except that only one-third of the recommended volumes were used in each step. For the primary tumor samples, rRNA was depleted from 1 µg of the total RNA for each sample using the Ribo-Zero™ rRNA Removal Kit (Epicentre Biotechnologies). The rRNA-depleted RNA was pipetted into the Illumina TruSeq™ RNA Sample Preparation Kit v2 by resuspending it in 6.5 µl of Elute, Prime, Fragment mix. (For a detailed description of protocol see Appendix) The resulting RNA-seq libraries were quantified using the Library Quantification Kit from KAPA Biosystems. The libraries were sequenced on an Illumina HiSeq 2000 at the EMBL GeneCore (Heidelberg, Germany). Reads were mapped using TopHat, assigned to genes using HTSeq-count, and differentially expressed genes and per-gene-dispersion estimates were called using DESeq. For gene ontology analysis, we used GeneCodis (197, 199), and DAVID bioinformatics GO annotation and signaling pathway tool. For Gene Set Enrichment Analysis (GSEA), we used the online molecular signature database (MSigDB) (200, 201) with gene sets from C2 database, which contains 1892 curated gene sets that are collected from various sources including online pathway databases and knowledge of domain experts. P-values thresholds were set to 0.01.

2.2.11 Methylated DNA immunoprecipitation sequencing (MeDIP-seq) and MeDIP-qPCR

Genomic DNA was isolated from cells of subclones (K7het, K7pos, and K7neg) using the PureLink Genomic DNA Mini Kit (Life Technologies). Purified genomic DNA (4 µg in 120 µl) was transferred to a Covaris microtube and sonicated in a Covaris S2 sonicator using the following settings: time 7 minutes, duty cycle 10 %, intensity 5, cycles per burst 200, temperature 4°C, and power mode frequency sweeping. The sheared DNA was precipitated with 1 volume of AMPure beads (Beckman Coulter) and 1 volume of 30 % PEG_{8000}, 1.25 M NaCl, washed 2 times with 75 % ethanol and resuspended in 41 µl of 10 mM Tris, pH 8.0, 0.1 mM EDTA. DNA ends were blunted, A-tailed and Illumina TruSeq adapters were ligated using an in-house made version of Illumina's TruSeq DNA sample

preparation kit. The methylated DNA immunoprecipitation was performed essentially as described (http://www.roadmapepigenomics.org/), except a short oligonucleotide (AGATCGGAAGAGCGTC) was added to the denaturation reaction to prevent DNA fragments from annealing together by their adapter sequences. The libraries were amplified with the Kapa HiFi DNA polymerase (Kapa Biosystems) and sequenced on an Illumina HiSeq 2000 (EMBL GeneCore, Heidelberg). The reads were mapped to the human genome using Bowtie and differentially methylated regions were identified with DiffReps (202). MeDIP-qPCR was used to validate the methylated DNA immunoprecipitation and differentially methylated regions identified by MeDIP-seq. The following components were added per reaction mix: 4µl H_2O, 1µl Primer Mix (5 pmol/l), 10 µl KAPA SYBR® FAST qPCR Master Mix (2X) (Kapabiosystems), and 5 µl input ultrasonicated DNA.

2.2.12 Bisulfite pyrosequencing

Quantification of CpG methylation at *KRT7* promoter was performed by pyrosequencing with predesigned assays (Human_*KRT7*_01_PM PyroMark CpG assay, Qiagen). Pyrosequencing was performed with QIAGEN PyroMark. Detection and quantitative mutation analysis were performed by the inbuilt software (Pyrogram). Genomic DNA was extracted using QIAamp DNA mini protocol (Qiagen) from either stable subclones (K7neg, K7pos) or FACS-sorted keratin-7-positive cells or keratin-7-negative cells from K7het subclones.

2.2.13 5-aza-2'deoxycytidine (5-Aza-dC), Trichostatin A (TSA) and suberanilohydroxamic acid (SAHA) treatment

K7neg subclones were treated with 0.1 - 10 µmol/L 5-aza-dC (stock dilution in dimethyl sulfoxide (DMSO), 1mmol/L) for 24 - 96 hours. For co-treatment, cells were incubated initially with 5-aza-dC (0.5 µmol/L) for 48 hours followed by TSA (stock dilution in ethanol 1mmol/L) or SAHA (stock dilution in DMSO, 10mmol/L) for 24 hours. All drugs were diluted in culture medium for final concentration.

2.2.14 Bromodeoxyuridine (BrdU) analysis

Cells were grown on microscopic slides in 6-well plates and incubated in a culture medium with 10 µM BrdU for 12 hours. Cells were fixed in BD Cytofix/Cytoperm™ Buffer,

washed with BD Perm/Wash Buffer (BD Pharmingen), and stained with anti-BrdU antibody (Abcam) and Alexa Fluor 488. Double staining was performed with anti-keratin-7 antibody (Dako) and secondary Alexa Fluor 594-conjugated antibody.

2.2.15 NOD/SCID tumor xenografts

For heterotopic xenograft transplantation, six-week-old NOD/SCID mice were subcutaneously inoculated into their lateral flanks with $5x10^6$ cells and monitored daily until the tumors became palpable. We used two different subclones in early (10-15 passages) and late passage (>20 passages), each representing the three keratin-7 phenotypes (K7het, K7pos, K7neg). At the end of the experiment the animals were euthanized, the tumors were excised, weighed, and fixed in formalin for IHC analysis. All the animals were housed in individually ventilated cages in the Animal House Facility of BRFAA (Biomedical Research Foundation of the Academy of Athens) in pathogen-free conditions, in full compliance with the recommendations of FELASA (Federation of Laboratory Animal Science Association). The Greek Ministry of Agriculture (European Directive 86/609) approved all procedures concerning the protection of animals used for experimental purposes.

2.2.16 Immunohistochemistry

IHC was performed on formalin-fixed and paraffin-embedded (FFPE) tissue sections (2 μm). Deparaffinized and hydrated sections were exposed to heat-induced antigen retrieval in citrate buffer (0.1 %, pH 6.0) for five minutes in a pressure cooker at 125°C and gradually cooled in the cooker for ten minutes. Endogenous peroxidase activity was blocked with hydrogen peroxide (1.0 %) for ten minutes at RT. Sections were incubated with primary antibodies overnight at 4°C. After washing in PBS, the slides were incubated with biotinylated secondary antibodies (a summary of all antibodies used is shown in Table 2-3). After washing in PBS, bound secondary antibodies were detected and developed with the ABC system (Dako) with AEC or DAB (0.06 % 3,3'-diaminobenzidine with PBS, 0.003 % H_2O_2) as chromogenic signal detectors. Slides were counterstained with Hematoxylin and mounted in water-soluble Aquafix solution. As a negative control, primary antibodies were replaced by PBS.

2.2.17 Immunostaining (double and triple Immunofluorescence) on cells

Cells were either grown on coverslips and stained or directly stained in 96-well plates. Cells were fixed in ice-cold methanol for five minutes and acetone for 30 seconds. Cells were washed in PBS and incubated with primary antibodies diluted in PBS for 30 minutes to one hour at RT. This was followed by three, five-minute washes with PBS. Secondary antibody, diluted in PBS, was added for an additional 45 minutes, followed by four five minute washes with PBS. Secondary antibodies used were either Alexa Fluor 488 or 594-conjugated antibodies (Invitrogen). Samples were mounted in a Dako Mounting medium with DAPI (Invitrogen). Immunofluorescence signals were detected by Zeiss AxioPhot fluorescence microscope or by Zeiss LSM 510 META scanning laser confocal microscope.

2.2.18 Ion Torrent Amplicon Sequencing and Array-CGH

Ion Torrent sequencing was performed on DNA from cryopreserved primary tumor samples, non-neoplastic liver, and cellular subclones. DNA was extracted with the QIAamp® DNA Mini kit (Qiagen). Ten nanograms of DNA were used for highly multiplexed PCR of 190 amplicons covering 739 mutations in 46 cancer-related genes (Ion AmpliSeq Cancer Panel Primer Pool, Life Technologies). Subsequent processing was performed according to the manufacturer's protocols for library construction and emulsion PCR using the Ion OneTouch system. Sequencing was done on the Ion Torrent Personal Genome Machine (Life Technologies). Data analysis, including base calling, alignment to the hg19 human reference genome, and variant calling, was done using Ion Torrent Suite Software (Life Technologies).

For Array-CGH, DNA from a cryopreserved primary-tumor sample, and clonal cell pellets and reference human DNA were labeled with different fluorescent dyes (Cy3 and Cy5, respectively) and co-hybridized on, 8x60K Human CGH array (Unrestricted SurePrint G3).

2.2.19 Western blotting

SDS PAGE stacking gels were cast into clean glass plates according to standard recipes. The gel was then immersed into 1X Running buffer and pre-run for a short period. Whole-cell protein extracts from each clonal cell type with Laemmli loading dye (BioRad) in 1:1 ratio were added, and the gel was run at 90V until the samples entered into the lower

portion of the gel. The samples were run to the bottom of the gel to ensure optimal separation. Proteins were transferred onto nitrocellulose pore membranes using a semi-dry iBlot transfer device (Life Technologies). Noting the orientation of the membrane, the membrane was washed twice in Tris buffered saline with Tween (TBST) for five minutes, followed by blocking in 5 % skimmed milk TBST for one hour. The blocking solution was removed with two five minute washes with TBST after which primary monoclonal keratin-7 antibody (Clone OV-TL 12/30, Dako) and anti-β-actin antibody (Clone AC-74, Sigma-Aldrich), as a loading control, were applied. The primary antibody incubation was left overnight at 4ºC. Next, prior to addition of secondary antibody and 1- hour incubation, the membrane was washed in TBST 3-4 times for five minutes. Four final five minutes TBST washes were performed.

ECL detection reagents (GE Healthcare Life Sciences) with a horseradish peroxidase-conjugated secondary antibody was used according to manufacturer's instructions and developed. The length of exposure on X-ray film was determined empirically.

2.2.20 Statistical analysis and software

Values are expressed as mean ± SD. qRT-PCR validation is represented in a $\log_2$ scale (Student's t-test: *p<0.05, **p<0.01, ***p<0.0001; n = 3-4 f1 clones). DiffReps (202) were used to associate differentially methylated regions (DMRs) with differentially expressed genes (DEGs). P-values were calculated on the hypergeometric distribution for DMRs. Cumulative tumor weights between different clonal phenotypes were calculated by One-way ANOVA with uncorrected Fisher's LSD test, **p<0.01, ****p<0.0001. DEGs were identified using the data analysis package DESeq (203). Statistical analysis was performed using the Prism 8 Graph Pad software.

2.3 Results

Please note that parts of this chapter have been published in (194).

2.3.1 Clinical features of the patient

A 72-year old female reported a weight loss of approximately 10 kg, within a six-month time frame, and increasing fatigue. The patient underwent an upper abdominal ultrasound showing a lesion about 5 x 6 cm, without bone metastasis, but with several nodules in the

lungs. CT showed a 9 cm expansion in the right liver lobe that was analyzed by biopsy surgically removed. Laboratory tests performed the day before the operation showed the following results: CA19-9, 5700U/ml; AFP, normal; slight anemia (Hb 10.9 g/dL), slight thrombocytosis (463,000 cells per mm^3) monocytosis (1,200 cells per mm^3), alkaline phosphatase 314 U/L, GGT 137U/L, fibrinogen 1,000 mg/L and CRP 147mg/L. The patient was not subjected to chemotherapy. One month after surgery, the patient was hospitalized again because of dyspnoea and thoracic pain. CT of the thorax showed a massive pleural effusion as well as multiple nodular lesions measuring up to 7 mm. Due to the patient's poor general condition, no chemotherapy was undertaken, and the patient died four weeks after the surgery.

2.3.2 Pathological features of the tumor

The tissues were fixed in buffered formalin and embedded in paraffin, and 2-4 µm thick sections were subjected to histopathological and IHC analyzes. Eight different tumor blocks showed heterogeneous histopathological attributes with regionally distinct invasive tumor components (satellite nodules) and lymph node metastasis (Figure 2-1A-D). Tumor components with tubular structures were classified as the carcinomatous component, while satellite nodules with predominant spindle cells, which were also present in the lymph node metastasis, were classified as the sarcomatoid component (Figure 2-1B). Furthermore, the carcinomatous part contained areas with mixed cells showing sarcomatoid and carcinomatous features (transitional component) (Figure 2-1A). Such transitional areas have been previously suggested to be the result of the metaplastic transformation of epithelial cells to mesenchymal-like cells (204-206). A lymph node metastasis (Met) showed sarcomatoid features (Figure 2-1B, C). The surgically resected liver tumor was diagnosed as intrahepatic cholangiocarcinoma with a sarcomatoid component.

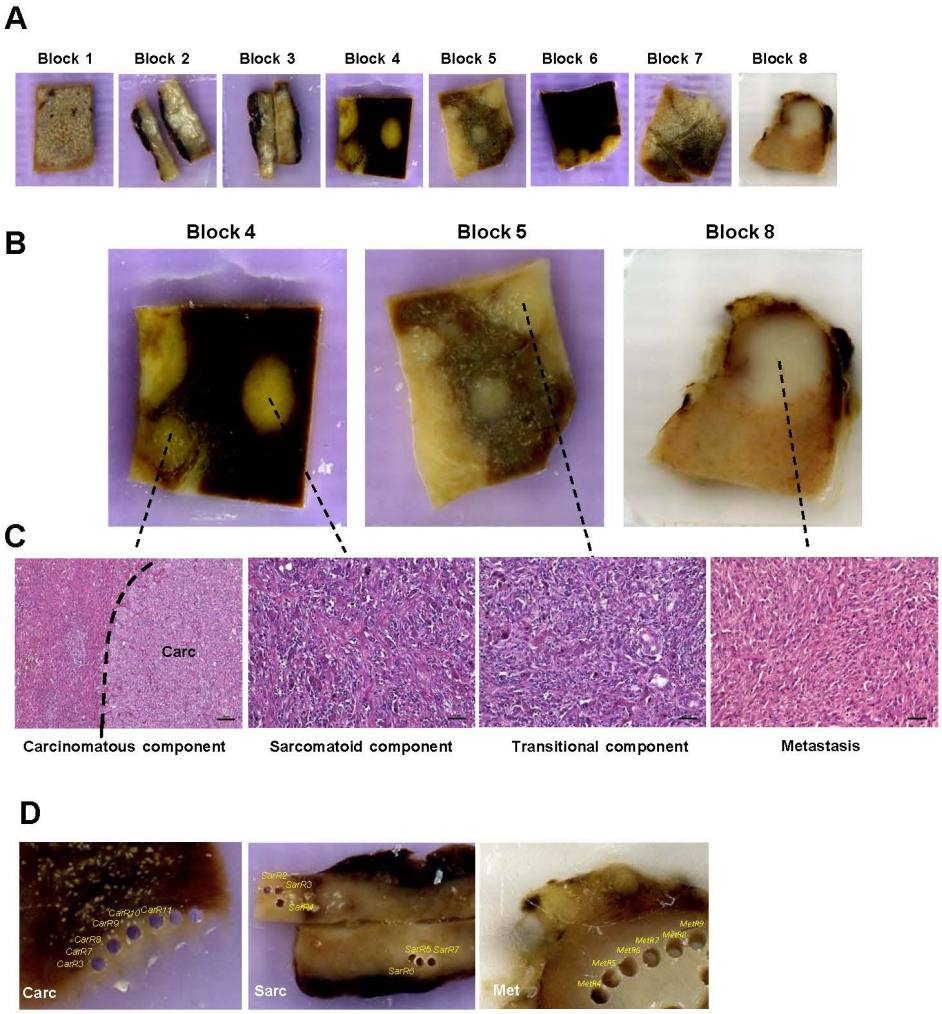


Figure 2-1. Overview of the FFPE blocks and their morphological features.

(A) FFPE blocks used for IHC and RNA-seq analysis. Carcinomatous features are found in tumor blocks 4, 5, 6 and 7; sarcomatoid features are found in tumor blocks 2, 3, 4, 5 and 8 (metastasis), transitional features are found in tumor blocks 3 and 5. Non-neoplastic liver = Block 1. (B and C) Higher magnification of distinct areas of the tumor displaying carcinomatous component (Block 4, dashed line marks the border between non-neoplastic liver and the carcinomatous component), sarcomatoid component (Block 4) and transitional component (Block 5) with mixed carcinomatous and sarcomatoid elements. Block 8 with the lymph node metastasis showing sarcomatoid differentiation. Scale bar: 50 µm. (D) FFPE blocks showing areas with core holes from which RNA was isolated for RNA-sequencing. (Adapted from (194)).

2.3.3 Transcriptome profiling of invasive and metastatic features inside sarcomatoid cholangiocarcinoma

To further characterize the heterogeneity in different tumor regions, transcriptional profiles by RNA-seq of each component were generated. Samples from tumor areas with defined morphological features were collected by core biopsies taken from the FFPE blocks (Figure 2-1D). Comparison of the sequencing profiles showed more similarities in gene expression between sarcomatoid and metastatic than between the carcinomatous and sarcomatoid or metastatic components, supporting the notion that metastasis has emerged from the sarcomatoid component (Figure 2-2A). Furthermore, the comparison between the carcinomatous and sarcomatoid tumor components (Carc-Sarc) and the carcinomatous and metastatic tumor component (Carc-Met) identified 585 and 1418 DEGs, respectively (log_2 fold change ≥ 2.0, p(adjusted) < 0.01), characteristic of the invasive and metastatic tumor components (14, 207). Gene ontology analysis with functional annotation clustering of common upregulated genes with sarcomatoid differentiation (Sarc and Met) were associated with mesenchymal differentiation (*MMP2, MMP11, CDH2, VIM, MCAM, SERPINE1*), invasive/migratory related genes (*PDPN, VEGFC, TLN2*), EMT-related genes (*TWIST1, TGFBR1, ZEB1, ZEB2*) as well as stem/self-renewal genes (*NCAM1, PROCR, ANPEP*), (Figure 2-2D, upper panel, Table 2-4), while common downregulated DEGs in components with sarcomatoid differentiation (Sarc and Met) identified cell-cell junction related genes (*DSP, JUP, PVRL4, GJB3, GJB4, GJB6, TJP2, TJP3, CGN*) and epithelial differentiation-related genes (*CDH1, CDH3, CRB3, OCLN, DDR1, EPCAM, CLDN3, FGFR2b*) (Figure 2-2D, lower panel; Table 2-4).

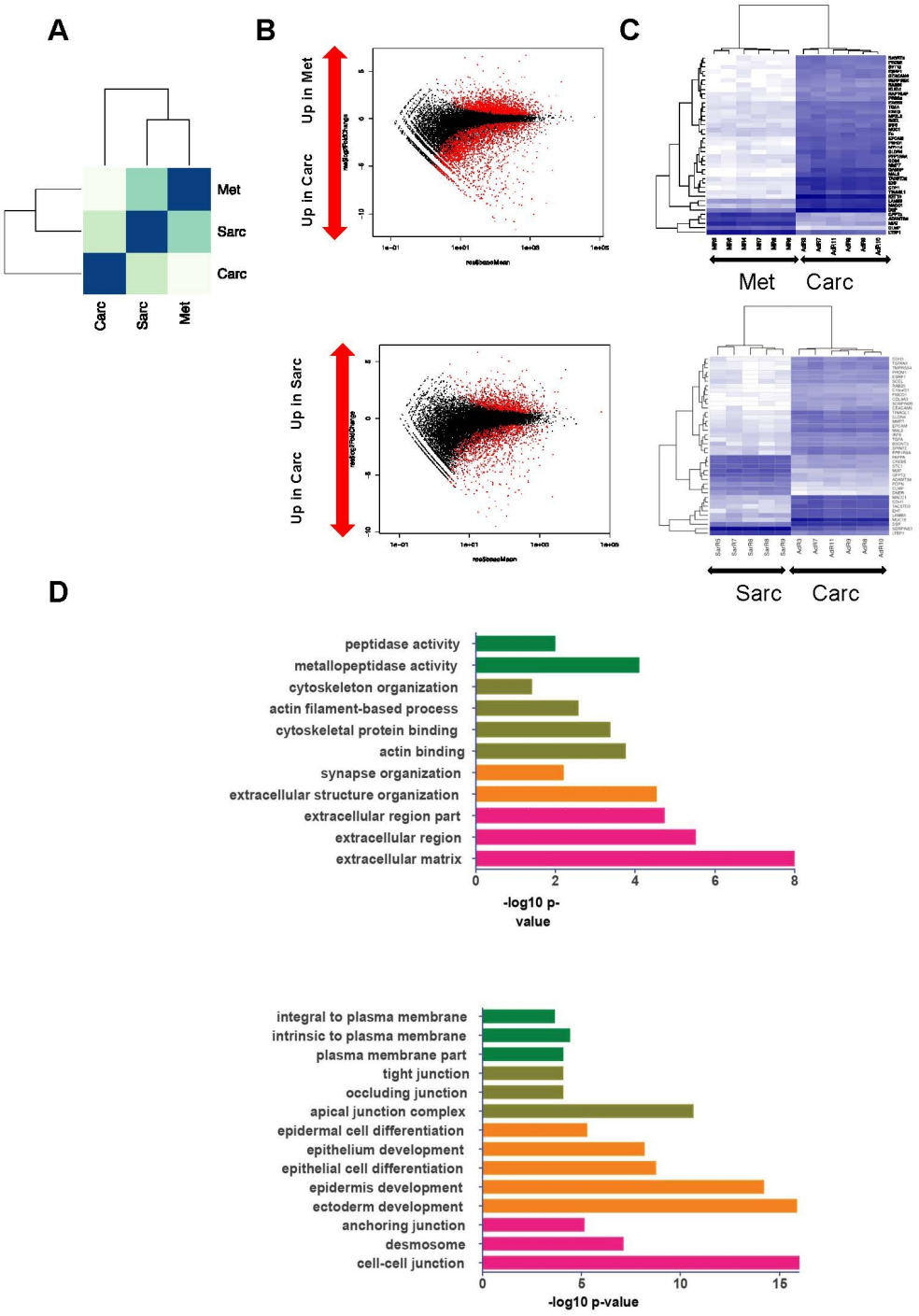


Figure 2-2. RNA-seq transcriptome analysis of different tumor regions in sarcomatoid cholangiocarcinoma.

(A) Heatmap and dendogram of sample-to-sample Euclidean distances of transcriptome profiles between different tumor components. (B) MA plot of differentially expressed genes identified between Carc, Sarc and Met tumor compartment. Dots represent individual gene expression. The x axis is the average expression over the mean of

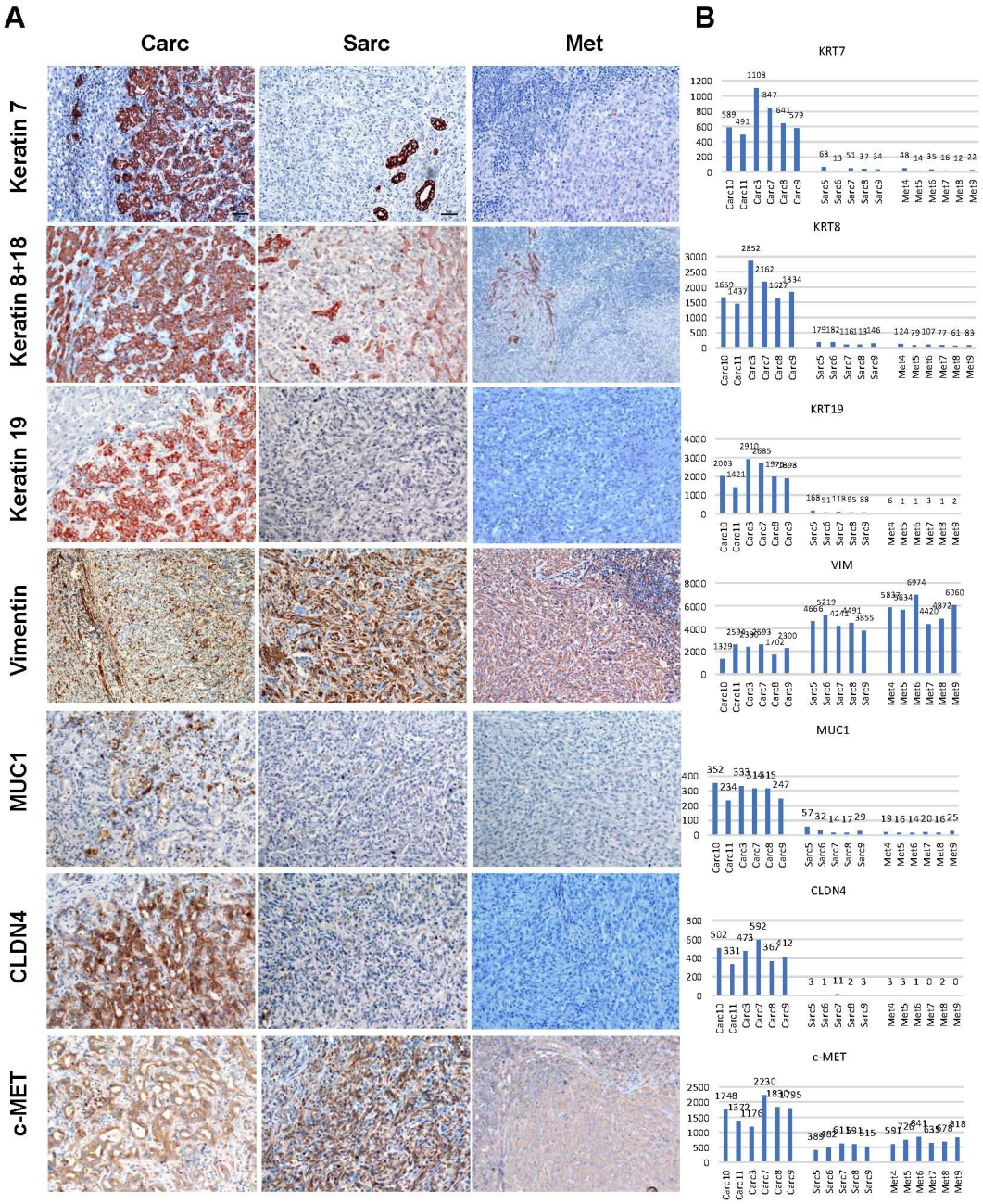
normalized counts (A-values), the y axis is the log2 fold change between Sarc/Carc groups or Met/Carc group (M-values). Genes with an adjusted p-value below a threshold are marked in red color. (C) Heatmaps showing the expression data of the top 40 differentially expressed genes. Each cell in the matrix represents the expression level of a gene feature in an individual sample. Dark-blue and white in cells reflect high and low expression levels, respectively. (D) Enriched gene ontology categories of common upregulated genes (n = 105) derived from (Sarc) component and lymph node metastasis (Met) relative to the carcinomatous (Carc) component (upper panel); Enriched gene ontology categories of downregulated genes (n = 391) derived from (Sarc) component and lymph node metastasis (Met) relative to the carcinomatous (Carc) component (lower panel). (Adapted from (194)).

2.3.4 Immunohistochemical characterization of the tumor

We validated RNA-seq data by performing IHC analysis on FFPE samples representing distinct tumor compartments used in the preparation of sequencing libraries as well as routinely used markers for differential diagnosis (Figure 3-3A, B).

Markers for biliary epithelial cells (keratin-8 and keratin-18, keratin-7, keratin-19, and MUC-1) and cell-cell adhesion proteins (CLDN4, E-cadherin, β-catenin) were strongly expressed in the carcinomatous component, indicating polarized and epithelial differentiation (Figure 2-3A). The sarcomatoid tumor component and the metastasis showed weak or negative expression of the above-mentioned proteins while displaying increased expression of the mesenchymal marker vimentin. The sarcomatoid component and the metastasis showed increased staining with the cellular proliferation marker Ki-67, which suggests an increased malignant phenotype (Figure 2-3A).

Furthermore, we analyzed markers reported to be associated with cholangiocarcinoma progression or poor patient survival. Membranous expression of the HGF receptor (c-MET) and CD44 and overexpressed p53 in the nucleus were observed in all tumor components. Carcinoembryonic antigen (CEA) staining was confined to the carcinomatous component with no expression in sarcomatoid component and the metastases (Figure 2-3A).



Continued on the next page.

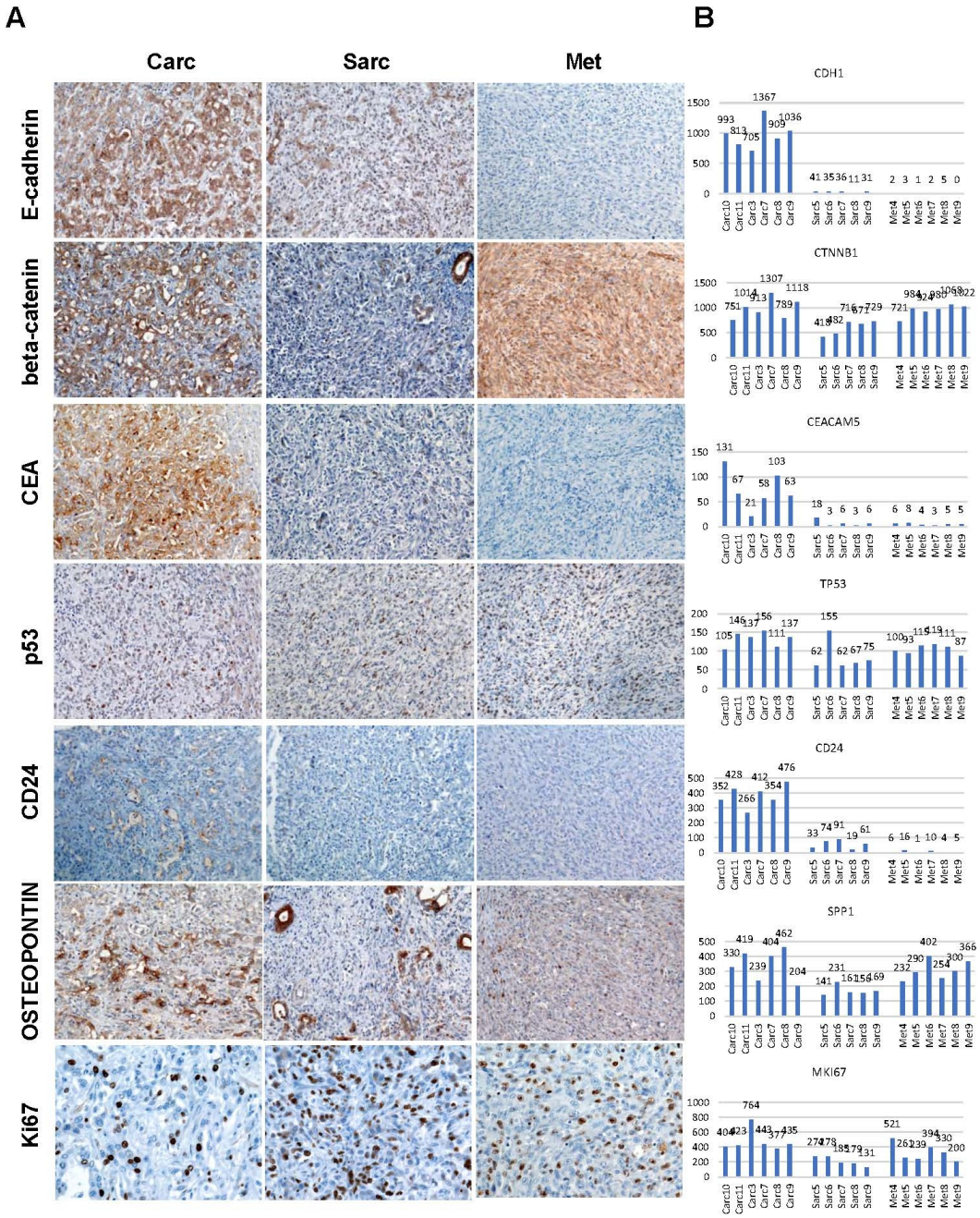


Figure 2-3. Immunohistochemical characterization and RNA-seq normalized read count data of selected differentially expressed genes in different carcinomatous, sarcomatoid and metastasis tumor compartments.

(A) Immunohistochemistry profiling of serial sections from primary tumor compartments. Immunohistochemistry was performed using specific antibodies against KRT8+18, KRT7, KRT19, Vimentin, MUC1, CLDN4, c-MET, E-cadherin, beta-catenin, CEA, and p53. Stained sections were developed with AEC (red color) or DAB (brown color) substrate-chromogen. Proliferative tumor score was based on average counts of Ki-67-positive cells in seven random high-power viewing. (B) Normalized RNA-seq read count values of

differentially expressed genes between Carc, Sarc and Met tumor compartments. Magnification A: 200x (Ki-67: 400x) (Adapted from (194)).

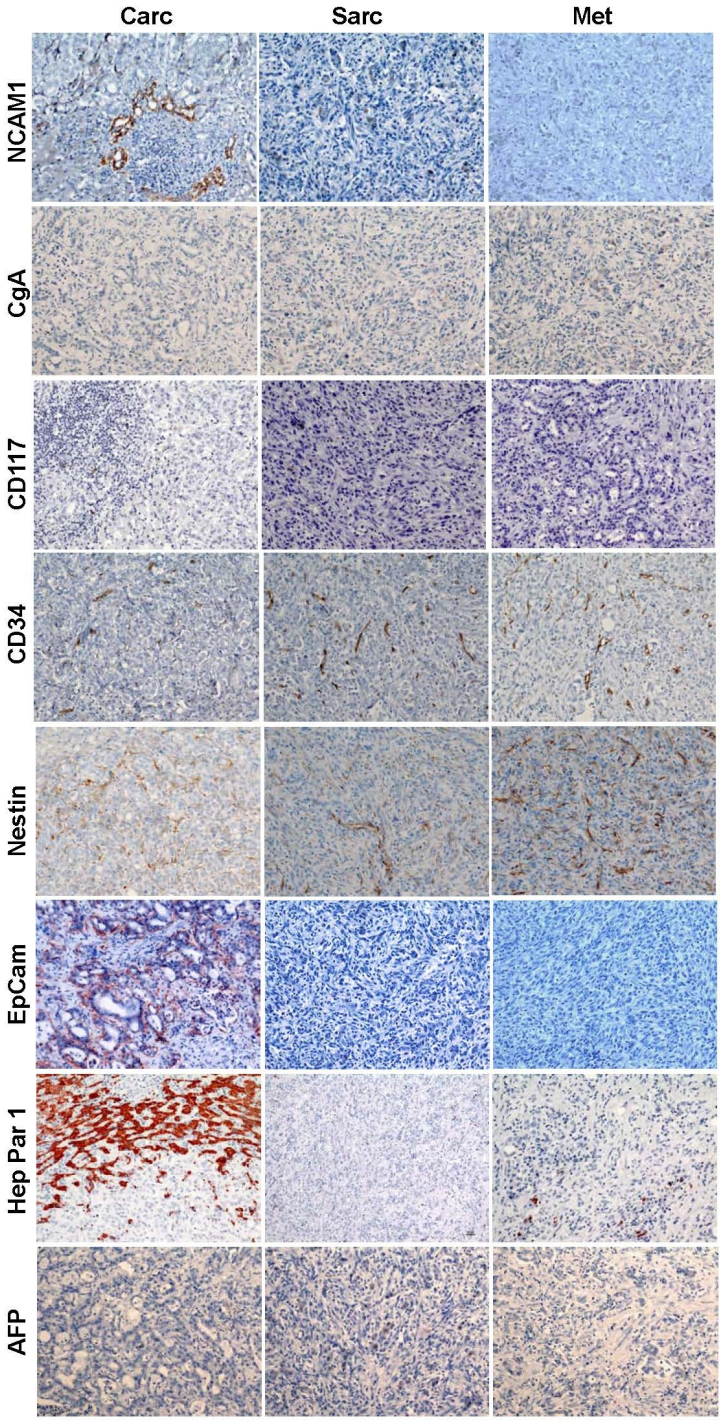


Figure 2-4. Immunohistochemical analysis of hepatic stem/progenitor cell markers and hepatocytic lineage associated markers.

Immunohistochemical staining of carcinomatoid, sarcomatoid and metastatic primary tumor compartments using antibodies against NCAM1, CgA, CD117, CD34, Nestin, EpCAM and HePar1 and AFP. Magnification x200. (Adapted from (194)).

IHC analysis with hepatic stem/progenitor cell markers (NCAM1, CD34, CgA, CD117, and Nestin) and hepatocytic lineage-associated markers (Hep Par 1 and AFP) were negative in tumor cells (Figure 2-4). These results suggest that the primary tumor was derived from the intrahepatic biliary duct epithelial cells with no apparent contribution of hepatic stem cells/progenitor cells to intratumor heterogeneity (Table 2-6). Also, a significant correlation between protein (immunohistochemistry scores) and mRNA expression was achieved between carcinomatous and sarcomatoid tumor compartments for selected cholangiocarcinoma-specific, EMT and pluripotent hepatic markers (data not shown).

2.3.5 Establishment and culture of primary cells

Initially, 5 culture dishes with cells outgrowing from the tumor were generated. The primary tumor cell culture most likely was a mixture of different cell types, such as malignant, non-malignant epithelial and mesenchymal cells. However, cells of only one of the dishes could be further propagated (dish C which gave rise to clone C), whereas cells of the other dishes stopped dividing, most likely because they were non-tumor cells (Figure 2-5).

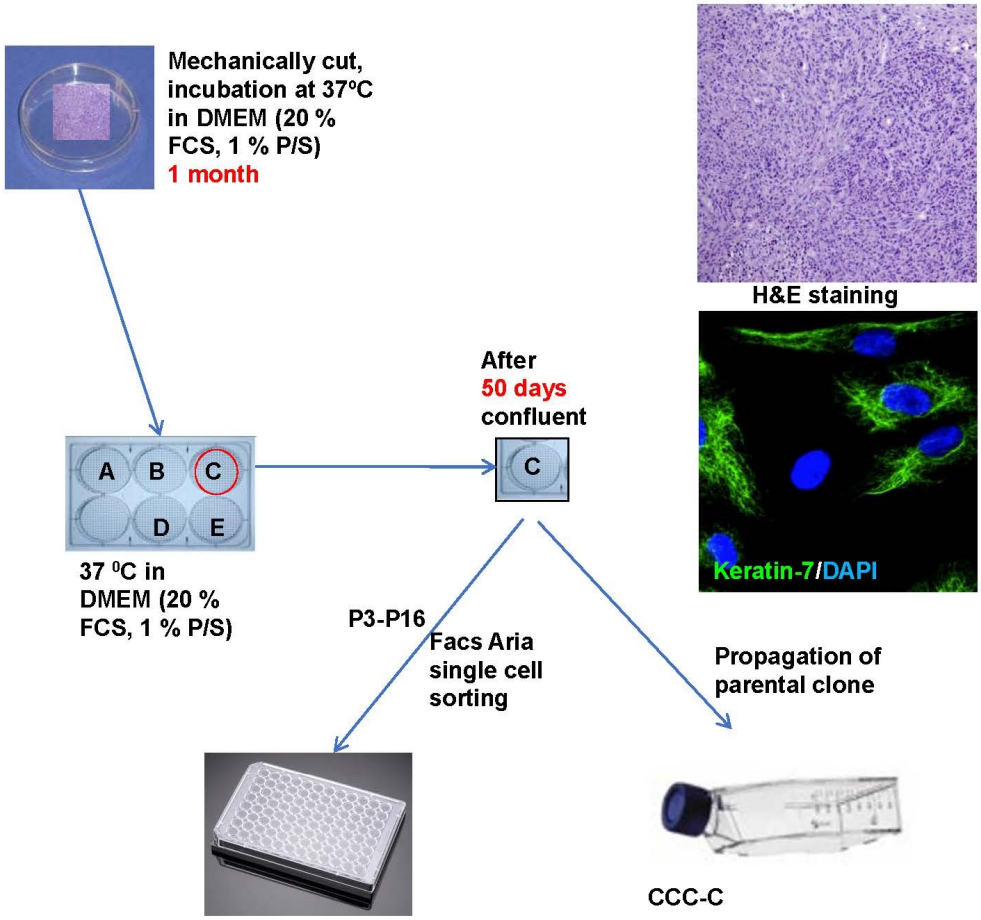


Figure 2-5. Establishment of cell culture derived from sCC case.

To exclude the possibility that the clone C isolated from tumor tissue is not of epithelial origin or is a mixture of epithelial, and non-epithelial cells, we performed FACS characterization of early passage clone C (passage 10) using the following antibodies: positive control: EpCam (epithelial cell marker), CD68 (macrophage marker), CD19 (B-cell marker), CD3 (T-cell marker), Ve-cadherin (endothelial marker). Furthermore, to exclude the presence of cells with muscle differentiation, we have performed immunofluorescence staining on cultured cells with antibodies directed to alpha-smooth muscle actin (alpha-SMA; data not shown). As expected, the EpCam antibody stained cells positively, while markers to lymphocytes, macrophages, endothelia, and smooth muscle cells were negative (Figure 2-6A-G).

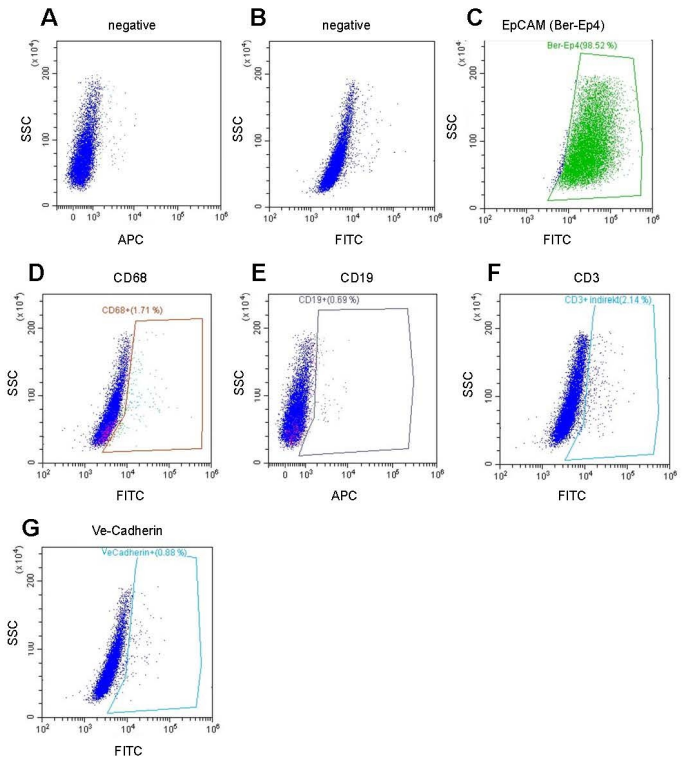


Figure 2-6. Cells culture of clone C were analyzed by flow cytometry for expression of common markers.

(A-B) Negative control: secondary fluorescence antibody only; (C) Positive control: epithelial marker (EpCam); D) macrophage marker (CD68); (E-F) immune cell marker (CD19 and CD3); G) endothelial cell marker (Ve-Cadherin). (APC: allophycocyanin; FITC: fluorescein isothiocyanate). (Adapted from (194)).

All further studies were performed with cells derived from the primary culture dish C and cellular subclones generated by single-cell cloning (Figure 2-5).

The IHC cancer- and lineage-specific markers (to distinguish biliary epithelial cells from hepatocytes), used for IHC profiling of the primary tumor were also used to characterize cells derived from dish C (Table 2-3, Figure 2-7A). Surprisingly, among cytokeratin markers (keratins-8, 18, 19 and 7), only keratin-7 expression presented a heterogeneous staining pattern, displaying mixed subpopulations of keratin-7 positive and keratin-7 negative cells (Figure 2-7A). We also investigated between keratin-7 subclones for differential expression of markers reported to be associated with CSC or involved in upregulated signaling pathways that contributed to cholangiocarcinoma proliferation and invasion. We evaluated IF staining for c-Met, CD44, beta-catenin, and EpCam. All markers showed no differences between the analyzed subclones, but did show heterogeneous staining between the cells inside the subclones (Figure 2-8). Our first

hypothesis to explain different phenotypes was that culture is a mixture of different cell types. To test this hypothesis, we have performed single cell cloning (Figure 2-5).

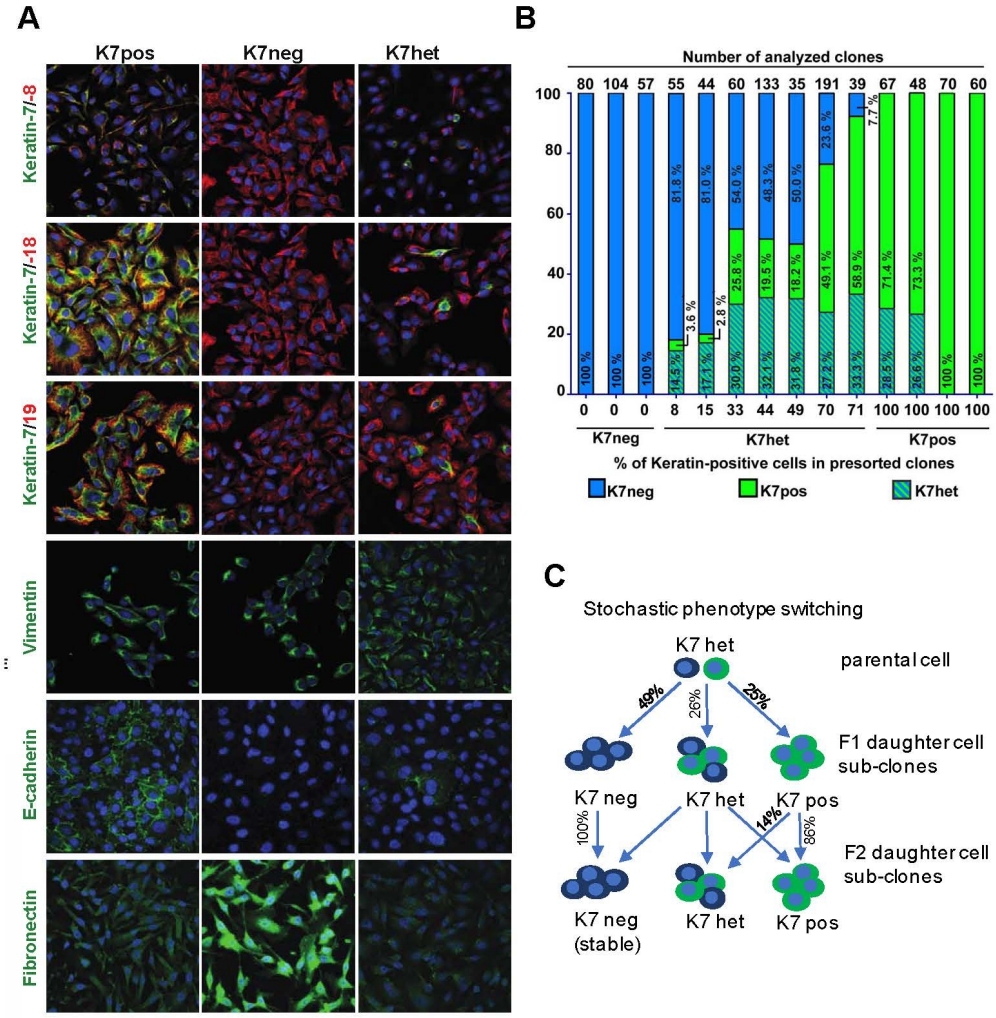


Figure 2-7. Characterization of K7pos, K7neg, K7het clonal cell populations.

(A) Immunofluorescence microscopy of different clonal phenotypes. K7pos, K7neg and K7het cells were analyzed with triple-label immunofluorescence staining (keratin-7 in combination with keratin-8, 18, 19 and DAPI as nuclear stain) or double-staining (DAPI in combination with vimentin). Scale bar = 20 µm; (B) Maintenance of keratin-7 phenotypes in daughter cell subclones. Proportions of keratin-7 phenotypes (y-axis) in cells derived from single-cell sorted parental cell clones with different K7-phenotypes (K7neg, K7het, K7pos) with a variable percentage of keratin-7 positive cells (numbers at x-axis). Each column represents results from a separate experiment for which the total number of analyzed clones is indicated at the top of the column; (C) Schematic demonstration of clonal phenotypes (K7neg, K7pos, K7het) in daughter cell subclones derived from single cells of a K7het parental clone showing stochastic phenotype switching. The average percentage of phenotype propagation to daughter cell subclones are derived from 1043 sub-clonal phenotype analyzes shown in (B); (D) Immunofluorescence microscopic characterization of different keratin-7 subclones using antibodies against E-cadherin and

fibronectin (green). Nuclei were stained with DAPI, 4,6- diamidino-2-phenylindole. Scale bar: 20 µm. (Adapted from (194)).

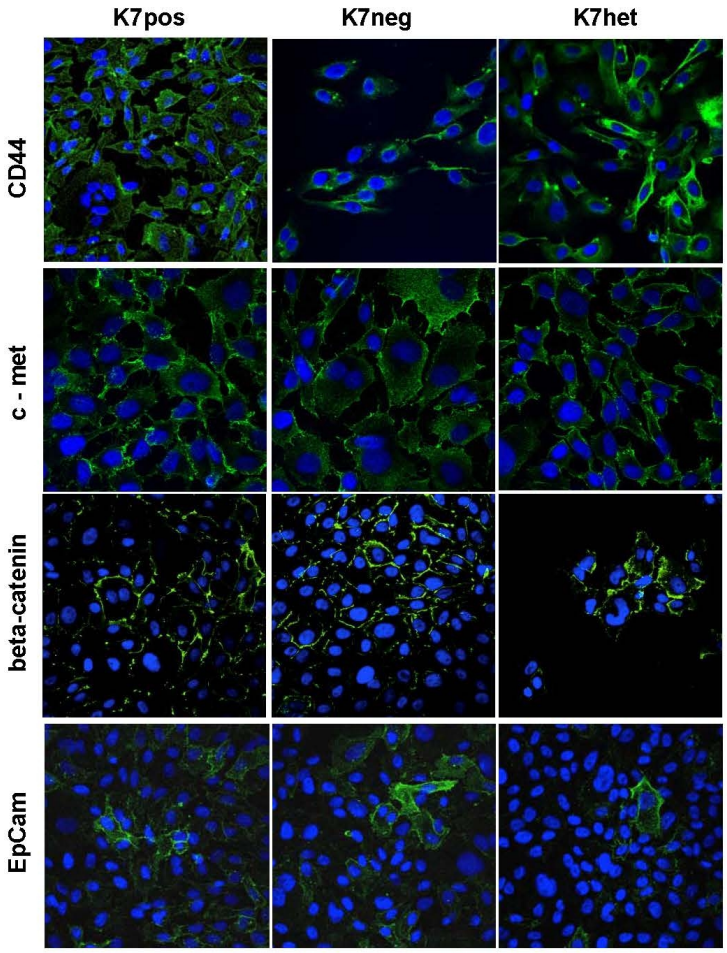


Figure 2-8. Immunophenotypes of K7pos, K7neg, and K7het subclones.

Immunofluorescence microscopy of different clonal phenotypes using antibodies against CD44, c-met, beta-catenin, EpCam (green). Nuclei were stained with DAPI, 4,6-diamidino-2-phenylindole, Magnification: 630x (c-met), 400x (CD44, beta-catenin, EpCam). (Adapted from(194)).

2.3.6 FACS Aria single-cell sorting

To characterize the mechanisms that govern intratumor phenotype heterogeneity, we used primary cell culture, Clone C (passage 16) for single-cell sorting (FACS-Aria, BD Biosciences) into 96-well plates. After 24 hours, plates were checked for wells containing single cells. After 2 weeks, expanded colonies (>50 cells) were either subcultured into 24-well plates or directly stained in the wells with anti-keratin-7 antibody or used for further single-cell sorting. We categorized the subclonal types into keratin-7 positive (K7pos), keratin-7 negative (K7neg) and keratin-7 heterogeneous (K7het) based on keratin-7

expression in single cell-derived subclones (Figure 2-7A). Parallel to this, we selected 47 subclones (f1 and f2 subclones) and propagated them separately for further clonal studies. Immunofluorescence profiling revealed the concurrent expression of epithelial-specific markers keratin-8, keratin-18, keratin-19, as well as expression of the mesenchymal marker vimentin (Figure 2-7A), indicating the metastable epithelial and mesenchymal nature of the cell subclones (136, 208, 209). We analyzed the stability of the keratin-7 phenotype in a total of 1043 single cell-derived daughter subclones (Figure 2-7B). K7het subclones yielded daughter subclones corresponding to all three keratin-7 sub-clonal types, that is, K7neg, K7pos and K7het (Figure 2-7B, C). On average, approximately 26 % of all K7het daughter subclones maintained the non-committed K7het phenotype with the ability of stochastic phenotype switching between K7pos and K7neg phenotypes. Interestingly, 14 % of K7pos subclones (which were originally derived from K7het parental clones) gave rise to K7het daughter subclones, which demonstrates the reversibility of phenotype switching (Figure 2-7B, C). In contrast, all K7neg subclones stably propagated their K7neg phenotypes to their daughter subclones, which suggests that K7neg subclones have acquired a stable phenotype (Figure 2-7B,C) Interestingly, we observed that in K7neg subclones, E-cadherin expression was downregulated (210) and independent of culture cell confluence, whereas fibronectin expression was upregulated, suggesting that K7neg subclones were residing in a more mesenchymal state (211) (Figure 2-7A).

2.3.7 Characterizing features of clonal cell populations.

Consistent with the keratin-7 staining patterns in keratin-7 clonal types, we detected high *KRT7* expression at the mRNA level in K7pos and K7het clones compared to negligible amounts found in the K7neg clone (Figure 2-9A). Western blot analysis supported concomitant expression in whole-cell lysates from corresponding clonal phenotypes and confirmed the specificity of the anti-keratin-7 antibody used in this study (Figure 2-9B). Morphologically, cells displayed dynamic changes in shapes and size, which were dependent on cell culture confluence (Figure 2-9C). There were no differences among keratin-7 clones in cell proliferation, migratory potential, or cell cycle occupancy, and all clonal phenotypes were capable of prolonged *in vitro* growth (>40 passages), (Figure 2-9D-F). K7neg subclones displayed robust heterogeneity where keratin-7 expression in this clonal population was highly consistent between measurements (Figure 2-10A). Also, there was no difference in the cell size and granularity between K7-positive and K7-negative

cells in K7het subclones (Figure 2-10B). We observed that K7neg subclones had significantly higher self-renewal capacity determined by the frequency of total established single-cell colonies compared to K7pos and K7het subclones under conventional culturing conditions (Figure 2-10C).

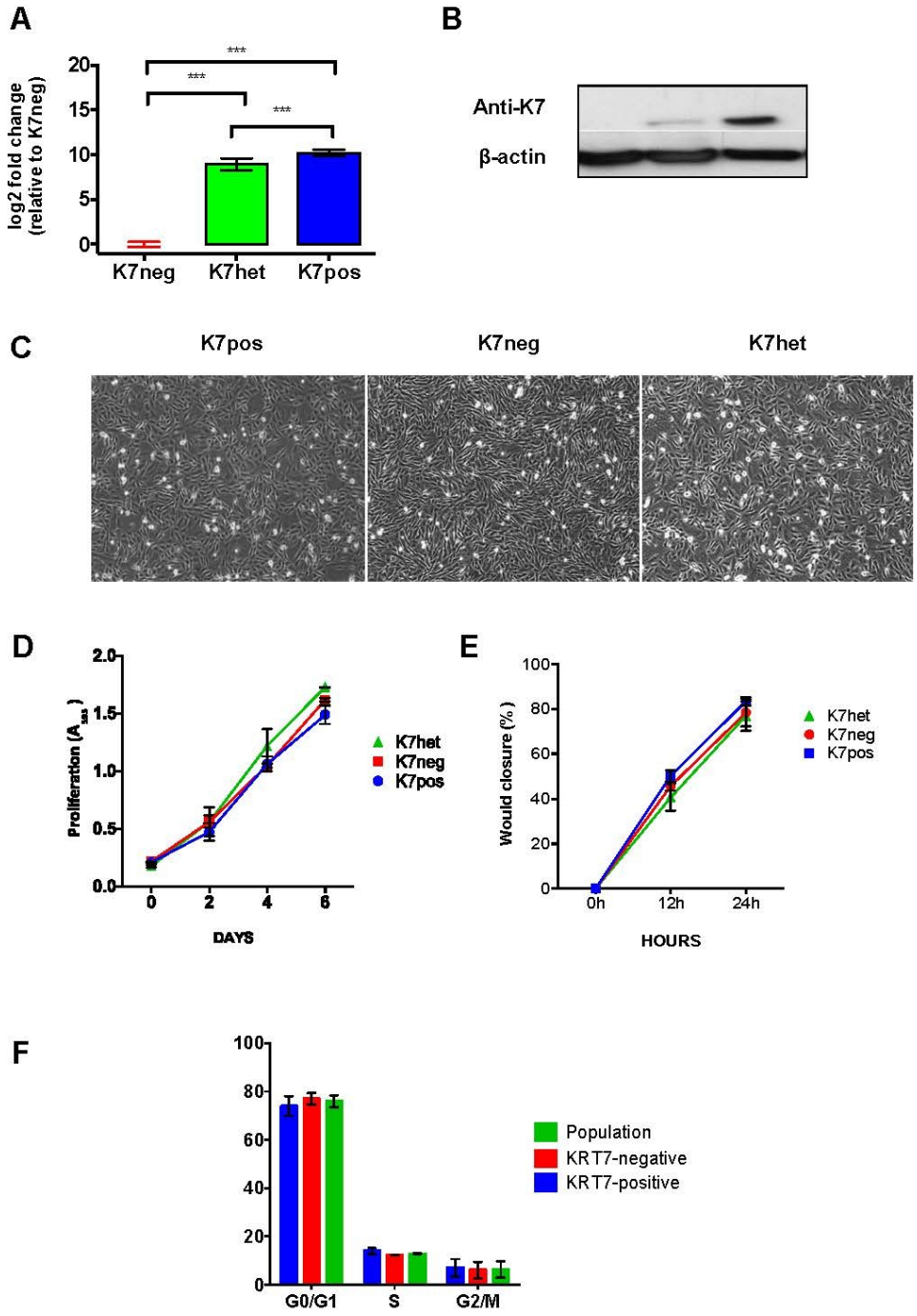


Figure 2-9. Characterization of cell features in K7 subclonal types.

(A) KRT7 mRNA levels in K7pos, K7neg, and K7het subclones. Results represent the mean and standard errors from eight independent qRT-PCR experiments (p-value < 0.05 by Student's t-test); (B) Western blot analysis of KRT7 expression in different clonal

subpopulations; (C) Phase-contrast images of K7pos, K7neg and K7het subclones; (D) *In vitro* proliferation curves of parental clone and differentially KRT7 expressing subclones. Cell content was determined by crystal violet staining; (E) Wound closure potential of subclones; (F) Cell-cycle analysis in keratin-7 positive and keratin-7 negative cells inside K7het subclones. Clonal cells in G0/G1, S, G2/M cell cycle phases analyzed by Hoechst stain.

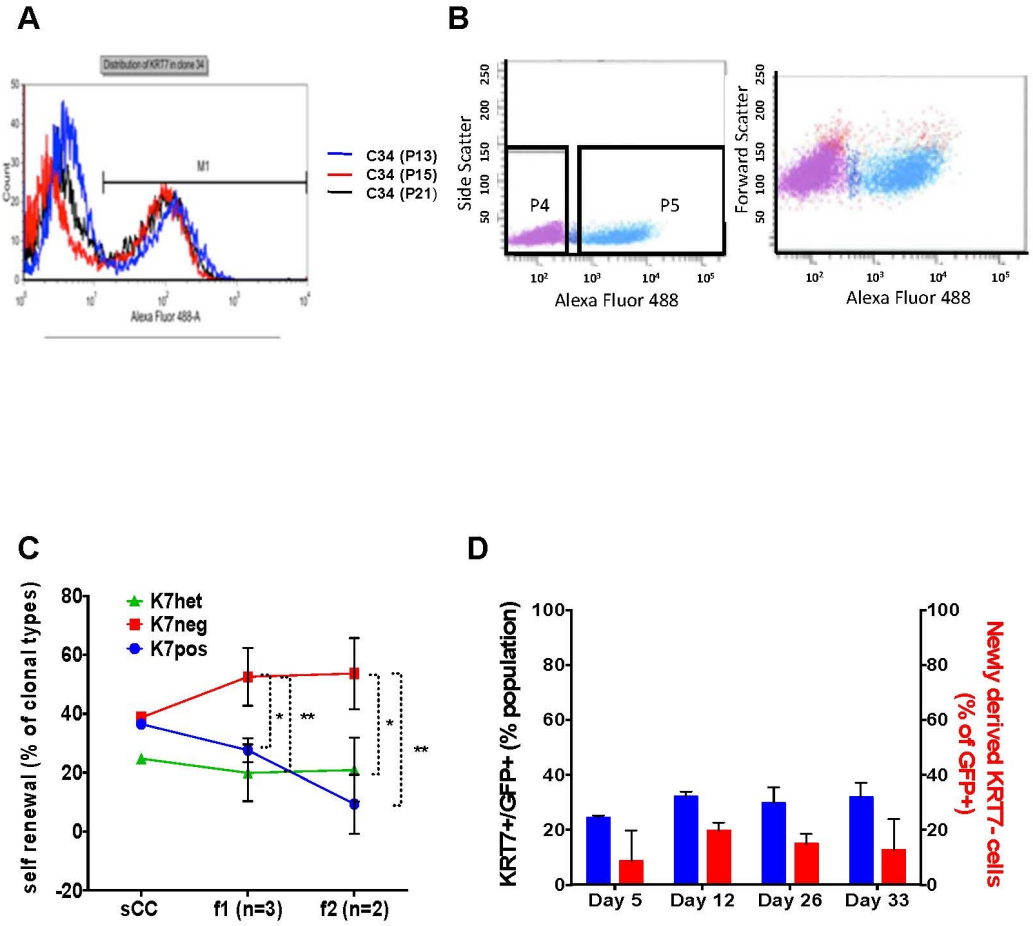


Figure 2-10. Robust heterogeneity in K7het subclones.

(A) Consistency between measurements of keratin-7 expression in K7het subclones (C1/34) in over four weeks period (Passage numbers are indicated in parenthesis); (B) Dot blot representation of cell size and granularity (FSC and SSC) of cell and Keratin-7 expression (Alexa Fluor-488); negative cells (F4) and positive cells (P5); (C) Percentage of cells of each subclonal type (K7pos, K7neg, K7het) from two clonal generations (f1 and f2) that formed colonies in cultures. Single cells were sorted into individual wells of 96-well plates and cultured for a period of two weeks; (D) Spontaneous loss of Keratin-7 expression in K7pos/GFP+ subclones pre-mixed with K7neg/GFP- subclones. Unlabeled K7neg cells from three independent single-cell clones (K7-/GFP-) were mixed with fluorescent (K7+/GFP+) cells creating a heterogeneous cell mixture. The percentages of K7+/GFP+ (blue bars) and newly derived K7-/GFP+ cells (red bars) were measured over a time course of 33 days Results are mean ± SD.

2.3.8 Spontaneous conversion of Keratin-7 expression in K7pos subclones

To further explore if individual cells from K7pos clones stochastically loose keratin-7 expression in a manner that does not occur when the cell types are propagated separately, we recreated heterogeneous populations by proportionately mixing cells from K7pos clones, labeled with green fluorescent protein (K7+/GFP+) and unlabeled cells from K7neg clones (K7-/GFP-) derived from three independently single-cells clones. During 33 days of monitoring, GFP+ cells maintained a steady proliferation rate, surprisingly, on day 5; a proportion of K7+/GFP+ started to convert into K7-/GFP+ cell fraction and was roughly maintained during a time course. Unlabeled K7-/GFP- cells never converted into K7+/GFP- cells, further supporting the spontaneous and one-way phenotype conversions of keratin-7 positive cells into keratin-7 negative cells (Figure 2-10D).

2.3.9 Targeted mutation hotspot analyzes between primary tumor (Sarc), K7pos and K7neg subclones

To demonstrate that the cell culture established is derived from the tumor, we performed targeted mutation hotspot analyzes of 46 cancer-related genes was performed. We detected seven variants above the call threshold including a BRAF V600E mutation, which has been previously described in cholangiocarcinoma with poor prognosis (212). Importantly, the same 7 variants were found in the original human tumor sample and the derived cell culture subclones (Figure 2-11A).

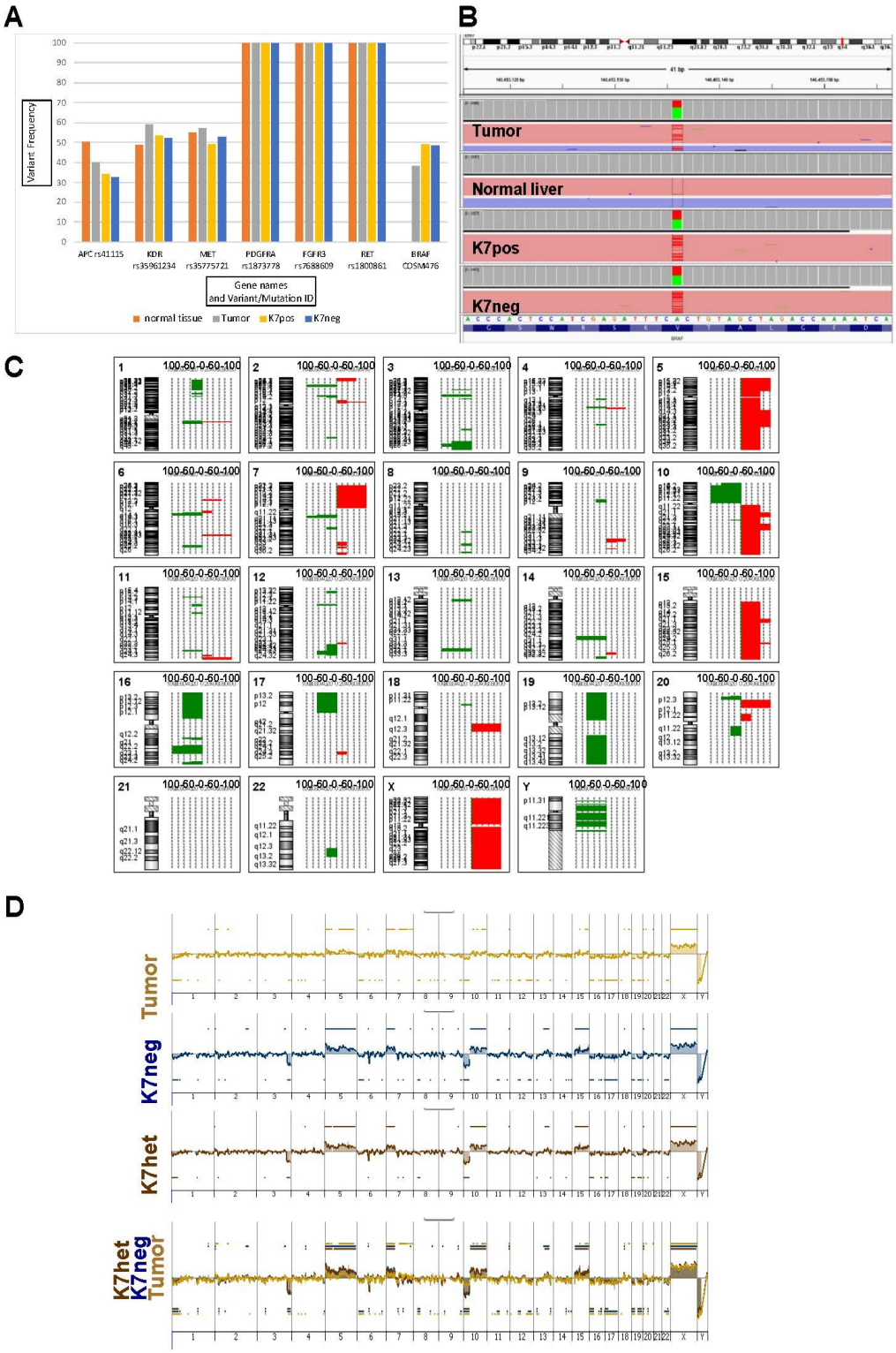


Figure 2-11. Genetic association between clonal keratin-7 phenotypes and primary tumor.

(A) Representative variant frequencies in K7pos and K7neg subclones and primary tumor (Sarc). Gene names and Variant/Mutation ID are displayed on the horizontal axis; (B) IGV snapshot of c.1799T>A substitution, (position 1799, T→A) in the BRAF gene (Chr 7: 7:140753331, hg19); (C) Summary of chromosomal gains and losses in clonal cell populations (K7neg and K7het (>43 % KRT7+ cells) and primary tumor measured by aCGH. Each chromosome marked with cytobands is represented with bars indicating the percentage of how often a gain or loss was found. Losses are displayed in green and gains are shown in red. Some changes, such as amplification of chromosomes 5, 7 and 10, or deletions of chromosome 10, 16, and 19 are common among all analyzed samples, while other changes are unique in different clones or patient tumor; (D) aCGH profiles of relative copy number aberrations (CNA) from patient tumor (Tumor; yellow-beige), keratin-7-heterogeneous (K7het; brown) and keratin-7-negative subclone (K7neg; blue). Blots are presented on a log2 scale (y-axis). (Adapted from (194)).

2.3.10 Comparative genomic hybridization in the primary tumor and cell clones with different keratin- 7 phenotypes

We extended the genetic data analysis and performed array-comparative genomic hybridization (aCGH) on DNA extracted from the sarcomatoid tumor component, as well as the primary cell culture (sCC) and the daughter K7neg clones (passage 29 and 30, respectively). While we found sporadic copy number aberrations specific for each clone and the primary tumor, several of the significant copy number aberrations detected in the original tumor (n = 26) were also present in both types of keratin-7 phenotype cell clones (K7neg: n = 26; K7het: n = 17) (Figure 2-11B, C). Genetic concordance between the samples analyzed showed that the clonal cell cultures were representative of the cellular diversity of the primary tumor and that the genetic relationship between K7het and K7neg cell subclones were essentially preserved (data not shown). Therefore, genetic variations do not explain the observed phenotypic heterogeneity in the different cell clones. Consequently, we explored the role of non-genetic mechanisms that underlie stochastic phenotypes switching associated with the process of sarcomatoid carcinoma dedifferentiation.

2.3.11 Clonal keratin-7 phenotypes constitute distinct transcriptional profiles

Transcriptional profiling (RNA-seq) of K7pos and K7neg f1 subclones identified 78 DEGs, of which 44 genes were significantly overexpressed and 34 underexpressed in K7neg clones compared to K7pos clones (fold change ≥ 1.0, p(adjusted) < 0.05). (Figure 2-12A). Furthermore, we generated transcriptome profiles from 10 independent daughter

subclones derived from a single f1 K7pos or K7neg subclones (f2 clones) and identified 276 DEG of which 143 genes were significantly downregulated and 133 upregulated in K7neg clones compared to K7pos clones (fold change ≥ 1.0, p(adjusted) < 0.05) (Figure 2-12B). The expression pattern of a set of randomly selected DEGs (n = 12, f1 clones and n = 7, f2 clones) were verified by qRT-PCR. The fold-changes in gene expression observed by qRT-PCR significantly correlated with the fold changes in gene expression observed by RNA-seq (f1 clones: r = 0.886; p < 0.001, and f2 clones: r = 0.962; p = 0.005) (Figure 2-12C). Almost 60 % (total of 47 genes) of DEG from f1 clones analysis was reproduced in f2 DEG, implying that reproducible transcriptional profiles were preserved at different time points during clonal propagation. Gene Ontology (GO) analysis identified commonly enriched terms in overexpressed genes in either K7pos or K7neg that included extracellular region, cell adhesion, and developmental processes. In addition, subtle differences of enriched categories based on overexpressed genes in K7neg clones related to germ-layer development, epithelial/mesenchymal differentiation, and regulation of cell growth and size. Reversely, overexpressed genes in K7pos clones were enriched in categories relating to metabolic processes and function (carbohydrate biosynthetic and aldo-keto reductase activity), and adult organ and system development (Figure 2-12D). Moreover, a literature-based query further categorized f1 DEG into functional categories. Differentially expressed genes enriched in either clonal phenotypes included cancer-specific and prognostic markers (*NRP1, COL7A1, IGDCC4, ACSL5, FGA, RBP4, GSTM3, SECTM1*), and genes associated with cancer proliferation, migration, and invasion (*THSD7A, FMNL3, NOV, CLDN3, AKR1C13, CA12, MAPK13, TNFSF10 SIX2, STK33*). Interestingly, overexpressed genes in K7neg clones compared to K7pos clones were implicated as inducers of EMT (*GALANT6, LEF1*) or inhibitors of differentiation (*MN1, MNX1*). Conversely, downregulated genes in K7neg clones included genes involved with development (DAAM2, ATOH8, PCSK9), differentiation (*ZNF268, AKR1B10, UBD, NKD2*), mesenchymal-epithelial transition (WISP2), and tumor suppressor genes (*PTPN6, RASL11A ADAP2, OLFML3*). Moreover, K7neg clones overexpressed a number of prometastatic genes (*FREM2, MGST1, NOG, TRO, ADAM22, INHBA*) and underexpressed several metastasis inhibitor genes (*NID1, CXCL14*). In addition, some f1 DEG genes were directly implicated in liver carcinogenesis (*AKR1B10, MAPK13, TNSF10, COL7A1*) and malignant transformation of hepatic progenitor cells (*DMBT1*) (a summary of DEGs is shown in Table 2-5).

To determine if the alterations in mRNA levels correlated with changes in protein levels at the single-cell level, we selected the *IL13RA2* gene, which negatively correlated with *KRT7* mRNA expression, and performed double-stained FACS analysis with anti-keratin-7 and anti-IL13RA2 antibodies. In stable K7pos subclones nearly 100 % of the cells were IL13RA2 negative, whereas K7neg subclones contained populations of both IL13RA2 positive and IL13RA2 negative cells (Figure 2-12E). Interestingly, K7het subclones contained a fraction of keratin-7/IL13RA2 positive cells, a characteristic that was not observed in either of the stable cell phenotypes (K7pos or K7neg) (Figure 2-12E). These results demonstrate that K7het subclones are not simply a mixture of K7pos and K7neg cells (Figure 2-12E, quadrant Q2), but they contain a phenotypically unstable subpopulation of cells which may generate daughter cells of different phenotypes.

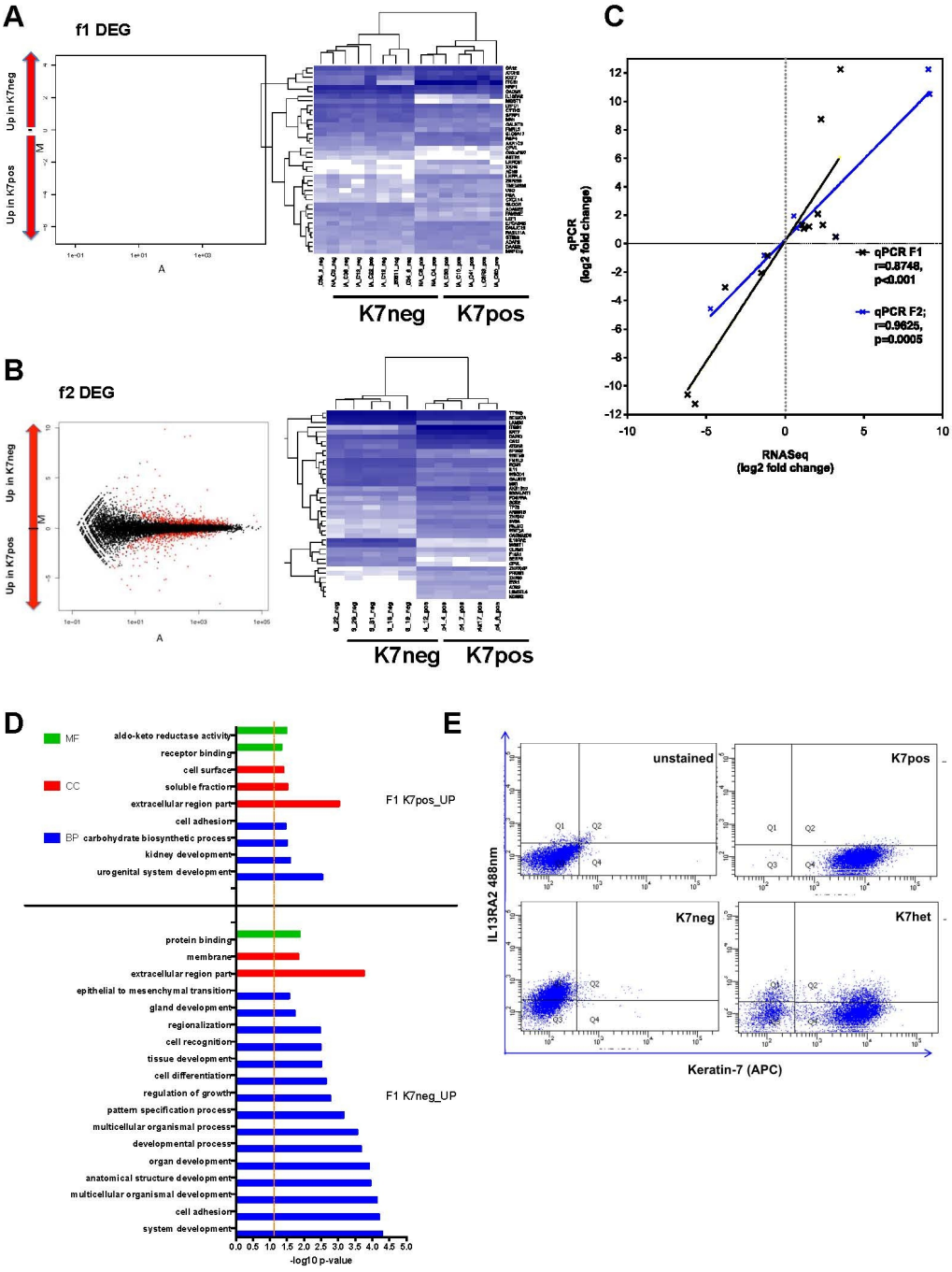


Figure 2-12. Differentially expressed genes in f1 and f2 subclones.

(A-B) MA plot of differentially expressed genes identified between K7pos and K7neg subclones in f1 and f2 subclones. Dots represent individual gene expression. The x-axis is the average expression over the mean of normalized counts (A-values), the y-axis is the log2 fold change between K7neg/K7pos (M-values). Genes with an adjusted p-value below a threshold are marked in red colour. Heatmaps showing the expression data of the top 40 differentially expressed genes. Each cell in the matrix represents the expression level of a gene feature in an individual sample. Dark-blue and white in cells reflect high and low expression levels, respectively; (C) qRT-PCR validation of 12 DEGs identified by

transcriptome sequencing. Differential expression is represented in log2 scale (Student t-test, *P < 0.05, **P < 0.01, ***P < 0.0001; n = 3-4 f1 or f2 clones); (D) Correlation of fold- changes of DEG between qRT-PCR and RNA sequencing in f1 and f2 K7pos and K7neg subclones; (E) Functional classification of differentially expressed genes in cell clones; Enriched gene ontology categories: genes up-regulated in K7pos cells (top) and up-regulated in K7neg cells (bottom); (F) FACS analysis of K7pos, K7neg, and K7het subclones co-immunostained with anti-keratin-7 and anti-IL13RA2 antibodies. Ten K7het and four different K7pos and K7neg cell subclones were individually analyzed and results are displayed as scatter plots in their respective panels. (Adapted from (194)).

2.3.12 Time and dose-dependent Keratin-7 mRNA and protein expression treated with 5-aza-dC

Previous studies have implicated variability in DNA methylation patterns in the context of stochastic gene expression in cancer cells (213). Treatment of K7neg subclones with 5-aza-dC caused keratin-7 protein re-expression, and increased the percentage of keratin-7 positive cells in a time- and dose-dependent manner (Figure 2-13A-C). Reactivation of keratin-7 expression occurred only in a sub-fraction of cells (Figure 2-13D), suggesting cell-to-cell variability in responses to 5-aza-dC treatment. In addition, we investigated the involvement of histone modifications on *KRT7* expression and protein levels, investigated by treatment of K7neg subclones with the histone deacetylase inhibitors (SAHA and TSA) in a time- and dose- dependent manner (Figure 2-13E, F). However, there was negligible reactivation of *KRT7* gene at the RNA level and no expression of the protein. We also investigated the synergistic effects of methylation and deacetylase inhibitors (SAHA and TSA) in K7neg clones and found a minor one-fold increase in expression with combined low-dose 5-aza-dC and TSA treatment (Figure 2-13G).

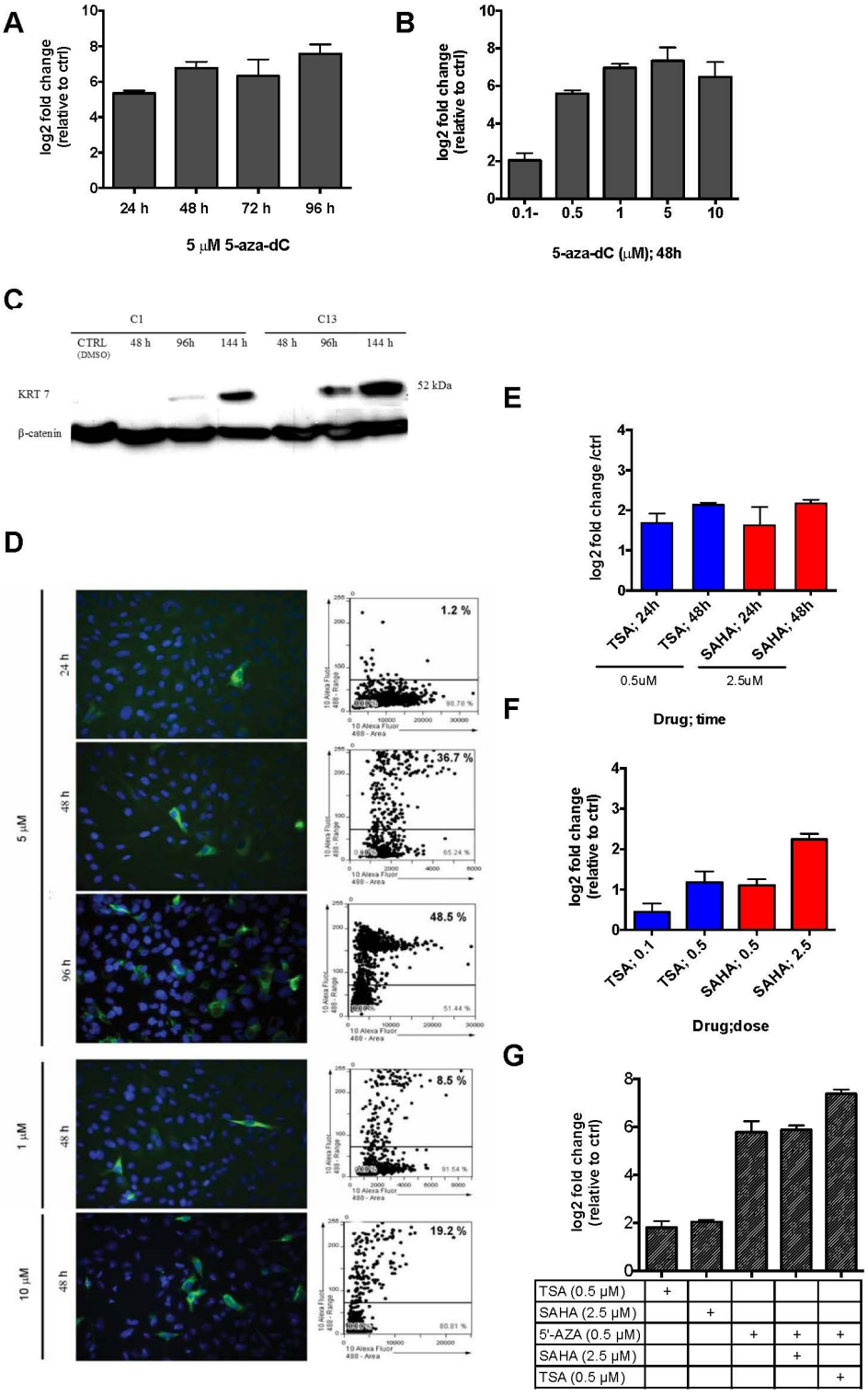


Figure 2-13. The keratin-7 expression is reactivated upon treatment with 5-aza-deoxycytidine (5-aza-dC).

(A-B) Time and dose-dependent re-expression of *KRT7* mRNA expression in K7neg subclones after 5-aza-dC treatment. Means ± standard error of triplicates shown; (C) Western blot of re-expressed keratin-7 in K7neg cells treated with 5-aza-dC at indicated time points. β-actin was used as a reference; (D) Immunofluorescence-stained cells representing time and dose-dependent re-expression of keratin-7 protein in K7neg clones after 5-aza-dC treatment. Right: Scatter plots representing single cells with re-expressed *KRT7* protein relative to nuclear DAPI positive cells. Cells were analyzed with TissueFAXS in each of the conditions. The percentages of positive cells were derived from the upper right corner in each scatterplot determined by the threshold lines and represent averages of at least 30,000 counted cells; (E-F) Time- and dose-dependent re-expression of Keratin-7 in K7neg clones after TSA and SAHA treatment; (G) Combined treatment with 5-aza-dC and histone deacetylase inhibitors (SAHA and TSA) in K7neg clones. Bar graphs represent averages of *KRT7* expression on log2 scale from two independent K7neg clones, (Student's t-test, *$p<0.05$). (Adapted from (194)).

2.3.13 MeDIP-seq analysis of KRT7 promoter and DMR and their correlation with DEG

For further investigation of the transcriptional regulation of keratin-7, whether KRT7 expression in different clonal phenotypes correlated with DNA methylation of the KRT7 promoter was analyzed. Using MeDIP-seq of three independently derived K7pos and K7neg f1 subclones showed a significant increase in DNA methylation at the CpG islands associated within the KRT7 promoter in all K7neg clones compared to K7pos cells, corroborating previous reports implicating DNA methylation in KRT7 gene regulation (185, 207) (Figure 2-14A, B). MeDIP-seq data was verified by performing MeDIP-qPCR using amplicons designed to an unmethylated locus (GAPDH), a highly methylated locus (SNRPN), and a differentially methylated locus (ADAM22) between stable K7pos and K7neg cell clones (Figure 2-14C; Figure 2-15A-E). Furthermore, we verified additional ten amplicons designed for highly methylated or differentially methylated loci between stable K7pos and K7neg cell clones (Figure 2-14D) were verified. To test the role of DNA methylation in global transcriptional stability, we used MeDIP-seq of different keratin-7 clonal types, and identified 3,344 differentially methylated regions (DMRs, false discovery rate <0.05). Differentially expressed genes demonstrated a much higher association with DMRs than non-differentially expressed genes, indicating that in cells with a stable phenotype the observed changes in gene expression were associated with changes in DNA methylation (Figure 2-14E).

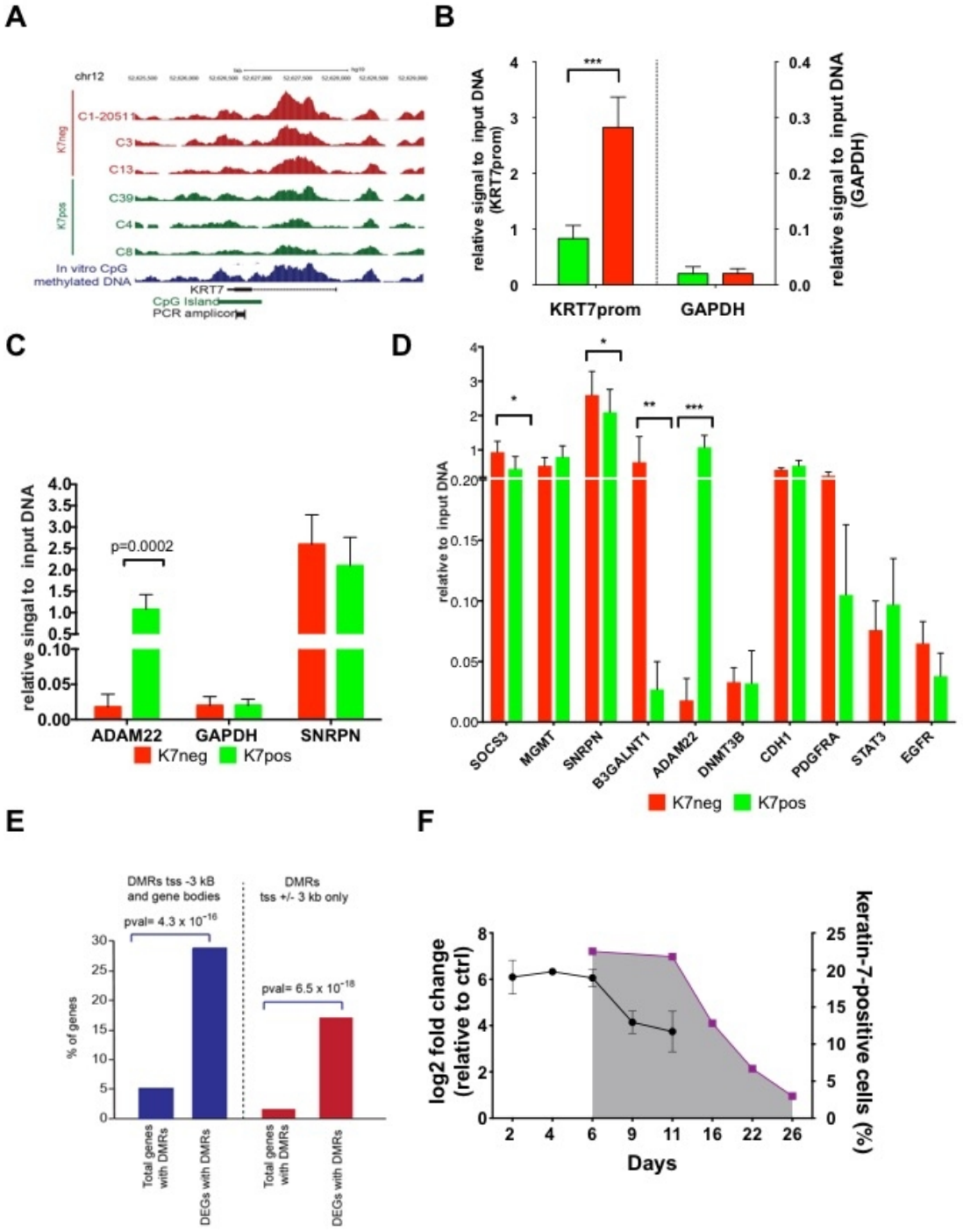


Figure 2-14. Differences in DNA methylation in stable cell subclones and subclones with stochastic phenotype switching.

(A) UCSC genome browser screenshot of representative *KRT7* promoter peaks from six independent subclones, where K7neg subclones (C12205, C3, and C13) (red) had increased 5-mC levels compared to K7pos subclones (C39, C4 and C8) (green). As a reference, MeDIP-seq of *in vitro* fully CpG methylated genomic DNA is shown in blue;

(B) MeDIP-qPCR verification of DMR at KRT7 promoter (KRT7prom) and GAPDH promoter (GAPDH) as a negative control; (C-D) Verification of MeDIP-seq data for specific DMR's (Student's t-test, *$p<0.05$, **$p<0.01$; (E) DMRs preferentially associated with promoters (right) and promoters and gene-bodies (left) of DEGs. tss: distances from transcription start site, (p-values, hypergeometric test); (F) Heritability of gene and protein expression was analyzed upon treatment K7neg subclones with 1 µmol/L 5-aza-dC, cells were treated with 1 µmol/L 5-Aza for 48 h. Cells were harvested for RNA extraction every second day for 10 beginning 48 h after drug removal. Protein expression was counted in single cells by TissueFAXS on cells grown and stained on glass slides from day 6 up to day 26. (Adapted from (194)).

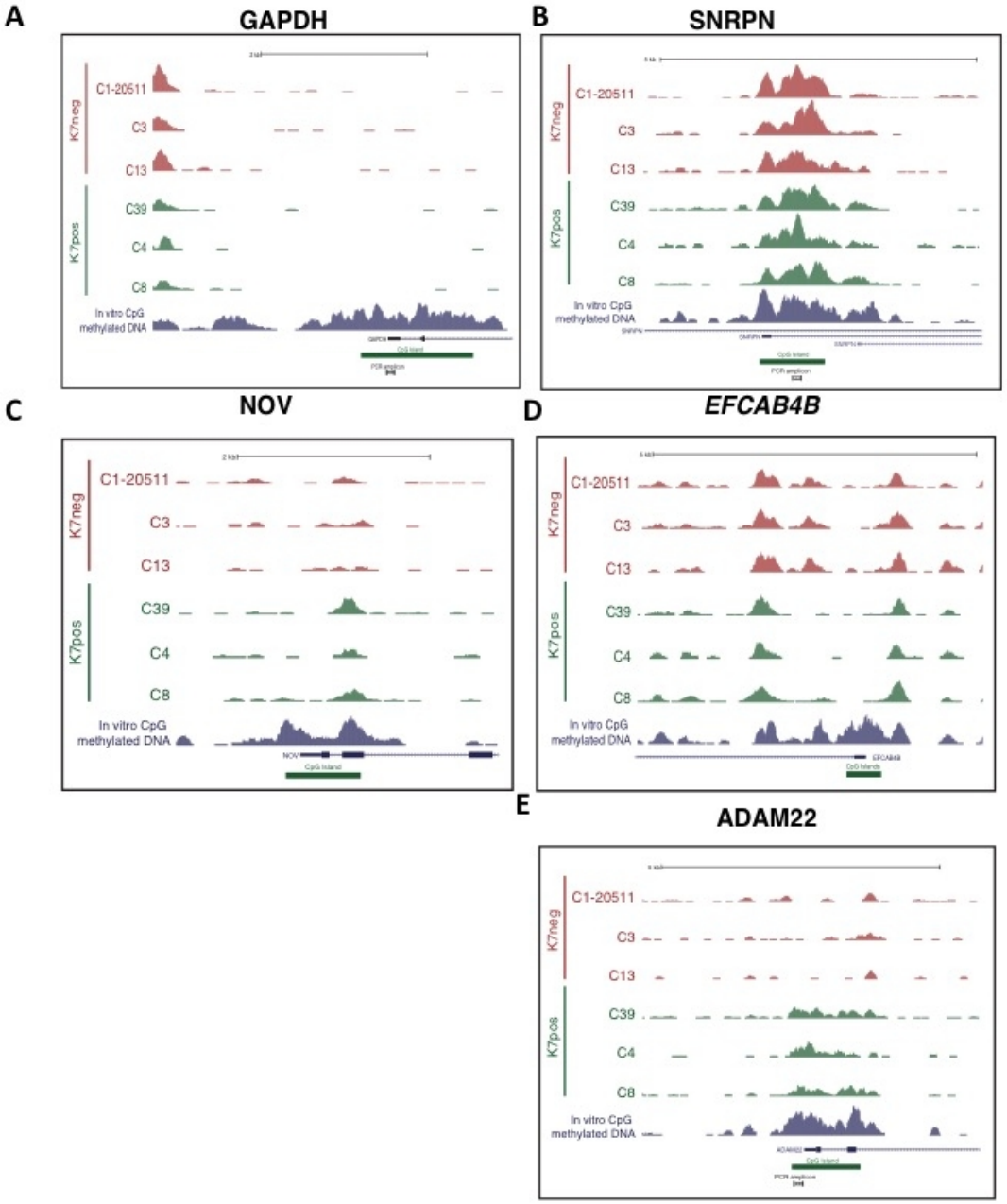


Figure 2-15. Genome browser view of DMR identified in keratin-7 positive and keratin-7 negative subclones.

(A) unmethylated locus (GAPDH), (B) highly methylated locus (SNRPN) and (C-E) three DMR's (NOV, EFCAB4B, ADAM22) are shown. Values are relative to a standard curve of input DNA. (Adapted from (194)).

2.3.14 Dynamics of Keratin-7 expression

To elucidate the dynamics of keratin-7 re-expression at mRNA and protein level after 5-aza-dC treatment (192), we initially treated K7neg cells with 5-aza-dC, and subsequently

propagated the cells in a drug-free environment while monitoring mRNA and protein expression. In a subsequent experiment, re-expression of keratin-7 protein levels were measured by counting cells in cell populations starting 6 days after treatment with 5-aza-dC (Figure 2-14F) . Treatment resulted in increased relative expression of KRT7 mRNA and protein. Both mRNA and protein levels gradually decreased during propagation in a drug-free medium, suggesting involvement of intrinsic, homeostatic mechanisms, which are able to reconstitute methylation markers after initial erasure of DNA-methylation on *KRT7* promoter (Figure 2-14F).

2.3.15 KRT7 promoter-specific pyrosequencing in FACS sorted keratin-7 positive and keratin-7 negative cells from K7het subclones

We further investigated if stable (K7pos and K7neg) and unstable (K7het) keratin-7 subclones differ in the way they utilize DNA methylation. Surprisingly, comparison of DNA methylation analysis results from phenotypically stable K7neg cells showed that methylation of the *KRT7* promoter was much more pronounced than in sorted keratin-7 negative cells of phenotypically unstable K7het subclones (Figure 2-16A, B). This implies that mechanisms not related to DNA methylation are involved in regulating keratin-7 expression in cells undergoing stochastic phenotype switching.

2.3.16 Clonal progeny derived from K7pos and K7neg clones and 5-aza-dC treatment stabilizes transcriptome variance

To further characterize the transcriptional stability of the K7pos and K7neg subclones, we compared the per-gene biological variance between replicates of K7pos and K7neg subclones derived from single cells from the primary sCC cell culture (f1 clones) and daughter subclones derived from a single f1 K7pos or K7neg subclones (f2 clones) (Figure 2-16C). Typical RNA-seq data analysis consists of applying statistical models to transcript read-count data sets. One advantage of the count models is their ability, for a given gene, to separate significant read count differences from biological and technical variability (203). Our experimental strategy has an advantage of accurately estimating biological variability for each gene between independently derived clones, by estimating the degree of natural stochastic gene expression variation (dispersion) on gene basis with a DEseq

analysis tool. We discovered that the per-gene variances between different f1 subclones were much higher compared to the average per-gene variances between their daughter f2 subclones (Figure 2-16C). Thus, by probing the gene expression variance between different clonal generations, we observed that not only *KRT7* gene expression was stabilized in the K7pos and K7neg subclones, but there also exists a general mechanism leading to global transcriptome stabilization during the transition from an unstable keratin-7 expressing phenotype (i.e. K7het) to a more stable keratin-7 phenotype (K7pos and K7neg) (Figure 2-16C). Furthermore, it appears that sequential cell divisions reduce the overall noise and globally stabilize the transcriptional program of inherently phenotypically unstable cells. Because variability of DNA methylation patterns has been associated with stochastic gene expression in cancer (213), we investigated whether gene expression in the different clonal phenotypes correlates with DNA methylated regions in the genome (214, 215). RNA-seq analysis of K7pos, K7neg, and K7het subclones (f1) treated with the DNA methyltransferase inhibitor 5-aza-2-deoxycytidine (5-aza-dC) revealed a remarkable decrease in variance; that is, the similarity between the 5-aza-dC-treated K7pos and the K7neg subclones was greater than that observed between untreated corresponding clones (Figure 2-16D). These observations of transcriptome "stabilization" due to rearrangements in the methylation patterns suggest an important implication of DNA methylation in the variability of gene expression, which may have a defining role in initiation and establishment of heterogeneous differentiation phenotypes.

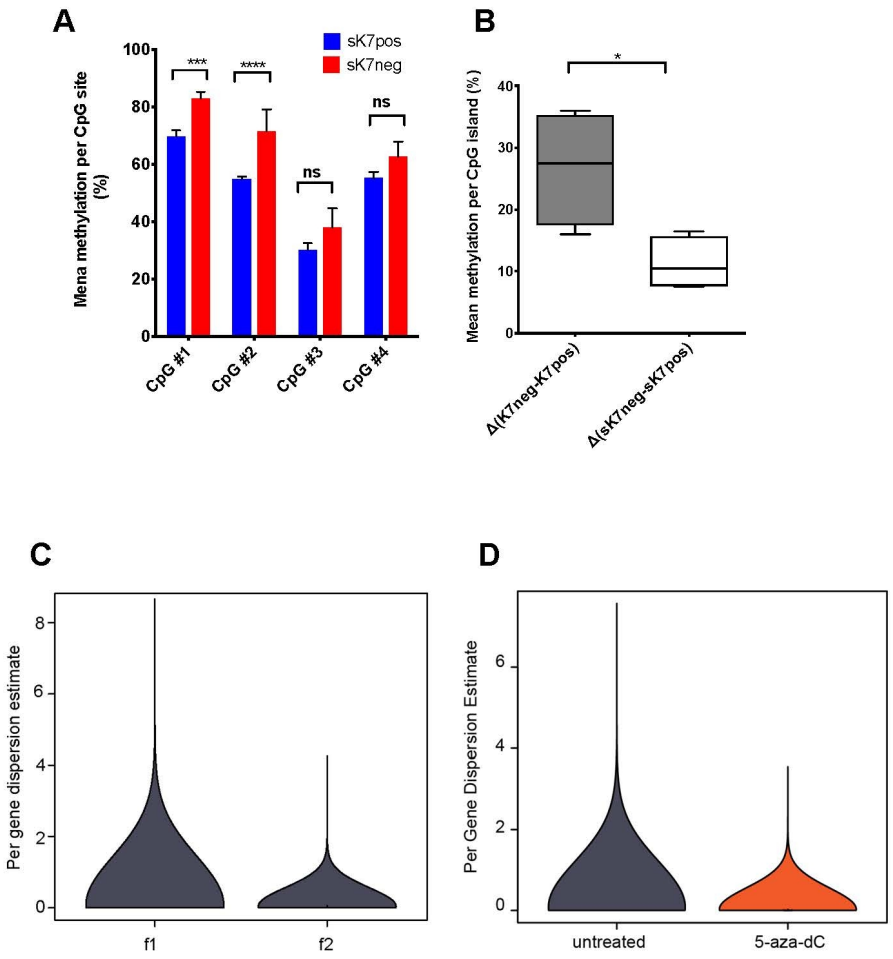


Figure 2-16. Quantification of CpG methylation at KRT7 promoter.

(A) Quantitative methylation differences at individual CpG sites in KRT7 promoter CpG island in sorted keratin-7-positive and keratin-7-negative cell fractions (sK7pos, sK7neg) of a K7het culture (n = 8). Two-way ANOVA with Bonferroni's multiple comparisons test, ***p<0.001, ****p<0.0001); (B) Relative methylation differences between stable K7pos and K7neg clones (n = 2) are significantly higher compared to methylation levels between keratin-7 expressing and non-expressing sorted cell subfractions of K7het clones (n = 4; Student's t-test, *p<0.05); (C) Violin plot representation of the per-gene variance distribution based on transcriptome sequencing of K7pos and K7neg f1 cell subclones compared to K7pos and K7neg f2 cell subclones. The f2 K7pos and K7neg subclones were derived from a single f1 cell each (n = 5 for each clonal type, Student's t-test, $p = 1.6 \times 10^{-15}$); (D) Violin plot representation of the per-gene variance distribution based on transcriptome sequencing of K7pos and K7neg f1 cell subclones compared to K7pos and K7neg f2 cell subclones. The f2 K7pos and K7neg subclones were derived from a single f1 cell each (n = 5 for each clonal type, Student's t-test, $p = 1.6 \times 10^{-15}$. (Adapted from (194)).

2.3.17 Phenotype switching occurs during mitosis

Asymmetric cell division is a characteristic property of gene expression noise (216). Therefore, we investigated if this stochastic modality is responsible for heterogeneous keratin-7 expression in daughter cells after mitosis. In several independent K7het clones, we observed different keratin-7 staining of daughter cells during cytokinesis or immediately after the cell division phase (Figure 2-17A). Such changes in keratin-7 phenotypes were never observed in dividing K7pos and K7neg cells (not shown). To further characterize the asymmetrical partitioning of keratin-7 expression that occurs between daughter cells, we pre-incubated cells with BrdU. Double staining for BrdU and keratin-7 confirmed unequal partitioning between daughter cells (Figure 2-17B).

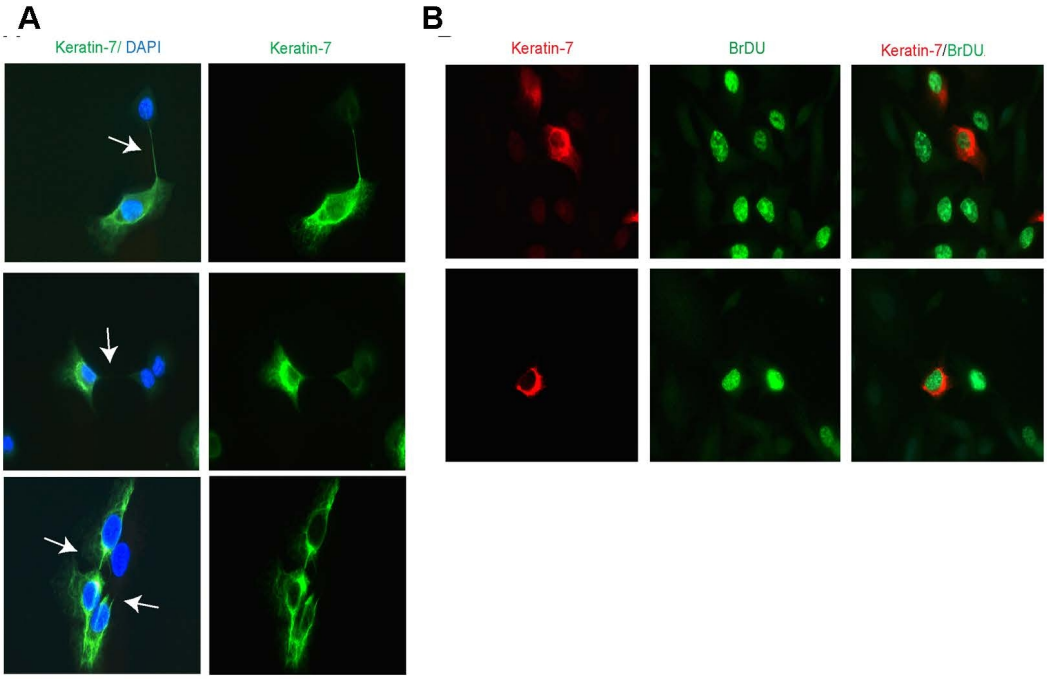


Figure 2-17. Stochastic phenotype switching and asymmetric cell division in K7het clones

(A) Immunofluorescence microscopy of keratin-7 expression in daughter cells immediately after cell division. White arrows point to connections between daughter cells. Lower panel: uneven expression of keratin-7 in daughter cells (right arrow) and daughter cells with same keratin-7 phenotypes (left arrow). Magnification: 630x; (B) Actively dividing K7het sub-clonal cell incubated with BrdU and double-stained with anti-Keratin-7 and anti-BrdU. (Adapted from (194)).

2.3.18 Loss of keratin-7 expression results in increased tumorigenicity *in vivo*

To investigate the biological relevance of the different keratin-7 phenotypes in cancer development, we evaluated their tumorigenic potential in a NOD/SCID mouse tumor xenograft model. The average time for the appearance of tumors after subcutaneous implantation of cells was 29±6 days. Only K7het and K7neg subclones established tumors whereas none of the K7pos subclones led to visible tumor growth at any of the injected sites (n = 8 per clonal type) (Figure 2-18A). We also observed variability between clonal types and weights of the tumors after resection (Figure 2-18B). Histological analysis of the developed tumors revealed mostly spindle cell morphology and an invasive growth pattern, which resembles the invasive sarcomatoid component of the primary patient tumor and the metastasis (Figure 2-18C). Importantly, IHC analysis showed that all K7het-derived tumors were comprised of keratin-7-negative cells (Figure 2-18C), except a few scattered keratin-7-positive cells, most of which showed features of apoptosis and were located in necrotic parts of the tumor (Figure 2-18C). This was a surprising result since the injected K7het subclones contained >60 % keratin-7 positive cells (clone CCC C: 78 % keratin-7-positive cells; CCC C1: 63 % keratin-7-positive cells). Both K7het- and K7neg- derived tumors concomitantly expressed keratin-8, vimentin, and Ki-67 markers. Additionally, in the xenograft environment, as in the primary patient tumor, we observed that some tumor cells lost keratin-8 expression while maintaining expression of vimentin (Figure 2-18D).

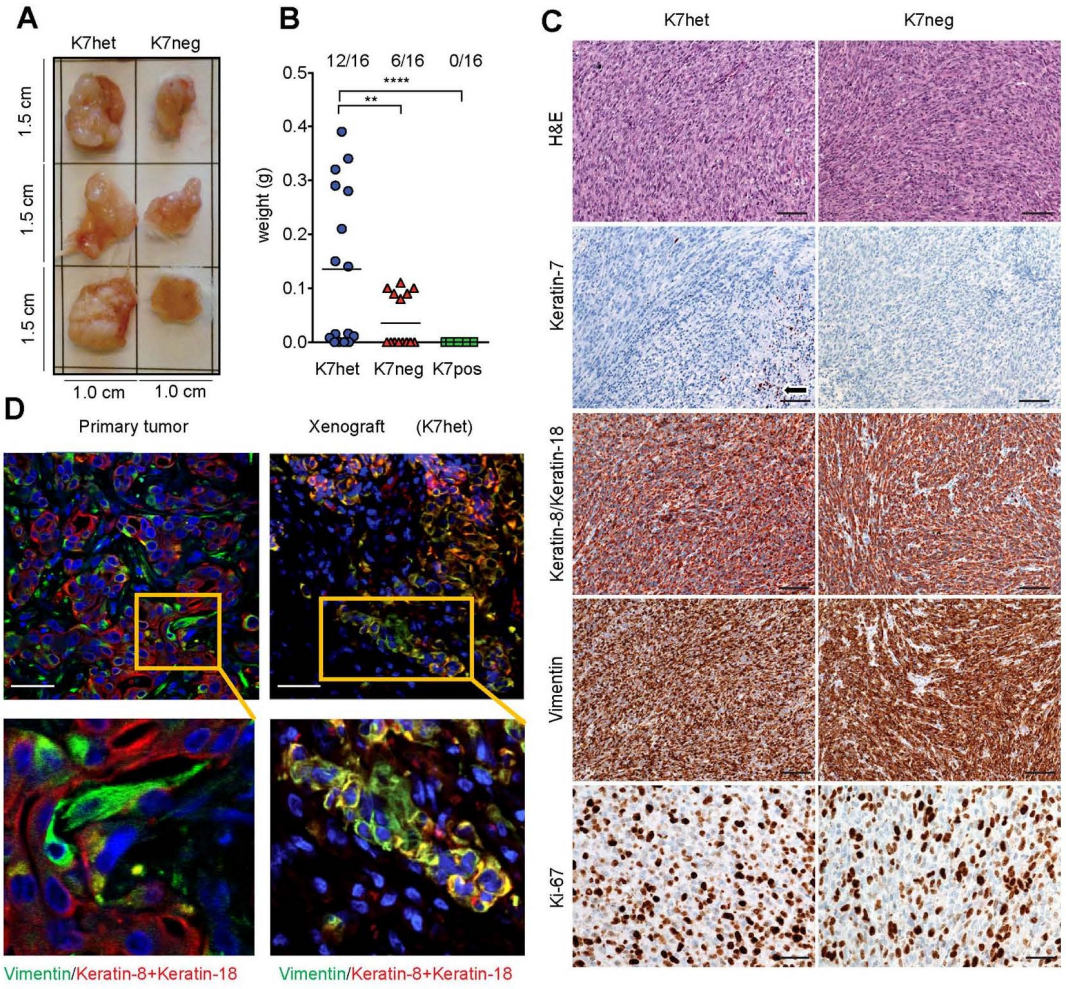


Figure 2-18. Keratin-7-negative subclones show increased tumorigenic potential in xenografts.

(A) Representative pictures of established tumor xenografts from K7neg and K7het subclones after 33 and 22 days, respectively; (B) Mean tumor weights developed from the different clonal phenotype. Two independent subclones per clonal phenotype have been injected into multiple mice (tumor weights are shown as mean ± SD; total number of developed tumors/total number of injection sites is shown at the top; One way ANOVA with uncorrected Fisher's LSD test (**p<0.01, ****p<0.0001); (C) Hematoxylin and eosin (H&E), and immunohistochemical staining of xenografted tumors. Tumors were derived from K7het and K7neg subclones and stained with anti-keratin-7 antibodies (the arrow indicates residual keratin-7-positive cells associated with necrotic and apoptotic tumor cells), anti-keratin-8+18 antibodies, anti-vimentin antibodies, and the Ki-67 antibody. Xenografts developed from K7het subclones showed a keratin-7 negative phenotype. Scale bar for H&E, keratin-7, keratin-8/keratin-18, vimentin images = 50µm. Scale bar for Ki-67 images = 100µm; (D) Triple immunofluorescence staining of the primary tumor and corresponding xenografts with anti-keratin-8 (red), anti-vimentin (green) antibodies and nuclear DAPI dye (blue). DAPI, 4,6- diamidino-2-phenylindole. Scale bar = 20µm; the insets in the upper panels indicate areas shown in higher magnification in the lower panels. (Adapted from (194)).

2.3.19 Expression of EMT markers in cultured subclones, primary tumor components, and xenografts

Further characterization of stochastic phenotype switching in culture, primary tumor components, and nude mice xenografts involved differentially analysis of EMT markers at gene expression and protein levels. Distinct gene expression signatures from different regions (Carc, Sarc, Met) provided unbiased proof for the involvement of EMT in tumor progression cascade from stationary to metastatic tumor. Significantly overexpressed genes in Sarc/Met compared to Carc suggest that tumorigenesis processes are functionally related to extracellular matrix remodeling and metallopeptidase activity, loss of cell-cell junction, and epithelial differentiated with concomitant involvement of mesenchymal differentiation, invasive and migratory genes, stem-cell/self-renewal genes, and finally EMT-inducing transcription factors Zeb1, Zeb2 (log_2 fold change < 2.0, p(adjusted) < 0.01), and Twist1 (p = 0.59), while Snai1, Snai2, and TCF3 were not significantly differentially expressed. We also confirmed EMT-related expression changes in the primary tumor by computing enrichment of genes upregulated in sarcomatoid and metastatic compartments within core EMT gene expression signature (217). Upregulated in Sarc were 9 out of 91, genes, (Fisher exact test, p < 0.0001), while upregulated in metastasis were 20 out of 91, (Fisher exact test, p<0.001). Among the DEGs in the primary tumor, we verified MCAM/CD146 expression in the primary tumor (Figure 2-19A, B), where it displayed progressive loss between different tumor components and was upregulated in K7het subclones (Figure 2-19D, E). MCAM/CD146 was previously implicated as an inducer of EMT in an aggressive type of breast cancer in clonal cultures and primary tumor, as well as in a promoter of EMT in HCC, which suggests its important role in EMT-like mechanisms associated with sarcomatoid dedifferentiation (218-220). In clonal cultures (f1 and/or f2 clonal generations), there was no deregulated classical EMT transcriptional factors between K7pos and K7neg subclones. Zeb1 showed similar protein expression in primary tumor compartments and xenografts (Figure 2-19F-I). However, in f1 and/or f2 clonal generations we identified upregulation of additional EMT-related set of genes (*MCAM, LEF1, GALNT6*), and downregulation of EMT inhibitors *AKR1B10* (221) and *WISP2* genes (222). *FN1* gene expression difference was not statistically significant in f1 and f2 clonal generations, but in f2 clonal generations, *CDH1* was significantly downregulated in K7neg subclones. On the individual cell level, no direct correlation was observed in the expression of EMT associated genes and keratin 7 expression (data not shown). In contrast,

at the cell population level, we immunohistochemically identified downregulation of EMT-related genes, E-cadherin and fibronectin in K7neg subclones (Figure 2-7A), including *AKR1B10*, which was significantly higher in K7pos clones compared to K7neg and K7het subclones (data not shown). Finally, we re-analyzed RNA sequencing data focusing on EMT associated genes selected from (223), and have summarized their expression in different tumor compartments and keratin-7 subclones (Table 2-7).

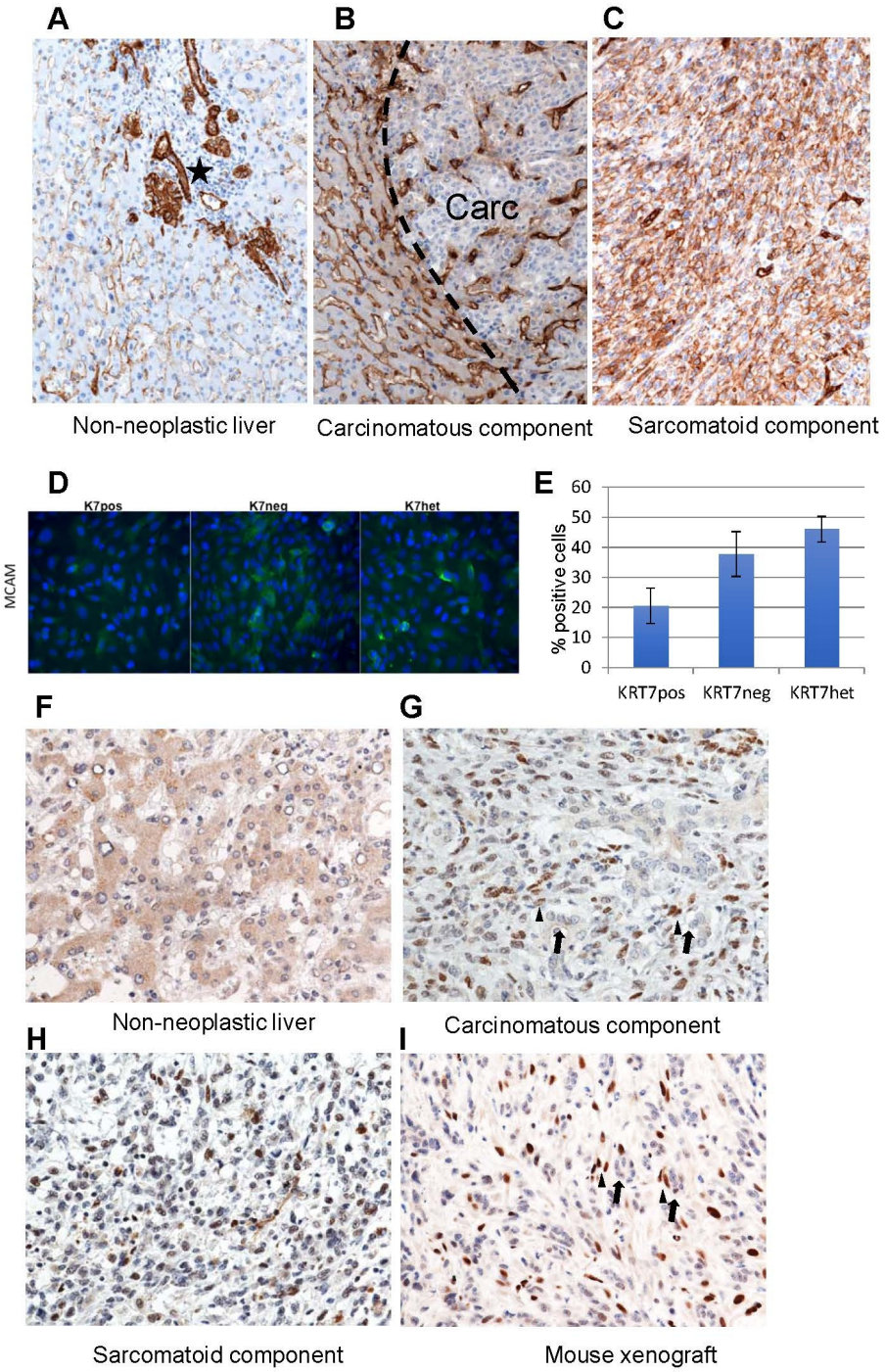


Figure 2-19. Expression of EMT markers in primary tumor components and cultured subclones.

(A-C) Immunohistochemical detection of CD146/MCAM in different primary tumor components A) Non-neoplastic liver, an asterisk indicates portal tract; (B) Carcinomatous tumor component [Carc]; dashed line indicates border between non-neoplastic liver and Carc; (C) Sarcomatoid component. CD146 shows no membranous staining in the Carc component whereas the sarcomatoid component shows strong membranous reaction (Magnification: 200x); (D) IF characterization of CD146/MCAM expression (green) in K7 subclones showing different phenotypes (KRT7pos, KRT7neg, KRT7het), and DAPI (blue) as nuclear stain; (E) FACS analysis of CD146/MCAM expression in clonal populations of the different keratin-7-expressing phenotypes (KRT7pos, KRT7neg, KRT7het), (n = 6); (F-I) Immunohistochemical detection of Zeb1 expression in different primary tumor components and mouse xenograft. (F) Non-neoplastic liver; (G) Carcinomatous tumor component; (H) Sarcomatois tumor component; (I) Nude mouse tumor xenograft. Arrows in (G) and (I) indicate tumor cells with epithelial differentiation which are negative for Zeb1, whereas cells in proximity with mesenchymal differentiation (arrowheads in G and I) are positive for Zeb1. Similar situations are seen in the carcinomatous component of the primary tumor and the mouse xenograft. Magnification 400x. (Adapted from (194)).

2.4 Discussion

Please note that parts of this section have been published in (194).

The observations presented in this work represent a significant advancement in our knowledge of intratumor heterogeneity in sarcromatoid cholangiocarcinoma, and thus broadens the knowledge base for understanding this particularly aggressive cancer. Our results and hypothesis accompany a large body of literature that explains intratumor heterogeneity in a repertoire of established models of intratumor heterogeneity (clonal genetic diversity, cancer stem cell model, and contribution of microenvironment signals).

By monitoring the fate of single cells derived from a primary culture of a human liver sarcomatoid cholangiocarcinoma, we determined the phenotypic and molecular history of 1043 single cell-derived subclones. We identified distinct self-propagating subclones characterized either as essentially stable (K7pos or K7neg), or as unstable subclones (K7het). K7het subclones can stochastically propagate to the stable phenotypes (K7pos or K7neg) or transfer their phenotypically unstable cell nature (K7het) to daughter generations. The morphology of each clonal phenotype appears stable in subsequent rounds of single-cell sorting, including the persistent unstable nature (i.e. stochastic phenotype switching) of K7het subclones. In K7het subclones, intraclonal heterogeneity stochastically occurs in individual cells following mitosis, and it is associated with adjustments of the cell's transcriptional program. Our findings implicate intraclonal epigenetic instability, manifested as stochastic phenotype switches, as a source of functional diversification of cancer cell populations with unique transcriptional programs between the investigated subclones. Importantly, in the investigated subclones of the primary tumor cell culture, epigenetic mechanisms such as DNA methylation stabilized stochastically generated phenotypes, rather than generating tumor heterogeneity. Stochastic phenotype switching is associated with aberrant trans-differentiation and acquisition of invasive and metastatic tumor phenotype through EMT-like mechanisms.

EMT is a classic example of cellular plasticity, and it may serve as an important prerequisite for invasion and metastatic spread of primary tumors (115, 136, 137). Importance of EMT in the plasticity of cancers comes from its role in the conversion of non-tumorigenic cells into tumorigenic cancer stem cells (116, 117). This process is either dependent on genetic or environmental influences in normal or transformed differentiated cell lines (51, 138, 224). Furthermore, cellular plasticity is associated with drug

insensitivity and is mediated by transient drug-tolerant states, which can be mediated through chromatin modifications (55).

Morphological and immunohistochemical characterization inside the primary tumor showed a progressive loss of epithelial phenotypes and acquirement of mesenchymal features characteristic of tumor progression as tumor cell morphology changed in distinct tumor regions (compartments) already inside the liver and in the distant lymph node metastases. We also identified a distinctive transitional tumor compartment, which suggests that progressive phenotype conversions occurred early in tumorigenesis and simultaneously at many sites inside well-differentiated tumor areas.

Transcriptome profiling in regionally separate tumor compartments provided unbiased evidence of gene expression signatures that implicated EMT during the tumor invasion and metastasis (58). Selected gene and protein expression analyzes were congruent in confirming the involvement of EMT in the carcinogenesis of sarcomatoid CC and highlight EMT's crucial role in cellular plasticity and tumor metastasis (225). Importantly, our comparative transcriptome signatures identified new genes, including keratin-7, which may be distinctive markers of EMT in sarcomatoid-differentiated tumors. Among the IHC verified DEGs in the primary tumor, we identified MCAM/CD146 as significantly upregulated in sarcomatoid compartment. MCAM/CD146 is potentially a new marker in invasive cholangiocarcinoma, previously implicated as an inducer of EMT (219, 220, 226). IHC of CD146 displayed progressive loss between different tumor compartments. We also showed that Zeb1 was differentially expressed in primary tumor compartments and showed that its protein expression was heterogeneously expressed (nuclear staining) in cells negative for keratin-8 expression, but maintained vimentin expression. However, on the single-cell basis no direct correlation between the expression of EMT markers and keratin-7 expression was found (not shown). This indicates that EMT and stochastic phenotype switching are related, but regulated by different mechanisms. EMT is an inherently dynamic process, therefore it can be that cells maintained in culture conditions may have undergone an EMT-independent conversion resulting in a more mesenchymal phenotype state (227). Moreover, tumor cell behavior can partially display transient EMT-like phenotype with variable levels of E-cadherin expression, and even disseminate without the involvement of EMT-associated gene expression patterns (228-230). Simultaneous expression of epithelial (e.g., keratin-8, keratin-18, keratin-19) and mesenchymal markers (e.g., vimentin, fibronectin) implies an incomplete EMT-like state, and that acquisition of mesenchymal features in K7neg clones occurred stochastically in

the absence of any external inducing factor. We cannot definitively exclude that EMT is the only mechanism involved in imparting mesenchymal-like phenotypes to K7neg clones. Absence of signaling pathways and gene inducers of EMT presumably reflects a common and non-stimulatory environment of *in vitro* cultures. Furthermore, phenotypes of established clones may not completely reflect cell types in the primary tumor. Still, upregulated genes in K7neg clones were associated with undifferentiated, mesenchymal state competent for invasion and metastasis, compared to K7pos clones. Finally, with the established role of EMT in generating CSC-like activity required for most of the steps of metastatic dissemination (116, 117), it is likely that during the tumorigenesis, keratin-7 negative cells adopt similar processes enabling them to complete the steps of invasive-metastatic cascade.

The biological significance of the different clonal types was shown in their capacity to initiate tumors in mice (231). Based on the expression of keratin-7 as an epithelial differentiation marker protein, we observed that only cells with a keratin-7 negative phenotype were capable of producing tumors in xenografts. Furthermore, all established tumor xenografts had an undifferentiated sarcomatoid morphology that morphologically resembled the invasive and metastatic components of the patient tumor. In this context, it was of particular interest that xenografts of K7het subclones, which contained more than 60 % keratin-7 positive cells, led to tumors which were essentially negative for keratin-7. This indicates that loss of keratin-7 expression correlates with *in vivo* tumorigenicity. Moreover, there was a greater take rate of xenografts after injection of K7het subclones (12 tumors developed at 16 injection sites) as compared to K7neg subclones (6 tumors developed at 16 injection sites) (Figure 2-18A, B), indicating that the ability of stochastic phenotype switching (as it is present in the K7het subclones) could be a feature required for better adaptation to the environment of a xenograft, thus resulting in greater tumor generation efficiency compared to cells with a stable keratin-7 negative phenotype. The essential absence of keratin-7 positive cells in xenografts developed from K7het subclones could result from a survival disadvantage of keratin-7 positive cells as compared to keratin-7 negative cells, or indicate that keratin-7 positive cells switched to a keratin-7 negative phenotype (Figure 2-18C). The observation that K7het-injected mice developed more and larger tumors than K7neg-injected mice favors the latter hypothesis.

Our results emphasize the involvement of deregulated and stochastic epigenetic processes in the phenotypic and functional heterogeneity of cancer cells (232). Mechanistically, we showed that hypermethylation of the *KRT7* promoter is involved in silencing of the *KRT7*

expression in stable K7neg subclones. Interestingly, we found significantly lower methylation of the *KRT7* promoter in the keratin-7 negative sorted cell fraction derived from the (phenotypically unstable) K7het subclones than in the stable K7neg subclones, suggesting the existence of mechanisms other than DNA methylation in the repression of keratin-7 expression in K7het cells showing stochastic phenotype switching. Dynamic stochastic variation in the epigenetic landscape, supported by genetic mutations or inflammatory tumor microenvironment may explain the stochastic loss of keratin-7 expression, in an EMT-like state (233).

Other epigenetic mechanisms like histone modifications, which tend to be relatively plastic, may be implicated in intrinsic stochastic nature of K7het clones (234, 235). Relatively lower methylation levels in sorted cell fractions derived from heterogeneous clones could initially explain the lower dependency on the dynamic changes of DNA-methylation patterns that would late become dominant in the subsequent stable clonal populations.

These results further affirm that cancer cellular plasticity is in its essence an epigenetic process (236), as did the findings from an inducible model of EMT, which showed that epigenetic mechanisms are effective regulators of cellular plasticity (237). Stochastic gain of DNA methylation continuously generate epigenetic diversity in single-cell methylomes that translate to phenotypic diversity, allowing cancer cells to articulate its sustainable phenotypes in the changing tumor environment (87, 238).

In addition, reconstituted heterogeneity with mixed stable keratin-7 clones suggests that inter-cellular communications between phenotypically divergent cells may be involved in their inter-conversions and in establishing equilibrium state between cancer cells populations. These results are supported by observations of cellular plasticity where cells can dynamically convert between different cellular identities either spontaneously or under induced environmental conditions (239-241).

We cannot exclude the possibility that stochastic phenotype instability may in the course of tumor progression facilitate additional mutations causing different tumor behaviors that cooperate with migratory and stem cells programs, or activate EMT-related processes dependent on inducers from the tumor's microenvironment (62). Beside stable covalent or transient epigenetic modifications, we also show that additional variability-inducing mechanisms, such as transient heritable partitioning of proteins in daughter cells (110), could account for heterogeneous keratin-7 expression in K7het clones.

Biological variability is a property of gene expression, which is inherently stochastic and defines cellular identity in clonogenic cell populations (98). RNA-seq data analysis consists of applying statistical models to account for the variance in gene expression and excludes true biological variability from non-biological sources of variation in RNA-seq data (shot and technical noise) (242, 243). Our RNA-seq data analysis revealed that variance in gene expression distinguished true biological variability between different tumor cells. By taking into account the coefficients of gene expression variation over biological replicates, we further implicated the role of cellular plasticity concerning the stability of genetic networks and heterogeneity of neoplastic phenotypes (17, 244). The transcriptional profiles of propagated subclones demonstrated a reduction of global variance in gene expression with an increasing number of clonal generations derived from either heterogeneous or stable phenotypes of parental subclones. These results suggest that progressive clonal history might be reflected in the dynamic transcriptional profiles supporting the transition from an unstable towards a more stable phenotypic state, thus generating distinct stable cell populations contributing to intratumor heterogeneity. Furthermore, 5-aza-dC treatments greatly reduced the per-gene variance between all clonal types supporting the role of global changes in methylation patterns in establishing variations in gene expression profiles.

Cellular reprogramming and stem cell differentiation implicate a stochastic and hierarchical phase in the transdifferentiating processes (245). It is conceivable that gene expression variability in early stages of transdifferentiation precedes a heritable and constant methylation state, with minimized relative gene expression variability associated with terminally differentiated cell populations. We also acknowledge that similar heritable epigenetic mechanisms may be involved in stabilization of gene expression and generation of phenotypic heterogeneity in regionally separated primary tumor compartments.

There has been recent recognition of the similarities in epigenetic mechanisms between cellular reprogramming and the transformation of normal cells. Although different epigenetic mechanisms exist in both scenarios (246), it is quite clear that global changes in methylation patterns are paramount in both oncogenic transformation and induced pluripotency events. More importantly, shared features between these cellular events imply that epigenetic reprogramming could be amenable for conversions of cancer cells into cell states with lesser tumorigenic properties. Insights into the phenotypic instability of cancer cells may further open opportunities for induced differentiation and/or homogenization therapies of poorly differentiated sarcomatoid liver cancers in the process described as

epigenome “re-shuffling” of cancer cells (234). With the increasing understanding of how epigenetic defects drive cancers, epigenetic therapy is showing clinical promise (247). In contrast to genetic abnormalities, epigenetic changes are potentially reversible (248, 249). Demethylation therapies have been applied in restoring normal expression of differentiation related genes by systematically inhibiting DNA methylation writers such as DNA methyltransferases (Dnmt1, Dnmt3a, and Dnmt3b) (250-252). This approach has proven to be successful in promoting the differentiation of cardiac progenitors or mesenchymal stem cells (253, 254).

Evaluating heterogeneity of cellular dedifferentiation can offer a clinically relevant diagnostic and prognostic biomarker of cancer invasion and metastasis as was recently suggested for genetic diversity of tumors (255). Epigenetic variability was predictive of patient prognosis and survival in ovarian, cervical, endometrial, and breast cancers (83, 256). Accordingly, intratumor heterogeneity of keratin-7 expression may offer a simple diagnostic read-out of tumor dedifferentiation, and a possible indicator of prognostic outcome. In addition, changes in gene promoter methylation levels may provide indicators for specific tumor behavior, or serve as markers for the effectiveness of pharmacological intervention (257). For example, hypermethylation markers in tumor biopsies can mark the effectiveness of the treatments or be used as a survival marker predicting the biological aggressiveness of tumors (258, 259).

In explaining stochastic phenotype changes, we cannot exclude other heterogeneity-inducing mechanisms, not tested in this study, such as chromatin modifications, noise in gene expression (260), or influence of tumor microenvironment (233) which may be involved in directing a phenotype that is more tumorigenic, and contributes to local invasion and metastasis of tumors. Cell diversity present in tumor cell populations will necessitate further research that will identify contributions of non-genetic mechanisms in determining functional behaviors of specific cellular subsets. Sequencing approaches of epigenetically mediated cancer transcriptomes will facilitate the discovery of specific mechanisms or diagnostic biomarkers that contribute to tumor heterogeneity with increased disposition to metastasis.

3 Expression of Cathepsin F in HCC

3.1 Aim of the study

Cathepsin F (CTSF) has a defined role in physiological and pathological processes that has studied in several disease models, including cancer. Conflicting results were generated concerning its role in carcinogenesis. Recently, CTSF was implicated as a tumor suppressor gene. We aimed to investigate CTSF expression in HCC and non-neoplastic liver samples and its association with clinicopathological characteristics and patient survival. Furthermore, we investigated if CTSF expression correlated with other prognostic parameters for HCC (Ki-67 and p53), and whether CTSF has a prognostic role in hepatocarcinogenesis. We examined CTSF expression in HCC by IHC, Western blot, and qRT-PCR.

3.2 Methods

Please note that parts of this section will be published in: Matak A, Kashofer K, Zoranovic T, Pichler M, Penninger J, Zatloukal K. Cathepsin F is underexpressed in invasive HCC.

3.2.1 Multi-cancer screen

The expression of forty-six candidate genes was screened in fifteen different human cancer tissues each consisting of 8-15 tumors, and in 3-6 normal tissues. In total, 31,000 qRT-PCRs were generated in the screen. Primers were designed in PerlPrimer (261). If several transcripts for one gene were found in the Ensembl database (262), we used only the mRNA region common in all transcripts for design. One of the primers spanned an exon-intron boundary. We selected two primer pairs for each gene, one with the least secondary structure and one in another exon. Primers were designed only in the translated region of the mRNA. The primers were tested on a 384 plate, each using 50 ng (100 %), 12.5ng (25 %) and 0 ng (0 %) cDNA mix (liver, lung, colon, placenta) in duplicates using SYBR® Green PCR Master Mix. After the run, the plates were analyzed in SDS 2.3 (Applied Biosystems 7500 Real-Time PCR Software). We used one detector per primer pair, the 100 % and 25 % samples were set as unknowns, and the 0 % as non-template control (NTC).

Automatic thresholding was used to determine the Ct values of all samples. The Ct values were imported into Openoffice Calc and the primers were assessed by the following criteria:

1) Estimation quality of the 25 % sample (i.e. are there two Ct values between 100 % and 25 %);

2) Lack of amplification in the NTC;

3) If there is an amplification in the NTC, are the primer dimers (PDs) distinguishable from the correct product in the dissociation curve.

Expression of selected genes was tested by qRT-PCR.

3.2.2 Study cohort

FFPE and frozen human liver specimens were retrieved from the Biobank of the Institute of Pathology (Medical University of Graz, Austria). One hundred and one specimens of HCC on two tissue microarrays (n = 83), including 18 HCC whole tissue blocks were immunohistochemically evaluated. Direct comparison of CTSF IH scores in malignant and adjacent non-malignant was available in 87 cases. In addition, we evaluated CTSF expression in whole tissue, benign liver tumors: hepatocellular adenoma (HCA, n=11), and focal nodular hyperplasia (FNH, n = 9). In addition, 3 frozen HCC corresponding to tumor grade G1, G2, and G3 and matching non-neoplastic liver samples were used for qRT-PCR and Western blot protein analysis. Clinicopathological variables of this study cohort were directly retrieved from *AURAWeb*, a web-based meta-search engine for diagnostic findings.

The analysis of human tissues in the multi-cancer screen and HCC cohort was approved by the ethics committee of the Medical University of Graz (Number: 20-404 ex 08/09). Approved applications were written by Prof. Kurt Zatloukal and Dr. Karl Kashofer.

3.2.3 Immunohistochemistry

FFPE sections were incubated in an oven at 66°C for 15 to 20 minutes, and deparaffinized for 15 minutes in Xylol, followed by rehydration in a graded concentration of EtOH (100 %-90 %-80 %-70 %-50 %), for three minutes each, and ending by washing in PBS for five minutes. For CTSF IHC, deparaffinized tissue sections were pre-treated in sodium citrate buffer (0.01 mM, pH = 6.0) for 40 minutes in a microwave, followed by an additional 20 minutes at RT. Slides were briefly washed with PBS (pH = 7.3). Blocking was done with 1

% H_2O_2 in methanol, followed by brief washing in PBS. Slides were incubated with the anti-CTSF antibody (10 μg/ml, R&D AF2075), diluted in Dako REAL™ (DAKO) antibody diluent for one hour at RT. Detection of the primary antibody was performed by applying biotinylated secondary antibody for 30 minutes at RT, washings in PBS, and application of StreptABC Complex (DAKO K0377), followed by AEC chromogen (ready to use, DAKO K3463) until the desired intensity of the signal was reached. After immunostaining, sections were counterstained with Mayer's Haematoxylin for 60 seconds and mounted with Aquatex (Merck). IHC for p53 and Ki-67 was performed similarly, except antigen retrieval was done in pressure cooker for 5 minutes at 125°C. Monoclonal anti-p53 antibody (1:100; clone DO-7, DAKO) and monoclonal human anti-Ki-67 antibody (clone MIB-1, 1:200; DAKO) were used following avidin–biotin based detection system (DAKO) and detected by DAB substrate kit (DAKO). Negative controls included omission of the primary antibody and incubation with DAKO REAL™ antibody diluent (S2022 DAKO).

3.2.4 Immunohistological evaluation

Immunohistological analysis was performed on FFPE liver samples, presented as core biopsies (2 to 6 spots per sample) spotted on two tissue microarrays (TMA). For cases where less than two cores were available, the samples were considered as 'not available'. In non-neoplastic and malignant hepatocytes, granular staining in cytoplasm corresponding to lysosomes was considered as positive staining. Each tissue specimen was evaluated based on a semi-quantitative scoring system (IH score) (263). Staining intensities (0 to 3) and percentage (%) of positive cells (0 to 3) were multiplied and averaged in cases where two or more cores were available from the same patient. The percentage of cells showing CTSF positivity were scored as score 0 = no positive cells; score 1 = 1 to 29 % positive cells; score 2 = 30 to 60 % positive cells; score 3 > 60 % positive cells. The intensity of staining was scored as 0 = negative; 1 = weak; 2 = moderate; and 3 = strong. Staining intensities and the percentage of positive cells were multiplied and averaged if more than one evaluated core belonged to the same patient. Expression of p53 was considered positive when there was nuclear staining detected, which could have been easily distinguished from the Hemalaon blue counterstain. Criteria used for p53 scoring was identical to the CTSF IH score. Proliferation activity in HCC was performed by immunohistochemical analysis of Ki-67 (264). Positive nuclei were counted in five high-

magnification viewing fields (400x) in 13 of 18 whole tissue slides of HCC cases. Between 500-1,200 cells were scored per sample. In addition, we analyzed Ki-67 staining in tissue microarray HCC samples, grouping patients in three proliferation categories: Low (<10 %), Medium (11-50 %), and High (>51 %) as described (265).

3.2.5 Kaplan–Meier curve analysis

Patient data from tissue microarray (n = 83) and isolated cases (n = 18) were used to generate survival curves. All patients who did not succumb due to primary disease, and patients who were still alive at the end of analysis, were censured. Among the analyzed censored patients, seventeen patients died within the first month after the operation. Cancer-specific survival was defined as the time from diagnosis until cancer-related death or end of follow-up.

3.2.6 Statistical Analysis

All statistical analyzes were performed using GraphPad Prism (version 8.00 for OS X, (GraphPad Software, La Jolla California USA, www.graphpad.com). Association between CTSF expression and presence of cirrhosis, HBV/HCV status, gender, and age were evaluated by chi-square test or Fisher's exact test. Patients were subgrouped according to the presence of CTSF expression and their outcomes were analyzed by the Kaplan-Meier method. Mann–Whitney U-test and the Student's t-test were used where appropriate. Spearman's correlation was calculated to compare IHC CTSF expression with tumor grade and p53 IH scores. A p-value of <0.05 was accepted as statistically significant.

3.3 Results

Please note that parts of this section will be published in: Matak A, Kashofer K, Zoranovic T, Pichler M, Penninger J, Zatloukal K. Cathepsin F is underexpressed in invasive HCC.

3.3.1 Differential expression of Ras-cooperating tumor suppressor genes in multiple cancer types

HCC displays low incidence of K-Ras mutations; however, activation of the Ras signaling pathways is a common event in HCC (266). In the first *in vivo* model of Ras-driven liver tumorigenesis, Nguyen *et al.* showed that high expression of activating Ras mutation ($KRAS^{V12}$) activated the ERK and Wnt-signaling pathway, and recapitulated the tumorigenesis process from hyperplasia to invasive HCC, which correlated with p53-dependent activation of senescence. In addition, the study revealed additional pathways that were activated in Ras-driven liver tumorigenesis (267).

In a collaboration with the Penninger's group, we analyzed mRNA expression of 46 novel candidate genes with tumor suppressor properties, recently identified in a *Drosophila* epithelial tumorigenesis screen and validated in Ras-driven pancreatic cancer (161).

We selected 15 different tumor entities, each consisting of 10 to 15 cancer samples and 3 to 6 normal tissue samples. The following tumors with corresponding surrounding normal tissues were included in the screen: liposarcoma (n = 10), adipose tissue; glioblastoma (n = 11), brain normal (n = 4); breast cancer (n = 15), breast fibroadenoma (n = 6), breast normal (n = 4); colon cancer (n = 10), colon normal (n = 3); endometrium cancer (n = 10), endometrium normal (n = 3); kidney cancer (n = 10), kidney normal (n = 5); HCC (n = 10), liver normal (n = 4); lung cancer (n = 10), lung normal (n = 6); B-cell lymphoma (n = 8), lymph node normal (n = 3); myosarcoma (n = 10), muscle normal (n = 3); ovary cancer (n = 10), ovary cystadenoma (n = 4); pancreas cancer (n = 10), pancreas normal (n = 3), prostate cancer (n = 8), prostate normal (n = 3); stomach cancer (n = 10), stomach normal (n = 3); testicular seminoma (n = 10), testis normal (n = 3); thyroid cancer (n = 10), thyroid gland normal (n = 3). Among the 45 analyzed genes, CTSF showed striking downregulation in all tumors analyzed in the screen. (Figure 3-1). Consequently, we chose to validate expression of CTSF in a separate cohort of HCC patients.

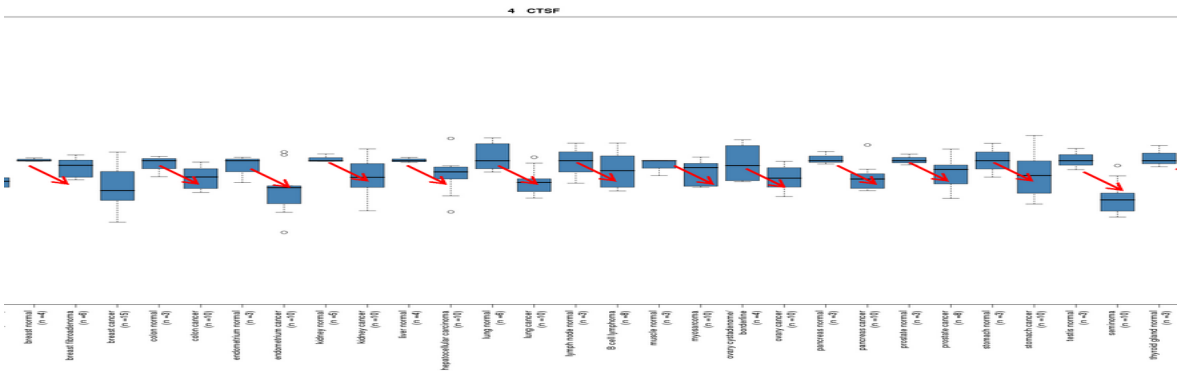


Figure 3-1. Comparative Cathepsin mRNA expression in a series of 15 tumor entities.

Magnified Box plot analysis displays CTSF mRNA expression in different cancers and their corresponding normal tissues. y axis represents the log2 values of 2ddCt value (Adapted from .

3.3.2 Evaluation of the specificity of anti-human CTSF antibody and expression of CTSF in HCC and non-neoplastic liver

Since we used commercially available polyclonal CTSF antibody, we validated the specificity of the antibody for IHC on FFPE liver samples from CTSF knock-out and wild-type mice. As expected, wild-type mouse liver exhibited strong, and granular staining throughout all hepatocytes with increased concentration around the central hepatic vein, while there were no staining in the gene knock-out mice (Figure 3-2A). In addition, we validated expression in different mouse organs based on published Northern blot analysis of CTSF RNA expression (268) (Figure 3-2B).

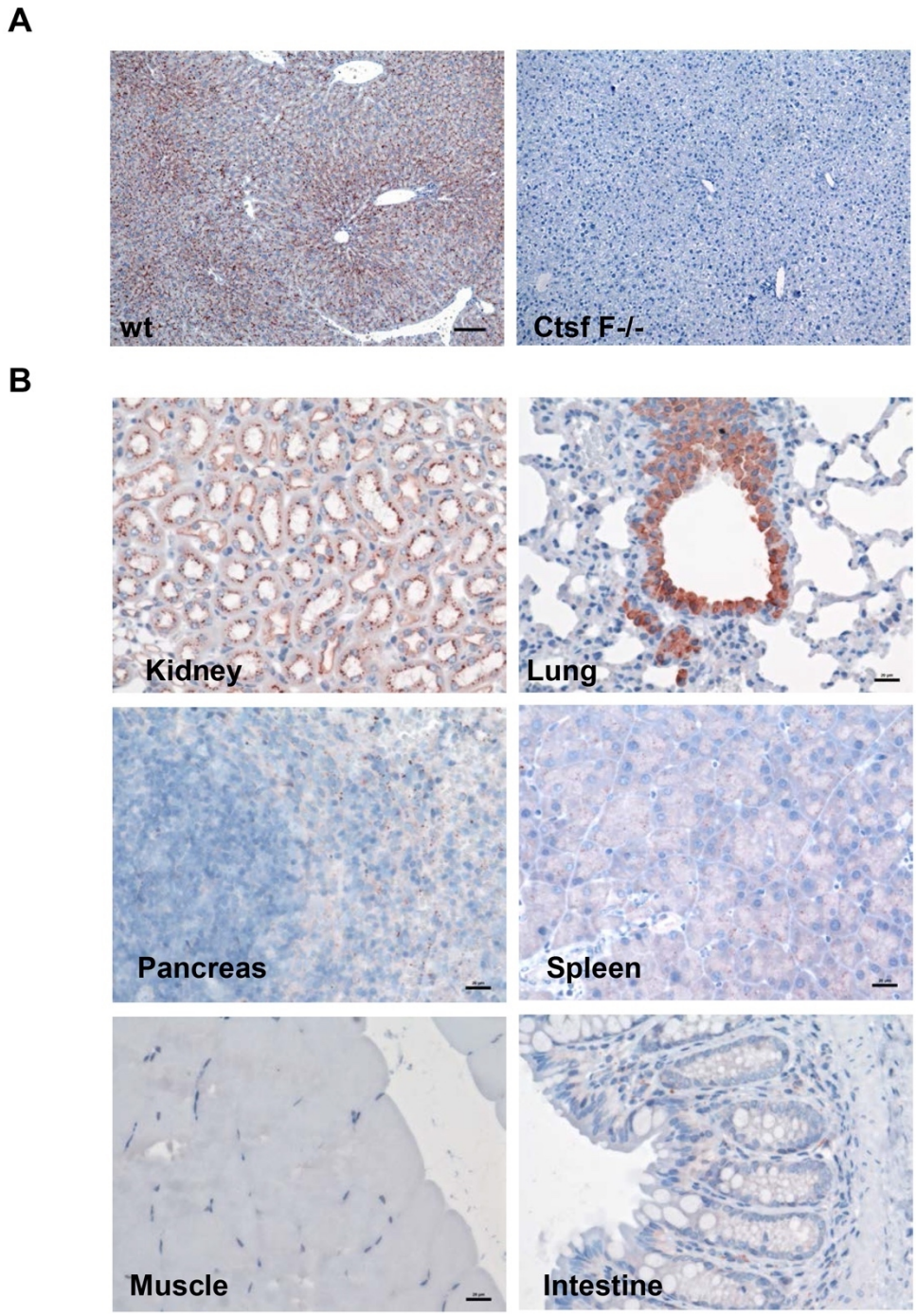


Figure 3-2. Validation of the anti-CTSF polyclonal antibody.

(A) Commercially available human anti-CTSF antibody (Polyclonal Goat IgG, Ala20-Asp484, # AF2075) was used on mouse normal and CTSF knockout liver. (B) Expression of CTSF in various mouse tissues and organs. Magnification in panel A is x100 and in panel B is x200. (Adapted from .

In non-neoplastic liver, CTSF was uniformly expressed inside cells cytoplasm in a granular pattern (Figure 3-3A). Its distribution and intensity were more pronounced around the central vein in the non-neoplastic liver (Figure 3-3A). When CTSF was expressed in HCC

samples, it formed aggresome-like inclusions with distorted distribution inside the cells (269) (Figure 3-3B). CTSF was also expressed in macrophages (Kupffer cells) in HCC (Figure 3-3C). There was no cross-reactivity with the anti-goat CTSF polyclonal antibody and auto-fluorescent lipofuscin, commonly occurring in aged hepatocytes (270) (Figure 3-3D). Staining patterns were concordant with the staining images available through The Human Protein Tissue Atlas (271).

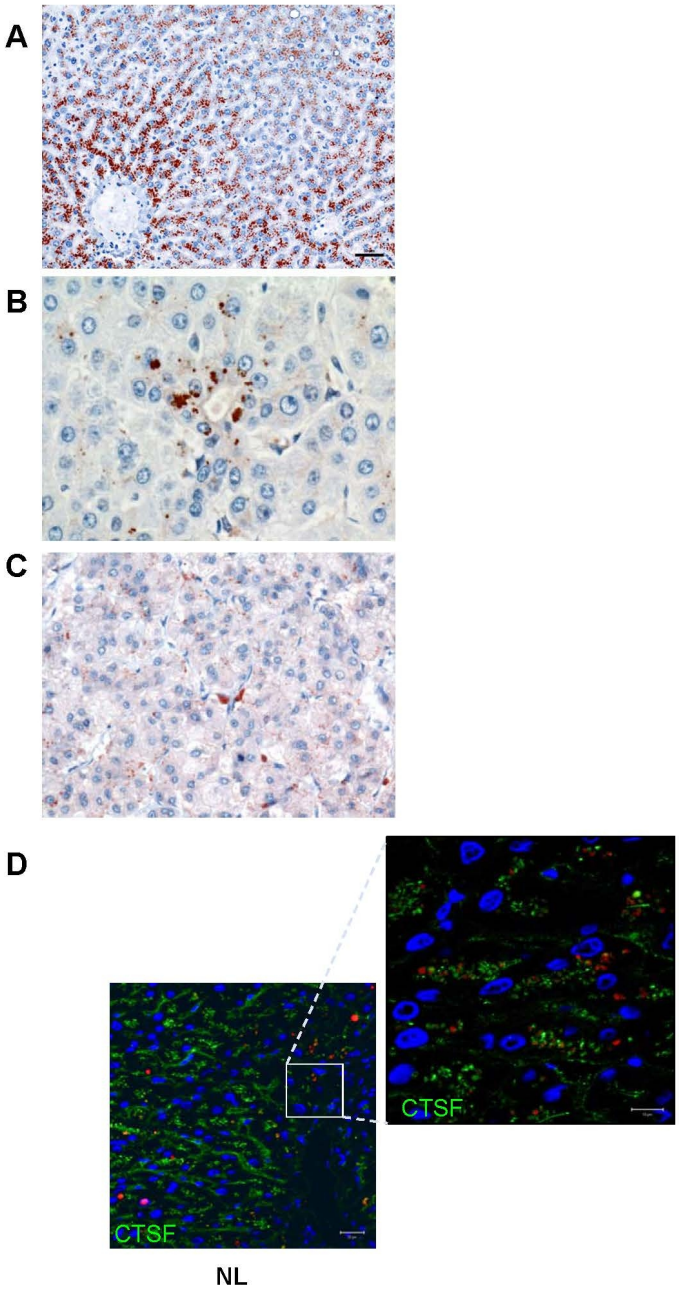
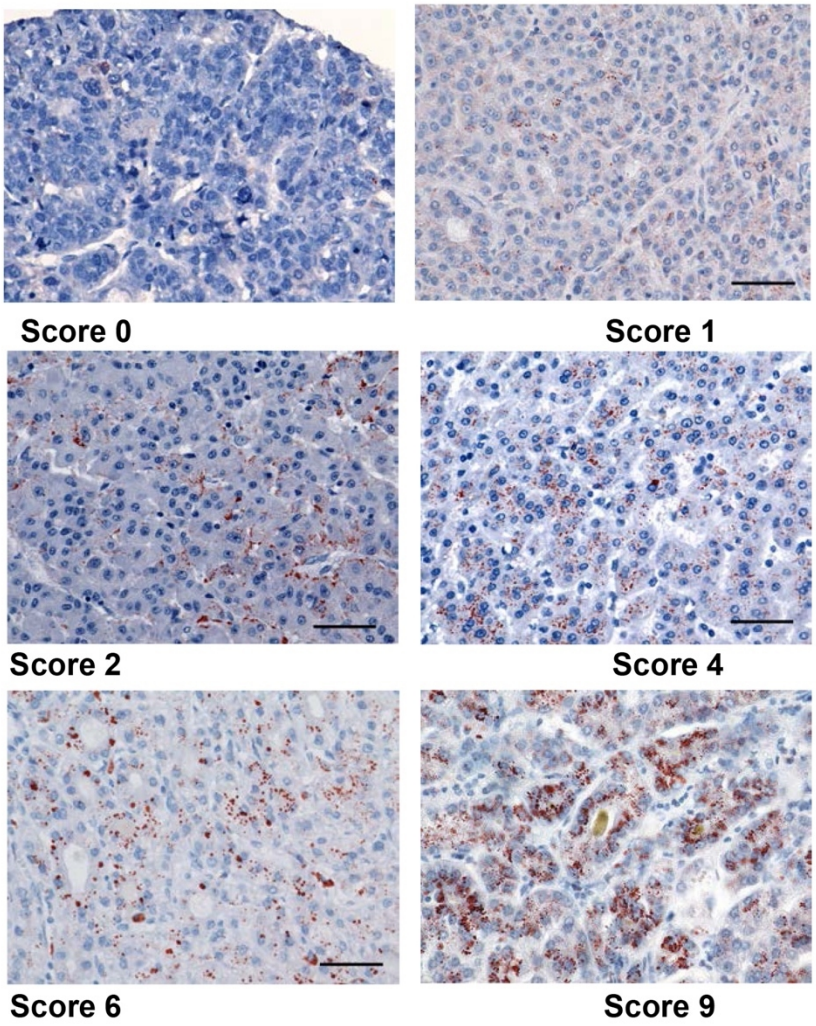


Figure 3-3. Expression features of CTSF in HCC and non-neoplastic liver.

(A) Expression of CTSF in normal hepatocytes is localized around central hepatic vein. (B) Malignant hepatocytes in well-differentiated HCC express CTSF which form aggresome-

like inclusions. (C) Expression of CTSF in macrophages. Magnification: Figure A: 200x; Figure B: 600x C: 400x. (D) Expression of CTSF in non-neoplastic liver (NL) (green) and autofluorescent lipofuscin (red) do not co-localize in lysosomes of hepatocytes. Scale bar 20 μm. Scale bar in the zoomed segment: 10 μm.

A



B

Intensity	Positive cells (%)
0=no staining	0=absence
1=weak	1=1-30%
2=mild	2=31-60%
3=strong	3=61-100%

IHC score= Intensity x positive cells

Figure 3-4. Downregulation of CTSF expression in HCC

(A) Representative immunohistochemistry pictures of CTSF IH scores (IH score = intensity x percentage). (B) IHC parameters used for calculating CTSF IH score. Magnification: x400.

Furthermore, CTSF expression was analysed in the HCC samples both in tissue microarray described before (272), and in separate sections from FFPE blocks. Calculation of the CTSF IH score and examples of HCC samples with different IH scores are shown in Figure 3-4 A, B. CTSF expression was absent in 70 out of 101 analyzed HCC samples with significantly higher CTSF IH scores associated with the adjacent non-neoplastic liver compared to HCC (Mann-Whitney test, $p<0.001$) (Figure 3-5 A, B). We then evaluated whether CTSF expression correlates with clinicopathological characteristic of HCC. We were able to partially retrieve verified patient data (n = 73), including gender, age, HBV and HCV status, presence of cirrhosis, tumor stage and tumor grade. There was no significant correlation between the data available and loss or presence of CTSF expression in HCC (all p-value>0.05, data not shown). Although CTSF knockdown was associated with increased proliferation and inhibition of apoptosis in gastric cancer (180) there was no correlation with tumor size in the analyzed samples (data not shown). However, when comparing HCC of different grades (G1, n = 29; G2, n = 47 and G3, n = 24) with CTSF expression, CTSF downregulation was associated with increasing HCC grade (Kruskal-Wallis test, G1-G2: p = 0.01; G1-G3: p = 0.01) (Figure 3-5C). These results imply that the loss of CTSF correlates with invasive behaviour of HCC.

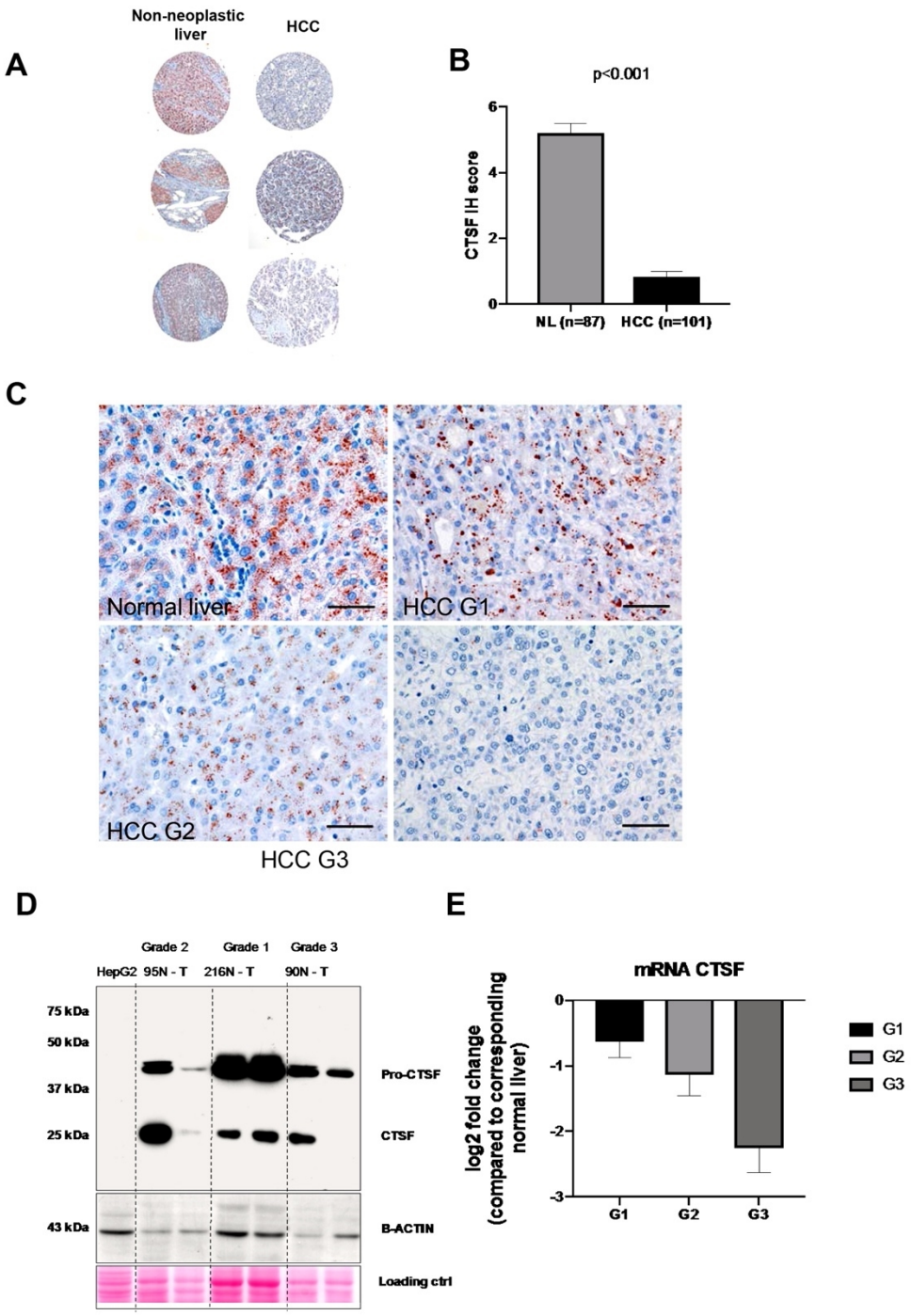


Figure 3-5. Immunohistochemical (IH) expression of CTSF in HCC and non-neoplastic liver.

(A) Representative photomicrographs of CTSF staining in tissue cores assembled on tissue microarrays. Magnification x100. (B) Expression of CTSF IH score in NL and HCC. Box-plot analysis by Mann-Whitney U test, $p<0.001$. (C) IH expression of CTSF in non-neoplastic liver and differentiation grades of HCC (HCC G1-G3). (D) Western blot of tumor lysates (T) and matched non-neoplastic liver (N) from three HCC samples representing different differentiation grade. HepG2 cell line was used as a negative control. Loading controls were evaluated against beta-actin and Ponceau S stain. (E)

Downregulation of CTSF mRNA levels is related to HCC differentiation grade (n = 2-3). Magnification in (A): 100x; Magnification in (C): 400x.

3.3.3 Western-blot analysis of CTSF protein levels in three differential HCC grades

Further validation of CTSF correlation with HCC differentiation grade was done by Western blot, where we also directly compared the specificity of anti-CTSF antibody and presence of different enzymatic forms of CTSF). Three HCC samples representing three differentiation stages and respective non-neoplastic liver were used (95N-T, 216N-T and 90N-T) (Figure 3-5D). In the non-neoplastic liver samples (90N, 216N and 95N), anti-CTSF antibody detected both activated and inactivated enzymatic forms of CTSF (pro-cathepsin F). The smallest band (~27 kDa) corresponds to active CTSF devoid of the pro-segment, while the two heavier bands (44, and 47 kDa) represent inactivated forms and are either products of two splice variants, or two glycosylated forms of CTSF (181) (Figure 3-5D). In higher grade HCC samples (90T and 95T), pro-cathepsin F bands protein bands at 47 kDa, and 27 kDa were downregulated, while the active form was completely missing (Figure 3-5D). In tumor sample (216T) all three bands were detected, confirming the results of IHC analysis in our HCC cohort (Figure 3-5C). The findings correspond to the published immunoblotting data with small discrepancies resulting from different experimental setup, such as SDS gel percentage formulations (269, 273). Based on the literature and computational predictions, full-length CTSF protein has a molecular weight of 53 kDa, while pepsin proteolytic activation results in activated CTSF with a molecular weight of 31 kDa. Furthermore, expression of CTSF mRNA originating from FFPE tissues corresponding to different tumor grade (G1-G3) was analyzed. CTSF mRNA expression was highest in low- grade HCC (G1) and progressively decreased in lower-differentiated HCC (G2-G3) (Figure 3-5E).

3.3.4 Expression of CTSF in benign liver tumors

Next, we analyzed if changes in CTSF expression occur in benign liver lesions Focal nodular hyperplasia (FNH) and Hepatic adenoma (HA) compared to adjacent non-neoplastic liver (NL). There was no statistical difference in CTSF expression between NL and FNH, or NL and HA. NL, FNH, and HA had significantly higher expression of CTSF compared to HCC (Kruskal-Wallis test, $p<0.001$) (Figure 3-6B).

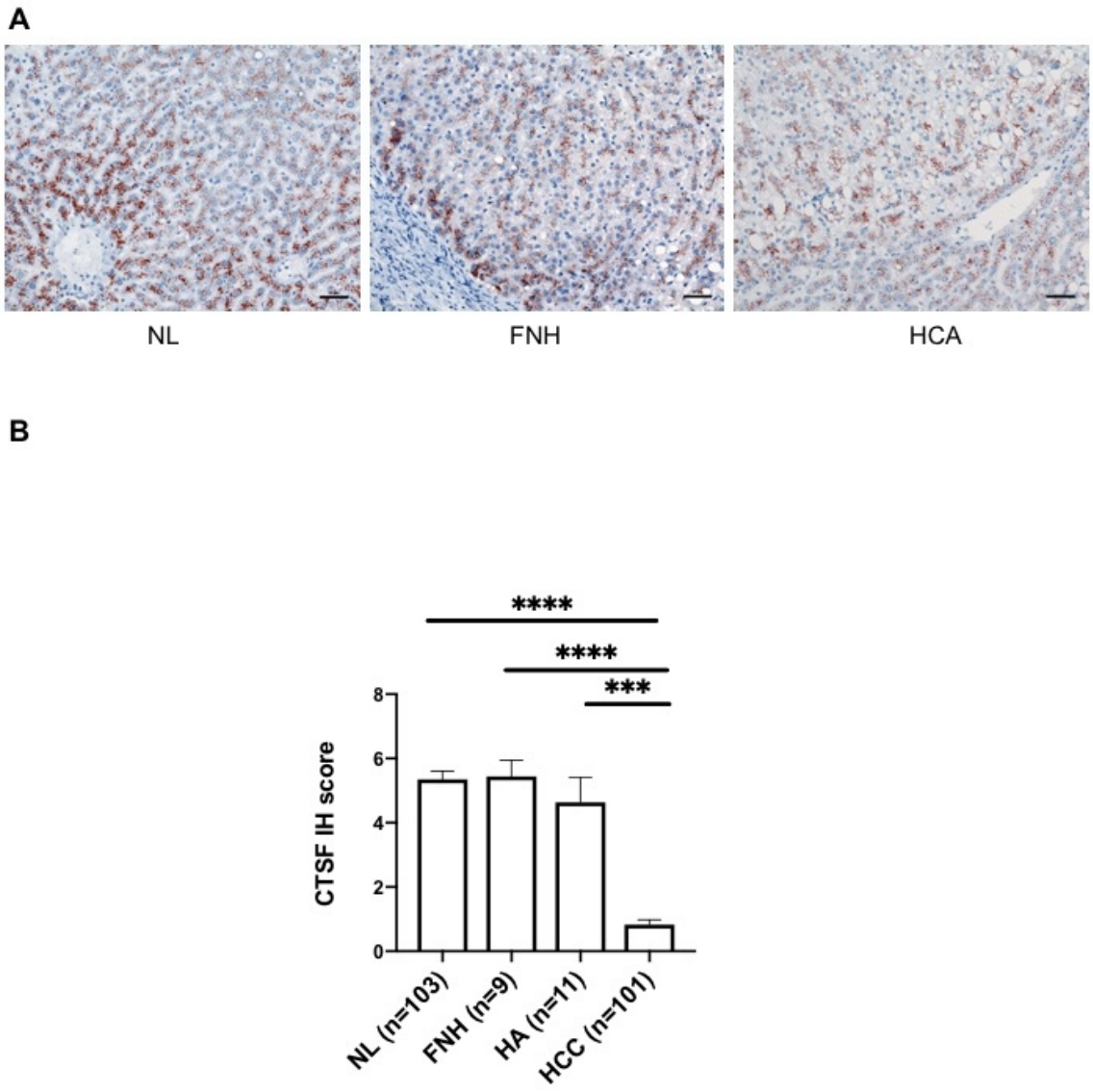


Figure 3-6. Expression of CTSF in non-neoplastic liver, Focal nodular hyperplasia (FNH) Hepatic adenoma (HA), and HCC.

(A) Representative pictures of CTSF staining in NL, benign liver lesions (FNH and HCA). (B) Bar graph of CTSF IH scores in NL, FNH, and HA and HCC. Scale bar: 0.05 mm.

3.3.5 IHC analysis of Ki-67 and p53 protein in HCC

Tumor staining with anti-Ki-67 antibody is a common diagnostic tool used to differentiate types of neoplasm. The antibody detects Ki-67 antigen, a marker of cellular proliferation, which is present in the nuclei of cells in the G1, S, and G2 phases of the cell division cycle as well as in mitosis (274). Ki-67 staining is an excellent operational marker in malignant neoplasm because it can detect proliferating growth fraction of cells in a given cell population (proliferation index), which is often difficult to assess in tumor diagnostics.

Ki-67 equivalent monoclonal antibody MIB-1 is today routinely used on paraffin sections after heat-induced antigen retrieval (275). A recent meta-analysis showed that Ki-67 expression is associated with cancer progression and overall survival of patients with HCC (276). To evaluate if there was any correlation between proliferating activity and CTSF expression, samples were categorized into three Ki-67 grading groups: Low (<10 %), Medium (11-50 %), and High (>51 %). We found that CTSF was significantly inversely associated in the three Ki-67 groups (chi-square test, $p<0.01$) (Figure 3-7A, B). Besides, we evaluated the Ki-67 proliferation index in isolated whole tumor tissue blocks (n = 13) according to the method reported previously (264). We found a trend in difference in mean percentages of Ki-67 proliferation index between CTSF positive (8.6±13.4 %) and CTSF negative HCC samples (23.7±11.9) (t-test, one-tailed, $p = 0.04$, data not shown).

P53 is a multifunctional transcription factor involved in the control of the cell cycle, programmed cell death, senescence, differentiation, development, transcription, DNA replication, DNA repair and maintenance of genomic stability (277). p53 is a tumor suppressor, and its gene is mutated in approximately 50 % of cancers (278). Mutational inactivation of p53 can result in loss of control of cell division, aberrant stimulation of cell growth, increased cell survival, and genetic instability which all contribute to the development of cancer (278). Environmental factors play an important role in the type of mutations found in the p53 gene. In Western countries, the incidence of p53-mutated HCC is more than 30 % and is usually associated with poorly differentiated HCC (279).

In the entire HCC cohort (including isolated cases), nuclear and cytoplasmic staining for p53 with monoclonal anti-human p53 antibody (DAKO, clone DO-7) was present in 32 out of 92 analyzed samples (34.7 %), while negative p53 expression was observed in 73 out of 92 samples (79 %) (280) (Figure 3-7C). There was no significant correlation found between CTSF and p53 IH scores (Spearman's test, $r = -0.14$; $p = 0.16$). However, when analyzing CTSF and p53 IH score expression in consecutively stained whole tissue samples (n = 9), using the same criteria of IH scoring, there was a significant inverse correlation between the expression of two markers (Spearman's test, $r = -0.7283$, $p = 0.02607$). Furthermore, the association between positive and negative expression of the p53 and CTSF markers was tested with chi-square test for independence, resulting in a significant relationship between the expression of the two markers (chi-square test, $p<0.05$).

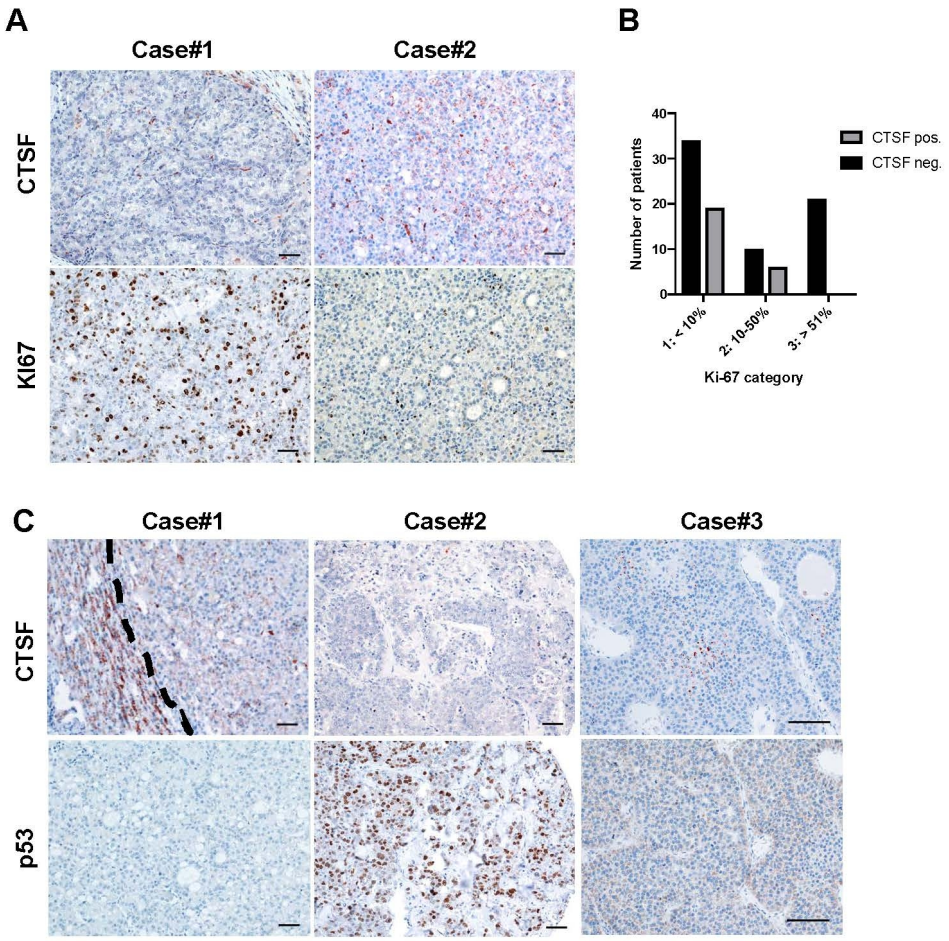


Figure 3-7. Relationship between CTSF expression and proliferating cells in HCC and p53.

(A) Representative pictures of high (Case #1) and low (Case #2) proliferating HCC and corresponding CTSF. (B) Significant inverse correlation between CTSF expression and Ki-67 grading groups (chi-square test, $p<0.01$). (C) Representative pictures of inverse expression between CTSF and nuclear and cytoplasmic expressed p53 in three isolated cases of HCC. Case#1 shows tumor border zone between non-neoplastic liver and HCC. Scale bar: 0.05 mm; Scale bar in Case#3: 0.1 mm.

3.3.6 Relationship between CTSF expression and patient outcome

We next inquired whether CTSF protein expression is associated with the outcome in patients with HCC. From our HCC cohort, we were able to retrieve clinical survival data from 73 patients. In addition, we used survival outcome data from HCC cohort collected as part of The Cancer Genome Atlas research network, n = 360 and analysed with Keplen-Meier plotter (Figure 3-5B) (281, 282). In both cohorts, subgrouping patients based on CTSF expression in the 10-year cancer-specific survival showed that low CTSF expression was correlated with poor patient survival (Figure 3-5A, B).

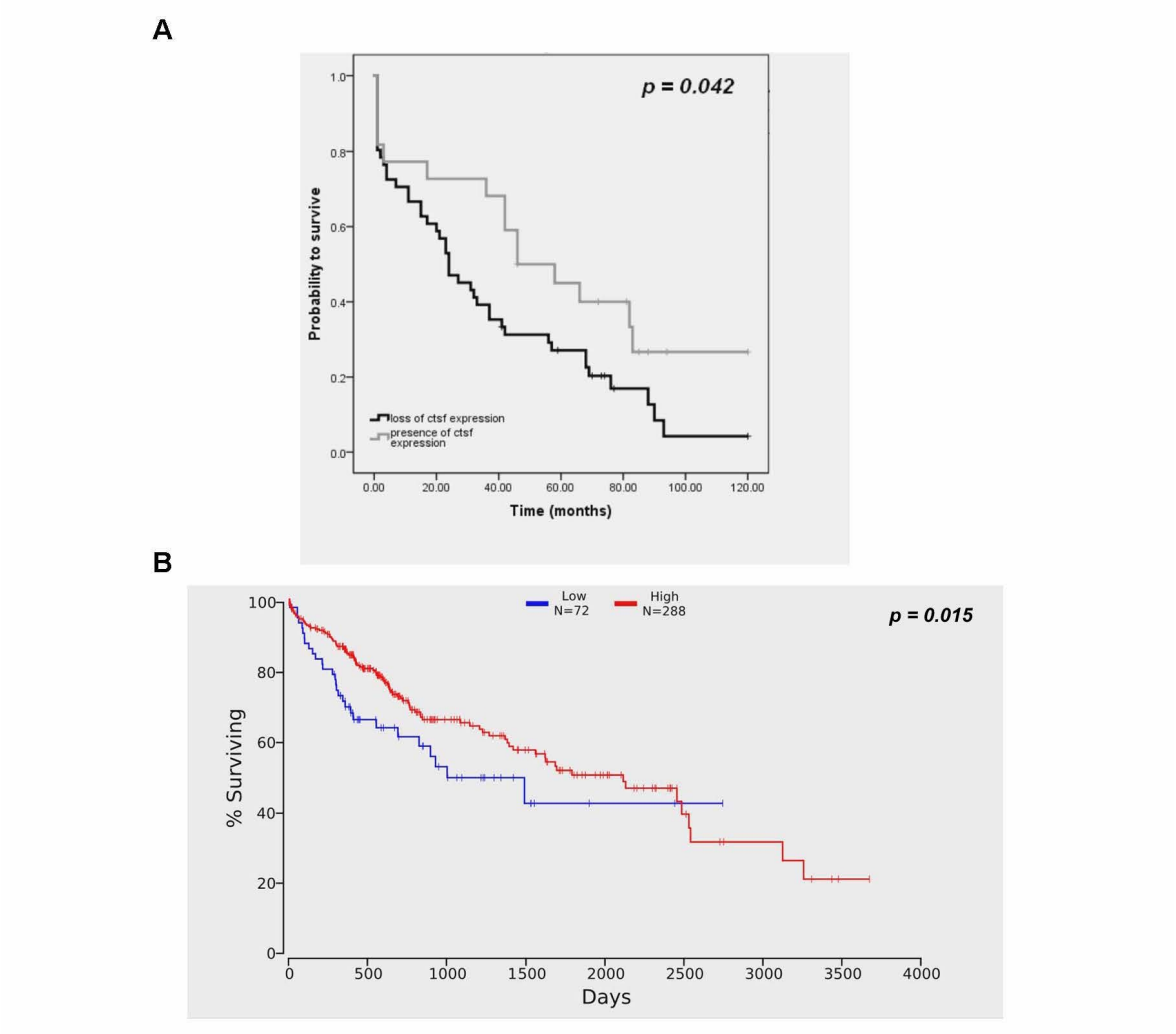


Figure 3-8. Relationship between CTSF expression and patient outcome.

(A) Kaplan–Meier plots for probability of 10-year survival in HCC patients (n = 73) that lost CTSF expression (black line) vs HCC patients with presence of CTSF expression. (B) Kaplan–Meier plots for survival data in TCGA HCC cohort (n = 360). Red line indicated high CTSF expression, blue line indicates low CTSF expression.

3.4 Discussion

Please note that parts of this section will be published in: Matak A, Kashofer K, Zoranovic T, Pichler M, Penninger J, Zatloukal K. Cathepsin F is underexpressed in invasive HCC.

We are the first to report immunohistochemical study on the loss of CTSF protein and mRNA expression in HCC. The downregulation of CTSF is associated with decreasing differentiation (grade) in HCC. Our studies are in line with an investigation of low mRNA CTSF levels in osteosarcoma and gastric cancer cell line and patient samples. Several mechanisms of CTSF downregulation can be discussed that are potentially involved in HCC. Loss of heterozygosity of the CTSF locus 11q13 was associated with increased risk of developing invasive carcinoma in breast cancer patients (169), supporting the study findings of progressive loss of CTSF protein in higher grade, undifferentiated HCC. A potential mechanism involved in CTSF downregulation in HCC is its reported role in cell death. Jeric *et al*. reported that N-truncated forms of CTSF displayed an aggresome-like formation inside cells, and were associated with autophagy markers rather than induction of apoptosis (269). This is supported by our findings that CTSF forms aggresome-like inclusions in CTSF-expressing HCC samples. Further support of CTSF involvement in programmed cell death was established in a necrosis model, where CTSF was upregulated in a p53-dependent manner (189). Altered DNA methylation of tumor suppressor gene in HCC has been previously reported (283). DNA methylation may likely be involved in CTSF downregulation in HCC. In an initial exploration of this assumption, we treated HepG2 cells with 5-Aza-dC and analyzed re-expression of CTSF mRNA at three-time points. After 48 hours incubation, we could already detect a 1-fold increase in CTSF expression, while, after 144 hours of incubation, CTSF mRNA expression increased over 3-fold (data not shown). CTSF was reported in several large microarray gene expression screen studies as well as *in vitro* studies investigating cellular senescence. CTSF was also reported as a senescence marker in β-catenin induced DNA damage in CAT-Tg DN4 thymocytes (188). Another study on human prostate epithelial cells showing permanent growth arrest and terminal differentiation reported that CTSF is a novel senescence marker (187).

In several recent reports, an interesting connection was made between CTSF involvement in p53 transcriptional networks crucial for tumor suppression by analyzing targeted knock-in p53 domain mutants (186). As CTSF contains p53 consensus sites, it is conceivable that

it is regulated in a p53-dependent manner in response to DNA damage, or by specific oncogene stress in both mouse and human cells (191, 284). CTSF was discovered to play a key role in effective p53-dependent tumor suppression transcriptional mechanisms that are distinct from robust transactivation of known p53 targets that mediate apoptosis, cell cycle arrest and cellular senescence (190, 285). In addition, CTSF was characterized as a tumor suppressor gene in c-MYC-driven lymphomagenesis, with functional Puma and p21 genes (191). In this study, no correlation was made between the CTSF expression and p53 gene mutation, even though it is established that IHC nuclear staining suggests a mutated p53 (286). The incidence of p53 mutations varies and depends on environmental and genetic risk factors. Also, p53 mutation rates are low in Western Europe (287). However, there was a significant relationship between the expression of the two markers in whole HCC cohort, where both cytoplasmic and nuclear overexpression of p53 and CTSF was scored. In addition, using the same criteria for IH scoring in consecutive tissues slides, both markers significantly correlated, suggesting that expression of both markers, due to their heterogeneous expression, is best accounted in whole tissue slides, instead of tissue microarray slide.

The findings suggests that CTSF may be involved in regulating proliferation in hepatocyte biology. In the HCC cohort, we detected that the number of cells staining positive for the proliferation marker Ki-67 was inversely correlated with CTSF expression. Supporting this observation was a study done on Cathepsin L being first cathepsin to be reported to have antiproliferative activities in HPv16-induced skin carcinogenesis mouse models (288). The protective role of CTSF was recently established in gastric cancer. High levels of CTSF autoantibodies in the serum of gastric cancer patients were predictive of longer survival (179). A follow up study provided evidence of anti-tumor properties of CTSF, being directly involved in cellular growth and apoptosis signaling pathways in gastric cancer cell lines (180). CTSF was also significantly reduced in gastric cancer patients, where CTSF correlated with tumor differentiation and invasiveness to lymph nodes (180). Our investigation also revealed that patients with resected CTSF-negative HCC had significantly poorer 10-year overall survival compared to CTSF-positive HCC.

In summary, our IHC results show a non-trivial downregulation of CTSF in a considerable number of HCC patients. Since this study is the first to propose a tumor suppressive role in HCC and with limited studies in other cancer types and available clinical data in CTSF downregulation, the exact role of CTSF in tumor progression and its correlations with p53 and Ras-mutation status requires further research.

4 General Conclusions

Today cancer still affects over 12 million people in Europe and accounts for 20.3 % overall human deaths, regardless of gender (3). Based on projections, cancer deaths will continue to rise with an estimated 9.5 million people dying from cancer in 2018, and 16.4 million dying in 2040 (3). Cancer of the liver and intrahepatic biliary tract is among the ten most common types of cancer in men and woman and the fourth death-inducing cancer (3), and current projections estimate that next year 80, 000 people will die from hepatobiliary cancers (3).

Despite the rising incidence of hepatobiliary cancers, there is an underappreciated understanding of intratumor heterogeneity and diversity in molecular and cellular processes that limit treatment options. Experimental cellular models are useful tools to analyze molecular mechanisms involved in tumorigenesis and to explore intratumor heterogeneity. Especially useful are models that recapitulate intratumor heterogeneity in human cancers.

In **Chapter 2** we describe an alternative source of intratumor heterogeneity which we describe as stochastic phenotype switching. We established a novel *in vitro* model from a case of sarcomatoid cholangiocarcinoma (sCC) that displayed extensive histological heterogeneity. Transcriptional and immunohistochemical profiles of the epithelial, mesenchymal, and metastatic tumor confirmed the involvement of dynamic epithelial-mesenchymal transdifferentiating events which correlated with loss of keratin-7 expression and showed involvement of EMT and increased competence for invasion and metastasis. At the cell population level, we identified distinct self-propagating clones characterized by stable expression of keratin-7 (K7pos), non-expressing keratin-7 clones (K7neg) and unstable clonal types with an interchangeable expression of keratin-7 (K7het). In further clonal propagations, K7het stochastically propagated either of the two stable clones and self-renewed its unstable nature. We also showed that no transitions occurred from K7neg to K7pos cells proving that K7neg clones represent stable states with epigenetically determined mechanisms of maintenance. However, stochastic and rare transitions occurred between K7pos clones to K7neg clones.

We were able to analyze differential transcriptome profiles in differential keratin-7 phenotypes. Transcriptional signatures were enriched in categories relating to extracellular space, cell adhesion, cellular development, Wnt-signaling and EMT-like transitions.

Mechanistically, stochastic and reversible hypermethylation at KRT7 promoter was involved in establishing stable transcriptional identities between differentially expressing keratin-7 clones. However, in K7het subclones we showed that additional epigenetic processes are involved and that these changes happen during unequal partitioning of keratin-7 in daughter cells during mitosis. Xenotransplantation essay revealed functionally consequences linked with a keratin-7 negative expression showing increased tumorigenic potential in unstable K7het clones. Stochastic phenotype switching is an alternative source of intratumor heterogeneity in sCC, which goes beyond clonal genetic diversity, hierarchical stem cell differentiation, and contribution of microenvironment signals. Increased variability in the epigenetic landscape may increase cell proclivity for stochastic phenotype switching or provide enduring stable epigenetic states to serve as a substrate for selection of advantageous cell properties under changing tumor environment. Accounting for this source of intratumor heterogeneity of cellular dedifferentiation may increase chances for finding clinically relevant diagnostic and prognostic biomarkers for cancer invasion and metastasis. Current studies on genomic analysis in HCC or CC discovered several mutations affecting both tumor suppressor genes and oncogenes. These include: telomerase reverse transcriptase (*TERT*) promoter mutations (60 %), catenin beta 1 (*CTNNB1*) mutations (30 %), axin inhibition protein 1 (*AXIN1)*, tumor protein p53 (*TP53*) mutations (12–48 %), AT-rich interaction domain 1A (*ARID1A*) mutations (10 %) (289). HCC is one of the rare solid tumors that does not have mutations which can be druggable, not to mention mutations that can be used for predicting therapeutic responses in clinical practice (290). Mutations in the Ras signaling pathway are common in cancers and account for approximately 30 % of human cancers. Since Ras mutations alone are not sufficient for malignant progression, it is important to discover cooperating genes and signaling pathways involved in Ras-driven cancers. *In vivo* RNAi-mediated knockdown screen in *Drosophila* presents a fast and affordable model having 75 % conserved human disease gene and epithelial tissue organization. *Drosophila* model has been used for the first time for large scale tumor suppressor screens allowing analysis of known and new genes involved in tumorigenesis of epithelial tumors. Gene candidates can easily be used for enrichment analysis of signaling pathways in multiple databases and reveal potentially new pathways in Ras-driven epithelial tumorigenesis.

In **Chapter 3**, we provide preliminary results of validating one of top candidate tumor suppressor gene recently discovered in *Drosophila* screen involved in Ras-driven tumorigenesis (161). We found that CTSF is significantly downregulated at protein and

mRNA levels in HCC and is inversely associated with differentiation grade and proliferation potential in HCC. We are the first to report that stratifying HCC patients based on CTSF expression can predict poorer 10-year survival. Another outlook of our study is that CTSF expression might be implicated in p53-mediated tumor-suppressing mechanisms linked to senescence.

5 References

1. Blechacz B. Cholangiocarcinoma: Current Knowledge and New Developments. Gut Liver. 2017;11(1):13-26.
2. Everhart JE, Ruhl CE. Burden of digestive diseases in the United States Part III: Liver, biliary tract, and pancreas. Gastroenterology. 2009;136(4):1134-44.
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
4. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B, et al. Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. J Natl Cancer Inst. 2017;109(9).
5. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet. 2018;391(10127):1301-14.
6. Nault JC, Villanueva A. Intratumor molecular and phenotypic diversity in hepatocellular carcinoma. Clin Cancer Res. 2015;21(8):1786-8.
7. Schulze K, Nault JC, Villanueva A. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. J Hepatol. 2016;65(5):1031-42.
8. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. Gastroenterology. 2015;149(5):1226-39.e4.
9. Kojiro M, Sugihara S, Kakizoe S, Nakashima O, Kiyomatsu K. Hepatocellular carcinoma with sarcomatous change: a special reference to the relationship with anticancer therapy. Cancer Chemother Pharmacol. 1989;23 Suppl:S4-8.
10. Nonnis R, Paliogiannis P, Giangrande D, Marras V, Trignano M. Low-grade fibromatosis-like spindle cell metaplastic carcinoma of the breast: a case report and literature review. Clin Breast Cancer. 2012;12(2):147-50.
11. Chrysikos D, Zagouri F, Sergentanis TN, Goutas N, Vlachodimitropoulos D, Flessas I, et al. Mucinous tubular and spindle cell carcinoma of the kidney: a case report. Case Rep Oncol. 2012;5(2):347-53.
12. Seok JY, Na DC, Woo HG, Roncalli M, Kwon SM, Yoo JE, et al. A fibrous stromal component in hepatocellular carcinoma reveals a cholangiocarcinoma-like gene expression trait and epithelial-mesenchymal transition. Hepatology. 2012;55(6):1776-86.
13. Yoon GS, Choi DL. Sarcomatoid carcinoma of common bile duct: a case report. Hepatogastroenterology. 2004;51(55):106-9.
14. Lee M-J, Yu G-R, Yoo H-J, Kim J-H, Yoon B-I, Choi Y-K, et al. ANXA8 down-regulation by EGF-FOXO4 signaling is involved in cell scattering and tumor metastasis of cholangiocarcinoma. Gastroenterology. 2009;137(3):1138-50, 50.e1-9.
15. Sato Y, Harada K, Itatsu K, Ikeda H, Kakuda Y, Shimomura S, et al. Epithelial-Mesenchymal Transition Induced by Transforming Growth Factor-β1/Snail Activation Aggravates Invasive Growth of Cholangiocarcinoma. The American Journal of Pathology. 2010;177(1):141-52.
16. Campbell LL, Polyak K. Breast tumor heterogeneity: cancer stem cells or clonal evolution? Cell cycle (Georgetown, Tex). 2007;6(19):2332-8.
17. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;12(5):323-34.
18. Melo FDSE, Vermeulen L, Fessler E, Medema JP. Cancer heterogeneity—a multifaceted view. EMBO reports. 2013;14(8):686-95.

19. Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. British Journal of Cancer. 2013;108(3):479-85.
20. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. The New England journal of medicine. 2012;366(10):883-92.
21. Farhangfar CJ, Meric-Bernstam F, Mendelsohn J, Mills GB, Lucio-Eterovic AK. The Impact of Tumor Heterogeneity on Patient Treatment Decisions. Clinical Chemistry. 2012;59(1):38-40.
22. Hiley C, de Bruin EC, McGranahan N, Swanton C. Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. Genome Biol. 2014;15(8):453.
23. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet. 2014;46(3):225-33.
24. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
25. Stoler DL, Chen N, Basik M, Kahlenberg MS, Rodriguez-Bigas MA, Petrelli NJ, et al. The onset and extent of genomic instability in sporadic colorectal tumor progression. Proc Natl Acad Sci U S A. 1999;96(26):15121-6.
26. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976;194(4260):23-8.
27. Cairns J. Mutation selection and the natural history of cancer. Nature. 1975;255(5505):197-200.
28. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481(7381):306-13.
29. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta. 2010;1805(1):105-17.
30. Tabassum DP, Polyak K. Tumorigenesis: it takes a village. Nat Rev Cancer. 2015;15(8):473-83.
31. Waclaw B, Bozic I, Pittman ME, Hruban RH, Vogelstein B, Nowak MA. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. Nature. 2015;525(7568):261-4.
32. Anderson AR, Weaver AM, Cummings PT, Quaranta V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. Cell. 2006;127(5):905-15.
33. Marusyk A, Polyak K. Tumor heterogeneity: Causes and consequences. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2010;1805(1):105-17.
34. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. Cancer Cell. 2015;27(1):15-26.
35. Bu P, Chen KY, Lipkin SM, Shen X. Asymmetric division: a marker for cancer stem cells in early stage tumors? Oncotarget. 2013;4(7):950-1.
36. Clevers H. The cancer stem cell: premises, promises and challenges. Nature Medicine. 2011:313-9.
37. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003;100(7):3983-8.
38. Abuetabh Y, Persad S, Nagamori S, Huggins J, Al-Bahrani R, Sergi C. Expression of E-cadherin and β-catenin in two cholangiocarcinoma cell lines (OZ and HuCCT1) with different degree of invasiveness of the primary tumor. Annals of clinical and laboratory science. 2011;41(3):217-23.

39. Yang ZF, Ho DW, Ng MN, Lau CK, Yu WC, Ngai P, et al. Significance of CD90+ Cancer Stem Cells in Human Liver Cancer. Cancer Cell. 2008;13(2):153-66.
40. Prince ME, Sivanandan R, Kaczorowski A, Wolf GT, Kaplan MJ, Dalerba P, et al. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. Proc Natl Acad Sci U S A. 2007;104(3):973-8.
41. Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. Prospective identification of tumorigenic prostate cancer stem cells. Cancer Res. 2005;65(23):10946-51.
42. Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, et al. Identification of human brain tumour initiating cells. Nature. 2004;432(7015):396-401.
43. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature. 1994;367(6464):645-8.
44. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med. 1997;3(7):730-7.
45. Eramo A, Lotti F, Sette G, Pilozzi E, Biffoni M, Di Virgilio A, et al. Identification and expansion of the tumorigenic lung cancer stem cell population. Cell Death Differ. 2008;15(3):504-14.
46. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015;347(6217):78-81.
47. Visvader JE. Cells of origin in cancer. Nature. 2011;469(7330):314-22.
48. Kim H, Park C, Han KH, Choi J, Kim YB, Kim JK, et al. Primary liver carcinoma of intermediate (hepatocyte-cholangiocyte) phenotype. J Hepatol. 2004;40(2):298-304.
49. Zhang F, Chen XP, Zhang W, Dong HH, Xiang S, Zhang WG, et al. Combined hepatocellular cholangiocarcinoma originating from hepatic progenitor cells: immunohistochemical and double-fluorescence immunostaining evidence. Histopathology. 2007;52(2):224-32.
50. De Sousa EMF, Vermeulen L, Fessler E, Medema JP. Cancer heterogeneity--a multifaceted view. EMBO Rep. 2013;14(8):686-95.
51. Schwitalla S, Fingerle Alexander A, Cammareri P, Nebelsiek T, Göktuna Serkan I, Ziegler Paul K, et al. Intestinal Tumorigenesis Initiated by Dedifferentiation and Acquisition of Stem-Cell-like Properties. Cell. 2013;152(1-2):25-38.
52. Visvader JE, Lindeman GJ. Cancer stem cells: current status and evolving complexities. Cell Stem Cell. 2012;10(6):717-28.
53. Roesch A, Fukunaga-Kalabis M, Schmidt EC, Zabierowski SE, Brafford PA, Vultur A, et al. A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. Cell. 2010;141(4):583-94.
54. Nguyen LV, Vanner R, Dirks P, Eaves CJ. Cancer stem cells: an evolving concept. Nature Reviews Cancer. 2012.
55. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. Cell. 2010;141(1):69-80.
56. Quintana E, Shackleton M, Foster HR, Fullen DR, Sabel MS, Johnson TM, et al. Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized. Cancer Cell. 2010;18(5):510-23.
57. Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. Mol Cancer. 2017;16(1):41.
58. Marjanovic ND, Weinberg RA, Chaffer CL. Cell Plasticity and Heterogeneity in Cancer. Clinical Chemistry. 2012.
59. Magee JA, Piskounova E, Morrison SJ. Cancer Stem Cells: Impact, Heterogeneity, and Uncertainty. Cancer Cell. 2012;21(3):283-96.

60. Altschuler SJ, Wu LF. Cellular Heterogeneity: Do Differences Make a Difference? Cell. 2010;141(4):559-63.
61. Sanz-Moreno V, Gadea G, Ahn J, Paterson H, Marra P, Pinner S, et al. Rac activation and inactivation control plasticity of tumor cell movement. Cell. 2008;135(3):510-23.
62. Kreso A, Dick John E. Evolution of the Cancer Stem Cell Model. Cell Stem Cell. 2014;14(3):275-91.
63. Sprouffske K, Athena Aktipis C, Radich JP, Carroll M, Nedelcu AM, Maley CC. An evolutionary explanation for the presence of cancer nonstem cells in neoplasms. Evol Appl. 2013;6(1):92-101.
64. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19(11):1423-37.
65. Pistollato F, Abbadi S, Rampazzo E, Persano L, Della Puppa A, Frasson C, et al. Intratumoral hypoxic gradient drives stem cells distribution and MGMT expression in glioblastoma. Stem Cells. 2010;28(5):851-62.
66. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell. 2010;140(6):883-99.
67. Chouaib S, Messai Y, Couve S, Escudier B, Hasmim M, Noman MZ. Hypoxia promotes tumor growth in linking angiogenesis to immune escape. Front Immunol. 2012;3:21.
68. Casciello F, Al-Ejeh F, Kelly G, Brennan DJ, Ngiow SF, Young A, et al. G9a drives hypoxia-mediated gene repression for breast cancer cell survival and tumorigenesis. Proc Natl Acad Sci U S A. 2017;114(27):7077-82.
69. Fogel O, Richard-Miceli C, Tost J. Epigenetic Changes in Chronic Inflammatory Diseases. Adv Protein Chem Struct Biol. 2017;106:139-89.
70. Kaukonen R, Mai A, Georgiadou M, Saari M, De Franceschi N, Betz T, et al. Normal stroma suppresses cancer cell proliferation via mechanosensitive regulation of JMJD1a-mediated transcription. Nat Commun. 2016;7:12237.
71. Feinberg AP. The Key Role of Epigenetics in Human Disease Prevention and Mitigation. N Engl J Med. 2018;378(14):1323-34.
72. Teschendorff AE, West J, Beck S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? Hum Mol Genet. 2013;22(R1):R7-R15.
73. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet. 2002;3(6):415-28.
74. Jones PA, Laird PW. Cancer epigenetics comes of age. Nature Genetics. 1999;21(2):163-7.
75. Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. New England Journal of Medicine. 2003;349(21):2042-54.
76. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89-92.
77. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. Science. 2003;300(5618):455.
78. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. Nature Reviews Cancer. 2013;13(7):497-510.
79. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, et al. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. Cancer Cell. 2014;26(6):813-25.
80. Shen H, Laird PW. Interplay between the Cancer Genome and Epigenome. Cell. 2013;153(1):38-55.

81. Timp W, Bravo HC, McDonald OG, Goggins M, Umbricht C, Zeiger M, et al. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. Genome Med. 2014;6(8):61.
82. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet. 2011;43(8):768-75.
83. Teschendorff AE, Widschwendter M. Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. Bioinformatics. 2012;28(11):1487-94.
84. Teschendorff AE, Jones A, Fiegl H, Sargent A, Zhuang JJ, Kitchener HC, et al. Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. Genome Medicine. 2012;4(3).
85. Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R, et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat Genet. 2012;44(11):1207-14.
86. Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, et al. Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol Cell. 2014;55(2):319-31.
87. Feinberg AP, Irizarry RA. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proceedings of the National Academy of Sciences of the United States of America. 2010;107 Suppl 1:1757-64.
88. Feinberg AP. Epigenomics reveals a functional genome anatomy and a new approach to common disease. Nature biotechnology. 2010;28(10).
89. Brock A, Chang H, Huang S. Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. Nature reviews Genetics. 2009;10(5):336-42.
90. Feinberg AP, Koldobskiy MA, Göndör A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. Nat Rev Genet. 2016;17(5):284-99.
91. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. Nature Reviews Genetics. 2006;7(1):21-33.
92. Pujadas E, Feinberg Andrew P. Regulated Noise in the Epigenetic Landscape of Development and Disease. Cell. 2012;148(6):1123-31.
93. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. Nature. 2010;467(7312):167-73.
94. Raj A, van Oudenaarden A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell. 2008;135(2):216-26.
95. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biology. 2006;4(10).
96. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature. 2008;453(7194):544-7.
97. Raser JM. Control of Stochasticity in Eukaryotic Gene Expression. Science. 2004;304(5678):1811-4.
98. Elowitz MB. Stochastic Gene Expression in a Single Cell. Science. 2002;297(5584):1183-6.
99. Kærn M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nature Reviews Genetics. 2005;6(6):451-64.
100. Graf T, Stadtfeld M. Heterogeneity of Embryonic and Adult Stem Cells. Cell Stem Cell. 2008;3(5):480-3.

101. Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, Garcia-Ojalvo J, et al. Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. PLoS Biology. 2009;7(7).
102. Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M. Open chromatin in pluripotency and reprogramming. Nat Rev Mol Cell Biol. 2011;12(1):36-47.
103. Hinohara K, Wu HJ, Vigneau S, McDonald TO, Igarashi KJ, Yamamoto KN, et al. KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance. Cancer Cell. 2018;34(6):939-53.e9.
104. Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev Cell. 2006;10(1):105-16.
105. Roesch A, Vultur A, Bogeski I, Wang H, Zimmermann KM, Speicher D, et al. Overcoming intrinsic multidrug resistance in melanoma by blocking the mitochondrial respiratory chain of slow-cycling JARID1B(high) cells. Cancer Cell. 2013;23(6):811-25.
106. Miller-Jensen K, Dey SS, Schaffer DV, Arkin AP. Varying virulence: epigenetic control of expression noise and disease processes. Trends Biotechnol. 2011;29(10):517-25.
107. Nguyen A, Yoshida M, Goodarzi H, Tavazoie SF. Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. Nat Commun. 2016;7:11246.
108. Hinohara K, Polyak K. Intratumoral Heterogeneity: More Than Just Mutations. Trends Cell Biol. 2019.
109. Niepel M, Spencer SL, Sorger PK. Non-genetic cell-to-cell variability and the consequences for pharmacology. Current Opinion in Chemical Biology. 2009;13(5-6):556-61.
110. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature. 2009;459(7245):428-32.
111. Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, Issaeva I, Sigal A, et al. Dynamic proteomics of individual cancer cells in response to a drug. Science. 2008;322(5907):1511-6.
112. Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, et al. Oscillations and variability in the p53 system. Mol Syst Biol. 2006;2:2006.0033.
113. Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nat Rev Clin Oncol. 2017;14(10):611-29.
114. Thiery JP. Epithelial–mesenchymal transitions in tumour progression. Nature Reviews Cancer. 2002;2(6):442-54.
115. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial-Mesenchymal Transitions in Development and Disease. Cell. 2009;139(5):871-90.
116. Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, et al. The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. Cell. 2008;133(4):704-15.
117. Morel A-P, Lièvre M, Thomas C, Hinkal G, Ansieau S, Puisieux A. Generation of Breast Cancer Stem Cells through Epithelial-Mesenchymal Transition. PLoS ONE. 2008;3(8).
118. Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016;529(7586):298-306.
119. Chaffer CL, Brennan JP, Slavin JL, Blick T, Thompson EW, Williams ED. Mesenchymal-to-Epithelial Transition Facilitates Bladder Cancer Metastasis: Role of Fibroblast Growth Factor Receptor-2. Cancer Research. 2006;66(23):11271-8.
120. Chaffer CL, San Juan BP, Lim E, Weinberg RA. EMT, cell plasticity and metastasis. Cancer Metastasis Rev. 2016;35(4):645-54.

121. Christiansen JJ, Rajasekaran AK. Reassessing epithelial to mesenchymal transition as a prerequisite for carcinoma invasion and metastasis. Cancer Res. 2006;66(17):8319-26.
122. Gotzmann J, Mikula M, Eger A, Schulte-Hermann R, Foisner R, Beug H, et al. Molecular aspects of epithelial cell plasticity: implications for local tumor invasion and metastasis. Mutat Res. 2004;566(1):9-20.
123. Jolly MK, Boareto M, Huang B, Jia D, Lu M, Ben-Jacob E, et al. Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. Front Oncol. 2015;5:155.
124. Jolly MK, Tripathi SC, Jia D, Mooney SM, Celiktas M, Hanash SM, et al. Stability of the hybrid epithelial/mesenchymal phenotype. Oncotarget. 2016;7(19):27067-84.
125. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. Cell. 2016;166(1):21-45.
126. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature. 2018;556(7702):463-8.
127. Chang HH, Oh PY, Ingber DE, Huang S. Multistable and multistep dynamics in neutrophil differentiation. BMC Cell Biol. 2006;7:11.
128. Latil M, Nassar D, Beck B, Boumahdi S, Wang L, Brisebarre A, et al. Cell-Type-Specific Chromatin States Differentially Prime Squamous Cell Carcinoma Tumor-Initiating Cells for Epithelial to Mesenchymal Transition. Cell Stem Cell. 2017;20(2):191-204.e5.
129. Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong ST, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature. 2015;527(7579):472-6.
130. Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature. 2015;527(7579):525-30.
131. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. Nat Rev Cancer. 2018;18(2):128-34.
132. Stankic M, Pavlovic S, Chin Y, Brogi E, Padua D, Norton L, et al. TGF-β-Id1 signaling opposes Twist1 and promotes metastatic colonization via a mesenchymal-to-epithelial transition. Cell Rep. 2013;5(5):1228-42.
133. van Denderen BJ, Thompson EW. Cancer: The to and fro of tumour spread. Nature. 2013;493(7433):487-8.
134. Pece S, Tosoni D, Confalonieri S, Mazzarol G, Vecchi M, Ronzoni S, et al. Biological and Molecular Heterogeneity of Breast Cancers Correlates with Their Cancer Stem Cell Content. Cell. 2010;140(1):62-73.
135. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An embryonic stem cell–like gene expression signature in poorly differentiated aggressive human tumors. Nature Genetics. 2008;40(5):499-507.
136. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nature reviews Cancer. 2009;9(4):265-73.
137. Brabletz T, Jung A, Spaderna S, Hlubek F, Kirchner T. Opinion: Migrating cancer stem cells — an integrated concept of malignant tumour progression. Nature Reviews Cancer. 2005;5(9):744-9.
138. Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proceedings of the National Academy of Sciences. 2011;108(19):7950-5.
139. Gupta PB, Chaffer CL, Weinberg RA. Cancer stem cells: mirage or reality? Nat Med. 2009;15(9):1010-2.

140. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
141. Acloque H, Adams MS, Fishwick K, Bronner-Fraser M, Nieto MA. Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. Journal of Clinical Investigation. 2009;119(6):1438-49.
142. Singh A, Settleman J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. Oncogene. 2010;29(34):4741-51.
143. Rubin H. Cellular epigenetics: effects of passage history on competence of cells for "spontaneous" transformation. Proc Natl Acad Sci U S A. 1993;90(22):10715-9.
144. Lavrovsky VA, Guvakova MA, Lavrovsky YV. High frequency of tumour cell reversion to non-tumorigenic phenotype. Eur J Cancer. 1992;28(1):17-21.
145. Zhou JX, Pisco AO, Qian H, Huang S. Nonequilibrium population dynamics of phenotype conversion of cancer cells. PLoS One. 2014;9(12):e110714.
146. Huang S. Tumor progression: chance and necessity in Darwinian and Lamarckian somatic (mutationless) evolution. Prog Biophys Mol Biol. 2012;110(1):69-86.
147. Huang S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. Cancer Metastasis Rev. 2013;32(3-4):423-48.
148. Blake WJ, KÆrn M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003;422(6932):633-7.
149. Paldi A. Stochastic gene expression during cell differentiation: order from disorder? Cellular and Molecular Life Sciences (CMLS). 2003;60(9):1775-8.
150. Ramirez M, Rajaram S, Steininger RJ, Osipchuk D, Roth MA, Morinishi LS, et al. Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. Nat Commun. 2016;7:10690.
151. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet. 2005;6(6):451-64.
152. Brumby AM, Richardson HE. Using Drosophila melanogaster to map human cancer pathways. Nat Rev Cancer. 2005;5(8):626-39.
153. Ntziachristos P, Lim JS, Sage J, Aifantis I. From fly wings to targeted cancer therapies: a centennial for notch signaling. Cancer Cell. 2014;25(3):318-34.
154. Jindal GA, Goyal Y, Burdine RD, Rauen KA, Shvartsman SY. RASopathies: unraveling mechanisms with animal models. Dis Model Mech. 8. England2015. p. 1167.
155. Pagliarini RA, Xu T. A genetic screen in Drosophila for metastatic behavior. Science. 2003;302(5648):1227-31.
156. Brumby AM, Richardson HE. scribble mutants cooperate with oncogenic Ras or Notch to cause neoplastic overgrowth in Drosophila. Embo j. 2003;22(21):5769-79.
157. Uhlirova M, Bohmann D. JNK- and Fos-regulated Mmp1 expression cooperates with Ras to induce invasive tumors in Drosophila. Embo j. 2006;25(22):5294-304.
158. Kang Y, Pantel K. Tumor cell dissemination: emerging biological insights from animal models and cancer patients. Cancer Cell. 2013;23(5):573-81.
159. Miles WO, Dyson NJ, Walker JA. Modeling tumor invasion and metastasis in Drosophila. Dis Model Mech. 2011;4(6):753-61.
160. Reiter LT, Potocki L, Chien S, Gribskov M, Bier E. A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. Genome Res. 2001;11(6):1114-25.
161. Zoranovic T, Manent J, Willoughby L, Matos de Simoes R, La Marca JE, Golenkina S, et al. A genome-wide Drosophila epithelial tumorigenesis screen identifies Tetraspanin 29Fb as an evolutionarily conserved suppressor of Ras-driven cancer. PLoS Genet. 2018;14(10):e1007688.

162. Young A, Lyons J, Miller AL, Phan VT, Alarcon IR, McCormick F. Ras signaling and therapies. Adv Cancer Res. 2009;102:1-17.
163. Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. RAS oncogenes: weaving a tumorigenic web. Nat Rev Cancer. 2011;11(11):761-74.
164. Fernandez-Medarde A, Santos E. Ras in cancer and developmental diseases. Genes Cancer. 2011;2(3):344-58.
165. Dimauro T, David G. Ras-induced senescence and its physiological relevance in cancer. Curr Cancer Drug Targets. 2010;10(8):869-76.
166. McGuire SE, Roman G, Davis RL. Gene expression systems in Drosophila: a synthesis of time and space. Trends Genet. 2004;20(8):384-91.
167. Rawlings ND, Barrett AJ. Families of cysteine peptidases. Methods Enzymol. 1994;244:461-86.
168. Wex T, Levy B, Wex H, Brömme D. Human cathepsins F and W: A new subgroup of cathepsins. Biochemical and biophysical research communications. 1999;259(2):401-7.
169. Newsham IF. The long and short of chromosome 11 in breast cancer. Am J Pathol. 1998;153(1):5-9.
170. Santamaría I, Velasco G, Pendás AM, Paz A, López-Otín C. Molecular cloning and structural and functional characterization of human cathepsin F, a new cysteine proteinase of the papain family with a long propeptide domain. J Biol Chem. 1999;274(20):13800-9.
171. Oorni K. Cysteine Protease Cathepsin F Is Expressed in Human Atherosclerotic Lesions, Is Secreted by Cultured Macrophages, and Modifies Low Density Lipoprotein Particles in Vitro. Journal of Biological Chemistry. 2004;279(33):34776-84.
172. Shi G-P, Bryant RAR, Riese R, Verhelst S, Driessen C, Li Z, et al. Role for cathepsin F in invariant chain processing and major histocompatibility complex class II peptide loading by macrophages. The Journal of experimental medicine. 2000;191(7):1177-86.
173. Lindstedt L, Lee M, Öörni K, Brömme D, Kovanen PT. Cathepsins F and S block HDL3-induced cholesterol efflux from macrophage foam cells. Biochemical and Biophysical Research Communications. 2003;312(4):1019-24.
174. Kaakinen R, Lindstedt KA, Sneck M, Kovanen PT, Öörni K. Angiotensin II increases expression and secretion of cathepsin F in cultured human monocyte-derived macrophages: An angiotensin II type 2 receptor-mediated effect. Atherosclerosis. 2007;192(2):323-7.
175. Saghizadeh M, Epifantseva I, Hemmati D, Ghiam CA, Brunken WJ, Ljubimov AV. Enhanced wound healing, kinase and stem cell marker expression in diabetic organ-cultured human corneas upon MMP-10 and cathepsin F gene silencing. Investigative ophthalmology & visual science. 2013.
176. Tang CH, Lee JW, Galvez MG, Robillard L, Mole SE, Chapman HA. Murine Cathepsin F Deficiency Causes Neuronal Lipofuscinosis and Late-Onset Neurological Disease. Molecular and Cellular Biology. 2006;26(6):2309-16.
177. Smith KR, Dahl HH, Canafoglia L, Andermann E, Damiano J, Morbin M, et al. Cathepsin F mutations cause Type B Kufs disease, an adult-onset neuronal ceroid lipofuscinosis. Hum Mol Genet. 2013;22(7):1417-23.
178. Maubach G, Lim MC, Zhuo L. Nuclear cathepsin F regulates activation markers in rat hepatic stellate cells. Mol Biol Cell. 2008;19(10):4238-48.
179. Yang L, Wang J, Li J, Zhang H, Guo S, Yan M, et al. Identification of Serum Biomarkers for Gastric Cancer Diagnosis Using a Human Proteome Microarray. Mol Cell Proteomics. 2016;15(2):614-23.

180. Ji C, Zhao Y, Kou YW, Shao H, Guo L, Bao CH, et al. Cathepsin F Knockdown Induces Proliferation and Inhibits Apoptosis in Gastric Cancer Cells. Oncol Res. 2018;26(1):83-93.
181. Wang B. Human Cathepsin F. MOLECULAR CLONING, FUNCTIONAL EXPRESSION, TISSUE LOCALIZATION, AND ENZYMATIC CHARACTERIZATION. Journal of Biological Chemistry. 1998;273(48):32000-8.
182. Vazquez-Ortiz G, Pina-Sanchez P, Vazquez K, Duenas A, Taja L, Mendoza P, et al. Overexpression of cathepsin F, matrix metalloproteinases 11 and 12 in cervical cancer. BMC cancer. 2005;5(1).
183. Oliveras-Ferraros C, Vazquez-Martin A, Cuyàs E, Corominas-Faja B, Rodríguez-Gallego E, Fernández-Arroyo S, et al. Acquired resistance to metformin in breast cancer cells triggers transcriptome reprogramming toward a degradome-related metastatic stem-like profile. Cell Cycle. 2014;13(7):1132-44.
184. Husmann K, Muff R, Bolander ME, Sarkar G, Born W, Fuchs B. Cathepsins and osteosarcoma: Expression analysis identifies cathepsin K as an indicator of metastasis. Molecular Carcinogenesis. 2008;47(1):66-73.
185. Di Rosa M, Sanfilippo C, Libra M, Musumeci G, Malaguarnera L. Different pediatric brain tumors are associated with different gene expression profiling. Acta Histochem. 2015;117(4-5):477-85.
186. Bieging KT, Attardi LD. Deconstructing p53 transcriptional networks in tumor suppression. Trends in Cell Biology. 2012;22(2):97-106.
187. Schwarze SR. Novel Pathways Associated with Bypassing Cellular Senescence in Human Prostate Epithelial Cells. Journal of Biological Chemistry. 2002;277(17):14877-83.
188. Xu M, Yu Q, Subrahmanyam R, Difilippantonio MJ, Ried T, Sen JM. Beta-catenin expression results in p53-independent DNA damage and oncogene-induced senescence in prelymphomagenic thymocytes in vivo. Mol Cell Biol. 2008;28(5):1713-23.
189. Tu HC, Ren D, Wang GX, Chen DY, Westergard TD, Kim H, et al. The p53-cathepsin axis cooperates with ROS to activate programmed necrotic death upon DNA damage. Proc Natl Acad Sci U S A. 2009;106(4):1093-8.
190. Brady CA, Jiang D, Mello SS, Johnson TM, Jarvis LA, Kozak MM, et al. Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor suppression. Cell. 2011;145(4):571-83.
191. Janic A, Valente LJ, Wakefield MJ, Di Stefano L, Milla L, Wilcox S, et al. DNA repair processes are critical mediators of p53-dependent tumor suppression. Nat Med. 2018.
192. Miranda TB, Cortez CC, Yoo CB, Liang G, Abe M, Kelly TK, et al. DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. Mol Cancer Ther. 2009;8(6):1579-88.
193. Ibragimova I, Ibanez de Caceres I, Hoffman AM, Potapova A, Dulaimi E, Al-Saleem T, et al. Global Reactivation of Epigenetically Silenced Genes in Prostate Cancer. Cancer Prevention Research. 2010;3(9):1084-92.
194. Matak A, Lahiri P, Ford E, Pabst D, Kashofer K, Stellas D, et al. Stochastic phenotype switching leads to intratumor heterogeneity in human liver cancer. Hepatology. 2018;68(3):933-48.
195. Liang CC, Park AY, Guan JL. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. Nat Protoc. 2007;2(2):329-33.
196. Zufferey R, Nagy D, Mandel RJ, Naldini L, Trono D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nat Biotechnol. 1997;15(9):871-5.

197. Kashofer K, Viertler C, Pichler M, Zatloukal K. Quality control of RNA preservation and extraction from paraffin-embedded tissue: implications for RT-PCR and microarray analysis. PloS One. 2013;8(7).
198. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods. 2001;25(4):402-8.
199. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome biology. 2007;8(1):R3.
200. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44-57.
201. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
202. Shen L, Shao NY, Liu X, Maze I, Feng J, Nestler EJ. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. PLoS One. 2013;8(6):e65598.
203. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010;11(10):R106.
204. Lao XM, Chen DY, Zhang YQ, Xiang J, Guo RP, Lin XJ, et al. Primary carcinosarcoma of the liver: clinicopathologic features of 5 cases and a review of the literature. The American journal of surgical pathology. 2007;31(6):817-26.
205. Malhotra S, Wood J, Mansy T, Singh R, Zaitoun A, Madhusudan S. Intrahepatic sarcomatoid cholangiocarcinoma. Journal of oncology. 2010;2010:701476.
206. Wang QB, Cui BK, Weng JM, Wu QL, Qiu JL, Lin XJ. Clinicopathological characteristics and outcome of primary sarcomatoid carcinoma and carcinosarcoma of the liver. Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract. 2012;16(9):1715-26.
207. Thiery JP. Epithelial-mesenchymal transitions in tumour progression. Nat Rev Cancer. 2002;2(6):442-54.
208. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science. 2013;339(6119):580-4.
209. Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, et al. EMT and dissemination precede pancreatic tumor formation. Cell. 2012;148(1-2):349-61.
210. Maeda M, Johnson KR, Wheelock MJ. Cadherin switching: essential for behavioral but not morphological changes during an epithelium-to-mesenchyme transition. Journal of cell science. 2005;118(Pt 5):873-87.
211. Frixen UH, Behrens J, Sachs M, Eberle G, Voss B, Warda A, et al. E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. J Cell Biol. 1991;113(1):173-85.
212. Loaiza-Bonilla A, Clayton E, Furth E, O'Hara M, Morrissette J. Dramatic response to dabrafenib and trametinib combination in a BRAF V600E-mutated cholangiocarcinoma: implementation of a molecular tumour board and next-generation sequencing for personalized medicine. Ecancermedicalscience. 2014;8:479.
213. Huh I, Zeng J, Park T, Yi SV. DNA methylation and transcriptional noise. Epigenetics Chromatin. 2013;6(1):9.
214. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;19(1A):A68-77.
215. Leonhardt H, Cardoso MC. DNA methylation, nuclear structure, gene expression and cancer. J Cell Biochem Suppl. 2000;Suppl 35:78-83.

216. Huh D, Paulsson J. Non-genetic heterogeneity from stochastic partitioning at cell division. Nature Genetics. 2010;43(2):95-100.
217. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proceedings of the National Academy of Sciences. 2010;107(35):15449-54.
218. Choi Y, Lee HJ, Jang MH, Gwak JM, Lee KS, Kim EJ, et al. Epithelial-mesenchymal transition increases during the progression of in situ to invasive basal-like breast cancer. Hum Pathol. 2013;44(11):2581-9.
219. Zeng GF, Cai SX, Wu GJ. Up-regulation of METCAM/MUC18 promotes motility, invasion, and tumorigenesis of human breast cancer cells. BMC Cancer. 2011;11:113.
220. Jiang G, Zhang L, Zhu Q, Bai D, Zhang C, Wang X. CD146 promotes metastasis and predicts poor prognosis of hepatocellular carcinoma. J Exp Clin Cancer Res. 2016;35:38.
221. Liu TA, Jan YJ, Ko BS, Wu YJ, Lu YJ, Liang SM, et al. Regulation of aldo-keto-reductase family 1 B10 by 14-3-3ε and their prognostic impact of hepatocellular carcinoma. Oncotarget. 2015;6(36):38967-82.
222. Ferrand N, Gnanapragasam A, Dorothee G, Redeuilh G, Larsen AK, Sabbah M. Loss of WISP2/CCN5 in estrogen-dependent MCF7 human breast cancer cells promotes a stem-like cell phenotype. PLoS One. 2014;9(2):e87878.
223. Giannelli G, Koudelkova P, Dituri F, Mikulits W. Role of epithelial to mesenchymal transition in hepatocellular carcinoma. J Hepatol. 2016;65(4):798-808.
224. Guo W, Keckesova Z, Donaher JL, Shibue T, Tischler V, Reinhardt F, et al. Slug and Sox9 cooperatively determine the mammary stem cell state. Cell. 2012;148(5):1015-28.
225. Tsai Jeff H, Donaher Joana L, Murphy Danielle A, Chau S, Yang J. Spatiotemporal Regulation of Epithelial-Mesenchymal Transition Is Essential for Squamous Cell Carcinoma Metastasis. Cancer Cell. 2012.
226. Choi Y, Lee HJ, Jang MH, Gwak JM, Lee KS, Kim EJ, et al. Epithelial-mesenchymal transition increases during the progression of in situ to invasive basal-like breast cancer. Human Pathology. 2013;44(11):2581-9.
227. Friedl P, Alexander S. Cancer invasion and the microenvironment: plasticity and reciprocity. Cell. 2011;147(5):992-1009.
228. Christiansen JJ. Reassessing Epithelial to Mesenchymal Transition as a Prerequisite for Carcinoma Invasion and Metastasis. Cancer Research. 2006;66(17):8319-26.
229. Gavert N, Ben-Ze'ev A. Epithelial–mesenchymal transition and the invasive potential of tumors. Trends in Molecular Medicine. 2008;14(5):199-209.
230. Wicki A, Lehembre F, Wick N, Hantusch B, Kerjaschki D, Christofori G. Tumor invasion in the absence of epithelial-mesenchymal transition: podoplanin-mediated remodeling of the actin cytoskeleton. Cancer Cell. 2006;9(4):261-72.
231. Kreso A, O'Brien CA, van Galen P, Gan OI, Notta F, Brown AM, et al. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. Science. 2013;339(6119):543-8.
232. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. Science. 2017;357(6348).
233. Niwa T, Ushijima T. Induction of epigenetic alterations by chronic inflammation and its significance on carcinogenesis. Advances in genetics. 2010;71:41-56.
234. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Developmental cell. 2010;19(5):698-711.

235. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. Nature. 2007;447(7143):433-40.
236. Dumont N, Wilson MB, Crawford YG, Reynolds PA, Sigaroudinia M, Tlsty TD. Sustained induction of epithelial to mesenchymal transition activates DNA methylation of genes silenced in basal-like breast cancers. Proceedings of the National Academy of Sciences. 2008;105(39):14867-72.
237. McDonald OG, Wu H, Timp W, Doi A, Feinberg AP. Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. Nature Structural & Molecular Biology. 2011;18(8):867-74.
238. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. Nature Genetics. 2011;43(8):768-75.
239. Scheel C, Eaton Elinor N, Li Sophia H-J, Chaffer Christine L, Reinhardt F, Kah K-J, et al. Paracrine and Autocrine Signals Induce and Maintain Mesenchymal and Stem Cell States in the Breast. Cell. 2011;145(6):926-40.
240. Chaffer CL, Thompson EW, Williams ED. Mesenchymal to Epithelial Transition in Development and Disease. Cells Tissues Organs. 2007;185(1-3):7-19.
241. Iliopoulos D, Hirsch HA, Wang G, Struhl K. Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. Proceedings of the National Academy of Sciences. 2011;108(4):1397-402.
242. Busby MA, Stewart C, Miller CA, Grzeda KR, Marth GT. Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression. Bioinformatics. 2013;29(5):656-7.
243. Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. Genome Biology. 2010;11(12).
244. Almendro V, Marusyk A, Polyak K. Cellular heterogeneity and molecular evolution in cancer. Annual review of pathology. 2013;8:277-302.
245. Buganim Y, Faddah Dina A, Cheng Albert W, Itskovich E, Markoulaki S, Ganz K, et al. Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase. Cell. 2012;150(6):1209-22.
246. Doi A, Park I-H, Wen B, Murakami P, Aryee MJ, Irizarry R, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nature Genetics. 2009;41(12):1350-3.
247. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. Nat Rev Cancer. 2011;11(10):726-34.
248. Kelly TK, De Carvalho DD, Jones PA. Epigenetic modifications as therapeutic targets. Nat Biotechnol. 2010;28(10):1069-78.
249. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. Nature. 2004;429(6990):457-63.
250. Issa JP, Kantarjian H. Azacitidine. Nat Rev Drug Discov. 2005;Suppl:S6-7.
251. Silverman LR, Mufti GJ. Methylation inhibitor therapy in the treatment of myelodysplastic syndrome. Nat Clin Pract Oncol. 2005;2 Suppl 1:S12-23.
252. Fenaux P, Mufti GJ, Hellstrom-Lindberg E, Santini V, Finelli C, Giagounidis A, et al. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. Lancet Oncol. 2009;10(3):223-32.
253. Choi SC, Yoon J, Shim WJ, Ro YM, Lim DS. 5-azacytidine induces cardiac differentiation of P19 embryonic stem cells. Exp Mol Med. 2004;36(6):515-23.

254. Zhou GS, Zhang XL, Wu JP, Zhang RP, Xiang LX, Dai LC, et al. 5-Azacytidine facilitates osteogenic gene expression and differentiation of mesenchymal stem cells by alteration in DNA methylation. Cytotechnology. 2009;60(1-3):11.
255. Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nature Genetics. 2006;38(4):468-73.
256. Zhuang J, Jones A, Lee S-H, Ng E, Fiegl H, Zikan M, et al. The Dynamics and Prognostic Potential of DNA Methylation Changes at Stem Cell Gene Loci in Women's Cancer. PLoS Genetics. 2012;8(2).
257. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer. 2003;3(4):253-66.
258. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med. 2005;352(10):997-1003.
259. Esteller M, Garcia-Foncillas J, Andion E, Goodman SN, Hidalgo OF, Vanaclocha V, et al. Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. N Engl J Med. 2000;343(19):1350-4.
260. Raser JM. Noise in Gene Expression: Origins, Consequences, and Control. Science. 2005;309(5743):2010-3.
261. Marshall OJ. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. Bioinformatics. 2004;20(15):2471-2.
262. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46(D1):D754-D61.
263. Klein M, Vignaud JM, Hennequin V, Toussaint B, Bresler L, Plénat F, et al. Increased expression of the vascular endothelial growth factor is a pejorative prognosis marker in papillary thyroid carcinoma. J Clin Endocrinol Metab. 2001;86(2):656-8.
264. Kitamura K, Hatano E, Higashi T, Narita M, Seo S, Nakamoto Y, et al. Proliferative activity in hepatocellular carcinoma is closely correlated with glucose metabolism but not angiogenesis. J Hepatol. 2011;55(4):846-57.
265. Ingolf JB, Russalina M, Simona M, Julia R, Gilda S, Bohle RM, et al. Can ki-67 play a role in prediction of breast cancer patients' response to neoadjuvant chemotherapy? Biomed Res Int. 2014;2014:628217.
266. Karnoub AE, Weinberg RA. Ras oncogenes: split personalities. Nat Rev Mol Cell Biol. 2008;9(7):517-31.
267. Nguyen AT, Emelyanov A, Koh CH, Spitsbergen JM, Lam SH, Mathavan S, et al. A high level of liver-specific expression of oncogenic Kras(V12) drives robust liver tumorigenesis in transgenic zebrafish. Dis Model Mech. 2011;4(6):801-13.
268. Deussing J, Tisljar K, Papazoglou A, Peters C. Mouse cathepsin F: cDNA cloning, genomic organization and chromosomal assignment of the gene. Gene. 2000;251(2):165-73.
269. Jerič B, Dolenc I, Mihelič M, Klarić M, Zavašnik-Bergant T, Gunčar G, et al. N-terminally truncated forms of human cathepsin F accumulate in aggresome-like inclusions. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2013;1833(10):2254-66.
270. Hoare M, Das T, Alexander G. Ageing, telomeres, senescence, and liver injury. J Hepatol. 2010;53(5):950-61.
271. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.

272. Aigelsreiter A, Janig E, Sostaric J, Pichler M, Unterthor D, Halasz J, et al. Clusterin expression in cholestasis, hepatocellular carcinoma and liver fibrosis. Histopathology. 2009;54(5):561-70.
273. Fonovic M, Brömme D, Turk V, Turk B. Human cathepsin F: expression in baculovirus system, characterization and inhibition by protein inhibitors. Biol Chem. 2004;385(6):505-9.
274. Lopez F, Belloc F, Lacombe F, Dumain P, Reiffers J, Bernard P, et al. Modalities of synthesis of Ki67 antigen during the stimulation of lymphocytes. Cytometry. 1991;12(1):42-9.
275. Cattoretti G, Becker MH, Key G, Duchrow M, Schlüter C, Galle J, et al. Monoclonal antibodies against recombinant parts of the Ki-67 antigen (MIB 1 and MIB 3) detect proliferating cells in microwave-processed formalin-fixed paraffin sections. J Pathol. 1992;168(4):357-63.
276. Luo Y, Ren F, Liu Y, Shi Z, Tan Z, Xiong H, et al. Clinicopathological and prognostic significance of high Ki-67 labeling index in hepatocellular carcinoma patients: a meta-analysis. Int J Clin Exp Med. 2015;8(7):10235-47.
277. Vousden KH, Lane DP. p53 in health and disease. Nat Rev Mol Cell Biol. 2007;8(4):275-83.
278. Muller PA, Vousden KH. p53 mutations in cancer. Nat Cell Biol. 2013;15(1):2-8.
279. Woo HG, Wang XW, Budhu A, Kim YH, Kwon SM, Tang ZY, et al. Association of TP53 mutations with stem cell-like gene expression and survival of patients with hepatocellular carcinoma. Gastroenterology. 2011;140(3):1063-70.
280. Aigelsreiter AM, Aigelsreiter A, Wehrschuetz M, Ress AL, Koller K, Salzwimmer M, et al. Loss of the putative tumor suppressor protein spinophilin is associated with poor prognosis in head and neck cancer. Hum Pathol. 2014;45(4):683-90.
281. wheeler@bcm.edu CGARNEa, Network CGAR. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017;169(7):1327-41.e23.
282. Menyhárt O, Nagy Á, Győrffy B. Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma. R Soc Open Sci. 2018;5(12):181006.
283. Fan G, Tu Y, Chen C, Sun H, Wan C, Cai X. DNA methylation biomarkers for hepatocellular carcinoma. Cancer Cell Int. 2018;18:140.
284. Zhang X, He Y, Lee KH, Dubois W, Li Z, Wu X, et al. Rap2b, a novel p53 target, regulates p53-mediated pro-survival function. Cell Cycle. 2013;12(8):1279-91.
285. Jiang D, Brady CA, Johnson TM, Lee EY, Park EJ, Scott MP, et al. Full p53 transcriptional activation potential is dispensable for tumor suppression in diverse lineages. Proc Natl Acad Sci U S A. 2011;108(41):17123-8.
286. Liu J, Li W, Deng M, Liu D, Ma Q, Feng X. Immunohistochemical Determination of p53 Protein Overexpression for Predicting p53 Gene Mutations in Hepatocellular Carcinoma: A Meta-Analysis. PLoS One. 2016;11(7):e0159636.
287. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology. 2007;132(7):2557-76.
288. Reinheckel T. The lysosomal cysteine protease cathepsin L regulates keratinocyte proliferation by control of growth factor recycling. Journal of Cell Science. 2005;118(15):3387-95.
289. Villanueva A. Hepatocellular Carcinoma. N Engl J Med. 2019;380(15):1450-62.
290. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients.

6 Appendix

Part of this section have been published in (194), or are available at Ethan Ford, Ethanomics, 2015, < https://ethanomics.wordpress.com/>

6.1 Overall research design.

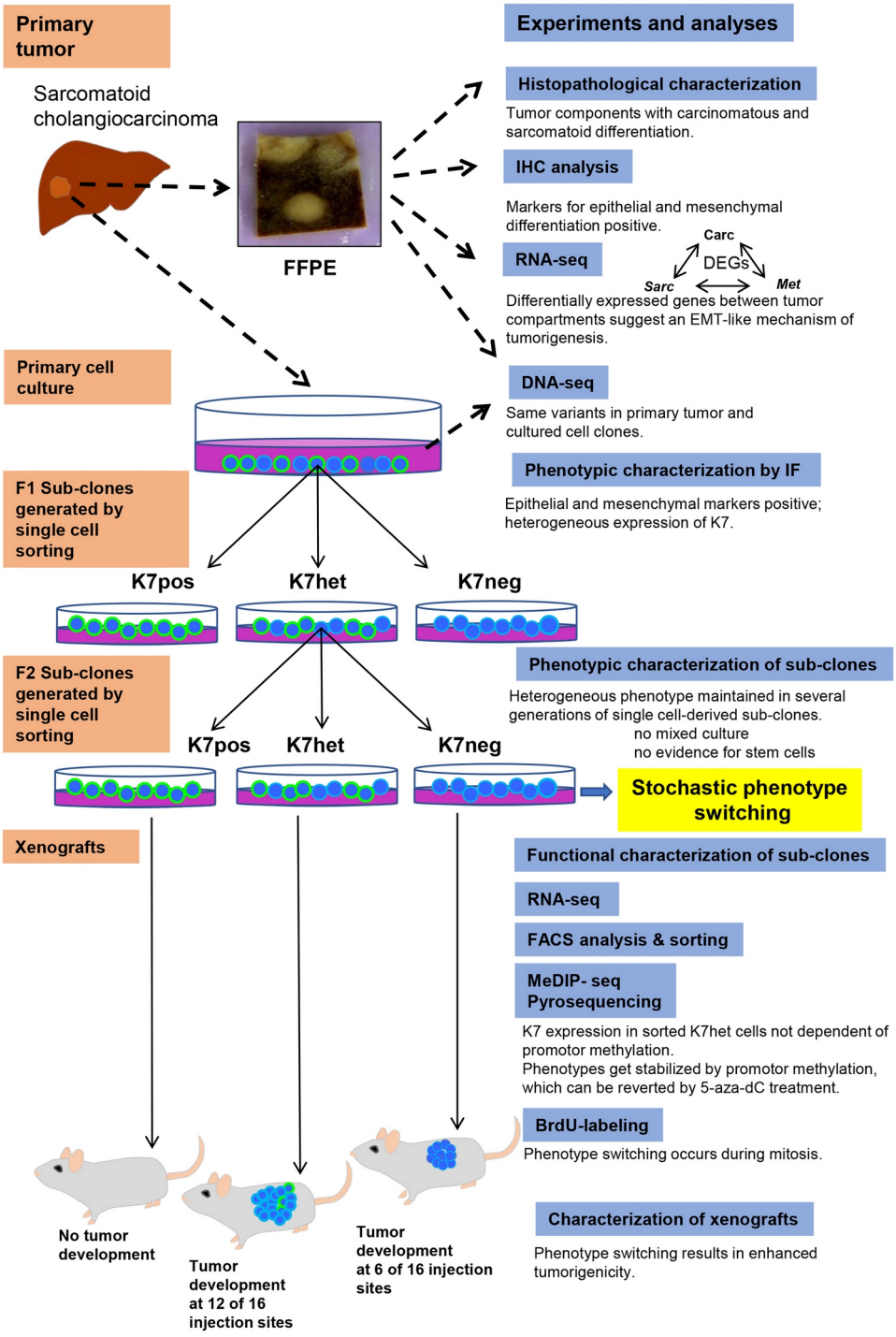


Figure 6-1. Overview of the phenotype characterization and functional studies of the derived subclones.

(Adapted from (194)).

6.2 Tables

Table 2-1. Primers used for qRT-PCR

Primer Name	Sequence	Size
CD73_cds_fwd	5' ACAGCAGCATTCCTGAAGATCCAAGC 3'	131bp
CD73_cds_rev	5' GCGGCATGATTGAGAGGAGCCATC 3'	
SEMA3E_cds_fwd	5' CCCCGGTTACGCCTGTCACATAAAG 3'	198bp
SEMA3E_cds_rev	5' CGGCCAGTGTATCTCTTTATAGCCG 3'	
MGST1_cds_fwd	5' CTGCATTGCGCGCGACCCG 3'	71bp
MGST1_cds_rev	5' CAATTTTGGTGGTGAGGCCGAGGAG 3'	
TRO_cds_fwd	5' TCAAGAAGACGAGGCCCATTCCCC 3'	152bp
TRO_cds_rev	5' ATATCTGGAGGGAAGGGAAGCCCC 3'	
LEF1_cds_fwd	5' GCTGCCTACATCTGAAACATGGTGG 3'	187bp
LEF1_cds_rev	5' GCGTCTCTAGCAGTGACCTCAGGG 3'	
SECTM1_cds_fwd	5' CAAGTCCCACCAGCCATGCAGACC 3'	173bp
SECTM1_cds_rev	5' ATGACGGTGTTCTCGCCCCAAGAC 3'	
SFRP1_cds_fwd	5' CCGGTCATGCAGTTCTTCGGCTTC 3'	151bp
SFRP1_cds_rev	5' CGTTGTCACAGGGAGGACACACCG 3'	
ADAM22_cds_fwd	5' ACACGTGGTCCGGGTGCATAATGG 3'	120bp
ADAM22_cds_rev	5' TGAAAAGGCAGGCACCACCTCCAC 3'	
ACN9_cds_fwd	5' TGGGCGACCAGTACGTGAAAGACG 3'	170bp
ACN9_cds_rev	5' TGGGAGGAAGGTGCCAAAACATGC 3'	
NRP1_cds_fwd	5' TCATGGATGACAGCAAACGCAAGGC 3'	167bp
NRP1_cds_rev	5' TTCACAGCCCAGCAGCTCCATTCTG 3'	
IL24_cds_fwd	5' ATTGAGGCTGCTTGGGAGGAAGGC 3'	117bp
IL24_cds_rev	5' CCAGCAAAGGAGGGCAGAAGGGTC 3'	

Table 2-2. Primers used for MeDIP-qPCR

Primers	Genomic location	Sequence 5'→3'	Tm ºC
SOCS3	hg19 chr17: 76354475F	CATGCCCTTTGCGCCCTTTACCC	72.7˚C
	hg19 chr17: 76354620R	TGGCCACTCTTCAGCATCTCTGTCG	71.0˚C
B3GALNT1	hg19 chr3: 160822804F	CTGCTCGTCACCGAGACCCGAGAC	72.5˚C
	hg19 chr3: 160822959R	GTATAGCCACCGGACCCGCCTTTC	71.1˚C
ADAM22	hg19 chr7: 87563370F	GGAGACAAGCGGGACACGGAGAG	70.6˚C
	hg19 chr7: 87563557R	GGGCGGTTGGCTGTAGCTAAGTGG	70.5˚C
MGMT	hg19 chr10: 131564511F	AGGCATGGCTGCGTTCCAGTAAAG	69.3˚C
	hg19 chr10: 131564682R	TGTGACGCTCTCAAACACTCCCCA	70.3˚C
CDH1	hg19 chr16: 68772240F	CTGGCTTTGACGCCGAGAGCTACAC	71.0˚C
	hg19 chr16: 68772428R	TTGAGCCAAGGAGGGAGCTTGCC	71.7˚C
STAT3	hg19 chr17: 40540170F	CATCTCTCCCGGCCCCACTG	69.6˚C
	hg19 chr17: 40540316R	GGCCTCTGCCGGAGAAACAGG	69.6˚C
PDGFRA	hg19 chr4: 55095141F	TCTTTCCCGGCAGAGCACCAA	69.4˚C
	hg19 chr4: 55095327R	TCTCTCAAACTCCCTGCGCTC	64.3˚C
DNMT3B	hg19 chr20: 31350997F	CCCTTGCAACTGGGATGTGGGGTC	73.3˚C

	hg19 chr20: 31351167R	**GATGGCTCAGCTCAGGAGCACCC**	**70.8˚C**
EGFR	**hg19 chr7: 55085465F**	**GCGGGCAGTGAGCAGACC**	**65.6˚C**
	hg19 chr7: 55085635R	**AGTAAGTGAAGCTACAGACCTGCCC**	**63.3˚C**

Table 2-3. List of antibodies used for immunohistochemistry, immunofluorescence and FACS analysis.

Antibodies used for Immunohistochemistry on FFPE primary tumor samples					
Antibody	**Cat #**	**Company**	**Isotype**	**Dilutions**	**Block**
K7	M7018	DAKO	Mouse / IgG1	1:200	block 2-7
K8+18	NCL-L-CK8-TS1/ MA5-12104	Novocastra/Thermo Fisher Scientific	Mouse / IgG1	1:200	block 4-8
K19	M0888	DAKO	Mouse monoclonal/ IgG1	1:200	block 4-8
VIM	61013	Progen	mouse monoclonal /IgG2a kappa	1:100	block 4-8
c-MET	ab51067	Abcam	Rabbit Monoclonal	1:200	block 1-8
Claudin 4	RB-9266-P0	Thermo Fisher Scientific	Rabbit Polyclonal	1:50	block 4-6, and 8
Muc1	NCL-MUC-1	Novocastra	Mouse monoclonal	1:100	block 1, 7 and 8
E-cadherin	610181	BD Transduction LaboratoriesTM	Mouse IgG2a	1:200	block 4-8
β-catenin	610153	BD Transduction LaboratoriesTM	Mouse IgG1	1:200	block 2-8
CEA	MS613-P0	Thermo Fisher Scientific	mouse monoclonal /IgG2a kappa	1:200	block 2-8
CD44	MS-668-P0	Thermo Fisher Scientific	mouse IgG2a	1:400	block 4-8
p53	M7001	DAKO	mouse monoclonal	1:100	block 2, 5, 7 and 8
Ki-67	M7240	DAKO	mouse	1:200	block

			monoclonal		2-8
NCAM1	NCL-L-CD56-1B6	Novocastra	mouse monoclonal IgG1	1:50	block 2-8
CgA	MS-382-P0	Thermo Fisher Scientific	mouse monoclonal/IgG1	1:200	block 1, 4-7
CD117	A4502	DAKO	rabbit polyclonal	1:200	block 3-7
CD34	MS-363-P0	Thermo Fisher Scientific	mouse monoclonal IgG1	1:200	block 4,5 and 7
Nestin	sc23927	Santa Cruz	mouse monoclonal	1:50	block 4-8
EpCam (ESA)	NCL-ESA	Novocastra	mouse monoclonal	1:100	block 4-8
Hep Par 1	M7158	DAKO	mouse monoclonal	1:50	block 4-7
AFP	A0008	DAKO	rabbit polyclonal	1:100	block 4-7
Zeb1	NBP1-05987	Thermo Fisher Scientific	Rabbit IgG	1:50	block 2,3
Muc18	ab75769	Abcam	Rabbit monoclonal	1:200	block 4,7 and 8
Antibodies used for staining cells grown on coverslips					
K7	M7018	DAKO	Mouse / IgG1	1:200	
K8	NCL-L-CK8-TS1/	Novocastra	Mouse / IgG1	1:200	
K18	MA5-12104	Thermo Fisher Scientific	Mouse / IgG1	1:200	
K19	M0888	DAKO	Mouse / IgG1	1:200	
VIM	61013	Progen	mouse monoclonal /IgG2a kappa	1:50	
β-catenin	610153	BD Transduction LaboratoriesTM	Mouse IgG1	1:200	
EpCam (ESA)	NCL-ESA	Novocastra	mouse monoclonal	1:100	
E-cadherin	MS-9470-P0	Thermo Fisher Scientific	Mouse monoclonal	1:100	

Zeb1	NBP1-05987	Thermo Fisher Scientific	Rabbit IgG	1:50	
Muc18	ab75769	Abcam	Rabbit monoclonal	1:200	
anti-BrdU	ab6326	Abcam	Rat monoclonal	1:250	
alpha-SMA	ab5694	Abcam	Rabbit polyclonal	1:100	
Antibodies used for FACS analysis					
Anti-Human IL-13 R alpha 2	#FAB614F	R&D Systems	goat	1:50	
K7	M7018	DAKO	Mouse / IgG1	1:100	
Muc18	ab75769	Abcam	Rabbit monoclonal	1:200	
Ve-Cadherin	sc9989	Santa Cruz	mouse monoclonal	1:50	
CD68	M0814	DAKO	mouse monoclonal	1:25	
CD19	555415	BD Pharmingen	Mouse monoclonal-APC-conjugated	1:20	
CD3	M7254	DAKO	mouse monoclonal	1:30	
EpCam/ Ber-Ep4	M0804	DAKO	mouse monoclonal	1:20	

Adapted from (194).

Table 2-4. List of differentially expressed genes (DEGs) in the tumor

Available upon request.

Table 2-5. The most significantly differentially expressed genes between K7pos and K7neg subclones

	downreulated genes in K7neg sub-	**Log2 fold change***	**p-value (adjusted)**	**downregulated genes in K7pos subclones**	**Log2 fold change***	**p-value (adjusted)**

	clones					
1	LRRC61	-5.9	0.0000969	GALNT6	1	0.00062003
2	ACN9	-5.7	0.00062003	MPP4	1.1	0.0048732
3	C7orf29	-4.7	0.00364877	THSD7A	1.1	0.03266506
4	XKR6	-4.6	0.0000382	NRP1	1.1	0.00000077
5	LOC84856	-3.9	0.00830947	COL7A1	1.1	0.00557708
6	KRT7	-3.8	0.000048	FMNL3	1.1	0.00032817
7	FGA	-3.8	0.00032817	LEF1	1.2	0.00179461
8	LHFPL4	-2.9	0.00000000	NOV	1.2	0.00238779
9	C4BPA	-2.8	0.00315505	NOG	1.2	0.00993124
10	LRRN2	-2.8	0.04295868	EVI2A	1.2	0.00326579
11	COL21A1	-2.6	0.02951149	SFRP1	1.3	0.00000369
12	ZNF268	-2.5	0.00087405	FREM2	1.3	0.00676192
13	CXCL14	-2.2	0.0007305	IGDCC4	1.3	0.00393465
14	NID1	-2	0.0139443	IL24	1.5	0.00231568
15	AKR1B10	-2	0.00728353	CLDN3	1.6	0.00333562
16	UBD	-2	0.00167271	MNX1	1.6	0.00478243
17	OLFML3	-1.9	0.01147409	FAM26E	1.6	0.0000382
18	NKD2	-1.8	0.013417	ACSL5	1.7	0.00940214
19	TMEM233	-1.8	0.00174979	LOC149837	1.7	0.0049401
20	DAAM2	-1.8	0.00062003	VEPH1	1.7	0.0033163
21	RBP4	-1.7	0.000328	TRO	1.7	0.004655
22	C17orf107	-1.7	0.04002703	GJB3	1.8	0.00467597
23	EMILIN2	-1.7	0.03027566	DMBT1	1.9	0.02140321
24	ATOH8	-1.6	0.00112841	ADAM22	2.1	0.00000118
25	GSTM3	-1.6	0.0023611	MN1	2.2	0.00000025
26	SGK2	-1.5	0.01263831	HS3ST3A1	2.3	0.00465564
27	SGSM1	-1.5	0.0049401	IL13RA2	2.3	0.00032817
28	SECTM1	-1.5	0.00260028	INHBA	2.4	0.01788492
29	AKR1C3	-1.5	0.00084269	ZP3	2.5	0.00315505
30	EFCAB4B	-1.5	0.00018373	C20orf197	2.5	0.0020407
31	PTPN6	-1.4	0.00676192	NTM	2.6	0.00355414
32	RASL11A	-1.3	0.0000382	CPVL	3.3	0.00000000
33	CA12	-1.3	0.0000382	MGST1	3.5	0
34	MAPK13	-1.3	0.00062003	SSTR1	4.2	0.00063413
35	ADAP2	-1.3	0.00137982			
36	B3GALNT1	-1.2	0.00333562			
37	TNFSF10	-1.2	0,00279829			
38	PCSK9	-1.2	0,03150805			
39	SLC6A17	-1.2	0.00069297			
40	SIX2	-1.1	0.01911192			
41	WISP2	-1.1	0.03528742			
42	LOC644538	-1.1	0.02382562			
43	STK33	-1.1	0.00062003			
44	PPP2R2B	-1.0	0.01488498			

***** The logarithm (to basis 2) of the fold change between samples K7pos and K7neg. A negative log2 fold change designates that the gene is upregulated in K7pos subclones and

downregulated in K7neg subclones, respectively. The opposite is true for the positive log2 fold change.
Adapted from (194).

Table 2-6. Summary of immunohistochemical characterization of cholangiocarcinoma with sarcomatoid component and corresponding lymph node metastasis

Antigen	Carcinomatous component	Sarcomatoid component	Lymph node metastasis
Keratin-7	+++	negative	negative
Keratin-8+18	+++	negative	negative
Keratin-19	+	negative	negative
Vimentin	+	+++	+++
c-Met	+++	+++	++
Claudin 4	+++	negative	negative
Muc1	+++*	negative	negative
E-cadherin	+++	negative	negative
Beta catenin	+++	+	+++ (nuclear staining)
CEA	+++	negative	negative
CD44	+++	+++	negative
P53	++	++	+++
Ki-67	21.6±4.8	44.8±3.9	29.8±9.8
NCAM1	negative	negative	negative
CgA	negative	negative	n.a.
CD117	negative	negative	negative
CD34	negative	negative	negative
Nestin	+*	+	++
EpCam	++	negative	negative
Hep Par 1	negative	negative	n.a.
AFP	negative	negative	n.a.

Adapted from (194).

Table 2-7. Relative expression of representative EMT genes between Carc, Sarc and Met tumor compartments.

Marker	Function in EMT	Relative expression in tumor and cell culture subclones
TWIST1	Transcription factor involved in reprogramming epithelial gene expression	Up in Met
ZEB1		Up in Met and Sarc
ZEB2		Up in Met and Sarc
DDR2	Collagen receptor that is expressed during epithelial-mesenchymal transition	Up in Met
CDH2	Mesenchymal marker, upregulated in EMT	Up in Sarc and Met and Sarc*
EPCAM	Negative regulator of EMT	Down in Sarc and Met compared to Carc
S100A4	EMT mediator protein, target of Wnt/beta-catenin signaling pathway	Down in Sarc and Met
MUC1	Epithelial marker	Down in Sarc and Met
FOXA1	Involved as TF in EMT in lung adenocarcinoma and pancreatic cancer	Down in Met compartment and upregulated in K7neg clones
ID2	Antagonist of the TGF-β–induced EMT in epithelial cells	Down in Met
MTDH	new EMT marker in HCC	Up in Carc, Met
Lipocalin-(Lcn)-2	Negative modulator of EMT in HCC	Downregulated in Sarc and Met
Muc15	Downregulated in aggressive HCC	Downregulated in Sarc and Met compartment
EDIL3	New regulator of EMT in HCC	Upregulated in Sarc and Met compartment
TCF4	Part of the Wnt/b-catenin signaling pathway	UP in Met in Carc_Met gene list
HNF4a	A suppressor of Wnt/-b-catenin pathway	Up in Carc in Carc_Met and Carc_Sarc gene lists
CLDN3	A suppressor of Wnt/-b-catenin pathway	Up in Carc in Carc_Met and Carc_Sarc gene lists
MMP9	TGF target gene	UP in Met in (Carc_Met) UP in Sarc (Carc_Sarc) gene list

Criteria for gene selection was 2.0 fold change between the two tumor compartments with P value of <0.01 (adjusted). (*) These genes were differentially expressed with the fold changes < 2.0)

6.3 TruSeq RNA kit protocol

Poly A select mRNA

1. Mix 1 µg total RNA and H_20 to a final volume of **50 µl**.
2. Vortex **RNA Purification Beads** and add **50 µl** to RNA sample.
3. Pipet up and down 6 times to mix.
4. Incubate in thermocycler: 65˚ C – 5 minutes 4˚ C -- hold
5. When thermocycler reaches 4˚ C remove sample and place on bench at RT for 5 minutes.
6. Place sample in magnetic separator for 5 minutes.
7. Remove and discard all the supernatant.
8. Remove sample from rack.
9. Add **200 µl** of **Bead Washing Buffer** and pipet up and down 6 times to mix sample.
10. Place the sample back in the magnetic separator for 5 minutes.
11. Remove and discard all the supernatant.
12. Add **50 µl** of **Elution Buffer** and pipet up and down 6 times.
13. Incubate in thermocycler: 80˚ C – 2 minutes 25˚ C -- hold
14. Remove sample from thermocycler when it reaches 25˚ C and keep at room temp.
15. Add **50 µl** of **Bead Binding Buffer** and pipet up and down 6 times.
16. Incubate at RT for 5 minutes.
17. Place sample in magnetic separator for 5 minutes.
18. Remove and discard all supernatant.
19. Remove sample from rack.
20. Add **200 µl** of **Bead Washing Buffer** and pipet up and down 6 times.
21. Place sample in magnetic separator for 5 minutes.
22. Remove and discard all supernatant.
23. Add **19.5 µl Elute, Prime, Fragment Mix** and pipet up and down 6 times.
24. Incubate in thermocycler: 94˚ C – 8 min 4˚ C – hold
25. Place sample in magnetic rack for 5 minutes.
26. Transfer **17 µl** of the supernatant to a new 0.2 ml PCR tube. First Strand Synthesis
27. Add **8 µl First Strand Master Mix/Super Script II mix** to sample.

28. Incubate in thermocycler: 25˚ C – 10 minutes 42˚ C – 50 minutes 70˚ C – 15 minutes 4˚ C – hold.

Second Strand Synthesis

29. Add **25 µl** of **Second Strand Master Mix** to sample.

30. Incubate in thermocycler at 16˚ C for 1 hour.

31. Remove sample from thermocycler and let warm to RT.

32. Add **90 µl** of well--mixed **AMPure XP** beads.

33. Incubate at RT for 15 minutes.

34. Place on magnetic rack for 5 minutes.

35. Remove and discard **135 µl** of the supernatant.

36. Keep sample in magnetic rack and add 200 µl of 80 % ethanol.

37. Incubate for 30 seconds. Remove and discard all supernatant.

38. Repeat step 36 and 37 once more for a total of two washes.

39. Let beads dry at RT for 2 minutes.

40. Add **62.5 µl Resuspension Buffer** and pipet up and down 10 times.

41. Incubate at RT for 2 minutes.

42. Place in magnetic rack for 5 minutes.

43. Transfer 60 µl of the supernatant to a new 0.2 ml PCR tube. Perform End Repair

44. Add **40 µl** of **End Repair Mix** to sample.

45. Incubate at 30˚C for 30 min.

46. Add **160 µl** of well--mixed **Ampure XP Beads**.

47. Incubate at RT for 15 minutes.

48. Place on magnetic rack for 5 minutes.

49. Remove and discard **127.5 µl** of the supernatant.

50. Repeat step 49 one more time.

51. Keep sample in magnetic rack and add 200 µl of 80 % ethanol.

52. Incubate for 30 seconds. Remove and discard all supernatant.

53. Repeat step 51 and 52 once more for a total of two washes.

54. Let beads dry at RT for 2 minutes.

55. Add **20 µl Resuspension Buffer** and pipet up and down 10 times.

56. Incubate at RT for 2 minutes.

57. Place in magnetic rack for 5 minutes.

58. Transfer **17.5 µl** of the supernatant to a new 0.2 ml PCR tube. Add 'A' bases to 3' ends

59. Add **12.5 µl** **A--Tailing Mix** to sample.

60. Incubate at 37˚ C for 30 minutes. Ligate Adapters to DNA fragments

61. Add: **2.5 µl** **DNA Ligase Mix**

2.5 µl **Resuspension Buffer**

2.5 µl **RNA Adapter Index**

62. Incubate at 30˚ C for 10 minutes

63. Add **5 µl** **Stop Ligase Mix**

64. Add **42 µl** of well--mixed **AMPure XP beads**.

65. Incubate at RT for 15 minutes.

66. Place on magnetic rack for at least 5 minutes.

67. Remove and discard **79.5 µl** of the supernatant.

68. Keep sample in magnetic rack and add 200 µl of 80 % ethanol.

69. Incubate for 30 seconds. Remove and discard all supernatant.

70. Repeat steps 68 and 69 one more time.

71. Let the beads dry at RT for 2 minutes.

72. Add **52.5 µl** **Resuspension Buffer** and pipet up and down 10 times.

73. Incubate at RT for 2 minutes.

74. Place in magnetic rack for 5 minutes.

75. Transfer **50 µl** of the supernatant to a new 0.2 ml PCR tube.

76. Add **50 µl** of wellmixed **AMPure XP beads**.

77. Incubate at RT for 15 minutes.

78. Place on magnetic rack for at least 5 minutes.

79. Remove and discard **95 µl** of the supernatant.

80. Keep sample in magnetic rack and add 200 µl of 80 % ethanol.

81. Incubate for 30 seconds. Remove and discard all supernatant.

82. Repeat steps 80 and 81 one more time.

83. Let the beads dry at RT for 2 minutes.

84. Add **22.5** µl **Resuspension Buffer** and pipet up and down 10 times.

85. Incubate at RT for 2 minutes.

86. Place in magnetic rack for 5 minutes

87. Transfer **20 µl** of the supernatant to a new 0.2 ml PCR tube. Amplify Library by PCR

88. Add: **5 µl <u>PCR Primer Cocktail</u> 25 µl <u>PCR Master Mix</u>**

89. Amplify with the following PCR protocol:

a. 98˚ C for 30 seconds

b. 5 to 18 cycles of (see note #1): 98˚ C for 10 seconds 60˚ C for 30 seconds 72˚ C for 30 seconds

c. 72˚ C for 5 minutes

d. Hold at 4˚ C

90. Add **50 µl** of well--mixed **<u>AMPure XP</u>** beads.

91. Incubate at RT for 15 minutes.

92. Place on magnetic rack for at least 5 minutes

93. Remove and discard **95 µl** of the supernatant.

94. Keep sample in magnetic rack and add 200 µl of 80 % ethanol.

95. Incubate for 30 seconds. Remove and discard all supernatant.

96. Repeat steps 94 and 95 one more time.

97. Let the beads dry at RT for 2 minutes.

98. Add **32.5 µl <u>Resuspension Buffer</u>** and pipet up and down 10 times.

99. Incubate at RT for 2 minutes.

100. Place in magnetic rack for 5 minutes.

101. Transfer **30 µl** of the supernatant to a new 1.5 ml PCR tube.

Notes:

1. Before PCR amplification take 1 µl of you adapter--ligated DNA and mix it with 1 µl of TruSeq PCR Quantification Primer Mix (25 µM), 8 µl H2O and 10 µl 2X Kapa SYBR FAST Master Mix. Run on qPCR machine with same perimeters as used for library amplification with 25 cycles. Determine the number of cycles it took to get to 50 % of amplication (late log phase) and use that number of cycles minus X. If you need more than 18 cycles you do not have enough DNA and unfortunately you need to do the experiment again. I don't now what X equals yet. For the Kapa HF polymerase it was 2. In my first try with the kit using 1 µg of HeLa cell RNA, the 15 cycles recommended by Illumina was way too many.

2. I the future I think I will divide all the reaction volumes (and starting material) in half to double the number of libraries I get out of the kit.

3. **PCR Primer Quantification Mix**

Resuspend TruSeq PCR Primer 1 and TruSeq PCR Primer 2 in TE/10 (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) to a concentration of 100 μM.

In new tube mix: 25 μl TruSeq PCR Primer 1 (100 μM), 25 μl TruSeq PCR Primer 2 (100 μM), 50 μl H_2O

TruSeq PCR 1 AATGATACGGCGACCACCGA*G

TruSeq PCR 2 CAAGCAGAAGACGGCATACGA*G

* = phosphorothioate bond

6.4 Bisulfite PCR library preparation using TruSeq (Illumina) enzymes and adapters from IDT (Integrated DNA Technologies, Inc)

1. Mix: 50 µl purified PCR product
2. Transfer to Covaris microTube.
3. Sonicate in Covaris S2 with the following settings:

Time	155 seconds
Duty Factor	20 %
Peak Incident Power (W)	175
Cycles per Burst	200
Temperature	4˚C
Power mode Frequency	Sweeping
Degassing mode	Continuous
AFA Intensifier	Yes
Water Level	12

4. Transfer DNA from microTube to 0.2 ml tube.
5. Add 100 µl of well-mixed Ampure XP beads.
6. Mix by pipetting up and down several times.
7. Incubate at RT for 10 minutes
8. Place on magnetic rack for 5 minutes.
9. Remove and discard the supernatant. When removing supernatant do so very slowly with pipetman being careful not to take any beads.
10. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.
11. Incubate for 30 seconds. Remove and discard ALL the supernatant.
12. Repeat steps 37 and 38 one more time.
13. Add 16 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.
14. Incubate at RT for 2 minutes.
15. Place in magnetic rack for 5 minutes.
16. Transfer 15 µl of the supernatant to a 0.2 ml PCR tube.

End Repair

17. Add 10 µl TruSeq End Repair Mix

18. Incubate at 30˚C for 30 minutes.
19. Add 50 µl of well-mixed Ampure XP beads.
20. Mix by pipetting up and down several times.
21. Incubate at RT for 10 minutes.
22. Place on magnetic rack for 5 minutes.
23. Remove and discard the supernatant. When removing supernatant do so very slowly with pipetman being careful not to take any beads.
24. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.
25. Incubate for 30 seconds. Remove and discard ALL the supernatant.
26. Repeat steps 37 and 38 one more time.
27. Add 5.375 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.
28. Incubate at RT for 2 minutes.
29. Place in magnetic rack for 5 minutes.
30. Transfer 4.375 µl of the supernatant to a 0.2 ml PCR tube.

A-tailing

31. Add 3.125 µl TruSeq A-Tailing Mix
32. Incubate at 37˚C for 30 minutes.

Ligate Adapters

33. Add: 0.625 µl TruSeq Ligation Mix and 1.25 µl Adapter Index (10 µM)
34. Immediately mix by pipetting up and down with P20 set to 10 µl.
35. Incubate at 30˚C for 10 minutes.
36. Add 1.25 µl TruSeq Stop Ligation Buffer
37. Add 15 µl of well-mixed Ampure XP beads.
38. Mix by pipetting up and down several times.
39. Incubate at RT for 10 minutes.
40. Place on magnetic rack for 5 minutes.
41. Remove and discard the supernatant. When removing supernatant do so very slowly with pipette being careful not to take any beads.
42. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.
43. Incubate for 30 seconds. Remove and discard ALL the supernatant.
44. Repeat steps 37 and 38 one more time.
45. Add 6.25 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.

46. Incubate at RT for 2 minutes.

47. Place in magnetic rack for 5 minutes.

48. Transfer 5.25 µl of the supernatant to a 0.2 ml PCR tube (use one 1 µl for cycle determination).

PCR amplify library

49. Mix: 1 µl PCR Primer Mix (25 µM)

6.25 µl TruSeq PCR Master Mix

50. Incubate in thermal cycler with the blow program:

1 cycle of

98˚C – 30 seconds

4 cycles of

98˚C – 10 seconds

63˚C – 30 seconds

72˚C – 30 seconds

1 cycle of

72˚C – 5 minutes

51. Add 17 µl of well-mixed Ampure XP beads.

52. Mix by pipetting up and down several times.

53. Incubate at RT for 10 minutes.

54. Place on magnetic rack for 5 minutes.

55. Remove and discard the supernatant. When removing supernatant do so very slowly with pipette being careful not to take any beads.

56. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.

57. Incubate for 30 seconds. Remove and discard ALL the supernatant.

58. Repeat steps 37 and 38 one more time.

59. Add 21 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.

60. Incubate at RT for 2 minutes.

61. Place in magnetic rack for 5 minutes.

62. Transfer 20 µl of the supernatant to a 0.2 ml PCR tube.

6.5 MeDIP-seq protocol

Prepare Genomic DNA

1. Harvest cells (<5x106 cells) by tyrpsinization.
2. Spin down at 3,000 rpm for 5 minutes.
3. Remove growth media.
4. Resuspend cell pellet in 200 µl PBS.
5. Add 20 µl PureLink Proteinase K.
6. Add 20 µl PureLink RNase A.
7. Mix well by brief vortexing.
8. Incubate at RT for 2 minutes.
9. Add 200 µl PureLink Genomic Lysis/Binding Buffer and mix well by vortexing.
10. Incubate at 55˚C for 10 minutes.
11. Add 200 µl 100 % ethanol. Mix well by vortexing for 5 seconds.
12. Transfer cell lysate to PureLink Spin Column
13. Centrifuge at 10,000 x g for 1 minutes.
14. Transfer column to new collection tube.
15. Add 500 µl PureLink Wash Buffer 1.
16. Centrifuge at 10,000 x g for 1 minutes.
17. Transfer column to new collection tube.
18. Add 500 µl PureLink Wash Buffer 2
19. Centrifuge at maximum speed for 3 minutes.
20. Place column in new 1.5 ml tube and add 70 µl PureLink Genomic elution buffer.
21. Incubate at RT for 1 minute.
22. Centrifuge at maximum speed for 1 minute.
23. Add an additional 70 µl PureLink Genomic elution buffer.
24. Incubate at RT for 1 minute.
25. Centrifuge at maximum speed for 1 minute.
26. Transfer DNA to new 1.5 ml tube.
27. Measure DNA concentration with NanoDrop.

Shear DNA with Covaris

28. Transfer <4 µg DNA to Covaris microTube with AFA fiber.
29. Add H2O to a final volume of 130 µl.

30. Sonicate in Covaris S2 with the following settings: Time 7 minutes
Duty Cycle 10 %
Intensity 5
Cycles per Burst 200
Temperature 4˚C Power mode Frequency Sweeping
Degassing mode Continuous
AFA Intensifier Yes
Water Level 14

31. Transfer DNA from microTube to 1.5 ml tube.
32. 130 µl of well--mixed Ampure XP beads and 130 µl 30 % PEG8000
1.25 M NaCl
33. Mix by pipetting up and down several times (See Note 9).
34. Incubate at RT for 15 minutes.
35. Place on magnetic rack for 5 minutes.
36. Remove and discard the supernatant. When removing supernatant do so very slowly with pipetman being careful not to take any beads.
37. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.
38. Incubate for 30 seconds. Remove and discard ALL the supernatant.
39. Repeat steps 37 and 38 one more time.
40. Let the beads dry at RT for 2 minutes.
41. Add 42.5 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.
42. Incubate at RT for 2 minutes.
43. Place in magnetic rack for 5 minutes.
44. Transfer 50 µl of the supernatant to a 0.2 ml PCR tube.
45. Quantitate DNA with nanoDrop.
46. Remove 2 µl and add 39 µl H2O to new tube for input control
Prepare SssI CpG methylated DNA control
47. Mix: 1 µl 32 mM SAM (see Note 10) 19 µl H20
48. Mix: 10 µl 10X NEB buffer #2
10 µl 1.6 mM SAM (freshly diluted from step 46) 7 µl DNA from step #45
5 µl SssI methyltransferase (add last) H2O to a final volume of 100 µl
49. Incubate at 37˚C for 1 hour.

50. 100 µl of well--mixed Ampure XP beads and 100 µl 30 % PEG8000

1.25 M NaCl

51. Mix by pipetting up and down several times (See Note 9).

52. Incubate at RT for 15 minutes.

53. Place on magnetic rack for 5 minutes

54. Remove and discard the supernatant. When removing supernatant do so very slowly with pipetman being careful not to take any beads.

55. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.

56. Incubate for 30 seconds. Remove and discard ALL the supernatant.

57. Repeat steps 37 and 38 one more time.

58. Let the beads dry at RT for 2 minutes.

59. Add 42.5 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.

60. Incubate at RT for 2 minutes.

61. Place in magnetic rack for 5 minutes.

62. Transfer 41 µl of the supernatant to a 0.2 ml PCR tube.

Blunt DNA Ends (End Repair)

63. Perform the following steps of the protocol with three sets of reactions:

a) Fully CpG methylated DNA control

b) Input DNA

c) Sheared genomic DNA for MeDIP

64. Mix: 41 µl genomic DNA from step #45, 41 µl Input DNA from step #46 or 41 µl SssI methylated DNA from step #61.

5 µl NEB T4 DNA ligase buffer 2 µl 10 mM dNTPs

2 µl End Repair Enzyme Mix H2O to a final volume of 50 µl

65. Incubate at 20˚ C for 30 minutes in PCR machine

66. 50 µl of well--mixed Ampure XP beads and 50 µl 20 % PEG8000 1.25 M NaCl

67. Mix by pipetting up and down several times.

68. Incubate at RT for 15 minutes.

69. Place on magnetic rack for 5 minutes

70. Remove and discard the supernatant. When removing supernatant do so very slowly with pipetman being careful not to take any beads.

71. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.

72. Incubate for 30 seconds. Remove and discard ALL the supernatant.

73. Repeat steps 53 and 54 one more time.

74. Let the beads dry at RT for 2 minutes.

75. Add 17.5 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.

76. Incubate at RT for 2 minutes.

77. Place in magnetic rack for 5 minutes.

78. Transfer 16 µl of the supernatant to a 0.2 ml PCR tube.Transfer eluted DNA to new 1.5 ml tube

Add 'A' Bases to the 3' Ends of the DNA fragments

79. Mix: 16 µl End--repaired DNA (from step 60) 2 µl 10X NEB Buffer #2

1 µl 4 mM dATP

1 µl 5 U/µl Klenow 3' to 5' exo minus

80. Incubate at 37˚ C for 30 minutes

Ligate Adapters to DNA fragments

81. Mix: 20 µl 'A'--tailed DNA (from step 62) 25 µl 2x Quick Ligase Buffer (NEB)

2 µl Annealed TruSeq Adapters (50 µM for genomic DNA – 5 µM for fully methylated DNA control and input control)

2 µl Quick Ligase (2,000 U/µl)

82. Incubate at room temp for 15 minutes.

83. Add 5 µl 0.5 M EDTA

84. Add 50 µl of well--mixed Ampure XP beads and 50 µl 20 % PEG8000

1.25 M NaCl

85. Mix by pipetting up and down several times.

86. Incubate at RT for 15 minutes.

87. Place on magnetic rack for 5 minutes.

88. Remove and discard the supernatant. When removing supernatant do so very slowly with pipetman being careful not to take any beads.

89. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.

90. Incubate for 30 seconds. Remove and discard ALL the supernatant.

91. Repeat steps 53 and 54 one more time.

92. Let the beads dry at RT for 2 minutes.

93. Add 27 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.

94. Incubate at RT for 2 minutes.

95. Place in magnetic rack for 5 minutes.

96. Transfer 25 µl of the supernatant to a 0.2 ml PCR tube.

97. Transfer 2.5 µl of DNA to a new 1.5 ml tube, add 47.5 µl Proteinase K Digestion Buffer and put aside for input control.

Perform anti--5--methylcytosine immunoprecipitation

98. Add 20 µl MeDIP-seq blocking oligo (100 pmols/µl) to adapter ligated DNA from step 96.

99. Denature DNA for 10 minutes at 95˚ C in thermocycler.

100. Transfer tubes directly from 95˚ C to ice water bath.

101. Keep on ice for 10 minutes.

102. Mix: 42.5 µl denatured adapter--ligated DNA

0.5 µl 1 µg/µl anti--5meC antibody (33D3 Diagenode) 500 µl MeDIP Wash Buffer (ice cold)

103. Incubate overnight at 4˚C.

104. Equilibrate beads. Prepare one 1.5 ml tube for each for each immunoprecipitation with 1 ml 1x IP Buffer and 10 µl of Sheep Anti--mouse M--280 Dynabeads to each tube. Note: The beads are usually a significant source of background and the minimum amount should be used. Invitrogen states that the binding capacity of the beads is 6.5--65 ng IgG/µl beads (yeah, that's a huge spread but that is what the product brochure obliquely says).

105. Transfer IP to the tube with the Dynabeads.

106. Incubate 2 hours at 4˚ C on tube rotator.

107. Spin briefly in microfuge at 3k RPM. Place in magnetic rack and remove liquid.

108. Add 1 ml IP Buffer, resususpend beads, spin briefly at 3.5k rpm, place in magnetic rack, and remove supernatant.

109. Repeat step 94 five more times for a total of six washes and change tubes two times during washes.

110. Resuspend beads in: 48 µl Proteinase K digestion buffer

2 µl 10 mg/ml Proteinase K

111. Process input and IP samples in parallel in the following steps.

112. Incubate 30 minutes at 50˚ C.

113. Place tubes in magnetic rack for 3 minutes.

114. Transfer supernatant to new 0.2 ml PCR tube.

115. Add 50 µl of well--mixed Ampure XP beads and 50 µl 20 % PEG8000

1.25 M NaCl

116. Mix by pipetting up and down several times.

117. Incubate at RT for 15 minutes.

118. Place on magnetic rack for 5 minutes.

119. Remove and discard the supernatant. When removing supernatant do so very slowly with pipetman being careful not to take any beads.

120. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.

121. Incubate for 30 seconds. Remove and discard ALL the supernatant.

122. Repeat steps 102 and 103 one more time.

123. Let the beads dry at RT for 2 minutes.

124. Add 15.5 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.

125. Incubate at RT for 2 minutes.

126. Place in magnetic rack for 5 minutes.

127. Transfer 14 µl of the supernatant to a 0.2 ml PCR tube.

Convert ssDNA to dsDNA (see note #2).

128. On ice mix 14 µl Adapter--ligated DNA Input DNA (from step 83) or 14 µl immunoprecipitated DNA (from step 111) 1 µl TruSeq PCR primer cocktail (25 µM) 15 µl 2X Kapa HiFi HotStart Ready Mix

129. Amplify with the following PCR protocol.

a. 45 sec at 98˚ C

b. 4 cycles of:

15 sec at 98˚ C 30 sec at 63˚ C 30 sec at 72˚ C

c. 1 minutes at 72˚ C

d. Hold at 4˚ C

130. Add 30 µl of well--mixed Ampure XP beads and 30 µl 20 % PEG8000 1.25 M NaCl.

131. Mix by pipetting up and down several times (See Note 9).

132. Incubate at RT for 15 minutes.

133. Place on magnetic rack for 5 minutes.

134. Remove and discard the supernatant. When removing supernatant do so very slowly with pipetman being careful not to take any beads.

135. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.

136. Incubate for 30 seconds. Remove and discard ALL the supernatant.

137. Repeat steps 121 and 122 one more time.

138. Let the beads dry at RT for 2 minutes.

139. Add 15.5 µl TE/10 and pipet up and down several times until pellet beads are completely resuspended.

140. Incubate at RT for 2 minutes. 141. Place in magnetic rack for 5 minutes.

142. Transfer 14 µl of the supernatant to a 0.2 ml PCR tube.

Size Select Library on Agarose Gel

143. Prepare a 2.5 % MetaPhor/SeaKem LE (3:1 ratio) agarose 1X TAE gel with ethidium bromide.

144. Load 200 ng of 100 bp ladder in two lanes at either side of gel.

145. Add 4 µl 5X loading dye (containing xylene cyanol and bromophenol blue) to sample.

146. Load sample(s) on gel with at least one lane between different samples and markers to avoid cross contamination.

147. Run gel at 120 V for 35 minutes.

148. Cut out slice of gel with clean scalpel that contains material between 200 and 400 bp and place in 1.5 ml tube (see note #3).

149. Determine volume of gel slice by zeroing scale with empty tube then weighing the tube with your gel slice (1 mg = 1 µl).

150. Add 5 volumes of Qiagen Buffer QG (e.g. if you gel slice is 0.2 g, add 1 ml of QG)

151. Incubate at RT until gel slice has completely dissolved. Mix continuously by hand or in tube rotator.

152. Add 1 volume isopropanol and mix by inverting several times.

153. Apply 650 µl to Qiagen MinElute column.

154. Spin for 30 seconds in microfuge. 155.Remove liquid from collection.

156. Repeat steps 139 to 141 until you have passed entire sample through column

157. Apply 500 µl Qiagen Buffer QG to column.

158. Spin in microfuge for 1 minutes.

159. Remove liquid from collection tube.
160. Apply 750 µl of Qiagen buffer PE to column.
161. Let stand for 2 minutes.
162. Spin in microfuge for 1 minutes.
163. Remove liquid from collection tube.
164. Spin in microfuge for 1 minutes.
165. Transfer to new 1.5 ml tube.
166. Add 15 µl Qiagen buffer EB.
167. Let stand for 5 minutes.
168. Spin in microfuge for 1 minute.
169. Transfer eluted DNA to 0.2 ml PCR tube.

Amplify Library by PCR

170. On ice mix : 14 µl Size--selected DNA (from Step 154).

1 µl TruSeq PCR primer cocktail (25 µM) 15 µl 2X Kapa HiFi HotStart Ready Mix.

171. Amplify with the following PCR protocol

a. 45 sec at 98˚ C
b. 5 to 14 cycles of (see note #4): 15 sec at 98˚ C 30 sec at 63˚ C 30 sec at 72˚ C.
c. 1 minute at 72˚ C
d. Hold at 4˚ C

172. Add 54 µl of well--mixed AMPure XP beads.
173. Incubate at RT for 15 minute with occasional vortexing.
174. Spin tube at 4,000 rpm for 5 seconds.
175. Place on magnetic rack for at least 5 minutes.
176. Remove and discard 105 µl of the supernatant.
177. Keep sample in magnetic rack and add 200 µl of freshly prepared 80 % ethanol.
178. Incubate for 30 seconds. Remove and discard all supernatant.
179. Repeat steps 163 and 164 one more time.
180. Let the beads dry at RT for 2 minutes.
181. Add 12.5 µl Resuspension Buffer and pipet up and down 10 times.
182. Incubate at RT for 2 minutes.
183. Place in magnetic rack for 5 minutes.
184. Transfer 10 µl of the supernatant to a new 1.5 ml tube.

Analyze library by running 1 µl on Bioanalyzer (see note #5) Quantitate Library with Kapa Biosystems Library Quantification Kit

Notes:

1. Some sequencing experiments require the use of fewer than 12 index sequences in a lane with a high cluster density. In such cases, select indexes carefully to ensure optimum base calling and demultiplexing by having different bases at each cycle of the index read. Illumina recommends the following sets of indexes for low--level pooling experiments. However, I have stopped following this advice.

Pool of 2 samples:

- Index #6 GCCAAT • Index #12 CTTGTA

Pool of 3 samples:

- Index #4 TGACCA • Index #6 GCCAAT • Index #12 CTTGTA

Pool of 6 samples: • Index #2 CGATGT • Index #4 TGACCA • Index #5 ACAGTG • Index #6 GCCAAT • Index #7 CAGATC • Index #12 CTTGTA

2. This step was added because the self--ligated Y--shaped adapter dimers appear to run very slow on a 2.5 % agarose gel. I cut above 250 bp and they were still in the sample. This also brings into question where exactly your sample ligated to the Y--shapped adapters is running. Furthermore, the relative migration of funny--shaped DNA molecules changes with the percentage of the gel, type of agarose, etc. So to know exactly where everything is running, I assumed it would be better to convert the Y--shaped DNA to double--stranded DNA. Everything should be converted to double--stranded DNA after two PCR cycles, but since the recovery of DNA during the gel extraction is low, I thought it would be a good idea to add a couple extra cycles of PCR. Theoretically you could do the entire PCR enrichment at this step but then you would have to worry about some of the sample running too high because of the formation of ‘bubbles’ and ‘daisy chains’ when primers become limiting.

3. The double--stranded adapters add 121 bp to the DNA fragments. Calculate the size of your DNA fragments by running a portion of the sheared DNA from step 46 on a 2.2 % agarose gel. Add 121 bp to that number and cut a gel slice containing your adaptor ligated DNA. Also be sure that you gel slice is well above 121 bp as that is where the self--ligated adapters run.

4. Before PCR amplification take 1 µl of you size--selected DNA and mix it with 1 µl of TruSeq PCR Primer Cocktail, 8 µl H2O and 10 µl 2X Kapa SYBR FAST Master Mix. Run on qPCR machine with same perimeters as used for library amplification with 25 cycles. Determine the number of cycles it took to get to 50 % of amplication (late log phase) and use that number of cycles minus 2.

5. The image from the Bioanalyzer should be a smear the same size as the band you cut out in the size selection step. If you have larger products, this is most likely due to over--amplification. When primers become limiting you get hybrid DNA molecules resulting from the adaptors ends of DNA molecules annealing. This creates bubble shaped molecules and daisy chains of molecules. Contrary to what is stated by Illumina, this should not be a problem, as your library will be denatured before attachment to the flow cell. On the other hand, if you have a sharp band around 121 bp, this is from self--ligated adaptors and is a problem.

6. All enzymes are purchased from NEB (except the Kapa polymerase).

7. The Kapa polymerase requires fewer cycles to achieve the same amplification as with Phusion polymerase.

8. Possible improvements:

b. Kapa makes their library amplifaction PCR mix with SYBR green. Perhaps you can monitor the library amplification in real--time and remove it from the machine as it becomes amplified but before it is over--amplified.

9. AMPure XP guidelines:

a) Make sure that you achieve a homogenous solution of the beads and PEG by pipetting up and down a sufficient number of times.

b) When removing the supernatant after the binding step, remove all but 5 µl with a P200 pipetman. Then go back, using a P20, and carefully remove as much of the last 5 µl as you can without taking any beads.

c) When washing with 80 % ethanol the beads stick to the wall of the tube better so you can remove the supernatant more quickly. However, I make sure I remove all traces of ethanol by going through each tube a second time with a new pipet tip.

d) Illumina recommends drying the beads, not only is this a waste of time but it probably reduces yield.

e) At the elution step, when transferring the eluted DNA to a new tube, it is important not to accidentally carryover any beads. I carefully look inside the tube as I am pipetting up the eluted DNA to make sure I don't accidentally take up and beads. Then I look at the eluted DNA in the pipet tip to see if I can see any beads. If there are, simply pipet the breads back into the tube you took them from and wait a few minutes for the sample to separate again in the magnetic rack. At this step it is important to leave at lease 1 µl behind, as it is not possible to remove all the liquid and not take any beads.

10. Three significant changes were made in this protocol from existing published protocols:

a. One issue with MeDIP-seq

seq is that it requires the DNA to be single stranded, which is a potential problem since the Illumina adapters have base-- pairing sequences. Thus, potentially every single stranded DNA molecule that is immunoprecipitated will bring down additional DNA molecules base--paired in the adapter region. This would substantially increase background. To circumvent this potential problem I add in a 50--fold excess of a 16 bp oligo that binds to the homologous region in the adapter, in order to block the adapter ligated DNA molecules from base pairing. The low melting temperature of the blocking oligo should prevent it from priming during subsequent PCR steps.

b. Four cycles of PCR are added before size selection by agarose gel electrophoresis, to convert the Y--shaped Illumina adapters into double stranded DNA. I have found that the Y--shaped adapters run much higher then the corresponding length of double--stranded DNA, thus making the identification of the correct size to excise from the get difficult.

c. If there was a third thing, I forget what it was.

Adapter Annealing Protocol

1. Resuspend adapter oligonucleotides at a concentration of 100 μM in:

10 mM Tris--HCl, pH 7.8

0.1 mM EDTA, pH 8.0 50 mM NaCl

2. Mix 25 μl of the TruSeq Universal Adapter with 25 μl of the indexed adapter.

3. Anneal oligos a thermal cycler with the following program:

a. 2 minutes at 95˚ C

b. 140 cycles of

30 sec at 95˚C (decrease temp 0.5˚ C every cycle)

c. Hold at 4˚ C

4. Dilute annealed adaptors 1:200 in H2O

PCR Primer Cocktail

1. Resuspend TruSeq PCR Primer 1 and TruSeq PCR Primer 2 in TE/10 (10 mM Tris--HCl, pH 8.0, 0.1 mM EDTA) to a concentration of 100 μM.

2. In new tube mix: 25 μl TruSeq PCR Primer 1 (100 μM)

25 μl TruSeq PCR Primer 2 (100 μM)

50 μl H2O

Oligonucleotide Sequences

* = phosphorothioate bond TruSeq Universal

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATC*T

TruSeq Adaptor, Index 2

/5Phos/

GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGT CTTCTGCTT*G

TruSeq Adaptor, Index 4

/5Phos/
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGT
CTTCTGCTT*G

TruSeq Adaptor, Index 5
/5Phos/
GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCCGT
CTTCTGCTT*G

TruSeq Adaptor, Index 6
/5Phos/
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGT
CTTCTGCTT*G

TruSeq Adaptor, Index 7
/5Phos/
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCGT
CTTCTGCTT*G

TruSeq Adaptor, Index 12
/5Phos/
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTGTAATCTCGTATGCCGT
CTTCTGCTT*G

TruSeq PCR 1 AATGATACGGCGACCACCGA*G

TruSeq PCR 2 CAAGCAGAAGACGGCATACGA*G

Note: The adaptor sequences are from Illumina. The PCR primer sequences are my best guess as to what Illumina is using in their PCR Primer Cocktail.

MeDIP-seq Blocking Oligo AGATCGGAAGAGCGTC

IMPORTANT!!!: The current batches of adapters from Illumina are methlylated so you absolutely cannot use them. You must have your own oligos synthesized.

End Repair Enzyme Mix
30 µl 3 U/µl T4 DNA polymerase 6 µl 1 U/µl Klenow Fragment
30 µl 10 U/µl T4 DNA Polynucleotide Kinase

MeDIP Wash Buffer
PBS
0.05 % Triton X--100
1 mM EDTA
Proteinase K digestion buffer
20 mM HEPES, pH 7.9
1 mM EDTA
0.5 % SDS
TE/10
10 mM Tris--HCl, pH 8.0
0.1 mM EDTA